Dottorato di ricerca in Digital Humanities

Università degli Studi di Genova

Ciclo XXXVIII

TITOLO TESI

Terminologia e ontologie per la storia dell'informatica: modelli linguistici e rappresentazioni della conoscenza per il patrimonio culturale digitale

Curriculum: LINGUE, CULTURE E TECNOLOGIE DIGITALI

Presentata da: Giada D'Ippolito

Tutor: Prof.ssa Micaela Rossi (Università di Genova), Prof.ssa Caroline Djambian (Università Grenoble Alpes, laboratorio GRESEC)

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani

Università di Genova

Borsa di dottorato cofinanziata con risorse dell'Unione europea-*NextGeneration EU* Piano Nazionale di Ripresa e Resilienza Missione 4, componente 1 "*Potenziamento dell'offerta dei servizi di istruzione: dagli asili nido all'Università*

Indice

Abstract

. La tesi affronta il problema della rappresentazione e della valorizzazione del patrimonio informatico attraverso l'integrazione di approcci terminologici, ontologici e corpus-based. Gli oggetti dell'informatica storica presentano infatti una forte complessità concettuale e una marcata variabilità terminologica, dovuta all'evoluzione tecnologica, ai diversi contesti d'uso e alle pratiche discorsive delle comunità scientifiche, tecniche e museali.

Il lavoro vuole esplorare, dunque, il potenziale del museo come laboratorio di risemantizzazione, capace di riscattare l'oggetto informatico dalla sua condizione di reperto muto attraverso la conversione di reperti tecnologici statici in dispositivi narrativi dinamici.

Il lavoro si colloca nel quadro teorico della terminologia, con particolare attenzione agli approcci descrittivi, sociocognitivi e basati su corpora, e dialoga con i modelli di rappresentazione della conoscenza propri del Web Semantico. A partire da corpora differenziati – accademici, storici, comunitari e museali – viene condotta un'analisi terminologica multidimensionale mediante strumenti di analisi testuale, statistica e distribuzionale grazie a strumenti come IRaMuTeQ, TermoStat, Tropes, Sketch Engine, al fine di osservare la variazione diacronica, diafasica, così come le dinamiche discorsive delle designazioni dei concetti del patrimonio IT.

Il caso di studio del progetto europeo ITinHeritage costituisce il contesto applicativo della ricerca. In tale ambito, la tesi documenta la costruzione di una risorsa terminologica e ontoterminologica orientata al patrimonio culturale, che integra dati linguistici, metadati museali e modelli concettuali formalizzati. Il contributo della ricerca risiede nella proposta di un modello metodologico che mette in relazione analisi linguistica, organizzazione semantica e modellizzazione ontologica, evidenziando il ruolo centrale della terminologia come interfaccia tra testi, concetti e oggetti patrimoniali. L'approccio ontoterminologico qui adottato non si limita alla mera catalogazione degli artefatti, ma mira a colmare lo scarto semantico tra la documentazione tecnica e la percezione del pubblico non esperto. In questo senso, la terminologia funge da interfaccia critica, capace di rendere intelligibile un patrimonio potenzialmente «opaco» come quello informatico, trasformando l'oggetto museale in un nodo di conoscenza dinamico e interconnesso.

Premessa

1. L’informatica come patrimonio: la sfida dell’opacità semantica

Nel contesto contemporaneo della ricerca scientifica e delle pratiche di valorizzazione del patrimonio culturale, la rappresentazione della conoscenza svolge un ruolo cruciale nel rendere accessibili, interpretabili e riutilizzabili saperi complessi. Tale esigenza si manifesta in modo particolarmente evidente quando l’oggetto di studio è costituito dal patrimonio informatico, un ambito che, pur avendo inciso profondamente sulla storia sociale, scientifica e culturale del XX e del XXI secolo, continua a presentare rilevanti difficoltà di documentazione, interpretazione e mediazione. In particolare, il patrimonio informatico pone sfide specifiche, legate non solo alla rapida obsolescenza tecnologica degli oggetti, ma anche alla difficoltà di renderne intelligibili i principi concettuali, le funzioni e le trasformazioni storiche.

Il punto di partenza di questa ricerca risiede nella consapevolezza che l’informatica non rappresenta più soltanto un insieme di strumenti funzionali, ma si è configurata come un pilastro fondamentale del patrimonio culturale contemporaneo.

Tuttavia, a differenza di altre forme di patrimonio materiale o tecnico-scientifico, l’artefatto informatico presenta una sfida epistemologica unica: la sua natura di “scatola nera”. Gli artefatti informatici risultano spesso opachi sul piano semantico: il loro funzionamento, le loro categorie e persino le loro denominazioni non sono immediatamente evidenti e variano sensibilmente nel tempo, nei contesti disciplinari e nelle diverse comunità di riferimento.

Hardware e software, una volta separati dal loro contesto d’uso, dai sistemi operativi, dalle interfacce e dalle pratiche che ne hanno determinato il funzionamento, perdono rapidamente la capacità di “parlare” di sé. Nello specifico, ad esempio, un calcolatore storico, una volta spento o privato del software originale, perde quasi istantaneamente la sua capacità di comunicare la propria funzione e il proprio contesto d'uso.

In quest’ottica, la ricerca indaga le strategie attraverso cui le istituzioni museali possono governare tale complessità: il museo evolve da mero deposito di artefatti a centro di mediazione culturale, dove l’adozione di modelli semantici avanzati diventa condizione necessaria per contrastare l’obsolescenza non solo fisica, ma soprattutto concettuale, dei propri beni.

Il valore culturale dell’oggetto informatico non è immediatamente leggibile nella sua materialità, ma risiede in una stratificazione di concetti, processi, modelli teorici e pratiche discorsive che ne hanno accompagnato la progettazione, l’uso e la diffusione.

In questo scenario, la terminologia non è un semplice accessorio linguistico, un apparato descrittivo o un elenco di etichette, ma diventa l'infrastruttura cognitiva e interpretativa necessaria per restituire intelligibilità all'oggetto e strutturare, esplicitare e mediare la conoscenza, sia tra gli specialisti sia tra

il pubblico non esperto, consentendo di collegare testi, concetti e oggetti materiali in modo coerente e interoperabile.

I termini informatici non si limitano a denominare entità tecniche, ma riflettono visioni del mondo, modelli computazionali, scelte progettuali e pratiche sociali che si sono evolute nel tempo.

La forte variabilità terminologica che caratterizza il dominio informatico – sia in senso diacronico sia diafasico – rappresenta dunque non un problema da eliminare, ma una risorsa conoscitiva da analizzare e valorizzare.

La presente tesi si colloca dunque all'incrocio tra studi terminologici, linguistica dei corpora e rappresentazione della conoscenza, muovendo dall'ipotesi che la valorizzazione del patrimonio informatico passi inevitabilmente attraverso una modellizzazione semantica rigorosa, fondata sull'osservazione empirica degli usi linguistici e sulla formalizzazione concettuale. In questo senso, la terminologia viene assunta come ponte tra il dato tecnico, la memoria storica e il linguaggio naturale.

2. Evoluzione del quadro teorico: dalla Teoria Generale della Terminologia all'approccio Sociocognitivo

Una parte consistente di questo lavoro è dedicata all'evoluzione del quadro teorico della terminologia, necessaria per comprendere come la denominazione degli oggetti informatici sia mutata insieme alla percezione sociale della tecnologia. La ricerca prende, in parte, consapevolmente le distanze dalla visione tradizionale della Teoria Generale della Terminologia di Eugen Wüster. Sebbene l'impostazione "wüsteriana" sia stata fondamentale per la standardizzazione tecnica, orientata a una normalizzazione prescrittiva e univoca (un concetto, un termine), essa appare oggi insufficiente per descrivere un dominio dinamico, stratificato e profondamente influenzato dall'uso sociale come quello dell'informatica. L'ideale di corrispondenza biunivoca tra concetto e termine appare inadeguato di fronte a fenomeni quali la polisemia, la sinonimia funzionale, l'obsolescenza terminologica e la coesistenza di denominazioni concorrenti.

In alternativa, la tesi adotta una prospettiva descrittiva e sociocognitiva, in linea con gli sviluppi della Terminologia Comunicativa e Sociocognitiva (Cabré, 1999; Temmerman, 2000) che riconosce nel termine un'unità di conoscenza soggetta a variazioni diacroniche, diafasiche e contestuali. Questo passaggio teorico è fondamentale per giustificare l'analisi condotta nel lavoro: non si cerca una "definizione corretta" universale o cristallizzata, ma si indaga come i concetti (si pensi a termini polisemici o in evoluzione come *Personal Computer*, *Mainframe*, *Memory* o *Server*) siano stati rinegoziati nel tempo e nei diversi contesti d'uso. L'attenzione è rivolta sia ai modelli teorici classici sia agli approcci più recenti basati su corpora, che permettono di osservare la terminologia "in vivo"

(Cabré, 1999), ovvero nel suo manifestarsi concreto all'interno dei discorsi dei laboratori di ricerca e dei manuali tecnici o spontanei dei forum di appassionati e collezionisti come Reddit.

3. L'Ontoterminologia e le tecnologie del Web Semantico

Il cuore metodologico della tesi è rappresentato dall'Ontoterminologia, teoria sviluppata da Christophe Roche con il gruppo Condillac (Roche, 2007; Roche *et al*., 2009)[1], che propone una sintesi tra terminologia e ingegneria della conoscenza. L'integrazione tra la dimensione linguistica (il termine, legato al linguaggio naturale e alla variazione) e quella logico-concettuale (l'ontologia, legata alla formalizzazione del sapere) permette di costruire modelli di conoscenza che sono al contempo leggibili dagli esseri umani e processabili dalle macchine.

In questo quadro, la tesi introduce l'uso di standard formali della modellizzazione ontologica e del Web Semantico. Viene esplorato l'uso di standard come il CIDOC-CRM (2021), specifico per i beni culturali, e di strumenti dedicati come TEDI (ontoTerminology editor). L'obiettivo finale non è la creazione di un glossario isolato, ma la genesi di un ecosistema di dati interconnessi (Linked Open Data) che permetta al patrimonio informatico di uscire dal proprio isolamento specialistico e di dialogare con altre istituzioni culturali, museali e accademiche a livello globale. La tesi mostra come la terminologia possa fungere da ponte tra le rappresentazioni semantiche e i sistemi formali di conoscenza.

4. Obiettivi e struttura del lavoro

A partire da questo quadro, la tesi propone un percorso che conduce dall'analisi dei testi alla strutturazione dei concetti, articolandosi in quattro momenti chiave che riflettono l'approccio interdisciplinare adottato:

- Il **Capitolo 1** introduce i principali riferimenti teorici e metodologici che costituiscono la base concettuale del lavoro. Esso ricostruisce il dibattito terminologico dalle sue origini fino alle moderne sfide della rappresentazione della conoscenza nelle Digital Humanities, analizzando il passaggio dal prescrittivismo al descrittivismo.
- Il **Capitolo 2** è dedicato all'analisi critica dello stato dell'esistente nel settore dei beni culturali digitali. Vengono esaminati gli standard di catalogazione esistenti, le ontologie di dominio e i sistemi di organizzazione della conoscenza, evidenziando le criticità specifiche poste dalla catalogazione degli artefatti informatici.
- Il **Capitolo 3** costituisce la fase empirica e analitica della ricerca. Attraverso la metodologia della linguistica dei corpora e l'ausilio di strumenti di analisi testuale (come IRaMuTeQ,

[1] http://ontoterminology.com

Sketch Engine, Tropes e TermoStat), viene mappata la variazione terminologica in tre ambiti discorsivi distinti: i discorsi storici/accademici, i discorsi delle comunità di pratica (social media e forum) e i discorsi istituzionali dei musei.

- Il **Capitolo 4** presenta l'applicazione pratica dei modelli teorici discussi nel contesto del progetto europeo ITinHeritage. Qui la teoria si traduce in una risorsa ontoterminologica multilingue e multimediale, progettata per la valorizzazione dei musei dell'informatica, dimostrando come la modellizzazione semantica possa trasformare un oggetto tecnico in un nodo di conoscenza dinamico.

Nel suo insieme, la tesi intende dimostrare che la terminologia, lungi dall'essere un mero strumento ancillare, costituisce una chiave di accesso privilegiata alla comprensione del patrimonio informatico. Attraverso una gestione consapevole della variazione linguistica e una formalizzazione semantica rigorosa, è infatti possibile restituire agli oggetti informatici la loro piena dimensione storica, culturale e concettuale, trasformandoli da artefatti opachi in nodi di conoscenza dinamici e interconnessi. Solo un approccio integrato, che metta in relazione la dimensione terminologica e quella ontologica, consente di preservare non soltanto l'oggetto informatico in quanto tale, ma anche il senso culturale e l'eredità intellettuale che esso incorpora e trasmette nel tempo.

Capitolo 1. Le ontologie e le terminologie negli studi in terminologia

Introduzione al quadro teorico

Nel panorama accademico e professionale, le ontologie e le terminologie rivestono un ruolo cruciale nel facilitare la comunicazione, la comprensione e lo sviluppo del sapere all'interno di discipline specifiche.

L'ontologia è una specificazione esplicita e formale della concettualizzazione di un dominio, in cui vengono definiti concetti (o classi), relazioni, funzioni e altri elementi, tra cui costrutti che specificano vincoli e proprietà del dominio modellato (Gruber, 1993). La terminologia è un campo interdisciplinare che si occupa dei concetti e delle loro rappresentazioni (termini, simboli, ecc.), includendo l'insieme dei termini che rappresentano il sistema dei concetti di un campo specifico e le pubblicazioni in cui tale sistema è rappresentato (Felber, 1984). Più specificamente, essa viene definita come l'insieme delle designazioni e dei concetti appartenenti a un singolo dominio o soggetto[2]. La loro importanza è evidente in una vasta gamma di settori, dalla medicina all'ingegneria, dalla giurisprudenza all'informatica, dove concetti e termini specialistici sono essenziali per trasmettere efficacemente le conoscenze e garantire una comunicazione chiara e precisa sia tra gli esperti del settore sia con il pubblico.

[2] Traduzione nostra della definizione ISO 1087:2019: "*set of designations and concepts belonging to one domain or subject*", 3.1.11. p.2.

1.1 Nascita della terminologia - le prime scuole

In questo capitolo, verranno esaminate le principali teorie che hanno influenzato lo studio della terminologia come disciplina. Si analizzeranno le fondamenta concettuali di tali teorie, le loro applicazioni pratiche e le relative sfide, al fine di comprendere le dinamiche linguistiche e terminologiche nei diversi ambiti del sapere.

La terminologia ha l'obiettivo di identificare e descrivere gli oggetti e i concetti che formano il sistema organizzato di conoscenze di un determinato settore specialistico, mirando a denominarli in modo il più possibile univoco (Adamo, 1999).

La terminologia opera, dunque, all'interno di linguaggi speciali. Il linguaggio speciale è definito nella norma ISO 1087 come «linguaggio utilizzato in un settore specifico, caratterizzato dall'uso di mezzi espressivi linguistici particolari». In questo contesto, essa si occupa di analizzare e organizzare il lessico settoriale al fine di garantire chiarezza e precisione nella comunicazione specialistica.

Secondo lo standard ISO 1087-1:2000, un termine è definito come «designazione verbale di un concetto generale» e «unità di conoscenza creata da una combinazione unica di caratteristiche». La designazione, a sua volta, è descritta come una «rappresentazione di un concetto mediante un segno che lo denota». La designazione è quindi strettamente legata alla terminologia. La designazione, ovvero il processo di assegnare nomi agli oggetti che ci circondano, è un fenomeno antico quanto il linguaggio umano stesso. Un esempio concreto di questo processo si trova in campi come la botanica e la zoologia, campi nei quali è noto anche come creazione di una *nomenclatura* e segue regole precise stabilite a livello internazionale. Le designazioni per descrivere oggetti naturali come piante o animali di norma seguono una logica tassonomica che riflette il loro ordine e classificazione scientifica.

La terminologia può essere considerata un processo di raffinamento della designazione, volto a organizzare, definire e sistematizzare semanticamente i riferimenti ai concetti in modo chiaro ed inequivocabile (Wüster, 1979). La terminologia nasce dal bisogno di categorizzare e sistematizzare la realtà che ci circonda. La terminologia serve quindi a organizzare e razionalizzare la conoscenza umana, che cambia e si evolve con il tempo.

Ad esempio, il termine *telefono* originariamente descriveva un dispositivo di comunicazione ideato per trasmettere la voce a distanza attraverso cavi. Col tempo, l'evoluzione tecnologica ha portato a significativi cambiamenti. Oggi, la parola 'telefono' non si limita più a indicare un apparecchio fisso collegato a una linea telefonica, ma può includere dispositivi mobili come telefoni cellulari e smartphone, che vanno ben oltre la semplice funzione di comunicazione vocale.

Ora, un telefono permette di navigare su Internet, scattare foto, utilizzare applicazioni, inviare messaggi istantanei, svolgere operazioni bancarie, e molto altro ancora.

La terminologia si evolve con il progresso e riflette l'evoluzione delle tecnologie e l'ampliamento delle loro funzionalità, riorganizzando la conoscenza per adattarla alla realtà mutevole che ci circonda.

Occorre quindi chiarire cosa si intenda per *termini* e quale sia il ruolo della terminologia.

La terminologia come disciplina nasce negli anni Trenta del Novecento con il suo primo grande esponente: Eugen Wüster.

Eugen Wüster (1898-1977) è considerato il padre fondatore della terminologia moderna ed è a lui che dobbiamo l'elevazione della disciplina allo status scientifico con la sua Teoria Generale della Terminologia (GTT). Ingegnere e linguista austriaco, egli riteneva importante fornire un insieme generale di principi terminologici.

Il suo libro "Internationale Sprachnormung in der Technik" (1931) è considerato il punto di partenza della terminologia moderna. Nel corso della sua carriera, Wüster ha posto le basi teoriche per la standardizzazione terminologica, collaborando con numerosi istituti e reti di terminologia per lo sviluppo di standard terminologici di scala globale, tra cui il Comitato Tecnico (TC) 37 dell'ISO (International Organization for Standardization), e il centro di documentazione Infoterm (Faber, L'Homme, 2022).

Questa visione scientifica e sistematica della terminologia è stata ulteriormente arricchita dai contributi di figure di spicco del panorama internazionale, quali Juan C. Sager e il linguista e terminologo giapponese Kyo Kageura.

Sager vede, infatti, la terminologia come l'insieme di pratiche e metodi a supporto per la raccolta, descrizione e presentazione dei termini, per la sistematizzazione della conoscenza in campi specifici (Sager, 1990).

Questo riflette una delle prospettive prevalenti nella letteratura, secondo cui i termini esprimono conoscenze specializzate e sono strettamente legati alle strutture concettuali di settori specifici, distinguendosi così dalle unità linguistiche comuni. Tuttavia, la terminologia non è stata considerata sin dall'inizio da tutti, soprattutto dai linguisti, come disciplina scientifica a sé, evidenziando il dibattito che ha accompagnato l'evoluzione di questo campo.

Secondo il terminologo Kyo Kageura (1999), la terminologia può essere considerata parte integrante del vasto campo della linguistica. Questa visione si allinea con un secondo pensiero che va a considerare i termini come unità linguistiche soggette a fenomeni linguistici generali, come la polisemia e la variazione. Egli sostiene che i precedenti tentativi di stabilire la terminologia come disciplina autonoma non hanno avuto successo poiché erano eccessivamente focalizzati sugli aspetti teorici, trascurando i termini come oggetti empirici. Kageura sottolinea l'importanza di distinguere tra questioni teoriche (*quid iuris*) e questioni di fatto (*quid facti*) per definire chiaramente lo status

della terminologia come disciplina indipendente. Inoltre, questo approccio offre un quadro teorico per studiare come la terminologia cambia nel tempo, concentrandosi sulle dinamiche interne che influenzano l'evoluzione dei termini all'interno di un dominio specifico, offrendo così una base per un'analisi più strutturata e autonoma della terminologia stessa (Kageura, 1999*)*.

Per comprendere lo status della terminologia classica sin dalle sue origini, è necessario esaminare l'impostazione wüsteriana, che la considera una rappresentazione delle espressioni e dei termini specifici all'interno di una visione concettuale. Questa visione implica che ogni termine non sia isolato, ma faccia parte di un sistema più ampio di denominazioni di *concetti* che definiscono e organizzano il significato all'interno di un campo specifico.

Secondo la norma ISO 1087:2019, un concetto è definito come «unità di conoscenza creata da una combinazione unica di caratteristiche». Tali caratteristiche devono essere essenziali e rappresentative delle proprietà fondamentali e dei tratti distintivi che definiscono l'identità di ciascun elemento, permettendo di distinguere ogni oggetto della conoscenza umana. Per evitare ambiguità, è necessario che queste caratteristiche abbiano una precisa collocazione.

In questo contesto, Lynne Bowker evidenzia l'importanza per un terminologo di identificare le caratteristiche (o dimensioni) di un concetto per comprendere appieno la sua natura e il suo significato. Secondo Bowker, la somma delle proprietà definisce l'intensione del concetto, rivelando anche le relazioni tra concetti. Infatti, gruppi di concetti possono essere associati tra loro quando condividono determinati tratti distintivi, il che facilita la loro classificazione e comprensione all'interno di un sistema di conoscenza (Bowker, 1995).

È la terminologia la disciplina che studia sistematicamente i concetti e le loro denominazioni, ossia i termini, in uso nelle lingue specialistiche di una scienza, di un settore tecnico, di un'attività professionale o di un gruppo sociale, con l'obiettivo di descriverne e/o prescriverne l'uso corretto (Riediger, 2023). Come afferma Hellmut Riediger, la terminologia è «l'insieme dei termini che rappresentano un sistema concettuale di un dominio particolare».

Sager (1999) evidenzia che una teoria della terminologia ha tre compiti fondamentali: innanzitutto, deve considerare insiemi di concetti come (1) entità distinte all'interno della struttura della conoscenza; (2) in secondo luogo, deve tenere in conto insiemi di elementi linguistici interconnessi, che si associano in qualche modo a concetti raggruppati e strutturati secondo principi cognitivi; (3) infine, deve stabilire un collegamento tra concetti e termini, tradizionalmente realizzato tramite definizioni. Tali elementi riflettono un rapporto bilaterale e costante tra concetti e termini.

Un esempio del concetto di algoritmo nel campo dell'informatica, secondo i tre passi suggeriti da Sager, potrebbe essere il seguente:

(1) Un algoritmo in informatica è definito come una sequenza finita di istruzioni ben definite, progettata per eseguire un compito specifico. Si tratta di un concetto (entità) distinto rispetto ad altri termini correlati (2), come:

Codice: rappresenta l'implementazione dell'algoritmo in un linguaggio di programmazione.

Linguaggio di programmazione: è un insieme di regole formali che consente di scrivere il codice e di tradurre le istruzioni in linguaggio comprensibile dalle macchine.

Programma: è l'implementazione di un algoritmo, cioè di una sequenza di istruzioni logiche, che viene scritta sotto forma di codice utilizzando un linguaggio di programmazione specifico. Questo codice viene poi eseguito da un computer per svolgere un compito.

Pseudocodice: è una rappresentazione informale e leggibile dell'algoritmo

Loop: sono una serie di istruzioni in grado di ripetere un'operazione o un insieme di operazioni all'interno dell'algoritmo.

(3) Questi termini formano un sistema concettuale che collega le proprietà e le applicazioni dell'algoritmo. Il termine 'algoritmo', dunque, è associato a un concetto ben definito che può essere utilizzato in modo preciso in contesti scientifici e tecnici, evitando ambiguità e garantendo chiarezza nella sua applicazione.

I concetti, secondo lo standard ISO 704/2009 «Principi e metodi del lavoro terminologico, linee guida per riconoscere i concetti», sono riassunti in tre punti fondamentali:

1. Descrizione di oggetti: I concetti descrivono o corrispondono a oggetti o gruppi di oggetti.
2. Rappresentazione linguistica: I concetti sono rappresentati o espressi attraverso designazioni linguistiche o non linguistiche, oppure da definizioni.
3. Relazioni tra concetti: I concetti sono legati da relazioni e organizzati in sistemi concettuali strutturati sulla base di tali relazioni.

Questi aspetti della terminologia tradizionale riflettono un rapporto biunivoco e permanente tra concetti e termini.

In questa visione sistematica, il concetto è definito in modo netto e rigoroso, «clear-cut» (Temmerman, 2000) e corrisponde a un unico termine, che viene assegnato in modo permanente. La teoria generale della terminologia (GTT) non considera le variazioni linguistiche, come la polisemia, e si concentra sulla normalizzazione dei concetti da un punto di vista sincronico. Questa impostazione, pur garantendo chiarezza e precisione nella definizione dei concetti, tende a privilegiare la stabilità e l'univocità del linguaggio, minimizzando, in qualche misura, la creatività e la dinamicità del linguaggio.

Wüster, in particolare, attribuisce una maggiore importanza al vocabolario rispetto alla grammatica, ritenendo che la standardizzazione e la chiarezza dei termini siano fondamentali per garantire precisione e coerenza nei contesti tecnici e scientifici. Sebbene la complessità della grammatica sia riconosciuta come importante, non è considerata centrale nel processo di normalizzazione.
In sintesi, Wüster concepisce la terminologia come un metodo di standardizzazione del lessico specialistico, mirato alla normalizzazione delle nozioni e dei termini e alla riduzione delle ambiguità.
Un aspetto fondamentale della terminologia, come delineato da Wüster e sostenuto dalla maggior parte dei terminologi, è l'adozione di un approccio onomasiologico, che parte dai concetti per arrivare ai termini. Questo processo si distingue da quello lessicologico, che invece si focalizza sulle parole per risalire ai concetti (processo semasiologico). Nel lavoro terminologico, il focus è sui concetti stessi, escludendo influenze da altri fattori linguistici e informazioni diacroniche (Cabré, 1999).
L'approccio onomasiologico, dunque, parte dalla realtà oggettiva per giungere alla denominazione. Gli oggetti e le entità astratte esistenti nel mondo vengono dapprima identificate e descritte attraverso un concetto, per poi essere rappresentate da un termine specifico che funge da identificatore univoco. Se un oggetto esiste nel mondo può essere nominato da nuovo concetto e inserito nel sistema concettuale. Di conseguenza, il terminologo, assumendo l'esistenza oggettiva del mondo e la possibilità di nominarlo, è chiamato ad esplorare inizialmente la dimensione concettuale di uno specifico ambito di studio, per poi associare a tali concetti le loro rappresentazioni linguistiche.

A partire dagli anni '70 il panorama degli studi subisce un'evoluzione; gli orientamenti della teoria terminologica in questo periodo vengono individuati da Helmut Felber e riassunti in modo esaustivo da Giovanni Adamo (1999), nei seguenti punti:

1) Approccio Interdisciplinare: Rappresentato dalla Teoria Generale della Terminología di Eugen Wüster, questo approccio, riassumendo ciò che è stato già detto, si concentra sul concetto e sulle sue relazioni all'interno di un sistema concettuale. L'obiettivo principale è la standardizzazione dei termini e la loro assegnazione ai concetti, cercando di mantenere una visione chiara e uniforme della terminologia.

2) Classificazione Filosofica dei Concetti: Questo approccio, discusso da Dahlberg (1974), si basa sulla classificazione dei concetti secondo categorie filosofiche. È un metodo che cerca di ordinare i concetti in base a una struttura logica e categoriale.

3) Studi Linguistici: Considerano le terminologie come sottocomponenti del lessico di un linguaggio speciale, rientranti nei sottocodici delle varie lingue.

Per tracciare un'evoluzione storica della disciplina, è opportuno fare riferimento alla classificazione pionieristica proposta da Pierre Auger, il quale per primo ha sistematizzato le diverse correnti terminologiche in base alle loro finalità e contesti d'uso. Questo schema è stato successivamente ripreso e arricchito da M. Teresa Cabré (1999), la quale offre una panoramica dettagliata delle principali scuole di terminologia e i relativi approcci emersi negli anni, che si possono riassumere nei seguenti punti:

1) approccio linguistico rappresentata dalle scuole di Vienna, Praga e Mosca;
2) approccio in traduzione-internazionale;
3) approccio alla pianificazione del linguaggio, con l'intento anche a livello legislativo di promuovere la valorizzazione di una lingua minoritaria considerata "a rischio", come ad esempio la lingua francese in Canada.

La **Scuola Austriaca** nasce dalla Scuola di Vienna o Circolo di Vienna, ed è caratterizzata da una visione interdisciplinare, che integra discipline tecniche e scientifiche, ma mantiene la terminologia come disciplina autonoma. L'importanza è data ai concetti e alle relazioni tra di essi; questa scuola è influenzata dalle idee di Eugen Wüster e dell'oggettivismo del Wiener Kreis. In questo contesto, si ritiene che i concetti esistano indipendentemente dal linguaggio, e l'obiettivo principale è la standardizzazione della terminologia, senza considerare l'evoluzione della lingua. Wüster, infatti, distingue tra terminologi, che si occupano dei concetti, e linguisti, che si concentrano sul contenuto delle parole. La terminologia tradizionale, in linea con i principi strutturalisti saussuriani del linguaggio, parte dall'oggettivismo, ovvero dalla concezione di un mondo oggettivo indipendente e prescindente dall'osservazione e dall'esperienza umana.

La **Scuola Ceca** legata alla Scuola di Praga o Circolo di Praga adotta una prospettiva linguistica, considerando la terminologia come una sottocomponente del lessico. Qui, l'enfasi è sulla descrizione strutturale e funzionale delle lingue speciali. Gli ultimi due approcci, quello russo e quello ceco, sono influenzati dalle teorie saussuriane, che vedono la *langue* e la *parole* come due aspetti complementari della lingua. In questo contesto, il termine rappresenta la forma e l'espressione pratica del concetto, ciò che abbiamo nella nostra mente.

La **Scuola Russa** usa un approccio di natura filosofica, focalizzato sulla classificazione logica dei concetti e sull'organizzazione della conoscenza. La Scuola di Mosca, operante nel contesto del plurilinguismo dell'ex Unione Sovietica, si occupa della normalizzazione linguistica, cercando di gestire la terminologia all'interno di un contesto linguistico diversificato.

La **Scuola canadese aménagiste** (francese-canadese) si propone di integrare le metodologie descritte, considerando il termine come un punto di partenza. Il termine non è più visto come un'entità statica o etichetta fissa, ma come un elemento dinamico da cui si può avviare un'analisi più approfondita sui concetti e sulle loro relazioni. Il suo obiettivo principale è preservare la lingua francese in Canada. Inoltre, questa scuola è tra le prime a esplorare i nuovi approcci sociocognitivi, che verranno analizzati in seguito.

L'evoluzione della terminologia come campo autonomo non può essere compresa appieno senza mapparne la dimensione propriamente storiografica, intesa come un processo stratificato di auto-riflessione scientifica. Come recentemente evidenziato nel volume curato da Warburton e Humbley, *Terminology throughout History. A Discipline in the Making* (2025), la transizione verso la modernità epistemologica della disciplina ha richiesto una costante rinegoziazione dei propri confini metodologici. L'analisi storica dimostra che le prime spinte alla standardizzazione, sebbene apparentemente rigide, rispondevano a precise contingenze di cooperazione internazionale e industriale.

In questa prospettiva, la nascita delle diverse scuole non va letta come una sequenza di paradigmi isolati, bensì come un percorso dialettico in cui l'esigenza prescrittiva wüsteriana si è progressivamente confrontata con le specificità culturali e testuali delle comunità di pratica. Il recupero di questa memoria storiografica permette di ridefinire lo statuto attuale della terminologia nelle Digital Humanities: essa non si configura più come un mero apparato tecnico di normalizzazione, ma come una disciplina intrinsecamente storica, il cui compito principale è documentare le modalità attraverso cui le comunità umane hanno formalizzato e trasmesso il sapere specialistico.

1.2 Dai testi alla cognizione: da un approccio prescrittivo ad un approccio descrittivo

Negli ultimi 30 anni, vi è stato un cambiamento di prospettiva nella terminologia, con nuovi dibattiti e rivalutazioni. Nonostante la GTT sia un punto di riferimento per la terminologia moderna, il suo approccio è oggi da diversi studiosi considerato limitato, specialmente per la sua idealizzazione della realtà, che non tiene conto dei dati empirici (Cabré, 2023) e per la ridotta applicabilità al di fuori della standardizzazione terminologica. La GTT è stata, infatti, oggetto di ampie critiche e rivalutazioni, tra cui l'adozione di un paradigma più semasiologico, basato sulla **teoria testuale** di Bourigault e Slodzian (1999).

Per l'approccio testuale**,** il testo è il punto di partenza ed è necessario comprendere alla base il legame tra la terminologia e la semantica.

La terminologia si basa sulla semantica in modo intrinseco e al tempo stesso fondamentale. La terminologia riguarda l'uso preciso dei termini in un campo specifico e la semantica studia il significato di questi termini e delle parole nell'uso generale. La semantica è lo studio del significato: si occupa di come le parole, le frasi e i testi rappresentano idee e concetti. L'approccio testuale nella terminologia si riferisce all'analisi e alla descrizione dei termini attraverso l'osservazione del loro uso reale nei testi.

Questo approccio riconosce che i termini non esistono in quanto entità isolate, ma sono parte integrante dei contesti comunicativi e discorsivi in cui vengono utilizzati. L'approccio testuale sottolinea che i testi non sono semplici contenitori di termini, ma piuttosto manifestazioni dinamiche delle pratiche comunicative. Per questo motivo grazie all'analisi dei testi specialistici (come articoli scientifici, documentazione tecnica, manuali) è possibile comprendere come i termini vengono effettivamente utilizzati. È possibile osservare le varianti terminologiche, le collocazioni (Sakr, 2022) e le sfumature di significato che possono emergere in contesti diversi. E ancora, è possibile identificare le relazioni semantiche tra i termini, come sinonimia, antonimia, iperonimia e iponimia e la loro evoluzione nel tempo. Questa prospettiva è particolarmente preziosa nei settori in rapida evoluzione, dove l'aspetto diacronico consente di monitorare il cambiamento dei termini e dei concetti correlati, contribuendo a una visione storica e sempre aggiornata. Grazie all'analisi dei testi specialistici, si possono documentare i cambiamenti lessicali e semantici, evitando che importanti informazioni linguistiche e concettuali vadano perse nel tempo e mantenendo una comprensione costante e aggiornata dell'evoluzione del settore. Un esempio significativo di un campo in continua evoluzione è quello informatico, dove si osservano termini più recenti come "Cloud Computing", "quantum computer", "EDGE AI" e "blockchain". Questi neologismi riflettono le innovazioni tecnologiche e i cambiamenti significativi nelle esigenze del settore.

L'emergere dell'interesse per la variazione diacronica nella terminologia rappresenta un cambiamento significativo, poiché questo campo ha storicamente trascurato l'aspetto temporale,

In passato, la terminologia si concentrava prevalentemente su studi sincronici, come evidenziato da Wüster. Sebbene le prime ricerche sulla terminologia diacronica siano emerse in modo limitato e tardivo, dalla fine degli anni '80 si sono verificati sviluppi significativi in questo ambito. In particolare, la conferenza di Bruxelles del 1988 e i successivi contributi di De Schaetzen e Moller hanno evidenziato l'importanza della prospettiva diacronica nella terminologia. Questi studi pionieristici hanno dimostrato che la variazione diacronica è cruciale per comprendere l'evoluzione dei termini e dei concetti nel tempo (Faber, L'Homme, 2022).

Nel corso degli anni '90 e 2000, la legittimazione della prospettiva diacronica è aumentata grazie a nuove correnti teoriche come la socioterminologia e la terminologia testuale. Inoltre, l'interesse crescente per la variazione diacronica ha condotto all'esplorazione di fenomeni precedentemente trascurati, quali l'obsolescenza dei termini e la condivisione di concetti tra diversi ambiti specialistici. Questa dinamicità dei termini può essere esplorata da diverse prospettive, poiché la nostra conoscenza dei campi specializzati evolve, così come i termini utilizzati.

L’analisi diacronica risulta essenziale per tracciare tali trasformazioni, evidenziando come i significati si evolvano in parallelo ai cambiamenti della società e della tecnologia.

Grazie all'approccio testuale è possibile dare visibilità ai neologismi e alle nuove sfumature semantiche che emergono in risposta alle esigenze della pratica professionale.

Il lavoro terminologico si integra quindi con l'analisi del discorso, enfatizzando il ruolo della terminologia nella comunicazione reale. Questo approccio sottolinea l'importanza di considerare come i termini vengano utilizzati nei discorsi quotidiani e specialistici, riconoscendo che la terminologia non è solo una questione di definizione di concetti, ma anche di pratica linguistica e di interazione comunicativa. Non si tratta più solo di definire termini, ma di esplorare come questi influenzino il nostro modo di pensare e comunicare.

Un approccio significativo nello studio della terminologia consiste nella combinazione della teoria terminologica tradizionale con le prospettive offerte dalla linguistica cognitiva. Questo tipo di integrazione, formalizzato da Dirk Geeraerts nel 1993 con il concetto di **terminologia cognitiva**, segna un importante passaggio metodologico.

Geeraerts propone un'integrazione tra la terminologia tradizionale e la linguistica cognitiva (semantica cognitiva), utilizzando dati reali provenienti da corpora linguistici e enfatizzando l’importanza della natura prototipica, esperienziale ed enciclopedica dei significati terminologici.

Ciò permette di valutare la comprensione dei termini più ad ampio raggio, analizzando come e perché differenti comunità scientifiche o tecniche interpretano e utilizzano i concetti. In questo modello, i

concetti sono strutture mentali dinamiche che riflettono la nostra comprensione e categorizzazione del mondo e che di conseguenza possono modificarsi nel tempo, contrastando di fatto, la visione strutturalista wüsteriana. Questo fenomeno evolutivo è ripreso anche dall'accademico giapponese Kyo Kageura, che introduce la nozione di "dinamica della terminologia" (2002).

Faber (2022) afferma, a questo proposito, che la linguistica cognitiva fornisce un quadro metodologico utile per analizzare le unità terminologiche nel linguaggio specialistico, in particolare per la loro capacità di trasmettere conoscenza nei testi specializzati.

Se si parla di cognitivismo si fa riferimento a tutti quei processi mentali coinvolti nell'acquisizione, nell'immagazzinamento, nella manipolazione e nel recupero delle informazioni.

La linguistica cognitiva, dalla sua nascita negli anni '70, si concentra sull'analisi del linguaggio reale e sulle modalità con cui esso rispecchia i processi cognitivi, rivelando come la mente umana organizza e comprende il mondo. Secondo Faber (2022), questo campo studia la relazione tra linguaggio e percezione, soffermandosi su concetti chiave come l'incorporamento (embodiment), espressione di schemi immagine, metafore e metonimie, e abilità cognitive che coinvolgono il cambiamento di attenzione e la categorizzazione. Come spiega Faber, questi strumenti cognitivi confermano che "la linguistica cognitiva mira a specificare le motivazioni dietro i fenomeni linguistici" e consente di creare "schemi coerenti attraverso fenomeni linguistici diversi"[3]. In tal modo, si costruisce un'immagine mentale del mondo che riflette il nostro modo di percepire e categorizzare la realtà.

Come evidenzia Faber (2022), la linguistica cognitiva considera la terminologia un'attività linguistica e cognitiva. I termini sono unità linguistiche di significato che riguardano aspetti morfologici, sintattici, semantici e pragmatici e sono fondamentali per la comunicazione specialistica. La grammatica cognitiva, inoltre, si basa sul principio dell'uso e sostiene che i significati terminologici sono il risultato di un'interazione continua tra i concetti e le esperienze degli utenti. In quest'ottica, i termini trasmettono conoscenza entro contesti specifici, in cui il mittente e il destinatario condividono non solo il linguaggio ma anche l'esperienza e la conoscenza del dominio trattato.

Faber propone tre prospettive fondamentali nella linguistica cognitiva: l'approccio esperienziale, la prominenza e l'attenzione. L'approccio esperienziale evidenzia come le associazioni che si formano attorno a una parola siano influenzate dal contesto e dai modelli mentali di chi comunica. La prominenza riguarda la selezione e la configurazione delle informazioni più rilevanti durante la comunicazione, mentre l'attenzione evidenzia come parlante e ascoltatore interpretano un evento comunicativo, ciascuno da una propria prospettiva. Queste tre prospettive sono supportate da evidenze psicologiche e considerano il linguaggio come una manifestazione della cognizione incarnata.

---

[3] Traduzione nostra (Faber, 2002, 2.2.4 Basic premises, p.37)

La linguistica cognitiva, come sottolineato da Faber (2022), è particolarmente utile per analizzare la **comunicazione specialistica** poiché "riconosce le caratteristiche uniche dei testi di linguaggio specializzato, che trasmettono conoscenza attraverso un uso ricco di termini." Il suo approccio lessicocentrico e basato sull'uso rende possibile uno studio dettagliato di come i termini riflettano il modo in cui interpretiamo e categorizziamo il mondo.

Le intuizioni di P. Faber si allineano con quelle di M.C. Cabré e con la sua **Teoria Comunicativa della Terminologia** (TCT), conosciuta nella versione inglese come *Communicative Theory of Terminology* (CTT). Cabré sottolinea l'importanza dell'aspetto comunicativo nella terminologia, sia nei suoi aspetti cognitivi e linguistici, (1998/1999), evidenziando come la terminologia emerga dalla comprensione del mondo e dalle esperienze degli individui. Si passa, dunque, ad una concettualizzazione che deriva direttamente dall'esperienza e dalla comprensione del mondo (Cabré (2000). Questa concezione viene accolta nelle scienze cognitive, tuttavia, Faber (2022) mette in luce alcune insufficienze della CTT, in particolare per quanto riguarda il suo legame con le teorie linguistiche e la definizione di significato concettuale. Inoltre, il modo in cui la CTT affronta la semantica specializzata risulta, secondo Faber, poco definito, poiché non chiarisce come si sviluppano le rappresentazioni concettuali e quali elementi le compongono. Questo rende complesso inquadrare la CTT nel panorama delle teorie sulla semantica lessicale.

1.3 Socioterminologia e semantica dei frame: approcci dinamici alla terminologia

La **socioterminologia** emerge concretamente negli anni ’90 come alternativa al rigido processo di standardizzazione tecnica adottato dalla GTT. La socioterminologia è un approccio che combina le pratiche linguistiche con le esigenze terminologiche, tenendo conto dell'influenza dei contesti sociali e culturali nella creazione e nell'uso dei termini. Essa considera la variazione linguistica come un aspetto essenziale e imprescindibile, distinguendone diversi tipi: diacronica, diatopica, diastratica, diafasica, diamesica e altre ancora.

Mentre l’approccio tradizionale della GTT promuove il principio di economia e la monorematicità (ovvero la preferenza per termini brevi e composti da una sola parola), Gambier evidenzia come le unità terminologiche non debbano necessariamente rispondere a criteri di concisione formale. Al contrario, esse riflettono la complessità del discorso specialistico, manifestandosi spesso attraverso strutture sintagmatiche estese e fenomeni di polisemia, che la norma ISO tende invece a escludere in favore di una standardizzazione univoca e statica.

In questo contesto, il confronto tra le raccomandazioni ISO 704:2009 e divergenze paradigmatiche del linguista Yves Gambier mette in luce differenze significative nella concezione della terminologia (Faber, L’Homme, 2022). Mentre le raccomandazioni ISO seguono un approccio più normativo e standardizzato, Gambier propone una maggiore apertura e flessibilità, tipica degli approcci sociolinguistici, sostenendo una visione che si oppone all'idea di una terminologia immutabile e fissa. Boulanger presta particolare attenzione alla creazione di nuove unità lessicali, ai neologismi, alle innovazioni lessicali, alla possibilità di intervenire nell’inserimento di nuovi item nei dizionari specialistici. Boulanger evidenzia, in più, come emergano nuovi termini, mentre i vecchi termini possano cambiare di significato o cadere in disuso.

Egli sottolinea l'importanza della diacronicità, definendo i neologismi secondo diversi criteri, tra cui la loro recente nascita (criterio diacronico), la loro assenza dai dizionari (criterio lessicografico), la loro instabilità e la percezione che i parlanti hanno di questi termini come nuove unità linguistiche (criterio psicologico, si veda anche Cabré, 1999).

Successivamente, Gaudin teorizza un **approccio descrittivo** basato sullo studio diacronico della storia della concettualizzazione e della denominazione, enfatizzando l'importanza del contesto sociale. I termini, secondo Gaudin, non sono solo strumenti linguistici neutri, ma riflettono le dinamiche sociali, culturali e professionali della comunità che li utilizza - si sviluppano e si modificano in risposta alle necessità comunicative.

Con la componente cognitiva teorizzata da Rita Temmermann (2000) e dal suo gruppo di ricerca Centrum voor Vaktaal en Communicatie (CVC) dell’Università di Vrije di Bruxelles, viene introdotto

il concetto *di unità di comprensione UoU*, unità concepite nella mente sulla base dell'esperienza e in continua evoluzione.
«La definizione di un'unità di comprensione x è una risposta alla domanda "che cos'è x?". Quali informazioni essenziali dipendono dal tipo di unità di comprensione» (Temmerman, 2000).
Questo approccio permette di integrare nuove espressioni linguistiche in base ai diversi contesti in cui i concetti vengono usati, rendendo la polisemia e la sinonimia funzionali nel campo specialistico per offrire prospettive diverse e una comprensione più profonda. Per Temmerman, infatti, i concetti non sono statici ma possono avere più significati, il che implica che i modelli di comprensione possono comprendere informazioni storiche, intracategoriali, intercategoriali e procedurali.

Se nella terminologia tradizionale si è parlato di identificatori univoci, dunque un termine unico per ogni concetto, nella **terminologia sociocognitiva** invece, la polisemia è funzionale nei discorsi specializzati in quanto fornisce informazioni utili, come ad esempio su eventuali cambiamenti sociali avvenuti per quel particolare dominio nel corso del tempo e attraverso il processo di variazione diacronica di un termine.
Nella terminologia sociocognitiva, la polisemia è il risultato dell'espressione ed evoluzione linguistica di una comunità nel corso del tempo e riflette come tale comunità abbia scelto il proprio lessico per comunicare e condividere informazioni in modo più efficiente possibile, senza ambiguità e perdita di informazioni. Questa teoria pone l'accento sugli aspetti sociali e culturali della terminologia, riconoscendo la possibilità che i termini subiscano cambiamenti nel tempo, nello spazio e nelle culture.
Kyo Kageura parla infatti di "questione di scelta" su come rappresentare i concetti o gli oggetti in un dominio specializzato. Secondo Kageura, la scelta terminologica può includere o escludere, ad esempio, i simboli extra-linguistici, come immagini e grafici, che possono arricchire la comprensione e la comunicazione del concetto. Ciò implica una considerazione attenta delle diverse esigenze comunicative e culturali che influenzano la definizione di un concetto.
Temmerman evidenzia come la teoria della Terminologia Sociocognitiva introduca l'UoU sulla base di tre elementi chiave: la percezione sensoriale della realtà, una mente funzionante e la padronanza di almeno una lingua umana.
Lo sviluppo delle UoU è influenzato da vari fattori, tra cui il desiderio di maggiore comprensione, l'interazione linguistica e i modelli cognitivi (Faber, L'Homme, 2022).
Temmerman distingue tre gruppi concettuali:

- Entità: concetti derivati da oggetti materiali o astratti.
- Attività: processi, operazioni e azioni svolte da o per le entità.

- Categorie collettive: categorie ombrello che includono altre entità e attività.

Temmerman, infatti, dimostra questo aspetto con un caso studio sui neologismi legali in Belgio riguardanti la definizione di "madre", identificando tre nuovi tipi di maternità (surrogata, anonima e co-maternità) che sfidano la definizione legale tradizionale, mostrando come la comprensione linguistica e categoriale possa variare culturalmente e linguisticamente.
La Terminologia Sociocognitiva introduce anche il concetto di triangolo semantico esteso, composto da simbolo (le parole), pensiero (il concetto) e referente (la realtà). Questo modello considera la comprensione come un processo dinamico, influenzato dalla lingua, dalle interazioni sociali e dalle esperienze sensoriali.

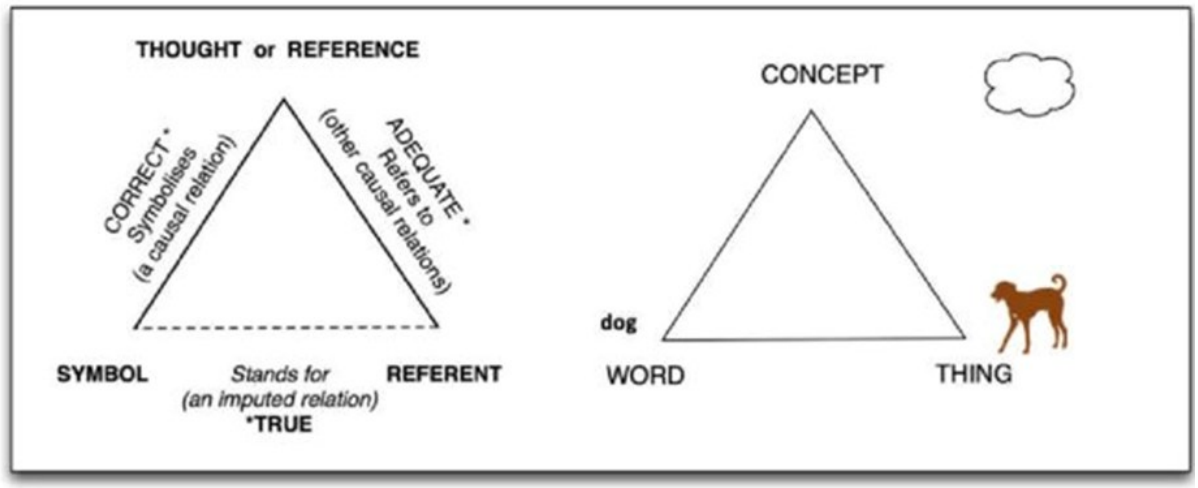


Figure 1 Illustrazione del triangolo semiotico secondo Ogden e Richards, adattata da Nuninger et al. (2020).

Un'altra teoria dinamica e rilevante nella terminologia è la teoria della **Semantica Cognitiva e dei Prototipi**, secondo cui il punto di partenza non è solo il concetto, ma piuttosto il termine presente nei testi scritti da specialisti del settore (Vezzani, 2022). Questa teoria suggerisce che i termini possano avere un nucleo centrale di significato, arricchito da estensioni più periferiche.
La Teoria dei Prototipi è stata sviluppata da Eleanor Rosch negli anni '70 e nasce nel contesto della Scienza Cognitiva e della Psicologia Cognitiva ed ha avuto un impatto significativo sulla linguistica e sulla terminologia, introducendo un nuovo modo di pensare alle categorie e ai concetti. Rosch suggerisce che le categorie non devono essere sempre organizzate gerarchicamente ma piuttosto devono essere considerate in una visione più flessibile, attraverso un concetto di "prototipo". Un prototipo è un esempio ideale o rappresentativo di una categoria. Non tutti i membri di una categoria sono ugualmente rappresentativi, alcuni possono essere più tipici per quella categoria rispetto ad altri. Nell'articolo *Principles of Categorization* (Rosch, 1978) la psicologa pone come esempio classico la categorizzazione degli uccelli: I pettirossi (*robin*) o i passeri (*sparrow*) sono considerati membri più

prototipici della categoria “uccello” rispetto ai pinguini. Questo perché i pettirossi possiedono molte delle caratteristiche tipiche degli uccelli (come volare e avere piume) che sono comunemente associate alla categoria “uccello”. I pinguini, sebbene siano tecnicamente uccelli, non condividono tutte le caratteristiche comuni degli uccelli, come la capacità di volare, e quindi sono considerati meno prototipici. Nelle risorse terminologiche, come glossari e dizionari, i termini possono essere organizzati in base alla loro rilevanza o frequenza d'uso piuttosto che solo in base a una struttura gerarchica rigida. Il processo di categorizzazione deve rispecchiare la struttura delle informazioni per come sono percepite dal mondo. I principi psicologici e cognitivi, dunque, influenzano la formazione e l'uso delle categorie.

Un ulteriore contributo significativo allo studio della linguistica e della terminologia nella dimensione cognitiva e dinamica è rappresentato dalla semantica dei frame (**Frame Semantics**), introdotta da Charles Fillmore nel 1976, e sulla base del quale è emerso successivamente l'approccio cognitivo alla terminologia chiamato Frame-based Terminology (FBT) sviluppato da Pamela Faber e dal suo team all'Università di Granada.

La semantica dei frame è un approccio alla semantica che pone l'accento sul significato delle parole in relazione ai “frame” cognitivi. I “frame” sono la più grande unità organizzativa della comprensione, ovvero strutture mentali che organizzano e rappresentano la nostra esperienza e conoscenza del mondo, in base a specifici contesti situazionali. Ogni frame rappresenta un concetto o un'esperienza, e include elementi chiave (slot) e le relazioni tra di essi.

Secondo Fillmore e la sua «semantica della comprensione», un termine come ‘ristorante’ può evocare una serie di relazioni e ruoli, come CAMERIERE e CLIENTE, che creano un quadro concettuale condiviso tra gli interlocutori. In questo senso, la terminologia non è solo un insieme di definizioni, ma riflette una conoscenza concettuale dinamica e interconnessa, costruita attraverso l’esperienza e l’uso.

La dinamicità della formazione dei concetti è un aspetto centrale; essa implica che i significati terminologici siano in continua evoluzione e richiedano un controllo terminologico per affrontare i cambiamenti nel loro significato (Faber, 2022)

Per illustrare questo concetto, consideriamo il frame del campo dell'informatica, il quale include non solo termini come ‘linguaggi di programmazione’, ‘software’, ‘hardware’ e ‘rete informatica’, ma anche le relazioni che intercorrono tra di essi. Ad esempio, il termine ‘software’ può evocare specifici ruoli, come ‘sviluppatore’ e ‘utente finale’, suggerendo un'interazione dinamica in cui il significato di “software” si arricchisce a seconda del contesto in cui viene utilizzato. Allo stesso modo, “rete informatica” non è solo un termine tecnico, ma un concetto che implica relazioni tra vari elementi,

come ‘server’, ‘clienti’ e ‘protocollo’, i quali sono essenziali per comprendere appieno il funzionamento delle comunicazioni digitali.

Faber (2022), con il concetto di “quadro” (*frame*) della semantica cognitiva, chiarisce che i significati terminologici dipendono dal contesto e dalle associazioni culturali, riconoscendo così l'importanza della conoscenza strutturata che va oltre il linguaggio.

I frame FBT sono definiti come *situated knowledge structures* e schematizzazioni dell'esperienza (Faber, 2015). Ciò che distingue la FBT dagli altri approcci cognitivisti è la metodologia basata su modelli sia linguistici che psicologici. Ad esempio, utilizza teorie come la teoria dei frame di Charles Fillmore e la Functional Lexematic Model, formulata dalla stessa Faber.

Un concetto può evocare un frame specifico nella mente dell'ascoltatore o del lettore. Il significato della parola è dunque determinato dalla rete di relazioni all'interno del frame, che non devono necessariamente limitarsi a relazioni di tipo semantico, ma possono aprire ad un quadro di relazioni molto più complesso e articolato. *Type_of, Affects, Location_of, Effected_by, Size_of, Delimited_of*, possono essere degli esempi e si differenziano in base al dominio di riferimento.

Ciò che rende la FBT particolarmente innovativa è la sua enfasi sulla modellazione concettuale nella creazione delle risorse di conoscenza. Questo approccio non si limita a strutturare singole voci terminologiche, ma mira anche a catturare le relazioni tra di esse, rivelando le combinazioni e le frequenze nei testi specializzati.

In questo caso, la progettazione delle risorse di conoscenza specializzata deve considerare sia i microcontesti che i macrocontesti dei concetti, aiutando così gli utenti a comprendere la fitta rete di significati (relazioni e concetti) di un determinato dominio.

La FBT, inoltre, propone modalità di rappresentazione dei termini che rispecchiano le strutture cognitive dei frame, attraverso l'uso di schemi, diagrammi o altre rappresentazioni visive. Queste visualizzazioni mettono in evidenza come i termini e i concetti siano collegati tra loro, offrendo una comprensione chiara della loro interconnessione e della loro organizzazione.

Se l'approccio cognitivo della FBT si concentra sulle strutture mentali e situazionali, un'altra critica alla rigidità tradizionale arriva dalla **“La terminologie culturelle”**. Mentre Faber guarda al 'frame' come struttura conoscitiva, Marcel Diki-Kidiri sposta l'accento sul soggetto parlante e sul suo sfondo etnolinguistico.

Nel 2000, Marcel Diki-Kidiri sottolinea l'importanza di considerare le peculiarità culturali e locali nella creazione e standardizzazione dei termini in una lingua, con un focus particolare sulle lingue africane. Per Diki-Kidiri ogni individuo e ogni comunità possiedono una storia unica che modella la cultura personale e collettiva. Questa storia spiega la diversità culturale e la sua evoluzione, mostrando come la cultura influenzi ogni aspetto della vita. Le esperienze accumulate dagli esseri

umani alimentano una memoria collettiva che mantiene viva la consapevolezza comunitaria, attraverso la trasmissione della storia, delle opere d'arte e delle tradizioni. Egli propone di non limitarsi esclusivamente alla terminologia tecnica e scientifica, ma di permettere ai parlanti nativi di scegliere naturalmente le espressioni più adatte. I concetti, infatti, possono essere percepiti culturalmente da diverse angolazioni, definite come “percept” (Diki-Kidiri, 2000).

La terminologia culturale, pertanto, pone l'essere umano al centro del processo terminologico, concentrandosi su come gli individui classificano, ordinano, nominano e categorizzano tutto ciò che percepiscono.

Tuttavia, Kageura (2000) evidenzia un problema significativo: quando si tratta di nominare nuove realtà o concetti nelle lingue africane, emergono delle limitazioni. Questi nuovi concetti spesso derivano dal contesto culturale e tecnologico occidentale e non erano presenti nelle tradizioni culturali africane. Di conseguenza, le lingue africane devono affrontare la sfida di adattare o creare nuovi termini per descrivere realtà che sono estranee alla loro esperienza culturale e tradizione preesistente.

1.4 Verso una nuova era della terminologia: tecnologie emergenti e approcci basati su corpora

Negli ultimi 20 anni il progresso tecnologico ha influenzato anche l'asset della teoria terminologica, ampliando i suoi metodi e le sue applicazioni, in particolare l'approccio testuale prende una nuova forma, più interessante e particolare, che trova spazio e punti di riflessione con Anne Condamines.
Prima degli anni '90, gli studi terminologici erano prevalentemente focalizzati sulla traduzione. Negli ultimi due decenni, invece, sono nati nuovi bisogni e spunti di analisi, insieme a strumenti innovativi per gestire e interpretare grandi volumi di dati. L'aumento della documentazione elettronica, la conseguente necessità di analisi dei corpus elettronici, le tecniche di machine learning (ML), natural language processing (NLP) e l'intelligenza artificiale (AI) hanno reso la disciplina della terminologia estremamente multidisciplinare, offrendo nuovi strumenti per affrontare le sfide di adattamento e innovazione, inclusi i problemi legati alla terminologia culturale.
Proprio Anne Condamines sostiene che i sistemi di intelligenza artificiale hanno il potenziale per simulare i processi di apprendimento e comprensione tipici degli esseri umani. (Condamines, 2022). Tuttavia, i corpora specializzati continuano a essere relativamente piccoli rispetto ai grandi corpora di linguaggio generale e presentano difficoltà significative nella loro raccolta.
In questo contesto è importante parlare di approccio terminologico basato su corpus, o **corpus-based terminology**, il quale offre la possibilità di sfruttare l'ampia disponibilità di testi digitali per estrarre, analizzare e comprendere empiricamente i termini specializzati nei vari domini.
Tali analisi sono supportate da modelli di 'semantica distribuzionale', i quali definiscono il significato dei termini attraverso i contesti in cui essi appaiono. Questo approccio non solo migliora la comprensione dei termini, ma sottolinea anche il legame tra semantica e terminologia, rendendo evidente come il significato dei termini possa variare a seconda del genere testuale in cui vengono utilizzati (Roche, 2005).

Con lo sviluppo della ricerca in terminologia negli anni '90, è emerso un legame intrinseco con i **generi testuali.** Per Condamines (2018) l'uso di corpora linguistici ha permesso di costruire terminologie partendo, appunto, dai testi, mentre gli strumenti di elaborazione del linguaggio naturale (ad esempio, concordancer e candidate-terms extractors/ candidate-relations extractors) hanno facilitato l'analisi delle relazioni concettuali e testuali e dunque una migliore comprensione dei generi testuali. I generi testuali sono diventati un elemento essenziale nella ricerca e nella pratica terminologica, specialmente nel contesto della *Corpus-based Terminology* e della *Textual Terminology*. Attraverso questi approcci, è stato possibile individuare caratteristiche terminologiche specifiche per diversi tipi di testi. I generi rappresentano infatti una dimensione fondamentale

attraverso cui i dati terminologici vengono analizzati nei corpora testuali (Faber, L'Homme, 2022). Questo ha reso possibile fare un'analisi comparativa che si concentra su generi, registri e sui rispettivi campi di specializzazione.

Ad esempio, termini che appaiono nei testi scientifici possono avere sfumature di significato e usi distinti rispetto agli stessi termini presenti in testi divulgativi o giuridici. Questo rende evidente come il genere testuale influenzi la multidimensionalità dei concetti, che non si limita a una dimensione linguistica, ma include anche variabili cognitive, sociali e culturali (aspetto, questo, già presente in Sager, 1990).

Nel contesto della terminologia, la multidimensionalità si riferisce alle diverse prospettive e variazioni che un termine può avere all'interno di un dominio specifico, come le varianti linguistiche, le sfumature di significato, le relazioni concettuali e le variabili culturali.

Il concetto di multidimensionalità è presente nell'approccio adottato da M.T. Cabré attraverso la **Teoria delle Porte**. La Teoria delle Porte concepisce il termine come un'unità composta da tre dimensioni fondamentali: quella semiotico-linguistica, quella cognitiva e quella comunicativo-pragmatica. Rispetto ai precedenti approcci semasiologici e onomasiologici, questo modello introduce una terza prospettiva che vede il termine anche in chiave pragmatica e comunicativa. Un approccio che tenga conto di questa interdisciplinarità dovrebbe consentire un'analisi più ricca e sfumata dei termini, rendendo possibile affrontarli in modo multidimensionale (Cabré, 2023).

Le tre dimensioni si intrecciano in tre diverse teorie, evidenziando la complessità delle unità terminologiche, le quali possono passare da un dominio di specialità a un altro, rendendo i termini dinamici e adattabili (Cabré, 2023).

Secondo Cabré, per comprendere pienamente un termine e il suo ruolo nella comunicazione, è necessario analizzarlo da questi tre punti di vista. La combinazione e l'interazione di queste componenti conferiscono al termine il suo valore comunicativo, e l'insieme delle interrelazioni fra di esse costituisce l'"unità terminologica". Durante il processo di comunicazione - sia nella creazione di testi sia nella loro comprensione - diventa essenziale considerare questa triplice struttura per un impiego preciso e funzionale dei termini.

Se il modello di Cabré stabilisce le coordinate fondamentali per un'analisi multidimensionale, la ricerca terminologica contemporanea ha cercato di mappare come queste dimensioni si articolino concretamente in sistemi più vasti. In questa direzione, Ágota Fóris, nel suo contributo 'Terminologia, modelli terminologici e reti' (2008), compie un passo ulteriore proponendo un modello teorico basato sulla rete a invarianza di scala.

La rete a invarianza di scala è un sistema dinamico che consente di descrivere il linguaggio in modo complesso e multidimensionale. Qui, il termine è visto come una rete che contiene *nodi* centrali

(concetti chiave) e *spigoli* (connessioni tra concetti). La rete terminologica è quindi una struttura complessa e multidimensionale, le cui parti sono costituite da reti stesse. Questa rete va oltre le tre componenti fondamentali, includendo aspetti fonetici, grammaticali e semantici, dimostrando come il termine sia immerso in una struttura intricata che garantisce il funzionamento globale della lingua.

Nello specifico, Fóris descrive così i tre aspetti:

- Cognitivo: Il ruolo di un termine dipende dal contesto e dalla rete concettuale a cui appartiene.
- Linguistico: Le unità terminologiche sono anche unità lessicali, con proprietà sintattiche e morfologiche che possono variare.
- Comunicativo: Il termine si adatta alla struttura del discorso e svolge un ruolo chiave nella coesione e coerenza testuale.

Prendendo in considerazione il termine ‘computer’, possiamo esplorare queste dimensioni.
A livello cognitivo, nella nostra mente, il termine ‘computer’ può evocare immagini diverse. Per un programmatore, il termine ‘computer’ è concepito come un'entità astratta e sistemica, più che un oggetto fisico. Viene interiorizzato come ambiente esecutivo, macchina logica e strumento creativo per la formalizzazione del pensiero. Diversamente, per un utente comune, un computer potrebbe rappresentare semplicemente il dispositivo utilizzato per navigare su Internet o per svolgere attività quotidiane come la scrittura di documenti e l'invio di e-mail.
A livello linguistico, il termine ‘computer’ presenta diverse varianti grammaticali. Possiamo riferirci a “i computer” (plurale) o utilizzare l'aggettivo “computazionale” quando parliamo di qualcosa legato al computer, come nel caso di “linguaggio computazionale”, che indica un linguaggio formale specificamente progettato per essere eseguito da un sistema computazionale sotto forma di istruzioni interpretabili dalla macchina.
Infine, a livello comunicativo, il termine ‘computer’ può essere utilizzato in diversi contesti tecnici. Ad esempio, in una conversazione specializzata, potremmo estendere il significato del termine per includere dispositivi come il telefono cellulare e la calcolatrice, o persino i classici calcolatori analogici, come la Macchina di Anticitera o il regolo calcolatore. È possibile associarlo ad altri termini specialistici come “CPU” (unità centrale di elaborazione) o “RAM” (memoria ad accesso casuale).

In questo contesto di interdisciplinarità e multidimensionalità, Kageura (1997) presenta un modello tridimensionale per rappresentare il sistema concettuale relativo al termine “computer”.

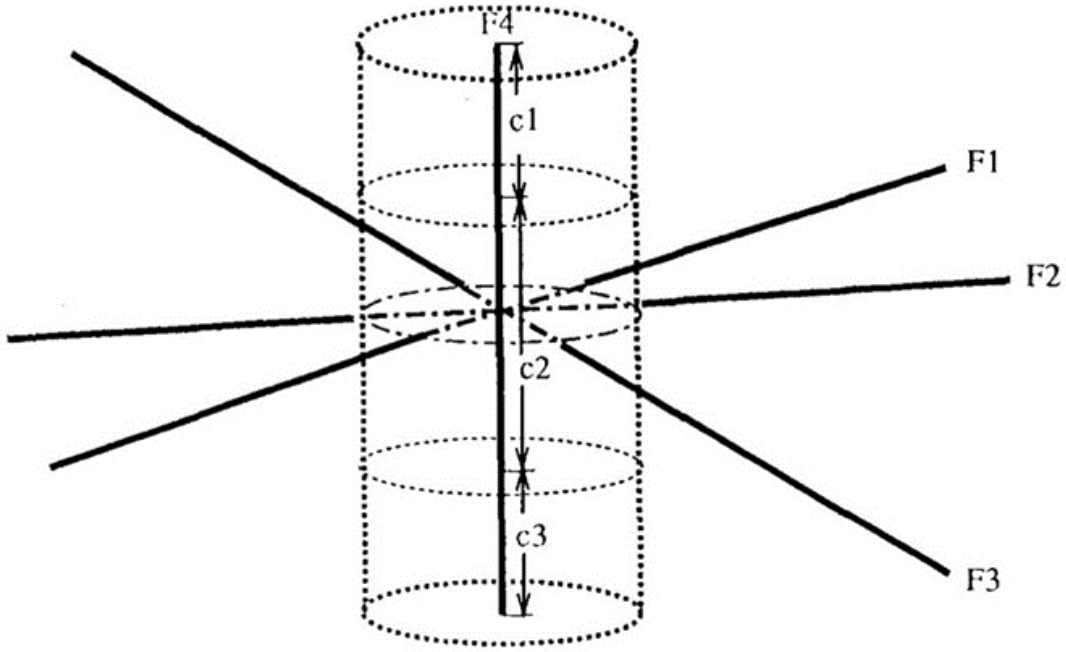


Figure 2 Modello tridimensionale di Kyo Kageura, in Handbook of Terminology Management: Volume 1: Basic Aspects of Terminology Management (Wright & Budin, 1997, pp. 119–132).

Tale modello illustra come diverse caratteristiche si intrecciano per formare un concetto complesso e multidimensionale. Ogni dimensione rappresenta una sfaccettatura o un tipo di caratteristica del computer stesso, come ad esempio il *mode of operation* (modalità di funzionamento), che è un elemento essenziale per distinguere tra diversi tipi di computer. Queste modalità presentano dei range che includono termini come analogico, ibrido e digitale, ognuno dei quali descrive un diverso tipo di computer in base alla sua operatività.

Un'altra dimensione che Kageura prende in considerazione è quella della dimensione, *size*. Ciò porta alla creazione di ulteriori spazi tridimensionali che includono concetti specifici come mainframe e workstation.

In aggiunta a queste, è possibile identificare altre dimensioni, come quella economica, di performance, applicativa ecc.

Nella dimensione economica, si possono analizzare i vari tipi di computer in base al loro costo e alla loro disponibilità. Ad esempio, i computer ad alto costo, come i mainframe, possono essere confrontati con soluzioni più economiche come i laptop o i computer desktop.

La dimensione di performance si concentra su misure quali la velocità di elaborazione, la capacità di memoria e la potenza di calcolo. Infine, la dimensione applicativa riguarda ciò per cui i computer sono progettati e come vengono utilizzati in diversi ambiti. I settori di applicazione possono includere

aree come militare, industriale, medico, educativo, commerciale, scientifico, artistico, di sicurezza, ambientale e di intrattenimento.

Tuttavia, è emerso che tale complessità dell'aspetto multidimensionale all'interno di un sistema concettuale può portare ad un'eccessiva flessibilità e ad un sovraccarico di informazioni (Faber, L'Homme, 2022). I terminologi hanno il compito di saper gestire le informazioni rilevanti in modo efficace. Per questo motivo, il compito dei terminologi di gestire tali flussi informativi in modo efficace ha trovato un alleato fondamentale nell'evoluzione delle tecnologie computazionali.

Negli anni '70 i computer venivano utilizzati principalmente nel campo della ricerca terminologica come archivi di dati; oggi, invece, consentono anche di processare e analizzare grandi quantità di testi in formato digitale. I testi digitali sono facilmente accessibili e manipolabili tramite l'uso di software specializzati che permettono di estrarre informazioni, identificare termini e relazioni, e comprendere il contenuto in maniera automatizzata. Questo utilizzo specialistico dei computer per l'elaborazione e l'analisi dei testi digitali ha trasformato il campo della terminologia, rendendo possibile l'analisi empirica su vasta scala e la progettazione di strumenti avanzati per supportare i terminologi nelle loro attività quotidiane, automatizzando e accelerando molti processi.

Il corpus-based approach facilita l'analisi della variazione terminologica, ossia le diverse forme e varianti di un termine che possono essere utilizzate nel linguaggio del dominio. Questo è particolarmente utile per comprendere la flessibilità e la dinamica del vocabolario specialistico, elementi che spesso si manifestano in modo diverso a seconda del genere testuale in cui i termini vengono utilizzati.

1.5 Le rappresentazioni della conoscenza negli approcci knowledge-based

Gli approcci knowledge-based mirano a strutturare concetti e relazioni in modo che possano essere gestiti, recuperati e interpretati automaticamente dai sistemi intelligenti.

Un contributo fondamentale in questo ambito è rappresentato dalle Terminological Knowledge Bases (TKB), nate negli anni '90 dalla convergenza tra terminologia e ingegneria della conoscenza. Una TKB si distingue per la sua capacità di gestire la duplice natura del termine: quella concettuale, legata alla strutturazione del sapere, e quella linguistica, legata alla sua manifestazione nei testi (Condamines, 2018).

Nel panorama terminologico contemporaneo, tali approcci si sono sviluppati principalmente in due direzioni: l'organizzazione della conoscenza (Knowledge Organization, KO) e la rappresentazione della conoscenza (Knowledge Representation, KR), con le ontologie come elemento centrale di entrambi.

Diversamente dagli approcci corpus-based, che interpretano le sfumature linguistiche dei termini tramite l'analisi del contesto e l'uso di tecniche di Natural Language Processing (NLP), che offrono strumenti per analizzare e interpretare dati testuali empirici, l'approccio knowledge-based mira ad una strutturazione sistematica dei concetti e delle relazioni tra essi.

Gli approcci knowledge-based si focalizzano sulla costruzione di sistemi in grado di elaborare, gestire e inferire nuove informazioni a partire, appunto, da una base di conoscenza (Knowledge Base, KB) strutturata. In questi sistemi, la conoscenza non è trattata come un semplice insieme di dati, ma come una rete complessa di concetti, relazioni e regole che descrivono un dominio specifico. Una base di conoscenza è il cuore degli approcci Knowledge-Based.

La Knowledge Organization, pur non essendo orientata all'automazione in ambito terminologico, si concentra sulla strutturazione della conoscenza in modo che sia comprensibile e accessibile agli utenti, fungendo da base per la creazione di KB. Questi sistemi sono progettati per inferire nuove informazioni o rispondere a domande in base ai dati organizzati in una base di conoscenza.

La KO, come campo di studio, si occupa di analizzare i processi e i sistemi utilizzati per organizzare concetti e documenti, includendo attività come descrizione, indicizzazione e classificazione, utilizzati nelle biblioteche, negli archivi e in altre istituzioni di memoria (Hjørland, 2008).

Il termine "Organizzazione della Conoscenza" ha avuto origine nel campo bibliotecario intorno al 1900 grazie a studiosi come Charles A. Cutter ed Ernest Cushington Richardson, e si è consolidato con le teorie di W.C. Berwick Sayers e Henry Bliss. Ogni disciplina, tra cui psicologia, linguistica, filosofia e intelligenza artificiale, studia i concetti da diverse prospettive, ponendo l'accento su aspetti differenti, ma tutte affrontano le stesse problematiche epistemologiche fondamentali, con un grande

beneficio per ciascuna disciplina quando viene sviluppata una teoria forte in questo campo (Hjørland, 2008).

Gli approcci knowledge-based sono principalmente orientati verso la gestione e l'elaborazione *automatica* delle informazioni. In particolare, la KO ha come principale obiettivo quello di rendere la conoscenza accessibile e comprensibile agli utenti attraverso l'organizzazione e la classificazione delle informazioni.

Pur trattando la conoscenza in modo più "umano" (ad esempio, per facilitare la ricerca e la navigazione), la KO può essere vista come una fase preliminare nella creazione di una base di conoscenza per sistemi knowledge-based.

La Knowledge Organization è strettamente legata all'emergere dei Knowledge Organization Systems (KOS) (Hjørland, 2016), che comprendono oltre a schemi di classificazione, thesauri, tassonomie, anche le ontologie.

Sebbene i KOS non comprendano dati empirici come fatti e ipotesi, questi sono archiviati in una knowledge base o in un database, e resi accessibili tramite il KOS stesso (Soergel, 2009).

Questi sistemi consentono di mappare e connettere concetti in modo coerente e navigabile. I KOS sono infatti un insieme ristretto di termini che si riferiscono ad un dominio di conoscenza specifico. Un KOS si occupa di concetti, categorie, classi, relazioni tra di essi e termini o altre designazioni per rappresentare questi concetti e relazioni (Soergel, 2009).

Nell'attuale "Era del digitale", dall'avvento di Internet alla nascita del World Wide Web e l'utilizzo massivo dei Social Network, la comunicazione è sempre più ambigua a causa della diffusione di informazioni non strutturate, che appartengono al nostro linguaggio naturale. Per far sì che le macchine possano interpretare questi dati, è importante strutturarli ed individuarne le relazioni e le proprietà. L'uso di vocabolari controllati nei KOS rappresenta un modo per evitare l'ambiguità del linguaggio naturale.

Il linguaggio controllato, infatti, è composto da termini standardizzati, che riducono la variabilità dei termini e rendono la ricerca più accurata. linguaggio speciale, che si diversifica dal linguaggio di uso comune e per questo, necessita di essere contraddistinto, standardizzato, delimitato, definito e soprattutto rappresentato.

Mentre il linguaggio naturale è ricco e flessibile, presenta ambiguità che complicano la ricerca di informazioni; i vocabolari controllati, quindi, fungono da filtro, consentendo di tradurre i concetti di linguaggio naturale in termini univoci e ben definiti, migliorando l'efficienza dei sistemi di recupero.

La creazione di KOS è nata, quindi, per rispondere alla crescente necessità di gestire la produzione scientifica e culturale, diventando una sottodisciplina delle scienze dell'informazione e delle

biblioteche (LIS). Pur differendo strutturalmente, tutti i KOS condividono l'obiettivo di migliorare l'accesso e il recupero delle informazioni (Mazzocchi, 2018).

Nella classificazione dei KOS, Hodge (2000) propone tre categorie principali:

- Liste: Sistemi semplici e lineari, come dizionari e file di autorità.
- Classificazioni e Categorie: Sistemi gerarchici che organizzano concetti in gruppi tematici, come tassonomie e schemi di classificazione.
- Liste di Relazioni: Sistemi complessi che enfatizzano le connessioni tra concetti, come thesauri, reti semantiche e ontologie.

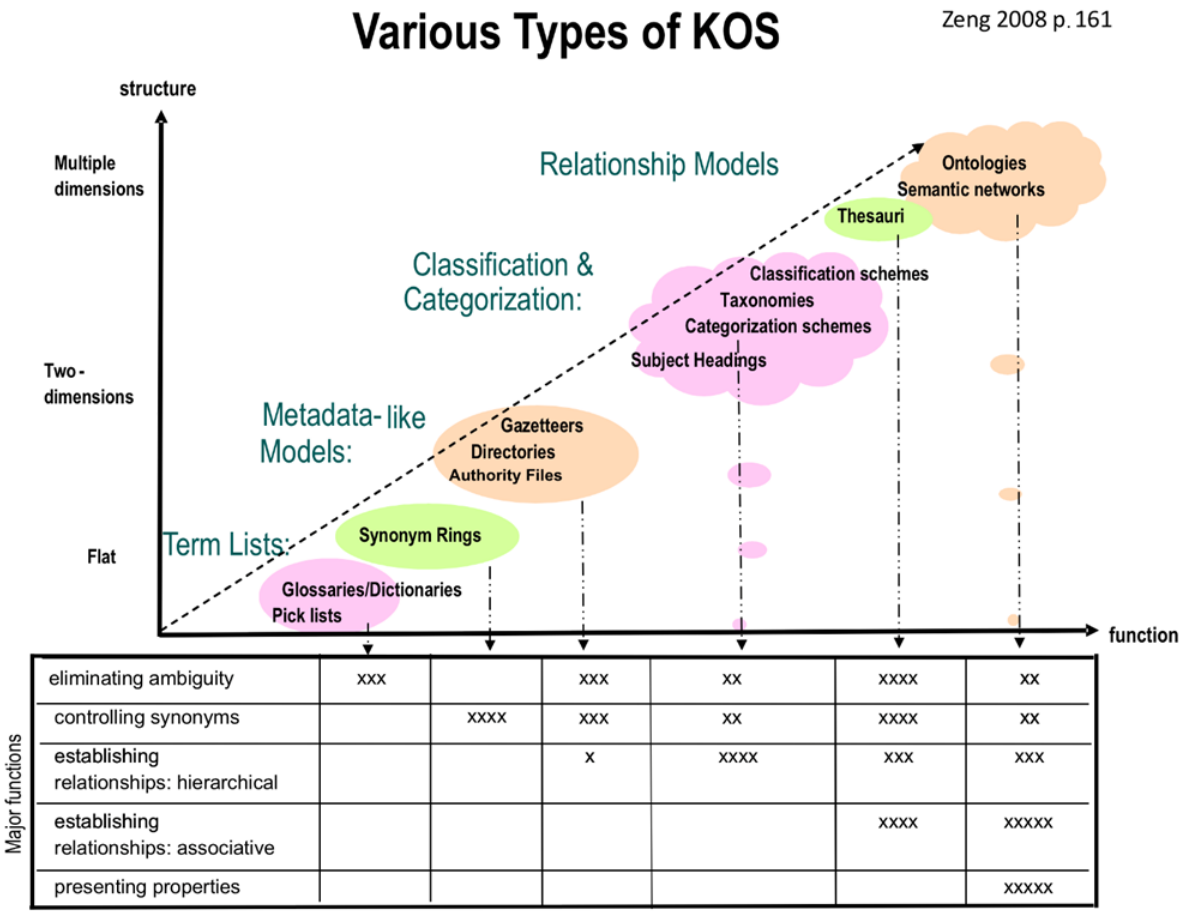


Figure 3 I KOS di M. Zeng, 2009.

I vocabolari controllati, come thesauri e tassonomie, sono strumenti chiave nei KOS, poiché forniscono un controllo linguistico che semplifica la ricerca e il recupero delle informazioni. I thesauri organizzano il vocabolario di un dominio in modo gerarchico, includendo sinonimi e relazioni semantiche, mentre le tassonomie offrono una classificazione gerarchica dei concetti, facilitando una navigazione intuitiva attraverso il dominio.

I sistemi KOS, come abbiamo visto, possono essere più o meno organizzati, più o meno formalizzati, più o meno controllati e di conseguenza più o meno complessi. Le **ontologie** si pongono ai vertici di questa piramide.

In una fase iniziale si riteneva che il trasferimento della conoscenza umana in una base di conoscenza potesse avvenire attraverso un processo relativamente semplice di raccolta, organizzazione e implementazione dei contenuti. Questa concezione nasceva dall'assunto, caratteristico dei primi approcci simbolici, che ogni forma di conoscenza potesse essere resa esplicita e tradotta in proposizioni logiche o regole formali, così da poter essere trattata automaticamente da un calcolatore. Ne derivava l'idea che il compito essenziale fosse quello di "trascrivere" i saperi specialistici in schemi strutturati e algoritmi, riducendo la complessità del sapere umano a rappresentazioni pienamente formalizzate. Con il tempo, tuttavia, tale impostazione si è rivelata riduttiva: non tutta la conoscenza è esplicita né immediatamente trasferibile in linguaggi formali. È emerso con chiarezza il ruolo della conoscenza implicita e tacita, fatta di abilità pratiche, intuizioni ed elementi contestuali che non si prestano a una codifica diretta. Inoltre, la conoscenza non è statica ma dinamica: si trasforma nei contesti sociali e comunicativi, rendendo insufficiente una concezione puramente "depositaria" delle basi di conoscenza. La sfida non consiste dunque soltanto nel raccogliere informazioni, ma nel modellare la complessità del sapere umano, distinguendo tra ciò che è formalizzabile e ciò che rimane necessariamente legato all'esperienza e all'interpretazione.

In risposta a questa sfida, negli anni '90 è emerso un nuovo campo chiamato Ingegneria della Conoscenza (Knowledge Engineering, KE), una disciplina che ha cercato di tradurre le competenze umane in modelli computabili. L'introduzione di modelli di rappresentazione della conoscenza ha permesso di formalizzare la conoscenza in modo tale che le macchine potessero compiere inferenze e generare nuove conoscenze. Le ontologie e altri approcci avanzati per la modellazione della conoscenza hanno risposto a tali limiti, offrendo soluzioni più sofisticate per la rappresentazione e gestione delle conoscenze (Studer, Benjamins, Fensel, 1998).

L'introduzione di strumenti come CODE (Conceptually Oriented Design Environment) ha rappresentato un passo significativo in questa evoluzione. CODE, sviluppato presso il Laboratorio di Intelligenza Artificiale dell'Università di Ottawa, facilita l'analisi concettuale, supportando l'uso dell'intelligenza artificiale e delle tecniche di linguaggio naturale per la gestione delle informazioni. Questo strumento, che risponde alla crescente necessità di supporto informatico nella terminologia, rappresenta un passo importante verso la costruzione di basi di conoscenza multifunzionali per la gestione delle informazioni e la traduzione automatica (Skuce, Meyer, 1990).

L'interazione tra l'ingegneria della conoscenza e la terminologia ha quindi portato alla creazione di basi di conoscenza terminologiche-ontologiche, dove i concetti vengono rappresentati attraverso reti

concettuali, al fine di fornire una rappresentazione più ricca e dinamica della conoscenza (Faber, L'Homme, 2022).

Le ontologie, infatti, rappresentano una forma avanzata, "evoluta", di KOS (M.T. Biagetti, 2020), che consentono non solo di organizzare la conoscenza in modo coerente, ma anche di rappresentarla formalmente, come avviene nei sistemi knowledge-based.

I sistemi di organizzazione della conoscenza si caratterizzano per la capacità di rappresentare in modo esplicito le relazioni tra concetti, spesso attraverso l'uso di markup di metadati che ne facilitano l'elaborazione automatizzata. Questi sistemi possono essere utilizzati in sinergia con strumenti di gestione terminologica oppure operare in modo indipendente (Faber, L'Homme, 2022).

L'obiettivo cambia: si passa ad una concettualizzazione e specificazione formale, dunque interpretabile dalle macchine ed esplicita, condivisa dalla comunità di riferimento. Se la loro struttura è complessa, bisogna considerarli non più semplicemente sistemi di organizzazione, ma sistemi di rappresentazione.

Il termine "ontologia" è l'etichetta più recente attribuita ad alcuni sistemi di organizzazione della conoscenza. Le ontologie vengono sviluppate come modelli concettuali specifici dalla comunità della gestione della conoscenza. Le ontologie, in particolare, si distinguono per la loro capacità di descrivere concetti e relazioni in modo rigoroso, facilitando la manipolazione automatica delle informazioni tramite inferenza e ragionamento.

In letteratura, le ontologie sono state descritte come «un insieme di singole istanze di classi che costituisce una base di conoscenza»[4]. Queste definizioni mettono in evidenza la duplice natura delle ontologie: da un lato ancorate alle istanze concrete, dall'altro orientate alla modellizzazione astratta e al ragionamento automatico. A differenza dei thesauri e delle tassonomie, le ontologie e le reti semantiche sono approcci più complessi che passano dalla semplice organizzazione alla *rappresentazione* formale della conoscenza.

La rappresentazione della conoscenza (KR), recentemente evoluta in **Knowledge representation and Reasoning** (KR&R), è un settore dell'intelligenza artificiale che si distingue dalla KO poiché si concentra sulla codifica formale dei concetti e delle relazioni per permettere inferenze e ragionamenti automatici. ello specifico, la KR&R mira a fornire «descrizioni di alto livello di un dominio, utilizzabili dalle applicazioni intelligenti per trarre conseguenze implicite da conoscenze esplicitamente rappresentate»[5].

Secondo R. Brachman e H. Levesque (2009), «la KR si occupa di come la conoscenza possa essere rappresentata simbolicamente e manipolata in modo automatizzato da programmi di ragionamento».

[4] Traduzione nostra (Nov & McGuinness, 2001, 2 What is in an ontology?, p.8)
[5] Traduzione nostra (Baader *et al.*, 2002, 1.1 Introduction, p.1)

Sia KO che KR utilizzano le ontologie per definire e standardizzare concetti e categorie, ma mentre la KO si limita alla classificazione e supporta solo interrogazioni basate sulle proprietà dei documenti, la KR va oltre, mirando a rappresentare formalmente la conoscenza in modo che i sistemi automatici possano compiere inferenze e rispondere in modo pertinente. A questo proposito, KR risulta essere molto potente e altamente espressiva, con un potenziale teoricamente molto ampio, poiché, grazie all'uso di ontologie descrittive, supporta interrogazioni basate su qualsiasi proprietà delle entità (Giunchiglia *et al*., 2014).

La KR è definita come «la parte dell'IA che si occupa del pensiero e di come il pensiero contribuisce ai comportamenti intelligenti […] ciò che dobbiamo studiare è ciò che sanno gli esseri umani» (Brachman e Levesque, 2009).

Gran parte dell'IA si concentra sulla costruzione di sistemi basati sulla conoscenza (KB), la cui capacità deriva in parte dal ragionamento su conoscenze rappresentate in modo esplicito. Un chiaro esempio di tale approccio è costituito dai sistemi esperti, ma le basi di conoscenza sono utilizzate anche in altre aree come la comprensione del linguaggio, la pianificazione, la diagnosi e l'apprendimento automatico (Brachman e Levesque, 2009).

In questi approcci, le ontologie svolgono un ruolo centrale nella modellazione dei domini di conoscenza, fornendo una struttura che permette ai sistemi informatici di comprendere, elaborare e ragionare su informazioni complesse.

Sebbene si sia discusso ampiamente della raffinatezza e complessità delle ontologie rispetto ad altri sistemi di rappresentazione della conoscenza, un'ulteriore distinzione significativa emerge all'interno del concetto stesso di ontologia. Giunchiglia *et al.* (2006) hanno evidenziato la differenza tra i modelli ontologici utilizzati nella Knowledge Organization e nella Knowledge Representation.

In particolare, nella KO, si utilizzano ontologie di classificazione, che organizzano concetti e termini in categorie gerarchiche per facilitare la ricerca e la classificazione dei documenti. Queste ontologie sono principalmente orientate alla strutturazione e al recupero delle informazioni.

Al contrario, nella KR si impiegano ontologie descrittive, che rappresentano in modo dettagliato le entità del mondo reale, le loro proprietà e le relazioni tra di esse. L'obiettivo della KR è supportare inferenze e ragionamenti più complessi, che vanno oltre la semplice organizzazione dei concetti. La maggiore capacità espressiva e il potenziale inferenziale delle ontologie descrittive rendono la KR più adatta a scenari in cui la comprensione semantica e l'elaborazione avanzata delle informazioni sono essenziali. Nel lavoro di Giunchiglia *et al.* (2006), vengono riportati esempi pratici per illustrare le differenze tra le due tipologie di ontologie, utilizzando domini differenti:

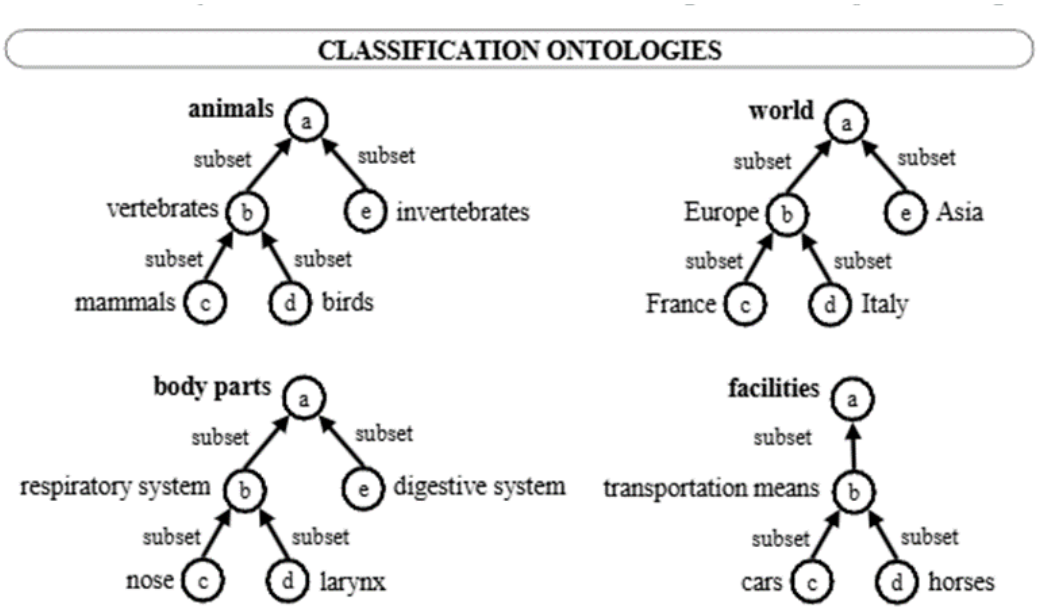


Figure 4 Esempio di ontologia di classificazione, come illustrato da Giunchiglia et al. (2006).

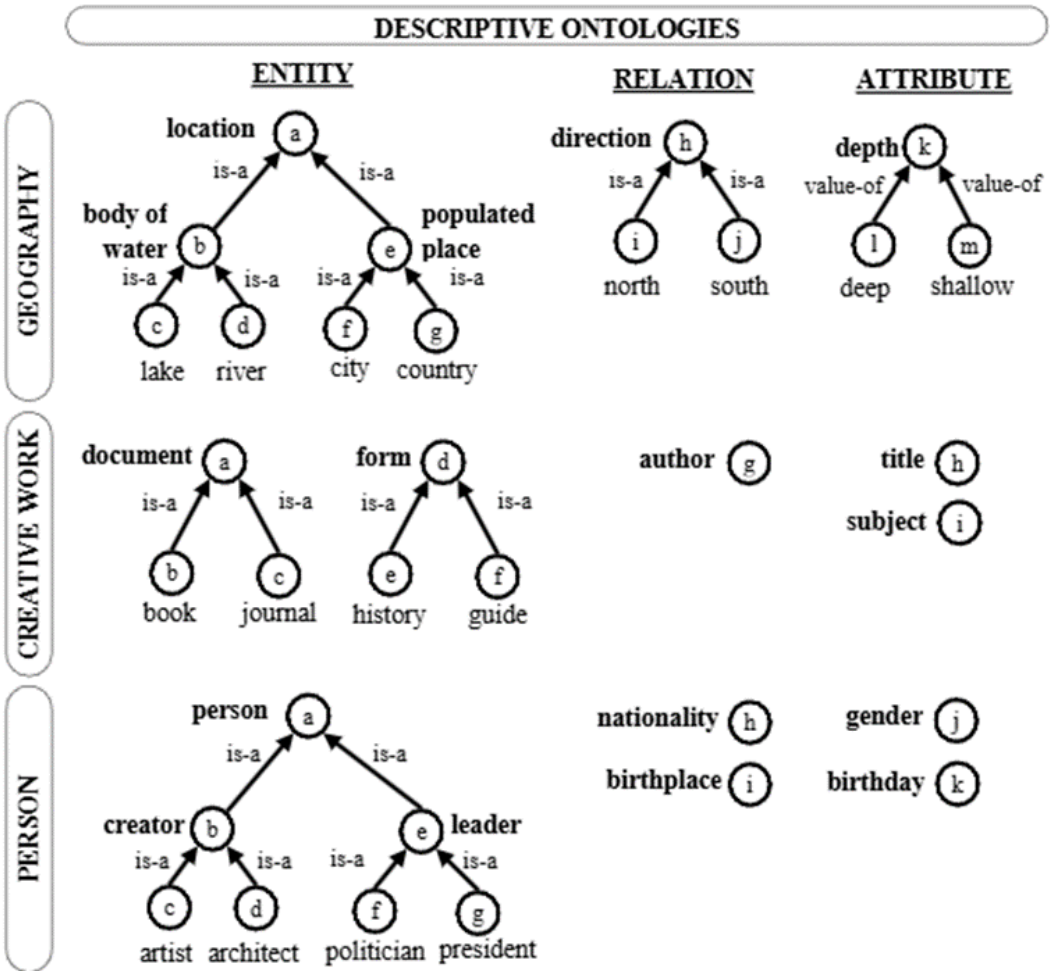


Figure 5 Esempio di ontologia descrittiva, come illustrato da Giunchiglia et al. (2006).

L'Organizzazione della Conoscenza impiega ontologie di classificazione per descrivere e ricercare documenti, dove i termini rappresentano gruppi di documenti e le relazioni gerarchiche indicano

connessioni di superclasse/sottoclasse (BT/NT) tra i termini. Gli individui sono i documenti stessi. Questo sistema è utilizzato in classificazioni, thesauri e liste di termini (authority files, glossari e dizionari).

Nel loro articolo del 2006, Giunchiglia *et al.* utilizzano l'esempio del termine "cavalli" in *Figura 4* per mostrare come, essendo posizionato sotto la categoria "mezzi di trasporto", possa riferirsi a documenti sui cavalli e, al contempo, stabilire una connessione semantica tra cavalli e trasporti. Questo approccio dimostra come il ragionamento nell'ambito della Knowledge Organization (KO) sfrutti la transitività delle relazioni per agevolare la ricerca semantica.

La KR, invece, adotta ontologie descrittive per rappresentare e analizzare entità del mondo reale. In queste ontologie, i termini rappresentano insiemi di entità reali, le relazioni "is-a" indicano una connessione di sottogruppo, e gli individui sono entità reali. Un esempio potrebbe essere 'cavallo is-a animale', che implica che i cavalli siano una sotto-categoria degli animali. Le ontologie descrittive offrono conoscenza riguardo a classi, attributi e relazioni, e KR ha un obiettivo più ampio rispetto alla KO, in quanto considera i documenti come una particolare entità e la ricerca di documenti come un tipo di ragionamento basato su queste ontologie (Giunchiglia *et al.* 2014).

Anche se le ontologie descrittive e la Knowledge Representation (KR) perseguono obiettivi differenti, entrambe sono rilevanti nel contesto del web semantico, dove il W3C ha sviluppato il Simple Knowledge Organization System (SKOS). Questo sistema consente di rappresentare thesauri, schemi di classificazione e tassonomie all'interno del web, con lo scopo di migliorare l'interoperabilità dei dati e nel rendere più efficaci le ricerche, facilitando la comunicazione tra sistemi diversi.

In sintesi, le ontologie rappresentano una componente essenziale degli approcci knowledge-based e del web semantico, fornendo un quadro formale e strutturato per la rappresentazione e l'elaborazione della conoscenza. La loro continua evoluzione, unitamente all'uso di tecniche avanzate come i Knowledge Graphs (o Grafi della conoscenza), sta trasformando il modo in cui le informazioni vengono gestite, condivise e utilizzate, con applicazioni che spaziano dall'intelligenza artificiale alla gestione delle informazioni nel web.

1.6 L'ontologia come strumento di rappresentazione della conoscenza

Secondo Cabré (1999), l'ontologia è la disciplina che si occupa di analizzare gli oggetti, la loro collocazione e le relazioni che li legano tra loro. In questo senso, le ontologie forniscono una struttura concettuale formale per i termini utilizzati in un dominio specifico, definendo concetti chiave e relazioni tra termini, e offrendo una base comune per la loro standardizzazione.

Come definito nel *W3C Linked Data Glossary* (2013), un'ontologia è "un modello formale che consente di rappresentare la conoscenza per un dominio specifico. Un'ontologia descrive i tipi di entità esistenti (classi), le relazioni tra di esse (proprietà) e i modi logici in cui tali classi e proprietà possono essere utilizzate insieme (assiomi)". Le ontologie, quindi, organizzano e descrivono concetti e relazioni di un dominio, fornendo una base per la comprensione e condivisione della conoscenza.

In un approccio basato sulla conoscenza (knowledge-based), le ontologie vengono utilizzate per rappresentare esplicitamente la conoscenza, permettendo ai sistemi di intelligenza artificiale (IA) e ad altri sistemi informatici di ragionare, inferire nuove informazioni e rispondere a domande complesse in modo coerente.
Un'ontologia si distingue per tre caratteristiche principali:

- Concettualizzazione esplicita: i concetti e le loro restrizioni devono essere chiaramente definiti.
- Formalità: deve essere interpretabile da un sistema informatico.
- Condivisione: rappresenta una conoscenza consensuale, non soggettiva.

Un'ontologia è composta da classi, proprietà (note anche come slot, ruoli) che descrivono le caratteristiche e gli attributi di ciascun concetto, e restrizioni sugli slot (chiamate anche *faccette* nei sistemi a frame o restrizioni di ruolo).
Le classi, spesso chiamate concetti, possono avere sottoclassi che rappresentano concetti più specifici rispetto alla superclasse (Noy, McGuinnes, 2001).

R. Studer, V. R. Benjamins, D. Fensel, (1998) identificano diversi tipi di ontologie:

- Ontologie di dominio: rappresentano la conoscenza specifica di un determinato settore, come nel caso della medicina, della meccanica o di altri ambiti specialistici.
- Ontologie generiche: sono strutture concettuali che possono essere applicate trasversalmente a diversi domini, come ad esempio le relazioni parte-tutto, che sono valide in molteplici contesti.

- Ontologie applicative: combinano ontologie di dominio con approcci metodologici per affrontare e risolvere problemi pratici in contesti specifici.
- Ontologie rappresentazionali: forniscono entità generali e astratte per la rappresentazione della conoscenza, senza essere legate a un dominio specifico.
- Ontologie di compito/metodo: si concentrano su concetti specifici per la realizzazione di compiti concreti, come la diagnosi, o per l'applicazione di metodi di problem-solving.

Attraverso approcci come il *top-down* e il *bottom-up* e utilizzando le domande di competenza, meglio conosciute come Competency Questions (CQ) si può guidare la progettazione ontologica per soddisfare le esigenze degli utenti e migliorare le applicazioni knowledge-based. Le CQ identificano domande chiave a cui il sistema ontologico deve rispondere, chiarendo obiettivi, risultati attesi ed esigenze da soddisfare[6].

Due figure centrali in questo contesto sono l'ingegnere dell'ontologia, che possiede competenze scientifiche e tecnologiche per il lavoro di conceptual modelling, e gli esperti di dominio, in genere clienti o partner industriali. Questo approccio aiuta a organizzare il lavoro, a fissare al meglio i requisiti dell'ontologia scegliendo un metodo appropriato (ad esempio, un approccio top-down o bottom-up) e, soprattutto, a soddisfare le necessità dell'utente e dell'esperto di dominio che consulta le fonti.

Con l'approccio bottom-up, si parte dall'osservazione e definizione di classi specifiche per poi procedere a generalizzazioni; con l'approccio top-down, si parte da classi generali e si scende verso definizioni più specifiche. È possibile adottare un approccio misto.

È, inoltre, spesso utile ricercare ed eventualmente utilizzare ontologie preesistenti per verificare se possano essere affinate o estese, soprattutto quando il sistema deve interagire con altre applicazioni già basate su ontologie o vocabolari controllati specifici. Molte ontologie sono disponibili in formato elettronico e possono essere importate (Noy, McGuinness, 2001).

[6] Ulteriori approfondimenti: Grüninger, M., Fox, M.S., *Methodology for the design and evaluation of ontologies*, 1995.

1.6.1 Ontologia e Terminologia

Negli ultimi anni, la terminologia ha vissuto una significativa evoluzione, spostandosi da un approccio basato puramente su lessicografia tecnico-scientifica a uno caratterizzato da un'attenzione crescente alla concettualizzazione dei termini e alla loro applicazione pratica (Roche, 2008).

In effetti, la terminologia, sebbene influenzata dalla linguistica, non può essere ridotta solo a una disciplina linguistica, poiché deve necessariamente anche considerare aspetti epistemologici e concettuali.

Proprio questo cambiamento di paradigma ha portato la terminologia a intersecarsi con le ontologie, due ambiti che, pur essendo correlati, sono distinti.

Se da un lato, infatti, la terminologia ha storicamente concentrato l'attenzione sullo studio e sulla gestione dei termini utilizzati in specifici domini del sapere, dall'altro le ontologie hanno aperto nuove prospettive, mirate a rappresentare in modo più formale e strutturato la conoscenza all'interno di questi stessi domini. Ne deriva che entrambi gli strumenti - ontologie e terminologie - risultino oggi essenziali per migliorare la comunicazione, sebbene operino su piani differenti.

Questa convergenza, come abbiamo visto nel paragrafo precedente, è fondamentale per applicazioni nell'Intelligenza Artificiale (IA), NLP e sistemi knowledge-based, facilitando il ragionamento e l'integrazione di dati eterogenei. Le ontologie, in particolare, aiutano a facilitare il dialogo tra software e utenti umani, mentre le terminologie rendono più chiara la comunicazione tra esperti in un campo specifico.

Tuttavia, nonostante questa sinergia, permane una distinzione metodologica netta: la terminologia si concentra sulla raccolta, definizione e gestione di termini specifici utilizzati in un determinato dominio o disciplina. L'ontologia, d'altra parte, mira a rappresentare la conoscenza di un dominio attraverso una struttura formale che possa essere compresa e processata dai computer. L'obiettivo è creare un modello concettuale che definisca le entità, le loro proprietà e le relazioni tra di esse, consentendo ragionamenti automatizzati e inferenze. È orientata alla conoscenza e alla sua rappresentazione, piuttosto che alla varietà linguistica. L'ontologia è definita come una specificazione esplicita di una concettualizzazione (Gruber, 1993), ovvero una specifica formale ed esplicita di una concettualizzazione condivisa (Studer, Benjamins e Fensel, 1998). Sotto il profilo storico ed etimologico, il termine 'ontologia' appare per la prima volta nel 1606 con la pubblicazione dell'opera di Jacob Lorhard (Øhrstrøm, Schärfe e Uckelman, 2008). Il termine deriva dal greco ὄντος, *òntos* (genitivo singolare del participio presente del verbo εἶναι, *èinai*, «essere») e da λόγος, *lògos* («discorso»), traducibile come "scienza dell'essere" (Bonomi, 2004–2008). Mutuato dalla filosofia, il termine ha acquisito un significato nei sistemi knowledge-based, dove ciò che "esiste" è esattamente ciò che può essere rappresentato (Gruber, 1993). L'ontologia rappresenta, quindi, uno studio

sistematico dell'esistenza, un concetto espresso da Parmenide di Elea con l'affermazione probabilmente più famosa della storia della filosofia: "L'essere è e non può non essere". Il termine è stato quindi mutuato dalla filosofia, in cui un'ontologia rappresenta una descrizione sistematica dell'Esistenza.

In questo contesto, le ontologie, intese come strutture concettuali formali, risultano fondamentali per l'organizzazione e la chiarificazione dei concetti che i termini specialistici rappresentano. Esse non solo forniscono una base teorica solida per la standardizzazione terminologica, ma contribuiscono anche a creare una mappa semantica che facilita la coerenza e la precisione nella comunicazione all'interno di specifici domini di conoscenza.

I primi studi terminologici si sono concentrati sulla standardizzazione del linguaggio, dunque calandola principalmente in contesti umani per migliorare la comunicazione tra esperti, per l'interpretazione dei testi, la traduzione e la documentazione. La convergenza con l'ontologia non è però casuale. Le ontologie possono trarre vantaggio dall'arricchimento linguistico per una rappresentazione concettuale più precisa e una comunicazione più efficace con gli utenti umani. Le ontologie offrono un'opportunità per estendere questa standardizzazione al mondo computazionale, consentendo una rappresentazione della conoscenza che possa essere compresa e utilizzata dai sistemi intelligenti e applicazioni computazionali, come il ragionamento automatizzato, il recupero di informazioni, l'integrazione di dati da diverse fonti e l'elaborazione del linguaggio naturale.

In questo contesto, le ontologie possono essere viste come un'evoluzione naturale degli studi terminologici, una forma più sofisticata e dinamica di organizzare e utilizzare le informazioni linguistiche.

L'introduzione delle ontologie negli studi terminologici rappresenta, quindi, non solo un ampliamento del campo, ma anche un modo per sfruttare nuove tecnologie e metodologie al fine di creare modelli di conoscenza più complessi e completi, in grado di migliorare la comunicazione tra sistemi e tra umani.

Nonostante le differenze esistenti tra questi due campi, vi sono sforzi crescenti per creare sinergie e sfruttare le somiglianze e consentire ai sistemi di IA di "comprendere" e utilizzare la conoscenza in aree come la Linguistica Computazionale, l'Elaborazione del Linguaggio Naturale (NLP) e l'Intelligenza Artificiale (IA).

In questi settori, l'integrazione di risorse linguistiche curate ha dimostrato di apportare notevoli benefici.

L'uso delle ontologie si estende, inoltre, in diversi campi come il knowledge management, l'integrazione dei dati, il content management, l'e-commerce, l'information filtering e il problem solving (Capuano, 2005) e soprattutto nel Web semantico.

1.6.2 Ontologie e Web Semantico

L'integrazione delle ontologie nel panorama digitale trova la sua massima espressione nel progetto del Semantic Web, teorizzato da Tim Berners-Lee nel 1998. Il Semantic Web viene inteso non come un Web separato, ma "un'estensione di quello attuale" (Berners-Lee et al, 2001) in chiave semantica, verso un sistema "machine-processable", proponendo strumenti come linguaggi standard e vocabolari condivisi per rappresentare la conoscenza, tra cui l'uso di ontologie (Capuano, 2005).

Sotto il profilo evolutivo, questa transizione segna il passaggio dal Web 2.0 - centrato sulla partecipazione attiva degli utenti e sui contenuti generati dai social media – ad una nuova fase definitiva che lo stesso Berners-Lee identificava originariamente come Web 3.0. In questo contesto, il termine designava un'architettura basata sull'integrazione di tecnologie semantiche - quali ontologie, vocabolari condivisi e linguaggi standard - finalizzate a migliorare l'interoperabilità tra sistemi e l'automazione dei processi informativi.

Tuttavia, l'univocità di questa definizione è venuta meno a partire dal 2014, quando il termine Web 3.0 è stato riutilizzato in un'accezione differente, non direttamente sovrapponibile a quella del Web Semantico. Introdotto da Gavin Wood nel 2014, esso viene oggi impiegato per designare un insieme eterogeneo di pratiche e infrastrutture digitali basate su tecnologie blockchain e su meccanismi di governance decentralizzata. In questo senso, Web 3.0 non corrisponde a un paradigma teorico unitario né a una specifica architettura semantica del Web, ma piuttosto a un costrutto discorsivo e tecnologico in fase di definizione, caratterizzato da una pluralità di interpretazioni.

Una definizione frequentemente citata in ambito divulgativo è quella proposta da Packy McCormick (2021), secondo cui «*Web3 is the internet owned by the builders and users, orchestrated with tokens»*. Tale formulazione mette in evidenza il ruolo dei token come elemento di coordinamento economico e organizzativo delle piattaforme decentralizzate, ma non fornisce una specificazione concettuale o semantica paragonabile a quella elaborata nell'ambito del Web Semantico.

Il concetto di semantica lo troviamo in diversi campi e in diverse modalità (Shethet *et al* 2005). In questo contesto si fa riferimento alla semantica esplicita, ovvero attraverso la rappresentazione della conoscenza in un formato strutturato e formalmente definito, per una comprensione non ambigua.

Il Semantic Web utilizza la semantica formale per attribuire significato ai dati, rappresentandoli in modo inequivocabile e permettendo alle macchine di elaborarli, attraverso diverse tecnologie (RDF, OWL, SKOS, SPARQL) e in un ambiente in cui le applicazioni possono interrogare i dati, trarre conclusioni e integrare informazioni da diverse fonti.

Tim Berners-Lee stesso ha sottolineato questa evoluzione, distinguendo il *Web dei documenti*, basato su pagine HTML pensate per essere lette dagli esseri umani, dal *Web dei dati*, progettato per essere elaborato e interpretato dalle macchine: «*We have HTML for the Web of documents that we can read.*

We have the Semantic Web for a Web of linked data… much more complicated. It's much more powerful than the Web of documents» (Berners-Lee, 2012). L'enfasi passa, dunque, dalle persone (documenti, Web di documenti) alle macchine (dati, Web di dati).

> "Il web di documenti (o web ipertestuale) è il web in cui i testi sono pubblicati in pagine HTML (HyperText Markup Language); il linguaggio con cui vengono scritti – l'HTML. Gli elementi di taggatura servono per la strutturazione del testo (o del contenuto) ai fini della sua rappresentazione visiva tramite un browser. Gli oggetti rappresentati sul web di documenti sono creati per essere letti, interpretati e, dunque, utilizzati dagli umani, non dalle macchine" (M. Guerrini, 2013).

Il Web Semantico consente l'interoperabilità semantica, utilizzando vocabolari condivisi e linguaggi formali per modellare entità e relazioni e "permettendo una migliore collaborazione tra computer e persone" (T.Berners-Lee et al, 2001). Grazie a strumenti come le reti semantiche e i Knowledge Graphs, le ontologie giocano un ruolo centrale nel collegare e armonizzare i dati, creando un ecosistema di informazioni interconnesse e accessibili sia agli utenti umani che ai sistemi intelligenti. Non solo, "le macchine sono molto più in grado di elaborare e "comprendere" i dati"[7].

Tra gli standard per garantire l'interoperabilità sintattica (formati dei metadati: grammatica e struttura dei dati condivisi), il Web Semantico si è basato nella sua proposta originale sull'XML (*eXtensible Markup Language*).

XML permette a chiunque di definire i propri tag, ovvero etichette utilizzate per annotare documenti e pagine Web. Tuttavia, XML non è sufficiente per attribuire un significato alle strutture che descrive. Questo compito è svolto da RDF (Resource Description Framewor) (T.Berners-Lee et al, 2001).

Per l'interoperabilità semantica il Web Semantico utilizza, dunque, il modello di rappresentazione RDF, il linguaggio RDFS per la definizione di vocabolari di base, e il linguaggio ontologico OWL per descrizioni semantiche più complesse. Il W3C definisce l'RDF come un modello standard per la rappresentazione dei dati sul Web (W3C, 2026). RDF descrive qualsiasi risorsa mediante sequenze di triple (talvolta anche definite *triplette*) che specificano informazioni sulla risorsa. Ogni tripla è costituita da soggetto, predicato e oggetto, dove il soggetto è la risorsa stessa che viene, mentre il predicato esprime una proprietà o relazione, e l'oggetto rappresenta il valore della proprietà o un'altra risorsa con cui il soggetto è collegato. RDF estende la struttura di collegamento del Web per utilizzare identificatori universali chiamati URI (Uniform Resource Identifier). Gli URI identificano le risorse

[7] Traduzione nostra (Berners-Lee *et a*l, 2001)

nelle triplette, ovvero soggetto, predicato e oggetto per garantire che i concetti rappresentati non siano semplici parole, ma abbiano un significato univoco accessibile sul Web.

1.6.3 RDF

RDF è uno standard sviluppato dal W3C, fondamentale per descrivere e collegare risorse sul Web in modo strutturato, e garantire l'interoperabilità tra applicazioni che gestiscono dati in formato *machine-readable*.

Spesso, in linguaggio naturale, utilizziamo lo stesso termine per riferirci a concetti diversi in contesti differenti. Nel concepire il Semantic Web, Tim-Berners Lee osservava che i dati presenti sul Web, pur essendo leggibili dalle macchine, non sono necessariamente comprensibili per esse.

Per esempio, come suggerito da Berners-Lee (2001), il termine "indirizzo" può riferirsi a un indirizzo fisico, come quello di un'abitazione, oppure un indirizzo di posta elettronica. Gli esseri umani, grazie al contesto, riescono facilmente a capire a cosa ci si stia riferendo.

La soluzione proposta nel Web Semantico è l'uso degli URI (Uniform Resource Identifier), che consentono di attribuire un identificatore univoco a ogni risorsa. Tuttavia, l'URI da solo non è sufficiente: esso permette infatti di distinguere due risorse, ma non di comprenderne il significato. Perché un concetto sia realmente interpretabile occorre associare all'URI anche una descrizione formale delle sue proprietà e delle relazioni con altri concetti, utilizzando linguaggi ontologici come RDF/S e OWL. In questo modo, ad esempio, un "indirizzo fisico" potrà essere distinto da un "indirizzo di posta elettronica": entrambi avranno il proprio URI, ma soprattutto saranno caratterizzati da attributi e relazioni specifiche che li rendono chiaramente identificabili e correttamente interpretabili dalle macchine, evitando ambiguità.

In sintesi, mentre il linguaggio umano è flessibile e può interpretare lo stesso termine in contesti diversi, l'automazione richiede definizioni precise e univoche per evitare errori e ambiguità. Ad esempio, un archivio che conserva metadati relativi a vari computer storici, dove ogni macchina ha attributi come "modello", "anno di produzione", "produttore" e "software compatibile", senza un sistema di standardizzazione, questi metadati potrebbero essere archiviati in formati incompatibili tra diverse istituzioni, rendendo difficile fare ricerche e connessioni tra i dati.

In conclusione, RDF e le tecnologie connesse, come OWL (Web Ontology Language) per ontologie complesse e SKOS (Simple Knowledge Organization System) per thesauri e vocabolari controllati, permettono la creazione di un Web interoperabile in grado di "parlare" il linguaggio dei dati strutturati, dove ogni concetto ha un significato preciso e univoco.

Questo approccio diventa fondamentale nel contesto del patrimonio informatico, dove la quantità e la diversità delle risorse richiedono sistemi efficaci di gestione e condivisione che superino le barriere del formato, del linguaggio, della provenienza, ecc.

RDF si basa su un modello di tripla, formato da:

- Soggetto (una risorsa identificata da un URI o un nodo anonimo se non è necessario identificare univocamente la risorsa);
- Predicato (una proprietà che descrive una caratteristica o una relazione);
- Oggetto (il valore della proprietà, che può essere un'altra risorsa o un dato letterale) (Miarka, Zacek, 2011).

Utilizzando RDF, ogni risorsa (in questo caso, ad esempio, ogni computer storico) viene identificata da un URI unico e i suoi attributi (come "modello", "produttore", "software") possono essere descritti con tripli RDF, creando una rete di informazioni interoperabili. Di seguito un esempio:

- Soggetto: "Apple I"
- Predicato: "è stato prodotto da"
- Oggetto: "Apple Inc."

Queste triple collegano ogni computer con altri concetti rilevanti nel dominio dell'informatica patrimoniale, come i produttori o i software compatibili, e possono essere estesi per includere altre risorse, come manuali digitalizzati, schemi tecnici o software originali che funzionavano con quel particolare modello.

Come affermato da Tim Berners-Lee et al (2001), il Web Semantico promuove un Web universale in cui "ogni concetto" può essere collegato in modo preciso grazie all'uso degli URI, facilitando l'interoperabilità e la connessione dei dati in maniera fluida e automatica. Gli URI garantiscono che i concetti non siano solo parole in un documento, ma siano legati a una definizione unica che chiunque può trovare nel Web.

La rappresentazione in RDF prevede la creazione di un grafo diretto ed etichettato, in cui i nodi rappresentano risorse e gli archi le proprietà e le relazioni tra esse. Il modello è indipendente dalla sintassi specifica utilizzata per la sua serializzazione, che può essere realizzata in diversi formati. Originariamente, è stato definito il formato RDF/XML, basato su XML. Successivamente sono stati introdotti altri formati di serializzazione, tra cui Turtle e JSON-LD, oggi tra i più utilizzati per la loro maggiore leggibilità e semplicità d'uso.

RDF, quindi, non descrive la semantica ontologica, ma fornisce una base comune per poterla esprimere, fornisce, in altri termini, un modello astratto per la rappresentazione di dati, che può essere serializzato utilizzando formati diversi, tra cui RDF/XML, in cui le informazioni sono rappresentate mediante tag XML (W3C, 1999). RDF/XML, basandosi su XML, fa uso dei namespace XML per associare in modo univoco ciascuna proprietà al vocabolario o schema che la definisce (W3C, 1999).

1.6.4 OWL

Il linguaggio OWL (Web Ontology Language), definito nel 2004 dal W3C, è costruito sopra il modello RDF, utilizzando RDF e RDFS come base per fornire una semantica ontologica più espressiva e strumenti avanzati per la rappresentazione formale di ontologie. Esso si presenta in tre versioni principali:

1. OWL-Lite: semplice, adatta per gerarchie di classi e vincoli elementari.
2. OWL-DL: intermedio, bilancia espressività e computabilità.
3. OWL-Full: massima espressività, ma senza garanzie di decidibilità

Le ontologie rappresentano una componente chiave per il Web Semantico, in quanto rendono espliciti i rapporti tra le risorse, fornendo un grado di struttura che consente alle macchine di comprendere e utilizzare i dati (M. Guerrini, 2013); armonizzano i dati provenienti da fonti multiple, supportando la costruzione di Knowledge Graphs e sono alla base dell'iniziativa Linked Data, che pubblica e collega dati strutturati sul Web. Come abbiamo discusso nei paragrafi precedenti, le ontologie sono modelli formali di rappresentazione della conoscenza, espressi attraverso linguaggi specifici, e costituiscono strumenti fondamentali per l'ingegneria della conoscenza, utili a modellare domini specifici a fini di manipolazione informatica (Roche, 2005).

1.7 Integrazione tra ontologie e terminologie: il contributo dell'ontoterminologia e della termontografia

Un tentativo di combinare gli aspetti della terminologia e delle ontologie in un'unica struttura integrata viene fornito da un approccio teorizzato da Christophe Roche e dal gruppo Condillac, l'Ontoterminologia, ovvero una terminologia il cui sistema concettuale è un'ontologia formale (Roche, 2012).

Anche se l'analisi dettagliata dell'Ontoterminologia e delle sue applicazioni specifiche verrà trattata più approfonditamente nel quarto capitolo, è importante sottolineare sin d'ora il suo ruolo cruciale nel dimostrare la connessione tra ontologie e terminologie.

Roche distingue tra termini di uso e termini normati. I primi sono quelli adottati nel linguaggio quotidiano o nel discorso corrente, mentre i secondi sono strettamente legati a un sistema concettuale formale.

L'Ontoterminologia ha due funzioni principali: da un lato, si concentra sulla costruzione del sistema concettuale, utilizzando principi epistemologici per definire i concetti in base a proprietà essenziali; dall'altro, si occupa dell'operazionalizzazione della terminologia, impiegando rappresentazioni formali e computazionali per garantire la coerenza, la condivisione e la verificabilità dei concetti, con l'obiettivo di permettere applicazioni pratiche e scientifiche. Questo approccio integra concetti empirici e razionali, contribuendo alla creazione di un linguaggio condiviso e verificabile tra diverse comunità di esperti (Roche, 2007). L'Ontoterminologia mira a ristabilire il concetto e la sua denominazione al centro della terminologia, senza però trascurare l'uso sociolinguistico dei termini all'interno delle lingue di specialità. Questo approccio cerca di preservare un equilibrio tra la rappresentazione formale e la flessibilità nell'uso dei termini nelle diverse comunità linguistiche e professionali.

Inoltre, la terminologia include una serie di pratiche interconnesse, come l'uso della lingua di specialità, che è fondamentale per identificare i termini specifici di un dominio, ma che dipende da un contesto extralinguistico condiviso tra gli attori del discorso. Essa si basa anche sulla «lingua dell'intellezione» (*langue de l'intellection*, Roche, 2007), ossia il concetto che consente di classificare oggetti condividendo proprietà comuni, e necessita dell'impiego di linguaggi formali per rappresentare in modo coerente e applicabile il sistema concettuale.

Questo è un aspetto essenziale per evitare ambiguità e per l'applicazione pratica della terminologia (Roche, 2007).

La terminologia va oltre la linguistica, in quanto si occupa della comprensione e della rappresentazione degli oggetti del mondo attraverso un approccio scientifico che richiede un linguaggio formale per evitare ambiguità e per applicazioni pratiche, come nel caso dell'ingegneria

collaborativa. Qui, la soluzione non sta nella traduzione dei termini, ma nella creazione di un linguaggio concettuale condiviso.

Anche Temmerman compie un passo significativo nell'integrazione delle ontologie nel campo della terminologia, introducendo la Termontologia e la Termontografia. In questo ambito, combina la teoria sociocognitiva con l'ingegneria ontologica (Faber, L'Homme, 2022). La termontografia ha superato le sue origini sociocognitive, incorporando tecniche ingegneristiche più avanzate (Faber, 2012). Essa rappresenta un tentativo sistematico di mappare e standardizzare le relazioni terminologiche in diversi ambiti disciplinari, creando collegamenti tra informazioni terminologiche multilingue e risorse concettuali (Faber, 2012).

In questo contesto, la termontografia svolge un ruolo fondamentale nel collegare le conoscenze linguistiche alla loro rappresentazione ontologica. Attraverso un'analisi approfondita delle strutture terminologiche e delle interrelazioni tra i concetti, la termontografia offre strumenti preziosi per comprendere come le terminologie possano essere tradotte in ontologie formali, facilitando così una migliore interazione tra linguistica e rappresentazione della conoscenza.

Un esempio significativo della rilevanza della termontografia nell'ambito del multilinguismo e delle conoscenze specifiche per cultura è fornito da Temmerman, insieme a Kerremans e Tummers, nel progetto *Representing Multilingual and Culture-Specific Knowledge in a VAT Regulatory Ontology: Support from the Termontography Method* (2023). In questo studio, gli autori affrontano le sfide e le metodologie necessarie per strutturare un database terminologico dedicato all'imposta sul valore aggiunto (IVA) europeo.

Il loro approccio si basa su un framework di unità di comprensione (UoU), precedentemente discusso (*vedi pag.*22), sviluppato in collaborazione con esperti di settore per migliorare la rappresentazione delle informazioni legislative in diverse lingue. La termontografia affronta le difficoltà legate alle differenze linguistiche e culturali che complicano l'allineamento delle terminologie IVA, articolando un workflow che comprende fasi di analisi, raccolta di informazioni, ricerca, raffinamento, verifica e validazione. Attraverso questa metodologia, i ricercatori sottolineano l'importanza di un framework di categorizzazione per gestire la diversità culturale e linguistica, rendendo il database una risorsa preziosa per l'ingegneria ontologica.

Questa diversità e complessità terminologiche non si limitano al contesto fiscale, ma si estendono anche al diritto processuale penale. Un esempio pertinente è fornito dallo studio di Gianluca Pontrandolfo, *La fase preliminare del processo penale. Il contributo dell'approccio sociocognitivo a un'indagine terminografica spagnolo-italiano* (2010), che applica l'approccio sociocognitivo all'analisi della terminologia giuridica spagnola e italiana nella fase preliminare del processo penale. Questo studio mette in luce le differenze concettuali tra i due sistemi giuridici, evidenziando

l'importanza della categorizzazione delle unità di significato per garantire una traduzione efficace e ridurre le asimmetrie terminologiche. La comprensione delle specificità terminologiche e delle loro interrelazioni è essenziale per facilitare il dialogo giuridico tra diverse culture legali e migliorando l’interoperabilità nei sistemi del diritto internazionale.

Capitolo 2. Stato dell’esistente

Il presente capitolo intende analizzare lo stato attuale delle risorse e degli strumenti impiegati per la rappresentazione, la gestione e la fruizione della conoscenza nel dominio dell'informatica attraverso il settore museale, analizzando come i musei valorizzino tale conoscenza. In questo contesto, l'impiego delle terminologie e delle ontologie si configura come il collegamento concettuale con le tematiche affrontate nel primo capitolo, fornendo gli strumenti indispensabili per l'organizzazione e l'interpretazione delle informazioni.

Dopo aver esaminato, nel primo capitolo, l'evoluzione teorica e metodologica dei concetti nella terminologia e nelle ontologie, il secondo capitolo si sposta sul panorama pratico, mediante un'analisi critica delle tecnologie, delle metodologie e delle applicazioni attualmente in uso. L'obiettivo è evidenziare i principali limiti, le potenzialità e le sfide che caratterizzano il settore, con particolare attenzione all’interoperabilità e alla gestione dei dati eterogenei.

Notevole rilevanza assume l'analisi di come le istituzioni culturali, in particolare i musei, organizzino e veicolino la conoscenza, sia attraverso modelli teorici sia mediante strumenti operativi. I musei non rappresentano unicamente depositari di oggetti e artefatti, ma assumono un ruolo etico fondamentale nella trasmissione del sapere, rispondendo alla necessità di educare e coinvolgere il pubblico. Tale funzione educativa e culturale richiede l'adozione di strumenti e strategie capaci di rendere le informazioni accessibili, nel rispetto dei valori scientifici e culturali che esse ricoprono.

In particolare, si analizzerà come gli standard e i modelli semantici possano supportare le istituzioni culturali nel migliorare l’accessibilità e la condivisione dei dati, affrontando le sfide legate all’interoperabilità e alla frammentazione delle risorse digitali. Le risorse, infatti, sono spesso rappresentate attraverso formati, strutture e metadati non standardizzati, che comportano difficoltà nell'accesso, nella gestione e nell'integrazione delle informazioni, ostacolando il processo di condivisione e collaborazione tra le istituzioni e limitando l'efficacia degli strumenti tecnologici.

La gestione dei dati museali si configura come un campo complesso e di notevole importanza per il patrimonio culturale, poiché include, dunque, non solo artefatti fisici, ma anche risorse digitali. Tale complessità è determinata da altri molteplici fattori, quali la varietà tipologica degli oggetti, l'adozione di sistemi di catalogazione che utilizzano standard differenti, nonché le difficoltà legate all'integrazione e alla gestione di dati e metadati.

Il dominio informatico nei musei, in particolare, si configura come estremamente articolato, richiedendo l'integrazione di sistemi eterogenei, la gestione di volumi crescenti di dati e la garanzia dell'interoperabilità tra piattaforme digitali e archivi tradizionali.

L'analisi delle soluzioni attuali, comprendente sia quelle consolidate sia quelle emergenti, mira a fornire un quadro chiaro delle opportunità e delle lacune presenti nel settore.

2.1 Strumenti per l'organizzazione della conoscenza: standard e modelli semantici del patrimonio culturale

In un contesto di gestione e rappresentazione della conoscenza, soprattutto quando si tratta di dati del patrimonio culturale, che sono spesso eterogenei e interconnessi, è fondamentale l'uso di standard e modelli di riferimento. Gli standard sono norme che regolano i processi, con l'obiettivo di renderli omogenei e coerenti. In ambito terminologico, gli standard svolgono un ruolo fondamentale nel rappresentare concetti in modo univoco, definendo protocolli e sistemi di organizzazione e rappresentazione della conoscenza che permettono di rappresentare, organizzare e condividere informazioni, superando le differenze tra i vari sistemi informatici e consentendo a diversi enti e sistemi di interagire e di scambiare dati semanticamente, garantendo l'integrazione di risorse provenienti da contesti eterogenei.

L'International Organization for Standardization (ISO)[8], fondata nel 1947 e con sede a Ginevra, Svizzera, è l'organismo internazionale che definisce e promuove gli standard a livello globale. Composta da esperti e rappresentanti degli organismi nazionali di standardizzazione provenienti da 164 paesi, l'ISO sviluppa e propone soluzioni basate sulle necessità emergenti in vari settori. Secondo l'ISO e l'International Electrotechnical Commission (IEC), uno standard è «un documento, stabilito tramite consenso e approvato da un organismo riconosciuto, che fornisce, per un uso comune e ripetuto, regole, linee guida o caratteristiche per attività o per i loro risultati, finalizzate al raggiungimento del grado ottimale di ordine in un dato contesto»[9].Tale definizione evidenzia l'importanza degli standard nel garantire uniformità e coerenza in ambiti diversi. Inoltre, l'ISO specifica che "gli standard sono la saggezza distillata di persone con esperienza nella loro materia e che conoscono le esigenze delle organizzazioni che rappresentano"[10],suggerendo come questi strumenti possano tradurre conoscenze specialistiche in soluzioni condivise, mirate al miglioramento delle pratiche e al facilitamento dell'interazione tra i diversi contesti.

Gli standard internazionali, come quelli definiti dall'ISO, forniscono principi generali per la rappresentazione delle informazioni. Tuttavia, nel contesto attuale, dove la digitalizzazione e la condivisione delle risorse culturali sono sempre più determinanti per la valorizzazione del patrimonio, è fondamentale considerare anche l'importanza degli standard specifici per le risorse sul Web, promossi dal W3C (World Wide Web Consortium), per migliorare la condivisione e l'integrazione dei dati su larga scala. Tra questi, il Resource Description Framework (RDF), descritto nel precedente capitolo.

[8] Il nome ISO deriva dal greco ISOS, che significa "uguale".
[9] Traduzione nostra (ISO/IEC Guide 2:2004).
[10] Traduzione nostra (sito ufficiale ISO).

Uno degli aspetti chiave di RDF è la gestione rappresentazione dei metadati, che sono dati strutturati che descrivono altre risorse. I metadati sono essenziali per migliorare il recupero e l'organizzazione delle informazioni sul Web, rendendole interpretabili dalle macchine. RDF introduce metadati strutturati che descrivono consente di descrivere in modo strutturato il contenuto delle risorse e le loro relazioni, supportando una gestione automatizzata più avanzata. Ad esempio, RDF consente di associare ai documenti web proprietà come autore, data di pubblicazione, lingua, licenza d'uso e molto altro.

Questo è particolarmente importante nei contesti in cui i dati provengono da fonti eterogenee, come nei musei e negli archivi digitali, dove l'integrazione tra dati di diversa origine può risultare complessa.

Inoltre, RDF è utilizzato in numerosi contesti, tra cui appunto la gestione dei dati museali e culturali. Due esempi rilevanti sono CIDOC-CRM ed EDM (Europeana Data Model), che impiegano RDF come modello di rappresentazione dei dati per supportare l'interoperabilità semantica e una descrizione formale e condivisa delle risorse culturali per garantire interoperabilità e rappresentazione semantica.

L'adozione di questi standard rappresenta la condizione necessaria per l'interoperabilità, ma la loro implementazione pratica richiede strumenti capaci di mediare tra la formalità tecnica e le necessità descrittive degli studiosi. In questo scenario si inserisce OntoMe (Ontology Management Environment), una piattaforma web-based progettata per la gestione collaborativa di ontologie e vocabolari. Sviluppata dal laboratorio LARHRA, OntoMe funge da ecosistema aperto dove i ricercatori possono modellare, allineare e condividere classi e proprietà specifiche per diversi domini di studio. Sebbene utilizzi spesso il CIDOC-CRM, uno dei principali modelli di riferimento nel dominio museale, come nucleo semantico per garantire la coerenza dei dati, la piattaforma non è limitata a un unico modello; essa facilita l'integrazione semantica tra dataset eterogenei e standard differenti (come Dublin Core o SKOS), permettendo di creare "profili applicativi" su misura per ogni progetto di ricerca.

La sezione che segue esplora i modelli di riferimento più utilizzati nel contesto delle ontologie e della rappresentazione della conoscenza, concentrandosi su quelli rilevanti per la gestione dei dati museali e dell'informatica.

2.1.1 CIDOC-CRM: Modello concettuale per la documentazione del patrimonio culturale

Un modello di riferimento è una rappresentazione astratta e standardizzata di un dominio di conoscenza, che fornisce una base concettuale per la definizione di ontologie specifiche e per l'interoperabilità tra sistemi. Fornisce, dunque, una struttura concettuale che descrive i principali elementi di un dominio e le relazioni tra di essi, utilizzando un linguaggio formale e un insieme di regole per interpretare e scambiare i dati. Tali modelli sono essenziali per risolvere il problema dell'interoperabilità semantica, che si verifica quando diverse istituzioni o sistemi utilizzano formati e terminologie differenti per rappresentare dati simili.

In questo contesto, diversi modelli ontologici sono stati sviluppati per affrontare specifiche esigenze del settore culturale, come la gestione dei dati museali, archivistici e bibliografici. Tali modelli non solo facilitano l'integrazione e l'interoperabilità, ma promuovono anche la creazione di sistemi di gestione più efficienti, capaci di adattarsi a nuove esigenze e tecnologie emergenti.

Il CIDOC Conceptual Reference Model (CIDOC-CRM) è uno dei modelli ontologici più riconosciuti a livello internazionale per la modellazione e l'integrazione dei dati culturali. Progettato inizialmente nel 1996 sotto l'egida dell'ICOM-CIDOC Documentation Standards Group, è ora riconosciuto come standard internazionale ISO 21127, con una revisione aggiornata al 2023[11]. È il risultato di oltre due decenni di sviluppo collaborativo da parte del CIDOC CRM Special Interest Group (SIG), un organismo tecnico operante sotto l'egida dell'International Council of Museums (ICOM), ed è stato progressivamente adottato da un'ampia gamma di istituzioni, incluse quelle museali, accademiche e archivistiche, in tutto il mondo.

Il suo obiettivo primario è fornire un'infrastruttura semantica per favorire lo scambio e l'integrazione di informazioni provenienti da fonti e sistemi eterogenei.

«Il ruolo principale del CRM è quello di consentire lo scambio di informazioni e l'integrazione tra fonti eterogenee di informazioni sul patrimonio culturale»[12] (Crofts *et al.*, 2009).

Il CIDOC-CRM definisce un'ontologia formale che fornisce una struttura concettuale modulare, estensibile e condivisa per descrivere relazioni complesse tra oggetti, eventi, persone e luoghi. Questo approccio consente di integrare ed esplorare i dati relativi al patrimonio culturale in modo indipendente da specifici software, applicazioni o schemi di archiviazione (Sanfilippo, Markhoff, Pittet, 2020). Fungendo come una sorta di "collante semantico" (*semantic glue*[13]), garantisce così una struttura semantica condivisa e standardizzata, trasformando informazioni locali in risorse globali

[11] ISO 21127:2023 Information and documentation — A reference ontology for the interchange of cultural heritage information

[12] La frase citata è stata tradotta dall'autore di questa tesi.

[13] https://cidoc-crm.org/

integrate (Crofts *et al.*, 2009). La natura dello standard CIDOC-CRM lo rende particolarmente adatto per descrivere e collegare conoscenze culturali complesse grazie ad un *CRMbase*, il nucleo dello standard, che definisce classi e relazioni fondamentali per la documentazione del patrimonio culturale, e una serie di estensioni modulari progettate per rispondere a specifiche esigenze di ricerca. La versione 6.2.1 del CIDOC-CRM comprende 94 classi tassonomiche e 168 relazioni organizzate gerarchicamente (genitore e figlio), per descrivere entità concettuali e fisiche.

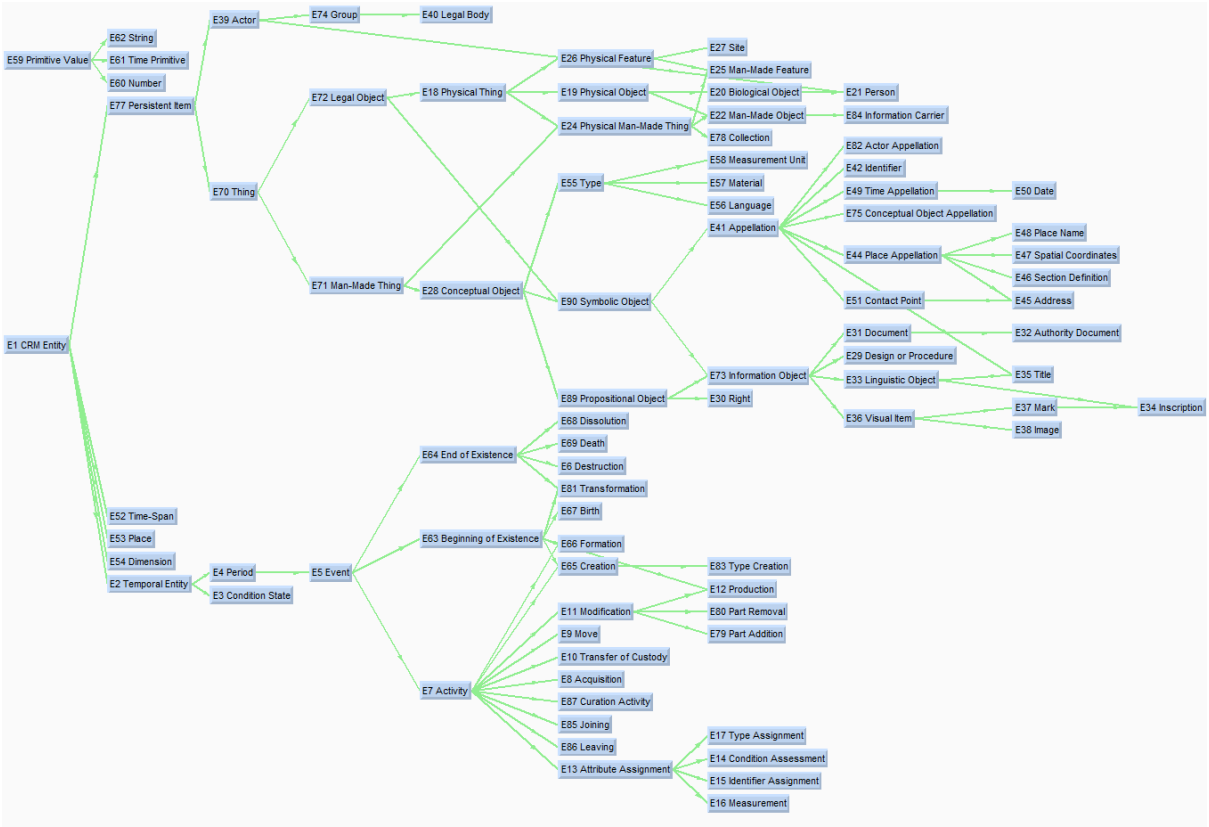


Figure 6 Gerarchia delle classi del CIDOC CRM, che illustra le principali classi e le loro relazioni gerarchiche all'interno del modello concettuale (Fonte: Sito ufficiale del CIDOC CRM, consultato il 02/12/2024).

La struttura modulare consente l'integrazione con estensioni specializzate per domini specifici, come la bibliografia o la geoinformatica, incoraggiando gli utenti a creare proprie estensioni per soddisfare esigenze particolari.

«Per la sua stessa struttura e formalismo, il CRM è estensibile e gli utenti sono incoraggiati a creare estensioni per soddisfare le esigenze di comunità e applicazioni più specializzate» (Crofts et al, 2009). Di seguito i modelli che si integrano con CIDOC-CRM:

Figure 7 Modelli Compatibili e Collaborazioni del CIDOC CRM (Fonte: Sito ufficiale del CIDOC CRM, consultato il 02/12/2024).

Questa figura mostra i modelli che si integrano con CIDOC-CRM, illustrando come le estensioni e le collaborazioni con altre iniziative possano ampliarne l'applicabilità e l'efficacia.

Sebbene il CIDOC-CRM sia uno strumento potente per garantire l'interoperabilità semantica, alcuni studiosi hanno individuato alcune criticità, come osservato da Sanfilippo *et al.* (2020), quali:

- Semi-formalità e ambiguità: alcuni concetti, come la classe *E5 Event*, e relazioni, come *P53 has former or current location*, possono essere interpretati in modi alternativi, compromettendo l'interoperabilità tra applicazioni. La distinzione tra *E39 Actor* ed *E70 Thing*, ad esempio, non è netta. *E21 Person* è sia un attore che un oggetto biologico, confondendo i limiti tra le categorie; tuttavia, non si tratta di un limite, bensì di una scelta specifica di progettazione, compatibile con la modellazione semantica in cui una classe può avere più super-classi (ereditarietà multipla) per riflettere la realtà in modo più preciso (nel caso specifico, si vuole poter rappresentare sia il corpo umano come oggetto biologico (es. resti archeologici, scheletri, mummie) sia come agente sociale (es. autore di un'opera, committente, testimone).
- Assenza di una definizione chiara delle cardinalità, che può portare a modelli non interoperabili, ad esempio se un evento può non avere partecipanti o può averne molti;
- Limitazioni nella gestione di dimensioni e valori: CIDOC non rappresenta in modo diretto valori qualitativi come colori o pesi in modo flessibile;

- Rappresentazione ridondante delle entità temporali: la necessità di modellare entità temporali anche per fenomeni semplici, come la data di creazione di un oggetto, aumenta la complessità senza benefici significativi, in contesti in cui magari non è necessario specificarla.

Per rispondere a queste sfide, sono state proposte diverse soluzioni:

1. Modularizzazione: suddivisione dell'ontologia in 18 moduli specifici, come quelli per oggetti fisici e attori, facilitando il riutilizzo selettivo;
2. Formalizzazione: revisione delle relazioni disgiuntive e definizione chiara dei valori qualitativi tramite proprietà OWL;
3. Introduzione di scorciatoie semantiche: ad esempio, proprietà come *createdAt* semplificano la rappresentazione delle date di produzione senza necessità di modellare entità temporali (Sanfilippo *et al.*, 2020).

In conclusione, sebbene il CIDOC-CRM non affronti direttamente le terminologie tecniche dell'informatica applicata ai beni culturali, né si concentri su domini specialistici come i beni digitali e la loro trasmissione linguistica, grazie alle sue estensioni e alla modularizzazione, esso consente un riutilizzo selettivo, una manutenzione semplificata e l'espansione dell'ontologia con moduli specifici a seconda degli ambiti. Questi miglioramenti rendono il CIDOC-CRM uno strumento potente e flessibile per la gestione dei dati museali e per l'integrazione semantica nel contesto culturale, rendendolo sempre più adattabile alle necessità di progetti complessi, come quelli legati ai dati archeologici (Sanfilippo *et al.*, 2020). Ad esempio, OpenArchaeo[14] utilizza moduli specifici per connettere dataset archeologici, semplificando lo sviluppo e favorendo una migliore comprensione dell'ontologia da parte degli utenti finali.

Tuttavia, nonostante i vantaggi offerti dal CIDOC-CRM, il suo utilizzo come base per l'integrazione con terminologie informatiche e museali in reti semantiche più ampie o specifiche può presentare delle sfide: il modello rimane, infatti, una risorsa generale e astratta, che può non soddisfare pienamente le necessità di tutti i contesti applicativi.

La crescente esigenza di un modello più flessibile per l'interoperabilità ha portato allo sviluppo dell'Europeana Data Model (EDM), che è stato adottato da Europeana come standard per la gestione dei dati culturali digitali. EDM si propone come un'alternativa più orientata alla specificità dei dati digitali e alle esigenze di interoperabilità tra i vari sistemi.

[14] https://open-archaeo.info/

2.1.2 EDM: Europeana Data Model per l'accesso ai dati culturali: Il caso del progetto *Amsterdam Museum Linked Open Data*

Europeana è una piattaforma digitale europea che raccoglie e mette a disposizione un vasto patrimonio culturale europeo proveniente da musei, biblioteche e archivi. La sua missione è quella di promuovere l'accesso a questi dati, rendendoli disponibili a scopi educativi, creativi e di ricerca. La piattaforma consente l'accesso ai metadati digitalizzati, che sono facilmente consultabili attraverso un'interfaccia user-friendly.

Un elemento chiave nella sua struttura è il Europeana Data Model (EDM), che serve a standardizzare e connettere i metadati provenienti da una molteplicità di fonti, garantendo un'elevata interoperabilità tra i sistemi. Grazie all'adozione del modello EDM, Europeana raccoglie e condivide metadati che provengono da numerosi ambiti culturali, creando una rete che consente di accedere e consultare contenuti provenienti da diverse istituzioni. Ciò facilita la costruzione di una risorsa centralizzata che, pur mantenendo l'autonomia delle singole entità, contribuisce a una visione globale del patrimonio culturale europeo.

EDM è stato progettato per essere più espressivo e flessibile rispetto al precedente modello ESE (Europeana Semantic Elements) e per essere un ponte tra i diversi standard. Si inserisce, dunque, nell'ambito più ampio del Web Semantico, facilitando la creazione di collegamenti tra dati e risorse esterne tramite l'uso di Linked Open Data (LOD). Il modello si basa su standard consolidati come RDF(S), OAI-ORE, SKOS, e Dublin Core, che ne garantiscono la compatibilità con le pratiche di catalogazione e la capacità di relazionarsi con altri sistemi di dati.

Nel 2021, Europeana ha avviato una strategia a medio termine per migliorare la disponibilità del multilinguismo sulla sua piattaforma, sfruttando i progressi tecnologici e l'uso di vocabolari multilingue per i metadati esistenti. L'obiettivo di questa strategia è rendere la navigazione, la ricerca e la lettura dei contenuti disponibili in più lingue, migliorando l'accessibilità e l'uso delle risorse culturali da parte di un pubblico globale. L'approccio adottato combina vocabolari affidabili con copertura multilingue e la traduzione in inglese come lingua ponte per le aree non coperte dai vocabolari (Silva, Terra, 2024).

I principali aspetti strutturali di EDM secondo Doerr *et al.* (2010) sono:

1. Compatibilità con il Web Semantico: il modello EDM adotta il framework RDF per rappresentare i dati come triple semantiche, creando una struttura che permette di associare metadati e contenuti a risorse esterne. Ciò facilita l'integrazione e la ricerca di dati in contesti ampi, come il Web Semantico e i Linked Data.
2. Integrazione di standard consolidati: EDM utilizza e integra ontologie e standard consolidati, come SKOS per la gestione della conoscenza, Dublin Core per le descrizioni basilari dei

metadati, e FOAF per la rappresentazione delle persone e delle loro connessioni nei dati digitali. Questa integrazione consente di mantenere una certa coerenza semantica e di garantire l'interoperabilità tra sistemi diversi.

3. Framework OAI-ORE: un altro aspetto fondamentale è l'integrazione con il framework OAI-ORE (Open Archives Initiative Object Reuse and Exchange), che permette la descrizione e la gestione di risorse composte. Questo è particolarmente utile per la gestione di dati complessi che comprendono più entità o che necessitano di essere rappresentati come risorse collegate, come avviene frequentemente nei contesti archivistici e museali.
4. Flessibilità ontologica: l'EDM funge da ontologia di alto livello, compatibile con altri modelli esistenti come LIDO[15], CIDOC-CRM ed EAD. La sua flessibilità consente di preservare la ricchezza dei modelli originali, adattandosi alle specifiche esigenze di diverse istituzioni e settori.
5. Focus sugli eventi e le relazioni storiche: un'altra caratteristica distintiva dell'EDM è il suo approccio *event-centric*, che mette l'accento sulle relazioni storiche e contestuali tra gli oggetti e gli eventi culturali. Questo approccio consente di valorizzare il contesto storico degli oggetti, migliorando la comprensione delle connessioni temporali e spaziali che li legano.

Un esempio pratico di applicazione dell'EDM è rappresentato dal progetto *Amsterdam Museum Linked Open Data* (De Boer *et al.*, 2013). Il Museo di Amsterdam ha reso disponibile la propria collezione di oltre 70.000 oggetti, comprendente dipinti, oggetti in vetro, argento, libri e costumi, è gestita digitalmente con il software *Adlib* e resa disponibile online dal 2010. La collezione è in formato Linked Data a cinque stelle (*5-star*), un'iniziativa che ha permesso di mappare i metadati museali in modo interoperabile.

Il termine *Linked Data a cinque stelle* fa riferimento a un sistema di valutazione proposto da Tim Berners-Lee per classificare la qualità dei dati pubblicati online come Linked Open Data:

[15] Nel contesto museale, LIDO ha assunto un ruolo centrale come standard di riferimento per il *harvesting* (raccolta) dei dati. Più che una sostituzione, si è consolidata una sinergia operativa: i musei utilizzano spesso LIDO per la descrizione granulare degli oggetti, che viene poi mappata e "tradotta" nel modello EDM per consentire l'integrazione e la pubblicazione su Europeana. Per ulteriori informazioni: https://icom-documentation.mini.icom.museum/working-groups/lido/lido-overview/

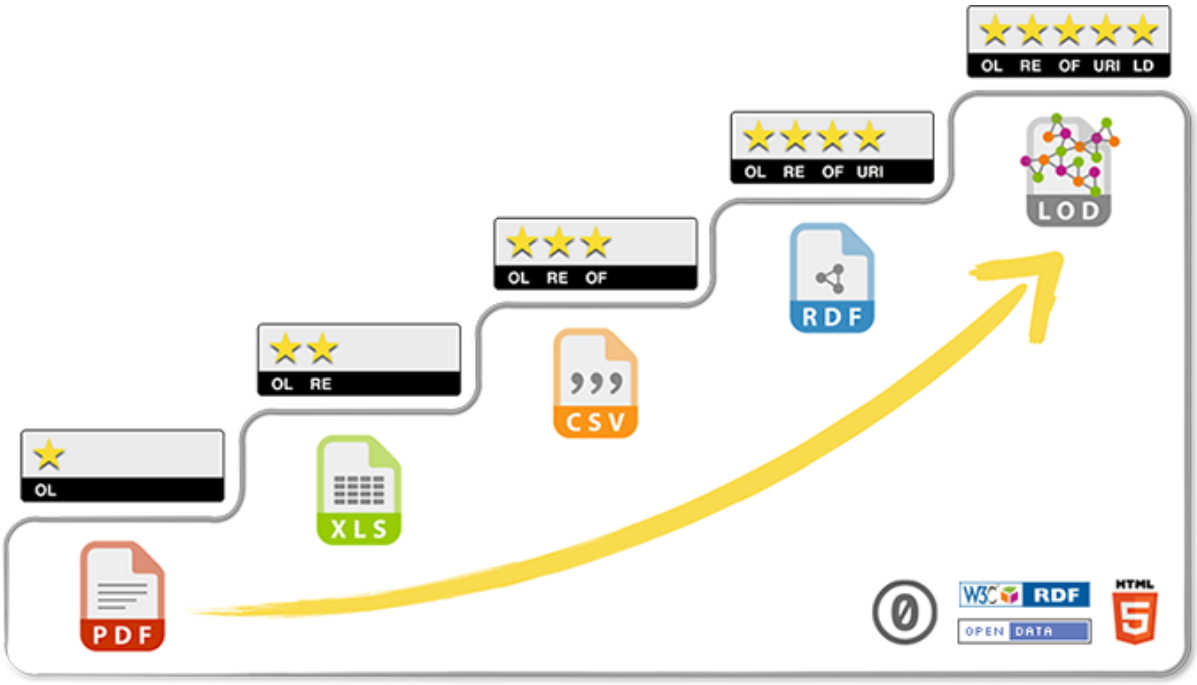


Figure 8 Diagramma che illustra i cinque livelli del piano per i dati aperti a 5 stelle di Tim Berners-Lee"The Five-Star Open Data." 5stardata.info, n.d., https://5stardata.info/en/.

La classificazione a 5 stelle per i Linked Open Data valuta quanto i dati siano strutturati, accessibili interoperabili (Ontotext, accesso dicembre 2024):

1 stella: Dati disponibili online in qualsiasi formato, con licenza aperta, ma non strutturati o facilmente elaborabili.

2 stelle: Dati strutturati in un formato proprietario (es. Excel), che possono essere elaborati solo con software specifici.

3 stelle: Dati strutturati in un formato aperto (es. CSV), facilmente elaborabili senza necessità di software proprietari.

4 stelle: Dati strutturati con URI e in un formato aperto, accessibili online tramite URL e utilizzabili in applicazioni.

5 stelle: Dati aperti e interconnessi con altri dataset tramite collegamenti, come i Linked Open Data, che consentono agli utenti di scoprire informazioni interconnesse e di esplorare nuove relazioni tra i dati.

Il dataset dell'Amsterdam Museum include informazioni dettagliate sugli oggetti, un thesaurus con 28.000 concetti e un file autoritativo con 66.966 persone legate agli oggetti. La mappatura al modello EDM ha consentito al museo di integrare i propri dati con risorse esterne, come *GeoNames*, *DBpedia* e il *Getty Union List of Artist Names* (ULAN), attraverso l'uso di URI e collegamenti Linked Open Data.

L'adozione di EDM ha permesso al Museo di Amsterdam di preservare la complessità dei dati originali, garantendo allo stesso tempo l'interoperabilità con altre risorse culturali a livello europeo, in linea con le pratiche promosse da Europeana.
Per rendere i dati del Museo di Amsterdam compatibili con l'EDM sono stati fatti due passaggi principali:

1. Coppie proxy-aggregazione: Ogni oggetto del museo è rappresentato da due risorse:
   - Proxy: contiene i dettagli dell'oggetto (ad esempio, chi lo ha creato, le sue dimensioni, ecc.).
   - Aggregazione: contiene informazioni sulla provenienza dell'oggetto (come il fornitore dei dati o i diritti d'uso) e sulla rappresentazione digitale (come le immagini dell'oggetto).
2. Mappatura delle proprietà: Le informazioni specifiche del Museo di Amsterdam sono collegate a quelle del modello EDM usando un sistema di "sottoproprietà" e "classi" che garantisce che i dati del museo possano essere facilmente utilizzati insieme ad altri dati compatibili con l'EDM.

Il dataset Linked Open Data del Museo di Amsterdam è diviso in tre parti:

- Metadati degli oggetti: 73.447 oggetti con informazioni dettagliate.
- Thesaurus: Un elenco di concetti utilizzati nei metadati.
- Lista di autorità delle persone: Elenco di 66.966 persone collegate agli oggetti del museo.

I metadati degli oggetti sono composti da circa 5,7 milioni di triplette RDF. Inoltre, 90 delle proprietà del museo sono collegate a quelle di Dublin Core, un altro standard di metadati, e 7 proprietà sono mappate direttamente all'EDM, rendendo i dati interoperabili e facilmente utilizzabili con altri dataset (De Boer *et al.*, 2013).

> «Europeana dà potere al settore del patrimonio culturale nella sua trasformazione digitale. Sviluppiamo competenze, strumenti e politiche per abbracciare il cambiamento digitale e incoraggiare partenariati che promuovono l'innovazione. Rendiamo più facile per le persone utilizzare il patrimonio culturale per l'istruzione, la ricerca, la creazione e la ricreazione. Il nostro lavoro contribuisce a una società aperta, informata e creativa» (Europeana Foundation, 2008)[16].

Questo esempio dimostra come EDM possa supportare la creazione di un'infrastruttura di dati condivisi tra istituzioni culturali, aumentando l'accessibilità e la fruibilità delle risorse. Il progetto, infatti, si inserisce in un contesto globale di scambio di dati culturali, valorizzando il patrimonio in

[16] Le frasi citate sono tradotte dall'autore di questa tesi.

modo efficiente e standardizzato, come parte della strategia di Europeana per la condivisione del patrimonio culturale digitale.

Il Web Semantico e il LOD offrono opportunità per la gestione della conoscenza, strutturando i dati in formati interoperabili leggibili sia da persone che da macchine e collegando informazioni provenienti da fonti diverse, come evidenziato da Bizer *et al.* (2009). Standard come RDF e OWL sono fondamentali per costruire reti di dati interconnesse. In questo contesto, il LOD consente di collegare informazioni provenienti da fonti diverse, favorendo l'accesso aperto e l'interoperabilità. Progetti come Europeana e il Getty Vocabulary Program mostrano il potenziale del LOD per il patrimonio culturale, consentendo l'accesso a milioni di risorse digitalizzate a livello globale.

Come CIDOC-CRM, anche Europeana presenta delle limitazioni quando si tratta di coprire aspetti tecnici più specifici, come quelli legati all'informatica applicata ai beni culturali. Sebbene l'EDM fornisca un importante strumento di collegamento e condivisione dei dati a livello europeo, non risponde completamente alle necessità di una terminologia tecnica, che è fondamentale per l'integrazione dei dati relativi all'informatica e alle scienze computazionali. Pur essendo un modello altamente efficace per l'aggregazione dei dati, l'EDM presenta una visione generalista, che non affronta in modo diretto o approfondito domini specialistici come l'informatica applicata ai beni culturali o la terminologia tecnica. Questo ne limita l'applicabilità in contesti che richiedono una precisione maggiore riguardo a specifici settori. Ciò rappresenta una sfida aperta che potrebbe beneficiare di soluzioni più specializzate, come quelle proposte da ontologie del settore informatico, per affrontare l'interoperabilità tra le diverse dimensioni della conoscenza.

Se modelli come il CIDOC-CRM e l'EDM sono fondamentali per l'organizzazione e l'interoperabilità dei dati museali, riflettendo un approccio pensato principalmente per il patrimonio materiale e documentale, esistono altri modelli che, pur essendo utilizzati in ambito informatico, si basano su rappresentazioni della conoscenza più formali e gerarchiche. Esempi di tali modelli includono il Computing Classification System (CCS) e la Computer Science Ontology (CSO), che offrono strutture concettuali orientate a classificare e organizzare il dominio dell'informatica.

2.2 Sistemi di classificazione e ontologie settoriali per la rappresentazione della conoscenza per il dominio informatico

Le risorse terminologiche esistenti, come glossari e thesauri, rappresentano strumenti essenziali per l'organizzazione e la classificazione della conoscenza.

Nel contesto delle discipline scientifiche, alcune aree, come la biologia e la fisica, sono descritte da tassonomie mature e su larga scala, come *MeSH* (Medical Subject Headings)[17] e *PhySH* (Physics Subject Headings)[18], che permettono di organizzare e classificare la ricerca in modo molto dettagliato. Questi sistemi consentono una comprensione profonda delle connessioni tra i vari concetti e la loro evoluzione nel tempo, facilitando così la navigazione e la scoperta di ricerche correlate.

Anche il contesto informatico offre una gamma di strumenti per la gestione e valorizzazione dei dati museali, tuttavia le tassonomie attualmente disponibili sono più generiche e tendono a evolversi in modo più lento (Salatino et al, 2020).

Molte soluzioni esistenti si concentrano sulla gestione di collezioni digitali e sull'interoperabilità semantica, senza approfondire gli aspetti legati al lessico tecnico o alla trasmissione linguistica della conoscenza scientifica.

Questo è dovuto alla natura dinamica e in rapida evoluzione della disciplina, che spesso rende difficile mantenere aggiornate e granulari le tassonomie. Le ontologie, rispetto alle tassonomie statiche, offrono un approccio più flessibile e consentono di modellare la conoscenza in modo più dinamico, integrando relazioni semantiche complesse (ad esempio proprietà e vincoli logici) che permettono una rappresentazione più precisa e adattabile dell'evoluzione dei concetti informatici.

Accanto a queste risorse, che mostrano i limiti delle tassonomie informatiche e il potenziale delle ontologie per la rappresentazione della conoscenza, è utile considerare anche strumenti settoriali già consolidati in altri ambiti disciplinari. Questa sezione prende in esame modelli di classificazione e rappresentazione della conoscenza che, pur nascendo in contesti diversi, presentano elementi utili anche per il dominio informatico. Tra questi rientra l'Art & Architecture Thesaurus (AAT), sviluppato nel contesto patrimoniale per l'organizzazione e la catalogazione dei dati artistici e architettonici e significativo come esempio di thesaurus specialistico capace di supportare la gestione semantica e multilingue delle collezioni. Accanto a esso, si trovano sistemi più specifici per l'informatica, come il Computing Classification System (CCS) e la Computer Science Ontology (CSO).

[17] https://meshb.nlm.nih.gov/
[18] https://physh.org/

2.2.1 Art & Architecture Thesaurus (AAT): Thesaurus per la catalogazione artistica e architettonica

Tra le innovazioni più significative implementate da Europeana vi è l'integrazione con vocabolari come l'Art & Architecture Thesaurus (AAT) del Getty Research Institute. Europeana ha collaborato con i propri partner per aggiornare i dataset, sostituendo le vecchie etichette AAT (semplici stringhe di testo) con gli URI AAT. In questo modo, è stato possibile accedere automaticamente ai dati semantici e multilingue associati ai concetti del vocabolario Getty, tramite i LOD. Questo processo ha permesso di collegare direttamente i metadati a risorse esterne, migliorando la qualità semantica dei dati e consentendo l'accesso a informazioni aggiuntive derivanti dal vocabolario Getty (sito ufficiale Europeana, aggiornato al 2023).

L'AAT è un vocabolario che contiene circa 44.000 concetti e 131.000 termini generici per la classificazione e l'organizzazione terminologica in ambito artistico e culturale. Fa parte del *Getty Vocabulary Program* (GVP), un insieme di vocabolari strutturati che include anche il *Getty Thesaurus of Geographic Names* (TGN), *l'Union List of Artist Names* (ULAN), il *Cultural Objects Name Authority* (CONA) e il *Getty Iconography Authority* (IA). Questi strumenti sono progettati per migliorare l'accesso alle informazioni relative all'arte, all'architettura e al patrimonio culturale materiale.

Gli obiettivi principali del vocabolario AAT, come descritto sul suo sito ufficiale, sono molteplici: in primo luogo, promuove una catalogazione coerente, utilizzando una terminologia uniforme e offrendo diverse opzioni per esprimere lo stesso concetto. L'AAT, oltre a favorire il collegamento dei dati, contribuisce anche alla riconciliazione delle informazioni provenienti da fonti diverse. Un altro aspetto importante riguarda il recupero delle informazioni, che è facilitato dall'uso di sinonimi, descrizioni generali o specifiche, e reti semantiche, rendendo più facile l'accesso a dati rilevanti. Inoltre, l'AAT è una risorsa preziosa per la ricerca storica e scientifica, poiché fornisce strumenti per l'approfondimento e la comprensione del contesto. L'accessibilità ai vocabolari è garantita tramite vari formati, tra cui LOD, XML, tabelle relazionali e API, con aggiornamenti periodici. La piattaforma online consente ricerche gerarchiche, ampiamente utilizzate da catalogatori e ricercatori per esplorare i dati in modo sistematico e intuitivo. La missione del *Getty Vocabulary Program* è quella di creare vocabolari completi, autorevoli e strutturati, rispettando gli standard internazionali. Il programma mira a favorire la ricerca e la comprensione delle arti visive attraverso una collaborazione internazionale, con l'obiettivo di ampliare la portata e la copertura dei vocabolari, rendendoli più multilingue, multiculturali e inclusivi. L'intento è rendere i vocabolari rappresentativi delle priorità globali nel campo della storia dell'arte (J. Paul Getty Trust, rivisitato il 26 luglio 2024).

L’AAT è stato creato negli anni '70, quando le biblioteche d'arte e i servizi di indicizzazione delle riviste artistiche, impegnati nella digitalizzazione dei loro processi, esprimevano la necessità di un vocabolario controllato. Ben presto, anche i catalogatori di oggetti museali e risorse visive riconobbero l'importanza di un sistema terminologico standardizzato per promuovere la coerenza e migliorare il recupero delle informazioni. Inizialmente, l'AAT fu costruito attingendo da terminologie già esistenti, come quelle presenti negli elenchi di autorità e nella letteratura accademica di storia dell'arte e architettura, con il supporto di esperti provenienti da vari ambiti disciplinari. Nel 1981 furono stabiliti i principi di base per la costruzione e gestione dell'AAT, che nel tempo ha subito revisioni per favorire un maggiore multilinguismo e inclusività. L'AAT si basa su una struttura gerarchica, ispirata al MeSH ed è continuamente aggiornato e validato dalla comunità accademica di storia dell'arte e architettura. Dal 1997 è distribuito solo in formato digitale, diventando liberamente accessibile sotto la licenza Open Data Commons Attribution (ODC-By) 1.0. Oggi, il programma continua a evolversi, con una gestione editoriale a cura di un team internazionale e di un supporto tecnico da parte del Getty Digital Division.

L'AAT, quindi, è capace di offrire una classificazione dettagliata dei settori artistici e culturali. L’AAT è suddiviso in varie categorie, chiamate *facets*, che raccolgono e organizzano i termini in modo sistematico. Ogni *facet* comprende termini legati a specifici aspetti dell'arte, dell'architettura e delle opere visive. Di seguito le principali categorie e le loro gerarchie:

- Concetti Associati: include termini che si riferiscono a caratteristiche mentali, sociali o di ruolo legate all’arte, all’architettura e alle opere visive;
- Attività: raccoglie termini relativi a settori di attività, azioni fisiche e mentali, eventi specifici, metodi e processi, come archeologia, ingegneria e analisi;
- Materiali: include termini per sostanze fisiche, naturali o sintetiche, utilizzate nella realizzazione di opere artistiche o architettoniche, come ferro, argilla e colori;
- Oggetti: la categoria più ampia, che comprende oggetti tangibili creati dall'uomo, come dipinti, sculture, documenti scritti, edifici e giardini;
- Marchi: contiene i nomi di marchi protetti da copyright, come Agfacolor e Araldite.

Ogni categoria è organizzata in gerarchie, dove un termine più ampio (genere) fornisce il contesto per termini più specifici (specie).

Le gerarchie possono essere suddivise in ulteriori sottoinsiemi, come:

- Attribuzioni fisiche (es. attributi, proprietà, effetti)
- Stili e periodi (es. movimenti artistici, periodi storici)
- Agenti (es. persone, organizzazioni, esseri viventi)

- Attività (es. discipline, tecniche)
- Materiali (es. tipi di materiali)
- Oggetti (es. gruppi di oggetti, generi di oggetti)

Questa organizzazione facilita la ricerca e la catalogazione nel contesto dell'arte e dell'architettura.
L'AAT utilizza una struttura poligerarchica, in cui ogni concetto, identificato tramite *Subject_ID*, può essere collegato a più *Parent_ID*. Ogni concetto è accompagnato da informazioni dettagliate, come il tipo di record e la nota [N] che descrive il termine nel contesto. Nel caso di concetti gerarchici, i termini sono generalmente ordinati alfabeticamente, ma possono essere anche organizzati in modo logico, ad esempio, cronologicamente. Le relazioni storiche sono indicate da un flag "H" (Historical), mentre quelle correnti non sono esplicitamente evidenziate (Getty Research Institute, 2024).
I concetti sono inoltre classificati in vari tipi di relazioni gerarchiche, come:

- Genus/Species (G): Un concetto appartiene a una categoria più ampia
- Whole/Part (P): Relazione di parte e tutto.
- Instance (I): Un esempio specifico di un concetto.

Oltre a queste relazioni gerarchiche, il thesaurus include anche relazioni associative che collegano concetti in modo diverso, come nel caso dei materiali utilizzati per la produzione di un oggetto.
Ogni voce dell'AAT è inoltre associata a un *contributore*, che rappresenta l'ente o il progetto che ha fornito il termine, e a una fonte che giustifica l'inclusione del concetto. Queste informazioni sono accessibili agli utenti per garantire trasparenza e un corretto riferimento alle origini dei dati.
Infine, il thesaurus conserva un registro delle revisioni che documenta le modifiche, i cambiamenti e le aggiunte ai record. Questa cronologia è fondamentale per chi gestisce l'aggiornamento delle informazioni, anche se non è visibile agli utenti finali (Getty Research Institute, 2024).
Quando si cerca, ad esempio, 'computer' nel sistema AAT, il vocabolario restituisce una serie di risultati e concetti correlati. La ricerca include 25 concetti, tra i quali si trova "computer analogico", descritto come "computer che eseguono calcoli manipolando variabili fisiche continue (come tempo e tensione) che sono analoghe alle quantità oggetto di calcolo"[19].Questo concetto è classificato come "data processing equipment" (equipaggiamento per l'elaborazione dei dati), sotto una gerarchia più ampia di "furnishings and equipment" (arredi e attrezzature), come è possibile visualizzare nella seguente immagine:

[19] Le frasi citate sono tradotte dall'autore di questa tesi.

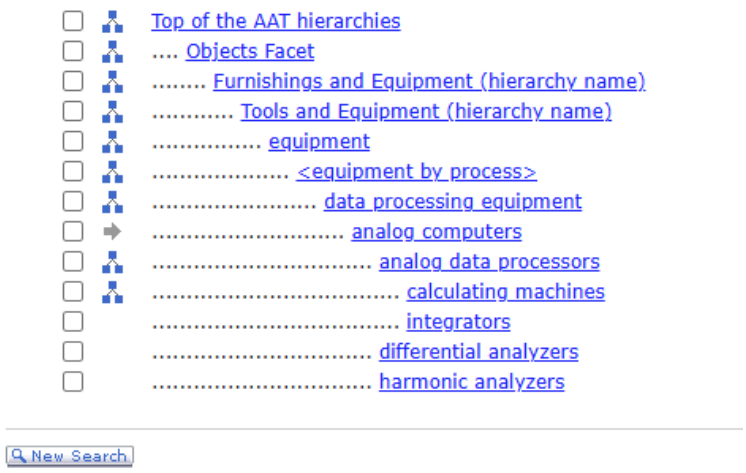


Figure 9 Struttura gerarchica di AAT: percorso terminologico per "analog computers".

Il sistema riconosce come termine preferito (preferred term) *analog computers* e tra i sinonimi come *analog computer*, *analogue computers*, *computers, analog*, riconosce anche *calculateur analogique* in lingua francese.

I progetti che integrano l'AAT sono numerosi, come nel caso del Castello di Torrelobatón, dove la modellazione 3D HBIM viene arricchita con dati semantici tratti dall'AAT (López *et al.*, 2018), e nel progetto AAT-Taiwan, che sviluppa una versione multilingue dell'Art & Architecture Thesaurus per sostenere l'integrazione terminologica a livello globale (Chen *et al.*, 2010).

Nonostante i notevoli progressi compiuti nel campo delle risorse terminologiche e ontologiche per il patrimonio culturale, esistono ancora significativi limiti che ne ostacolano un uso pieno e integrato. Pur essendo utile per descrivere oggetti materiali e concetti artistici, il vocabolario non si estende in modo specifico al dominio informatico o alle sue applicazioni nel patrimonio culturale.

Alcune risorse, come il Computing Classification System (CCS) dell'ACM, si rivolgono specificamente all'informatica, offrendo una classificazione dettagliata che può potenzialmente dialogare con l'AAT per creare un'integrazione più robusta.

2.2.2 Computing Classification System (CCS): Sistema di classificazione per i concetti informatici

Il Computing Classification System (CCS), sviluppato dall'Association for Computing Machinery (ACM), rappresenta un sistema tassonomico avanzato per organizzare e classificare i concetti della disciplina informatica. La versione più recente, pubblicata nel 2012, contiene solo circa 2K di ricerca argomenti ed è stata concepita come vocabolario controllato poligerarchico, ottimizzato per applicazioni nel web semantico. Questa versione ha sostituito il sistema del 1998, chiamato *Computing Reviews Classification System*, o CRCS fino al 1995, che era il riferimento standard per la classificazione nel campo dell'informatica.

Integrato nelle funzionalità di ricerca e nelle visualizzazioni tematiche della Digital Library dell'ACM, il CCS si basa su un vocabolario semantico che riflette lo stato dell'arte del settore. La sua struttura è progettata per adattarsi alle evoluzioni della disciplina, includendo strumenti per l'applicazione di categorie CCS ai nuovi articoli e per mantenerlo aggiornato. La visualizzazione della tassonomia offre sia una modalità interattiva che una vista piatta, ed è disponibile anche un file SKOS scaricabile per integrare la risorsa in sistemi esterni. Inoltre, il CCS è diventato un elemento chiave per l'evoluzione delle funzionalità di ricerca, come la creazione di interfacce dedicate al reperimento di esperti, affiancando i metodi tradizionali di ricerca bibliografica (ACM, 2012).

Il CCS 2012 dell'ACM è strutturato come un albero gerarchico a sei livelli, progettato per offrire una tassonomia dettagliata che consente di organizzare e recuperare in modo preciso informazioni nel campo dell'informatica, ad esempio all'interno della Digital Library dell'ACM. Le categorie principali sono[20]:

- General and reference
- Hardware
- Computer systems organization
- Networks
- Software and its engineering
- Theory of computation
- Mathematics of computing
- Information systems
- Security and privacy
- Human-centered computing
- Computing methodologies
- Applied computing
- Social and professional topics

Di seguito un esempio di organizzazione a sei livelli per la categoria *Software and its engineering*:

1. Software and its engineering (*Software e la sua ingegneria*)

[20] https://dl.acm.org/ccs

2. Software organization and properties (*Organizzazione e proprietà del software*)
    3. Contextual software domains (*Domini contestuali del software*)
        4. Software infrastructure (*Infrastrutture software*)
            5. Middleware (*Middleware*)
                6. Message-oriented middleware (*Middleware orientato ai messaggi*)

A seguire un esempio di classificazione completa della categoria *Hardware*:

Table 1 Classificazione completa della categoria "Hardware" nel CCS.

Classe	**Termini più specifici** (Narrower terms-NT)	**Termini correlati** (Broader terms-BT)
Hardware	Printed circuit boards	•Electromagnetic interference and compatibility •PCB design and layout
	Communication hardware, interfaces and storage	•Signal processing systems: Digital signal processing, Beamforming, Noise reduction •Sensors and actuators •Buses and high-speed links •Displays and imagers •External storage •Networking hardware •Printers •Sensor applications and deployments •Sensor devices and platforms •Sound-based input/output •Tactile and hand-based interfaces: Touch screens, Haptic devices •Scanners •Wireless devices •Wireless integrated network sensors •Electro-mechanical devices
	Integrated circuits	•3D integrated circuits •Interconnect: Input/output circuits, Metallic interconnect.. •Semiconductor memory: Dynamic memory, Static memory.. •Digital switches: Transistors, Logic families •Logic circuits: Arithmetic and datapath circuits.. •Reconfigurable logic and FPGAs: Hardware accelerators, High-speed input/output..
	Very large scale integration design	•3D integrated circuits •Analog and mixed-signal circuits: Data conversion, Clock generation and timing.. •Application-specific VLSI designs: Integrated circuits, Instruction set processors, Specific processors •Design reuse and communication-based design: Network on chip, System on a chip.. •Design rules •Economics of chip design and manufacturing •Full-custom circuits •VLSI design manufacturing considerations •On-chip resource management •On-chip sensors •Standard cell libraries •VLSI Packaging: Die and wafer stacking,Input / output styles,Multi-chip modules,Package-level interconnect •VLSI system specifications and constraints
	Power and energy	•Thermal issues:Temperature monitoring,Temperature simulation and estimation.. •Energy generation and storage: Batteries,Fuel-based energy, Renewable energy, Reusable energy storage •Energy distribution: Energy Metering, Power conversion,Power Networks, Smart grid •Impact on the environment •Power estimation and optimization: Switching devices power issues, Interconnect power issues..
	Electronic design automation	•High-level synthesis: Datapath optimization, Hardware-software codesign, Resource binding and sharing.. •Logic synthesis: Combinational synthesis, Circuit optimization,Sequential synthesis.. •Modeling and parameter extraction •Physical design: Clock-network synthesis, Packaging,Partitioning and floorplanning, Placement.. •Timing analysis: Electrical-level simulation,Model-order reduction,Compact delay models.. •Methodologies for EDA: Best practices for EDA, Design databases for EDA, Software tools for EDA
	Hardware validation	•Functional verification: Model checking,Coverage metrics,Equivalence checking,Semi-formal verification.. •Physical verification: Design rule checking, Layout-versus-schematics, Power and thermal analysis.. •Post-manufacture validation and debug: Bug detection, Localization and Diagnosis, Bug fixing,.
	Hardware test	•Analog, mixed-signal and radio frequency test •Board- and system-level test •Defect-based test •Design for testability: Built-in self-test, Online and diagnostics, Test data compression •Fault models and test metrics •Memory test and repair •Hardware reliability screening •Test-pattern generation and fault simulation •Testing with distributed and parallel systems
	Robustness	•Fault tolerance:Error detection and error correction,Failure prediction,Failure recovery.. •Design for manufacturability :Process variations,Yield and cost modeling,Yield and cost optimization •Hardware reliability: Aging of circuits and systems,Circuit hardening, Early-life failures and infant mortality.. •Safety critical systems
	Emerging technologies	•Analysis and design of emerging devices and systems: Emerging architectures,Emerging languages and compilers.. •Biology-related information processing:Bio-embedded electronics,Neural systems •Circuit substrates:III-V compounds,Carbon based ectronics,Cellular neural networks.. •Electromechanical systems:Microelectromechanical systems,Nanoelectromechanical systems •Emerging interfaces •Memory and dense storage •Emerging optical and photonic technologies •Reversible logic •Plasmonics •Quantum technologies: Single electron devices,Quantum computation(Quantum communication and cryptography..).. •Spintronics and magnetic technologie

Entrando nella sezione di ogni categoria, vengono presentati gli articoli pertinenti al concetto, pubblicati su ACM.

Sebbene il CCS sia una risorsa specialistica solida e ampiamente adottata, esso non affronta direttamente l'intersezione tra informatica e patrimonio culturale, limitandone l'applicabilità a iniziative che richiedono una visione interdisciplinare. Ciò crea un vuoto nella definizione di un lessico specializzato in grado di supportare la rappresentazione della conoscenza in questo settore specifico.

Anche all'interno del dominio informatico, il CCS presenta alcune limitazioni, in particolare perché la sua ultima versione risale al 2012. La tassonomia copre un numero relativamente ristretto di argomenti e presenta una granularità limitata in alcune aree, come la gestione delle reti e dei servizi (Santos *et al.*, 2016). Di conseguenza, sebbene il CCS offra una struttura utile per molte applicazioni, emergono esigenze di strumenti più avanzati per rispondere alle sfide dei domini interdisciplinari.

2.2.3 Computer Science Ontology (CSO): ontologia per la classificazione delle scienze informatiche

Come già accennato nei paragrafi precedenti, tra gli strumenti che offrono strutture concettuali orientate alla classificazione e organizzazione del dominio dell'informatica si annovera anche la *Computer Science Ontology* (CSO), sviluppata presso il Knowledge Media Institute dell'Open University, disponibile nella sua versione aggiornata sul sito ufficiale [21]. La CSO è un'ontologia completa che mappa e organizza le principali aree e sub-discipline dell'informatica, ed è generata automaticamente utilizzando l'algoritmo *Klink-2* applicato al dataset *Rexplore*. Questo dataset include 26.000 argomenti e 226.000 relazioni semantiche, su circa 16 milioni di pubblicazioni, principalmente nel campo dell'informatica. L'algoritmo *Klink-2* combina tecnologie semantiche, apprendimento automatico e conoscenze provenienti da fonti esterne per creare un'ontologia popolata con concetti e relazioni specifici delle aree di ricerca. Sebbene la generazione sia principalmente automatizzata, alcune relazioni sono state riviste manualmente da esperti durante la preparazione di due survey assistite da ontologie nei campi del *Semantic Web* e del *Software Architecture* (Salatino *et al.*, 2020). La CSO ha come radice principale l'informatica (*Computer Science*), ma include anche radici secondarie, come *Linguistics, Geometry, Semantics*, e altre.

La CSO offre due vantaggi principali rispetto a classificazioni manuali come CCS dell'ACM del 2012:

- Dettaglio e granularità: consente di caratterizzare aree di ricerca ad alto livello mediante centinaia di sotto-argomenti e termini correlati, mappando termini specifici a contesti di ricerca più ampi. Ad esempio, come vedremo nella *Figura 13* la CSO caratterizza il *Semantic Web* con circa 40 sottocategorie, tra cui *Linked Data*, *Semantic Web Services*, *Ontology Matching, SPARQL, OWL, SWRL* e molti altri. Questo livello di dettaglio offre una rappresentazione più granulare rispetto ad altre classificazioni, come quella dell'ACM, che include solo tre concetti correlati: *Semantic web description languages*, *Resource Description Framework* (RDF) e *Web Ontology Language* (OWL);
- Aggiornamento dinamico: può essere facilmente aggiornata eseguendo l'algoritmo Klink-2 su un nuovo set di pubblicazioni, garantendo che l'ontologia rimanga attuale e pertinente nel tempo.

Il modello di dati della CSO è un'estensione di SKOS e include otto relazioni semantiche principali, che permettono di esprimere connessioni tra concetti in modo chiaro e preciso:

[21] Computer Science Ontology (CSO): https://cso.kmi.open.ac.uk/

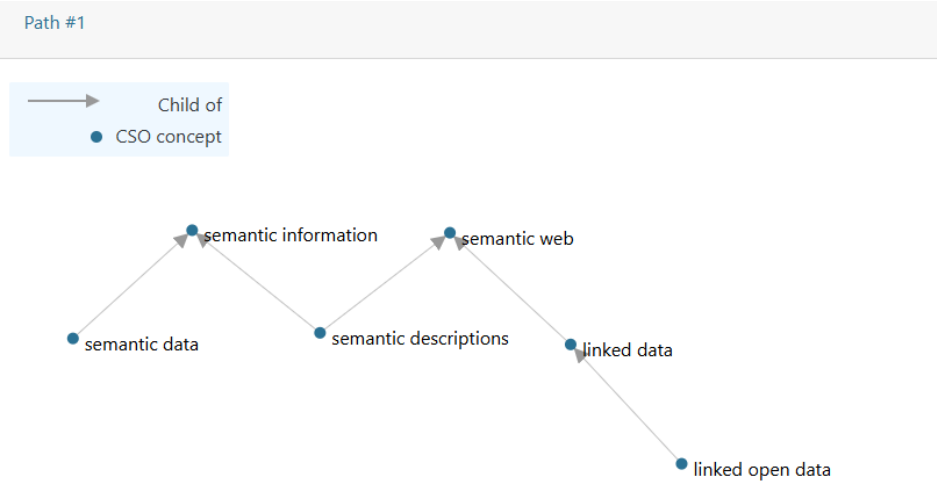
- relatedEquivalent: indica che due argomenti possono essere trattati come equivalenti per esplorare i dati di ricerca (es. *Ontology Matching*, *Ontology Mapping*).
- superTopicOf: indica che un argomento è una sotto-area di un altro (es. *Linked Data*, *Semantic Web*).
- contributesTo: indica che i risultati di ricerca di un argomento contribuiscono a un altro, senza che quest'ultimo sia necessariamente una sotto-area (es. *Ontology Engineering* contribuisce al *Semantic Web*).
- preferentialEquivalent: specifica l'etichetta principale per argomenti correlati (es. *ontology* e *ontologies* avranno entrambi come *preferentialEquivalent* il termine *ontology*).
- rdf:type: dichiara che una risorsa è un'istanza di una classe (es. un argomento è un'istanza del concetto *topic*).
- rdfs:label: fornisce una versione leggibile del nome della risorsa.
- owl:sameAs: collega concetti della CSO ad altre ontologie del Linked Open Data Cloud (es. DBpedia, Wikidata).
- schema:relatedLink: collega i concetti della CSO a pagine web correlate, come articoli di Wikipedia o risorse su Microsoft Academic (Salatino *et al.*, 2020).

Ogni risorsa è accessibile tramite il proprio URI. Ad esempio, la risorsa "semantic web" è navigabile al seguente indirizzo: https://cso.kmi.open.ac.uk/topics/semantic_web. Il portale CSO permette di accedere alle risorse in diversi formati, adattandosi alle esigenze degli utenti e delle applicazioni, tra cui:

- HTML (esplorabile via browser).
- RDF/XML, Turtle, JSON-LD, N-Triples (compatibili con formati per Linked Data).

La flessibilità e l'automazione della CSO la rendono uno strumento avanzato per la rappresentazione della conoscenza nel dominio accademico e per il supporto a ricerche interdisciplinari. Inoltre, grazie alla sua struttura, facilita l'integrazione con altre ontologie, contribuendo a migliorare la scoperta di nuove connessioni tra aree di ricerca (Salatino *et al.*, 2020).

Una delle caratteristiche distintive della CSO è la capacità di calcolare la distanza tra i concetti, tracciando il percorso attraverso le relazioni semantiche che li legano. Ad esempio, se si inseriscono due argomenti nella barra di ricerca, come 'semantic data' e 'linked open data', l'algoritmo è in grado di identificare e presentare i percorsi (path) tra i due concetti, evidenziando i passaggi intermedi e le relazioni semantiche che li collegano:



Subject	Predicate	Object
semantic data	child of	semantic information
semantic information	parent of	semantic descriptions
semantic descriptions	child of	semantic web
semantic web	parent of	linked data
linked data	parent of	linked open data

Figure 10 Risultato di ricerca nel sistema CSO, che mostra il calcolo della distanza semantica tra i concetti "semantic data" e "linked open data", evidenziando i percorsi intermedi e le relazioni semantiche tra di essi.

L'immagine sottostante mostra una visualizzazione grafica del concetto di "Semantic Web" nella CSO ed esplora le relazioni tra il "Semantic Web" e altri concetti ad esso correlati.

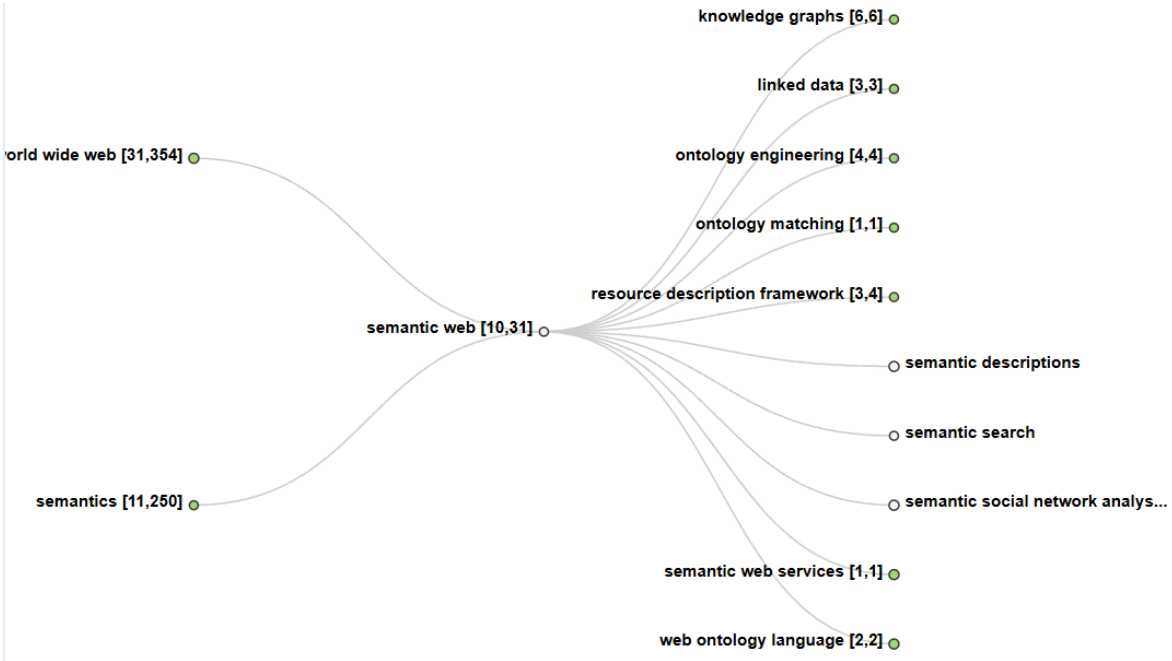


Figure 11 Visualizzazione grafica del concetto di Semantic Web nella CSO.

Nel grafo, il nodo centrale rappresenta il concetto di “Semantic Web”, che è connesso a una serie di concetti secondari, tra cui *World Wide Web*, *Semantics, Linked Data*, *Ontology Engineering*, *Ontology Matching*, e *Resource Description Framework*. Ogni nodo è accompagnato da un numero, che indica rispettivamente il numero di sotto-argomenti (figli diretti) e discendenti complessivi a partire dal concetto principale.

La visualizzazione è interattiva, permettendo di esplorare le relazioni tra i vari concetti e di approfondire la comprensione delle interconnessioni in ambito informatico. Inoltre, sono previste diverse opzioni per personalizzare la visualizzazione del grafo, consentendo di esplorare il concetto in modalità compatta o dettagliata. È anche possibile espandere i nodi per visualizzare ulteriori dettagli, facilitando un'analisi più approfondita delle relazioni tra i concetti.

La creazione automatica di ontologie richiede algoritmi capaci di comprendere il contesto e le relazioni semantiche profonde tra i concetti, dati di alta qualità, ricchi e ben strutturati su cui addestrare i modelli ed una capacità di generalizzazione che vada oltre le semplici correlazioni statistiche.

Tuttavia, il problema non è solo tecnico. Anche dal punto di vista umano, la costruzione di ontologie implica scelte soggettive e contestuali che sono difficili da replicare con precisione tramite strumenti automatizzati.

Una possibile soluzione potrebbe essere la combinazione di metodi automatici e contributi manuali, soprattutto da parte di esperti del dominio, sfruttando la velocità e la scalabilità degli algoritmi insieme alla profondità e alla creatività dell'intervento umano.

2.3 Integrazione di ontologie e Web Semantico nel patrimonio culturale: Il caso del MuseumFinland

La crescente complessità dei dati e la necessità di garantire un'interoperabilità sempre maggiore tra sistemi hanno portato allo sviluppo di approcci avanzati per la rappresentazione della conoscenza. L'applicazione delle tecnologie semantiche nel settore dei beni culturali ha trasformato radicalmente le modalità di gestione, rappresentazione e fruizione della conoscenza. L'uso di ontologie e linked data ha permesso di superare le tradizionali limitazioni dei sistemi informativi chiusi, favorendo l'interoperabilità tra database museali, archivi e collezioni digitali.

Un esempio significativo di applicazione delle tecnologie del Web Semantico nel dominio dei beni culturali è il *MuseumFinland*[22], un portale semantico che sfrutta ontologie e standard semantici per armonizzare e interconnettere le informazioni museali provenienti da diverse istituzioni. Sviluppato dal Semantic Computing Research Group (SeCo), il portale, attraverso RDF e OWL, mira a strutturare le collezioni museali eterogenee secondo un modello condiviso, consentendo agli utenti di effettuare ricerche avanzate basate su relazioni concettuali piuttosto che su semplici parole chiave.

Questo approccio non solo potenzia il recupero delle informazioni, ma favorisce anche nuove forme di analisi e connessione tra oggetti culturali, che altrimenti rimarrebbero frammentati. In tal modo, arricchisce le modalità di esplorazione e fruizione del patrimonio, migliorando significativamente l'esperienza degli utenti.

Secondo Hyvönen *et al.* (2005) MuseumFinland è stato concepito con lo scopo di:

• Integrare collezioni distribuite: creare una vista globale delle collezioni museali, trattandole come un archivio unificato nonostante la loro eterogeneità.

• Supportare la ricerca basata sui contenuti: offrire strumenti di ricerca intelligenti che utilizzano concetti ontologici piuttosto che una semplice corrispondenza di parole chiave.

• Evidenziare le relazioni semantiche: rendere visibili agli utenti le connessioni implicite tra gli oggetti delle collezioni.

• Fornire un canale di pubblicazione economico: permettere ai musei di condividere i propri dati senza necessità di investimenti tecnologici significativi.

Inoltre, l'architettura del MuseumFinland si distingue per l'uso di un modello basato su ontologie condivise, che consente di superare le barriere semantiche tra le diverse collezioni museali. In particolare, il progetto si avvale di:

[22] Disponibile su: https://seco.cs.aalto.fi/applications/museumfinland/tutorial/

• Navigazione multidimensionale: gli utenti possono esplorare le collezioni attraverso molteplici dimensioni, come il materiale degli oggetti, il periodo storico, o il luogo di provenienza, migliorando significativamente l'esperienza di ricerca.
• Collegamenti dinamici: le pagine del portale sono collegate dinamicamente in base alle relazioni semantiche tra gli oggetti, permettendo un'esplorazione fluida e contestuale.
• Ontologie come base comune: tutti i dati sono organizzati utilizzando un insieme di ontologie standard, che garantiscono l'interoperabilità tra le collezioni.

Nel 2004, MuseumFinland è stato premiato con il *Semantic Web Challenge Award*, a testimonianza del suo contributo innovativo nel campo delle tecnologie semantiche. Il portale non solo ha migliorato la fruizione delle collezioni museali finlandesi, ma ha anche fornito un modello replicabile per altre istituzioni culturali interessate a implementare soluzioni basate sul Web Semantico.
Sebbene risalga al 2004, MuseumFinland rappresenta ancora oggi un esempio significativo di come l'uso di ontologie e piattaforme interoperabili possa contribuire a superare le limitazioni strutturali nella gestione delle risorse culturali, come discusso nelle sezioni precedenti di questo capitolo. Il progetto ha anticipato molte delle attuali pratiche nel campo del patrimonio culturale digitale, mostrando il potenziale di un approccio fondato su standard condivisi e strumenti di interrogazione semantica per promuovere la collaborazione tra istituzioni e rendere i contenuti culturali più accessibili, interconnessi e intelligibili a livello globale.

2.4 Ontoterminologia applicata al patrimonio culturale: il caso dell'archeologia di al-Andalus

Un ulteriore caso di studio rilevante è costituito dal progetto di rappresentazione ontologica della ceramica islamica nell'archeologia di al-Andalus, progetto iniziato nel 2012, Roche (2012) conducendo allo sviluppo dell'ontologia OntoAndalus (Almeida e Costa, 2021). Tale iniziativa evidenzia il ruolo delle ontologie nella formalizzazione e organizzazione delle conoscenze archeologiche, permettendo una modellizzazione dettagliata delle categorie ceramiche e delle loro caratteristiche distintive. L'utilizzo del linguaggio OWL consente di rappresentare, in maniera rigorosa e computabile, le relazioni tra classi di manufatti, stili decorativi e tecniche di produzione, facilitando non solo la ricerca ma anche il ragionamento automatico sui dati delle collezioni archeologiche.

L'ontoterminologia, affrontata nel primo capitolo, come sottolineato da Roche (2012), permette di separare la dimensione linguistica da quella concettuale, standardizzando la conoscenza pur preservando al contempo la diversità linguistica. Tale approccio si rivela particolarmente utile nell'affrontare le sfide legate alla rappresentazione della conoscenza nel dominio dei beni culturali. Nel progetto relativo all'archeologia di al-Andalus, si pone l'accento sull'importanza di integrare terminologie specialistiche con strutture ontologiche formali. Se da un lato la terminologia si focalizza sulla descrizione linguistica dei concetti, dall'altro le ontologie forniscono un quadro strutturato che consente di definire relazioni gerarchiche e semantiche tra gli elementi, come discusso nel primo capitolo. Questo metodo ibrido non solo preserva la ricchezza semantica dei termini specialistici, ma ne migliora anche la coerenza e la fruibilità nei sistemi informativi digitali.

La ceramica, una delle risorse più durature e significative della cultura islamica in al-Andalus, riveste un ruolo centrale nello studio archeologico, specialmente in Spagna e Portogallo. Lo studio della ceramica fornisce informazioni sulla vita quotidiana, le abitudini alimentari, le relazioni commerciali, lo sviluppo tecnico e la simbologia di quella società. I termini usati per classificare i tipi di ceramica in al-Andalus sono strettamente legati all'analisi tipologica degli artefatti, come il caso del lavoro pionieristico di Rosselló-Bordoy in Spagna, che ha identificato varie serie di ceramiche. Il lavoro di Rosselló-Bordoy e del gruppo CIGA in Portogallo ha portato alla creazione di una tipologia e di una terminologia specifiche per la ceramica islamica, che include descrizioni testuali e illustrazioni di artefatti tipici. Ogni tipo di ceramica viene definito da caratteristiche formali e funzionali (Almeida, Roche, Costa, 2018).

Attualmente, la conoscenza condivisa da archeologi portoghesi e spagnoli è rappresentata attraverso l'ontologia OWL in quanto consente l'utilizzo di ragionatori per la classificazione automatica dei concetti. Questa ontologia permette di rappresentare i vari tipi di ceramica, distinguendo classi con

ampiezze diverse, come “oggetti da tavola” o “oggetti da illuminazione”. Un esempio pratico di formalizzazione, fornito da Almeida *et al*. (2018), riguarda la tipologia degli oggetti da illuminazione, che include categorie come “candil” e “lanterna”.

Nella tipologia proposta dal gruppo CIGA, gli oggetti da illuminazione in ceramica sono suddivisi in varie classi, tra cui “candil”, “candeia”, “candeia de pé” e “lanterna”, formalizzate nell'ontologia OWL tramite il principio di “genere-differenza”. Tale principio permette di distinguere vari tipi di lampade ceramiche, sia chiuse che aperte, e tra varianti con o senza becco per la fiamma. Il sistema concettuale proposto offre una rappresentazione precisa delle caratteristiche essenziali degli artefatti, come la funzionalità e la parte del corpo, facilitando una classificazione automatica e accurata.

Questo lavoro dimostra come la classificazione archeologica tradizionale possa essere formalizzata e arricchita tramite l'ontoterminologia, permettendo di rappresentare in modo preciso e interconnesso la conoscenza sugli artefatti archeologici.

Progetti come MuseumFinland e lo studio di Almeida *et al.* evidenziano il potenziale delle ontologie e delle tecnologie semantiche nella gestione dei dati culturali. Questo approccio si allinea perfettamente con gli obiettivi del progetto ItinHeritage, discusso nel quarto capitolo, che mira a sviluppare una piattaforma interdisciplinare in grado di integrare tecnologie semantiche e terminologie specifiche, ottimizzando la gestione e la fruizione del patrimonio culturale digitale.

2.5 Mondo IT e gestione dei dati museali

Le ontologie esistenti di cui si è discusso nei paragrafi precedenti, come il Computing Classification System dell'ACM, e i modelli ontologici applicabili al patrimonio culturale, come l'Art & Architecture Thesaurus del Getty Research Institute o il progetto MuseumFinland, rappresentano risorse fondamentali per l'organizzazione e la gestione della conoscenza. Tuttavia, tali strumenti si concentrano prevalentemente sulla classificazione e gestione di dati tangibili e intangibili, senza affrontare in modo specifico le esigenze interdisciplinari legate all'informatica applicata ai beni culturali o alla trasmissione linguistica della conoscenza tecnica in questo settore.

Un approccio innovativo risiede nell'integrazione di ontologie e terminologie, come si è potuto constatare dallo studio di Almeida *et al.*, ma con il fine di rappresentare il patrimonio culturale pienamente interoperabile con il mondo informatico, arricchendo il tutto con una logica interdisciplinare. Infatti, come abbiamo visto, uno degli ambiti meno esplorati del dominio informatico riguarda proprio le risorse museali. Il patrimonio culturale non solo raccoglie una molteplicità di significati, ma offre anche l'opportunità di integrare il mondo informatico nelle sue sfaccettature, conferendo un senso alla rappresentazione concettuale attraverso la valorizzazione degli strumenti tecnologici che hanno accompagnato il nostro sviluppo e che continuano a evolversi. In tale contesto, si rende sempre più rilevante l'esigenza di sviluppare strumenti e metodologie in grado di tracciare e documentare l'interazione tra tecnologie emergenti e patrimonio culturale, monitorando l'evoluzione di tali strumenti e valutandone l'impatto sulla conservazione e sull'accessibilità del patrimonio stesso. Questo è l'aspetto innovativo. Tali necessità trovano una rispondenza pratica nel bisogno di colmare il divario tra gli strumenti esistenti e le richieste di un dominio complesso e altamente specializzato.

L'analisi dello stato dell'esistente ha evidenziato, oltre alle risorse e agli strumenti già consolidati per la gestione terminologica e ontologica, anche le lacune che limitano il potenziale dell'informatica applicata al patrimonio culturale. Nonostante le opportunità di innovazione, le sfide sono molteplici. Come evidenziato in precedenza, uno dei principali ostacoli nell'ambito dell'informatica applicata al patrimonio culturale è la carenza di una classificazione terminologica coerente. L'assenza di un framework interdisciplinare che integri linguistica, informatica e patrimonio culturale rappresenta una lacuna significativa da colmare.

Gli attuali sistemi di classificazione, dunque, tendono a includere termini fuori contesto e privi di una strutturazione tematica coerente. L'applicazione di modelli concettuali e ontologie informatiche al settore museale è ancora limitata. I domini mappati tendono a fare riferimento esclusivamente alla teoria generale dell'informatica, senza approfondire le specificità del settore. I glossari e i thesauri esistenti, come abbiamo visto, risultano spesso incompleti o inadeguati, trattando termini che si

riferiscono limitatamente a software, tecnologie o teorie generali, o in maniera frammentata, senza un'organizzazione semantica che colleghi i vari sottodomini. Si assiste così alla creazione di un conglomerato di concetti che spaziano da quelli legati a software e applicazioni aziendali fino a quelli teorici e concettuali. Da un lato, schemi tassonomici come il **CCS** tendono a classificare rigidamente le discipline informatiche, imponendo gerarchie fisse tra concetti. Dall'altro, modelli come il **CIDOC-CRM** privilegiano una rappresentazione narrativa e relazionale della conoscenza, descrivendo processi, eventi e relazioni tra oggetti culturali. Il **CSO**, invece, adotta un approccio diverso: non mira a fornire una rappresentazione narrativa né a classificare rigidamente l'intero dominio informatico, ma si concentra sulle aree di ricerca disciplinari. In questo senso, le relazioni tra concetti nel CSO riflettono gerarchie e correlazioni tematiche tra campi scientifici specifici, senza estendersi agli aspetti applicativi, storici o concreti del dominio. Questo squilibrio solleva interrogativi sulla possibilità di un'integrazione più efficace tra questi due approcci, al fine di migliorare la gestione delle collezioni digitali e del patrimonio informatico.

Inoltre, emerge una significativa lacuna per quanto concerne la prospettiva diacronica, necessaria per connettere la dimensione storica a quella teorica del patrimonio informatico. Le risorse terminologiche attualmente disponibili nel dominio IT appaiono spesso limitate a strutture tassonomiche rigide, basate su gerarchie di categorie e sottocategorie che non beneficiano di un'analisi lessicografica sistematica o di una mappatura rigorosa delle relazioni semantiche tra i termini. In assenza di una metodologia terminologica strutturata, tali strumenti faticano a riflettere l'evoluzione del lessico nel tempo, mancando di una reale integrazione con le fonti storiche. Questo limite impedisce di documentare le trasformazioni concettuali dei termini, rendendo difficile la ricostruzione dei legami semantici che uniscono le diverse fasi dello sviluppo tecnologico.

Anche in un paese come l'Italia, ricco di patrimonio culturale, la mancanza di piattaforme integrate e aggiornate per la gestione terminologica e ontologica specifica del settore, in lingua italiana, rappresenta un ulteriore limite. La scarsità di un contesto specifico per il patrimonio culturale, unita alla frammentazione dei sistemi di catalogazione, contribuisce a ridurre l'efficacia degli strumenti disponibili. Pur esistendo risorse e standard internazionali di riferimento – come il Sistema Informativo Generale del Catalogo (SIGECweb), utilizzato anche per la catalogazione di collezioni specifiche come quelle del Museo degli Strumenti per il Calcolo di Pisa – l'adozione di soluzioni realmente integrate e adattate al contesto nazionale rimane incerta e ancora incompleta.

Come si è visto, il MuseumFinland rappresenta un modello di successo nell'uso di tecnologie semantiche per la valorizzazione del patrimonio culturale. Attraverso l'uso di ontologie, il progetto ha permesso di collegare informazioni provenienti da diverse istituzioni, rendendole accessibili attraverso una piattaforma unificata. Tuttavia, un esempio pratico che potrebbe essere esplorato è

come tali approcci potrebbero essere adattati al contesto italiano, dove la frammentazione tra piccole realtà museali è una problematica diffusa.

Una grande sfida a valle, da tenere in considerazione, è l'eterogeneità degli oggetti conservati nei musei. Gli artefatti museali possono variare notevolmente in termini di tipologia (dipinti, immagini, reperti storici, oggetti tecnologici, ecc.) e natura, ovvero tra oggetti materiali e oggetti immateriali. In Italia, il Codice dei Beni Culturali (D.Lgs. 42/2004, Art. 2) distingue tra beni mobili e immobili con valore artistico, storico, o archeologico, identificandoli come rappresentativi di civiltà.

Il progetto di ricerca ITinHeritage, che sarà approfondito nel quarto capitolo, esplora specificatamente il patrimonio informatico. Anche questo dominio include sia beni materiali, come computer, stampanti, mouse e hardware in generale, guide e documenti, sia beni immateriali, come programmi software, app, videogiochi, formati, protocolli, sistemi operativi e così via.

Questa varietà rende difficile l'integrazione e l'interoperabilità tra i sistemi, specialmente quando i dati devono essere trasferiti o condivisi agli utenti o con altre istituzioni.

I modelli come AAT, CIDOC-CRM ed EDM hanno già formalizzato la terminologia museale garantendo una chiara distinzione tra i concetti, tuttavia il problema dell'integrazione accurata con modelli informatici rimane aperto. Il principale ostacolo non è solo la differenza di significato di alcuni termini tra i due ambiti, ma la difficoltà di creare un allineamento semantico tra questi modelli. Ad esempio, il concetto di *sistema* nell'informatica si riferisce a un insieme di componenti software e hardware, mentre nel contesto museale può indicare una collezione di oggetti correlati o una struttura organizzativa. Anche se entrambi i domini definiscono chiaramente questi termini, la loro interoperabilità non è automatica e richiede un mapping esplicito. Allo stesso modo, il termine *architettura* può designare la struttura di un software in ambito informatico, mentre nel settore culturale è associato agli stili e alle costruzioni storiche: senza un vocabolario interdisciplinare condiviso, il rischio di errori di interpretazione nei processi di data mapping rimane elevato.

Il patrimonio informatico si distingue dal patrimonio culturale tradizionale per la sua natura dinamica e in continua evoluzione, una caratteristica che spesso lascia poco tempo per uno studio approfondito e per un'adeguata valorizzazione (Djambian et al, 2024). Inoltre, la sua percezione come "recente" contribuisce a una sottovalutazione della sua importanza, portando a una carenza di patrimonializzazione e a una riflessione critica sistematica ancora insufficiente. Questa percezione di "novità" contribuisce a una sottovalutazione della rilevanza storica e sociale degli artefatti IT, i quali, nonostante la loro genesi cronologicamente vicina, hanno già accumulato una stratificazione significativa di evoluzioni, fallimenti e innovazioni che meritano un'analisi sistematica.

Negli ultimi 70 anni si è passati da computer di prima generazione, che utilizzavano tubi a vuoto, a computer di quinta generazione, includendo concetti come intelligenza artificiale, robotica, reti

neurali e computer sperimentali come computer chimici, quantici, ottici e organici. La conservazione deve includere non solo oggetti fisici, ma anche il contesto umano e sociale che li circonda.
Ogni tipologia di oggetto richiede metadati specifici e controllati (come dimensioni, materiale, epoca, autore, luogo di origine), che rendono complessa la creazione di modelli di dati standardizzati. Tuttavia, la scarsità di standard comuni ostacola l'interoperabilità e la comunicazione tra le macchine. Intorre (2008) sottolinea come i metadati costituiscano la "chiave della comunicazione museale", essenziali per trasformare i dati in conoscenza condivisa.
Sia la metadatazione che la catalogazione fanno parte nel processo di gestione museale. Infatti, un ulteriore problema potrebbe risiedere nell'obsolescenza dei Sistemi di Catalogazione.
Molti musei utilizzano ancora infrastrutture tecnologiche e sistemi di gestione dei dati obsoleti o personalizzati che possono non essere compatibili con gli standard attuali di metadati, come i modelli CIDOC-CRM o Dublin Core, limitando l'integrazione con altre piattaforme e la possibilità di aggiornamenti delle collezioni.
Come evidenziato da Intorre (2008), «la creazione di infrastrutture tecnologiche standardizzate garantisce la durata e l'efficacia dei dati nel tempo».
Un altro aspetto riguarda proprio la conservazione e sostenibilità a lungo termine. La rapida evoluzione tecnologica può rendere obsoleti formati di file e infrastrutture di archiviazione, mettendo a rischio la conservazione dei dati. Il mantenimento dei dati digitali degli artefatti richiede politiche e tecnologie per garantire l'accessibilità a lungo termine. Tale scenario è aggravato dalla difficoltà nel reperire fonti documentarie su temi in costante aggiornamento, in un ambito che si pone come punto di intersezione tra molteplici discipline.
A queste sfide di ordine tecnico si aggiunge un ostacolo di natura strutturale: la formazione spesso insufficiente del personale e la carenza di esperti IT dotati di competenze specifiche applicate al settore museale. L'assenza di standard condivisi comporta che la qualità dei metadati vari sensibilmente in base alla soggettività del catalogatore, generando discrepanze nelle descrizioni e nell'uso della terminologia specialistica. Tali incongruenze portano a discrepanze nei campi compilati, nelle descrizioni e nell'uso di termini specifici, ostacolando la ricerca e il recupero accurato delle informazioni.
Secondo Bianchi (2021), i musei tipicamente utilizzano ad esempio file Word o Excel non centralizzati e non accessibili. La frammentazione delle informazioni e la scarsa centralizzazione rappresentano ostacoli significativi.
Secondo un'indagine condotta da Ekosaari e Pekkola (2019) su 58 esperti tra direttori, curatori e professionisti IT, le principali sfide includono anche:

- Budget insufficienti e costi elevati per la digitalizzazione;

- Problematiche legate alla privacy e ai diritti d'autore: alcuni dati, come quelli relativi a opere d'arte contemporanea o a reperti con un significato culturale sensibile, potrebbero essere soggetti a restrizioni di accesso e diritti d'autore, complicando la gestione dei metadati e la loro distribuzione;
- Risorse limitate che impediscono l'aggiornamento e la manutenzione delle collezioni digitali. La gestione dei dati e dei metadati richiede competenze tecniche e risorse economiche. Molti musei, soprattutto quelli di piccole dimensioni, potrebbero non avere personale adeguato o fondi sufficienti per aggiornare e mantenere le loro collezioni digitali in modo efficace.

Inoltre, la cultura organizzativa dei musei, spesso radicata in tradizioni, può ostacolare l'adozione di nuove tecnologie.

Ekosaari e Pekkola (2019) suggeriscono l'adozione di piattaforme tecnologiche avanzate, come modelli concettuali e l'uso di applicazioni del web semantico, per semplificare l'archiviazione e l'analisi dei dati. Queste tecnologie riducono i costi operativi e garantiscono una maggiore sostenibilità.

Perry (2024) stima che circa 80 milioni di record museali siano ancora offline, limitando la collaborazione e l'innovazione.

L'adozione dell'ontoterminologia come metodologia di standardizzazione linguistica e concettuale emerge come risposta efficace per affrontare queste sfide, soprattutto in un contesto di web semantico, dove sono richiesti dati interoperabili e strutturati.

Le ontologie sono strumenti essenziali per strutturare, rappresentare e gestire la conoscenza in modo organizzato e interoperabile, soprattutto nel contesto museale. Tuttavia, la loro efficacia non può prescindere dall'integrazione con la linguistica, che consente di preservare la diversità terminologica e culturale e di facilitare l'accesso ai contenuti da parte di un pubblico eterogeneo, in vista, soprattutto in un settore in cui vi è una mancanza di corrispondenze tra termini e oggetti, oltre ai problemi di fluidità terminologica. Il dominio informatico, infatti, non dispone di un unico pubblico.

Questa integrazione è particolarmente rilevante anche sotto un profilo etico. La gestione museale implica una responsabilità nella conservazione e divulgazione del patrimonio. I musei, infatti, hanno il compito di divulgare dati accurati e accessibili, offrendo testimonianze del passato e salvaguardando la memoria storica per le generazioni future. La rappresentazione del patrimonio culturale non deve essere limitata a una prospettiva puramente tecnica o concettuale; è essenziale mantenere vivo il legame con le dimensioni culturali e linguistiche che caratterizzano la conoscenza umana.

Roche e Papadopoulou (2020) sottolineano che la capacità di mettere in relazione il livello concettuale e quello linguistico tramite un ambiente tecnologico progettato ad hoc permette di normalizzare ciò che può essere normalizzato, ovvero le conoscenze del dominio, e di tutelare ciò che è fondamentale preservare, cioè la diversità linguistica.

Questo approccio consente ai musei di adempiere alla loro funzione di custodi del patrimonio, garantendo che la memoria e la cultura del passato siano rappresentate in modo inclusivo e universale.

In sintesi, le sfide della gestione dei dati museali e del patrimonio culturale informatico richiedono un approccio integrato, che combini terminologie specifiche, ontologie avanzate e tecnologie semantiche. Queste riflessioni costituiscono il punto di partenza per il capitolo successivo, dedicato all'analisi comparativa dei corpora accademici, comunitari e patrimoniali. Attraverso questa indagine, il progetto ITinHeritage potrà affrontare in modo innovativo la complessità del dominio, traducendo la variabilità linguistica e concettuale osservata in strumenti di rappresentazione terminologica e ontologica capaci di valorizzare il patrimonio informatico in una prospettiva multilingue, comparativa e interoperabile.

Capitolo 3. Approccio comparativo alla terminologia informatica

Nei capitoli precedenti è stato delineato il quadro teorico della terminologia e dell'ontologia (Capitolo 1) e sono stati esaminati gli standard semantici necessari per l'interoperabilità dei dati nel settore dei beni culturali (Capitolo 2). Tuttavia, la definizione di modelli formali e l'adozione di standard internazionali non possono prescindere da una verifica empirica di come la conoscenza venga effettivamente veicolata attraverso il linguaggio naturale.

Il passaggio all'analisi dei corpora in questo terzo capitolo si rende necessario per colmare il divario tra la norma (lo standard) e l'uso (il discorso). Se l'obiettivo finale è la costruzione di una risorsa ontoterminologica per il patrimonio informatico, non è possibile limitarsi a una classificazione *a priori* dei concetti; occorre innanzitutto indagare la variabilità terminologica così come si manifesta nei testi prodotti dalle diverse comunità di riferimento. L'obiettivo principale di questa sezione è indagare le modalità con cui la terminologia informatica si è evoluta e trasformata nel tempo, in relazione ai diversi contesti d'uso e alle comunità coinvolte. L'analisi adotta una prospettiva inclusiva, volta a considerare la pluralità dei registri discorsivi e delle pratiche comunicative, attraverso la selezione di testi provenienti da fonti eterogenee: contributi accademici (articoli scientifici, atti di convegni), materiali divulgativi (riviste di settore, manuali per utenti), e contenuti generati dagli utenti su piattaforme digitali come blog, forum e community online.

Per raggiungere tale scopo, la metodologia adottata non si limita a una ricognizione statistica isolata, ma segue un percorso iterativo volto a colmare il divario tra l'uso linguistico e la formalizzazione concettuale. L'analisi si articola su due direttrici principali – la semasiologia e l'onomasiologia – che connettono l'osservazione dei testi *in vivo* alla costruzione della risorsa ontoterminologica *ITinHeritage* (Capitolo 4) attraverso tre livelli metodologici interdipendenti:

1. Il livello semasiologico (Uso): incentrato sull'osservazione dei termini all'interno dei testi. Attraverso l'impiego sinergico di strumenti di linguistica computazionale (IRaMuTeQ, Tropes, TermoStat e Sketch Engine), viene mappata la variazione terminologica in quattro ambiti discorsivi eterogenei: i discorsi storico-accademici, quelli tecnico-divulgativi, le dinamiche comunicative delle comunità online (come Reddit) e le pratiche istituzionali dei musei.
2. Il livello concettuale (Mediazione): in cui la variabilità, le sinonimie, le polisemie e i fenomeni di obsolescenza emersi dall'analisi testuale non vengono considerati "rumore" statistico, ma la base empirica per risalire ai concetti sottostanti. In questa fase si evidenzia come termini quali *personal computer, home computer, PC* o *laptop* non siano etichette intercambiabili, ma riflettano prospettive culturali e pratiche d'uso sensibilmente differenti che richiedono di essere raccordate.

3. Il livello onomasiologico e formale (Ontologia): lo strato finale in cui i pattern estratti in vivo in questo capitolo abbandonano la fluidità del discorso per essere stabilizzati e formalizzati nell'ambiente ontoterminologico (TEDI). Il concetto formale agisce così da hub meta-linguistico, neutralizzando l'ambiguità nei sistemi informativi ma preservando la ricchezza storiografica emersa dai testi.

L'analisi comparativa dei corpora si configura pertanto come il presupposto operativo diretto del progetto *ITinHeritage*: la terminologia funge da interfaccia critica fondamentale, estraendo la conoscenza dai testi per permettere di formalizzarla e ancorarla, infine, alla materialità degli oggetti patrimoniali.

In una prima fase semasiologica, l'analisi si concentra sull'osservazione dei termini "in vivo" all'interno dei testi. Attraverso l'impiego sinergico di strumenti di linguistica computazionale - quali IRaMuTeQ, Tropes, TermoStat e Sketch Engine - è possibile mappare la variazione terminologica in quattro ambiti discorsivi eterogenei: i discorsi storico-accademici, quelli tecnico-divulgativi, le dinamiche comunicative delle comunità online (come Reddit) e le pratiche istituzionali dei musei[23]. Questa prospettiva inclusiva permette di far emergere sinonimie, polisemie e fenomeni di obsolescenza che caratterizzano il dominio IT, evidenziando come termini quali *personal computer, home computer, PC* o *laptop* non siano semplici etichette intercambiabili, ma riflettano prospettive culturali e pratiche d'uso sensibilmente differenti.

La verifica di tale variabilità costituisce il presupposto indispensabile per la successiva transizione all'onomasiologia in chiave patrimoniale. Una volta identificata la fluidità delle denominazioni, i risultati dell'estrazione semasiologica offrono la base empirica per risalire ai concetti sottostanti e stabilizzarli all'interno di una rete concettuale formale. È proprio in questo passaggio che l'oggetto informatico viene riscattato dalla sua condizione di "reperto muto" o "scatola nera": attraverso una definizione onomasiologica rigorosa, il termine viene ancorato a caratteristiche essenziali che ne spiegano la funzione, la tecnologia e il valore storico.

La scelta dei corpora è stata guidata da una duplice esigenza: da un lato, coprire un ampio arco temporale; dall'altro, rappresentare la varietà dei registri comunicativi e delle comunità che partecipano alla costruzione e alla trasmissione della conoscenza tecnica.

I testi analizzati provengono dalla letteratura scientifica così come dalla divulgazione digitale e coinvolgono soggetti eterogenei, quali storici, storici dell'informatica, musei, collezionisti,

[23] La metodologia d'analisi e la costituzione dei corpora presentati in questo capitolo sono state consolidate durante due soggiorni di ricerca effettuati presso il laboratorio AGORA dell'Università CY Cergy Paris, sotto la supervisione del Prof. Julien Longhi. Questi periodi di mobilità (febbraio-marzo 2025 e settembre 2025) hanno permesso di approfondire l'uso del software IRaMuTeQ e di affinare l'approccio della linguistica computazionale e dell'analisi terminologica applicata al patrimonio informatico, grazie anche alla partecipazione a seminari specialistici e alla collaborazione con i ricercatori del laboratorio.

appassionati di retrocomputing e utenti generici delle piattaforme digitali. Tale pluralità consente di osservare fenomeni di stratificazione comunicativa (dal linguaggio tecnico-specialistico a quello destinato al grande pubblico) e di variabilità terminologica, che si manifestano tanto sul piano diacronico quanto in relazione al tipo di comunità e al mezzo di espressione.

In questa sezione l'analisi sarà circoscritta alla lingua inglese, che, pur rappresentando una lingua veicolare e ampiamente condivisa in ambito tecnico e scientifico, mostra già al proprio interno una notevole complessità terminologica. Tale complessità risulta amplificata quando si considerano le sfide legate al multilinguismo, come accade nel dominio museale, dove la rappresentazione concettuale dei dispositivi richiede un livello di astrazione superiore rispetto alla sola denominazione. Ad esempio, come viene definito oggi un laptop nei diversi contesti museali? La risposta può variare in funzione dell'epoca di riferimento, del tipo di istituzione, della funzione attribuita all'oggetto e della prospettiva curatoriale adottata. Come vedremo nel progetto ITinHeritage, la costruzione di una terminologia condivisa in ambito museale implica una riflessione profonda sulla natura concettuale degli oggetti e sulle modalità con cui essi vengono raccontati, classificati e resi accessibili a pubblici diversi.

3.1 Corpora e sottocorpora

La strutturazione del corpus complessivo e la successiva estrazione dei sottocorpora d'indagine hanno risposto a criteri metodologici stringenti, volti a garantire la rappresentatività sociolinguistica, la pertinenza concettuale e l'equilibrio diacronico dei dati. L'obiettivo non era la costituzione di una library cumulativa o indifferenziata, bensì la creazione di un campione testuale bilanciato capace di riflettere le asimmetrie e le dinamiche di risemantizzazione del dominio informatico.

A tal fine, la selezione è stata governata da tre macro-criteri di campionamento:

1. Criterio di diversificazione diafasica (I Registri): Per coprire l'intero spettro dell'ecosistema discorsivo del patrimonio IT, sono stati isolati tre registri comunicativi ben distinti. Il registro *tecnico-scientifico e storiografico* (letteratura accademica dalle library DBLP, ACM e IEEE) fornisce i termini nel loro massimo grado di formalizzazione e astrazione logica. Il registro *istituzionale-curatoriale* (schede descrittive del Science Museum di Londra e del HomeComputerMuseum) cattura la terminologia orientata all'oggetto, alla classificazione patrimoniale e alla divulgazione controllata. Infine, il registro *spontaneo-comunitario* (piattaforma Reddit, nello specifico r/retrobattlestations) intercetta la lingua *in vivo* degli utenti e degli appassionati, dominata da dinamiche affettive, slang, metafore e fenomeni di risemantizzazione nostalgica.
2. Criterio di ripartizione temporale (La Diacronia): La cronologia dei testi risponde alla necessità di mapparne l'evoluzione storica. Nel caso del corpus accademico, l'arco temporale esteso (1961–2024) è stato suddiviso in due macro-blocchi (1961-1984 e 1985-2024). Questa segmentazione non è arbitraria, ma risponde a una precisa contingenza storiografica: isolare la fase di emersione e prima concettualizzazione della tecnologia (in cui i confini semantici di termini come *computer* erano ancora fluidi) rispetto alla fase di stabilizzazione e contemporanea autoriflessione storica. Per i corpora museali e comunitari, la concentrazione sul periodo più recente (2010–2024) è legata al momento di massima proliferazione delle pratiche di patrimonizzazione digitale e *retrocomputing* sul web.
3. Criterio di focalizzazione e pertinenza terminologica (La selezione dei Sottocorpora): Data la vastità dei dati grezzi estratti dalle piattaforme e dalle library, si è reso necessario applicare un filtro di pertinenza per la creazione di sottocorpora mirati. Nel caso del corpus comunitario (Reddit), si è scelto di restringere l'estrazione alla stringa mirata *"home computer"* anziché al termine generico *"computer"*. Questa restrizione risponde a un vincolo scientifico: isolare una categoria terminologica ibrida, intrinsecamente socio-tecnica, che si colloca esattamente all'intersezione tra l'evoluzione industriale e la memoria culturale di massa, evitando al contempo che il software di computazione linguistica (IRaMuTeQ) venisse saturato da

milioni di occorrenze banali o fuori fuoco (come i riferimenti ai moderni computer da ufficio o server aziendali).

3.1.1 Corpus accademici e storici

Il primo corpus preso in considerazione è stato costruito a partire dalla digital library DBLP**,** (https://dblp.org), con l'obiettivo di raccogliere testi in lingua inglese dedicati alla storia dell'informatica. La selezione è stata articolata in due sottoinsiemi temporali, ciascuno corrispondente a un diverso periodo di pubblicazione, così da consentire l'osservazione delle trasformazioni terminologiche e concettuali del campo.

In una prima fase sono stati inclusi i conference papers pubblicati tra il **1961** e il **1984** dall'ACM (*American Federation of Information Processing Societies*) Journal on Computing and Cultural Heritage, Volume 17 (2024). Come dichiarato nella propria missione istituzionale, l'organizzazione mirava a «advance knowledge in the field of information science». Questo sottoinsieme (**Corpus 1**) raccoglie contributi di autori rilevanti per la storia dell'informatica, tra cui Ceruzzi, Kuhn e altri, con una media di circa 15 articoli per anno.

Successivamente, è stato costruito un secondo sottoinsieme (**Corpus 2**), costituito da articoli pubblicati tra il **1985** e il **2024** negli *IEEE Annals of the History of Computing*, rivista specializzata in history of computing, information technology e computer hardware. In questo caso la media è di circa 10 articoli per anno, ma su un intervallo temporale più esteso.

Complessivamente, i due corpora contano oltre 700 documenti. Per ciascun anno incluso negli intervalli temporali sono stati scaricati i PDF disponibili su DBLP; uno script Python ha poi automatizzato la conversione dei file in formato .txt, organizzandoli in cartelle ordinate per anno. Ogni file testuale segue una convenzione coerente di denominazione, utile a garantire l'allineamento delle successive analisi:

- riga 1: titolo dell'articolo
- riga 2: autore/i
- riga 3: anno di pubblicazione
- righe successive: contenuto testuale dell'articolo

Questa formattazione uniforme ha reso possibile non solo la corretta suddivisione del corpus in sottocartelle annuali, ma anche l'indicizzazione automatica dei metadati associati a ciascun documento. Nelle fasi successive, il corpus è stato elaborato attraverso diversi strumenti informatici, tra cui **IraMuTeQ,** *Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires* (Ratinaud, 2014), software specificamente progettato per l'analisi statistica multidimensionale di testi e descritto più in dettaglio nel paragrafo seguente dedicato agli applicativi. Questa struttura ha permesso di tracciare in modo preciso la distribuzione temporale dei termini chiave, facilitando l'analisi comparativa della variazione terminologica nel tempo.

Di seguito vengono presentati alcuni esempi significativi:

A COMPUTER FOR DIRECT EXECUTION OF ALGORITHMIC LANGUAGES
James P. Anderson
1961
Burroughs Corporation Paoli, Pennsylvania INTRODUCTION The machine outlined in this paper repre sents another of a recent series of effort
cated in the title, has long been recognized by some [1] but ALGOL '60 was chosen as the basis for the exploitation of the relationship ha
as well as those associated with the definition of subroutines (proce One area yet to be exploited by machine dures). The imperative stat
e procedure or function for obtaining a data loops, and to determine the scope of other value may be used in any expression where control
In addition to abstracting more conven variable is defined in the structure of the tional coding structures, ALGOL '60 is program statemen

A GENERAL PURPOSE SYSTEMS SIMULATION PROGRAM
Geoffrey Gordon
1961
International Business Machines Advanced Systems Development Division White Plains, N. Y. 1. INTRODUCTION difficult to foresee all eventua
t is possible the studies that are being conducted have that the process of producing a simulation presented several difficulties in organ
ulties are created by the fact that there it forms a natural choice on which to base are often many design variations to be stud the input
m that has been 87 88 / Computers - Key to Total Systems Control developed from this early work and applied 2. PRINCIPLES OF THE PROGRAM s
detailed in this paper. No knowl block may have several lines entering it to edge of the computer operation is assumed. represent the fact

A SIMULATION MODEL FOR DATA SYSTEM ANALYSIS
Leon Gainen
1961
The Rand Corporation Dayton, Ohio SUMMARY The author asserts that designing a data system to support manage ment objectives is, at present
duction worthy of serious study because of the great potential inherent in present and promised The main object of this paper is to promot
ment in any proposed manage and indeed, to bring our profession closer to ment information system. a science. The technique of system simu
municated simultaneously to the processing titative way assure the "best" system has and storage function. For convenience, the been achie
handi Generally speaking, the major constraints work. are limitation in the number of dollars and I contend that data systems can be model
problem. This means that hardware, then it would appear that simulation each system function is provided just the of system operation is r
system giving due consideration to the effect of these factors II. The Data System on the operation of the system. IV. Structure and Opera
per ric shape. It is not necessary for each of formance and its contribution to system costs. these functions explicitly to occur in the T

Figure 12 Estratti del Corpus 1 (1961), convertito e strutturato per l'analisi testuale.

Considerata la dimensione consistente dei due corpora, l'elaborazione ha richiesto un notevole impiego di risorse computazionali, nonostante una preliminare fase di pulizia testuale (comprendente la rimozione di punteggiatura e stopwords, operazione che peraltro IRaMuTeQ esegue comunque automaticamente).

3.1.2 Sottocorpora tematici

Per ottimizzare l'analisi, è stata impiegata una funzione specifica di IRaMuTeQ che consente la creazione di sottocorpora tematici basati su parole chiave. In particolare, è stato selezionato il termine *computer*, che costituisce il nodo centrale dell'indagine: si colloca, come emergerà dall'analisi, tra i lemmi con frequenza più alta in entrambi i corpora e rappresenta un concetto-chiave per l'analisi dell'evoluzione semantica del lessico informatico.

La scelta di costruire sottocorpora dedicati risponde anche a esigenze di ordine tecnico: la dimensione complessiva dei due corpora, nonostante le operazioni preliminari di pulizia testuale, risultava troppo estesa per permettere un'elaborazione fluida. In diversi casi, infatti, gli applicativi utilizzati non supportavano la grandezza dei file o richiedevano tempi di calcolo eccessivi. La creazione di sottocorpora tematici ha quindi permesso di circoscrivere l'analisi, garantendo una gestione più efficiente dei dati e al tempo stesso una maggiore focalizzazione sugli aspetti lessicali di interesse.

Sono stati pertanto costruiti due sottocorpora focalizzati sul termine *computer*:

- Sottocorpus Computer1: relativo al periodo 1961–1984 (derivato dal Corpus 1);
- Sottocorpus Computer2: relativo al periodo 1985–2024 (derivato dal Corpus 2).

L'analisi ha approfondito le co-occorrenze, i contesti d'uso e le trasformazioni semantiche del termine lungo l'arco temporale considerato.

3.1.3 Corpus della comunità di Reddit: il caso degli home computer

La variazione terminologica non dipende soltanto dall'evoluzione storica, ma anche dal pubblico di riferimento: i testi rivolti a esperti, appassionati o utenti generici mostrano strategie e scelte lessicali sensibilmente diverse. Mentre i sottocorpora accademici e storici hanno consentito di osservare l'evoluzione del termine *computer* nei testi specialistici lungo un arco temporale esteso, l'inclusione di un corpus non standardizzato, costituito da testi di utenti non professionisti, amplia la prospettiva in una dimensione socio-discorsiva, indagando come le trasformazioni terminologiche si manifestino in ambienti digitali contemporanei, caratterizzati da pratiche comunicative meno formali e da un forte legame con la dimensione patrimoniale e culturale degli oggetti tecnologici.

Si è scelto di arricchire il progetto focalizzando l'attenzione sul concetto di *home computer*. La ricerca del termine *computer* su Reddit avrebbe restituito un numero eccessivo di occorrenze, spesso generiche e poco significative dal punto di vista terminologico. Per questo motivo si è scelto di restringere l'analisi al concetto di *home computer*, che rappresenta la forma di calcolatore più direttamente associata all'esperienza quotidiana e all'immaginario del grande pubblico. In questo modo, il corpus Reddit si distingue dai corpora accademici non solo per il registro discorsivo informale, ma anche per la prospettiva d'uso: centrata sull'ambiente domestico e sulla dimensione culturale e affettiva legata agli oggetti tecnologici.

Gli *home computer* rappresentano un oggetto di studio al contempo tecnico e culturale: da un lato, rimandano a componenti hardware, software e usi pratici; dall'altro, si collocano in una dimensione sociale e comunicativa che riguarda l'informatica domestica, l'educazione e il tempo libero. Anche il lessico che li descrive ha subito trasformazioni significative nel tempo: espressioni come *personal computer*, *home computing*, *desktop* e *PC* riflettono non solo l'evoluzione tecnologica, ma anche i mutamenti nei modi di nominare e rappresentare questi dispositivi.

Si potranno individuare attraverso i vari software e gli esempi almeno tre fasi principali di trasformazione:

- negli anni '80, gli home computer in senso stretto coincidono con i dispositivi programmabili a basso costo e destinati all'uso domestico (Commodore, ZX Spectrum, Amiga, ecc.);
- tra anni '90 e 2000, il termine si estende per sovrapposizione semantica al personal computer con sistema Windows (desktop e successivamente notebook), utilizzato in casa ma ormai assimilabile per funzioni a quello professionale;
- dagli anni 2010 in poi, si assiste a una transizione dell'home computing verso dispositivi mobili o ibridi (laptop, tablet, smartphone), con una progressiva dissoluzione dell'etichetta

originaria, ma con persistenze simboliche nel lessico (es. PC di casa, computer domestico, retrocomputer).

Proprio per questa natura ibrida, l'oggetto si presta particolarmente bene a un'analisi della variazione linguistica, osservando come il linguaggio cambi in relazione ai contesti d'uso (ambiti tecnico-specialistici o divulgativi) e ai destinatari cui i testi si rivolgono, siano essi esperti del settore, appassionati o utenti generici.

Per osservare come questa dimensione si rifletta nei discorsi degli utenti, è stato costruito un corpus automatizzato di post provenienti da Reddit, tramite uno script Python. L'estrazione è focalizzata sul subreddit *r/retrobattlestations*. I subreddit su Reddit sono sottosezioni tematiche della piattaforma, ciascuna dedicata a un argomento specifico. In particolare, il subreddit "retrobattlestations", una comunità tematica dedicata ai *retro computers*, ovvero postazioni informatiche vintage, e agli *home computing setups*, composti da computer storici (anni '70, '80, '90), spesso ancora funzionanti, documentati da appassionati con foto, descrizioni e discussioni. Per una ricerca più pertinente, sono stati associati una serie di termini di ricerca correlati a ["home computer", "personal computer", "pc", "retro computer", "vintage computing", "old computers"].

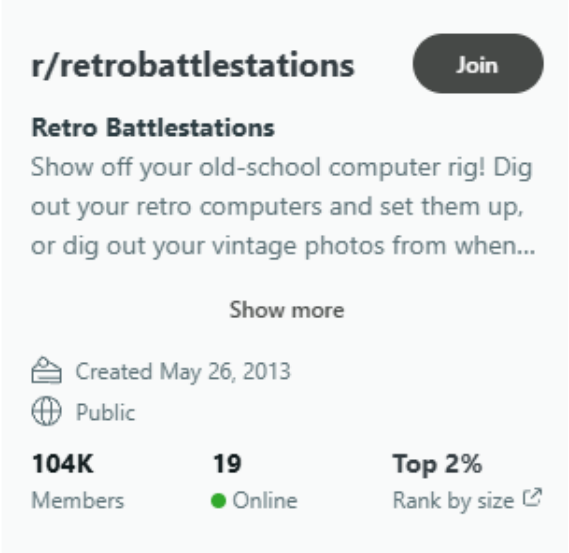


Figure 13 Home page del subreddit r/retrobattlestations, dedicato alla condivisione e valorizzazione di retro-computer.

Il pezzo di codice inizia stabilendo un ponte con Reddit tramite PRAW (Python Reddit API Wrapper), la libreria Python dedicata all'API di Reddit: sfrutta le credenziali (*client_id, client_secret* e *user_agent*) per autenticarsi e prepararsi a navigare i contenuti del sito.

I parametri di ricerca sono i seguenti:

- Intervallo temporale: 2010–2024 (Reddit nasce nel 2005);
- Post per anno: 5 per anno (*posts_per_year = 5*);
- Commenti per post: fino a 5 commenti (*comments_per_post = 5*);

- Subreddit di riferimento: r/retrobattlestations;
- Termini di ricerca: *"home computer"*, *"personal computer"*, *"pc"*, *"retro computer"*, *"vintage computing"*, *"old computers"*.

Il funzionamento del codice può essere riassunto nei seguenti passaggi:

- Per ciascun anno, il codice calcola i timestamp di inizio e fine;
- Per ciascun termine di ricerca, si effettua una query nel subreddit selezionato, ordinata per post più votati (*sort="top"*) e limitata a 100 risultati per query;
- I post vengono filtrati per anno e analizzati solo se pubblicati nell'intervallo temporale specificato;
- Se il post soddisfa questi criteri vengono estratti: titolo, testo (se presente) e commenti (massimo 5);
- queste informazioni vengono salvate in un'unica stringa, formattata per l'output e compatibile con IRaMuTeQ.

I dati così raccolti sono stati organizzati per anno e successivamente ripuliti da elementi non rilevanti (es. caratteri di formattazione superflui come gli asterischi). È stato inoltre sperimentato un codice alternativo, che prevedeva la ricerca su reddit.subreddit("all"); tuttavia, i risultati si sono dimostrati più pertinenti e mirati quando circoscritti al subreddit r/retrobattlestations, come confermato anche dalle elaborazioni effettuate con IRaMuTeQ e TermoStat.

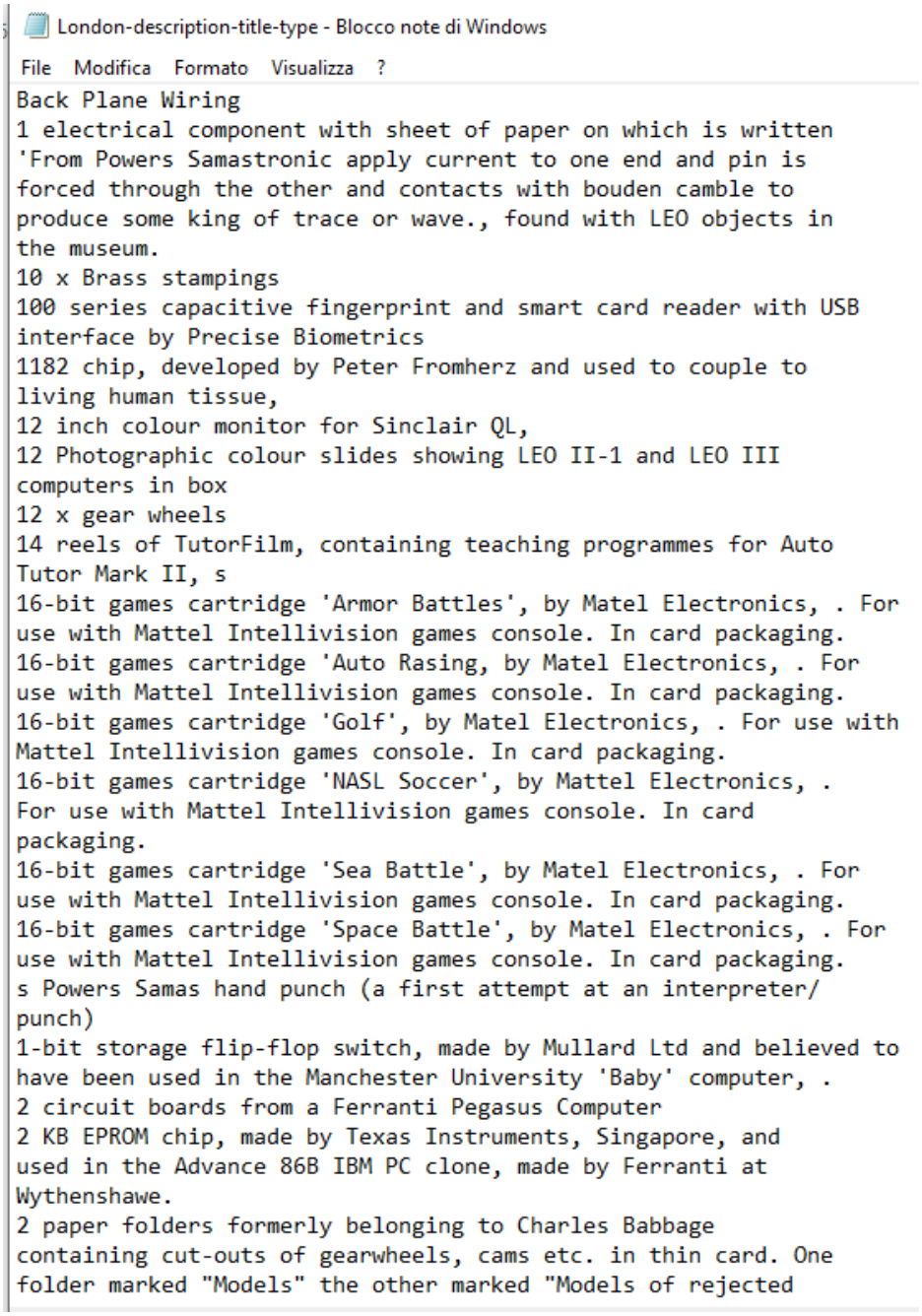
3.1.4 Corpus patrimoniale

Il corpus museale (o patrimoniale) raccoglie testi prodotti da istituzioni culturali con funzione espositiva e divulgativa. Questi testi rappresentano una forma di mediazione istituzionale della storia dell'informatica, in particolare attraverso il racconto museale degli oggetti del *computing*.

Per approfondire la variazione terminologica in questo dominio, si è scelto di includere dati museali in lingua inglese utilizzati per l'analisi nel progetto *ITinHeritage* presentato nel capitolo successivo, in particolare quelli relativi al **Science Museum** di Londra e all'**HomeComputerMuseum** dei Paesi Bassi. Le schede museali di cui si dispone sono strutturate in formato tabellare, con metadati ben definiti, che variano leggermente tra le due istituzioni.

I campi disponibili sono diversi, tuttavia, ai fini dell'analisi linguistica sono stati selezionati soltanto quelli più significativi dal punto di vista lessicale. In particolare:

- per il Science Museum: *title, type* e *description*


London-description-title-type - Blocco note di Windows

File Modifica Formato Visualizza ?

```
Back Plane Wiring
1 electrical component with sheet of paper on which is written
'From Powers Samastronic apply current to one end and pin is
forced through the other and contacts with bouden camble to
produce some king of trace or wave., found with LEO objects in
the museum.
10 x Brass stampings
100 series capacitive fingerprint and smart card reader with USB
interface by Precise Biometrics
1182 chip, developed by Peter Fromherz and used to couple to
living human tissue,
12 inch colour monitor for Sinclair QL,
12 Photographic colour slides showing LEO II-1 and LEO III
computers in box
12 x gear wheels
14 reels of TutorFilm, containing teaching programmes for Auto
Tutor Mark II, s
16-bit games cartridge 'Armor Battles', by Matel Electronics, . For
use with Mattel Intellivision games console. In card packaging.
16-bit games cartridge 'Auto Rasing, by Matel Electronics, . For
use with Mattel Intellivision games console. In card packaging.
16-bit games cartridge 'Golf', by Matel Electronics, . For use with
Mattel Intellivision games console. In card packaging.
16-bit games cartridge 'NASL Soccer', by Mattel Electronics, .
For use with Mattel Intellivision games console. In card
packaging.
16-bit games cartridge 'Sea Battle', by Matel Electronics, . For
use with Mattel Intellivision games console. In card packaging.
16-bit games cartridge 'Space Battle', by Matel Electronics, . For
use with Mattel Intellivision games console. In card packaging.
s Powers Samas hand punch (a first attempt at an interpreter/
punch)
1-bit storage flip-flop switch, made by Mullard Ltd and believed to
have been used in the Manchester University 'Baby' computer, .
2 circuit boards from a Ferranti Pegasus Computer
2 KB EPROM chip, made by Texas Instruments, Singapore, and
used in the Advance 86B IBM PC clone, made by Ferranti at
Wythenshawe.
2 paper folders formerly belonging to Charles Babbage
containing cut-outs of gearwheels, cams etc. in thin card. One
folder marked "Models" the other marked "Models of rejected
```

Figure 14 Estratto del corpus Science Museum

- per l'HomeComputerMuseum: *title, creator* e *type*


HomeComputerMuseDataE2 - Blocco note di Windows

File Modifica Formato Visualizza ?

```
Sega Model2 Rally PCB   Sega    Component
Nintendo DSi XL Nintendo        Handheld Game Console
Nintendo 3DS    Nintendo        Handheld Game Console
Nintendo DS Lite        Nintendo        Handheld Game Console
Nintendo DS Lite        Nintendo        Handheld Game Console
Nintendo DS Lite        Nintendo        Handheld Game Console
Nintendo DS Lite        Nintendo        Handheld Game Console
Dendy 8 Bit Super Entertainment System  Dendy   Game Console
Philips CDI 450 Philips Multimedia computer
Sol-20 Terminal Computer        Processor Technology    Micro-Computer
IBM PCjr 4860   IBM     Desktop computer
Vtech Little Talking Scholar    vTech   Micro-Computer
IBM A21p (Type 2629)    IBM     Portable computer
IBM ThinkPad T42p (Type 2372)   IBM     Portable computer
IBM ThinkPad T42p (Type 2372)   IBM     Portable computer
Lenovo W701 (2541-27G)  Lenovo  Portable computer
Amiga 500       Commodore       Desktop computer
Philips 14GR1220/10B - Discoverer       Philips Televisie
Acorn BBC Model B       Acorn   Desktop computer
IBM Model M4-1 keyboard IBM     Input device
Toshiba T1000   Toshiba Portable computer
MSi U123 MS-N033        MSI     Portable computer
Apple iMac G3 (Blueberry)       Apple   Desktop computer
Apple iMac G4   Apple   Desktop computer
Apple iMac G3 (Indigo)  Apple   Desktop computer
Nintendo Game & Watch Donkey Kong II    Nintendo        Handheld Game Console
Apple iMac (Mid 2007)   Apple   Desktop computer
Toshiba 620CT   Toshiba Portable computer
Dell Latitude X1        Dell    Portable computer
Toshiba Equium 2000     Toshiba Desktop computer
Chessman Schaakcomputer Onbekend        Game Console
Minitel Server  Cooler Master   Desktop computer
SyQuest EZ 135 Drive    Syquest Media
Philips SBC 1745 Rekenmachine   Philips Desktop computer, Calculator
Merlin m4 Digitaal Kookboek     Proval  Portable computer
Merlin m4 Digitaal Kookboek     Proval  Portable computer
Anbonn AM-09 Monochrome Monitor Anbonn  Output device
Anbonn AM-09 Monochrome Monitor Anbonn  Output device
Visicom 2 (Multicare Systems)   Goedhart        Micro-Computer
Visicom 2 (Multicom)    Goedhart        Micro-Computer
Syquest Mirror 44 MB drive      Syquest Input device
```

Figure 15 Estratto del corpus HomeComputerMuseum

I testi di partenza non sono lunghi: nella maggior parte dei casi si tratta di termini singoli o brevi frasi, con l'eccezione del campo *description* del museo londinese, che presenta descrizioni più articolate. Per garantire la compatibilità con strumenti di analisi lessicometrica (IRaMuTeQ, TermoStat, Tropes), tutti i contenuti sono stati aggregati in un unico file testuale, in modo da ottenere un formato adatto a segmentazione e trattamento linguistico.

3.2 Strumenti e metodi applicati su ogni corpus

Per affrontare la complessità e la stratificazione dei dati linguistici emersi dai corpora selezionati, si è scelto di ricorrere a una serie di strumenti informatici orientati all'esplorazione lessicale e semantica dei testi, con particolare attenzione alle forme terminologiche ricorrenti, alle variazioni contestuali e all'organizzazione concettuale implicita nelle diverse comunità discorsive. Gli strumenti selezionati, ovvero IRaMuTeQ, Tropes e TermoStat, rispondono a esigenze analitiche differenziate.

In primo luogo, IRaMuTeQ (Ratinaud, 2014), consente di descrivere la distribuzione delle forme lessicali, di individuare nuclei di co-occorrenza e di evidenziare, attraverso l'analisi delle similitudini, le configurazioni concettuali che emergono nei vari corpora. Questo livello permette di identificare lo "scheletro" del discorso, ossia le strutture ricorrenti e le associazioni terminologiche che caratterizzano ciascun contesto d'uso.

In secondo luogo, Tropes[24] (Molette, Landre, 2021) permette di ricostruire le reti semantiche e le categorie concettuali sottese ai testi, rendendo visibili le relazioni tra temi, attori, pratiche e ruoli discorsivi. Tale analisi è cruciale per cogliere come gli oggetti informatici – in particolare il *computer* – vengano narrati, interpretati e semantizzati in modi differenti in relazione ai pubblici di riferimento e ai generi comunicativi.

Infine, TermoStat[25] (Drouin, 2003), consente di isolare le unità terminologiche rilevanti e di osservare le oscillazioni denominative nei diversi corpora, distinguendo tra termini stabilizzati e forme variabili, nonché tra usi specialistici, divulgativi e comunitari.

Ciascuno di questi applicativi è stato impiegato in funzione della struttura e delle caratteristiche linguistiche del corpus su cui è stato applicato.

L'analisi si sviluppa attraverso una progressione metodologica coerente: si parte dall'identificazione delle occorrenze e delle co-occorrenze lessicali, per passare alla ricostruzione dei cluster semantici e delle reti concettuali, fino all'individuazione delle scelte terminologiche che segnalano trasformazioni di significato.

L'integrazione dei tre livelli (distribuzionale, semantico e terminologico) consente non solo di descrivere la distribuzione delle forme, ma di interpretare la variazione terminologica come esito di pratiche culturali, tecniche e sociali situate nel tempo. In questo modo, l'analisi mette in evidenza pattern ricorrenti e divergenze tra i diversi contesti discorsivi, offrendo una base empirica solida per comprendere le dinamiche di costruzione, negoziazione e trasmissione della conoscenza informatica.

[24] http://www.tropes.fr/
[25] https://termostat.ling.umontreal.ca/

3.2.1 IraMuTeQ

Per l'analisi lessicometrica dei corpora è stato utilizzato IRaMuTeQ[26], un software open source sviluppato da Pierre Ratinaud presso il Laboratoire LERASS dell'Università di Toulouse-Le Mirail. IRaMuTeQ consente l'analisi quantitativa e multidimensionale di corpora testuali attraverso tecniche statistiche e classificatorie basate su R, con estensioni in Python per la gestione di input complessi. Il software è progettato per elaborare grandi volumi di dati testuali, offrendo funzionalità specifiche per:

- l'analisi delle frequenze lessicali,
- la classificazione automatica (analisi delle corrispondenze, clustering tematico),
- l'analisi di co-occorrenze e forme specifiche,
- la costruzione di nuvole semantiche e dendrogrammi lessicali.

I file testuali (.txt) sono stati preparati e codificati in formato UTF-8, seguendo la sintassi richiesta dal software. Ogni documento è preceduto da una riga di metadati strutturati, come nel seguente esempio:
***** *source_AFIPS_Joint_Computer_Conferences *year_1961*

Questa intestazione consente a IRaMuTeQ di segmentare correttamente i testi e di associare a ciascun segmento le relative variabili categoriali, facilitando l'analisi comparativa tra sottogruppi del corpus (es. per anno, fonte, tipologia testuale).

[26] Iramuteq: www.iramuteq.org

3.2.1.1 Statistiche descrittive di base e forme attive

Una prima fase ha previsto l'estrazione di statistiche di base relative al numero di testi, parole e occorrenze.

Corpus 1:

```
Nombre de textes : 8434
Nombre d'occurrences : 36850539
Nombre de formes : 231721
Nombre d'hapax : 440 (0.00% des occurrences - 0.19% des formes)
Moyenne d'occurrences par texte : 4369.28
```

Figure 16 Statistiche generali del corpus 1 generate da IRaMuTeQ: numero di testi, occorrenze, forme, hapax e media di occorrenze per testo.

Nel *Corpus 1* si osserva un totale di 8.434 testi per oltre 36 milioni di occorrenze, distribuite su 231.721 forme uniche. Il numero di *hapax*, ovvero le forme che compaiono una sola volta, è molto contenuto (440, pari allo 0,19% delle forme), a fronte di una media di circa 4.369 parole per testo. Questo dato suggerisce un lessico relativamente standardizzato, caratterizzato da alta ripetitività e bassa dispersione, compatibile con la natura tecnico-scientifica e normativa del linguaggio utilizzato nei testi conferenziali delle prime decadi.

Corpus 2:

```
Nombre de textes : 1366
Nombre d'occurrences : 8349147
Nombre de formes : 186019
Nombre d'hapax : 114137 (1.37% des occurrences - 61.36% des formes)
Moyenne d'occurrences par texte : 6112.11
```

Figure 17 Statistiche generali del corpus 2 generate da IRaMuTeQ: numero di testi, occorrenze, forme, hapax e media di occorrenze per testo.

Il *Corpus 2*, pur presentando un numero minore di testi (1.366), mostra invece una media di parole per testo più elevata (6.112) e un vocabolario significativamente più disperso: su 186.019 forme uniche, ben 114.137 sono *hapax*, pari al 61,36% del totale. Tali valori indicano una maggiore densità lessicale e una marcata variabilità terminologica, probabilmente dovute alla maggiore eterogeneità tematica e stilistica della scrittura, al registro più riflessivo o divulgativo e alla progressiva specializzazione dei contenuti.

Nel confronto, il *Corpus 1* evidenzia quindi un uso linguistico più uniforme e formalizzato, mentre il *Corpus 2* rivela un'evoluzione verso una terminologia più ricca, frammentata e articolata.

Nel confronto tra i due corpora, emerge quindi un passaggio significativo: dal linguaggio specialistico e normalizzato degli albori dell'informatica a una lingua della riflessione, più aperta, variabile e concettualmente stratificata. Queste osservazioni statistiche trovano riscontro nella trasformazione più ampia che ha interessato il discorso sull'informatica: Se negli anni '50–'70 la comunicazione accademica era dominata da un linguaggio formalizzato e orientato alla risoluzione di problemi computazionali, dagli anni '80 in poi si è assistito all'emergere di una prospettiva più interdisciplinare, storica e teorica. Come osserva Mahoney, «the history of computing has yet to establish a significant presence in the history of science and technology» (Mahoney, 1988), evidenziando come solo a partire da quel periodo la storia dell'informatica abbia iniziato a trovare spazio nel più ampio panorama storiografico.

Sottocorpus Computer1:

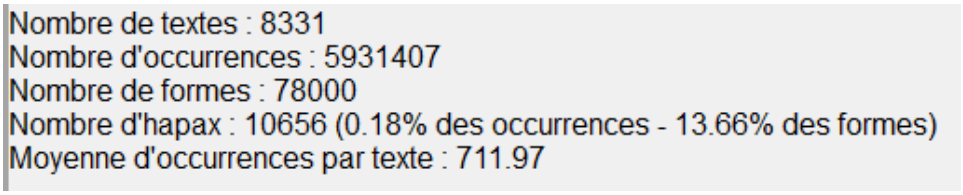


Figure 18 Statistiche generali del sottocorpus computer 1 generate da IRaMuTeQ: numero di testi, occorrenze, forme, hapax e media di occorrenze per testo.

Composto da 8.331 testi, il sottocorpus presenta 5.931.407 occorrenze, distribuite su circa 78.000 forme uniche. Gli hapax sono 10.656, pari allo 0,18% delle forme, con una media di 711,97 parole per testo.

Questi valori confermano la tendenza a un uso ripetitivo e standardizzato della terminologia.

Sottocorpus Computer2:

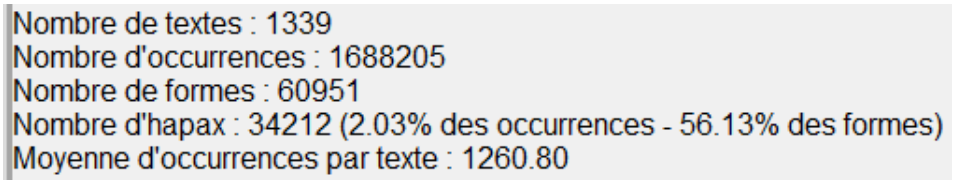


Figure 19 Statistiche generali del sottocorpus computer 2 generate da IRaMuTeQ: numero di testi, occorrenze, forme, hapax e media di occorrenze per testo.

Il Sottocorpus Computer2 comprende 1.339 testi, per un totale di 1.688.205 occorrenze, distribuite su 60.951 forme uniche. Gli hapax risultano essere 34.212, pari al 2,03% delle forme, con una media di 1.261 parole per testo.

Anche in questa suddivisione si conferma il trend già evidenziato nella comparazione tra i corpora principali: una progressiva evoluzione verso una maggiore varietà terminologica e una più ricca articolazione semantica, che riflette il cambiamento di statuto del computer nella cultura tecnico-scientifica e nella riflessione interdisciplinare contemporanea.

Corpus Reddit:

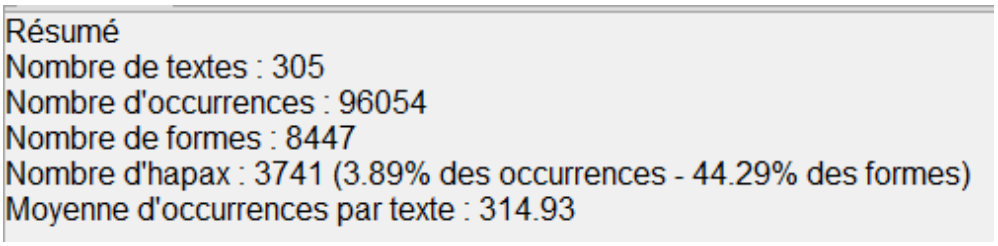


Figure 20 Statistiche generali del corpus Reddit generate da IRaMuTeQ: numero di testi, occorrenze, forme, hapax e media di occorrenze per testo.

Il corpus Reddit è molto frazionato (305 testi di lunghezza relativamente breve). La percentuale molto alta di hapax mostra una forte variabilità lessicale, tipica di discorsi spontanei, dove entrano in gioco neologismi, slang, espressioni soggettive e riferimenti situati.

Corpus Musei:

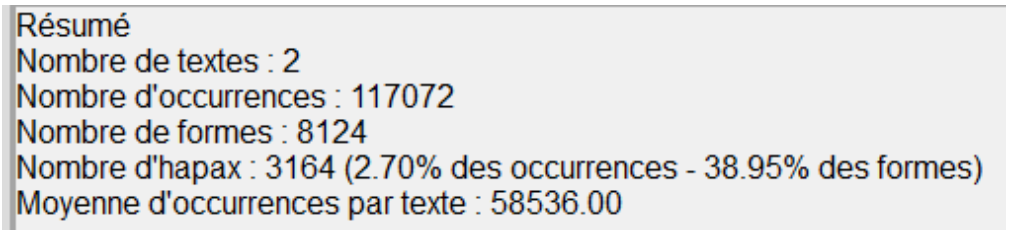


Figure 21 Statistiche generali del corpus museale generate da IRaMuTeQ: numero di testi, occorrenze, forme, hapax e media di occorrenze per testo.

Il corpus museale (2 testi, 117.072 occorrenze, 8.124 forme, 38,95% hapax) presenta invece una struttura molto diversa: i due "testi" corrispondono in realtà a repertori terminologici e descrittivi, organizzati in maniera standardizzata e coerente con pratiche di documentazione istituzionale. Questo spiega la frequenza relativamente bassa di hapax: il vocabolario impiegato è controllato e ripetitivo, costruito per garantire stabilità, precisione e uniformità nella denominazione degli oggetti.

Forme attive e categorie grammaticali

Un'ulteriore fase dell'analisi ha riguardato l'esame delle forme attive, ovvero delle unità lessicali con frequenza almeno pari a uno, classificate secondo le principali categorie grammaticali. I dati evidenziano, in entrambi i corpora, una netta prevalenza dei sostantivi, a indicare la centralità della nominazione e della classificazione concettuale nei testi analizzati. Tale tendenza risulta coerente con la natura dei contenuti conferenziali, spesso caratterizzati da un'esposizione tecnica e concettuale che privilegia l'identificazione precisa di oggetti, funzioni e processi.

Nel Corpus 1 le forme attive sono le seguenti:

Forme	Freq.	Types
system	291441	nom
data	195500	nom
computer	177324	nom
program	167231	nom
time	136355	nom
figure	100508	nom
process	100330	nom
control	95459	nom
user	89477	nom
memory	75571	nom
design	73635	nom
numb	73350	adj
set	70998	ver
information	69627	nom
function	69475	nom
problem	66817	nom
require	62963	ver
file	59352	nom
operation	56822	nom
language	54535	nom
line	53196	nom
input	52365	nom
type	51216	nom
result	50854	nom
show	49894	nom
test	48586	nom
word	48174	nom
output	47782	nom
model	47332	nom
point	46244	nom
structure	45927	nom
level	45423	adj
provide	45041	ver
processor	44518	nom
form	42803	nom
base	42626	adj
cost	41268	ver
storage	40647	nom
network	40326	nom

conference	40150	nom
machine	39841	nom
large	39827	adj
table	39795	nom
case	39649	nom
record	39321	nom
application	39314	nom
software	38799	nom
order	37788	nom
code	37786	nom
work	37206	nom
access	37048	nom
tion	36689	nr
call	35868	nom
address	35062	nom
state	34976	nom
error	34964	nom
unit	34823	nom
high	34062	adj
display	33999	nom
store	33826	nom
device	33265	nom
instruction	33075	nom
include	32604	ver
part	32018	nom
technique	31902	nom
area	31839	nom
method	31761	nom
block	31676	nom
analysis	31503	nom
terminal	31308	nom
write	31220	ver
operate	30912	ver
variable	30886	nom
element	30765	nom
define	30763	ver
procedure	30691	nom
list	30034	nom
register	29913	nom

Figure 22 Formes Actives corpus 1

In cima alla lista si trovano termini come *system*, *data*, *computer*, *program*, *process*, *user*, *memory*, *information*, tutti concetti centrali nel discorso informatico e computazionale. Molti termini, infatti, fanno riferimento a concetti chiave del software (*program, code, application, instruction*), dell'hardware (*computer, machine, device, processor*), e dei processi computazionali *(input, output, process, operation*).

Questa prevalenza nominale è coerente con il registro tecnico-scientifico del genere testuale, ampiamente documentata in letteratura, in cui la nominalizzazione rappresenta una strategia ricorrente per astrarre, classificare e condensare concetti complessi. L'Homme (2004) osserva infatti

che, nei dizionari terminologici tecnico-scientifici, le unità linguistiche registrate sono quasi esclusivamente sostantivi, poiché il nome designa entità concrete e astratte all'interno di un sistema concettuale; al contrario, termini verbali, aggettivali o avverbiali trovano raramente spazio accanto a quella che definisce la "parte del discorso regina" (L'Homme, 2004, p. 50). A titolo di esempio, in un dizionario dedicato alla microinformatica si privilegiano termini che denotano oggetti concreti, come tipi di stampanti, microcomputer, periferiche e le loro componenti. Si descrive, per esempio, come una "stampante laser" faccia parte della categoria "periferiche" e sia composta da elementi quali "cartuccia" o "sistema di alimentazione della carta". Il termine viene analizzato in una prospettiva onomasiologica, cioè a partire dal concetto che rappresenta, piuttosto che dal contesto linguistico in cui è impiegato.

In modo coerente, Sager (1990) sottolinea che i concetti terminologici sono prevalentemente espressi tramite sostantivi; i concetti che in lingue tecniche si manifestano come aggettivi o verbi appaiono spesso esclusivamente nella forma nominale corrispondente, e alcuni studiosi arrivano persino a negare l'esistenza di concetti espressi da aggettivi o verbi.

Accanto a queste entità, nei testi scientifici compaiono anche termini legati a modalità discorsive e strutture del testo scientifico (*figure, table, result, show, define, method*), a testimonianza della dimensione comunicativa della scrittura scientifica.

I **verbi** (*require*, *provide*, *show*, *include*, *write*, *operate*) ricorrono meno frequentemente e sono legati a azioni operative o a processi di sistema, spesso in forma impersonale o passiva.

Gli **aggettivi** (*large*, *high*, *base*) compaiono raramente nella lista delle 100 forme più frequenti, confermando una tendenza a minimizzare la valutazione soggettiva o descrittiva, a favore della classificazione oggettiva.

Alcune forme lasciano intuire ambiguità o errori di lemmatizzazione, per esempio:

- *numb* è marcata come adj, ma appare probabilmente come forma troncata o mal rilevata di *number*.
- *tion*, classificata come nr, sembra derivare da un'errata segmentazione (forse residuo di una tokenizzazione fallita o taglio da suffisso).

Questi indizi suggeriscono che il processo di elaborazione linguistica automatica (lemmatizzazione, POS-tagging) può necessitare di una verifica qualitativa, specialmente per garantire l'accuratezza nell'analisi semantica e terminologica.

Corpus 2:

Forme	Freq.	Types
computer	51438	nom
system	29534	nom
compute	24771	nom
history	23523	nom
program	21018	nom
work	21005	nom
machine	20111	nom
ieee	19122	nr
time	16938	nom
ibm	15750	nr
software	13938	nom
pp	13284	nr
design	12608	nom
technology	12530	nom
data	12517	nom
year	12488	nom
early	12418	adj
research	12251	nom
company	11739	nom
process	11403	nom
development	10756	nom
university	10732	nom
science	10620	nom
apply	10613	ver
vol	10374	nr
information	10295	nom
engineer	9905	nom
limit	9891	nom
project	9724	nom
annals	9677	nr
develop	9094	ver
license	8601	nom
february	8391	nr
di	8298	nr
problem	8135	nom
restriction	8131	nom
ing	8074	nr
authorize	7969	ver
utc	7888	nr

xplore	7881	nr
studi	7856	nr
degli	7855	nr
universita	7851	nr
genova	7811	nr
include	7780	ver
download	7588	ver
language	7567	nom
industry	7397	nom
control	7376	nom
group	7238	nom
build	7204	nom
user	7152	nom
write	7131	ver
numb	6986	adj
person	6807	nom
part	6777	nom
figure	6766	nom
network	6716	nom
business	6570	nom
large	6545	adj
report	6479	nom
product	6456	nom
digital	6345	nom
application	6322	nom
begin	6319	ver
tion	6188	nr
market	6181	nom
set	6079	ver
study	5997	nom
interest	5994	nom
state	5936	nom
call	5911	nom
paper	5882	nom
press	5837	nom
article	5832	nom
model	5747	nom
de	5713	nr
memory	5551	nom

Figure 23 Formes Actives corpus 2

Anche nel secondo corpus, come nel primo, la classe nominale domina nettamente tra le forme più frequenti. Termini come *computer (51.438), system (29.534), data (12.517), information (10.295), software (13.928), technology (12.530), machine (20.111),* emergono chiaramente e indicano la centralità dei concetti informatici. Tuttavia, si nota una maggiore enfasi storica e contestuale, con frequenze rilevanti per termini quali:

- *history (23.523), development (10.576), early (12.418), research (12.251), project (9.724), work (21.005), article (5.832), time (16.938)*

Sono, di conseguenza, più presenti che nel corpus 1, concetti di temporalità e sviluppo: *early (12.418), late (3.556), time (16.938), period (2.815), begin (6.319)*.
Inoltre nel corpus 2 emergono più chiaramente:

- nomi propri o quasi-propri (ad es. *ibm (15.750), bell (1.425), mit (3.706), unix (807)*): indicano luoghi, aziende, sistemi e contesti storici determinati;
- nomi legati alla ricerca e alla pubblicazione (ad es. *paper (5.882), article (5.832), journal (1.451), issue (4.804), vol (10.374*): segnalano la dimensione editoriale;

L'analisi delle forme attive mostra una presenza significativa di aggettivi valutativi e temporali, come *early* (12 418 occorrenze), *important* (4 105), *major* (3.724)*, significant* (1.325) e *historical* (2.519), che assumono un ruolo cruciale nel definire il tono interpretativo del discorso.
L'esame delle concordanze relative agli aggettivi *early, important, major* e *significant* conferma il ruolo di questi termini come marcatori linguistici del discorso storiografico e interpretativo che caratterizza il sottocorpus contemporaneo (IEEE Annals of the History of Computing).
L'aggettivo *early* è frequentemente impiegato per collocare nel tempo le origini di innovazioni, teorie o paradigmi, contribuendo a costruire una narrazione genealogica del progresso tecnologico. Gli estratti mostrano l'uso di early in espressioni come:
"*Still the **early** computers are variations on the protean design of a limited Turing machine*" (IEEE Annals, 2018),
"*By the **early** 1970s Codd actively argued the merits of the relational approach*" (IEEE Annals, 2012),
"*The **early** 1950s were no less important in terms of their architecture and utility*" (IEEE Annals, 2006).
Allo stesso modo, gli aggettivi *important, major* e *significant* funzionano come indicatori valutativi che enfatizzano la rilevanza o la portata di scoperte e applicazioni.
Le concordanze testimoniano questa tendenza:
"*It is **important** to understand how the Geromati I was valued as a meaningful object for them to design*" (IEEE Annals, 2020),
"*The photocopy machine itself is, of course, an **important** invention with great social consequence*" (IEEE Annals, 2009),
"*The logic of relational databases, which became a **major** IBM project later in that decade*" (IEEE Annals, 2021),

"*Libraries reporting their government led to* ***significant*** *practical applications*" (IEEE Annals, 2002),

La co-occorrenza ricorrente di *important, major* e *significant* con sostantivi come *development, project, invention, advance* o *application* evidenzia un modello ricorrente di valutazione scientifica, attraverso il quale gli autori esprimono giudizi di rilevanza e priorità sugli eventi tecnologici[27]. Attraverso questa struttura lessicale, i testi non si limitano a descrivere tecnologie, ma attribuiscono loro valore e funzione narrativa, enfatizzando la dimensione di progresso, influenza e impatto storico.

Anche in questo corpus si rilevano anomalie nella segmentazione o nel lemmatizing automatico. Alcuni esempi ricorrenti:

- *tion* (6.188) appare ancora, come nel corpus 1, indicando un probabile errore di tokenizzazione;
- *net* (1.225) o abbreviazioni simili possono, invece, rappresentare forme troncate da URL, nomi di dominio o codici.

Come osservato nell'analisi del corpus di partenza, anche nel *sottocorpus computer1* si conferma la netta predominanza della categoria nominale, coerente con la natura dei testi analizzati:

[27] Per un approfondimento teorico sul ruolo dell'evaluation nella costruzione del punto di vista autoriale e nella marcatura della rilevanza discorsiva, si veda Hunston & Thompson (2000), Evaluation in Text: Authorial Stance and the Construction of Discourse, Oxford University Press.

Forme	Freq.	Types
computer	177324	nom
system	64366	nom
conference	35096	nom
program	32126	nom
data	29718	nom
joint	22500	nom
time	20909	nom
figure	18194	nom
control	17419	nom
process	16208	nom
design	14887	nom
user	14059	nom
national	13240	nom
problem	12666	nom
pp	12169	nr
information	11579	nom
memory	11525	nom
network	9878	nom
digital	9808	nom
language	9344	nom
fall	9334	nom
require	9281	ver
numb	9038	adj
function	8959	nom
operation	8656	nom
application	8531	nom
large	8363	adj
communication	8154	nom
model	7933	nom
software	7853	nom
vol	7614	nr
cost	7595	ver
test	7587	nom
set	7561	ver
machine	7477	nom
input	7448	nom
provide	7384	ver
spring	7304	nom
proceeding	6977	nom

Forme	Freq.	Types
result	6968	nom
development	6807	nom
output	6752	nom
work	6693	nom
processor	6618	nom
type	6603	nom
line	6473	nom
analog	6451	nom
file	6384	nom
display	6346	nom
terminal	6274	nom
base	6233	adj
science	6188	nom
hardware	6170	nom
analysis	6142	nom
simulation	6115	nom
operate	6105	ver
technique	5938	nom
structure	5934	nom
show	5908	nom
word	5870	nom
point	5689	nom
storage	5683	nom
tion	5648	nr
form	5621	nom
service	5593	nom
table	5579	nom
research	5528	nom
device	5352	nom
method	5339	nom
level	5321	adj
unit	5295	nom
high	5269	adj
develop	5260	ver
report	5209	nom
instruction	5163	nom
general	5134	nom
university	5134	nom
part	5126	nom

Figure 24 Formes Actives Sottocorpus 1

I sostantivi più frequenti fanno riferimento a componenti e concetti fondamentali dell'hardware e dell'architettura computazionale dell'epoca: *machine (7.477), memory (11.525), processor (6,618), terminal (6.274), system (64.366), input (7.448), output (6.752)*. In particolare, il sostantivo *mainframe* compare con una frequenza significativa, riflettendo l'egemonia dei grandi sistemi centralizzati nel panorama informatico del periodo.

Si rileva inoltre la presenza di nomi legati all'ambito applicativo o istituzionale, come *university (5.134), laboratory (2.574), research (5.528), application (8.531)*, che denotano il contesto d'uso prevalente dei computer in quegli anni: ambienti accademici, governativi e industriali, piuttosto che il mercato di consumo.

L'aggettivo *digital* (9.808) risulta particolarmente frequente, a testimonianza della necessità, in questo periodo, di distinguere in modo marcato tra tecnologie digitali e analogiche — distinzione evidenziata anche dal minor numero di occorrenze di *analog* (6.451).

Le forme verbali, d'altra parte, sono rare e prevalentemente legate a funzioni operative (*run (4.397), operate (6.105), connect (1.954), store (4.581*).

Forme	Freq.	Types
computer	51438	nom
system	6918	nom
history	6543	nom
compute	6513	nom
ieee	5497	nr
program	5206	nom
science	4840	nom
work	4081	nom
pp	3644	nr
machine	3602	nom
early	3517	adj
design	3480	nom
ibm	3475	nr
technology	3233	nom
time	3161	nom
university	3092	nom
software	3059	nom
research	2999	nom
digital	2856	nom
development	2725	nom
industry	2610	nom
vol	2588	nr
engineer	2533	nom
process	2517	nom
annals	2395	nr
year	2349	nom
company	2332	nom
data	2306	nom
information	2281	nom
project	2271	nom
develop	2231	ver
electronic	2204	nr
society	1984	nom
build	1917	nom
apply	1901	ver
user	1748	nom
publish	1679	ver
limit	1677	nom
press	1670	nom

Forme	Freq.	Types
network	1670	nom
control	1641	nom
include	1636	ver
center	1632	nom
business	1609	nom
application	1585	nom
large	1583	adj
ing	1554	nr
department	1490	nom
group	1488	nom
personal	1461	nom
begin	1451	ver
market	1448	nom
problem	1423	nom
study	1420	nom
language	1402	nom
person	1391	nom
national	1383	nom
license	1382	nom
report	1350	nom
restriction	1341	nom
february	1341	nr
article	1337	nom
authorize	1329	ver
interest	1322	nom
xplore	1290	nr
utc	1254	nr
di	1250	nr
memory	1246	nom
security	1245	nom
part	1238	nom
world	1236	nom
universita	1225	nr
write	1210	ver
degli	1204	nr
pro	1191	nom
studi	1183	nr
paper	1179	nom
service	1177	nom

Figure 25 Formes Actives Sottocorpus 2

Come negli altri corpora, i sostantivi costituiscono la stragrande maggioranza delle forme più frequenti. I termini tecnici principali includono: *computer (51.438), system (6.918), software (3.059), technology (3.233), network (1.670), machine (3.602), information (2.281), program (5.206)*. Sono molto ricorrenti anche i termini che esprimono concetti temporali e storici — *history (6.543), early (3.517), development (2.725), article (1.337), research (2.999), time (3.161)* - a testimonianza del carattere riflessivo e cronologico dei testi.

Si rileva inoltre la presenza di nomi propri o sigle *ieee (5.497), ibm (3.475), pp (3.644), vol (2.588), annals (2.395)*) che rimandano a riviste e conferenze, e la comparsa di termini legati a istituzioni e contesti accademici, come *university (3.092), science (4.840), study (1.420), education (670), faculty (277), student (693)*.

Nel *sottocorpus 2* (IEEE), pur restando forte la componente accademica, si osserva una maggiore articolazione: emergono riferimenti specifici alla carriera universitaria, alla ricerca scientifica e alla storia dell'informatica. A ciò si affiancano nuovi assi tematici, come quello culturale e ludico, rappresentato da termini legati all'informatica domestica, ai giochi e al fenomeno degli hobbisti degli anni '80 - si vedano *Apple (418)*, *game (893)*, *hobbyist (226*). Inoltre, acquisisce rilievo anche la dimensione economica, con un lessico riferito a mercato, industria e produzione.

Nel corpus *Reddit* il risultato è il seguente:

Forme	Freq.	Types
https	826	nr
computer	765	nom
www	715	nr
retrobattlestations	713	nr
reddit	680	nr
comment	621	nom
week	486	nom
titolo	305	nr
corpo	305	nr
commenti	305	nr
work	286	nom
x86	238	nr
machine	227	nom
drive	212	nom
pc	207	nr
year	206	nom
run	203	ver
time	194	nom
find	183	nom
apple	180	nom
retro	169	nr
game	166	adj
portable	165	nom
monitor	155	nom
keyboard	155	nom
thing	149	nom
love	145	nom
http	141	nr
back	133	nom
ram	130	nom
system	128	nom
pizza	128	nom
post	127	nom
home	126	nom
ibm	118	nr
pravetz	116	nr
contest	114	nom
mac	114	nom
box	113	nom

good	109	adj
card	109	nom
vintage	107	nom
build	103	nom
case	102	nom
memory	99	nom
picture	98	nom
window	97	nom
person	97	nom
crt	94	nr
floppy	93	adj
photo	91	nom
video	91	nom
software	91	nom
stuff	90	nom
model	90	nom
play	89	nom
screen	89	nom
nice	89	adj
bite	89	nom
great	88	adj
cool	86	adj
basic	85	nom
original	85	nom
ii	85	nr
program	84	nom
day	84	nom
power	81	nom
jpg	81	nr
board	81	nom
display	79	nom
early	77	adj
part	77	nom
hp	77	nr
compute	76	nom
processor	75	nom
personal	73	nom
start	72	nom
imgur	72	nr

Figure 26 Formes Actives corpus Reddit

I termini tecnici centrali del dominio informatico: *computer (765), pc (207), machine (227), drive (212), system (128), processor (75), ram (130)*, sono comunque presenti, ma appaiono circondati da parole che rimandano a un contesto d'uso concreto, esperienziale e spesso nostalgico come *keyboard (155), monitor (155), floppy (93)*[28]*, apple (180), mac (114), retro (169), vintage (107)*. Questi termini, oltre a denotare oggetti fisici, evocano anche l'estetica e la memoria collettiva dell'informatica storica, in linea con la natura del subreddit *r/retrobattlestations.*

Si osserva inoltre un'alta frequenza di termini associati alla dimensione visiva e digitale della comunicazione online: *https (826), www (715), jpg (81), video (91), picture (980), photo (91), screen (89)*, che riflettono l'importanza di contenuti multimediali (immagini, screenshot, setup fotografici)

[28] È interessante notare che floppy (93) viene classificato automaticamente come aggettivo, nella sua traduzione comune dall'inglese di "molle" o "floscio" invece che come termine tecnico, probabilmente per la mancanza di contesto lessicale che consenta di ricostruire la locuzione floppy disk. Questo caso evidenzia i limiti dell'analisi automatica quando le unità terminologiche composte vengono trattate come parole isolate.

all'interno delle discussioni. Termini come *comment (621), titolo (305), corpo (305)* vanno invece interpretati come metadati strutturali, introdotti dallo script di estrazione.

Interessante è anche la presenza di parole a forte carica valutativa o affettiva, come *love (145), nice (86), cool (86), great (88), good (109)*, tipiche di un linguaggio spontaneo e apprezzativo che valorizza l'interazione sociale e il riconoscimento all'interno della community. Allo stesso modo, il lessico include riferimenti diretti alla vita online e partecipativa, come *post (127), week (486), contest (114), pizza (128)*, che rinviano a rituali digitali, competizioni e consuetudini del subreddit.

In sintesi, il corpus Reddit mostra un profilo linguistico ibrido, in cui la terminologia tecnica convive con un registro discorsivo informale. A differenza dei corpora accademici, dove domina la descrizione concettuale e normativa, qui prevale l'esperienza vissuta, l'identificazione soggettiva con gli oggetti e la componente partecipativa della memoria digitale.

Nel *corpus museale* il risultato è il seguente:

Forme	Freq.	Types
computer	3351	nom
game	1302	adj
desktop	1072	nom
device	766	nom
commodore	743	nom
apple	682	nom
model	628	nom
data	594	nom
philips	578	nr
console	532	nom
ibm	529	nr
micro	515	nom
system	494	nom
output	465	nom
video	438	nom
storage	437	nom
bbc	431	nr
calculator	422	nom
â	414	nr
personal	413	nom
electronic	398	nr
transfer	391	nom
atari	367	nr
machine	359	nom
portable	359	nom
nintendo	359	nr
monitor	358	nom
drive	326	nom
cassette	326	nom
unit	303	nom
design	299	nom
bite	287	nom
board	283	nom
card	280	nom
power	278	nom
disk	264	nom
series	252	nr
software	241	nom
sinclair	238	nr

print	230	nom
keyboard	215	nom
part	215	nom
manufacture	214	nom
work	213	nom
program	212	nom
printer	212	nom
tandy	211	nr
year	210	nom
icl	209	nr
sony	203	nr
box	200	nom
tape	198	nom
punch	197	nom
compute	189	nom
packard	186	nr
digital	176	nom
hewlett	176	nr
ferranti	175	nr
input	174	nom
pc	173	nr
microcomputer	171	nom
ii	170	nr
prototype	169	nom
world	161	nom
circuit	160	nom
include	160	ver
project	157	nom
school	157	nom
user	156	nom
launch	155	nom
publish	155	ver
case	153	nom
build	152	nom
time	151	nom
manual	150	nom
component	149	nom
manchester	149	nr
type	140	nom

Figure 27 Formes Actives del corpus museale

Nel corpus museale, le forme attive riflettono un linguaggio fortemente oggettuale e descrittivo. I termini più frequenti appartengono quasi esclusivamente alla categoria dei sostantivi, a conferma della funzione primaria museali di nominare, classificare e descrivere dispositivi e componenti fisici in modo preciso e accessibile, evidenziando così la centralità degli oggetti e delle loro componenti fisiche.

Ad esempio, un oggetto come il *Commodore 64* viene documentato nei musei attraverso una scheda che include il titolo dell'oggetto (es. *Commodore 64*), la classificazione funzionale (es. *home computer*), una descrizione tecnica e storica (es. "microcomputer domestico a 8-bit diffuso negli anni '80"), l'elenco delle componenti principali (tastiera integrata, porte I/O, processore), le misure fisiche e i dati di produzione (produttore, anno, provenienza).

Tra i termini centrali figurano *computer* (3351 occorrenze), *game* (1302), *desktop* (1072), *device* (766), *console* (532), *system* (494), *output* (465), *monitor* (358), *keyboard* (215), *printer* (212), *input*

(174), *microcomputer* (171), *component* (149), tutti associati alla materialità dei dispositivi digitali storici.

È notevole la presenza di nomi propri di aziende produttrici come *commodore (743)*, *apple (682)*, *ibm (529)*, *philips (578)*, *atari (367)*, *nintendo (359)*, *bbc (431)*, *sony (203)*, *tandy (211)*, *hewlett (176)*, *packard (186)*, *ferranti (175)*, *icl (209)*, che evidenziano l'approccio documentale e storico della narrazione museale, in cui il marchio è un dato rilevante quanto la funzione o la tipologia dell'oggetto.

Emergono, inoltre, termini che rinviano a componenti tecnici: *circuit (160), tape (198), storage (437), transfer (391), card (280), unit (303), board (283), disk (264*), funzioni (*output (465), input (174), display (88), power (278), component (149), manual (150*), forme di supporto e interazione con l'oggetto digitale (*cassette (326), box (200), calculator (422), launch (155*). La descrizione tende quindi a essere oggettiva, funzionale e orientata alla ricostruzione dell'uso storico del dispositivo.

Infine, la presenza di termini come *school (157), university (139), educational (64), user (156), project (157)* suggerisce che una parte dei testi museali affronti anche il contesto d'uso sociale e pedagogico dei computer storici.

3.2.1.2 Analisi di similitudine (con indice “sur les arêtes”)

L’analisi di similitudine è stata condotta per osservare le co-occorrenze tra termini e la loro frequenza relativa all’interno dei due corpora principali (*computer 1* e *computer 2*). Attraverso l’uso di grafi visualizzati con IRaMuTeQ, sono stati identificati dei cluster tematici rappresentati con colori distinti, attraverso l’uso della libreria *igraph* di R basata su indici.

Ogni grafo include:

Nodi: rappresentano i termini più frequenti (forme attive), la cui dimensione è proporzionale alla loro frequenza nel corpus. Come possiamo vedere nei due corpus, il termine ‘computer’ risulta essere più centrale nel secondo corpus rispetto al primo, che vede ‘system’ e ‘data’ occorrere più frequentemente.

Archi (halo): rappresentano i legami di co-occorrenza tra i termini, calcolati in base alla loro prossimità contestuale. Lo spessore degli archi riflette l’intensità del legame (più spesso = co-occorrenza più forte).

I cluster tematici sono stati evidenziati mediante colori distinti, e ciascun cluster riflette un ambito concettuale omogeneo emergente dalla distribuzione delle parole nel corpus.

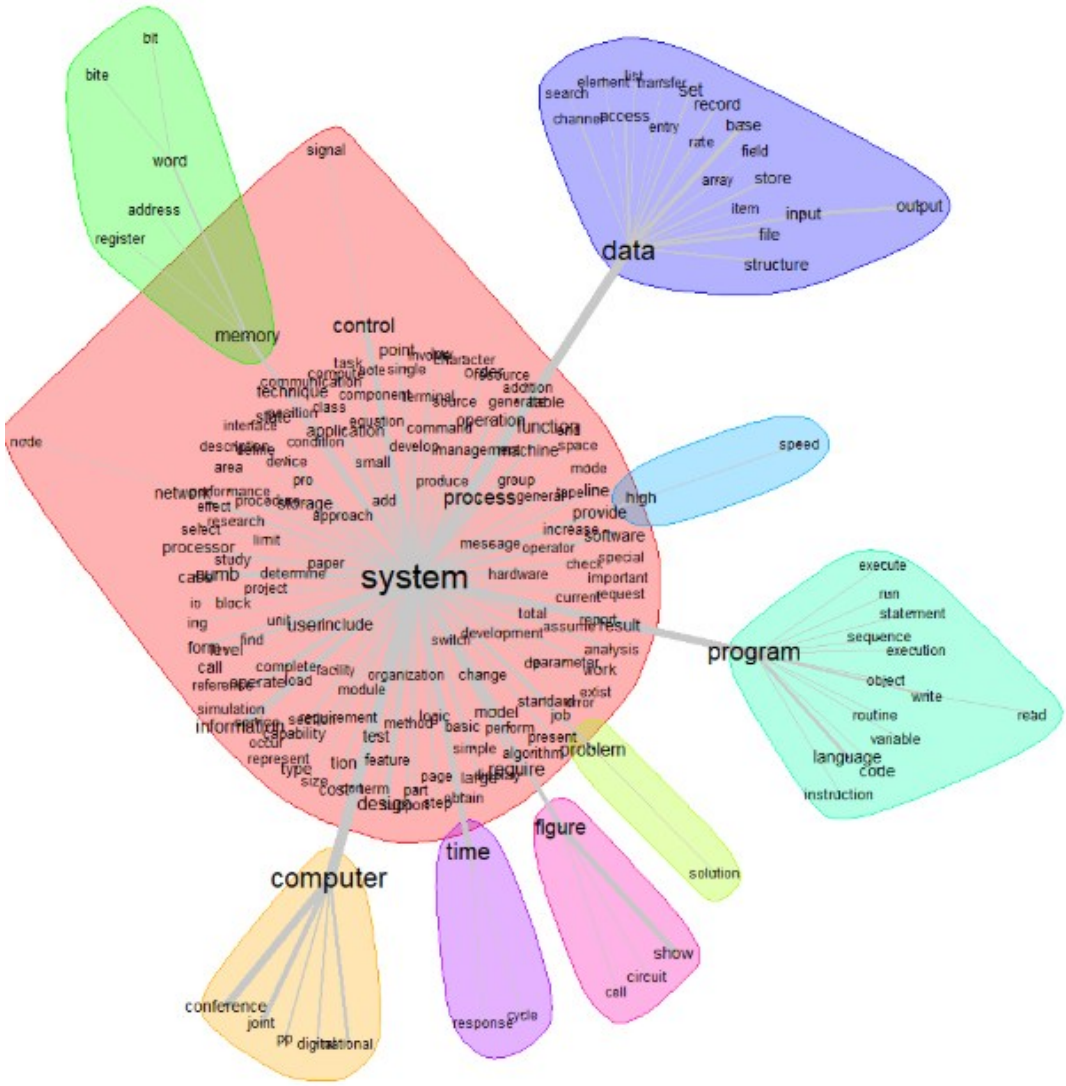


Figure 28 Analyses de similitude del corpus 1

Nel Corpus 1, i termini centrali (*system, data, processor, function*) delineano un nucleo lessicale fortemente orientato agli aspetti operativi e architetturali dell'informatica, con una prevalenza di termini che rimandano sia ai processi computazionali (*function, execute, process*) sia alla struttura interna dei sistemi *(system, memory, processor*). Intorno a questo nucleo si distribuiscono cluster satelliti come *memory, computer, program* e *figure*, che riflettono i principali campi applicativi e metodologici del linguaggio tecnico dell'epoca.

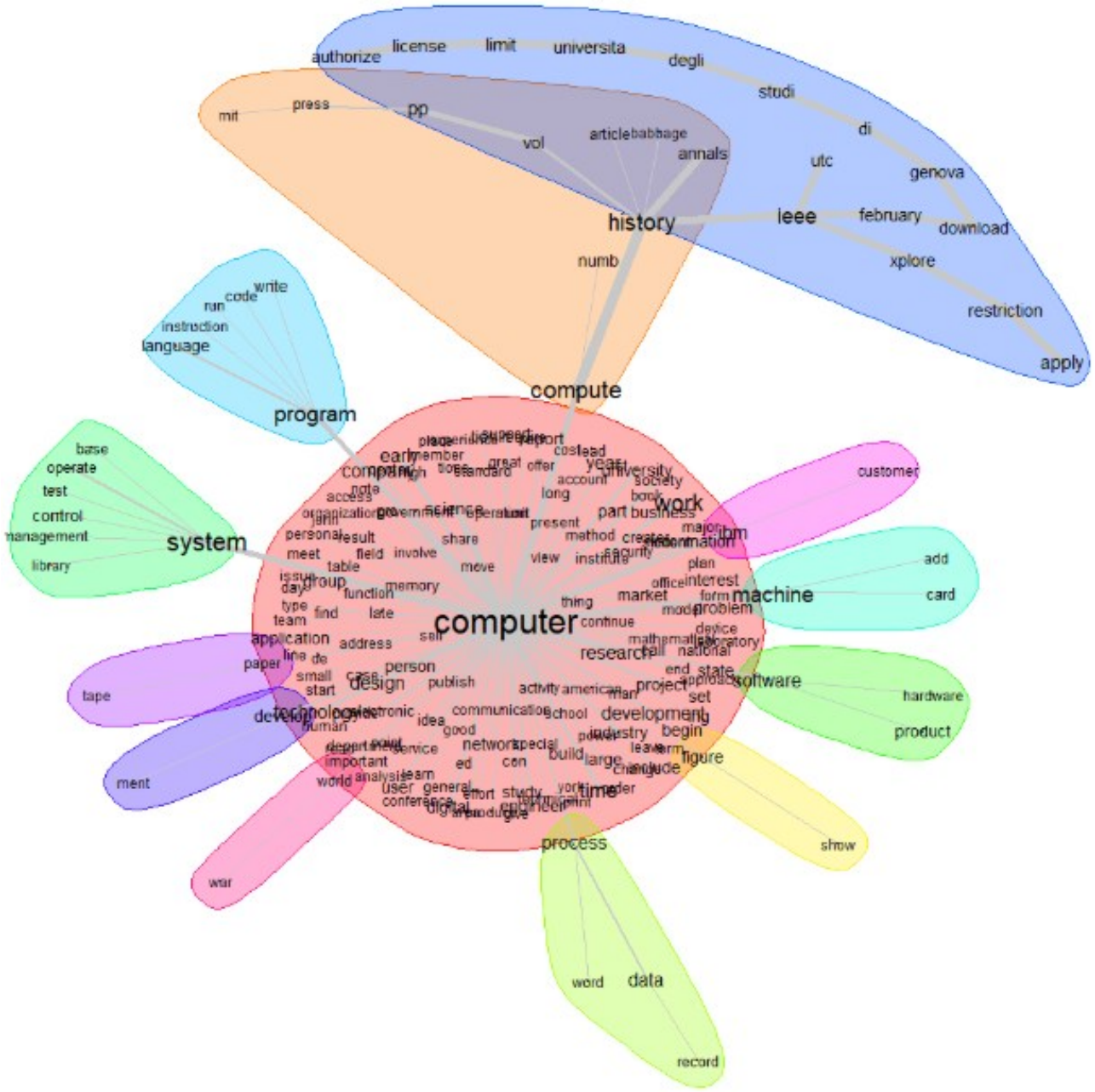


Figure 29 Analyses de similitude del corpus 2

Nel Corpus 2, il nodo *computer* acquisisce maggiore centralità semantica, configurandosi come hub attorno al quale si aggregano cluster legati non solo alla tecnica (*system*, *machine*, *program*), ma anche alla storia e diffusione della tecnologia, come suggerito dai gruppi *history*, *work*, *development* e *research*.

Dall'analisi emergono associazioni ricorrenti tra i termini *analog* e *digital* in relazione diretta al nodo *computer* in entrambi i corpora. Questo riflette la storica dicotomia tra le tecnologie analogiche e digitali nel campo dell'informatica.

Rispetto al primo corpus, come abbiamo già notato nelle statistiche precedenti, si osserva una maggiore attenzione agli aspetti istituzionali e autoriflessivi della comunicazione scientifica,

testimoniata dalla frequente menzione di sigle accademiche e fonti editoriali come *IEEE*, *Annals* e *pp*.

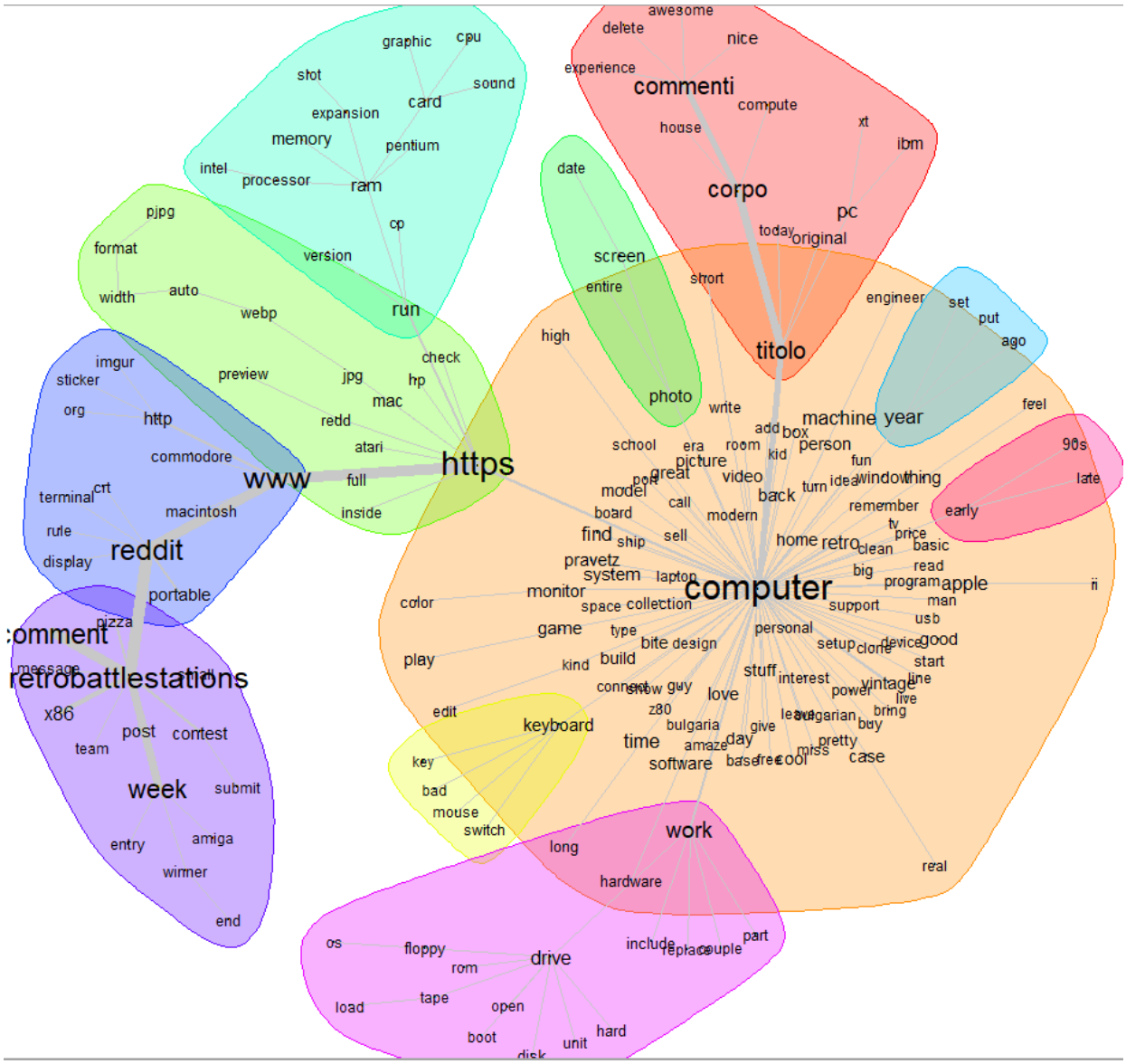


Figure 30 Analyses de similitude del corpus Reddit

- Cluster rosa, in basso a destra: termini come *drive, floppy, rom, disk, boot* evidenziano un'area dedicata ai dispositivi di memorizzazione e al caricamento del sistema.
- Cluster viola, in basso a sinistra: *retrobattlestations, amiga, x86, contest, week* e *pizza* rimandano chiaramente al contesto del subreddit dedicato alle macchine vintage e alle competizioni tra collezionisti.
- Cluster verde-chiaro/azzurro, in alto a sinistra: *https, www, jpg, preview, atari, macintosh, ram, pentium* formano un ponte tra contenuti web e specifiche hardware dell'epoca.
- Cluster rosso, in alto a destra: *commenti, corpo, titolo, pc, compute, nice, awesome* indicano il linguaggio più generico dei post, il "contenuto parlato" piuttosto che la materia tecnica.

- Cluster arancio, al centro: la parola computer funge da hub principale, circondata da *game, monitor, keyboard, software, case, work, retro, time*, ecc., racchiude il nucleo semantico più ampio e "neutro".
- Cluster blu chiaro, lato destro: *year, 90s, early, late, ago* segnalano una piccola sottorete incentrata sugli aspetti cronologici e sui decenni di interesse.

Nel corpus Reddit, la struttura del grafo mostra una distribuzione più radiale e dispersa. Il termine *computer* resta centrale, ma circondato da cluster tematici legati all'esperienza d'uso, alla descrizione visiva (*photo*, *screen*, *keyboard*) e all'interazione sociale (*comment*, *week*, *contest*, *retrobattlestations*). Il lessico tecnico convive con elementi valutativi (*great*, *cool*, *nice*) e simbolici (*mac*, *atari*, *https*), segnalando una forma discorsiva più narrativa, affettiva e partecipativa.

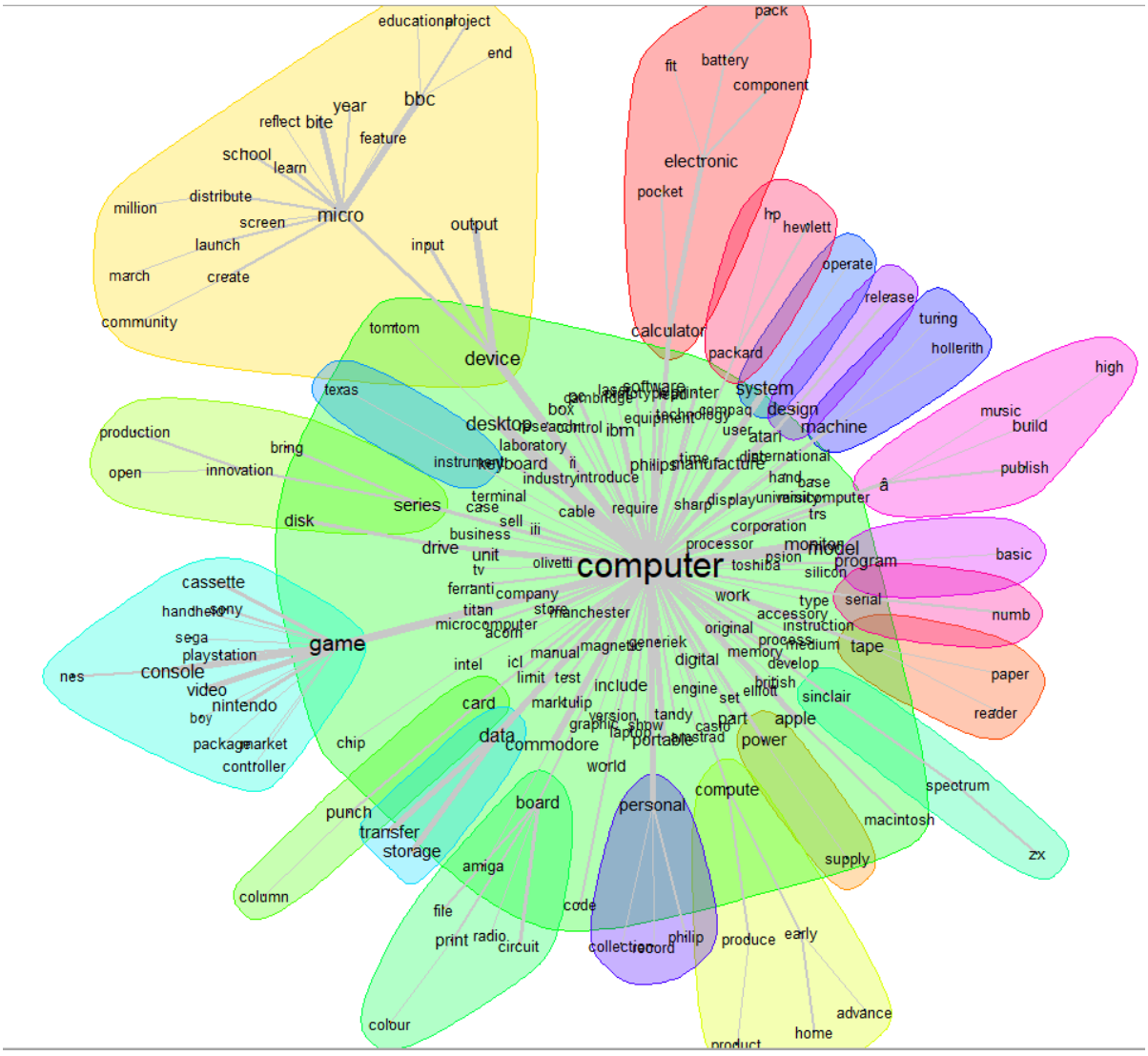


Figure 31 Analyses de similitude del corpus museale

Nel corpus museale, l'organizzazione lessicale è fortemente orientata all'oggetto fisico e alla sua catalogazione: il termine *computer* è associato a nodi che rinviano a componenti materiali (*device*,

component, *monitor*, *keyboard*), produttori storici (*Commodore*, *Apple*, *Philips*, *Ferranti*) e classificazioni funzionali (*calculator*, *micro*, *personal*, *console*). La rete semantica è più densa e compatta, riflettendo un lessico descrittivo, nominale e finalizzato alla documentazione museale standardizzata.

Il passaggio dal computer come oggetto tecnico-specialistico (sottocorpus 1) al computer come artefatto culturale, sociale e storico (sottocorpus 2) emerge chiaramente dall'evoluzione dei termini co-occorrrenti, come abbiamo visto nelle analisi precedenti.
Nel sottocorpus 1, il termine computer co-occorre principalmente con *system, processor* e *memory*, componendo un'area semantica orientata all'architettura e al funzionamento:
• «The computer ***system*** executes programs stored in memory through processor-controlled operations».
(AFIPS Joint Computer Conferences, 1973)
• «The ***processor*** coordinates the execution of instructions and manages access to memory resources».
(AFIPS Joint Computer Conferences, 1978)
• «Data are transferred between the central ***memory*** and input-output devices according to system scheduling policies».
(AFIPS Joint Computer Conferences, 1969)

Nel *sottocorpus 2*, invece, computer si associa più spesso a *history, development, research / laboratories*, indicando una narrazione riflessiva e contestuale, come abbiamo osservato nelle analisi di specificità e AFC. Gli estratti seguenti provengono direttamente dal corpus:
• «…how the process of calculating has been performed throughout ***history*** as well as the people, events and forces involved in the development of the electronic computer…»
source_IEEE_AnnalsOfTheHistoryOfComputing, year_1985, 0_1
• «…interested [in the] ***development*** of the electronic digital computer and in ***historical*** topics…»
source_IEEE_AnnalsOfTheHistoryOfComputing, year_1985, 0_7
«…later computer ***developments***, the IBM 700 series…»
source_IEEE_AnnalsOfTheHistoryOfComputing, year_1985, 0_21

• «…he spent the war years in radar ***research***…»
source_IEEE_AnnalsOfTheHistoryOfComputing, year_1985, 3_584
«…visited several ***research laboratories*** and computer manufacturers in the United States…»
source_IEEE_AnnalsOfTheHistoryOfComputing, year_1985, 3_734

• «…a ***history*** of computing more in the nature of a practical ***project***…»
source_IEEE_AnnalsOfTheHistoryOfComputing, year_1985, 0_45
«…during its seven-year life the System R project published more than 40 technical papers…»
source_IEEE_AnnalsOfTheHistoryOfComputing, year_2012, 850_131243

Questo spostamento segnala una ridefinizione discorsiva del calcolatore: da tecnologia da progettare, ottimizzare e controllare, a oggetto da interpretare, documentare e contestualizzare all'interno di pratiche sociali e culturali. Per comprendere come tali configurazioni discorsive si consolidino all'interno delle reti terminologiche, è necessario osservare la struttura delle co-occorrenze nei grafi prodotti *dall'Analyse de similitude* e, in particolare, il ruolo giocato dagli *Indices sur les arêtes* nel definire la forza dei legami tra i nodi.

3.2.1.3 Indices sur les arêtes

Gli *Indices sur les arêtes* ovvero valori numerici che rappresentano il numero esatto di co-occorrenze tra coppie di termini.

Questa funzionalità consente di quantificare la forza del legame tra i termini all'interno del grafo, andando oltre la semplice visualizzazione grafica delle comunità e di individuare con maggiore precisione quali termini fungono da ponte tra cluster diversi o quali sono i nodi fortemente coesi.

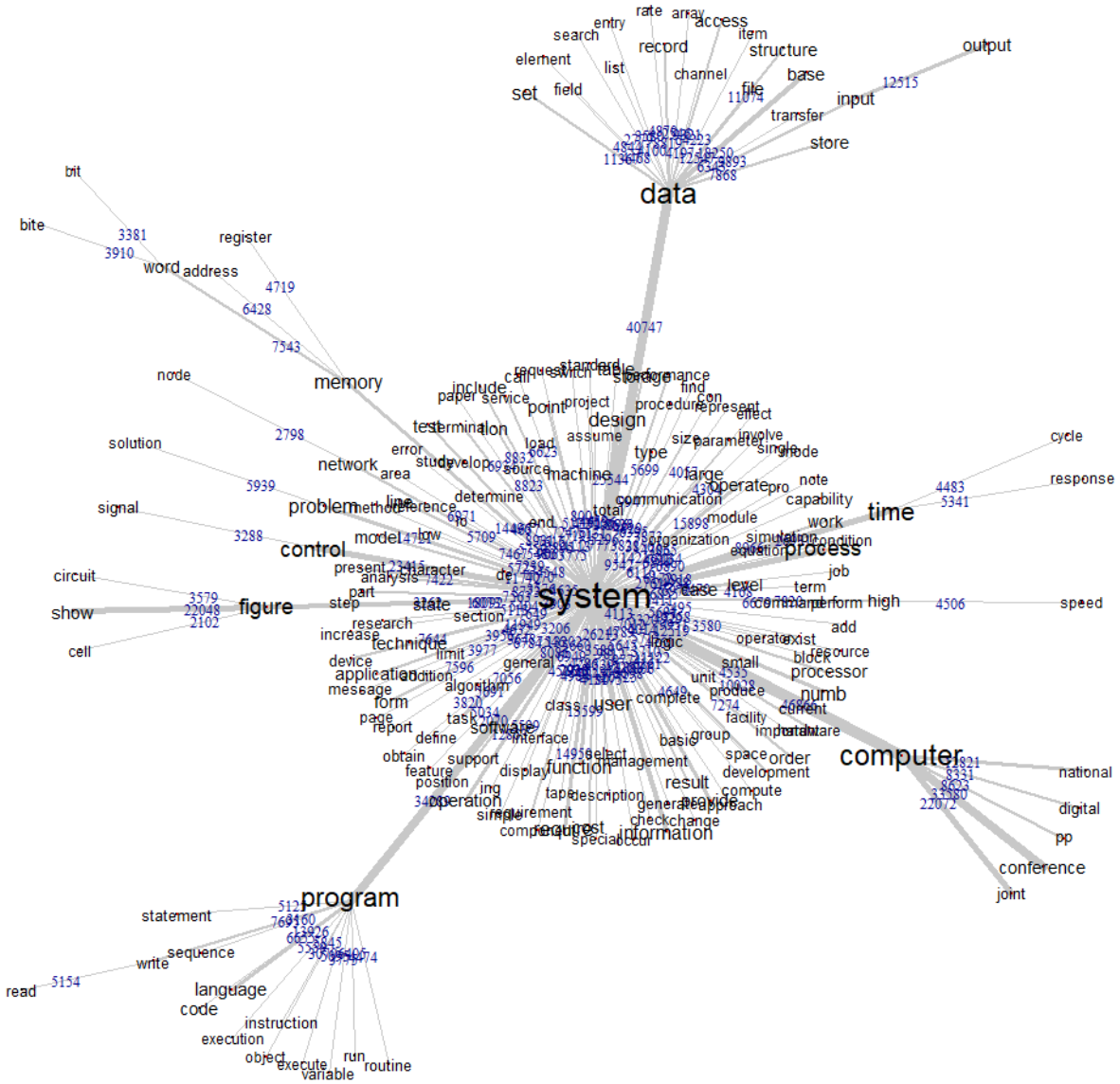


Figure 32 Analyses de similitude con Indices sur les arêtes del corpus 1

La connessione tra *system* e *data* ha indice 40747, indicando una co-occorrenza estremamente frequente, quindi un forte legame semantico o discorsivo relativamente al resto del grafo, ovvero nel

confronto con gli altri valori presenti nel dataset. La maggior parte degli archi visibili nei due grafi mostra valori tra 500 e 5000. Alcuni picchi superano i 10000 (es. data e output = 12815; ieee e history = 10785), ma 40747 è un outlier marcato.

Alcuni termini fungono da connettori inter-cluster, termini nodo che hanno legami significativi con più cluster, come *system* collega molteplici concetti: *program*, *control*, *data*, *computer*, ecc.

Termini come *program*, *memory*, *control* appaiono fortemente interconnessi con *system*, suggerendo un **c**luster tecnico-operativo.

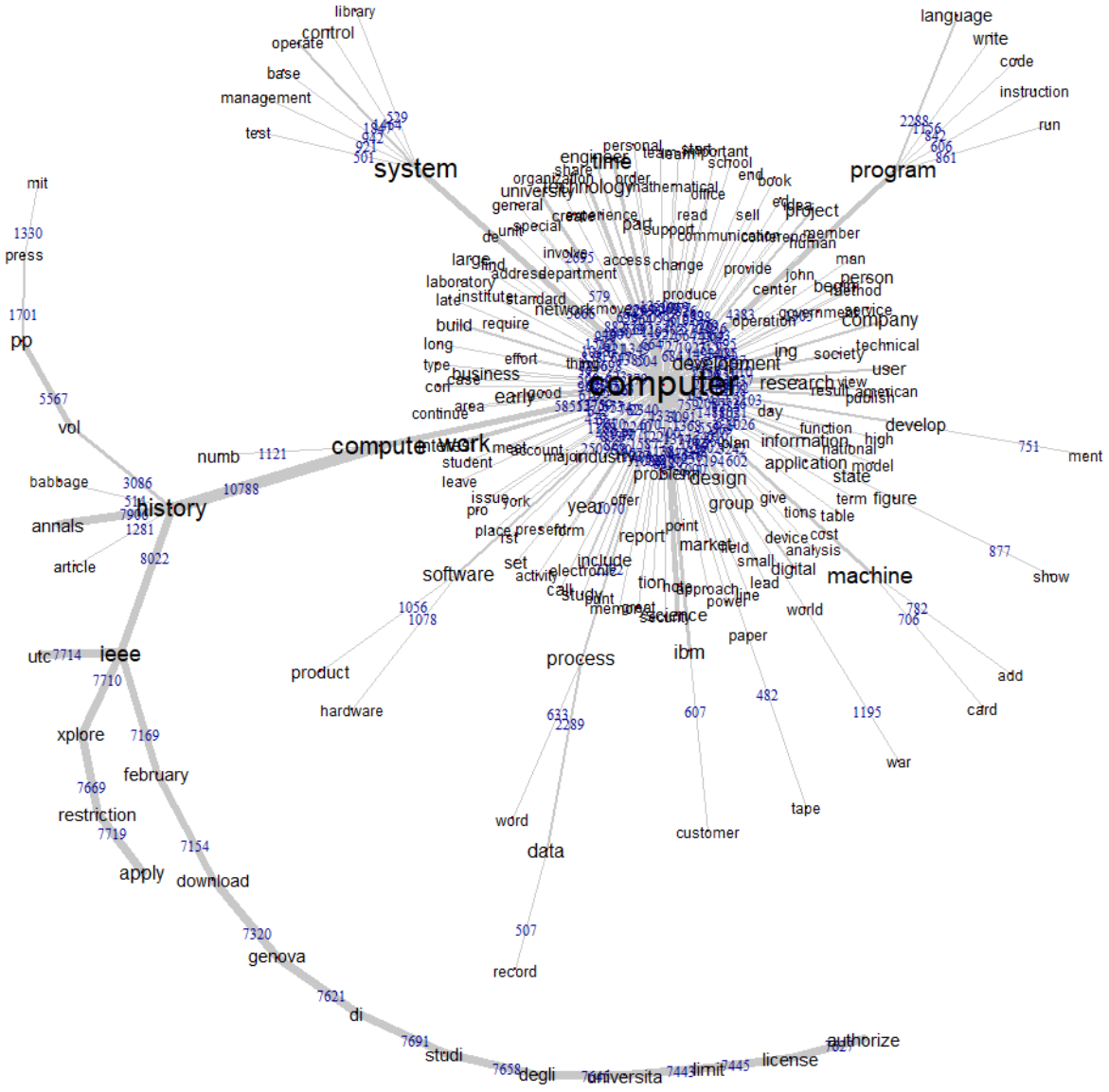


Figure 33 Analyses de similitude con Indices sur les arêtes del corpus 2

Computer nel secondo grafo si collega sia ad ambiti tecnici (*process, machine*) sia storici (*ieee, history*).

Per comprendere l'importanza della lemmatizzazione nella costruzione di reti semantiche, è stata effettuata un'analisi parallela e comparativa del *Corpus 1*, questa volta senza applicare alcuna normalizzazione linguistica.

L'obiettivo di questo confronto è evidenziare come la mancanza di lemmatizzazione comporti una maggiore frammentazione concettuale e una riduzione della coesione semantica tra i termini.

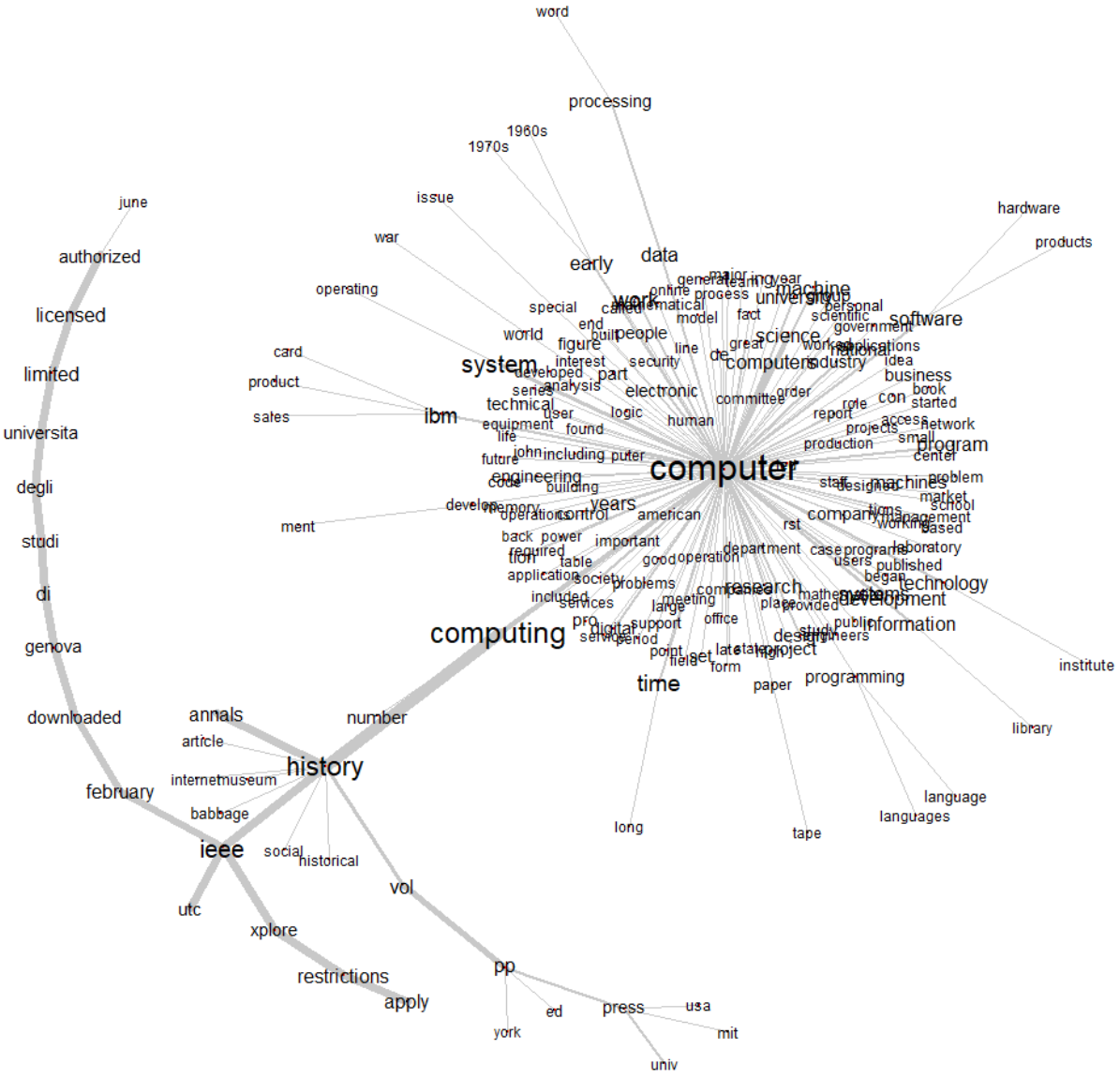


Figure 34 Analyses de similitude senza lemmatizzazione del corpus 1

Con la lemmatizzazione, il grafo è più compatto e semanticamente coeso. Termini come *computing, computer, computers* vengono ricondotti a un'unica forma (computer), aumentando la densità dei legami e la visibilità dei concetti forti.
Senza lemmatizzazione, invece, queste forme restano separate: *computing* e *computer* appaiono come nodi diversi e possono finire in cluster differenti, disperdendo l'informazione.
Nel grafo senza lemmatizzazione, aumentano i nodi "sporchi" o ridondanti, come *pp, vol, press, licensed, authorized, limited*, che possono essere residui di metadati o note bibliografiche.
La rete si frammenta, e alcuni cluster si confondono o perdono definizione per via della frammentazione terminologica.

Di seguito il risultato de l'*Indice sur les aretes* sul corpus *Reddit*:

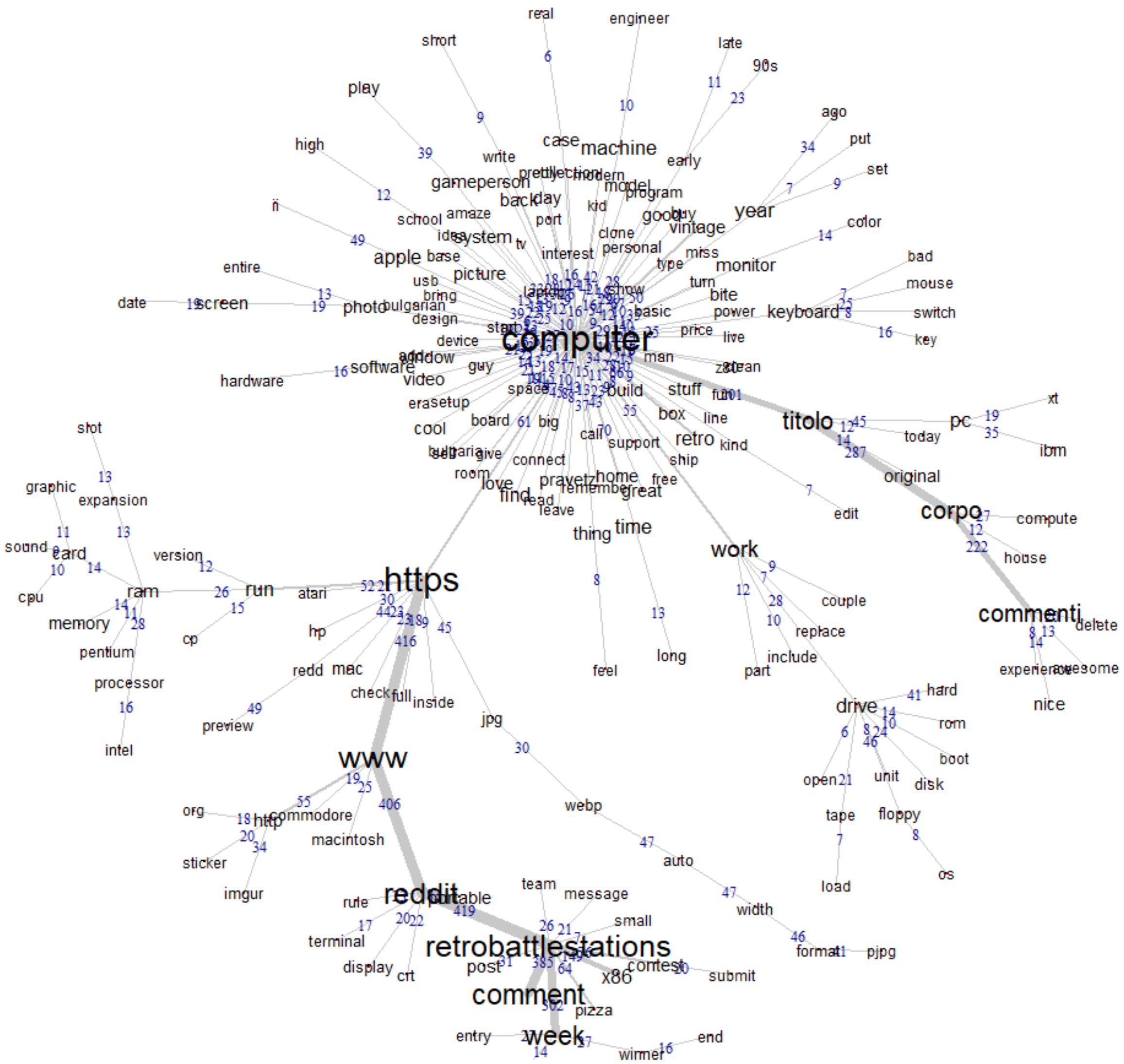


Figure 35 Analyses de similitude e indice sur les aretes senza lemmatizzazione del corpus Reddit

Sebbene il grafo risulti visivamente denso, il software consente di esportare i dati tabellari in formato CSV, permettendo un'analisi quantitativa più precisa delle co-occorrenze. Il nodo *computer* rappresenta il fulcro della rete, con legami particolarmente forti verso termini come *monitor* (145), *cool* (101), *keyboard* (25), *machine* (42), *retro* (66) e *game (33)*, che testimoniano una narrativa centrata sull'estetica e funzionalità dei dispositivi storici.

Un cluster significativo è rappresentato da termini tecnici come *ram, cpu, floppy, drive, boot,* connessi da frequenze che riflettono discussioni ricorrenti sulle caratteristiche hardware.

3.2.1.4 Visualizzazione mediante Word Cloud

Una word cloud (nuvola di parole) è una rappresentazione grafica delle frequenze dei termini in un corpus testuale. Le parole più ricorrenti appaiono con dimensioni maggiori, facilitando l'individuazione visiva dei concetti predominanti. Questo tipo di visualizzazione viene spesso impiegato per esplorazioni preliminari nei progetti di analisi testuale e lessicometrica.

Il *sottocorpus Computer1* include esclusivamente segmenti testuali in cui il lemma *computer* figura in posizione centrale. L'obiettivo è osservare i co-termini più ricorrenti e le costellazioni concettuali che si organizzano attorno a tale nodo lessicale.

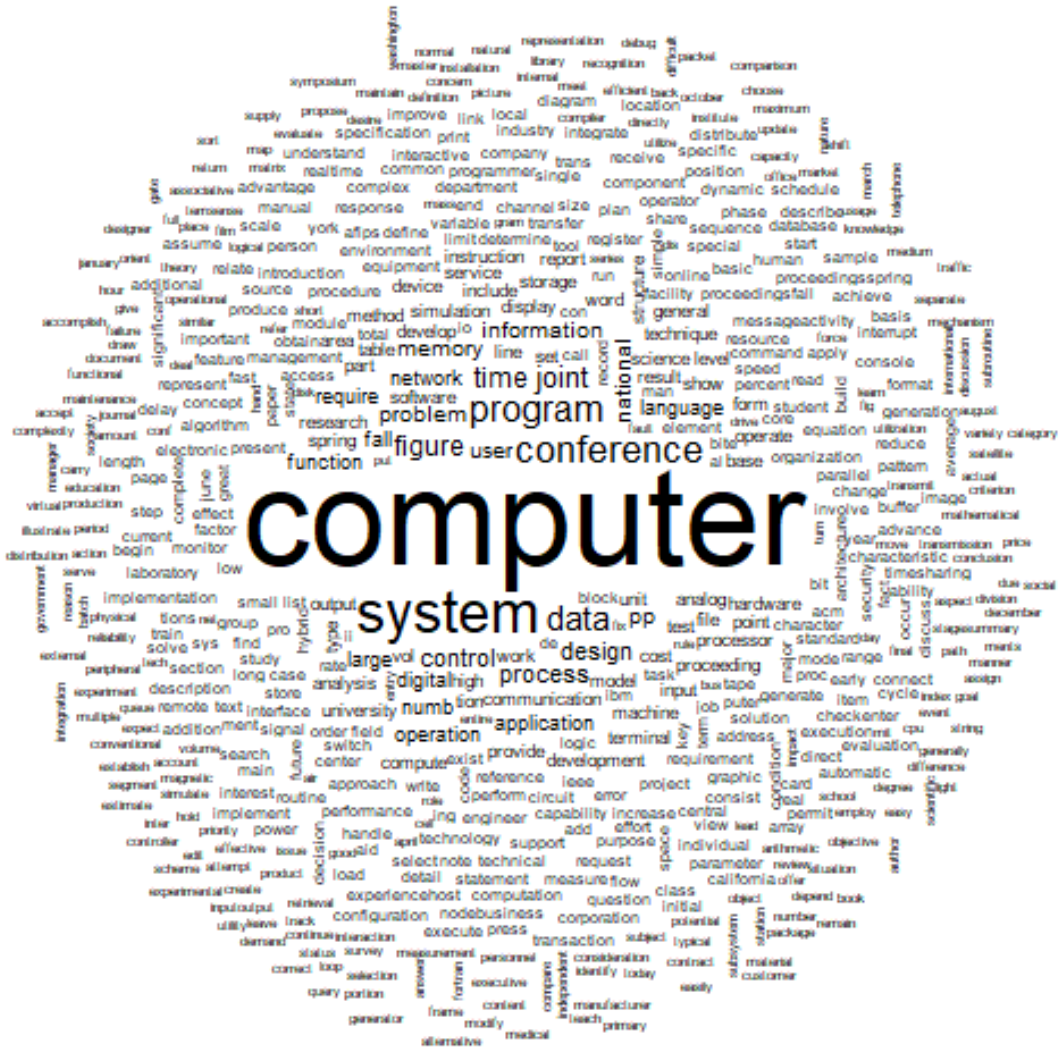


Figure 36 WordCloud del sottocorpus Computer1

Spiccano parole come *system*, *program*, *data*, *conference*, *language*, *network*, a indicare i principali ambiti tematici in cui il termine si inserisce.

L'analisi comparativa tra le due *word cloud* offre una rappresentazione sintetica ma efficace delle trasformazioni lessicali e concettuali nella trattazione del termine *computer* lungo l'arco temporale coperto dal corpus.

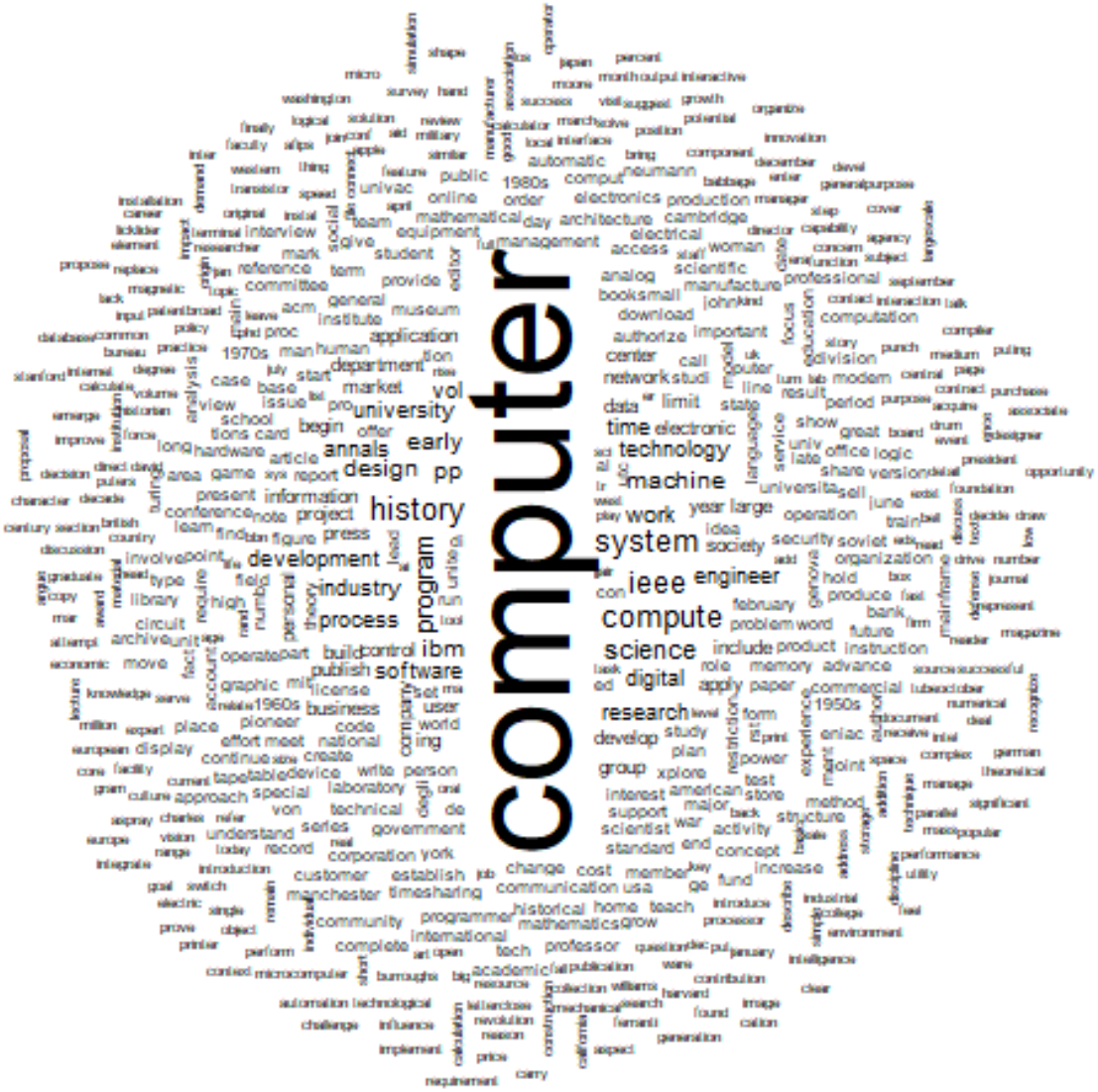


Figure 37 WordCloud del sottocorpus Computer2

Le word cloud generate per entrambi i sottocorpora confermano visivamente questa tendenza: i termini rappresentati, oltre a essere numerosi, rivelano la forte densità di sostantivi specifici del dominio informatico.

Un altro aspetto rilevante è la valorizzazione del contesto storico: parole come *history, historical, early, 1960s, development, museum* suggeriscono un uso più riflessivo e documentario del termine, coerente con l'orientamento retrospettivo della disciplina della *history of computing*.

Il confronto tra i due sottocorpora evidenzia, come abbiamo visto reti di co-occorrenza lessicale, nei grafici sulle analisi di similitudine e nelle word cloud, uno spostamento progressivo della funzione discorsiva del termine computer.

Nel *sottocorpus 1*, computer si colloca in un'area terminologica orientata all'architettura e al funzionamento del sistema *(system, processor, memory, program*), mentre nel sottocorpus 2 il termine si inserisce in contesti storici, documentari e riflessivi (*history, development, museum, early*), assumendo un ruolo interpretativo e culturale.

Il confronto conferma quindi il passaggio dal computer come oggetto tecnico-specialistico da progettare e ottimizzare, al computer come artefatto culturale e storico da documentare, interpretare e contestualizzare.

In altri termini, si passa da una semantica della macchina a una semantica del significato e dell'impatto.

Corpus museale:

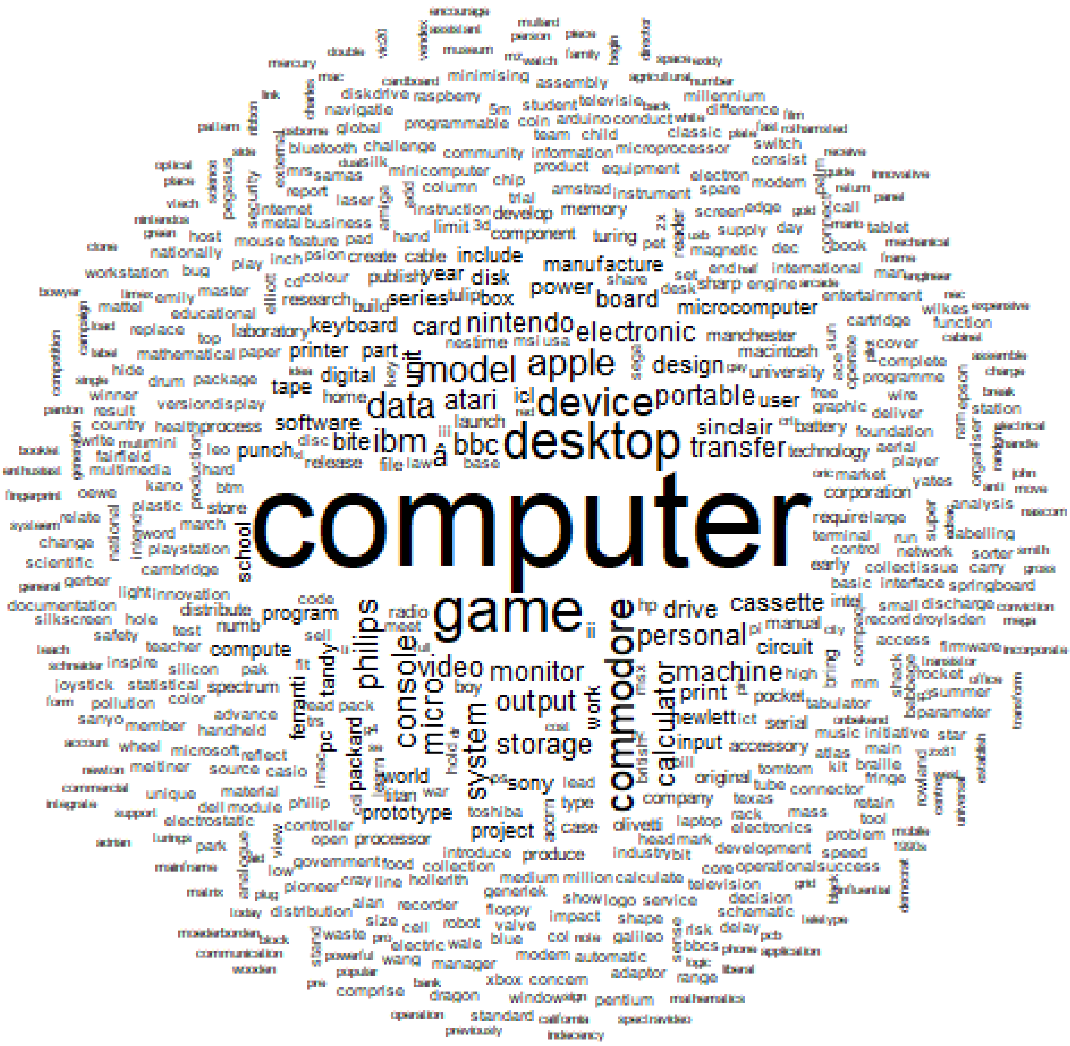


Figure 38 WordCloud del corpus museale

Pur provenendo da contesti istituzionali e culturali, il lessico della word cloud museale risulta fortemente orientato su termini di natura tecnico-materiale. La visualizzazione mostra in modo evidente come i termini centrali siano legati alla descrizione fisica e funzionale dei dispositivi: *computer*, *game*, *system*, *console*, *monitor*, *model*, *storage*, *device*.

Al contrario, ci si sarebbe potuti aspettare una maggiore presenza di vocaboli propri del registro museologico, come i seguenti, che confermano una bassa frequenza, dunque sono marginali all'interno del corpus:

- *exhibit*: 5
- *interactive*: 5
- *museum*: 27

- *original*: 115
- *store*: 106
- *display*: 33
- *archive*: 12

A titolo di confronto, il lemma *computer* compare 3351 volte, e *game* oltre 1300, collocandosi tra le parole più prominenti e visivamente centrali nella word cloud.

3.2.1.5 Analisi di specificità

L'analisi di specificità ha l'obiettivo di identificare parole caratteristiche che risultano statisticamente rilevanti all'interno di categorie del corpus (nel mio caso, suddivisioni per anno o per fonte). È importante sottolineare che la specificità non coincide con la semplice frequenza: un termine può essere poco frequente, ma altamente specifico di un determinato periodo o contesto.

Il calcolo è stato effettuato utilizzando test statistici quali il *chi-quadrato* e, in presenza di frequenze basse, *Fisher-Yates exact test*. L'output consiste in una tabella che riporta, per ciascun termine significativo, la frequenza osservata, il valore *chi-quadrato* e il *p-value* associato. L'AFC è utile per visualizzare relazioni tra termini (o documenti) su una mappa bidimensionale, dove l'asse x e l'asse y rappresentano le componenti principali che spiegano la varianza nei dati.

Tale analisi consente di individuare veri e propri marker linguistici, ossia termini che fungono da indicatori distintivi per specifiche epoche, contesti o comunità discorsive. Inoltre, è utile per segmentare il corpus e isolare tendenze linguistiche o tematiche legate a particolari autori, periodi storici o temi di ricerca, o ancora risorse diverse.

Nel caso specifico, l'analisi è condotta su base conferenziale, considerando una conferenza per volta: per questo motivo non si è ritenuto opportuno suddividere ulteriormente il corpus, ad esempio per autore o tema. L'attenzione si concentra dunque sulle tendenze emergenti **anno per anno**, così da evidenziare eventuali variazioni lessicali e concettuali nel tempo.

Per illustrare concretamente il funzionamento di questo strumento - e per mantenere coerenza con il tema generale della tesi - si prende come caso di studio la diacronia del termine *personal computer*, analizzato nei due sottocorpora computer 1 (1961-1984) e computer 2 (1985–2024).

Questo esempio è emblematico perché mostra come l'analisi di specificità possa essere applicata a qualsiasi termine d'interesse per osservare la sua emergenza, diffusione e ridefinizione nel tempo, restituendo una rappresentazione quantitativa dell'evoluzione terminologica nei testi scientifici, accademici e divulgativi selezionati per la tesi.

Un limite strutturale di IRaMuTeQ, tuttavia, è la mancata gestione delle unità lessicali composte: il software considera ogni parola come un token isolato e non riconosce automaticamente le espressioni multi-parola come un'unica forma lessicale. In assenza di un intervento manuale, termini come *personal computer* verrebbero scomposti in *personal* e *computer*, con conseguente dispersione delle frequenze e perdita di specificità.

Per superare questo problema, il corpus è stato pre-processato sostituendo le parole composte con forme normalizzate unite da underscore (es. *personal_computer*), in modo da conservarne l'unità semantica e permettere al software di trattarle come un'unica voce. Questa soluzione, pur efficace, comporta inevitabilmente un effetto collaterale: le forme così trattate diventano graficamente

“innaturali” e non sempre intercettano le varianti grafiche presenti nei testi originali (ad esempio *personal-computer* o *Personal Computer*).

Ciononostante, questa procedura di normalizzazione rappresenta un passaggio imprescindibile per garantire la coerenza delle misure statistiche e la possibilità di seguire nel tempo l’evoluzione dei termini composti tipici del linguaggio informatico.

Statistiques | **Formes** | Formes banales | Types | Fréquences des formes | Fréquences des types | Fréquences relatives des formes | Fréquences relatives des types | AFC

formes	*year_1961	*year_1962	*year_1963	*year_1964	*year_1965	*year_1966	*year_1967	*year_1968	*year_1969	*year_1970	*year_1971
personal_computer	-0.7924	-1.502	-2.6298	-4.9253	-4.2775	-1.6148	-1.879	-6.9616	-6.4076	-5.3172	-4.8249
personal	-2.6507	-3.1178	-5.5727	-8.2706	-2.3853	-2.3295	-6.2855	-10.3089	-8.3642	-10.0459	-3.6331
person	-5.7574	-7.7278	-15.5432	-2.946	0.5901	-7.5766	-5.4434	-7.201	-1.8994	-13.4938	2.284
persistent	1.1829	2.0997	1.9314	0.3286	-0.9838	-0.5614	-0.4322	-1.6012	-0.2332	-1.2229	0.3411
persistence	1.4665	0.9829	-0.4208	-0.788	-0.2669	1.9426	-0.3006	-1.1138	-1.0252	-0.8507	-0.772
persist	-0.0713	-0.1352	1.003	1.1334	0.6708	-0.2197	1.2529	-0.2282	-0.5767	-0.4785	-0.4342
perry	-0.0951	-0.1802	-0.3156	-0.591	-0.1686	8.0642	-0.2255	-0.8354	-0.7689	0.7743	-0.579
perq	-0.0396	-0.0751	-0.1315	-0.2463	-0.2139	-0.122	-0.0939	-0.3481	-0.3204	-0.2659	-0.2412
perplex	-0.0396	-0.0751	-0.1315	-0.2463	-0.2139	-0.122	-0.0939	4.4766	-0.3204	-0.2659	-0.2412
perpetuate	-0.0555	-0.1051	-0.1841	-0.3448	-0.2994	-0.1709	0.5829	-0.4873	3.6282	-0.3722	-0.3377
perpetual	-0.0436	-0.0826	-0.1446	-0.2709	1.0213	-0.1342	-0.1033	1.3263	-0.3524	-0.2924	-0.2654
perpetrator	-0.2655	-0.5032	-0.881	-1.65	-1.433	-0.8177	-0.6295	-2.3321	-2.1465	-1.7812	-0.5357
perpetrate	-0.0674	-0.1277	-0.2235	-0.4186	-0.3636	-0.2075	-0.1597	-0.5917	-0.5446	-0.452	-0.4101
perpendicular	-0.099	-0.1878	0.275	0.3933	-0.5347	0.8189	-0.2349	1.4058	2.8518	0.7372	-0.6031
perone	-0.0475	-0.0901	-0.1578	-0.2955	-0.2566	-0.1464	-0.1127	-0.4177	-0.3844	-0.319	9.1286
permuter	-0.0753	-0.1427	-0.2498	-0.4679	-0.4064	-0.2319	-0.1785	0.7535	-0.6087	3.2175	-0.4584
permute	-0.0515	-0.0976	2.2176	-0.3201	-0.278	-0.1587	-0.1221	-0.4525	-0.4165	-0.3456	-0.3136
permutation	-0.4477	0.8857	-1.4858	-2.7828	-1.5904	-1.3791	-0.5212	-0.6491	0.3845	2.9697	-0.8703
permit	0.3072	1.8405	6.406	8.966	0.9733	1.0167	3.8695	2.933	-0.4351	0.4541	1.7409
permission	-0.2179	0.6143	-0.7232	-1.3544	0.9072	-0.6712	-0.5167	-1.9144	1.8257	-0.232	0.2391
permissible	-0.206	0.2269	2.3544	3.9796	1.4218	-0.2398	-0.1587	-0.372	1.5284	1.4774	-0.3356
permin	-0.0594	-0.1127	-0.1972	18.8798	-0.3208	-0.1831	-0.1409	-0.5221	-0.4806	-0.3988	-0.3619
permeate	-0.0475	-0.0901	-0.1578	-0.2955	-0.2566	1.3753	-0.1127	-0.4177	-0.3844	-0.319	1.6194
permeability	-0.0792	-0.1502	-0.263	-0.4925	-0.4277	-0.2441	-0.1879	-0.6962	0.7787	-0.5317	18.5563
permanently	-0.2615	0.9811	0.5023	-0.9395	-0.4205	0.956	0.7748	1.1825	0.4837	3.332	0.5456
permanent	-0.6146	-0.9546	9.1129	-0.2914	4.9873	6.3752	-2.8185	-0.5323	-3.1914	0.7404	-0.2639
permalloy	-0.1268	0.9829	0.607	-0.788	2.4216	1.9426	-0.3006	-1.1138	5.1294	-0.8507	-0.772
perma	-0.0951	-0.1802	0.287	-0.591	-0.5133	-0.2929	-0.2255	-0.8354	2.2447	-0.6381	-0.579
perlis	-0.1704	-0.3229	-0.5654	0.6751	2.3823	0.4746	-0.404	0.3716	-1.3776	0.6007	-1.0373
periphery	-0.1268	-0.2403	-0.4208	-0.788	0.704	-0.3905	-0.3006	0.345	-0.4874	1.9817	0.6049
peripheral	34.2045	0.4711	18.596	5.1544	0.7242	-1.074	1.0887	0.7665	-0.4715	0.6064	-0.4794

*year_1972	*year_1973	*year_1974	*year_1975	*year_1976	*year_1977	*year_1980	*year_1981	*year_1982	*year_1983	*year_1984
-10.5553	-1.998	-4.5799	-1.4792	-2.7444	3.2109	1.5563	-0.5305	17.9835	7.6141	68.9244
0.5676	0.5846	1.6924	3.1967	15.1576	21.3822	1.775	1.0684	1.4634	3.3028	9.8635
-5.9909	8.2224	25.211	-1.655	25.5728	30.9543	-2.2009	0.5048	-1.8919	1.5123	1.1918
1.4847	-1.0104	-1.0534	9.5908	-0.6432	-0.9728	-0.6474	-0.4827	-0.5334	-0.3986	-0.3356
8.9691	-0.7029	-0.7328	-0.5853	-0.862	-0.6768	1.0912	-0.3358	-0.3711	-0.2773	-0.2335
-0.4281	-0.3954	-0.4122	-0.3292	-0.1521	-0.3807	-0.2533	2.0713	-0.2087	-0.156	-0.1313
-1.2666	-0.5271	0.9099	-0.439	-0.6465	-0.5076	-0.3378	-0.2519	1.6111	-0.208	-0.1751
-0.5278	-0.2196	-0.229	-0.1829	-0.2694	-0.2115	-0.1407	-0.1049	-0.116	17.0438	-0.073
1.0125	-0.2196	-0.229	-0.1829	0.3352	-0.2115	-0.1407	-0.1049	-0.116	-0.0866	-0.073
0.6771	-0.3075	-0.3206	-0.2561	1.3183	-0.2961	-0.197	-0.1469	-0.1623	0.6131	-0.1021
0.439	-0.2416	-0.2519	-0.2012	-0.2963	1.8766	-0.1548	-0.1154	-0.1276	-0.0953	-0.0802
-3.536	-1.4716	-1.5343	-1.2255	60.4184	1.4022	-0.9429	-0.7031	-0.7769	-0.5805	-0.4888
0.5125	-0.3734	-0.3893	-0.3109	14.3347	-0.3595	-0.2393	-0.1784	-0.1971	-0.1473	-0.124
-1.3194	-0.188	-0.5725	-0.4573	-0.6734	-0.177	-0.3518	0.3435	-0.2899	-0.2166	-0.1824
-0.6333	-0.2636	-0.2748	-0.2195	-0.3232	1.7671	-0.1689	-0.1259	-0.1391	-0.104	-0.0875
-1.0027	-0.4173	4.5618	-0.3475	-0.5118	1.2301	-0.2674	-0.1994	-0.2203	-0.1646	-0.1386
-0.6861	-0.2855	2.4659	1.8328	-0.3502	1.6686	-0.183	-0.1364	-0.1507	-0.1126	-0.0948
0.8353	-0.7293	5.9011	0.4903	-0.3217	-1.0396	-1.5904	1.2651	1.9863	-0.2122	-0.8244
-4.6583	-0.5339	-0.711	-0.8725	-1.6994	-7.9078	-2.7167	-5.3158	-3.6227	-4.2599	-1.5507
-0.6365	-1.2081	0.5051	-1.006	2.8277	-0.2915	2.0967	1.3733	0.7529	-0.4766	-0.4012
-1.8568	-0.5742	-1.1908	-0.9511	-0.2091	-1.0997	-0.7318	-0.5457	-0.603	-0.4506	-0.3794
-0.7916	-0.3295	-0.3435	-0.2744	-0.404	-0.3172	-0.2111	-0.1574	-0.1739	-0.13	-0.1094
0.8231	-0.2636	-0.2748	-0.2195	-0.3232	1.7671	-0.1689	-0.1259	-0.1391	-0.104	-0.0875
-1.0555	-0.4393	-0.458	-0.3658	-0.5387	-0.423	-0.2815	-0.2099	-0.2319	-0.1733	-0.1459
0.467	-1.4497	-0.8511	-1.2072	-1.0599	-1.3958	1.2254	-0.6926	-0.7653	-0.5719	-0.4815
0.557	-2.1229	-2.3009	-1.8371	-0.2532	-0.4838	-1.4383	-0.2409	-3.4788	-0.809	-0.911
-1.6888	-0.278	-0.7328	-0.5853	-0.378	-0.6768	-0.4504	-0.3358	0.2407	-0.2773	-0.2335
4.5811	-0.5271	-0.5496	-0.439	-0.6465	0.9852	1.4	-0.2519	-0.2783	-0.208	-0.1751
1.2877	-0.9445	-0.9847	-0.7865	0.3164	-0.9094	1.3221	-0.1391	0.9835	-0.3726	-0.3137
-0.3065	0.6816	-0.7328	-0.5853	0.5198	0.7136	-0.4504	-0.3358	1.2887	-0.2773	-0.2335
-0.7016	-2.6481	-3.156	-3.2939	-0.4584	-2.7513	-3.3206	-0.3356	-2.333	-2.5132	-5.2377

Figure 39 Analyse de spécificités del sottocorpus computer 1 (1961–1984)

L'analisi del termine *personal_computer* nei due sottocorpora rivela una traiettoria terminologica significativa.

Nel *sottocorpus 1* (1961–1984), i valori di specificità risultano negativi per tutti gli anni Sessanta e Settanta, segnalando la marginalità del concetto nel lessico tecnico-scientifico dell'epoca. A partire dal 1977, tuttavia, si osserva una prima crescita dei valori, che culmina con un picco tra il 1982 e il 1984 (rispettivamente +17.98, +7.61 e +68.92), in corrispondenza della diffusione dei primi computer destinati all'uso personale (Apple II, Commodore PET, IBM PC), come confermeranno gli estratti testuali analizzati più avanti.

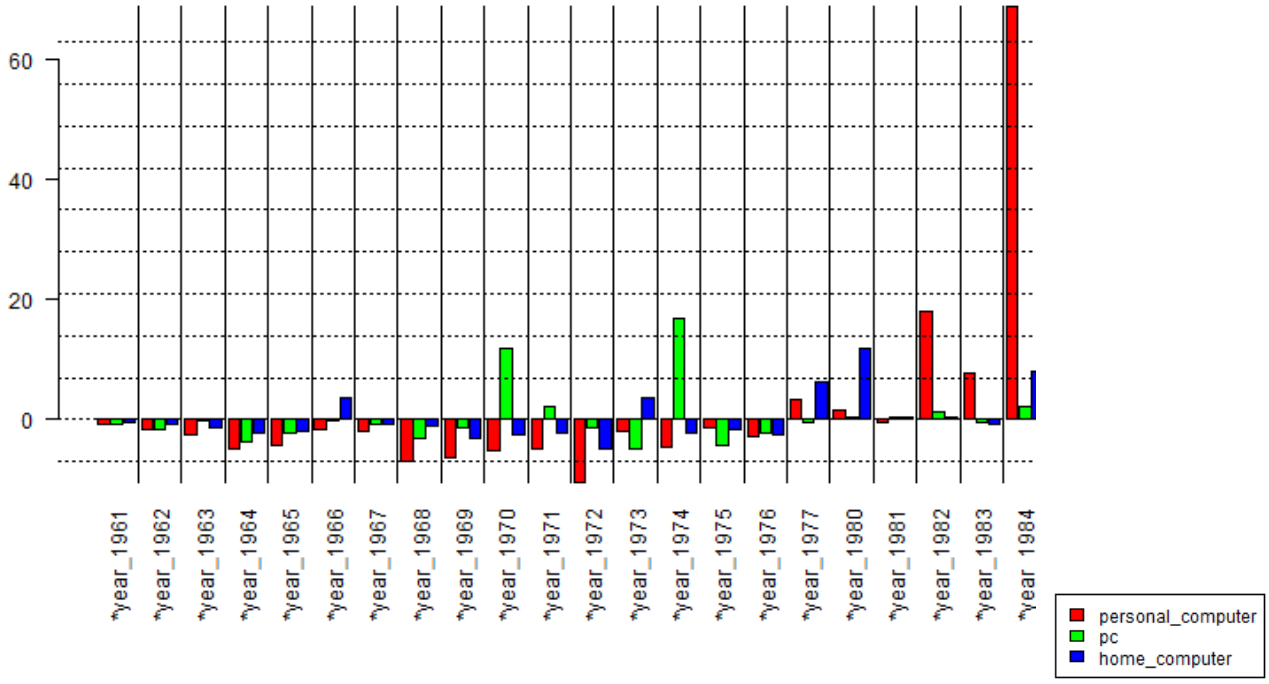


Figure 40 Andamento dei valori di specificità dei termini personal_computer, pc e home_computer nel sottocorpus 1 (1961–1984).

Questa fase coincide con l'emersione dei personal computer e delle sue prime varianti (*home computer, pc),* come categoria discorsiva centrale, indicatrice di un cambiamento di paradigma: dall'informatica come dominio sperimentale e aziendale a una tecnologia accessibile, individuale e domestica.

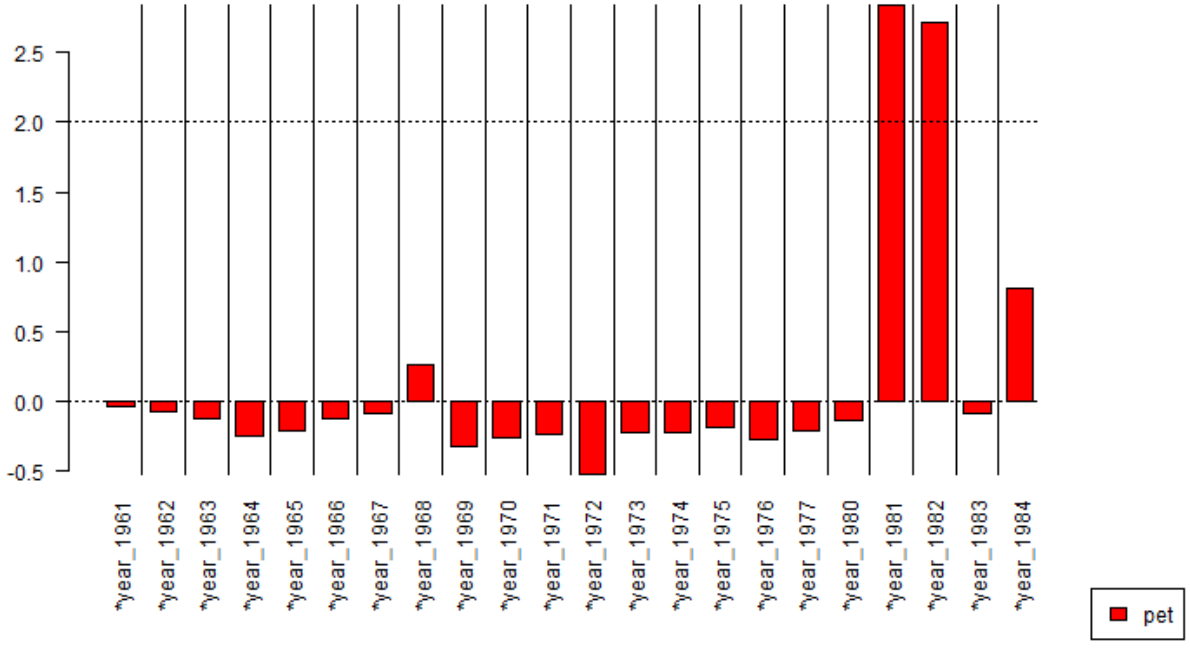


Figure 41 Andamento dei valori di specificità del termine pet nel sottocorpus 1 (1961–1984).

Il termine *pet* mostra i primi picchi di specificità positivi già dal 1981, segnalando l'impatto lessicale del Commodore PET, tra i primi modelli a incarnare la nozione stessa di "personal computer" accessibile:

- ***** *source_AFIPS_Joint_Computer_Conferences *year_1981 *id_5352_573621*

Many home computers provide graphics-oriented character sets; two examples of this are the Commodore PET; [29]

- ***** *source_AFIPS_Joint_Computer_Conferences *year_1981 *id_2545_274184*

Many home computers provide graphics-oriented character sets; two examples of this are the Commodore PET and the Radio Shack.[30]

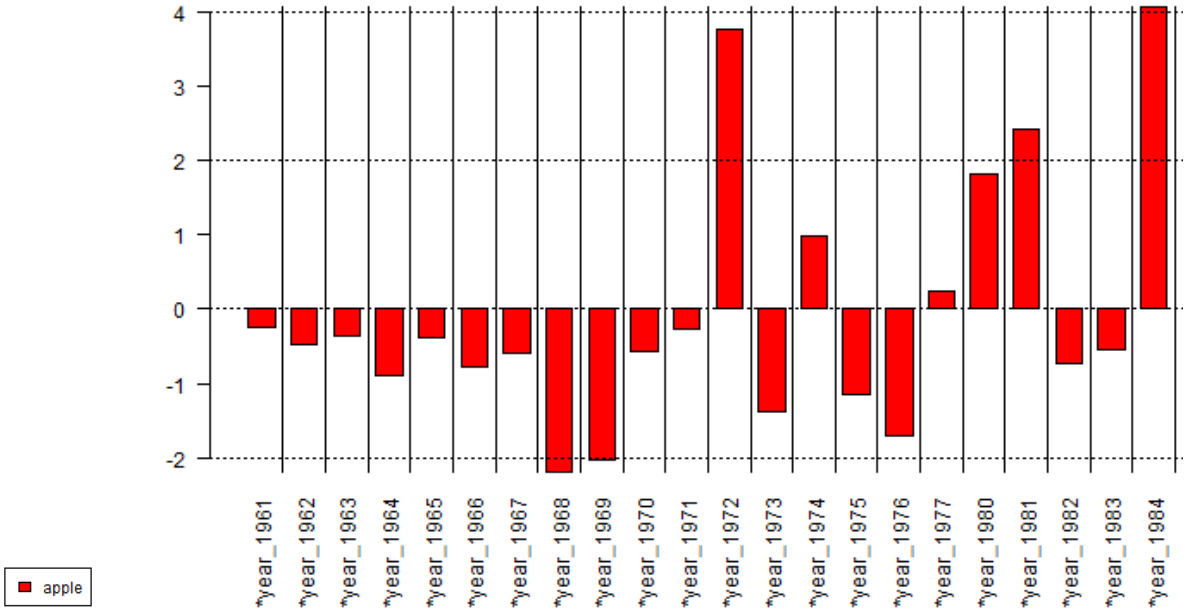


Figure 42 Andamento dei valori di specificità del termine apple nel sottocorpus 1 (1961–1984).

In parallelo, il termine *apple* rimane in negativo fino alla fine degli anni Settanta, ma registra una crescita dal 1970 in poi, con un marcato picco proprio nel 1984:

- ***** *source_AFIPS_Joint_Computer_Conferences, year_1980, id_5236_561738*

..to Apple II computers at less than each approaches used to provide them and all of these configurations can run virtually the same hardware and software;

[29] La frase è significativa non solo per la menzione diretta del marchio, ma per l'uso esplicito della denominazione "home computers", che inquadra il PET come prototipo di computer domestico.
[30] Variante pressoché identica al primo, ma proveniente da un documento diverso.

- ***** *source_AFIPS_Joint_Computer_Conferences, year_1984, id_5576_595027*

By Edward Birss, Apple Computer Inc., Cupertino, California. Abstract: in Apple began to develop Lisa workstation to enhance the productivity of office workers.

Statistiques | **Formes** | Formes banales | Types | Fréquences des formes | Fréquences des types | Fréquences relatives des formes | Fréquences relatives des types | AFC

formes	*year...	*year_1986	*year_1987	*year_1988	*year_1989	*year_1990	*year_1991	*year_1992	*year_1993	*year_1994	*year_1995
personal_computer	-2.6466	-5.0294	-3.2623	-0.3841	-1.9649	-1.2168	-0.8994	-0.8528	-2.8501	-5.9616	-1.4299
however	-2.6635	-1.1884	-0.4216	-0.4065	-0.7786	-0.4831	0.4208	-0.2627	1.3331	1.3174	0.5552
online	-2.6695	-3.9584	-2.3501	-5.4929	-1.9851	-1.7894	-0.9113	4.1845	-2.8779	-5.0554	-0.5816
apple	-2.6825	-3.7273	-2.4177	-1.3303	-1.8194	-1.9439	0.3025	-1.1141	-1.8651	-5.3459	0.3688
babbage	-2.6889	0.6262	8.4209	25.1045	0.5627	-0.7537	3.4856	-0.278	-1.0041	0.6201	-0.4836
punch	-2.7339	-0.4395	-1.6371	-1.3788	-0.3903	0.3229	2.9225	-1.1355	-1.4183	1.1261	-0.9092
develop	-2.7431	-0.3377	0.5487	-3.167	-0.4427	-1.4292	-1.6373	0.5289	3.374	-1.0192	0.4381
customer	-2.7442	-1.6584	-1.2906	0.3938	0.6272	-1.2741	0.328	-0.888	-0.3445	0.2842	3.9649
calculator	-2.7595	8.6453	-1.6568	1.2476	2.6953	1.0356	2.0058	1.2946	-0.5386	-2.8484	8.1237
library	-2.7614	-0.7927	-1.3022	-1.9544	-0.4228	0.5434	-1.3771	-0.8943	-4.8725	-3.2101	-3.8948
their	-2.8424	0.7699	-2.679	-0.8431	-2.8251	-2.2894	-0.879	-0.314	-1.3853	-0.9516	-0.3041
account	-2.9056	-2.2606	-0.2331	-2.5892	-1.2472	-0.9527	-2.0688	-0.9466	0.3089	-0.6447	1.6507
rand	-2.997	-1.408	-0.2384	-1.6339	-1.4866	-2.1718	-0.6273	-1.2448	-2.1886	-1.4167	0.4493
card	-3.0097	0.6738	-1.9765	-1.8187	-0.7237	0.2995	2.6544	-1.6713	-1.0432	1.1441	-0.6461
firm	-3.0548	-1.1057	-1.885	-1.3085	0.4018	-0.9297	-1.0004	-0.6743	11.2739	1.3944	-2.3367
focus	-3.0793	-2.4518	-0.2764	-2.3077	-4.0933	-1.4734	-0.7941	-1.01	-1.3917	-5.8669	-0.865
archive	-3.196	-1.5856	-0.6734	0.2699	9.183	-0.9999	-0.7111	-0.385	-0.5705	0.3208	-1.6565
download	-3.2046	-4.2444	-5.4919	-2.6705	-3.1943	-0.5087	3.0829	-0.2334	2.4439	-1.9078	-1.3922
soviet	-3.2538	-1.7609	-3.8698	0.2947	-2.5052	-0.7948	-2.3251	-1.0741	-5.5819	4.0303	-7.7913
bank	-3.2538	-2.1068	-2.9326	-1.1551	-0.6054	-1.5477	-1.6408	1.8556	9.3143	-2.0737	9.8202
woman	-3.3244	-1.3189	-2.096	-3.9067	-3.3181	-2.4091	4.5176	-0.7586	-3.2784	-2.6083	-3.9755
fund	-3.3372	2.2467	0.2733	-0.536	0.6283	-2.4184	-2.5312	-0.4168	1.3617	4.4876	-0.4197
press	-3.5451	-1.7294	-1.3487	0.995	0.3949	-2.4967	0.3054	-1.2315	-0.762	-1.7691	0.4512
usa	-3.5619	-4.9492	-3.2103	-2.2146	-0.2403	-1.1867	-1.2736	-1.4794	-3.5599	-7.0983	-5.2569
graphic	-3.6284	-2.1142	-2.4686	-2.4699	-1.7571	-3.4092	-1.4314	-1.9539	-6.116	2.2316	-8.5368
1980s	-3.6582	-5.083	-1.7195	-1.4923	-3.6512	-1.7974	-0.6214	-1.5194	-2.2826	-5.0781	-5.4208
company	-3.7632	-3.6247	-8.0266	-13.3389	-0.5675	-2.8286	-2.4323	0.8799	-0.4009	-3.5571	2.4526
between	-3.8009	-1.2229	-1.3762	0.4298	-2.1057	-0.3249	-0.3942	-0.606	-0.8064	-1.6593	0.9589
p	-3.835	-3.5647	-2.3817	0.323	-0.6943	-3.3125	6.0507	-5.8117	-4.5919	-7.6123	-5.0398
service	-3.8527	-2.9765	-0.6531	-2.1247	0.5525	0.5583	-0.6176	2.1525	-5.6831	1.883	0.9054
society	-3.9824	13.9155	-2.4952	-3.4039	-2.8355	2.9876	-4.2955	-0.8378	-6.5903	-8.6494	-1.2368

*year_1996	*year_1997	*year_1998	*year_1999	*year_2000	*year_2001	*year_2002	*year_2003	*year_2004	*year_2005	*year_2006	*year_2007	*year_2008	*year_2009
-2.2773	-0.8101	-0.1582	-1.8362	-3.1957	-3.6206	-3.3413	-2.5765	-1.6037	-0.7174	10.2578	1.0302	-0.8272	-2.4165
-0.9076	1.5619	-0.4074	-1.7973	0.4374	-0.2983	-0.3484	-0.8787	1.0451	1.4518	1.4041	-1.3867	0.6675	-0.3534
-2.3085	-2.9078	-0.16	-1.5112	-3.9292	-3.6675	13.4712	0.6662	-1.6325	-2.4612	-1.3057	-1.8787	-0.3355	-0.7283
-1.9834	-2.2032	0.2505	-4.5176	-1.19	-4.849	-2.9156	-4.6948	-0.3944	-2.6263	3.9998	0.7721	-1.8921	-0.7834
1.8931	0.7429	-0.359	-3.4654	2.8471	1.2803	-0.3367	0.6361	0.8551	1.2591	-0.3752	-2.4684	-1.5449	-1.0269
-0.4192	3.5449	-0.365	-0.9985	1.8018	33.6549	2.9806	-0.6733	6.5961	0.9945	-1.4275	-1.2521	-0.8176	0.8114
0.432	-0.4542	-0.2583	8.3283	-1.487	-0.4926	0.897	-0.4192	-0.3715	1.7993	1.149	6.0735	-1.626	1.135
11.605	0.6495	0.9642	-1.5893	-0.3251	-0.3346	4.4946	-0.4425	3.0796	-1.2162	-1.151	0.4986	3.1004	0.8174
2.363	8.0659	-0.3684	-1.0205	1.2412	0.329	-1.7517	-1.5599	3.6829	-0.3563	-0.4152	-0.992	-1.3253	-0.8466
-4.8591	-5.2939	-0.5004	1.6983	-0.9687	-2.5604	200.2787	4.0613	-0.4161	0.6502	-2.359	0.7591	-0.7606	-1.4932
0.7647	-0.9287	-0.6271	-3.0662	-0.7525	1.8393	-0.8444	-0.3728	0.3268	-2.2801	-0.9345	0.6883	2.1313	-2.2507
4.7036	0.5449	-0.1788	-1.7615	-0.395	9.9653	0.9676	-0.5577	1.1851	-1.1328	-0.7345	-0.3834	0.6201	-2.3577
1.9657	-0.3176	0.629	-3.134	-1.4924	5.2709	-2.0029	-1.1294	10.874	-2.5377	-1.7504	0.6875	11.9481	-1.6609
-0.2614	2.7726	-0.5373	-2.2641	0.9117	27.3575	7.7477	-1.078	8.8922	4.1546	-1.4339	-1.9343	-2.6251	1.8839
1.8787	3.9531	-0.4079	-0.7632	-3.1025	5.8363	2.108	0.3602	13.3281	5.4829	0.8814	-2.4397	4.0322	-1.3896
-0.3779	-0.2607	-0.5476	-1.0687	-3.2104	-0.6573	-0.3002	-1.3475	-3.3791	3.9921	1.2246	-2.3962	1.3585	-2.5947
-1.4356	-4.4008	-0.4267	-2.7594	1.7277	0.4624	-0.2987	-1.1489	-1.1547	-0.9931	-1.3625	0.5506	1.2875	-0.9862
-1.5706	-0.6203	-0.4767	1.6438	0.5343	-0.3249	0.3669	2.5554	2.0479	1.1774	0.4817	1.7311	5.2452	2.7935
-8.0674	-6.185	-0.5733	40.0824	-2.4367	-6.1478	-6.2761	-10.9612	-7.1266	-6.0212	27.3381	55.6918	-1.2125	-8.3225
-0.4346	-2.9648	-0.4344	-2.8353	-2.1913	-0.5737	-1.3989	-1.996	14.4902	-0.3708	-4.1692	-0.3104	-1.0947	0.7146
31.5296	-1.6386	-0.4439	-4.4507	-1.8266	-1.3776	-2.9775	32.1051	-2.6711	-3.6155	-5.079	-1.8916	0.4206	-4.3326
1.9918	-0.4758	-0.4456	-0.4276	13.7394	-3.1134	-0.6587	1.2924	-2.6903	-2.0749	-2.615	-0.3456	0.7917	-1.1085
0.9099	2.6389	-0.7979	-7.6748	-2.6304	-0.9076	-2.5402	0.2869	0.4425	-1.9503	0.4318	-1.328	2.0984	0.5171
-4.5509	-3.3903	-0.4756	-3.2454	-4.6762	-7.9813	-4.0885	-7.817	-6.6422	-6.9219	-7.8134	-6.6317	-6.9239	-3.9784
-3.6772	-4.3309	21.9967	-1.0535	-3.3939	5.9956	-0.6704	-1.6136	-3.8208	-1.5843	-2.4947	-2.6957	23.679	1.7254
-0.2964	0.3232	0.9825	0.6152	-3.2461	-2.424	-0.3566	-10.6008	-1.9586	-0.5847	0.6779	0.826	0.3451	1.2334
-0.3568	-0.9909	0.5039	-4.3369	-1.1114	1.6392	2.2764	1.4471	-0.3366	1.6476	-2.1318	-2.3938	7.9986	0.4977
1.0268	0.4495	-1.2008	-2.8427	-1.2871	-1.6316	-4.6975	2.5594	-1.1933	-0.3606	-1.2495	-1.2692	0.5269	0.3069
-4.7194	20.6469	-0.9678	-10.8979	-9.3517	-4.0993	-9.2848	0.5952	-2.5465	-5.5235	-0.4149	-3.1698	0.3796	-1.59
-0.4913	-3.3861	-0.2282	-3.1637	-2.0555	-3.982	12.7224	-0.4568	14.8117	2.8644	-1.7317	0.5044	10.7773	-1.6118
-3.1905	-4.9355	-1.0086	-4.2494	-4.612	-4.08	-4.7684	1.156	0.7514	8.988	-0.8869	-2.6393	0.5115	2.5627
-4.9696	-0.7479	-1.2941	-4.3763	-4.3347	-8.0196	-2.3794	-0.3259	0.3071	0.5604	-1.3003	0.6207	2.8013	1.1341
-2.8383	2.1556	-0.5433	-2.313	-0.3505	-1.0967	-0.2902	1.3649	0.3113	-0.5412	-0.8497	-0.9147	2.7291	-1.8199
-8.1156	-7.4938	-0.5767	-7.2741	-7.0832	-11.3284	-6.3136	-12.5168	-9.6212	3.8728	4.075	7.6896	12.6644	8.0387
2.8785	-1.085	-0.3406	-2.0291	-2.6042	-3.8854	4.5915	-4.8424	-0.369	-0.6155	-0.329	-1.4514	8.3171	1.3017
-0.3297	-0.4262	0.2689	-1.9556	2.7537	0.4274	0.6134	0.8097	-0.5431	-0.318	-0.5019	-0.594	1.7216	-0.3133
0.9983	-0.311	-0.4945	-3.0206	2.7452	-0.3812	26.2163	-6.5309	1.91	1.3768	-1.8738	-0.6156	20.1521	5.9691
0.7682	-0.395	-0.9405	-1.8732	-2.3484	3.1266	-0.523	0.4967	-2.7825	-0.7087	0.3096	-0.4374	-1.2771	1.0661

*year_2010	*year_2011	*year_2012	*year_2013	*year_2014	*year_2015	*year_2016	*year_2017	*year_2018	*year_2019	*year_2020	*year_2021	*year_2022	*year_2023	*year_2024
1.1527	3.9137	-1.8632	-0.2999	0.3813	0.317	1.7713	13.4016	1.112	0.9231	16.3041	-1.1955	4.0045	-1.6316	0.8328
0.976	1.4537	0.3649	0.43	1.1744	-0.6416	-0.3916	-0.2688	0.8986	-0.2707	-1.887	-0.8057	-0.785	-1.2099	-0.2912
-3.4027	-1.41	-0.3403	0.6286	-3.136	-1.9123	0.527	-2.0035	-0.7212	4.6183	3.8864	4.717	4.3443	27.3049	16.0013
-2.8311	41.1069	-0.9715	14.9902	-2.0546	-2.0004	-0.7608	-0.5737	7.4429	7.1807	0.7269	-4.1068	0.5624	0.9908	2.1991
-2.8422	-1.3757	-0.5383	-1.2679	-3.0159	0.6918	-0.5846	-3.2299	-1.03	-2.0708	-0.9145	2.144	-2.8884	-3.9564	-0.9493
-1.9843	-1.4316	-1.3268	-0.439	-2.5763	-2.5152	-4.6054	-2.1276	-0.33	0.2982	-0.7384	-1.4202	0.383	-1.6486	-1.6234
2.6479	-0.5148	-0.5063	0.2908	-2.0072	1.0335	-0.2925	-3.7644	0.656	0.9332	-5.823	-0.9503	-0.4517	-1.0897	0.7807
-0.6127	-1.261	0.4342	2.4008	-0.4536	-0.4266	-1.8053	-0.952	-1.0317	-1.0265	-6.3613	-3.6977	-0.2656	-2.7705	0.6081
-2.0215	-2.6201	0.9263	0.2743	-2.6194	-4.4301	-1.0565	3.2382	-1.5211	-2.8496	-0.4328	-1.8887	-1.3003	-1.1254	-1.3016
-1.6372	-0.5825	-1.9979	-1.7336	-5.769	-2.0439	-3.0139	-0.4759	0.4629	-2.1164	-2.0376	0.4814	-1.4421	-0.4861	-2.8997
0.7953	2.5786	1.552	0.4623	2.0222	-0.6719	-0.6767	2.1616	-1.2076	4.1088	-0.5186	3.1143	0.3991	3.0906	-1.778
-1.5461	0.7561	1.489	-0.5436	-0.8428	0.5495	1.2212	-0.71	-1.1406	1.1068	-0.7116	1.2228	0.6497	-1.4637	0.3008
-1.0827	3.5291	5.9739	4.1971	-2.1307	-0.4514	8.0206	-1.6283	-1.2487	-0.7945	-4.0658	-0.5495	-2.5447	-3.9338	-2.3958
-5.2838	-1.5738	-2.7525	-0.368	-1.8428	-1.7789	-5.307	-3.3986	0.365	-0.5718	0.2985	-0.9764	0.9132	-1.152	-2.7768
4.7122	-2.2168	2.0561	-0.8267	-7.3588	-3.0538	-4.3313	-0.8462	-1.2908	-4.2592	-6.2564	-4.6768	-4.4202	-6.533	-0.7941
-0.375	1.5334	1.1025	2.9152	1.0183	1.2354	0.6074	0.7233	0.4708	0.9633	2.9813	2.2334	0.861	2.2255	1.1129
-0.5324	0.7987	-0.3737	1.6247	0.5293	-1.3724	-0.7885	5.1797	-0.3497	-0.3592	22.9666	-0.2498	-1.706	-1.1458	-1.1446
-17.2898	0.306	-12.6309	-4.4549	-4.3373	1.0496	2.4846	0.637	-0.3053	2.6775	0.4427	-0.3211	0.9426	-0.8129	3.8008
3.628	-7.8411	-0.3866	-5.0426	-7.8396	1.2016	28.4873	-6.9366	-4.7267	-1.1439	-7.46	-0.4382	-5.0236	0.2888	-4.9459
-1.9637	-0.4019	11.8353	-4.1163	-2.9635	0.4094	4.2391	-1.0092	0.5156	-0.5323	1.7137	6.6055	1.3507	-5.168	-4.0412
0.7893	-6.7133	-5.6392	-0.2751	0.9908	-3.0067	0.3733	32.5164	1.3509	-0.767	3.2742	-2.5344	-0.8441	0.4305	-2.3455
5.3675	-0.5719	-0.439	-0.2799	-1.8955	0.2778	0.475	-0.5173	-2.0126	-1.0262	-0.3838	-2.0341	0.6643	0.785	-0.6238
0.317	0.9421	1.1096	1.5473	2.6069	-0.4632	0.7311	0.9036	1.7916	6.8694	0.8729	0.3823	-0.5805	-0.6942	-0.9708
-5.953	-7.257	-6.0421	-4.6294	-8.5803	-3.9454	-5.2755	-5.6084	0.9241	45.082	19.7402	39.4506	32.4661	48.4319	31.8325
-7.4274	-3.925	-6.6878	1.2309	7.1266	-4.8374	2.5783	-5.0322	23.3158	2.0013	-0.5031	-3.5776	-1.0292	-0.514	30.4284
-0.2677	2.4151	-0.2527	3.1035	0.9783	6.8162	5.0035	0.2767	-0.9899	0.8954	0.5845	1.2353	3.5574	0.7777	2.3167
1.428	-4.2704	4.1962	4.6769	-0.8136	-0.2934	-0.4401	2.2923	0.5075	1.0125	1.1268	-2.3453	2.3718	-4.2676	7.4872
0.7536	0.7665	-1.8645	-1.0822	3.1946	1.0341	3.2656	0.7159	0.6969	1.1488	0.7397	4.1786	-0.6481	1.2564	0.9152
0.5026	12.1867	28.2655	4.1614	0.9923	0.7289	3.7525	8.9049	-8.7666	-0.8125	0.4818	5.3133	-0.7111	4.0636	0.4614
-2.5838	-4.1212	1.2076	-0.4411	-4.4809	-0.4219	-3.2316	5.878	-0.8269	-1.2605	-0.7674	-0.2773	0.2862	3.3213	-0.9448
-0.5284	1.2413	1.0462	2.7551	-0.7745	1.0493	1.1674	0.4153	0.7873	2.48	3.121	0.6671	4.6613	1.9539	0.9826
7.9461	1.6829	3.8177	3.5089	1.1874	0.9494	1.2931	-0.6702	3.3253	-3.1023	-1.1318	1.5196	1.9739	1.4134	17.3706
0.6187	1.5737	1.1299	0.507	2.4451	1.1023	0.6277	-1.4841	-2.6638	0.4999	0.5784	0.7364	0.8824	5.0789	0.6649
-10.0295	3.002	-7.327	-1.1744	-1.1944	4.8833	9.0562	3.3544	1.3062	7.0027	2.6172	1.4925	3.1536	2.1006	11.2971
1.0833	0.6203	0.6734	9.6174	0.4763	3.1566	-1.0806	1.103	-0.5489	1.2828	-1.6824	-3.6853	0.9699	0.8511	-2.0901
0.6067	2.5818	-0.3002	-0.9394	1.9888	-0.7444	-1.5645	2.8155	-0.4147	0.6087	0.8108	-0.8828	-1.0789	-1.0354	0.3471
2.4258	-7.802	1.7246	5.7961	-4.9231	0.576	-6.3539	-2.684	-0.8122	-2.4449	-7.3771	-8.2077	-5.5029	-16.0433	-1.3031
1.9393	-2.0326	0.4151	-1.2178	1.6037	-1.3537	0.574	8.8178	0.4586	-0.8152	-1.6581	2.2527	-1.6159	4.4301	-0.8559

Figure 43 Analyse de spécificités del sottocorpus computer 2.

Nel *sottocorpus 2* (1985–2024), il termine mantiene valori elevati per tutto il decennio 1985–1995, coerenti con la sua consolidazione come etichetta classificatoria nei testi accademici e divulgativi, per distinguere la categoria del personal computer rispetto ai mainframe e ai sistemi aziendali.

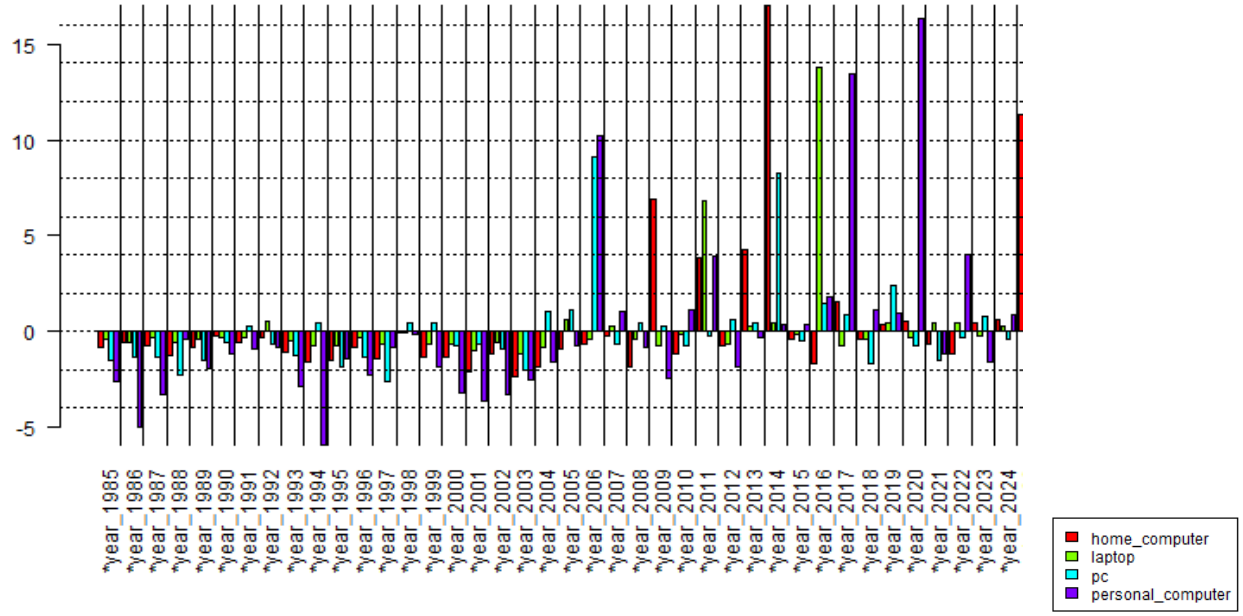


Figure 44 Analisi diacronica delle varianti terminologiche (pc, laptop, home_computer e personal_computer) nel sottocorpus 2 (1985–2024).

Tra la fine degli anni Novanta e i primi Duemila si registra invece un calo progressivo, dovuto all'adozione generalizzata di forme (*PC, laptop, home computer*), che non sostituiscono completamente la denominazione estesa, ma ne ridefiniscono l'uso nei diversi contesti comunicativi. *Pc* si afferma come abbreviazione economica e neutra nel linguaggio tecnico e quotidiano. *Personal_computer* e *pc* non seguono semplicemente traiettorie parallele ma dialogano diacronicamente: il calo dell'uno coincide con l'ascesa dell'altro, fino a quando, nell'ultimo decennio, entrambi riemergono con funzioni diverse, l'uno come termine di riferimento storico, l'altro come abbreviazione funzionale nel linguaggio tecnico e quotidiano. La forma *pc*, infatti, registra i valori più alti di specificità proprio a partire dai primi anni Duemila, con picchi intorno al 2006 e al 2014.
Laptop emerge in riferimento alla portabilità e alla riorganizzazione spaziale del lavoro e dello studio e *home_computer* permane soprattutto in contesti storici e divulgativi, legati alla memoria dei primi modelli domestici.
Questa divergenza funzionale tra le varianti sarà analizzata più nel dettaglio a p. 230, attraverso l'uso di Sketch Engine e del confronto tra contesti d'impiego.

A seguire il risultato dell'analisi della specificità sul corpus *Reddit*:

formes	*year...	*year_2014	*year_2015	*year_2016	*year_2017	*year_2018	*year_2019	*year_2020	*year_2021	*year_2022	*year_2023	*year_2024
occurences	1617	6609	5948	8346	9473	6693	5295	13392	17135	6940	7964	6642
formes	612	1325	1552	2084	2070	1884	1624	2939	2746	1991	1865	1725
hapax	87	142	268	531	315	472	298	825	380	474	313	317
segments	47	186	169	236	277	194	153	379	489	200	230	189
hapax/formes	0.14216	0.10717	0.17268	0.2548	0.15217	0.25053	0.1835	0.28071	0.13838	0.23807	0.16783	0.18377

Figure 45 Statistiche del corpus Reddit

ormes	*year...	*year_2014	*year_2015	*year_2016	*year_2017	*year_2018	*year_2019	*year_2020	*year_2021	*year_2022	*year_2023	*year_2024
personal_computer	0.3293	0.943	-0.2247	-1.4153	0.5916	1.9741	0.8712	-0.9397	-3.0823	0.5868	-0.3851	0.6109
key	0.3208	-0.3046	-0.2376	-1.4547	-1.6103	-1.1516	0.8396	-2.3524	4.1019	0.5603	0.4277	0.3123
real	0.3208	-0.3046	-0.2376	-0.434	-0.2798	0.3223	-0.4113	-0.6212	0.4706	0.5603	0.4277	0.5839
until	0.3208	-0.3046	-1.0566	0.3982	-0.5203	-0.5776	-0.1765	-0.6212	1.6245	1.8186	-0.7607	-0.29
short	0.3208	0.2975	-0.2376	0.6718	-0.2798	-0.2808	-0.4113	-0.6212	0.2996	-0.3045	1.996	-1.1712
thing	0.3177	0.2912	1.0299	0.5106	-0.4844	-0.7344	-0.2497	-0.4641	-0.4828	-0.3108	-0.3924	1.2228
bring	0.3126	-0.6374	2.0332	2.3611	0.3043	0.9236	-0.428	-1.0273	-0.7814	-1.2339	-0.2168	-0.6147
late	0.3126	-0.6374	-0.5309	0.3769	-0.5459	-0.2959	0.4569	-1.0273	0.6405	0.8688	-0.4263	1.3211
three	0.3126	-1.2342	-1.0852	-0.2365	-0.5459	-0.2959	-0.1866	2.977	-0.5184	0.8688	-0.4263	0.2965
can_t	0.3126	0.5349	-0.5309	0.3769	0.5317	-0.6002	-0.428	0.5954	-0.5184	-0.3207	0.6821	-0.3056
wish	0.3126	0.5349	1.0392	0.3769	0.3043	-0.6002	0.8094	-0.656	-0.5184	0.2822	-0.7896	0.2965
man	0.2971	-1.2991	-0.5718	0.581	-0.5983	0.2766	0.7527	0.5249	0.7816	0.2539	-0.4687	-0.3375
next	0.2971	0.488	-0.5718	-0.2672	0.478	-0.6459	-0.9846	-0.2736	1.0534	0.4881	0.3648	-0.6614
getting	0.2971	-0.6856	0.3258	-0.5009	0.478	0.5252	-0.2074	-0.2736	-0.593	0.4881	-0.4687	1.229
inside	0.2897	1.1442	-1.1709	-0.5238	-1.0583	0.8211	-0.2182	0.4927	-0.2641	0.2408	0.5931	-0.3539
always	0.2826	0.7384	0.8983	-0.547	-1.0933	-0.359	3.4053	-0.5143	-1.3658	0.7386	-0.5123	0.2412
plus	0.2826	0.7384	-0.6134	-0.954	0.4297	-0.6922	-0.496	-0.316	-0.2879	-0.1928	3.2931	-0.7087
oh	0.2826	-0.7345	0.8983	0.819	-0.3739	-0.359	0.7005	0.6876	-0.2879	-0.7343	-0.907	0.2412
design	0.2757	-0.7591	0.535	0.5029	-0.3939	0.7597	-0.2402	-0.5438	1.4722	-0.7589	-0.5345	-0.3873
about	0.2754	0.2891	0.4943	-0.3279	-0.9834	0.4604	-0.6401	-0.9852	1.1338	0.7857	0.5437	-0.6445
check	0.2691	-0.4237	-1.2566	-0.5943	-0.7067	-0.7391	0.6523	1.1662	-0.3384	1.4369	0.516	-0.2035
almost	0.2691	-1.4291	-0.6554	0.2712	-0.4143	-0.392	-0.5308	-0.574	1.3952	-0.4236	2.023	0.427
show	0.2691	-0.2162	0.8359	1.0855	-1.1638	0.4407	-0.2515	-0.8764	-0.5145	-0.2161	-0.3068	1.0658
fun	0.2626	-0.4417	1.6447	-0.3504	-0.7345	1.4724	0.6296	-0.2331	-2.0908	-0.4416	0.7695	-0.2149
school	0.2564	-0.4599	-0.698	-1.8086	0.3471	0.4039	0.9682	0.5457	0.5207	-0.4598	-0.3395	0.6576
enough	0.2564	-0.4599	-0.3638	-1.0794	-0.4562	0.6767	0.3291	-0.4088	1.5744	-0.8336	1.8967	-1.4562
really	0.2459	1.0298	2.8317	-0.5784	-1.019	0.8761	0.4041	-1.5087	-2.8687	0.7914	-0.2414	1.0905
set	0.2445	-0.4971	1.9893	-1.8873	-0.4995	-0.8344	0.9095	-0.7002	-0.9264	-0.2659	1.3579	1.7485
didn_t	0.2388	1.6403	0.6998	0.3768	0.4902	0.6032	0.2915	-1.0732	-1.7833	-0.279	-0.3908	0.3423
doesn_t	0.2388	5.2202	-0.2075	-0.2366	1.0425	0.6032	-0.6198	-0.485	-2.3659	-0.9095	-0.2144	-0.8785
collection	0.2333	1.1808	4.2666	0.5811	0.7086	-0.4963	-0.6379	-1.5841	-1.8449	0.5364	-1.149	-0.2761
laptop	0.228	-1.6565	2.3453	-0.2635	-1.4151	0.3251	1.6749	-0.5386	-2.5054	-1.6562	-0.2393	5.1204
10	0.228	-0.5543	-0.4411	-0.7675	-0.3399	0.3251	0.5102	0.597	1.5221	-0.3059	-0.4264	-0.5299
color	0.2228	-0.5737	2.2875	0.326	-0.9364	-0.5326	-1.28	0.3832	1.1585	0.7711	-1.2107	-0.5486
since	0.2228	-0.9866	-0.2394	1.4807	-1.4515	0.538	-0.3483	-0.232	-1.1112	-0.3197	1.1845	1.1521
ever	0.2177	-1.0124	0.924	0.766	-0.9662	-0.3041	1.5916	-0.5944	-0.4251	0.2705	-0.4632	0.7785
while	0.2128	-0.3483	-0.8728	-0.3063	0.386	0.4985	0.2385	-0.2654	0.4321	-0.613	-0.2788	1.0819
20	0.2128	-1.754	2.7105	0.2958	0.5969	0.4985	-0.3741	-0.623	-2.7166	-1.7537	0.5306	2.9347
please	0.2081	0.4396	-0.2731	1.3432	0.5718	-0.5884	0.4442	-1.3224	-1.6582	1.7936	0.5091	-0.6059

Figure 46 Analyse de spécificités del corpus Reddit

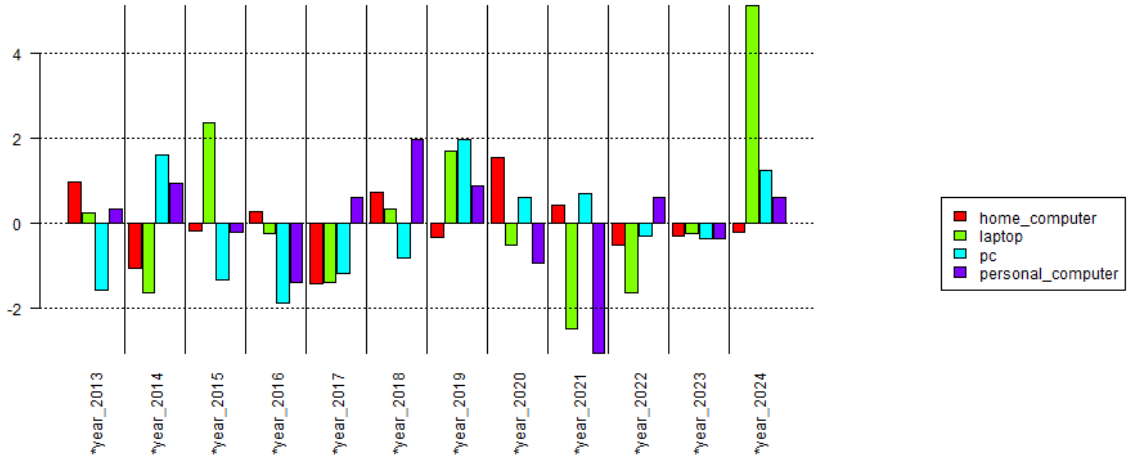


Figure 47 Analisi diacronica delle varianti terminologiche (pc, laptop, home_computer e personal_computer) nel corpus Reddit (2010-2024).

Nel corpus Reddit (2014–2024), *personal_computer* e le varianti prese in considerazione (*pc, home_computer, laptop*) non mostrano traiettorie lineari, ma oscillazioni moderate e spesso discontinue. *Personal computer* riemerge con oscillazioni moderate come oggetto di memoria e narrazione affettiva:

- *source: **** year_2018 subreddit_retrobatlestations audience_appassionati id_110_1031*

*"It could arguably be said that the first **personal_computer** software ran on this computer… portable for the time with battery backup power…" (2018*). Il termine compare in contesti rievocativi, legati alla classificazione storica dei primi dispositivi e alla celebrazione delle loro peculiarità tecniche.

- *source: **** year_2024 subreddit_retrobatlestations audience_appassionati id_280_2585*

*"My first experience of CP on one of these **personal_computer**…" (2019).* La prospettiva si fa autobiografica, e il riferimento tecnico diventa supporto a una memoria personale.

- *source: **** year_2019 subreddit_retrobatlestations audience_appassionati id_136_1159*

*"IBM **personal_computer** XT model 5160 enhanced color display…" (2024*). Il riferimento nominale al modello è accompagnato da formule emotive ("*I inherited my wife's late grandfather's…*"), che situano l'oggetto nel racconto domestico e affettivo.

Il termine *home_computer* presenta valori variabili, con picchi tra il 2018 e il 2020, in corrispondenza di discussioni legate alla storia della microinformatica domestica:

- ***** year_2021 subreddit_retrobattlestations audience_appassionati id_278_2576*
  *sir sinclair invented the pocket calculator but was best known for popularising the **home_computer**, bringing it to british high street stores at relatively affordable prices the following year;*
- ***** year_2020 subreddit_retrobattlestations audience_appassionati id_162_1283*
  *looking kind of like a box that might be used for pizza delivery, unix workstations from sun and next were some of the earliest examples of the pizza box with **home_computer** companies jumping in not long after;*
- ***** year_2024 subreddit_retrobattlestations audience_appassionati id_299_2711*
  *the pravetz 16 could be purchased as a **home_computer**, and the price was again outrageous as 3000 levs. versions were produced between 1984 and 1988 and some sources claim production went into the 90s as well.*

Il termine *laptop* presenta frequenze più irregolari e compare soprattutto in contesti pratici, riferiti all'uso quotidiano o alla valutazione delle prestazioni:

- ***** year_2015 subreddit_retrobattlestations audience_appassionati id_340_326*
  *I do have a 13 year old **laptop** that is my primary around the house computing device. I have a modern work **laptop** that goes with me for work reasons so rarely need to bring my old **laptop** with me.*
- ***** year_2023 subreddit_retrobattlestations audience_appassionati id_256_2381*
  *intel because they feared they could negatively affect the sale of more expensive ultraportable business **laptops**.*

- ***** year_2023 subreddit_retrobattlestations audience_appassionati id 104 938 my hp 100lx hinge has a bit more play… I'd love to see a modern laptop that small…"*

A seguire l'analisi della specificità sul corpus museale in base alle tipologie di risorse:

formes	*source_HomeComputerMu...	*source_LondonMuseum
computer	1482	1869
desktop	1007	65
commodore	609	134
device	606	160
philips	549	29
apple	491	191
output	448	17
data	409	185
and	389	2695
storage	386	51
transfer	384	7
or	383	81
game	327	975
portable	279	80
console	278	254
atari	266	101
ibm	253	276
monitor	195	163
tandy	191	20
drive	172	154
nintendo	161	198
input	157	17
sony	136	67
tulip	136	0
sinclair	124	114
disk	121	143
packard	118	68
micro	110	405
hewlett	108	68
2	91	272
macintosh	90	39
1	87	271
olivetti	80	37
generiek	79	0
accessory	78	39
ii	77	93
handheld	75	7
model	74	554
80	74	86

Figure 48 Analyse de spécificités del corpus museale

Mentre nei sottocorpora precedenti (sottocorpus *computer1*, sottocorpus *computer2*, *Reddit*) l'analisi si è concentrata sulla variazione diacronica dei termini (anno per anno), nel caso del corpus museale è stato possibile esplorare un'altra dimensione dell'analisi lessicometrica: il confronto tra risorse distinte. In particolare, è stato utilizzato il metadato **source_* per distinguere i testi provenienti da diverse istituzioni, dimostrando così come IRaMuTeQ consenta non solo di tracciare l'evoluzione temporale del lessico, ma anche di mettere in relazione i discorsi prodotti da soggetti diversi su uno stesso dominio.

L'analisi della frequenza dei termini per ciascuna risorsa museale rivela differenze lessicali significative tra il HomeComputerMuseum e il LondonMuseum, due istituzioni entrambe impegnate

nella documentazione e valorizzazione del patrimonio informatico, ma con approcci discorsivi differenti.

Nel caso del HomeComputerMuseum, prevale un lessico di taglio tecnico-descrittivo, fortemente orientato alla nomenclatura hardware e alla specificità dei marchi:

- *desktop (1007),*
- *commodore (609),*
- *device (606),*
- *philips (549)*

Le frequenze elevate anche per termini relativi al funzionamento (*storage (386), transfer (384), portable (279*). Si tratta di un linguaggio che richiama quello della manualistica tecnica o della collezione privata, volto a restituire una descrizione dettagliata dei dispositivi.

Il LondonMuseum, invece, mostra una diversa configurazione lessicale: pur condividendo termini comuni (*computer, console, monitor*), emergono con maggiore evidenza parole come *game*, *model*, *micro*, nonché una frequenza insolitamente alta di congiunzioni (*and*, *or*), indizio di una narrazione più articolata e discorsiva.

Il confronto evidenzia come anche all'interno dello stesso dominio tematico, l'analisi dei corpora tramite strumenti lessicometrici consenta di mettere in luce differenze discorsive tra soggetti produttori del contenuto, aprendo così la strada a considerazioni più ampie sulla stilizzazione linguistica e ideologica della tecnologia nei contesti espositivi.

formes	*source_HomeComputerMu...	*source_LondonMuseum
occurences	20225	96847
formes	2723	7350
hapax	1151	2444
segments	514	2712
hapax/formes	0.4227	0.33252

formes	*source_HomeComputerMu...	*source_LondonMuseum
nom	9748	36836
nr	7295	14084
num	1560	4717
sw	1089	34867
adj	522	3271
ver	11	3072

Figure 49 Confronto morfosintattico e lessicometrico tra le fonti museali

Il corpus del *HomeComputerMuseum* presenta, su circa 20.000 occorrenze, solo 11 verbi, a fronte di oltre 3000 nel *LondonMuseum*, che risulta più esteso (quasi 97.000 occorrenze). La struttura linguistica del primo è quindi marcatamente nominale e numerica, con un'elevata presenza relativa di numeri (7,7% contro 4,8%), che rimanda a un'enumerazione di modelli, codici e versioni.

A livello di varietà lessicale, nonostante la minore estensione, il *HomeComputerMuseum* mostra un indice hapax/formes più elevato (0.4227), a indicare una maggiore diversificazione relativa del vocabolario impiegato, probabilmente dovuta alla descrizione di oggetti estremamente eterogenei. Tuttavia, questa varietà non si traduce in una narrazione discorsiva: l'assenza di verbi e di connettivi limita la costruzione testuale, che rimane ancorata a schemi enumerativi e nominali.

3.2.1.6 Analisi fattoriale delle corrispondenze (AFC)

L'AFC è stata applicata per esplorare le relazioni tra le parole e le unità di contesto (in questo caso i segmenti temporali o annuali del corpus), riducendo la matrice di contingenza a un numero limitato di dimensioni interpretabili.

Il risultato di questa procedura viene rappresentato su piani cartesiani, dove ogni punto corrisponde a una forma attiva (termine) o a una unità di contesto (anno o blocco di anni) e le loro rispettive posizioni indicano vicinanza o lontananza semantica.

L'analisi consente di mettere in luce strutture tematiche latenti all'interno del corpus, che difficilmente sarebbero individuabili con una semplice analisi di frequenza,

Permette inoltre di osservare contrapposizioni stilistiche o semantiche tra i periodi, evidenziando l'evoluzione del lessico tecnico e delle priorità scientifiche lungo l'asse cronologico.

Attraverso l'AFC emergono:

- cluster lessicali: gruppi di termini che tendono a comparire insieme in specifici periodi storici,
- relazioni cronologiche: associazioni tra specifiche epoche e campi semantici predominanti (es. anni '70 associati a *mainframe*, anni 2010 a *cloud* e *cybersecurity*).

A seguire si riportano le proiezioni sul piano fattoriale delle forme e delle variabili temporali per il *Sottocorpus 1*:

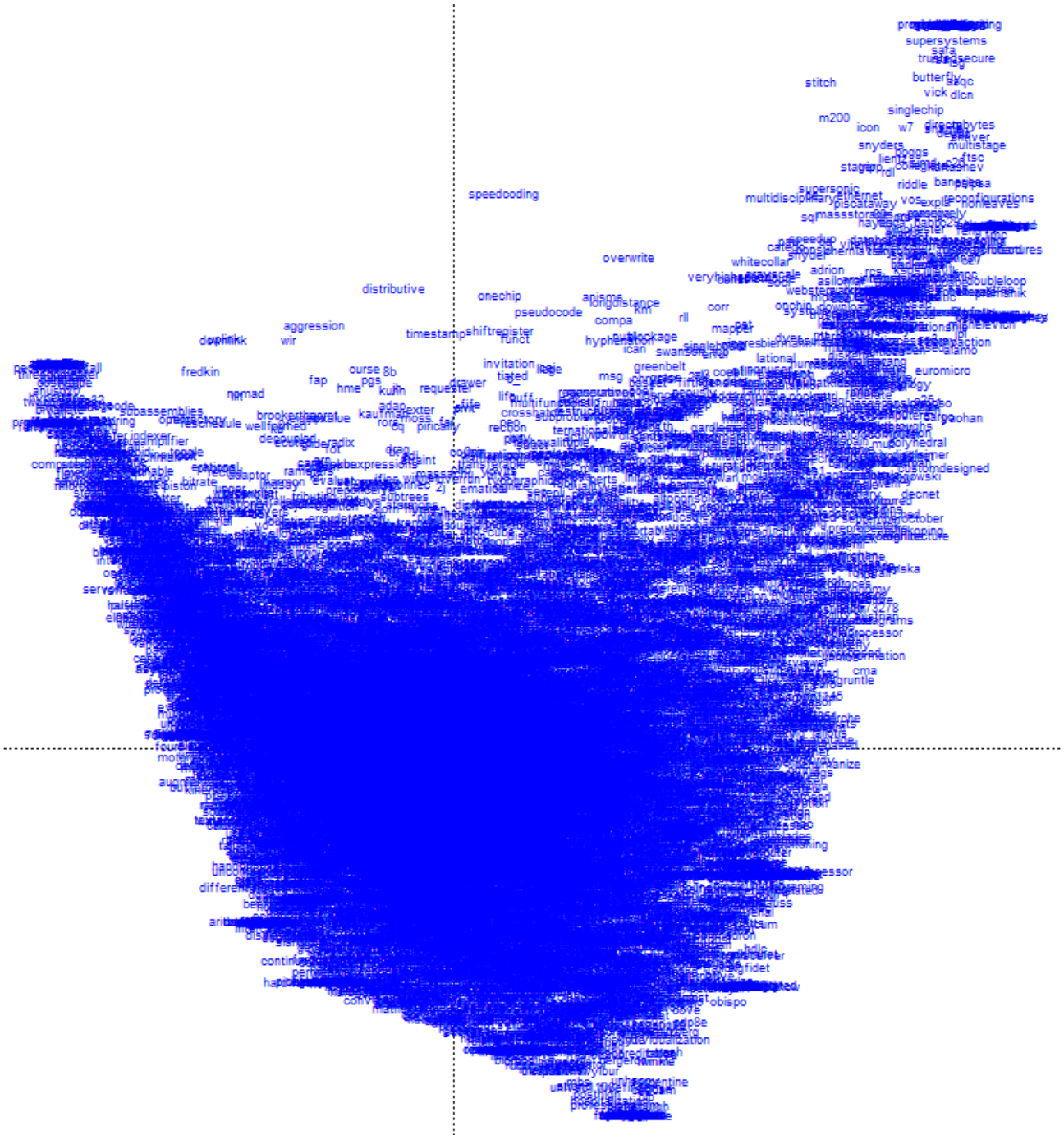


Figure 50 Mappa delle parole su piano fattoriale (AFC) per il sottocorpus 1

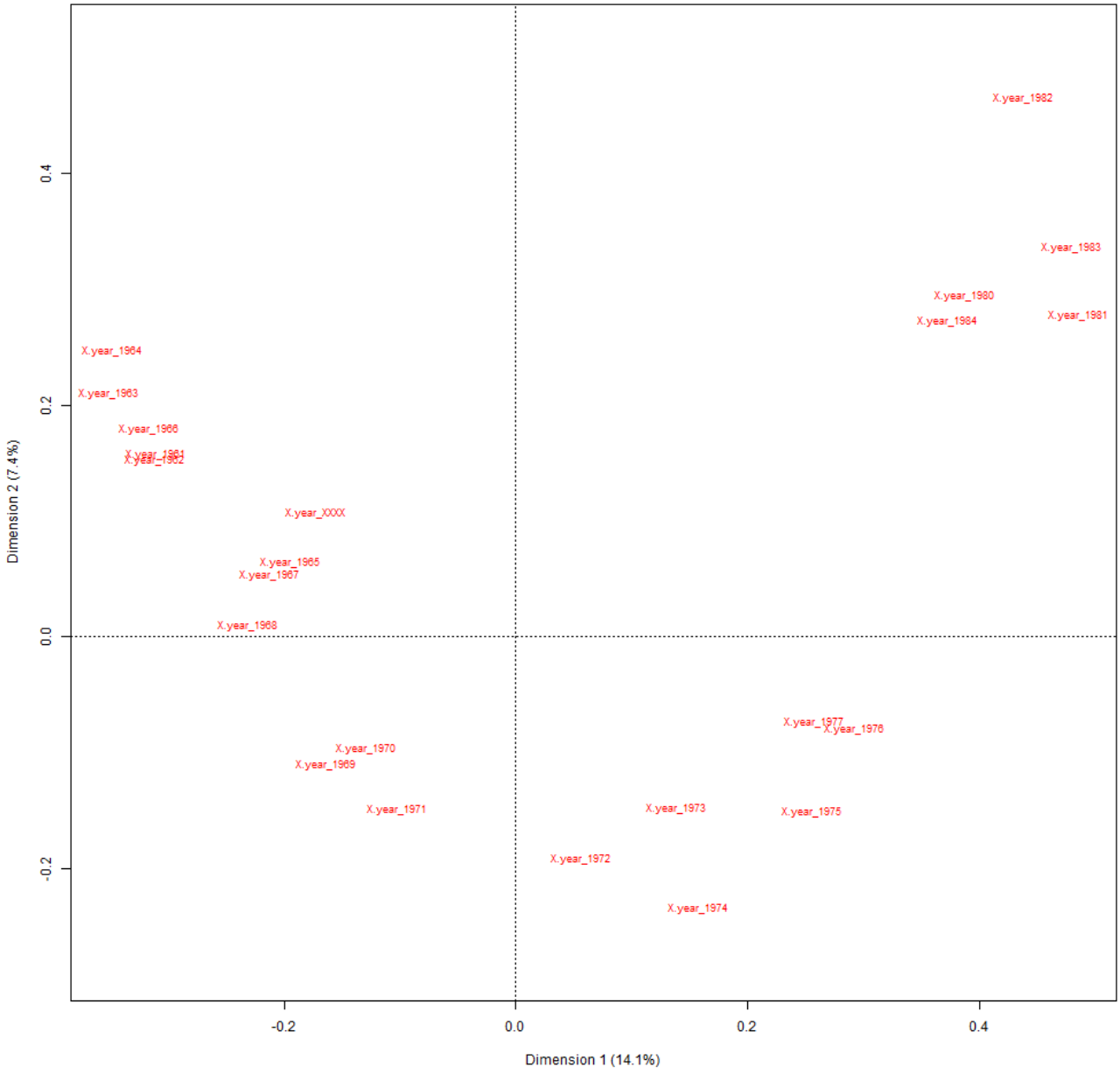


Figure 51 Proiezione degli anni sul piano fattoriale (AFC) del sottocorpus 1.

A seguire, si riportano i risultati ottenuti per il *Sottocorpus 2*:

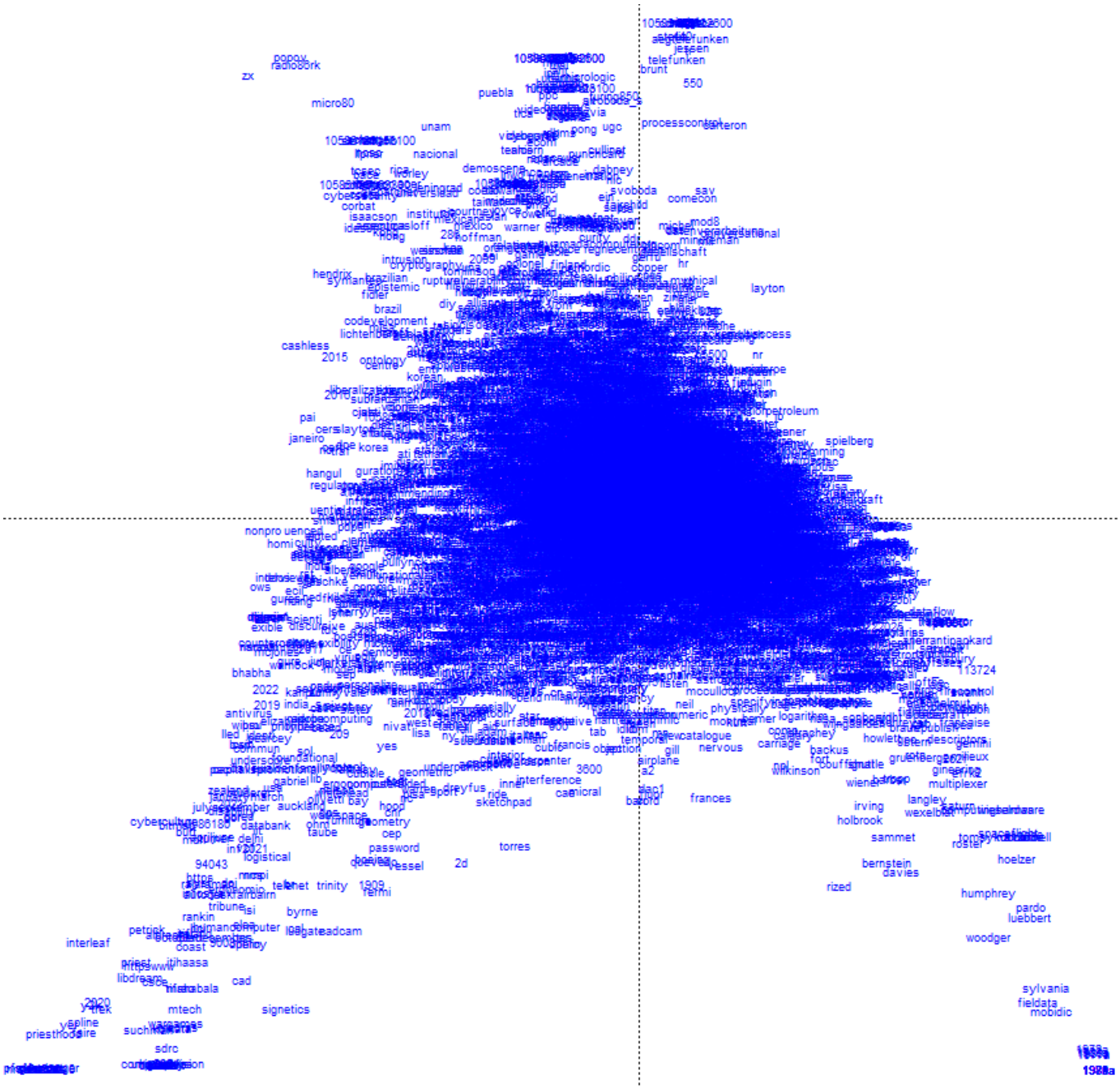


Figure 52 Mappa delle parole su piano fattoriale (AFC) per il sottocorpus 2

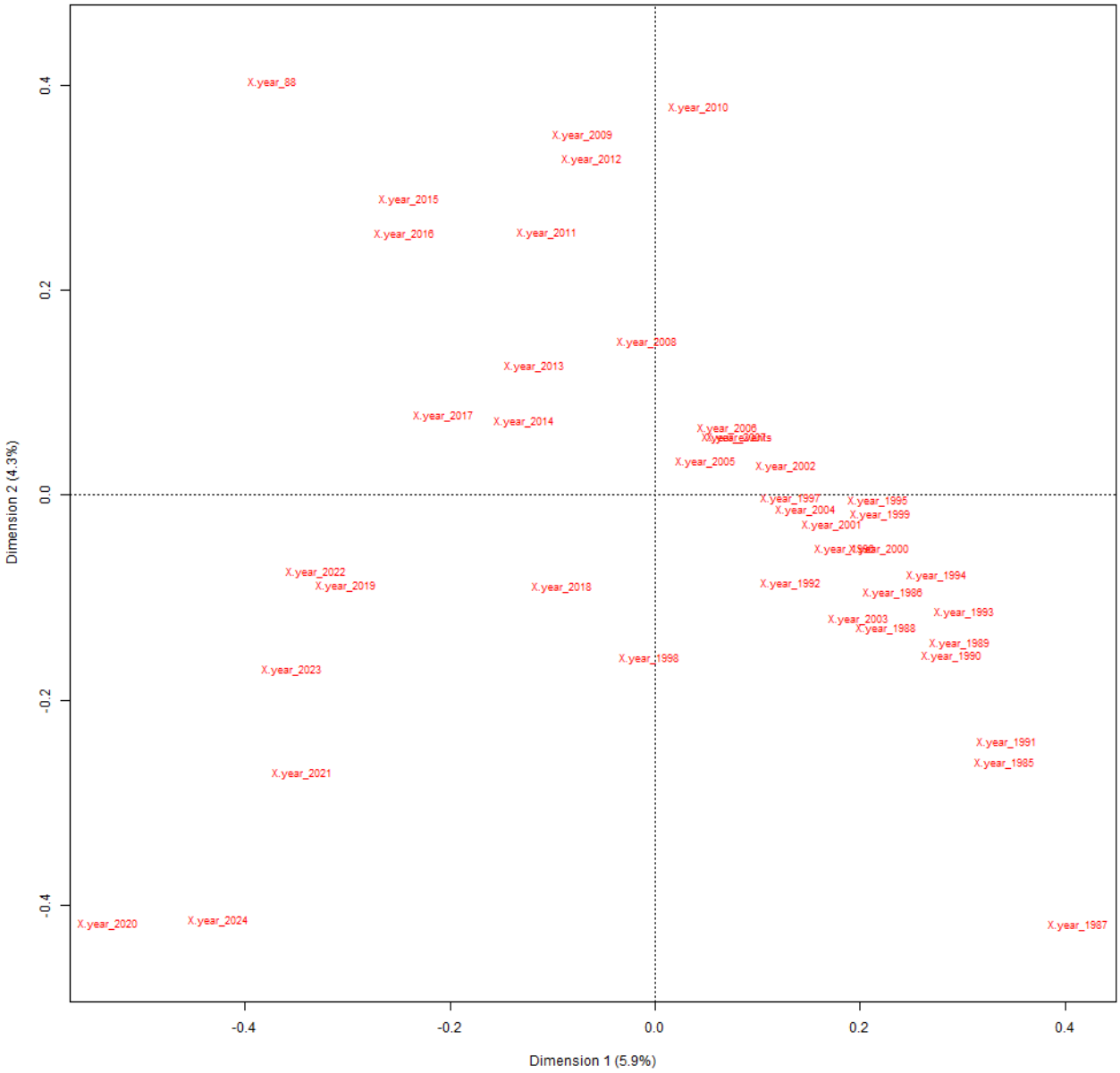


Figure 53 Proiezione degli anni sul piano fattoriale (AFC) del sottocorpus 2.

Nel sottocorpus 1, la prima dimensione AFC spiega il 14.1% della varianza, contro appena il 5.9% nel sottocorpus 2. Questo dato suggerisce una maggiore distinzione lessicale tra i periodi storici nel primo sottocorpus, rispetto a una relativa omogeneità nei testi contemporanei.

In termini statistici, la varianza rappresenta la quantità di informazione (cioè di dispersione o dissimilarità) tra righe e colonne della matrice che viene catturata lungo quel particolare asse.

Nel grafico AFC generato con IRaMuTeQ, si evidenzia una distanza tematica significativa tra i vari periodi del corpus. La distribuzione a "V" o 'U', o rovesciata, dei periodi lungo i fattori principali suggerisce un *effetto Guttman* inverso, come descritto da André Salem (1991). Questo fenomeno,

frequentemente osservato nei risultati delle analisi fattoriali applicate a serie cronologiche, indica una discontinuità tra le fasi centrali e quelle iniziali e finali del corpus.

L'articolo di André Salem si concentra sull'analisi delle serie testuali cronologiche, ovvero sui corpora composti da testi prodotti nel tempo da un unico emittente (individuo o collettivo) o in situazioni comunicative simili, che quindi presentano caratteristiche lessicali comuni. L'obiettivo principale è indagare le variazioni lessicali nel tempo. Il metodo delle specificità valuta l'uso di termini in specifiche sotto-sezioni del corpus rispetto a un modello probabilistico. Tuttavia, Salem sottolinea che nei corpora cronologici questo approccio può risultare limitato, in quanto tende a rilevare solo anomalie marcate tra sottoinsiemi (ad esempio, tra i periodi iniziali e finali), senza cogliere appieno i cambiamenti progressivi. Per superare questa limitazione, Salem suggerisce di calcolare le specificità "connesse" tra gruppi contigui di periodi (ad esempio, tra il periodo 1-2, 2-3, e così via), invece di considerare solo singole unità temporali isolate.

Salem osserva che la metodologia classica delle specificità potrebbe "perdere" informazioni rilevanti in un corpus cronologico, poiché valuta ogni periodo (o anno) separatamente, ignorando la progressione temporale. Per risolvere questo problema, introduce il concetto di "triangolo delle specificità connesse": ovvero non si guarda solo alle specificità di ogni singolo anno, ma si calcolano anche le specificità su gruppi di anni consecutivi (ad esempio, anni 1-2, 2-3, 3-4, ecc.).

Questo approccio crea una struttura triangolare di calcoli che permette di tracciare meglio l'evoluzione lessicale nel tempo.

Analisi per blocchi quinquennali

- Per rilevare le tendenze lessicali all'interno del mio corpus, ho scelto di suddividerlo in blocchi quinquennali, secondo la seguente articolazione:
- 1985-1989
- 1990-1994
- 1995-1999
- 2000-2004
- 2005-2009
- 2010-2014
- 2015-2019
- 2020-2024

Questa segmentazione consente di osservare successivamente le variazioni lessicali in intervalli regolari e di analizzare l'evoluzione del vocabolario all'interno di periodi storici distinti.
La forma a V nei grafici *Figure 50* e *Figure 52* rivela che i periodi centrali (gli anni intermedi) si collocano al vertice della V, mentre i periodi estremi (gli anni iniziali e finali) si trovano alle estremità superiori della parabola. Questo andamento suggerisce una coerenza cronologica del corpus, con una graduale sostituzione dei termini nel tempo.
Nel sottocorpus 1, la configurazione a "V", infatti, riflette un vocabolario che evolve significativamente tra le fasi iniziali e finali, con scarsa sovrapposizione lessicale. La dinamica osservata suggerisce che il lessico tecnico segue un ciclo evolutivo: nuovi termini emergono in risposta a innovazioni, si consolidano con la diffusione delle tecnologie corrispondenti, e infine tendono a scomparire o a essere sostituiti. Questa traiettoria lessicale riflette la natura dinamica del sapere tecnico, dove il vocabolario si adatta continuamente alle trasformazioni concettuali, tecnologiche e culturali. Il sottocorpus 2, al contrario, mostra una maggiore densità centrale, indicando una maggiore continuità tematica nel tempo.
L'effetto Guttman nelle serie cronologiche si manifesta quando i punti (nel nostro caso, gli anni) si dispongono lungo la curva, con gli anni adiacenti che risultano vicini nello spazio fattoriale, mentre quelli più distanti nel tempo si collocano alle estremità opposte del grafico. Come spiegato da Salem, tale fenomeno è dovuto a vari fattori:

- Il vocabolario evolve gradualmente nel tempo, determinando una progressione lessicale.
- I periodi consecutivi tendono a condividere una maggiore quantità di termini rispetto a quelli più distanti.
- L'AFC "traduce" questa progressività temporale in una curva, dove l'asse principale (fattore 1) rappresenta il gradiente temporale.

Il grafico a seguire mostra il risultato dell'analisi AFC sul corpus *Reddit*:

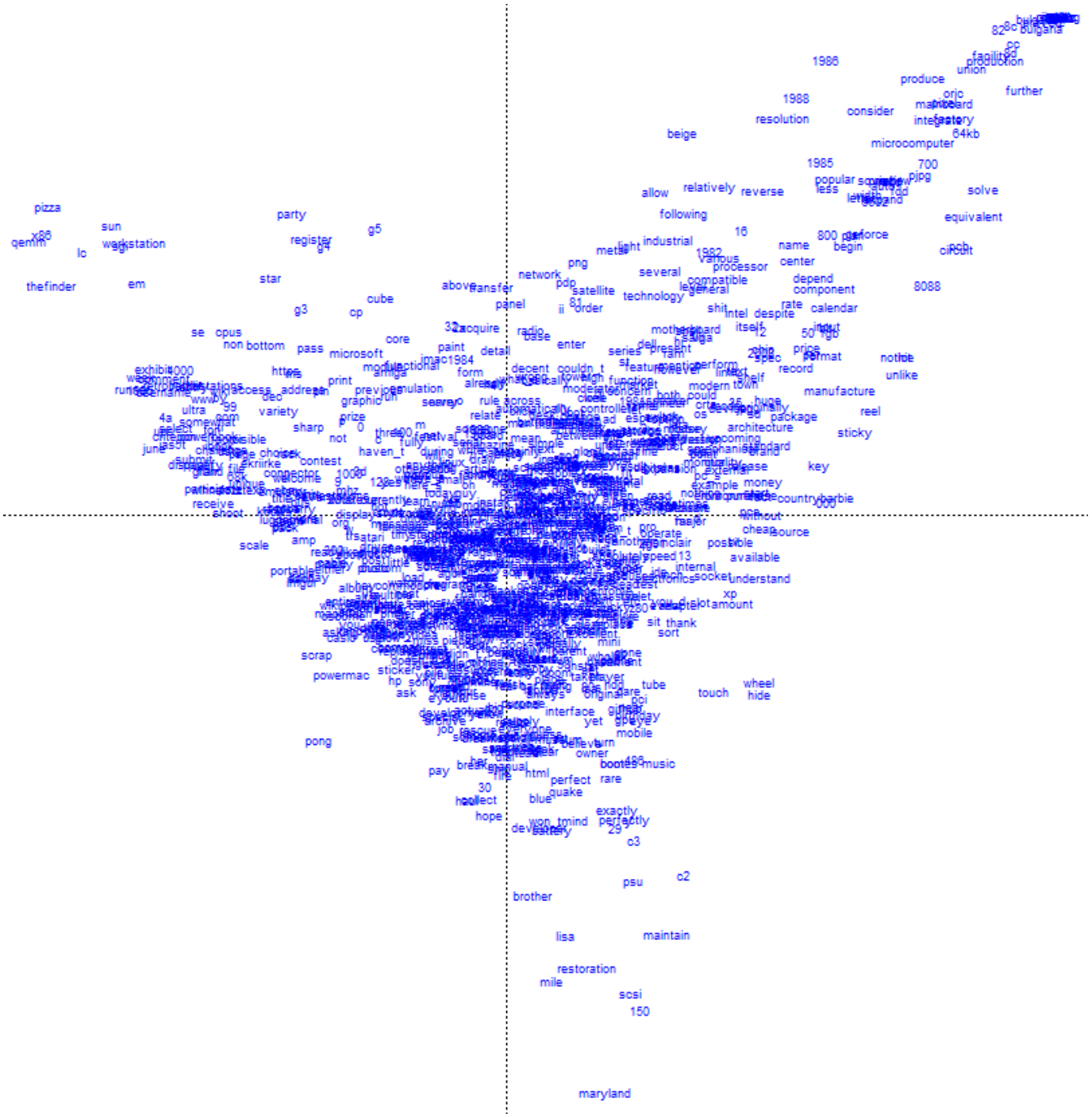


Figure 54 Mappa delle parole su piano fattoriale (AFC) per il corpus Reddit

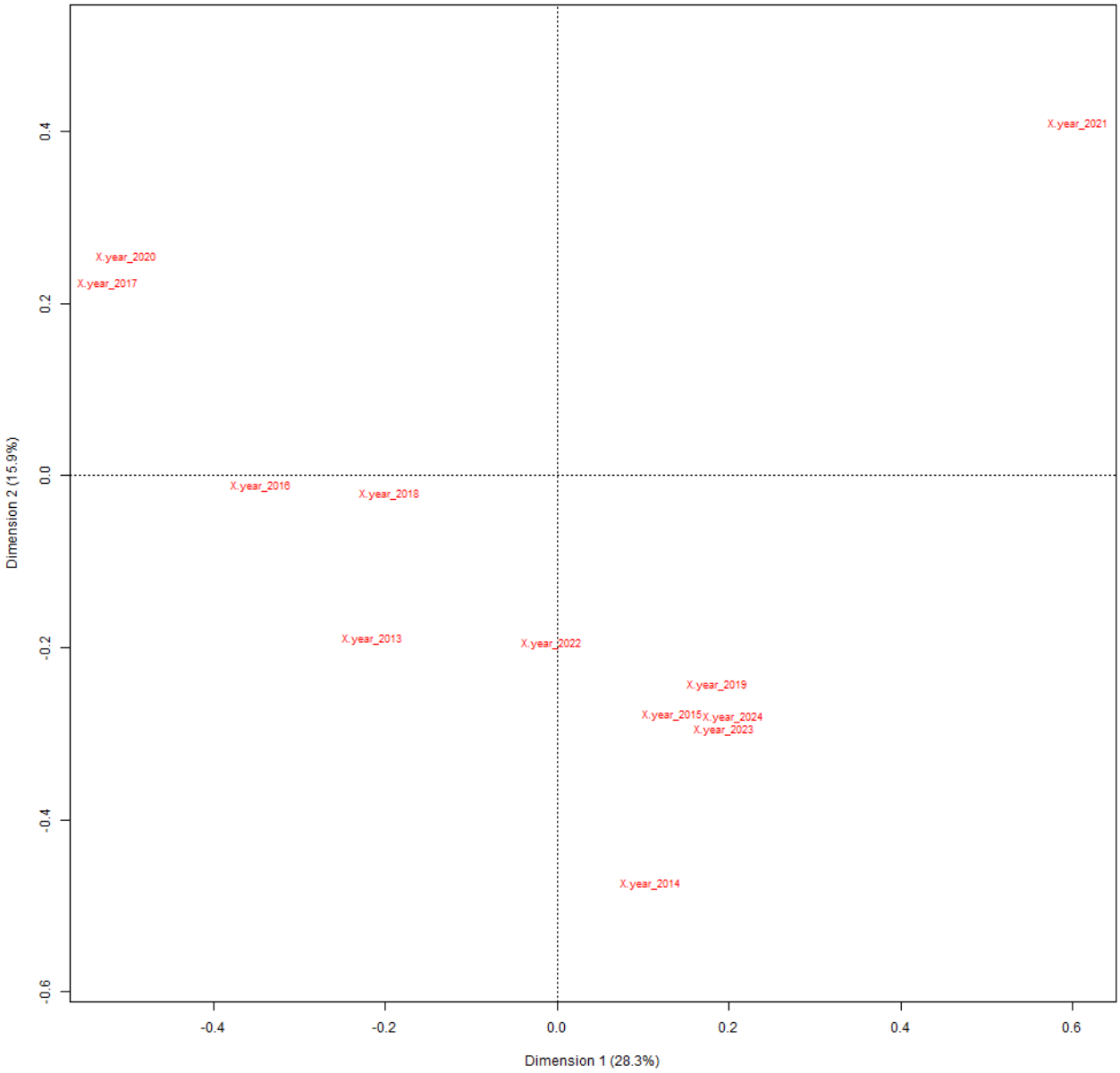


Figure 55 Proiezione degli anni sul piano fattoriale (AFC) del corpus Reddit.

L'AFC del corpus Reddit mostra sempre una 'V' semantica: le fasi più remote e più recenti emergono come zone lessicali periferiche, mentre gli anni intermedi confluiscono in un nucleo di frequenza condivisa.

A differenza degli altri corpora esaminati finora, qui gli anni non si susseguono in modo lineare: il 2020 occupa un picco isolato su un versante della V, il 2021 appare completamente distaccato sull'altro, e al centro convivono anni apparentemente lontani nel tempo – il 2014, il 2019 e persino il 2024 – come se condividessero un medesimo lessico "neutro". Questo disallineamento temporale suggerisce la presenza di ondate tematiche distinte, piuttosto che un'evoluzione continua e uniforme del discorso.

3.2.1.7 Classificazione automatica dei segmenti di testo (méthode Reinert)

Nel presente lavoro è stata adottata la *Classification Hiérarchique Descendante* (CHD), o classificazione gerarchica discendente, secondo la *méthode Reinert* (Reinert, 1983), come implementata nel software IRaMuTeQ. Tale metodo, inizialmente sviluppato da Max Reinert negli anni '80 per il software ALCESTE (*Analyse Lexicale par Contexte d'un Ensemble de Segments de Texte*), rappresenta una tecnica di analisi statistica dei dati testuali che consente di segmentare un corpus in unità di contesto (ad es. paragrafi, frasi o blocchi di testo) e di raggrupparle in classi lessicali omogenee sulla base della co-occorrenza delle parole. L'impiego della *méthode Reinert* nel presente lavoro si inserisce in una tradizione consolidata di analisi lessicometriche, già applicata in ambiti diversi, dal discorso politico (Longhi *et al.*, 2018) alla comunicazione scientifica, per l'individuazione di classi discorsive e strutture tematiche. L'analisi CHD si basa sull'ipotesi che sia possibile estrarre classi lessicalmente omogenee a partire dalla distribuzione delle forme attive (IRaMuTeQ, 2014).
Questa rappresentazione consente di visualizzare le distanze lessicali tra le classi sotto forma di rami divergenti, che riflettono la dissimilarità tra i gruppi lessicali individuati. Ogni classe risultante rappresenta un ambito tematico o discorsivo, caratterizzato da un vocabolario specifico.
A seguire, si riporta il dendrogramma riferito al *Sottocorpus 1* (*setting* con un valore massimo 10 mila forme):

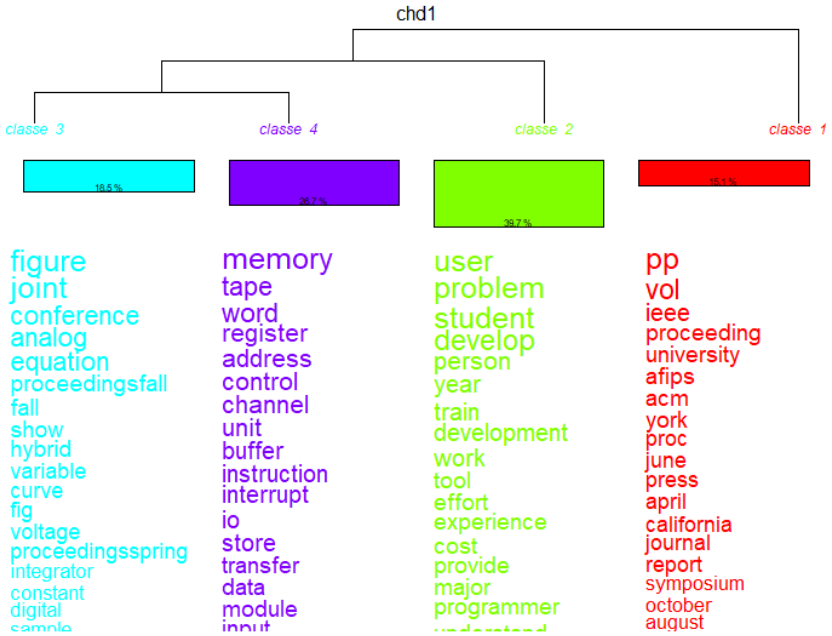


Figure 56 Dendogramma Sottocorpus 1

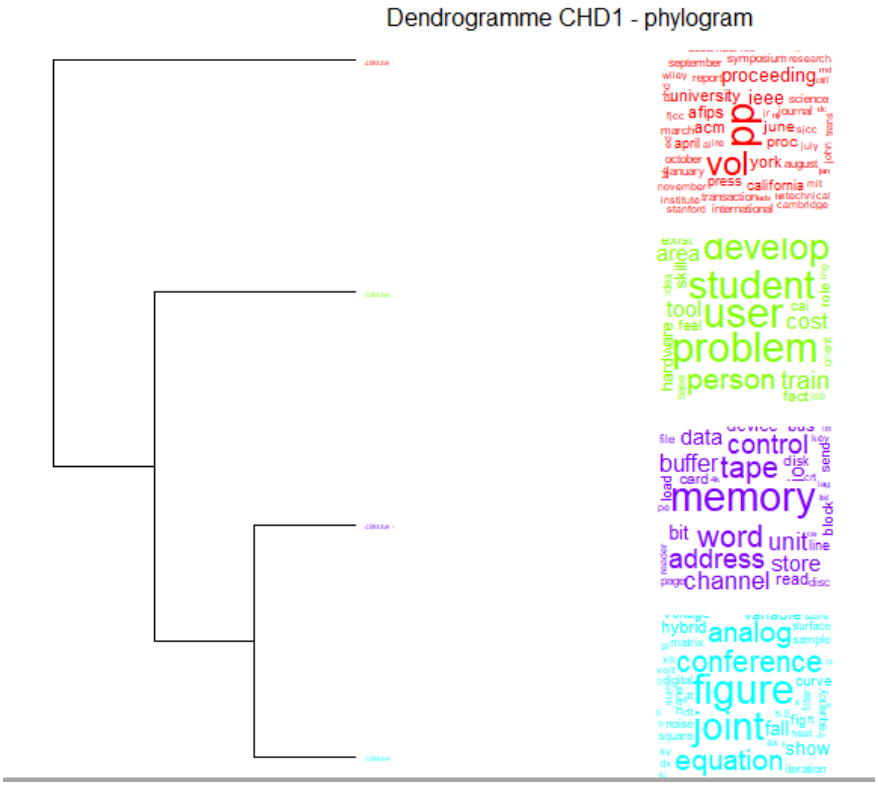


Figure 57 Phylogramma delle classi lessicali – Sottocorpus 1

Il primo dendrogramma (CHD1) rappresenta la struttura gerarchica risultante dalla classificazione automatica dei segmenti di testo secondo il metodo Reinert.

Il dendrogramma visualizza la partizione del corpus e fornisce un'indicazione della dimensione relativa di ciascuna classe (espressa in percentuale del corpus classificato). Ogni classe è caratterizzata da forme attive specifiche, la cui rilevanza statistica è calcolata tramite il test del χ^2 (chi quadrato). A partire da un'unica classe iniziale, il corpus viene segmentato iterativamente in sotto-classi lessicalmente omogenee, ciascuna contraddistinta da una distribuzione lessicale distintiva. Le classi identificate sono quattro:

Le classi identificate sono quattro:

Classe 1 (rossa, 16,7%) – pubblicazioni, atti di conferenze, riviste scientifiche.

Parole chiave: *pp, vol, ieee, proceedings, university, acm, report, symposium, journal, california, october, august.*

Questa classe evidenzia la centralità della letteratura scientifica e accademica nella documentazione della storia dell'informatica.

Il lessico include riferimenti a conferenze di rilievo (*ACM, IEEE, AFIPS*), riviste e pubblicazioni.

Classe 2 (verde, 26,7%) – utenti, studenti, sviluppo e programmazione

Parole chiave: *user, problem, student, develop, train, tool, experience, effort, cost, programmer.*

Indica un focus sul ruolo dell'utente, della formazione e dello sviluppo di software.

Potrebbe riferirsi sia agli aspetti educativi che a quelli professionali della programmazione.

Classe 3 (azzurra, 31,7%) – conferenze, elementi di elettronica e circuiti.

Parole chiave: *figure, joint, conference, analog, equation, voltage, hybrid, variable, curve, proceedingsspring, integrator.*

Mostra un legame con l'elettronica e la matematica applicata.

Potrebbe riflettere una fase iniziale della storia dell'informatica in cui gli aspetti hardware e ingegneristici erano dominanti.

Classe 4 (viola, 24,7%) – memoria, registri, controllo hardware.

Parole chiave: *memory, tape, word, register, buffer, instruction, interrupt, channel, data, module, input, output.*

Indica un'enfasi sui componenti hardware della memoria e sulla gestione dei dati.

Potrebbe essere legata agli anni '60-'80, quando la gestione della memoria era un problema tecnico di rilievo.

La percentuale relativa delle classi è simile tra loro, indicando che il corpus è bilanciato.

Il secondo grafico, detto "phylogramma", è semplicemente una rappresentazione alternativa della stessa struttura CHD, organizzata come albero filogenetico. In questa visualizzazione i rami non solo mostrano le gerarchie, ma indicano anche la distanza quantitativa (cioè la dissimilarità) tra le classi: più il ramo è lontano dal nodo comune, più diversa è la classe rispetto al resto del corpus.

Segue il grafico di Analisi delle Corrispondenze Multiple (AFC), che consente di rappresentare graficamente sul piano fattoriale le classi lessicali individuate tramite il metodo di Reinert (CHD). Le parole chiave associate a ciascuna classe vengono proiettate in base alla loro co-occorrenza, e le distanze spaziali tra i gruppi di termini riflettono la dissimilarità semantica.

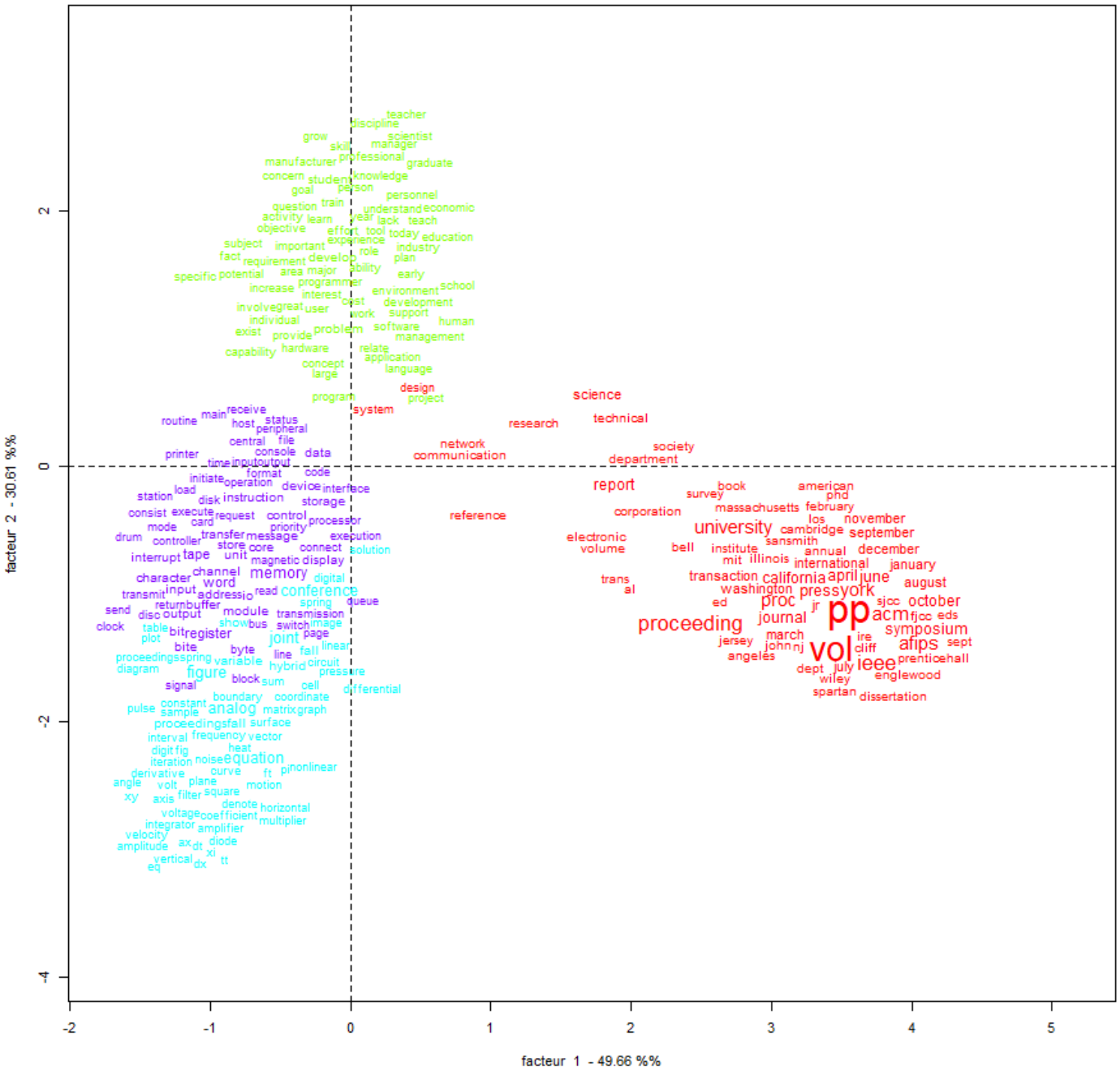


Figure 58 Analisi delle Corrispondenze Multiple (AFC) – Sottocorpus 1

In questo modo, l'AFC conferma visivamente la distinzione tematica delle classi emerse dalla classificazione automatica mostrando aree lessicali ben separate, suggerendo coerenza interna e differenziazione tra gli ambiti discorsivi rappresentati. In particolare, i termini legati alla pubblicazione scientifica (*press, vol, journal, conf*) si dispongono in un'area semanticamente autonoma, ben separata dai cluster tecnici e ingegneristici (es. *memory, signal, circuit*), evidenziando una netta separazione tra discorso accademico e linguaggio operativo-tecnologico.

Il grafico che segue mostra la proiezione sul piano fattoriale dei metadati “year” e “source”, relativi al subcorpus computer1, ottenuta tramite l’AFC.
I punti corrispondono ai diversi anni di pubblicazione (*year_1961*, *year_1981*, ecc.) e all’unica fonte presente (*source_AFIPS_Joint_Computer_Conferences*), colorati in base alla classe lessicale con cui sono stati associati.

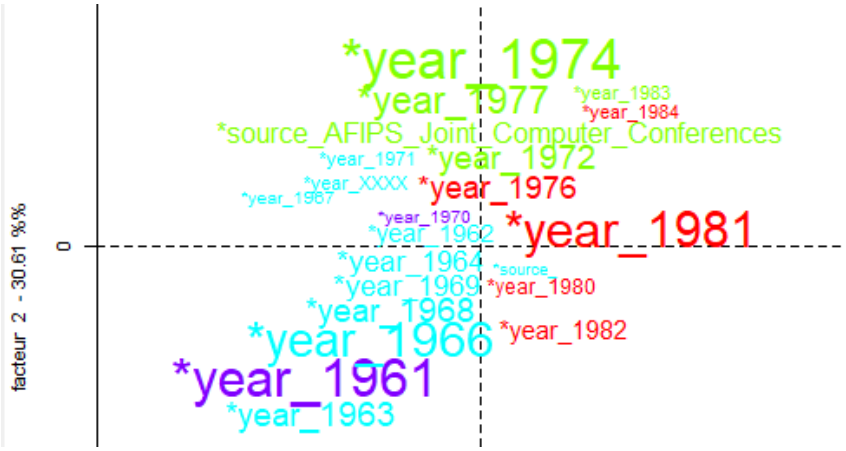


Figure 59 Analisi delle Corrispondenze Multiple (AFC) – Metadati temporali Sottocorpus 1

La disposizione spaziale dei punti nel piano fattoriale riflette somiglianze e divergenze lessicali tra gli anni. Alcuni anni, come il 1961 (in viola) e il gruppo 66,’68,’69,’64, (in azzurro), risultano associati a un vocabolario differente rispetto agli anni 1981–1982 (in rosso), suggerendo un’**evoluzione terminologica** netta nel tempo. Altri anni, come il cluster 1972–1977 (in verde), occupano una posizione intermedia tra i due poli, indicando transizioni semantiche o convergenze lessicali tra fasi successive. Alcuni punti si collocano in posizioni isolate rispetto al proprio cluster: ad esempio, il 1967 (azzurro) risulta più distante dagli altri anni ’60, suggerendo un profilo terminologico atipico o meno coerente internamente. Similmente, il 1984 (verde) appare separato dagli altri anni ’80, indicando una possibile variazione tematica o discorsiva. È inoltre rilevante il caso degli anni 1961 e 1963, vicini spazialmente ma appartenenti a classi differenti (viola e azzurro): ciò suggerisce una certa prossimità semantica, nonostante una classificazione statistica divergente legata alla diversa distribuzione delle forme attive nei rispettivi segmenti.

Dopo aver illustrato il funzionamento della classificazione automatica e dell’AFC applicate al primo sottocorpus (1961–1984), possiamo ora estendere l’analisi al secondo sottocorpus, relativo al periodo 1985–2024. Questo confronto consente di mettere in luce eventuali cambiamenti nei domini lessicali, emergenze concettuali e nuove aree discorsive che caratterizzano la produzione recente.

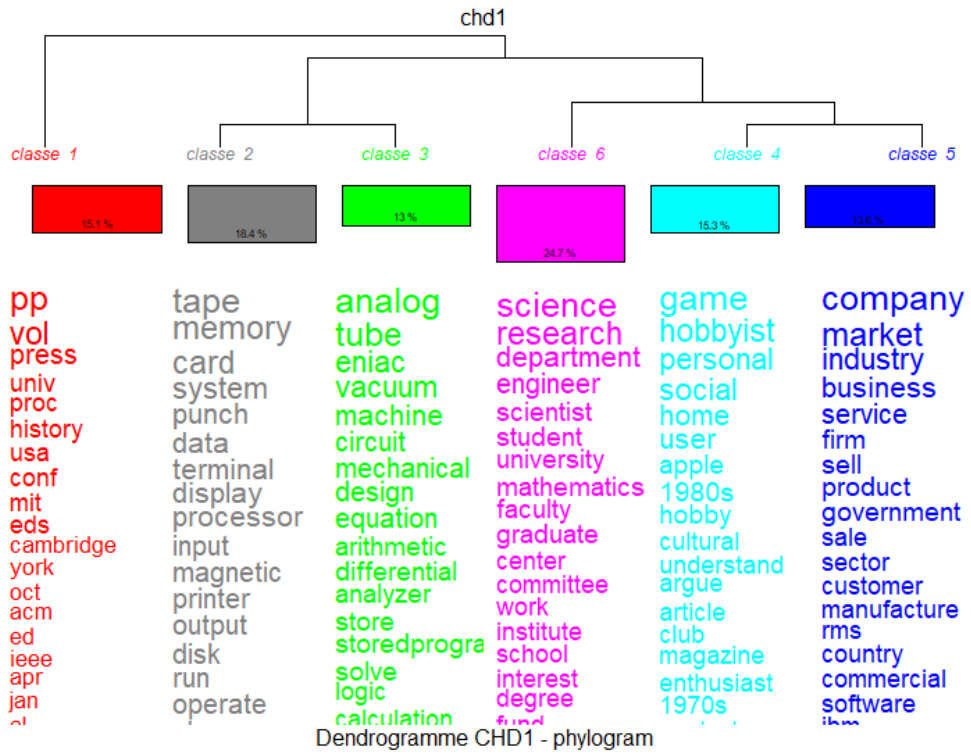


Figure 60 Dendogramma Sottocorpus 2

In questo dendrogramma, le classi sembrano più dettagliate e includono:

Classe 1 (rossa, 15,1%) – pubblicazioni, atti di conferenze, riviste scientifiche, come per il primo corpus.

Classe 2 (grigia, 18,4 %) - memorie di sistema, hardware di base.

Classe 3 (verde, 16,3%) – elettronica analogica, circuiti, strumenti di calcolo.

- Parole chiave: *tube, analog, vacuum, machine, circuit, mechanical, equation, arithmetic, analyzer, store, solve, calculation.*

Qui appare un chiaro riferimento alla storia dell'informatica, con parole chiave legate ai calcolatori analogici e ai primi sistemi di calcolo (*tube, vacuum, machine*).

Classe 4 (azzurra, 15,5%) – giochi, hobbisti e informatica domestica (anni '80 e '70).

- Parole chiave: *game, hobbyist, personal, home, user, apple, 1980s, article, club, software, magazine.*

Questa classe non era presente nel primo dendrogramma ed è molto significativa: introduce la dimensione culturale della storia dell'informatica, con riferimenti all'epoca dell'informatica domestica e al movimento degli hobbisti. Il riferimento agli *anni '80* e a *Apple* suggerisce una segmentazione temporale ben distinta.

Classe 5 (blu, 16,3%) – mercato, industria e impatto economico.

- Parole chiave: *company, market, industry, business, service, product, firm, customer, manufacture, country, rms, software.*

Questa è un'altra novità rispetto al primo dendrogramma: cattura l'evoluzione dell'informatica come settore economico, sottolineando il passaggio da una disciplina accademica a un'industria globale.

Classe 6 (rosa, 21,7%) – scienza e ricerca accademica.
Parole chiave: *science, research, department, engineer, scientist, university, mathematics, faculty, degree, committee, institute, phd.*

Questa classe corrisponde a una specializzazione della precedente classe "sviluppo software e utenti" (verde nel primo dendrogramma).
L'aggiunta di termini accademici (*PhD, faculty, department*) suggerisce un'analisi più dettagliata della comunità scientifica.

Il dendrogramma relativo al secondo sottocorpus evidenzia una classificazione più articolata rispetto al primo. Questa maggiore complessità può essere ricondotta sia all'ampiezza dell'arco temporale coperto (1985–2024), sia a un vocabolario più eterogeneo.
Un aspetto significativo è l'emergere di nuove classi tematiche, assenti nel primo corpus: in particolare, una dedicata all'informatica hobbistica e una al settore industriale. Ciò suggerisce un ampliamento del dominio discorsivo analizzato, che ora include non solo gli ambiti più tradizionali della ricerca accademica e delle conferenze, ma anche l'elettronica, la cultura informatica e l'evoluzione del mercato tecnologico.
Il dendrogramma distingue chiaramente tra epoche diverse: i termini *tube, vacuum* indicano l'era dei primi computer, *home, Apple, 1980s* richiamano la rivoluzione dell'informatica personale, e *market, industry* mostrano il passaggio all'era commerciale. Nel complesso mentre il primo corpus appare più focalizzato su concetti tecnici (hardware, memoria, software), il secondo corpus mostra una visione più ampia, includendo storia, cultura e industria.
Le tendenze emerse saranno ora esaminate più nel dettaglio attraverso l'analisi dei grafici successivi, con particolare attenzione alla disposizione spaziale delle classi lessicali nei piani fattoriali e alla segmentazione semantica risultante dall'AFC.

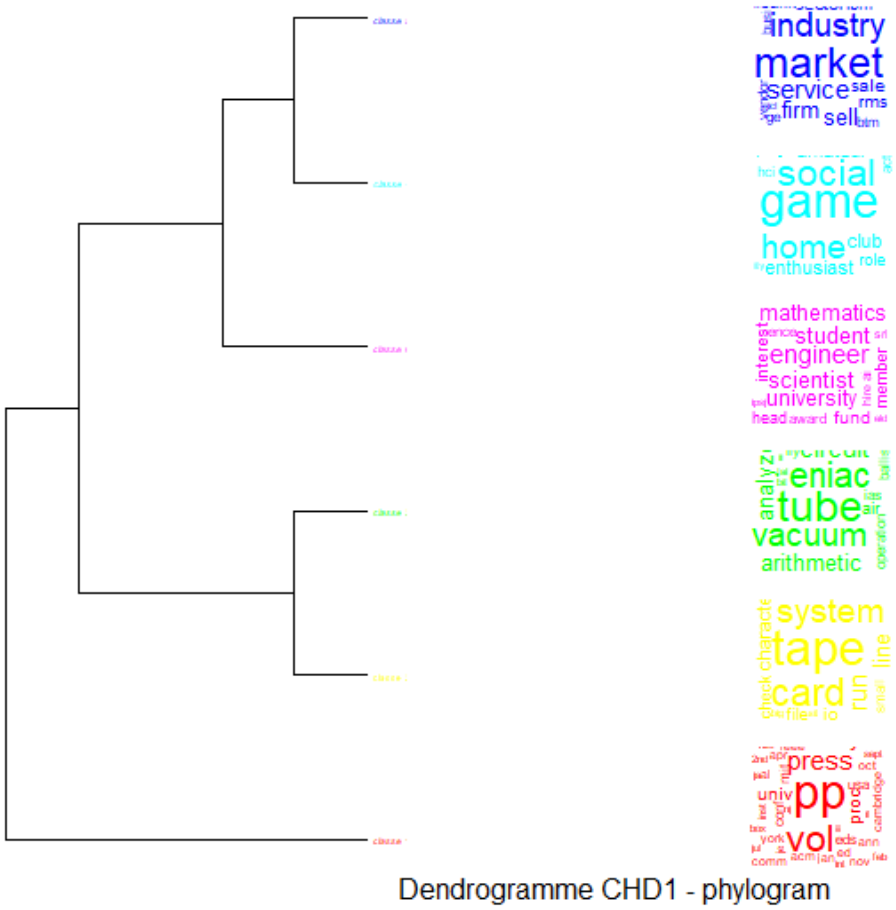


Figure 61 Phylogramma delle classi lessicali – Sottocorpus 2

Nel phylogramma, le sei classi evidenziate mostrano una maggiore articolazione rispetto al primo corpus. La disposizione dei rami evidenzia una distanza significativa tra alcuni ambiti tematici: ad esempio, la Classe 1, legata alle pubblicazioni scientifiche, risulta lessicalmente la più distante dalle altre, mentre si osserva una maggiore prossimità tra campi affini come l'industria (Classe 5) e l'informatica hobbistica (Classe 4), suggerendo una certa convergenza discorsiva tra la dimensione commerciale e quella sociale dell'informatica.

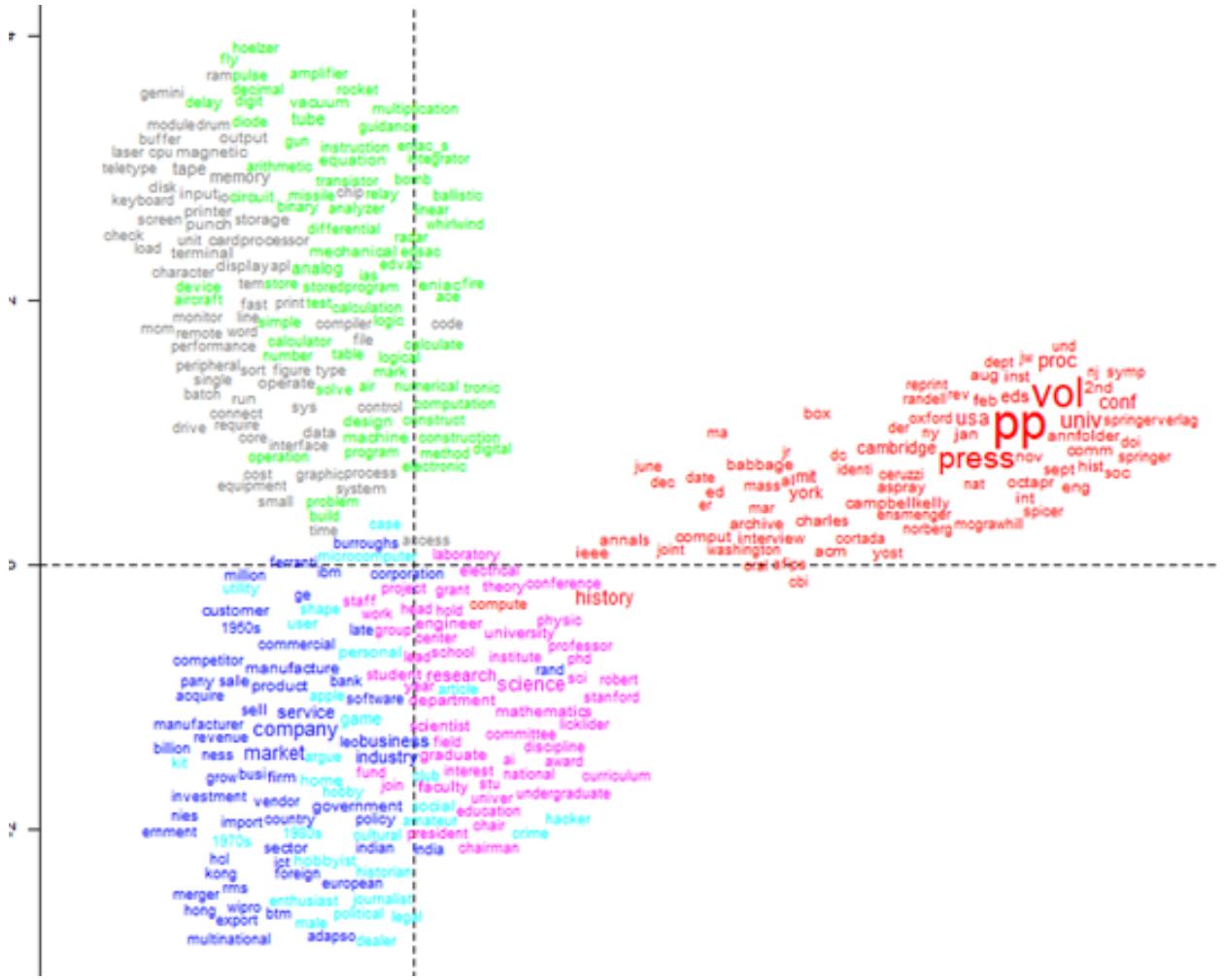


Figure 62 Analisi delle Corrispondenze Multiple (AFC) – Sottocorpus 2

Anche nel grafico dell'AFC relativo al secondo sottocorpus, i termini associati alla comunicazione accademica e scientifica (area rossa: *press, vol, journal, pp, conf*) si collocano lungo un asse semantico ben distinto, risultando chiaramente separati dagli ambiti lessicali, come già evidenziato nel primo sottocorpus.

L'area blu (con termini come *market, company, business, sell*) riflette la dimensione economica e industriale dell'informatica e appare parzialmente contigua all'area azzurra, centrata sull'informatica domestica e culturale (*hobbyist, apple, 1980s, club*), indicando una convergenza tra innovazione tecnologica, consumo e pratiche hobbistiche. Inoltre, si osserva una certa vicinanza tra queste due aree e l'area viola, che rappresenta il dominio accademico-scientifico (*science, research, university, mathematics*).

Questa disposizione spaziale suggerisce che nel periodo contemporaneo si sia sviluppata una interconnessione tra ricerca, mercato e cultura informatica, in linea con l'evoluzione dell'informatica da disciplina accademica a fenomeno pervasivo, con ramificazioni culturali, economiche e sociali.

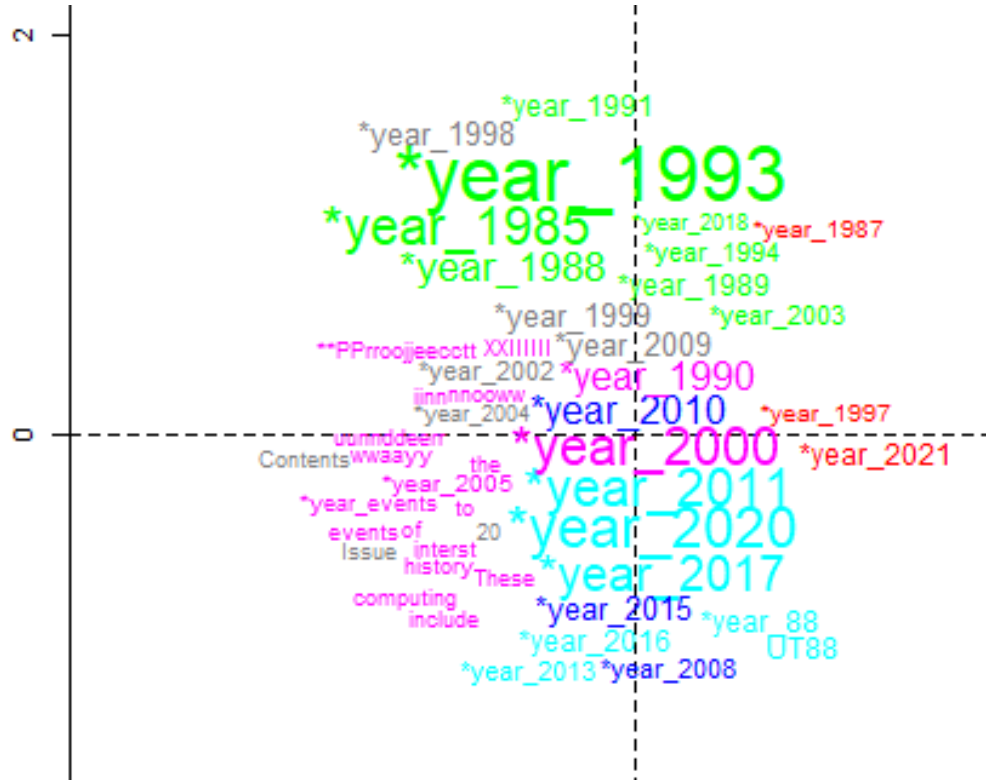


Figure 63 Analisi delle Corrispondenze Multiple (AFC) – Metadati temporali Sottocorpus 2

Nel secondo sottocorpus, i pattern di distribuzione sono meno marcati rispetto al primo corpus. L'area verde, corrispondente agli anni 1985-1993, si colloca in una posizione relativamente compatta sul quadrante superiore sinistro, suggerendo una coerenza terminologica. Gli anni più recenti (in azzurro, es. 2011, 2015, 2017, 2020), distribuiti lungo l'asse orizzontale, rappresentano una nuova articolazione lessicale, potenzialmente connessa all'ampliamento del discorso informatico in direzione economica, educativa e culturale. L'area rosa (anni 2000-2005) sembra fungere da zona intermedia tra i due poli, riflettendo una fase di riorientamento discorsivo.

Alcune etichette isolate, come *year_1997* e *year_2021* in rosso, appaiono più distaccate dal nucleo centrale, indicando specificità terminologiche marcate per quei singoli anni, probabilmente legate a eventi o contesti editoriali particolari.

L'osservazione delle concordanze relative alle forme *press, vol* e *pp* (v. Figura X, Analisi delle Corrispondenze Multiple (AFC) – Metadati temporali, Sottocorpus 2, e ricerca delle parole nello strumento di analisi di specificità, p. 53) integra i risultati quantitativi con un riscontro qualitativo sui contesti d'uso, confermando che gli anni 1997 e 2021, isolati nel piano fattoriale, corrispondono a momenti di particolare densità editoriale e metadiscorsiva.

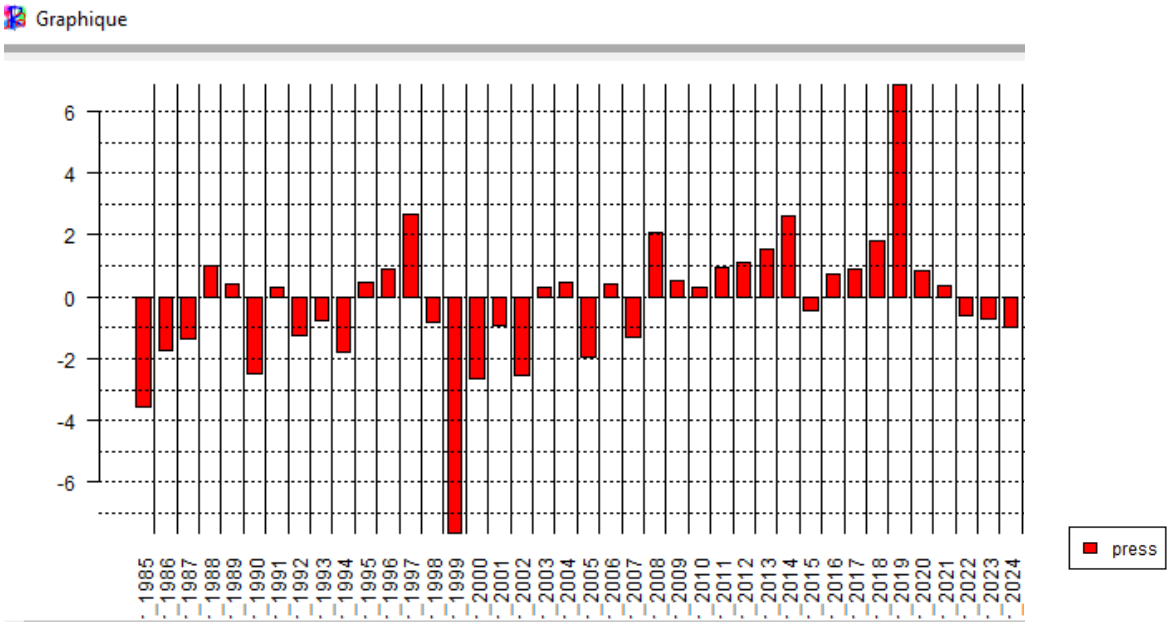


Figure 64 Distribuzione di specificità della forma press (1985–2024) – picchi nel 1997 e 2021.

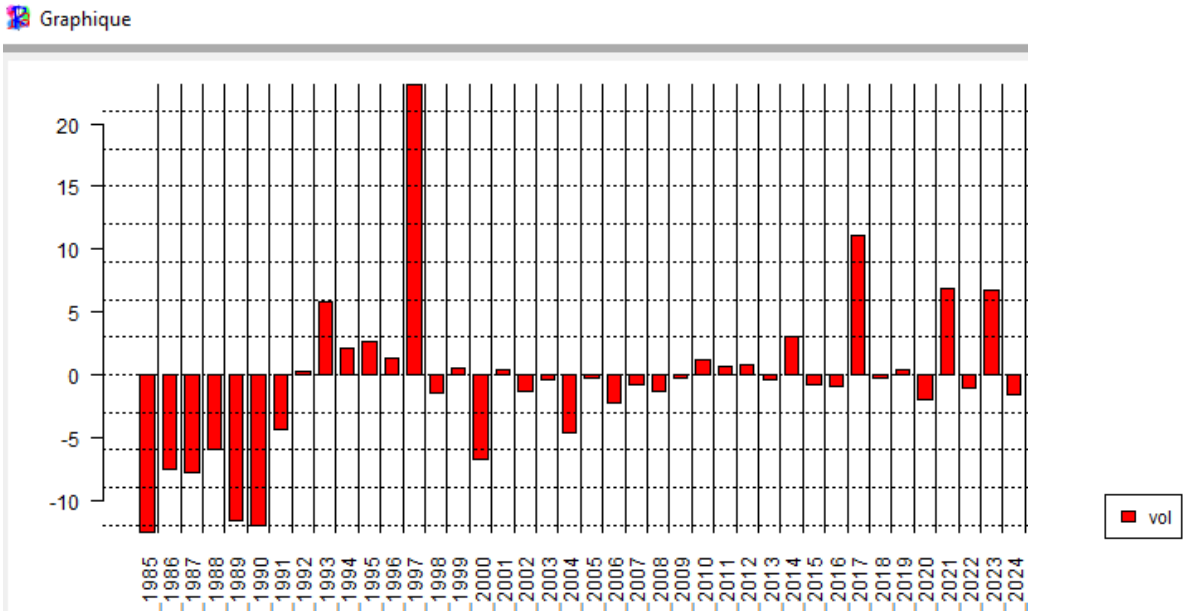


Figure 65 Distribuzione di specificità della forma vol (1985–2024) – picchi nel 1990 e 1997.

I grafici di specificità per le forme *press* (*Figure 64*) e *vol* (*Figure 65*) mostrano chiaramente due picchi significativi negli stessi anni, indicando un incremento nella frequenza relativa di riferimenti editoriali e citazioni bibliografiche all'interno degli articoli.

Nel 1997, le occorrenze di *press* compaiono prevalentemente in contesti di citazione bibliografica e accademica (MIT Press, Cambridge University Press, Oxford University Press), come mostrano gli esempi seguenti:

"Computers London; A guide to the history of computing, and the Her Majesty's Stationery Office (1960); Hills Information Processing Industry, Westport, Conn.: Greenwood Press.";
"*History Oxford, England: Oxford University Press, 1989; United States, Princeton, N.J.: Princeton University Press, 1962; Hendry, Innovating for Failure, Government Policy and the British Computer Industry, Cambridge, Mass.: MIT Press, 32 ff."*

Tale distribuzione riflette il carattere istituzionale di questa rivista con contributi come Shapiro 1997, "The Historical Necessity of Synthesis in Software Engineering" e Katz 1997, "Historical Content in Computer Science Texts".
Nel 2021 (vol. 43), le concordanze mostrano invece l'inserimento di *press* in un discorso più ampio e interdisciplinare (AI criticism, logic, culture of computing, algorithmic rationality):
con numerosi contributi che collegano informatica, etica e cultura come *Garvey 2021, "The General Problem Solver Does Not Exist: Mortimer Taube and the Art of AI Criticism"*; *Kneese 2021, "Your Computer is on Fire"* e *De Mol 2021"Logic, Programming, and Computer Science: Local Perspectives"*.

Anche per il corpus Reddit, le rappresentazioni grafiche tratte dalla classificazione automatica e AFC permettono di individuare aree lessicali ben differenziate, legate a specifici registri discorsivi.

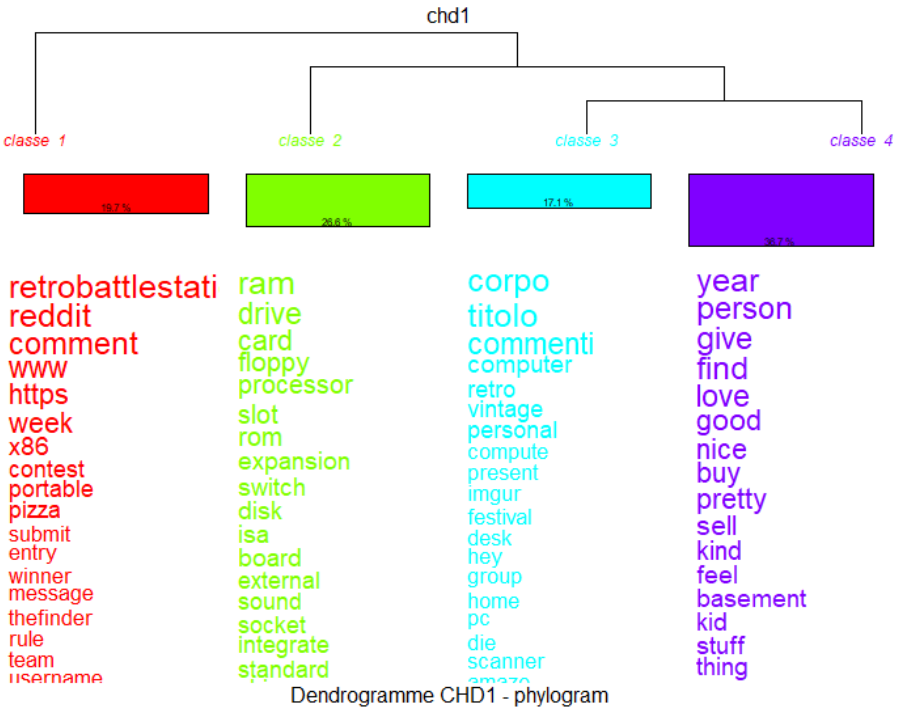


Figure 66 Dendogramma – Corpus Reddit

- **classe 1** (10,7 %, in rosso): raccoglie terminologia legata specificamente al contesto Reddit e al concorso (*retrobattlestation, reddit, comment, https, week, contest, entry, pizza*, ecc.).
- **classe 2** (26,6 % in verde): è centrata su componenti hardware e periferiche (*ram, drive, card, floppy, processor, slot, rom, expansion, disk, sound, socket,* ecc.).
- **classe 3** (17,1 % in azzurro): contiene lessico di discorso generico e riferimenti a contenuti web/nostalgici (*corpo, titolo, commenti, computer, retro, vintage, personal, festival, imgur, desk, pc, scanner,* ecc.).
- **classe 4** (36,7 % in viola): ospita vocaboli di tono maggiormente colloquiale o commerciale (*year, person, give, find, love, good, buy, sell, stuff, thing*, ecc.).

Il raggruppamento semantico della classe 4, essendo la più distante semanticamente dalle altre, avviene solamente nell'ultima fusione che chiude l'albero gerarchico.

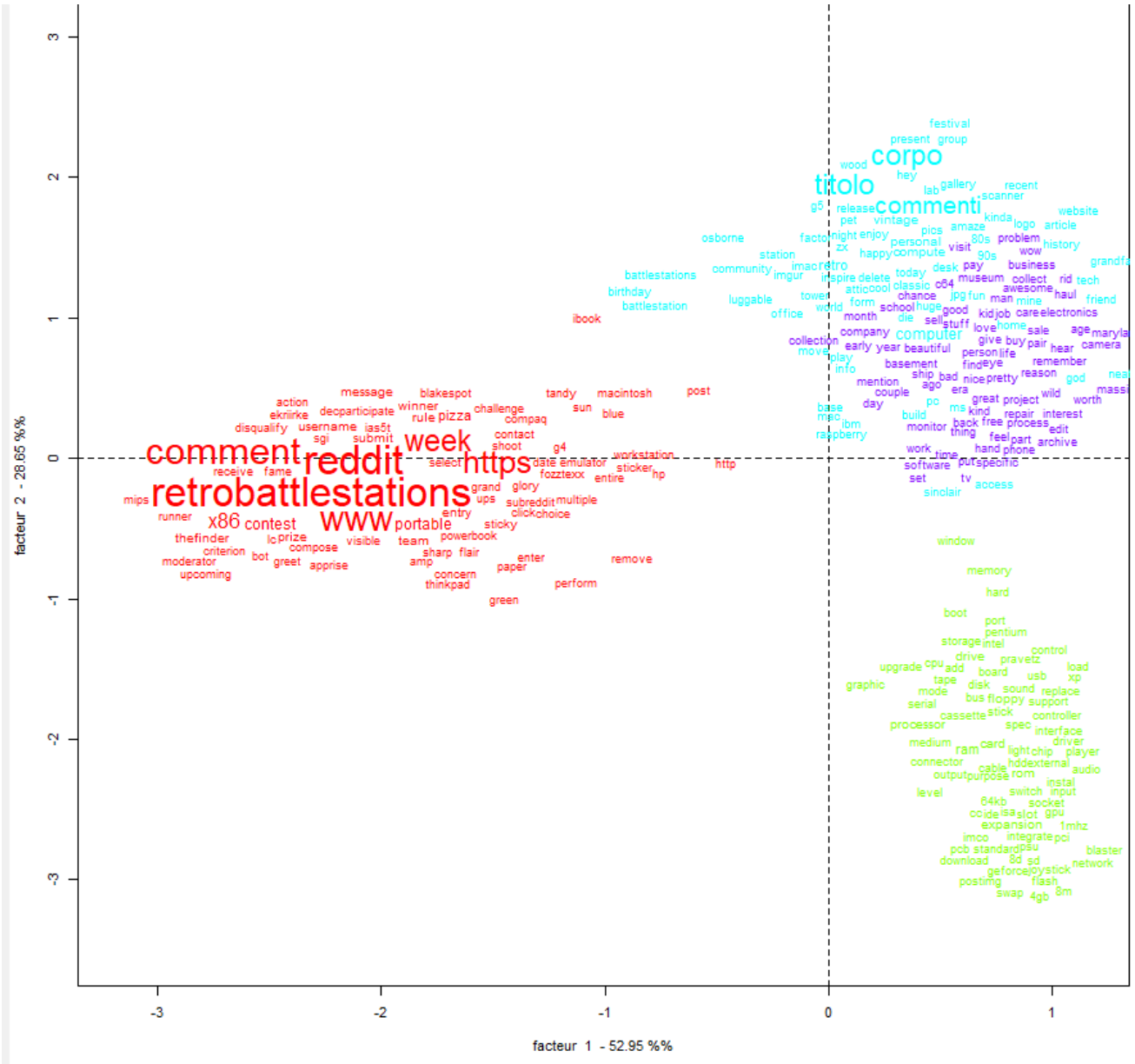


Figure 67 Analisi delle Corrispondenze Multiple (AFC) – Corpus Reddit

L'AFC conferma questa articolazione: le parole chiave si distribuiscono spazialmente in regioni ben separate (figura Y), evidenziando l'eterogeneità terminologica del corpus. L'area rossa è fortemente compatta e separata dalle altre, indicando un cluster coerente e molto specifico, mentre le aree azzurra e viola mostrano una sovrapposizione parziale, legata a una condivisione di forme linguistiche comuni, ma con sfumature semantiche diverse (culturale vs. affettivo-commerciale). L'area verde (tecnica) si mantiene nettamente separata lungo l'asse semantico principale.

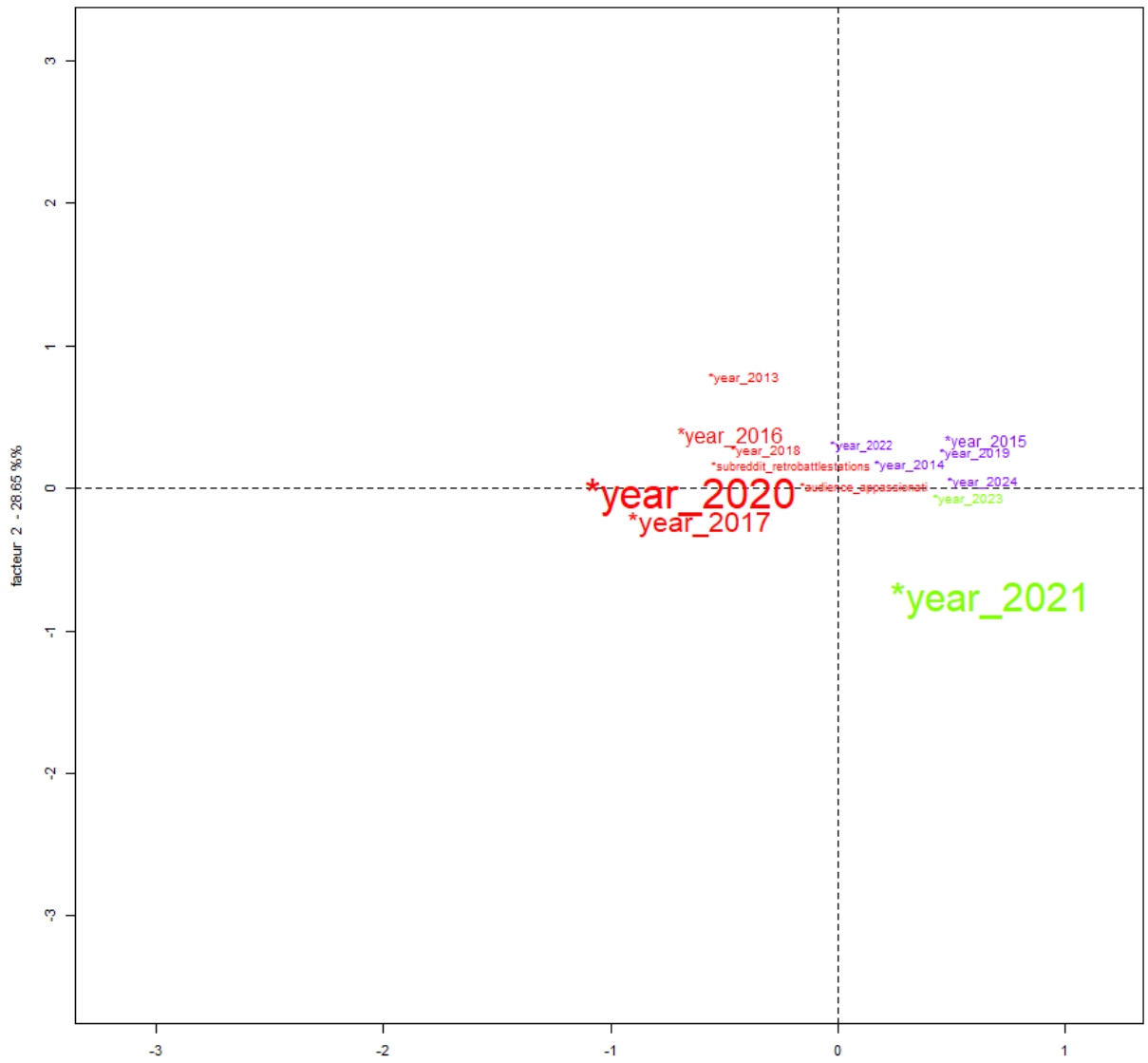


Figure 68 Analisi delle Corrispondenze Multiple (AFC) – Metadati temporali Corpus Reddit

Anche l'AFC per anni suggerisce una progressiva evoluzione terminologica: i termini più recenti (2021, 2023, 2024) si distanziano dal blocco centrale (2013–2019), mostrando un possibile slittamento tematico o stilistico nei contenuti pubblicati.

In sintesi, le rappresentazioni grafiche mostrano come il discorso informatico all'interno della comunità Reddit sia polifonico e stratificato: da un lato tecnico-specialistico, dall'altro culturale, narrativo o affettivo. Questa complessità si riflette in modo netto nelle mappe e nei dendrogrammi, che costituiscono strumenti efficaci per mettere in evidenza le differenze terminologiche interne al corpus.

Dopo aver analizzato i corpora storici e comunitari attraverso la combinazione di dendrogramma, AFC e test di specificità, si passa ora al corpus museale, costituito da testi descrittivi e informativi provenienti da siti di musei o archivi digitali dedicati alla storia dell'informatica. Questo corpus

presenta caratteristiche distintive, sia sul piano stilistico che contenutistico, legate alla sua finalità divulgativa e istituzionale.

A differenza degli altri insiemi, per il corpus museale IRaMuTeQ ha generato un'unica rappresentazione grafica: il dendrogramma delle classi lessicali:

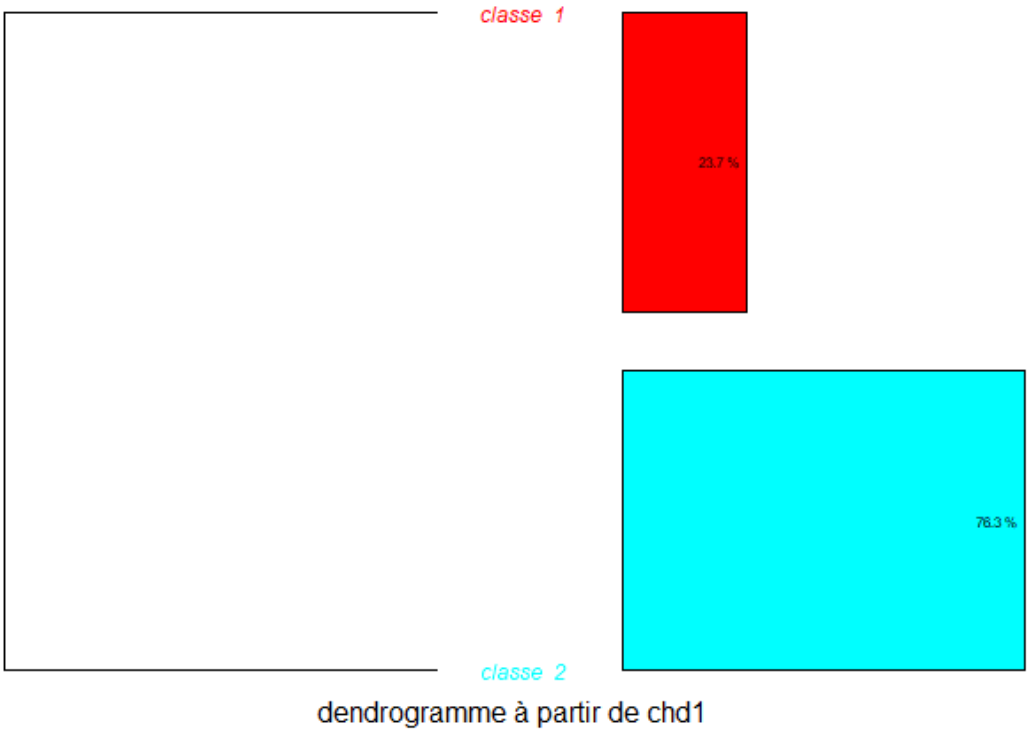


Figure 69 Dendrogramma semplificato della classificazione lessicale automatica (CHD1) – Corpus museale

L'assenza di un piano fattoriale potrebbe dipendere dalla composizione e dimensione del corpus, che rispetto agli altri presenta una minore varietà strutturale nei segmenti analizzati oppure una più marcata polarizzazione tematica.

Tuttavia, anche in assenza di una visualizzazione sul piano fattoriale, l'analisi di specificità tramite *test del chi-quadrato* consente di ottenere indicazioni significative sulla struttura terminologica del corpus.

In particolare, lo strumento denominato ***Segments de texte caractéristiques*** restituisce, per ciascuna classe individuata dalla classificazione automatica, i segmenti di testo che contengono i lemmi più tipici della classe, insieme a una serie di indicatori statistici: la frequenza del lemma all'interno della classe, il valore del chi quadrato (χ^2) che esprime la forza dell'associazione tra il lemma e la classe, e il relativo p-value, che ne indica la significatività statistica.

Questo approccio permette di interpretare la composizione semantica di ciascuna classe in modo approfondito, andando oltre la semplice visualizzazione gerarchica del dendrogramma e offrendo

evidenze quantitative a supporto delle differenziazioni terminologiche tra i sottoinsiemi del corpus museale.

1 Classe 1 555/2345 23.67% | 2 Classe 2 1790/2345 76.33%

n...	eff. s.t.	eff. total	pourcentage	chi2	Type	forme	p
0	389	435	89.43	1278.305	nom	desktop	< 0,0001
1	297	360	82.5	814.819	nom	device	< 0,0001
2	238	250	95.2	792.587	nom	output	< 0,0001
3	223	228	97.81	768.418	nom	transfer	< 0,0001
4	221	260	85.0	608.882	nom	storage	< 0,0001
5	187	206	90.78	562.995	nr	philips	< 0,0001
6	219	269	81.41	560.842	nom	commodore	< 0,0001
7	199	264	75.38	440.339	nom	apple	< 0,0001
8	177	221	80.09	429.966	nom	portable	< 0,0001
9	233	372	62.63	371.615	nom	data	< 0,0001
10	108	122	88.52	299.654	nom	input	< 0,0001
11	77	88	87.5	206.214	nr	tandy	< 0,0001
12	139	231	60.17	189.022	nom	console	< 0,0001
13	93	125	74.4	188.111	nr	atari	< 0,0001
14	464	1373	33.79	188.048	nom	computer	< 0,0001
15	72	85	84.71	181.887	nom	macintosh	< 0,0001
16	55	55	100.0	181.648	nom	tulip	< 0,0001
17	52	52	100.0	171.515	nr	generiek	< 0,0001
18	128	227	56.39	148.942	nom	monitor	< 0,0001
19	43	47	91.49	122.116	nr	compaq	< 0,0001
20	74	112	66.07	117.065	nom	micro	< 0,0001
21	52	66	78.79	114.211	nr	trs	< 0,0001
22	45	53	84.91	112.561	nr	amiga	< 0,0001
23	33	34	97.06	102.861	nom	multimedia	< 0,0001
24	119	238	50.0	101.668	nom	drive	< 0,0001
25	62	92	67.39	101.332	nom	accessory	< 0,0001
26	31	31	100.0	101.321	nr	televisie	< 0,0001
27	61	92	66.3	96.357	nr	sony	< 0,0001
28	59	91	64.84	88.815	nr	packard	< 0,0001

Figure 70 Segments de texte caractéristiques del corpus museale – Classe 1

1 Classe 1 555/2345 23.67% | 2 Classe 2 1790/2345 76.33%

n...	eff. s.t.	eff. total	pourcentage	chi2	Type	forme	p
0	256	263	97.34	72.349	nom	machine	< 0,0001
1	224	230	97.39	62.598	nr	electronic	< 0,0001
2	184	186	98.92	57.076	nom	manufacture	< 0,0001
3	222	231	96.1	55.445	nom	unit	< 0,0001
4	186	193	96.37	46.753	nom	card	< 0,0001
5	149	151	98.68	44.597	nom	punch	< 0,0001
6	171	177	96.61	43.574	nom	design	< 0,0001
7	135	136	99.26	42.025	nom	print	< 0,0001
8	120	120	100.0	39.213	nom	university	< 0,0001
9	114	114	100.0	37.153	nom	manual	< 0,0001
10	123	126	97.62	33.397	nom	microcomputer	< 0,0001
11	102	102	100.0	33.064	nom	serial	< 0,0001
12	106	107	99.07	32.071	nom	part	< 0,0001
13	94	94	100.0	30.362	adj	numb	< 0,0001
14	126	131	96.18	30.264	nom	user	< 0,0001
15	107	109	98.17	30.161	nom	software	< 0,0001
16	99	100	99.0	29.708	nr	ferranti	< 0,0001
17	90	90	100.0	29.019	ver	develop	< 0,0001
18	86	86	100.0	27.68	nr	manchester	< 0,0001
19	86	86	100.0	27.68	nom	cable	< 0,0001
20	110	114	96.49	26.953	nom	work	< 0,0001
21	90	91	98.9	26.692	nom	circuit	< 0,0001
22	90	91	98.9	26.692	nom	original	< 0,0001
23	83	83	100.0	26.679	nom	store	< 0,0001
24	88	89	98.88	26.025	nom	company	< 0,0001
25	101	104	97.12	26.018	nom	program	< 0,0001
26	81	81	100.0	26.013	nom	paper	< 0,0001
27	81	81	100.0	26.013	nom	instruction	< 0,0001
28	298	342	87.13	25.86	nom	system	< 0,0001

Figure 71 Segments de texte caractéristiques del corpus museale – Classe 2

Nel caso specifico, il dendrogramma ha individuato due classi principali:

- **Classe 1** (23,67%): connotata da lemmi legati a prodotti, brand e oggetti fisici (es. *desktop, device, commodore, storage, tandy, packard*), suggerendo un orientamento più concreto e legato all'oggetto materiale della memoria tecnologica.
- **Classe 2** (76,33%): dominata invece da termini relativi a concetti tecnici, processi produttivi o contesti istituzionali e storici (*machine, manufacture, electronic, punch, university, instruction, ferranti*), con un lessico orientato alla spiegazione dei meccanismi storici e delle infrastrutture culturali legate all'informatica.

L'analisi mostra una netta separazione tra i due filoni terminologici: da un lato la dimensione oggettuale e collezionistica, dall'altro quella tecnico-storica, più esplicativa e orientata alla mediazione del sapere. La presenza di un elevato numero di forme con significatività statistica elevata ($p < 0,0001$) conferma la solidità della classificazione ottenuta. Il *p-value* (o valore p) è un indice statistico utilizzato per valutare la significatività di un risultato ottenuto in un test d'ipotesi. Esso rappresenta la probabilità. Dal punto di vista statistico, la significatività dell'associazione tra il lemma e la classe è molto elevata, mentre IRaMuTeQ, coerentemente con le soglie comunemente accettate in ambito statistico, non considera più rilevanti i lemmi con $p > 0,05$.

Questa stessa metodologia si è rivelata altrettanto efficace nell'analisi del **sottocorpus 2**, in cui il numero di classi individuate è risultato maggiore e la distribuzione tematica più articolata. Anche in questo caso, lo strumento dei *Segments de texte caractéristiques* ha consentito di isolare lemmi fortemente rappresentativi all'interno delle singole classi, fornendo una chiave di lettura approfondita delle traiettorie lessicali.

1 Classe 1 5896/39106 15.08%	2 Classe 2 7176/39106 18.35%	3 Classe 3 5075/39106 12.98%	**4 Classe 4 5979/39106 15.29%**	5 Classe 5 5318/39106 13.6%	6 Classe 6 9662/39106 24.71%

n...	eff. s.t.	eff. total	pourcentage	chi2	Type	forme	p
0	377	581	64.89	1120.204	adj	game	< 0,0001
1	181	207	87.44	836.429	nom	hobbyist	< 0,0001
2	565	1317	42.9	802.258	nom	personal	< 0,0001
3	288	467	61.67	785.039	nom	social	< 0,0001
4	274	469	58.42	681.879	nom	home	< 0,0001
5	561	1519	36.93	571.576	nom	user	< 0,0001
6	182	321	56.7	428.491	nom	apple	< 0,0001
7	251	544	46.14	405.399	nr	1980s	< 0,0001
8	86	109	78.9	341.479	nom	hobby	< 0,0001
9	95	133	71.43	324.745	nr	cultural	< 0,0001
10	215	482	44.61	323.844	ver	understand	< 0,0001
11	146	268	54.48	319.972	ver	argue	< 0,0001
12	387	1143	33.86	313.462	nom	article	< 0,0001
13	100	153	65.36	297.324	nom	club	< 0,0001
14	138	264	52.27	280.697	nom	magazine	< 0,0001
15	63	75	84.0	273.918	nom	enthusiast	< 0,0001
16	244	630	38.73	271.657	nr	1970s	< 0,0001
17	133	254	52.36	271.303	nom	context	< 0,0001
18	65	81	80.25	264.437	nom	amateur	< 0,0001
19	129	250	51.6	256.137	nr	popular	< 0,0001
20	181	424	42.69	248.462	nom	historian	< 0,0001
21	164	374	43.85	237.832	nom	practice	< 0,0001

Figure 72 Segments de texte caractéristiques del sottocorpus 2 – Classe 1

Un esempio emblematico è rappresentato dalla **Classe 4**, dominata dal lemma *game*, che raggiunge un valore di chi² pari a 1120,20, il più alto rilevato tra tutte le classi del sottocorpus e nettamente superiore persino rispetto al secondo termine più caratteristico, che si attesta a 836,429. Tale valore, evidenzia la centralità del concetto di ***game*** nella costruzione della classe e suggerisce una forte specificità tematica. I segmenti associati a questa classe descrivono contesti di home computing, club informatici, riviste specializzate e pratiche ludiche, restituendo un quadro in cui l'informatica viene narrata come esperienza sociale, culturale e personale.

Di seguito esempi del concetto *game* nel concordancier:

Figure 73 Segments de texte caractéristiques– Classe 4, Sottocorpus 2

Il confronto con le classi del corpus museale mostra, dunque, una chiara divergenza nei regimi discorsivi: se il lessico museale tende a oggettivare e storicizzare l'informatica, negli altri corpora, come quello di *Reddit* e, come appena dimostrato, per alcuni aspetti il *sottocorpus 2* che è invece di tipo scientifico e divulgativo, emerge una dimensione relazionale e narrativa, più vicina all'uso quotidiano e alle pratiche comunitarie. In entrambi i casi, però, l'analisi di specificità e la rappresentazione dei segmenti tipici si dimostrano strumenti cruciali per rilevare le differenziazioni terminologiche e per comprendere come il linguaggio si adatti ai diversi contesti di produzione discorsiva.

3.2.2 TROPES

3.2.2.1 Analisi semantica

Per approfondire ulteriormente la varietà terminologica e concettuale del dominio informatico, ho utilizzato il software Tropes (Molette, Landre, 2021), applicandolo ai due sottocorpora *computer1*, *computer2*, *Reddit* e *Musei*, selezionati in formato .txt ed esportati da IRaMuTeQ tramite la funzione *concordancier*. Questo ha permesso di salvare sottocorpora in formato .txt e riutilizzarli in altri software.

Tropes consente di ricostruire reti semantiche e scenari concettuali a partire dalla classificazione automatica delle unità lessicali, evidenziando le relazioni tra lemmi e domini tematici.

A seguito dell'elaborazione con Tropes, è stato possibile ottenere per ciascun sottocorpus due classificazioni semantiche distinte. La prima, denominata Reference 1, raccoglie le categorie concettuali generali associate ai testi, mentre la seconda, Reference 2, fornisce un dettaglio più fine delle unità lessicali, aggregandole in insiemi semantici specifici.[31] Le due rappresentazioni, pur basandosi sugli stessi dati testuali, evidenziano livelli diversi di granularità: Reference 1 è utile per una visione d'insieme del campo semantico, Reference 2 permette invece di cogliere con maggiore precisione le sfumature lessicali che attraversano il discorso informatico.

Di seguito si riportano le due classificazioni per entrambi i sottocorpora (computer1 e computer2):

Sottocorpus computer1

Reference 1 (categorie generali)

computer_science	255704
communication	83419
time	58257
system	47335
device	41892
language	29667
education	26994
business	23536

[31] In Tropes, le References rappresentano il contesto semantico: aggregano i principali sostantivi del testo in classi equivalenti e sono visualizzate su tre livelli (Reference fields 1, Reference fields 2, References). La visualizzazione delle References e delle loro relazioni porta "al centro del discorso" attori, oggetti e concetti in ordine decrescente di importanza; elementi pertinenti possono essere fissati nello Scenario per una classificazione personalizzata. (*cfr. Documentazione Tropes, sezione "References & Scenario"->"Explain").*

electronics 20019
north_america 16675
person 15451
control 15345
characteristic 14602
location 13462
technology 12984
math 12787

Reference 2 (categorie più specifiche)

computer 153638
system 47333
time 44639
communication 43629
software 24637
computer_science 22953
mechanism 18589
person 17678
computer_hardware 17107
u_s_a 16166
telecommunication 15032
electronic_device 15029
month 13773
device 13295
processor 12409
area 11284
control 11147
language 10513
organization 10313
information 9910
colleges_and_universities 9736
grammar 9370
design 9156
work 8148

computer_industry 8091
business 7920
science 7629
expense 7379
programming_language 7320
text 7234
sameness 7168
exploitation 6573
location 6450
difference 6393
furniture 5973
technology 5959
technique 5938
method 5929
way 5764
equipment 5636
company 5319
mistake 5011
cognition 5007
newspapers 4941
bank 4415
money 4368
database 4278
math 4242
important_person 4217
students 3985
evaluation 3805
body 3658
test 3459
security 3389
weight 3377
industry 3311
study 3227
spatial_property 3161

french_text	3021
production	3002
container	2998
operating_system	2955
speed	2873
connection	2805
environment	2787
electronic_equipment	2764
graphic_art	2651
projects	2643
tool	2565
education	2542
architecture	2463
transport	2429
electrical_device	2378
logic	2328
music	2262
people	2254
courses	2209
state	2134
air_travel	2130
characteristic	2070
perceptibility	2051
postal_service	2035
medicine	1861
quantity	1840
man	1835
social_gathering	1833
press	1757
trustworthiness	1714
topology	1655
failure	1646
research	1583
army	1570

administration 1554
advantage 1545
movie 1496
physical_property 1495
office 1495
electricity 1487
game 1477

Sottocorpus computer 2

Reference 1

computer_science 71679
time 16298
education 14904
business 12911
communication 12612
device 11315
language 8085
history 7339
social_group 6295
Europe 6250
north_america 6043
science 5847
organization 5284
technology 5070
foreign_text 4765
system 4567
electronics 3808
media 3652
fight 2992
location 2923
industry 2815

cognition 2611
work 2415
money 2375
person 2183
transport 2157
math 2155
security 1998
society 1973
behavior 1960
art 1926
politics 1851
entertainment 1744
characteristic 1505
asia 1411
control 1376
production 1283
law 1205
body 1129
animal 1124
feeling 0999
world 0979
electricity 0933
furnishings 0858
ecc

Reference 2

computer 44281
time 9770
colleges_and_universities 8111
computer_science 7898
mechanism 7382
history 6987
computer_industry 6859
month 6617

software 5495
u_s_a 5293
system 4567
organization 4130
business 4112
science 3934
technology 3894
company 3762
person 3518
italian_text 3052
text 2927
telecommunication 2833
industry 2743
west_europe 2706
communication 2638
exploitation 2626
computer_hardware 2610
grammar 2530
electronic_device 2402
important_person 2366
work 2169
bank 2081
information 2000
society 1973
location 1942
device 1907
language 1844
design 1801
programming_language 1726
projects 1578
united_kingdom 1563
security 1532
german_text 1510
area 1424

money 1422
newspapers 1419
way 1394
processor 1334
production 1302
ecc…

Nel *sottocorpus computer1*, le categorie più frequenti rientrano nell'ambito strettamente informatico: *computer science*, *system*, *device*, *software*, *processor*, *control*, *language*, *technology*, *computer hardware*. Emergono anche riferimenti a contesti produttivi o istituzionali, come *business*, *company*, *organization*, *colleges and universities*. Il lessico è fortemente tecnico e sistemico, con un'elevata presenza di concetti ingegneristici e funzionali. Termini come *programming language*, *telecommunication*, *database*, *electronic device* contribuiscono a delineare un dominio concettuale centrato sulla costruzione, gestione e formalizzazione dei sistemi informatici. La presenza di categorie come *sameness*, *difference*, *logic* e *control* suggerisce inoltre un impianto teorico e classificatorio tipico della trattazione specialistica.
Nel *sottocorpus computer2* il quadro semantico si amplia. Accanto alle categorie ancora legate al dominio tecnico (come *computer science*, *software*, *system*, *mechanism*, *technology*), emergono con forza nuove aree concettuali: *history*, *media*, *society*, *art*, *behavior*, *politics*, *entertainment*, *social group*. Queste categorie indicano una progressiva apertura della trattazione informatica verso ambiti culturali, storici e sociali. L'informatica non viene più soltanto descritta nei suoi aspetti funzionali, ma anche narrata, contestualizzata, interpretata come fenomeno storico e culturale. Inoltre, la presenza di categorie geolinguistiche (*italian text*, *german text*, *west europe*, *united kingdom*) e mediali (*newspapers*, *press*) suggerisce un uso più discorsivo e localizzato del lessico informatico, spesso connesso a pratiche comunicative e interpretative.

Un ulteriore elemento rilevante è dato dalla ricorrenza, nel corpus, di forme linguistiche che esprimono approssimazione e incertezza, come *almost*, *probably*, *possibly*, *perhaps*, *approximately*. Questo tipo di lessico, rilevato in particolare nei testi storici o speculativi, è indicativo di un registro riflessivo e probabilistico, in contrasto con l'assertività tipica dei testi tecnici. Si osservano, ad esempio, formulazioni come "*which is **probably** the aspect of the computer revolution that has had the most direct effect on the everyday lives of ordinary people*" o "*computers have been in use for more than eight decades, personal computers for **approximately** 35 years*". Nel 2013 compaiono inoltre espressioni come *"...the growing commitment of IBM to R&D and **perhaps** even the early*

*recognition that electronics and computers **might** be critical to its future*" (IEEE Annals of the History of Computing, 2013) e *"computer scientists are **perhaps** justified in treating the computer solely in terms of the platonic*" (ibid.). Tali formulazioni, che attenuano l'enfasi dichiarativa e introducono una valutazione prudente o stimata, testimoniano l'inserimento dell'informatica in ambiti di ricostruzione storica, ipotesi epistemologiche o previsioni scientifiche, dove il discorso assume forme narrative e metadiscorsive.

L'integrazione dei due ulteriori corpora consente di precisare la direzione della variazione.

Corpus Reddit

Reference 1 (categorie generali)

computer_science 3223
time 1248
device 0844
language 0798
communication 0552
electronics 0546
business 0247
entertainment 0244
location 0215
food 0190
social_group 0189
europe 0160
money 0148
work 0141
feeling 0127
media 0124
beverage 0121
family 0121
electricity 0119
transport 0117
photograph 0112
system 0106

education 0104
clothing 0096
quantity 0092
art 0087
housing 0078
animal 0073
north_america 0072
plant 0070
body 0062
cognition 0058
fight 0058
man 0052
technology 0051
space 0051
child 0047
characteristic 0047
industry 0046
music 0043
sport 0043
success 0042
person 0040
ecc.

Reference 2
computer 1335
time 1145
computer_industry 0603
computer_hardware 0432
mechanism 0343
electronic_device 0289
device 0270
processor 0265
operating_system 0256
area 0195

game 0194
food 0171
software 0166
container 0155
work 0140
beverage 0127
text 0120
east_europe 0116
electrical_device 0113
telecommunication 0110
photograph 0110
equipment 0109
system 0106
month 0106
money 0103
electronic_equipment 0101
location 0101
way 0088
video 0077
fashion_industry 0075
business 0074
people 0074
whatchamacallit 0071
internet_provider 0068
parent 0060
body 0060
memory_unit 0059
japanese_trust 0058
connection 0056
u_s_a 0055
team 0055
ecc.

Nella **Reference 1** del corpus Reddit le categorie più frequenti includono *computer_science* (3223), *time* (1248), *device* (844), *language* (798), *communication* (552), *electronics* (546), con una cospicua quota di domini esperienziali/valutativi (*entertainment* (244), *feeling* (127), *media* (124), *family* (121), *food* (190), *beverage* (121), *clothing* (96), *sport* (43)). In **Reference 2** oltre a *computer* (1335) e *time* (1145), compaiono con forza etichette tecnico-oggettuali (*computer_industry* (603), *computer_hardware* (432), *processor* (265), *operating_system* (256), *memory* (59), ma affiancate da molte categorie di vita quotidiana e giudizio soggettivo (*game* (194), *food* (171), *beverage* (127), *money* (103), *love* (25), *friendship* (26), *fashion_industry* (75), *press* (27), *web* (26), *social_network* (9), *social_occasion* (12)). Reddit mostra, appropriatamente, un registro comunitario e conversazionale: al lessico tecnico si sovrappongono campi semantici valutativi/affettivi e di consumo (cibo, bevande, moda, intrattenimento), indizio di pratiche discorsive che combinano uso, esperienza, opinione e consiglio (how-to, acquisti, confronti). Il computer è trattato come oggetto d'uso e oggetto culturale nel quotidiano; le categorie *game/game_consoles* e *press/web/social_network* segnalano la forte integrazione con la cultura pop e i media.

Corpus museale

Reference 1 (categorie generali)

computer_science 8478

device 2777

entertainment 2049

electronics 1519

business 1104

communication 0628

education 0624

europe 0572

time 0501

social_group 0354

system 0350

plant 0350

electricity 0307

clothing 0297

media 0281

language 0272

music 0256

transport 0239
north_america 0233
fight 0211
science 0209
animal 0205
location 0169
person 0158
production 0157
politics 0152
math 0143
body 0140
technology 0137
industry 0135
money 0132
world 0129
power 0123
characteristic 0112
security 0111
substance 0103
organization 0102
quantity 0102
space 0101
creation 0096
ecc.

Reference 2
computer 4025
computer_industry 2519
mechanism 1374
game 1197
device 1030
game_consoles 0741
computer_hardware 0727
electronics_industry 0631

japanese_trust 0519
software 0489
time 0443
electronic_device 0421
united_kingdom 0417
container 0417
processor 0382
system 0350
electronic_equipment 0285
electrical_device 0284
fashion_industry 0280
company 0271
text 0225
telecommunication 0225
music 0218
u_s_a 0199
computer_science 0191
person 0189
equipment 0187
colleges_and_universities 0187
science 0178
business 0167
production 0158
schools 0157
design 0151
important_person 0148
projects 0147
bulbous_plant 0136
industry 0132
television 0129
world 0129
area 0124
ecc.

Per quanto riguarda i risultati Tropes del **Reference 1** del corpus museale accanto a *computer_science* (8478) emergono *device* (2777), *entertainment* (2049), *electronics* (1519), *business* (1104). Si osserva inoltre la presenza di domini descrittivo-materiali, che includono termini legati alla composizione, produzione e classificazione degli oggetti (ad es. *device* (2777)*, electronics* (1519)*, production* (157)*, industry* (135)*, equipment* (187)), e tracce storico-documentarie (*history* (49)*, publishing* (49)). Mentre nella **Reference 2** dominano denominazioni oggettuali e di filiera: *computer* (4025), *computer_industry* (2519), *mechanism* (1374), *game* (1197), *game_consoles* (741)*, computer_hardware* (727)*, electronics_industry* (631)*, processor* (328)*, system* (350)*, electronic_device* (284), *electronic_equipment* (285), oltre a marcatori di provenienza/istituzionali (*united_kingdom* (417)*, u_s_a* (199)*, company* (271)*, colleges_and_universities* (187)*, students* (47)*, projects* (29)). Nel corpus museale, dunque, il computer è, anche in questo caso appropriatamente, oggetto materiale da catalogare, esporre e attribuire (marca, modello, industria, Paese), con forte attenzione a componenti, meccanismi e provenienze. Il lessico è descrittivo-classificatorio e istituzionale; *game/game_consoles* non è "ludico" in senso Reddit, ma una categoria espositiva.

All'interno del software Tropes, un'indicazione interessante deriva dalla categoria **Scenario**, che serve a rappresentare il contesto generale in cui si svolgono le azioni o in cui sono collocati i concetti di un testo. Non si tratta solo di un luogo fisico, ma anche di ambienti concettuali o tematici come ambiti scientifici, tecnologici, storici ecc.

Lo Scenario consente di personalizzare e stabilizzare la classificazione semantica[32]: quando una Reference, un verbo o un aggettivo viene incluso nello Scenario, esso viene riconosciuto come centrale e risulta evidenziato nei risultati (checked), rendendo più chiara l'organizzazione concettuale del testo. L'uso di uno Scenario personalizzato è quindi particolarmente utile in corpora eterogenei come quelli considerati in questa ricerca, perché permette di mantenere coerenti le categorie interpretative anche quando i discorsi cambiano registro (tecnico, storico, comunitario, museale).

[32] (cfr. Documentazione ufficiale software Tropes, sezione "Scenario"->"Explain")

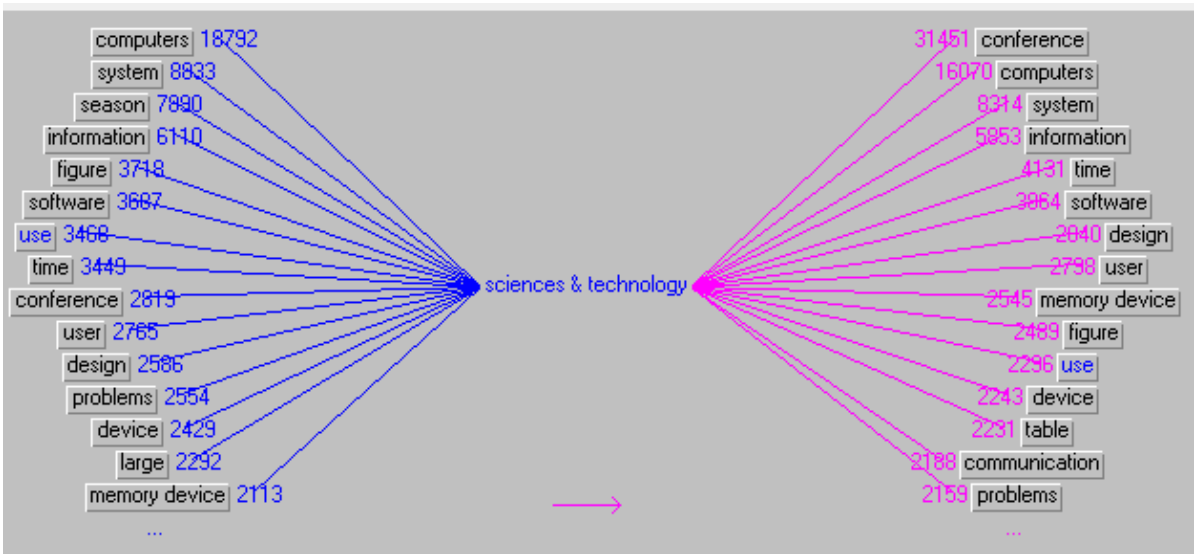


Figure 74 Rete semantica della categoria "sciences & technology" nel sottocorpus 1 (1961–1984).

Nel corpus 1961–1984, la rete semantica di *sciences & technology* si struttura attorno a termini funzionali come *system, software, information, device, design*, che denotano un lessico tecnico-operativo e un uso strumentale del linguaggio tecnologico

In questo scenario si colloca la rete semantica costruita attorno alla categoria *automation & robotics*, particolarmente significativa per la sua trasversalità e polivalenza concettuale:

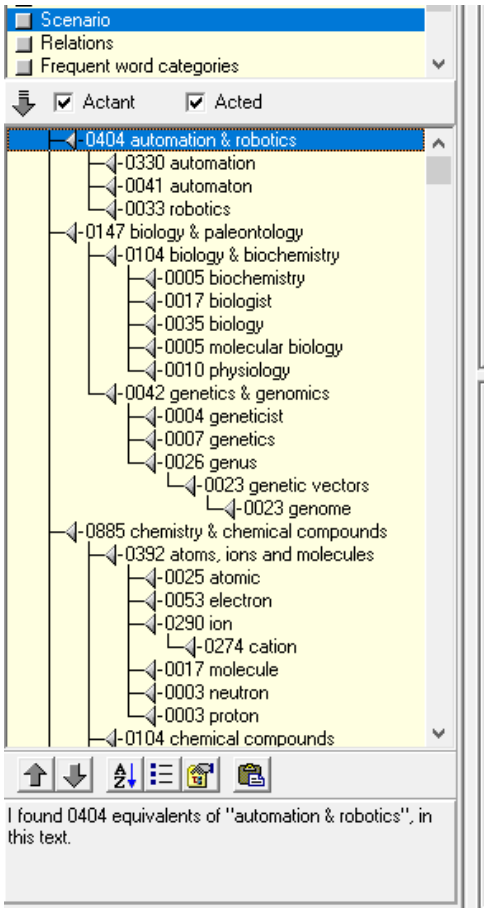


Figure 75 Gerarchia semantica dei concetti individuati nel sottocorpus 1 tramite Tropes. Il dominio "automation & robotics" (codice 0404) comprende i termini automation, automaton e robotics.

L'analisi semantica mostra, infatti, che *automation & robotics* si collega sia a discipline scientifiche (*biologia, genetica, chimica*), sia a contesti applicativi (*informatica, ingegneria, software, office, university*), configurandosi come un nodo concettuale ibrido, capace di unire domini tradizionalmente separati. A sinistra e a destra si vedono le relazioni concettuali più forti:

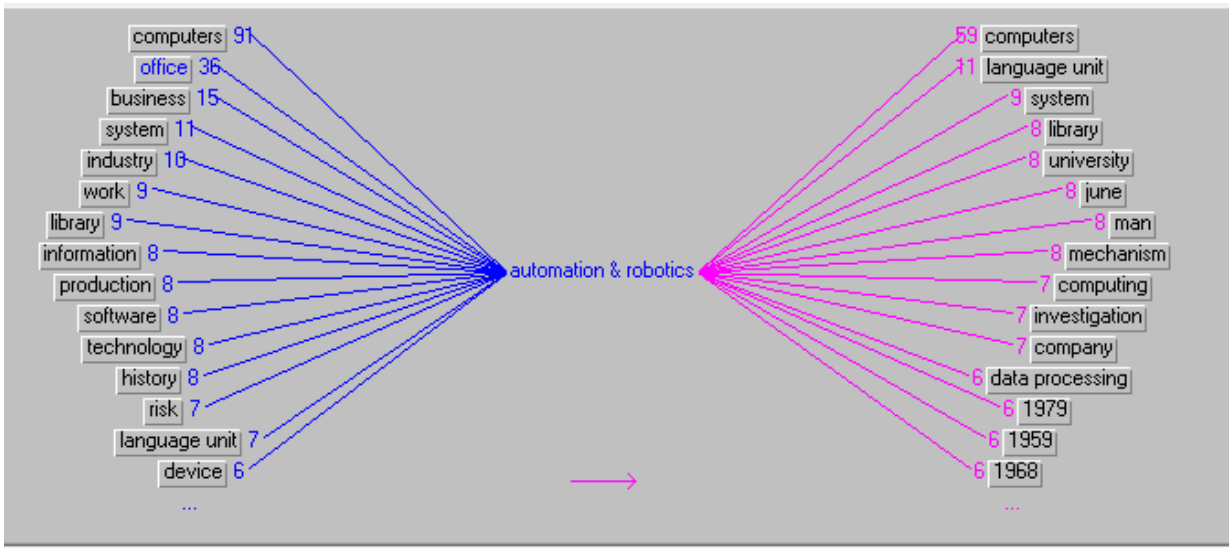


Figure 76 Rete semantica dei co-occorrenziali del concetto "automation & robotics" nel sottocorpus 1

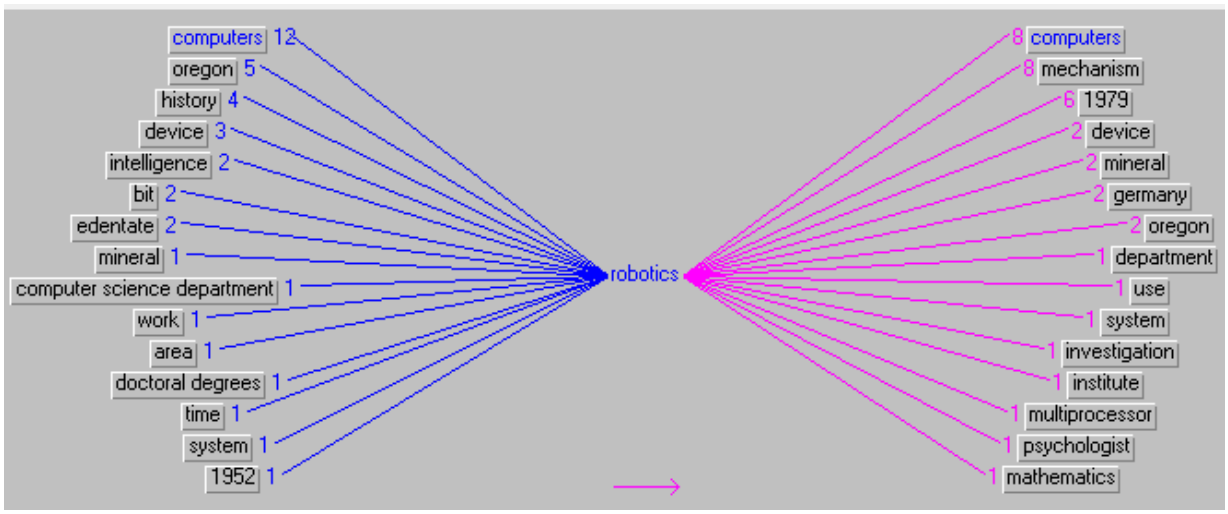


Figure 77 Rete semantica delle relazioni co-occorrenziali per il termine "robotics" nel sottocorpus 1

Tale rete concettuale è visibile nelle rappresentazioni generate da Tropes, dove i legami tra i termini non si limitano a relazioni tematiche, ma includono anche affinità discorsive e culturali.

Di seguito i contesti d'uso del termine *robotics*:

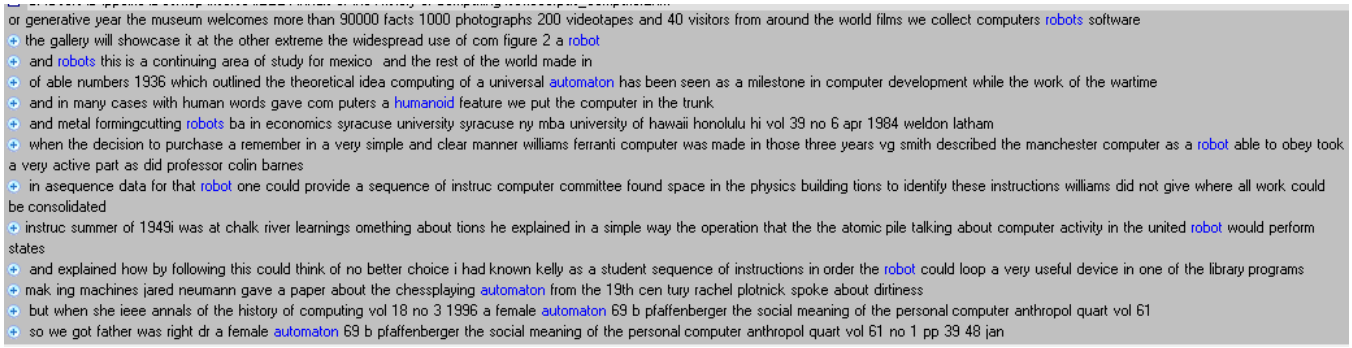


Figure 78 Estratti di concordanze nel sottocorpus 1 per i termini "robot", "robots" e "automaton"

Nel corpus 1985–2024, la rete semantica si amplia e si riorganizza.

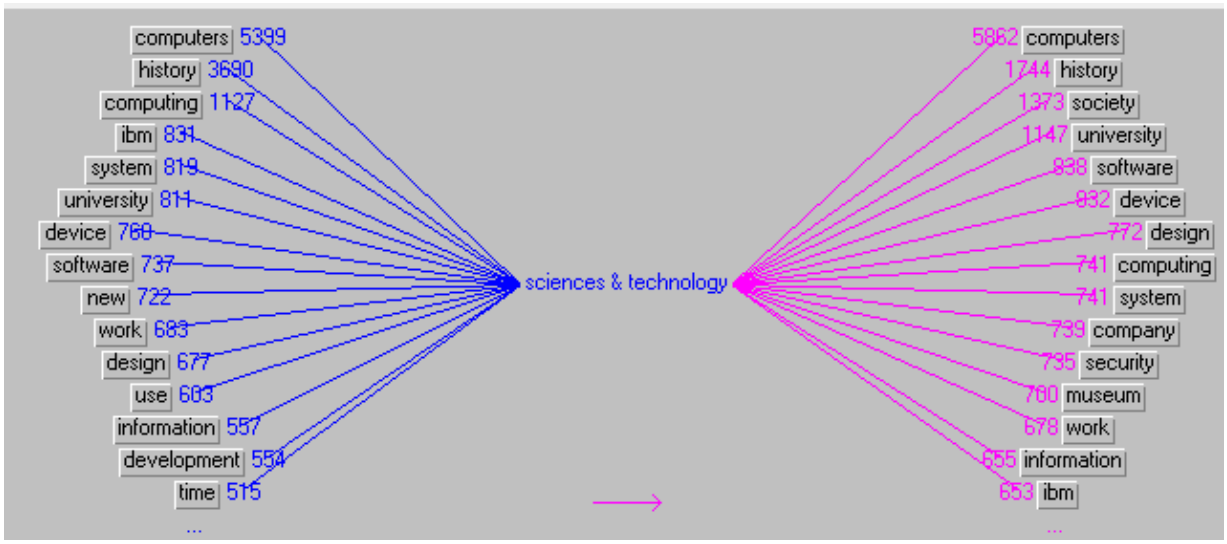


Figure 79 Rete semantica della categoria "sciences & technology" nel sottocorpus 2 (1985–2024).

L'evoluzione terminologica individuata attraverso IRaMuTeQ trova ulteriore conferma nell'analisi Tropes.

Nel sottocorpus più recente, la stessa categoria si ristruttura: accanto ai termini tecnici compaiono *history, society, museum, university, design, computing.*

Questo spostamento segnala un processo di culturalizzazione del discorso scientifico, in cui la tecnologia non è più soltanto oggetto di produzione o ricerca, ma anche di memoria (*history*), analisi e riflessione (*university, society*) e patrimonializzazione (*museum*).

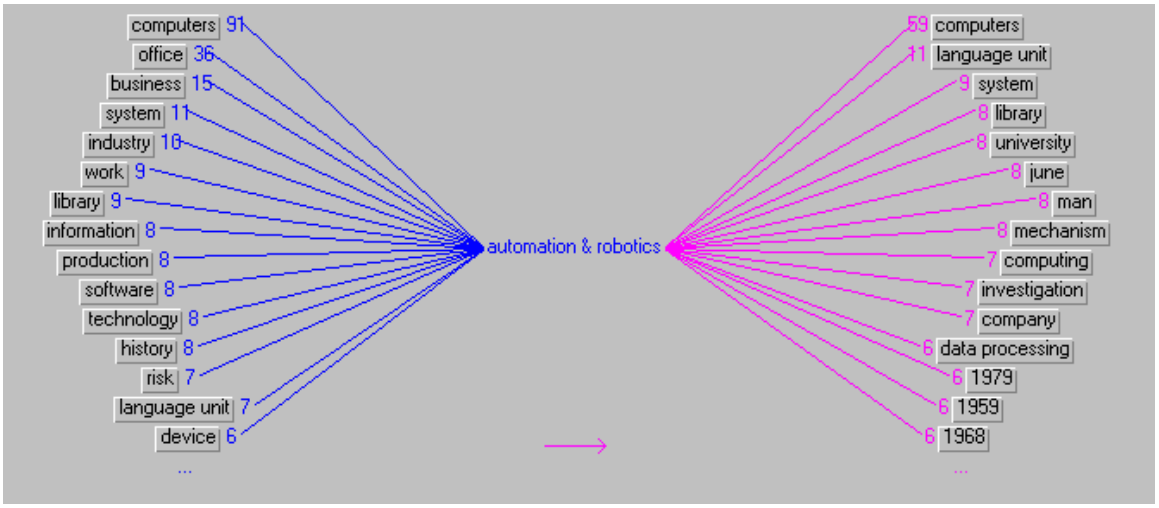


Figure 80 Rete semantica delle relazioni co-occorrenziali per il termine "robotics" nel sottocorpus 2

Questo passaggio avviene anche nella categoria *automation & robotics*, che nel primo corpus si associa a *computers, office, business, production*, mentre nel secondo corpus si collega a *university, library, company, data processing*.

Questa eterogeneità semantica rappresenta un tratto distintivo del lessico informatico contemporaneo, che non si limita a descrivere strumenti o processi, ma si adatta e si ibrida con diversi regimi discorsivi.

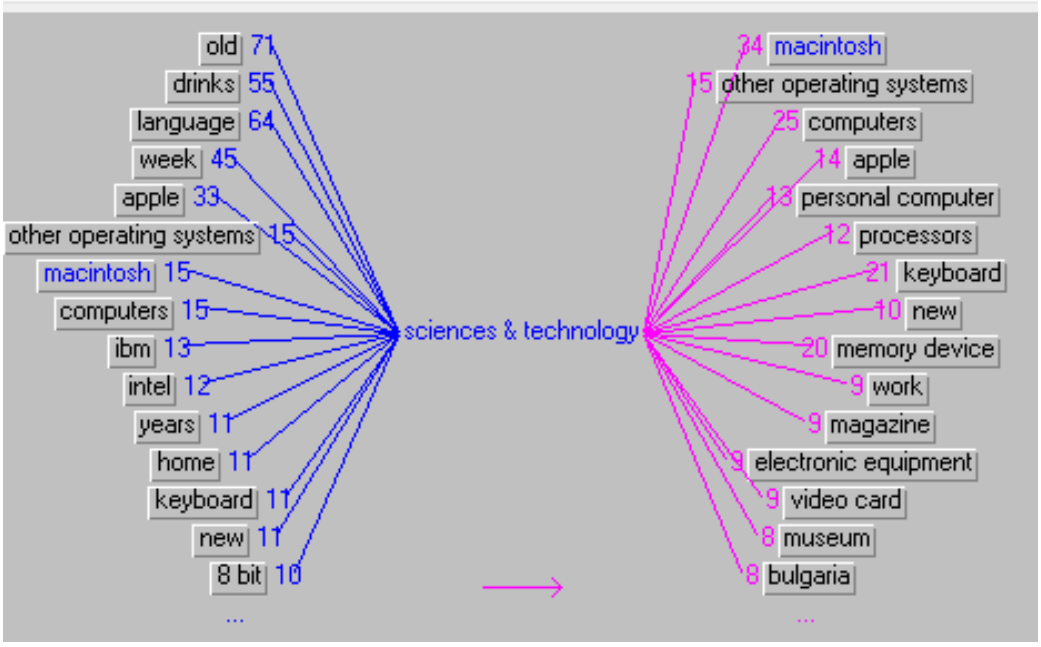


Figure 81 Rete semantica della categoria "sciences & technology" nel corpus Reddit.

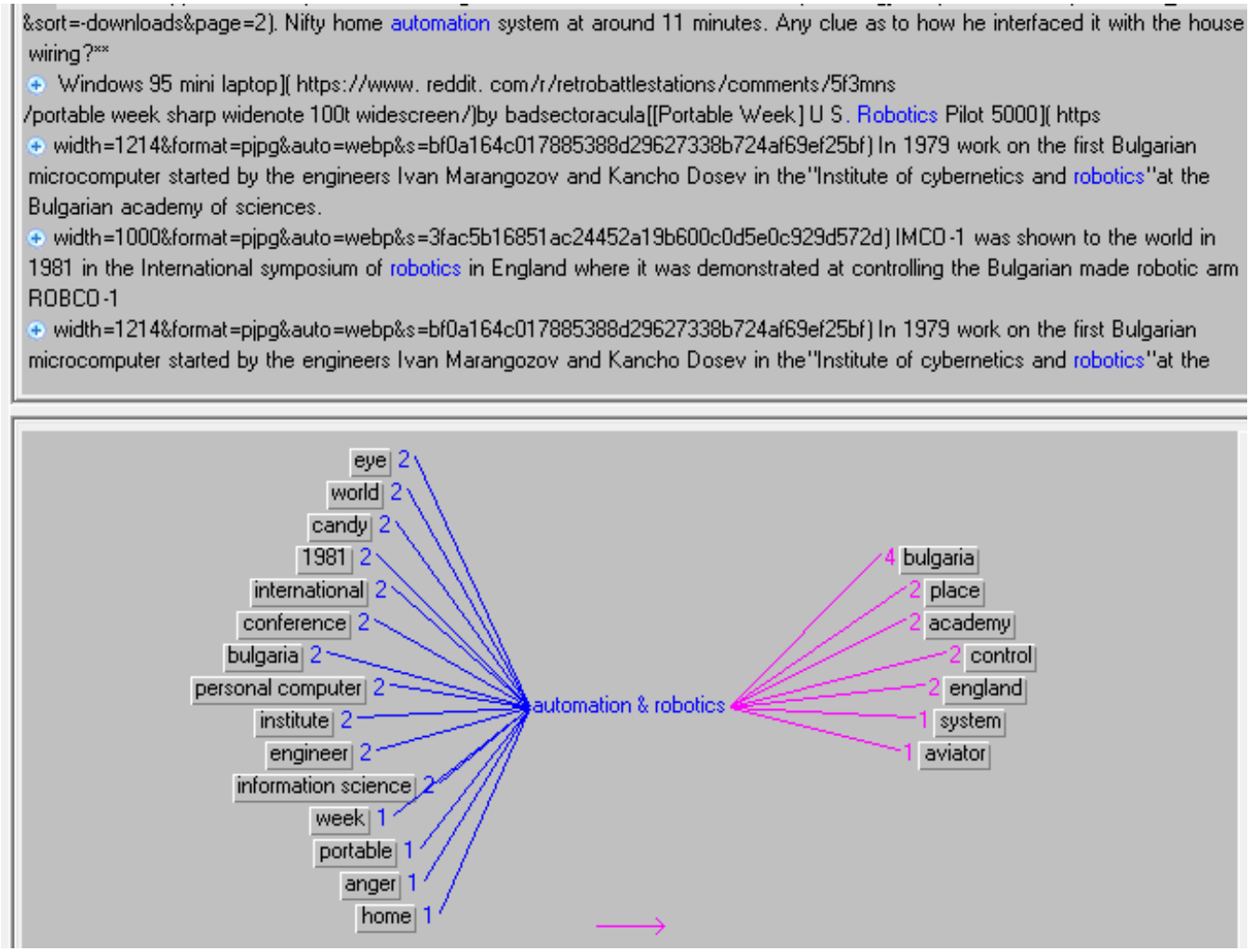


Figure 82 Rete semantica delle relazioni co-occorrenziali e contesti per il termine "robotics" nel corpus Reddit.

Nel corpus Reddit, invece, la categoria *automation & robotics* assume una configurazione semantica intermedia tra la dimensione storica e quella esperienziale. Le reti Tropes mostrano co-occorrenze quali *personal computer, information science, portable, international conference, institute, bulgaria, control*, *anger* che rimandano a contesti di discussione tecnica, rievocazione storica e condivisione di conoscenze pratiche.

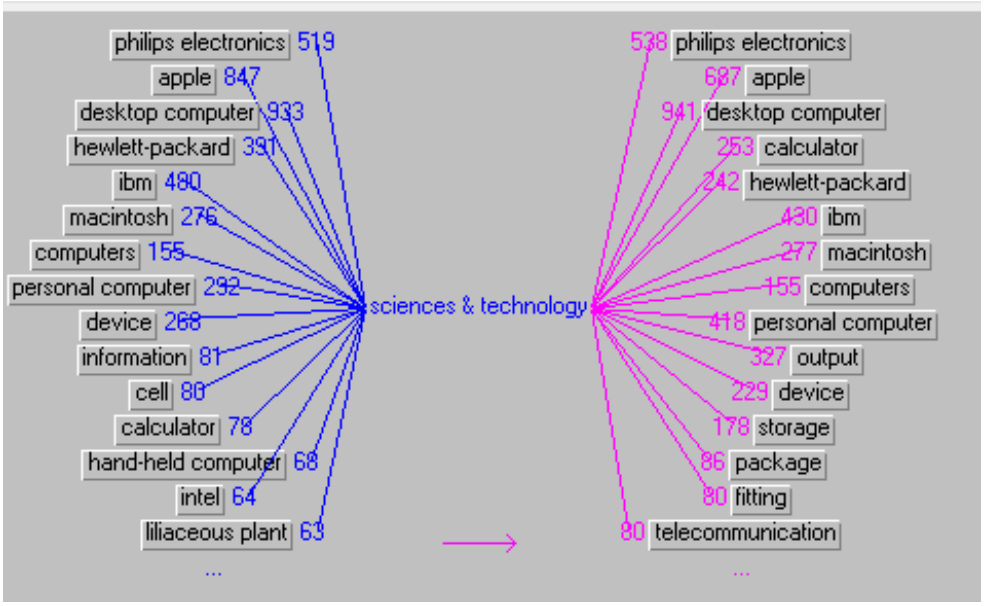


Figure 83 Rete semantica della categoria "sciences & technology" nel corpus museale.

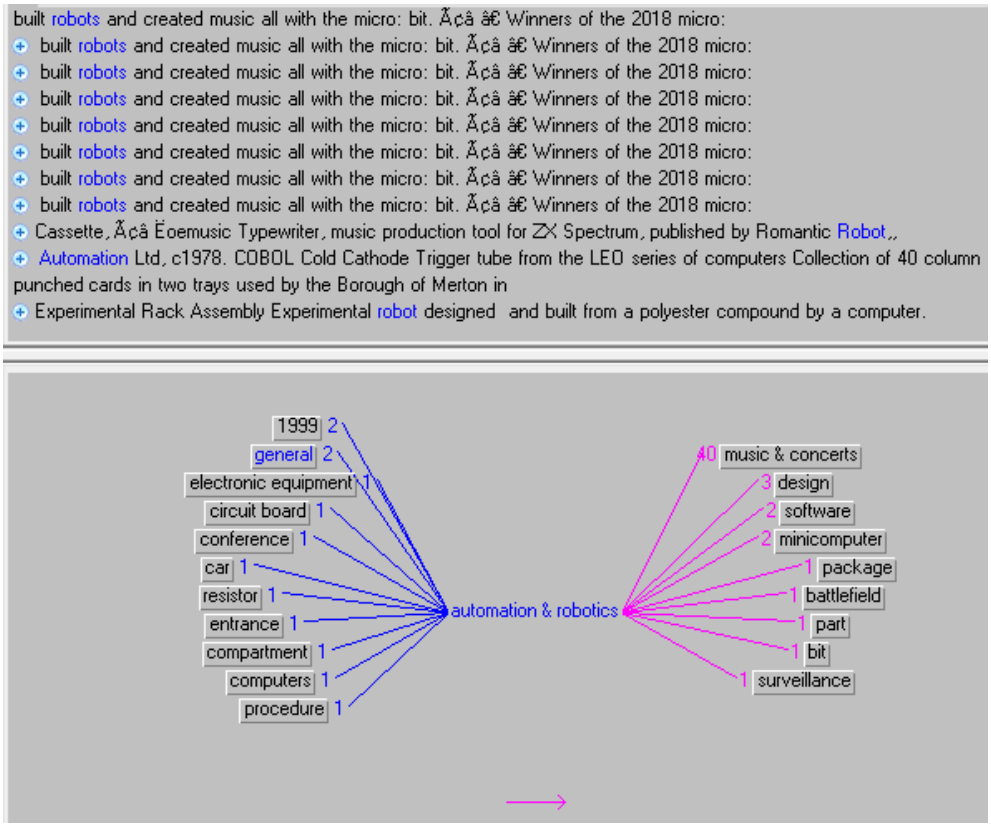


Figure 84 Rete semantica delle relazioni co-occorrenziali e contesti per il termine "robotics" nel corpus museale.

Nel corpus museale, infine, la rete semantica assume una forma curatoriale, confermando le analisi con Iramuteq e lo scopo stesso dei musei. *Automation & robotics* si lega ad *electronic equipment, minicomputer, package, surveillance, battlefield*. Il robot è trattato come artefatto materiale, classificato secondo provenienza, struttura, filiera produttiva e contesto d'uso. Il discorso non è pragmatico (uso), né nostalgico (comunità), ma documentario e attributivo.

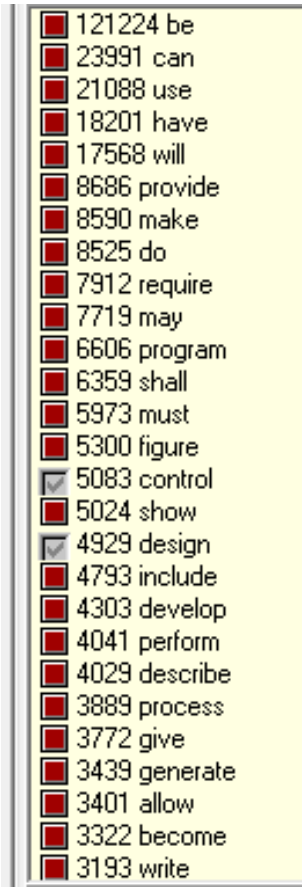
In sintesi, l'analisi Tropes mostra che le reti semantiche non si limitano a riflettere contenuti diversi, ma traducono veri e propri cambiamenti negli usi della tecnologia e nei modi di parlarne. Nei testi storici e tecnici, l'informatica è rappresentata come processo (progettare, calcolare, controllare); nei testi accademici recenti diventa oggetto di interpretazione (studiare, valutare, comprendere); nelle comunità online emerge come esperienza (condividere, ricordare, personalizzare); infine, nei contesti museali si configura come patrimonio materiale (classificare, conservare, esporre).

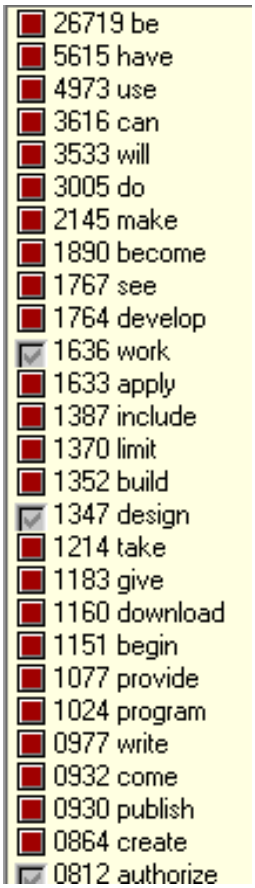
Queste trasformazioni mostrano che il lessico informatico non evolve solo in relazione allo sviluppo tecnologico, ma in stretta connessione con i regimi discorsivi che lo utilizzano: tecnico, scientifico, comunitario, curatoriale. La terminologia, quindi, non "nomina" soltanto gli oggetti, ma contribuisce a costruirne i significati culturali, definendo chi li usa, come, dove e per quali scopi.

È proprio su questo piano, nella relazione tra forma linguistica e funzione discorsiva, che si colloca l'analisi delle parti del discorso, che segue nella sezione successiva, e che consentirà di osservare come gli stessi lemmi assumano ruoli diversi a seconda del contesto comunicativo.

3.2.2.2 Analisi delle parti del discorso nei corpora

Attraverso l'analisi delle parti del discorso, con il software Tropes, è possibile osservare come gli stessi lemmi vengano impiegati con funzioni diverse, talvolta come sostantivi, altre volte come verbi o aggettivi, a seconda del tipo di discorso e del genere testuale.

Corpus1 

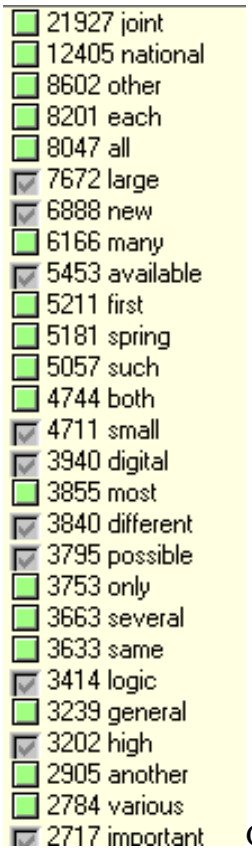
Corpus2 

Per quanto riguarda i verbi, il *sottocorpus computer1* presenta una forte densità di forme modali e operative tipiche della scrittura tecnica, come *can, use, have, provide, make, require, program, control, perform, develop, generate*. L'alto numero di occorrenze di *must, shall, may* evidenzia un linguaggio prescrittivo e normativo, spesso usato nella documentazione tecnica, nei protocolli o nei manuali. I verbi sono fortemente orientati all'azione tecnica e alla regolazione di procedure. Si tratta di una struttura verbale che supporta un discorso assertivo, sistemico e formale.

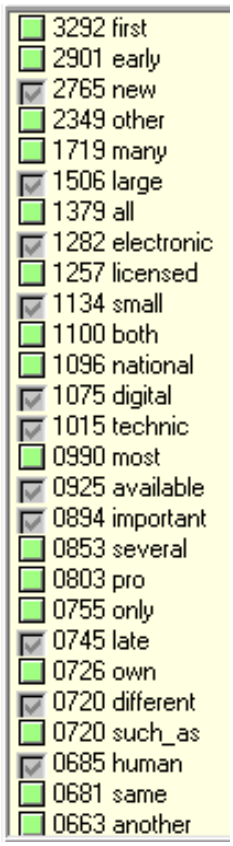
Nel sottocorpus computer2, invece, i verbi rivelano una maggiore varietà e apertura concettuale. Accanto a forme comuni (*be, have, use, make, do*), troviamo verbi legati al cambiamento e alla progettualità (*become, develop, work, apply, build, create*), ma anche all'interazione sociale e alla condivisione (*see, take, give, download, publish, authorize*). Il registro è meno prescrittivo e più

descrittivo o esperienziale, indicativo di un contesto comunicativo più narrativo e orientato alla fruizione, tipico della divulgazione o della riflessione interdisciplinare.

Corpus1



Corpus2



Anche sul piano degli aggettivi, le differenze risultano marcate. Nel corpus computer1, si nota una prevalenza di aggettivi classificatori e quantitativi, connotati da logiche dicotomiche e sistemiche: *national, digital, available, different, same, possible, important, logic, general, high*. Questi aggettivi sono utilizzati per categorizzare, tipizzare, descrivere proprietà tecniche e distinguere tra varianti. La loro funzione è informativa, descrittiva e sistematizzante.

Nel corpus computer2, al contrario, emergono aggettivi temporali (*early, first, late*), attributivi e valutativi (*important, own, pro, licensed, electronic*), insieme a elementi indicativi di contestualizzazione cronologica o storica *(techn(ic), available, national*). La frequenza di *first, early, late e licensed* suggerisce un'attenzione alla dimensione evolutiva e legale del discorso, spesso legata alla ricostruzione di eventi, alla storia dei prodotti informatici o alla descrizione museale. È evidente una maggiore varietà e contestualizzazione, coerente con l'approccio narrativo e retrospettivo del corpus.

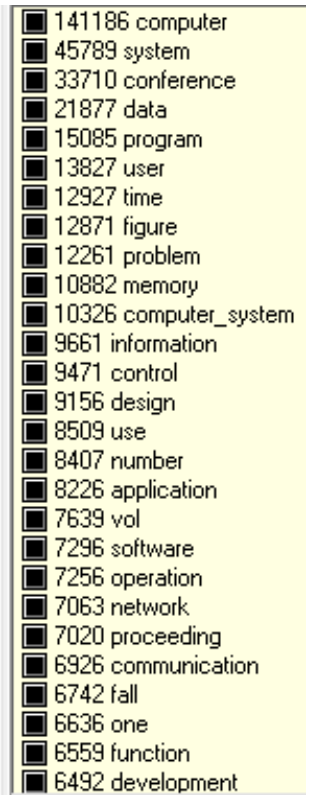


Figure 85 Classifica delle forme verbali più frequenti nel corpus 1

Figure 86 Classifica delle forme verbali più frequenti nel corpus 2

Infine, l'analisi dei sostantivi mostra una concentrazione terminologica diversa nei due insiemi. Nel corpus *computer1*, il lessico nominale è dominato da termini tecnico-scientifici: *computer*, *system*,

conference, *data*, *program*, *memory*, *design*, *application*, *software*, *network*, *operation*. Il lessico ruota intorno a componenti architetturali, funzionali e procedurali, in un contesto che privilegia la descrizione dell'oggetto e della sua ingegnerizzazione.

Nel corpus *computer2*, invece, accanto a *computer*, *software*, *development*, *technology*, *information*, si evidenzia la comparsa di termini legati al contesto storico-sociale e all'evoluzione della disciplina: *history*, *society*, *university*, *research*, *company*, *project*, *year*, *industry*, *ibm*, *machine*. Questi lemmi segnalano un cambiamento nel punto di vista: non solo cosa è il computer, ma chi lo produce, quando, dove e in quale contesto. La presenza di nomi propri (*ibm*) e di strutture sociali (*company*, *university*, *society*) suggerisce una narrazione più umanizzata e localizzata, orientata alla contestualizzazione storica e culturale.

Parti del discorso – Reddit

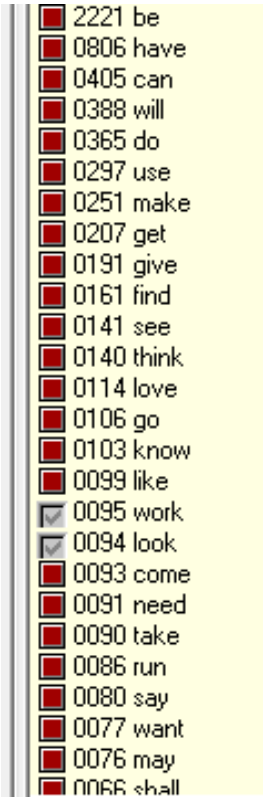


Figure 87 Classifica delle forme verbali più frequenti nel corpus Reddit

I verbi più frequenti mostrano un profilo discorsivo informale, soggettivo ed esperienziale:

Verbi comuni e quotidiani: *be, have, can, do, make, get, give, find, see, know.*

Verbi affettivi e valutativi: *love, like, think, want, say, look, need.*

Azioni concrete ma non tecniche: *run, come, go, take, work.*

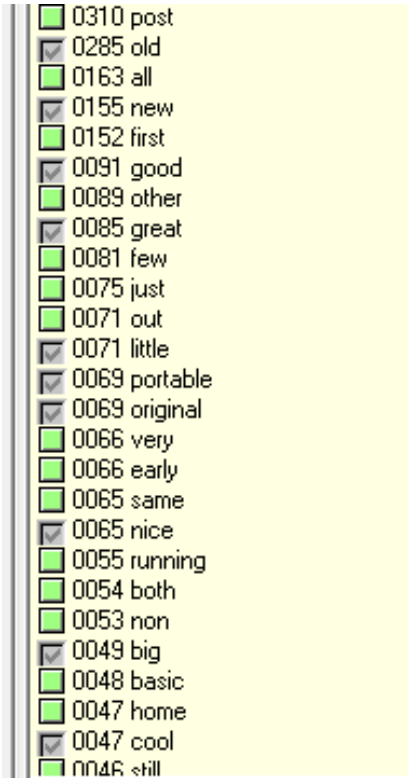


Figure 88 Classifica delle forme aggettivali più frequenti nel corpus Reddit

Gli aggettivi riflettono una dimensione soggettiva e valutativa, spesso connotata affettivamente:

- Aggettivi valutativi o enfatici: *good, great, nice, cool, big, original, very, still.*
- Aggettivi contestuali o temporali: *old, new, early, first, running, basic, portable.*
- Quantificatori o determinanti: *few, all, other, same, both.*

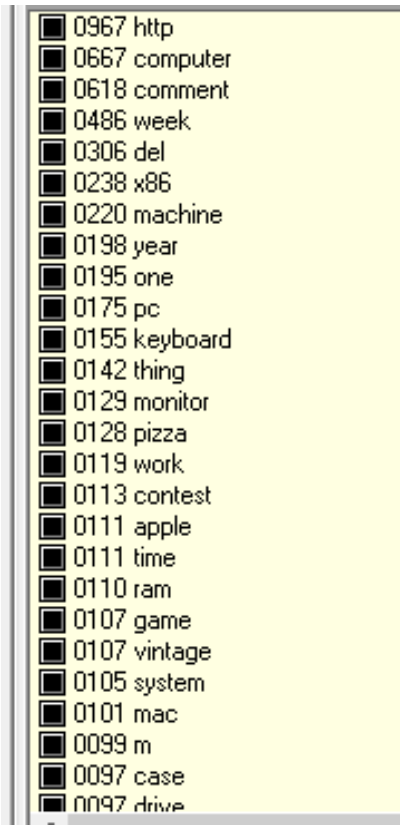


Figure 89 Classifica delle forme nominali più frequenti nel corpus Reddit

I sostantivi più frequenti coprono un'area ibrida tra tecnologia e cultura digitale pop:

- Tecnologia concreta: *computer, machine, x86, monitor, keyboard, ram, system, mac, drive.*
- Contesto comunicativo: *comment, week, post, time, http.*
- Oggetti specifici e cultura web: *pizza, game, apple, contest, thing.*

La nominalizzazione è ibrida, quotidiana e concreta, con enfasi su oggetti tangibili e artefatti tecnologici legati alla pratica e alla cultura utente. Termini come *pizza, game, contest* denotano un ambiente discorsivo ludico, spesso informale, ben diverso dalla nomenclatura funzionale dei corpora accademici.

Nel corpus tratto da Reddit emerge un registro discorsivo fortemente connotato in senso soggettivo e partecipativo. I verbi rivelano un linguaggio orientato all'esperienza e al giudizio personale, con ampio uso di forme psicologiche e affettive come *like*, *love*, *think*, *want*. Gli aggettivi esprimono valutazioni, nostalgia o descrizioni esperienziali (*cool*, *old*, *portable*, *first*), mentre i sostantivi oscillano tra tecnologia concreta (*ram*, *x86*, *keyboard*) e riferimenti alla cultura della rete e dell'interazione sociale (*pizza*, *comment*, *contest*). Questa composizione lessicale riflette un discorso informatico non più specialistico, ma relazionale, culturale e non standardizzato, tipico dei forum online.

Parti del discorso – corpus museale

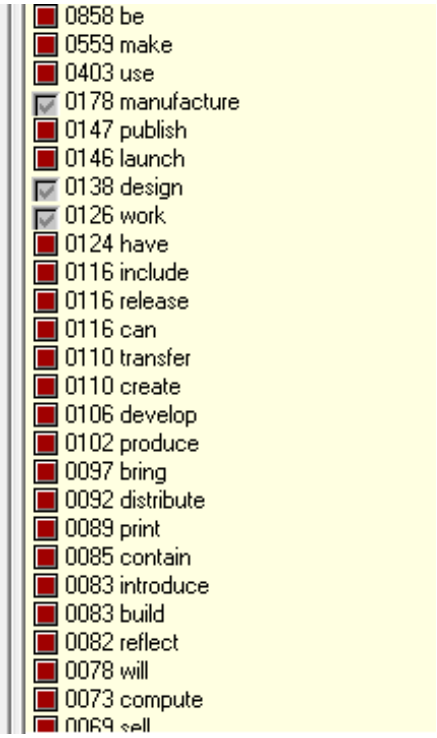


Figure 90 Classifica delle forme verbali più frequenti nel corpus museale

Il lessico verbale mostra un chiaro orientamento alla documentazione storica e descrittiva:

- Azioni produttive o industriali: *manufacture, design, launch, publish, develop, introduce, build, produce.*
- Azioni legate alla distribuzione e fruizione: *transfer, distribute, contain, print, sell.*
- Verbi cognitivi o descrittivi: *reflect, work.*

Il discorso verbale è centrato sulla storia degli oggetti tecnologici, con enfasi su progettazione, produzione e diffusione. L'assenza di verbi modali e prescrittivi (come must, shall, can) distingue chiaramente questo corpus dai testi tecnici. Il registro è cronachistico, catalografico, museale, orientato alla narrazione dell'evoluzione tecnologica.

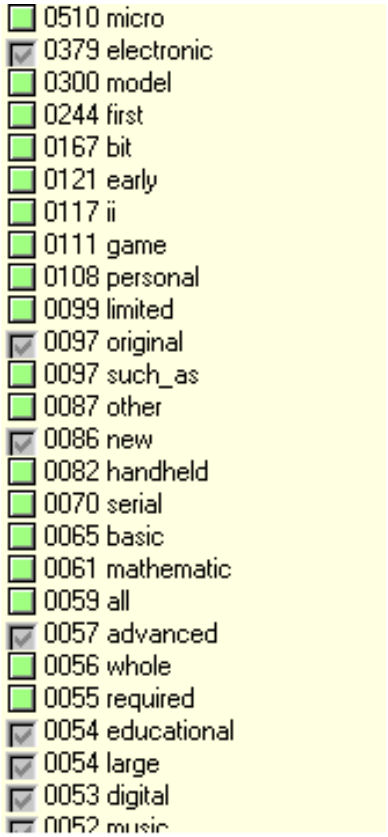


Figure 91 Classifica delle forme aggettivali più frequenti nel corpus museale

Gli aggettivi denotano una funzione classificatoria e temporale, spesso associata a prodotti informatici storici:

- Aggettivi tecnico-categoriali: *electronic, digital, model, serial, unit, portable, handheld, advanced.*
- Aggettivi temporali o storici: *first, early, ii* (riferimento a generazioni di console).
- Aggettivi di specificazione o uso: *personal, educational, mathematic, game.*

La funzione principale degli aggettivi è tipologica: descrivono versioni, modelli, epoche, formati.

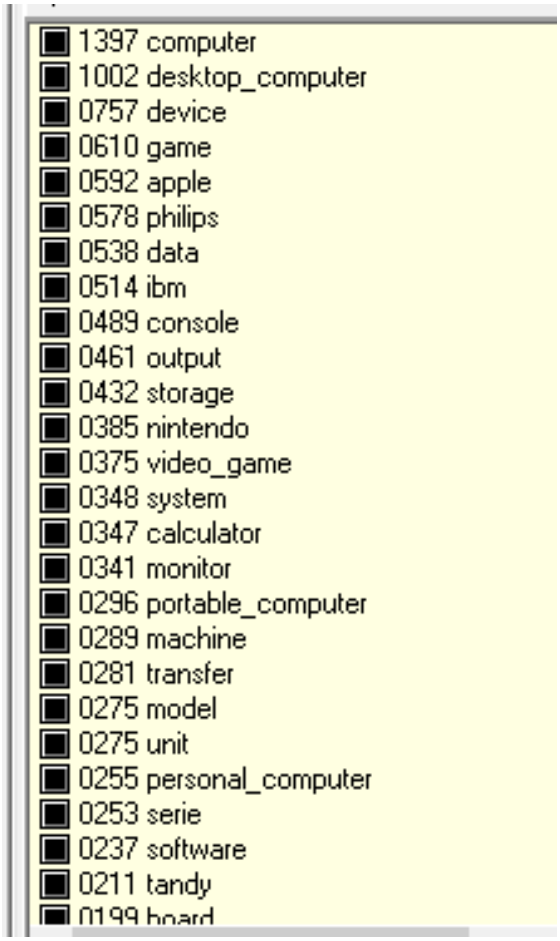


Figure 92 Classifica delle forme nominali più frequenti nel corpus museale

Il lessico nominale è dominato da artefatti materiali e brand storici:

- Oggetti informatici specifici: *computer, desktop_computer, portable_computer, personal_computer, device, console, calculator, monitor, unit, board.*
- Brand o marchi iconici: *ibm, apple, nintendo, philips, tandy.*
- Componenti o funzioni tecniche: *output, storage, software, system, model.*

Nel corpus museale si riscontra una notevole varietà lessicale nella denominazione dei dispositivi informatici ad uso personale, che riflette l'evoluzione storica e culturale del concetto di computer domestico. Le forme più ricorrenti includono:

- desktop_computer (1002 occorrenze)
- console (489 occorrenze)
- portable_computer (296 occorrenze)

In questo caso si vede chiaramente che la rete semantica è centrata su console, con estratti testuali contenenti riferimenti ai videogiochi (*videogame* con 375 occorrenze):

- Atari, Commodore, Amiga, Sega, Sony, Nintendo, IBM

L’analisi delle parti del discorso nei sottocorpora conferma che il dominio informatico non è caratterizzato da un unico registro linguistico, ma da una pluralità di stili e lessici che riflettono la natura multiforme della sua evoluzione. Da un lato, il discorso tecnico-normativo, orientato alla funzionalità e alla precisione; dall’altro, il discorso narrativo, storico e interpretativo, che fa emergere la dimensione sociale, culturale e simbolica dell’informatica. All’interno di questa polarità si inseriscono, con tratti distintivi, il discorso soggettivo e partecipativo di Reddit — centrato sull’esperienza, sull’affettività e sulla condivisione comunitaria — e quello storico-museale, volto alla classificazione, alla descrizione tipologica e alla documentazione cronachistica dei dispositivi. Queste prospettive coesistono all’interno del dominio, ma si manifestano con strutture e priorità lessicali profondamente diverse a seconda del contesto.

In conclusione, l’analisi semantica condotta con Tropes conferma che la terminologia informatica non si configura come un sistema monolitico, ma come una rete dinamica e articolata di significati. La presenza di categorie come *automation & robotics* in scenari sia tecnici che interdisciplinari, e l’emergere di elementi simbolici accanto a termini funzionali, mostrano chiaramente come il dominio informatico si presti a rappresentazioni molteplici, riflesso di comunità, contesti e pratiche discorsive differenti. Tropes si rivela dunque uno strumento utile per mettere in evidenza questa varietà semantica, rafforzando l’idea di una terminologia profondamente contestuale e in continua evoluzione. Nel complesso, i risultati di Tropes confermano che la terminologia informatica non è uniforme né stabile, ma si adatta alle trasformazioni storiche, ai contesti discorsivi e alle finalità comunicative. Mentre nei testi tecnici prevale una strutturazione funzionale del lessico, nei testi più recenti o generalisti si osserva una contaminazione con categorie culturali, simboliche e affettive. Il confronto tra i due sottocorpora mostra dunque una trasformazione progressiva dell’informatica da dominio tecnico-specialistico a fenomeno culturale e sociale, con ricadute dirette sul piano lessicale e concettuale.

3.2.3 TermoStat

3.2.3.1 Denominazioni del personal computer: un'analisi diafasica e diacronica

Dopo aver analizzato la struttura delle co-occorrenze e delle reti di significato mediante IRaMuTeQ e la categorizzazione semantica dei contesti tramite Tropes, si è proceduto a un'analisi terminologica mirata con TermoStat (Drouin, 2003).

L'obiettivo, in questo passaggio metodologico, è quello di individuare le unità terminologiche salienti, osservare le variazioni diafasiche e diacroniche e confrontare le scelte denominate nelle diverse comunità discorsive.

TermoStat consente infatti di calcolare le specificità lessicali, evidenziare termini caratteristici di ciascun corpus e mettere in relazione le denominazioni attraverso la misura delle co-occorrenze lessicali e della loro stabilità nei diversi contesti d'uso.

L'analisi diafasica e diacronica delle denominazioni del *personal computer* nei corpora rivela l'esistenza di un insieme di quasi-sinonimi termini che nel tempo sono stati impiegati per indicare i computer ad uso personale e domestico, utilizzati in maniera non sovrapponibile.

Un primo elemento di differenziazione risiede nell'epoca storica di riferimento. La forma *home computer* emerge con particolare forza negli anni '70 e '80, come mostrato dall'analisi di specificità nel sottocorpus 1961–1984, in corrispondenza della diffusione domestica dei primi modelli commerciali.

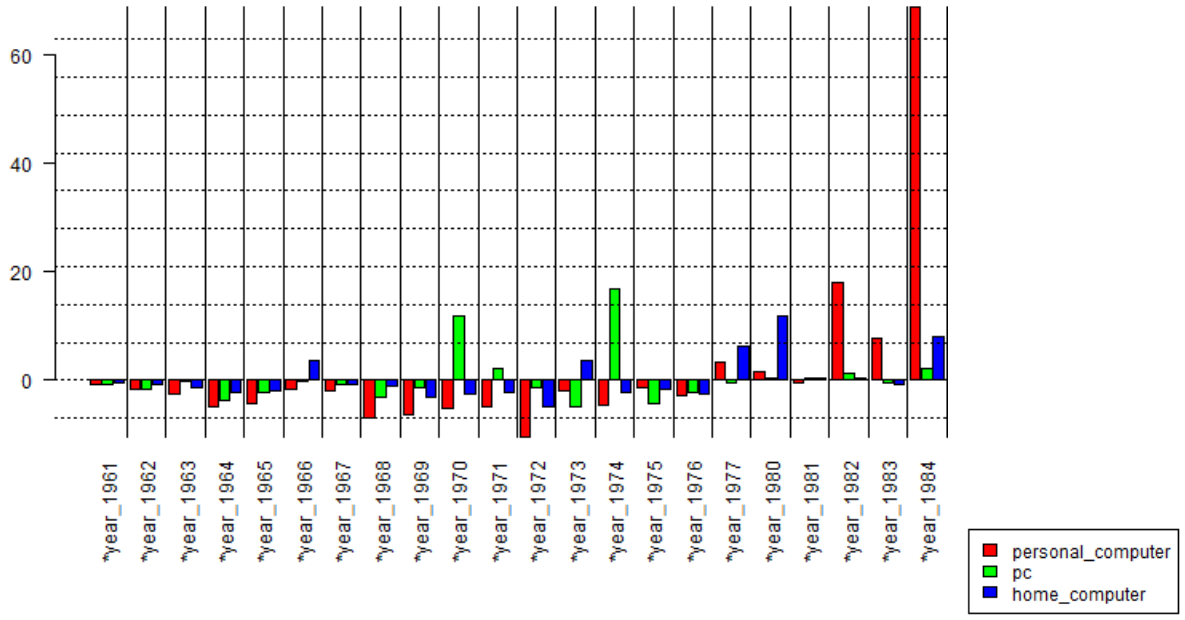


Figure 93 Andamento dei valori di specificità dei termini personal_computer, pc e home_computer nel sottocorpus 1 (1961–1984) (riproposta da pag. 274)

Il caso del Commodore PET esemplifica questo passaggio: l'oggetto entra nello spazio familiare e scolastico, legittimando la denominazione *home*.

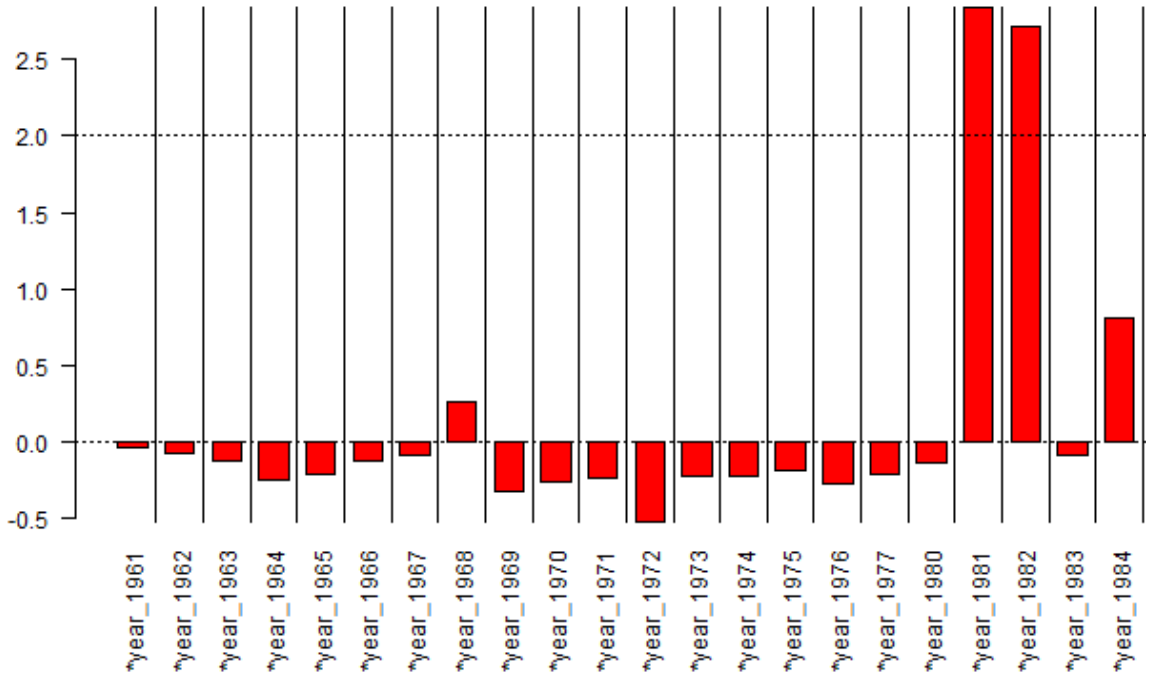


Figure 94 Andamento dei valori di specificità del termine pet nel sottocorpus 1 (1961–1984) (riproposta da pag. 274)

Nel sottocorpus 1985–2024, la frequenza relativa di *home computer* si riduce progressivamente, mentre *PC* acquisisce centralità a partire dagli anni 2000, con picchi di specificità nei periodi di standardizzazione Microsoft/Intel.

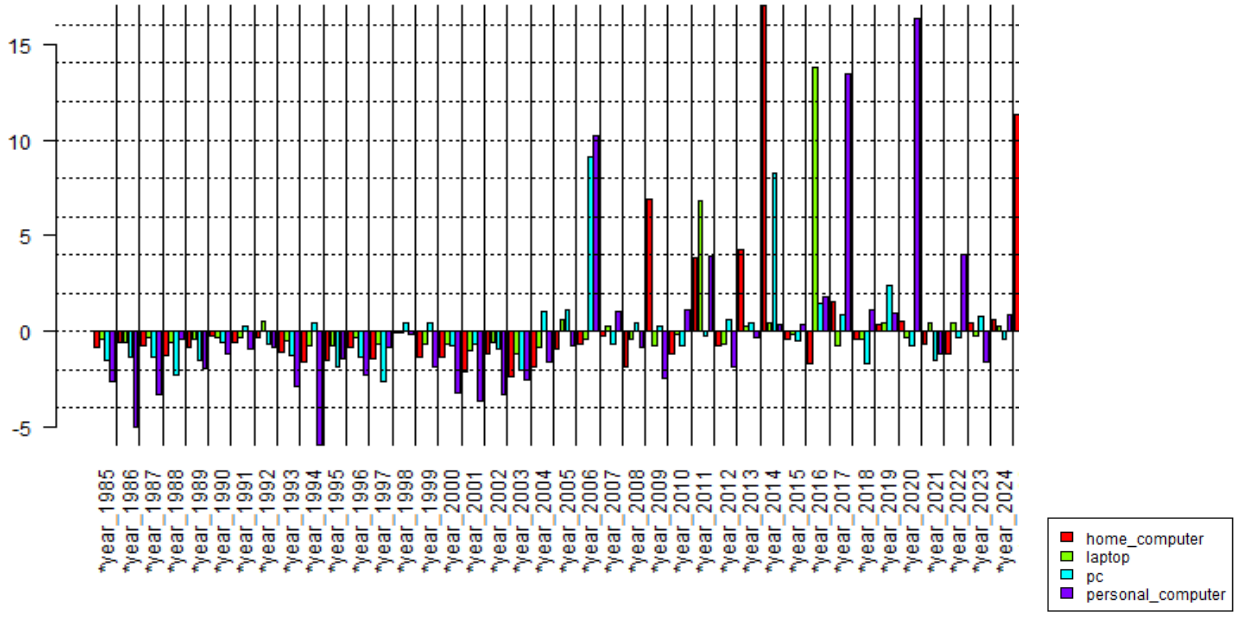


Figure 95 Analisi diacronica delle varianti terminologiche (pc, laptop, home_computer e personal_computer) nel sottocorpus 2 (1985–2024) (riproposta da pag. 283)

Una dinamica analoga si osserva anche nelle pratiche discorsive comunitarie: nel corpus Reddit la variante *home computer* è riattivata in chiave memoriale (*sir sinclair invented the pocket calculator but was best known for popularising the home_computer, bringing it to british high street stores at relatively affordable prices the following year;*).
Strettamente legato al dato temporale è il ruolo delle comunità d'uso, che orientano la scelta terminologica non solo in base a obiettivi semantici, ma anche seguendo il principio dell'economia del linguaggio. Nel corpus Reddit (2014–2024) emergono prevalentemente termini come *retro computer* e *vintage computing*. Tali termini definiscono l'identità di comunità esplicitamente orientate alla memoria e alla conservazione, come nelle r/retrobattlestations.

L'analisi delle forme attive (*vedi pag.*119) e la struttura dei cluster discorsivi (*vedi pag.126*) mostrano chiaramente che *retro* (169 occorrenze) e *vintage* (107 occorrenze) non funzionano come etichette tecniche, ma come marcatori identitari e affettivi: indicano oggetti che appartengono a una storia personale e collettiva, non soltanto a una classificazione.
Parallelamente, la scelta terminologica appare influenzata dal genere testuale. La forma *personal computer* rimane preferita nei testi accademici, istituzionali e documentali, dove svolge una funzione descrittiva e neutra. Questo andamento è evidente nell'analisi diacronica del primo sottocorpus (1961–1984), dove *personal computer* costituisce la denominazione standardizzata del settore progettuale (*vedi Figure 93*).

Anche nel secondo sottocorpus (1985–2024), pur in presenza di un'espansione terminologica significativa, *personal computer* rimane stabile come forma lessicale istituzionale e tecnica, mentre *PC* diventa progressivamente la variante dominante nei contesti ordinari di comunicazione (*vedi Figure 95)*.
La stessa tendenza emerge nel corpus museale, dove *personal computer* compare 239 volte, contro 173 occorrenze di *PC*. L'uso è associato a descrizioni espositive, schede inventariali e narrazioni storico-documentarie, spesso accompagnate da riferimenti a marche e linee produttive (IBM, Apple, Commodore). In questo caso, la denominazione non risponde a un'esigenza comunicativa immediata, ma alla necessità di collocare l'oggetto in una genealogia tecnica e materiale, coerente con le pratiche classificatorie del mondo museale.
Infine, l'analisi evidenzia come le dinamiche legate all'economia del linguaggio giochino un ruolo cruciale, specialmente nei contesti informali. Nel corpus Reddit, al contrario, la forma abbreviata PC è nettamente dominante: 183 occorrenze, risultando la quinta parola più frequente del corpus, contro appena 29 occorrenze di personal computer.

La densità terminologica evidenzia un focus documentario su oggetti concreti e identificabili. I sostantivi funzionano da etichette classificatorie e identificative, spesso connesse a marchi e modelli. Un esempio emblematico di variazione terminologica è, quindi, rappresentato dalla costellazione di forme che designano il concetto di *personal computer*. Nei testi accademici prevalgono forme standardizzate come *computer* e *system*; nei corpora museali e divulgativi compaiono espressioni come *portable computer*, *micro*, *unit*, *console* e *model*, spesso associate a nomi propri (es. *IBM*, *Apple*, *Commodore*), mentre nel corpus Reddit emergono denominazioni più informali o affettive, come *pc*, *mac*, *x86*, *game*, tipiche di una narrazione culturale e soggettiva dell'informatica

Per esplorare ulteriormente la distribuzione e la variabilità lessicale delle denominazioni legate al "personal computer" in un contesto comunitario e contemporaneo, è stato utilizzato anche il software TermoStat, applicando l'analisi sia al corpus Reddit sia al corpus museale. Poiché il software non supporta corpora di dimensioni eccessive, non è stato possibile estendere l'analisi ai sottocorpora accademici, di ampiezza superiore ai limiti gestibili dall'applicativo. A differenza degli strumenti precedenti, TermoStat consente di isolare in modo efficiente collocazioni lessicali stabili e combinazioni ricorrenti legate a uno stesso termine-chiave.

Ad esempio, nel corpus Reddit, *PC* appare con maggiore frequenza (183 occorrenze), mentre *home computer* e *personal computer* compaiono rispettivamente 31 e 29 volte, spesso in contesti di descrizione o confronto nostalgico. I termini non si equivalgono pienamente: *home computer* richiama una dimensione storica e familiare, *PC* una categoria commerciale e funzionale, *retro computer* una rielaborazione culturale e identitaria.

Il lemma "computer" risulta essere, dopo "reddit", il più frequente del corpus (662 occorrenze), a testimonianza della sua centralità concettuale.

Décomposition	
computer	
Tête	--
Expansion	--
Apposition gauche	--
Apposition droite	--
Adjectif	--
Termes en relation	vintage (104.1) personal (76.21) old (74.33) first (50.99) home (26.95) bulgarian (10.17) electronic (8.47) retro (7.29) favorite (6.77) system (5.31) case (4.28)
Inclus dans	bit computer bulgarian computer computer case computer magazine computer museum computer scene computer stuff computer systems electronic computer favorite computer first computer home computer old computer personal computer portable computer retro computer vintage computer

Figure 96 Scomposizione sintattica e lessicale del termine "computer" nel corpus Reddit

Gli aggettivi più frequentemente associati al sostantivo 'computer' sono *vintage, personal, old, first* e *home*, suggerendo un marcato orientamento esperienziale e nostalgico nel discorso degli utenti. Nella sezione "Inclus dans" emergono numerose collocazioni composte che riflettono tale tendenza: *vintage computer*, *home computer*, *retro computer*, *favorite computer*.

Per approfondire ulteriormente questa prospettiva, sono state analizzate le **concordanze** dei termini che consentono di osservare più da vicino i contesti d'uso e le sfumature semantiche attribuite dagli utenti al termine.

Concordanze di ***Home computer*** (31 occorrenze):

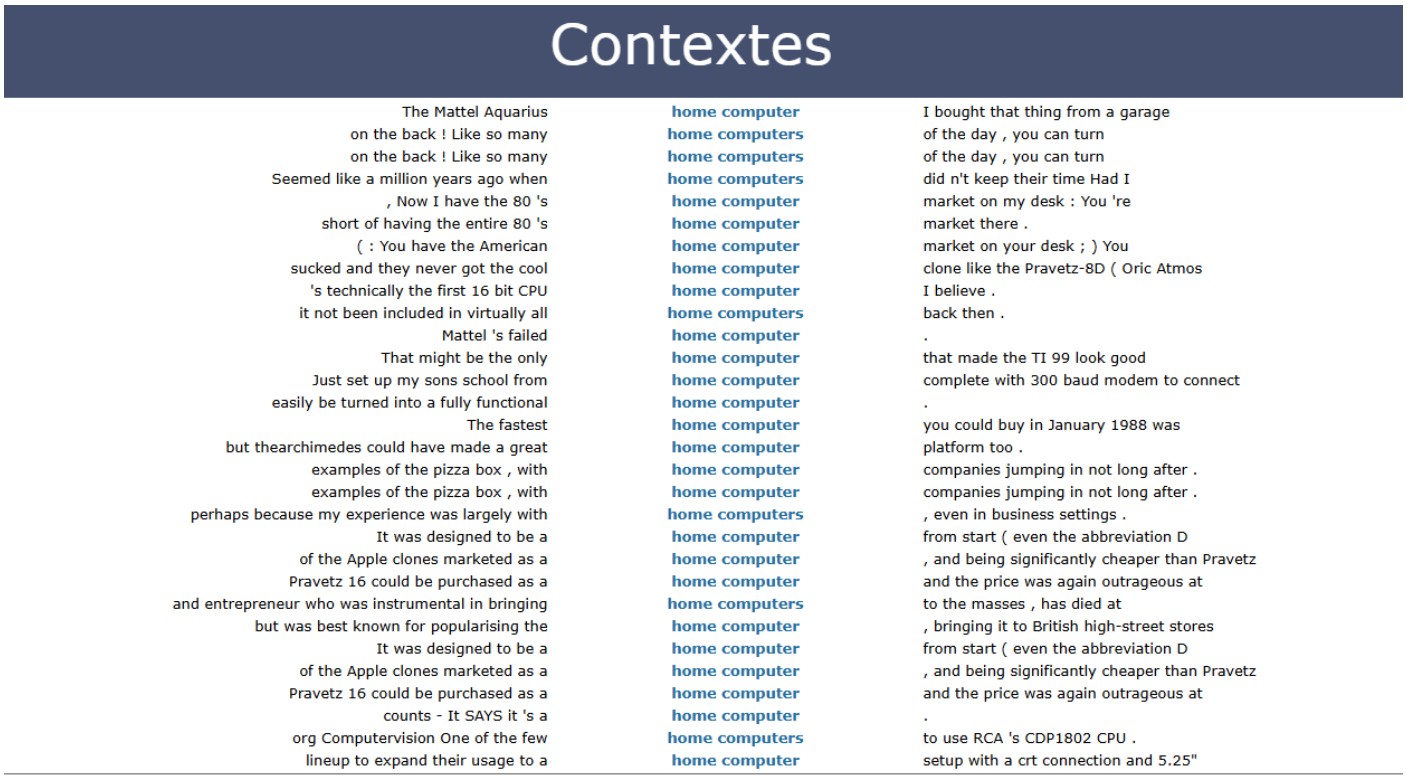


Figure 97 Concordanze del termine Home Computer nel corpus Reddit

Le concordanze di *home computer* mostrano che il termine è spesso impiegato in un contesto nostalgico e narrativo, legato agli anni ’80 e ai primi esperimenti domestici (es. riferimenti al *TI-99, Mattel Aquarius, Pravetz, Apple clones*). Emergono descrizioni di setup, componenti e ricordi personali (*my sons school from home computer, market on my desk, 300 baud modem*), che sottolineano la funzione sociale e memoriale di questi oggetti.

Concordanze di ***Personal computer*** (29 occorrenze):

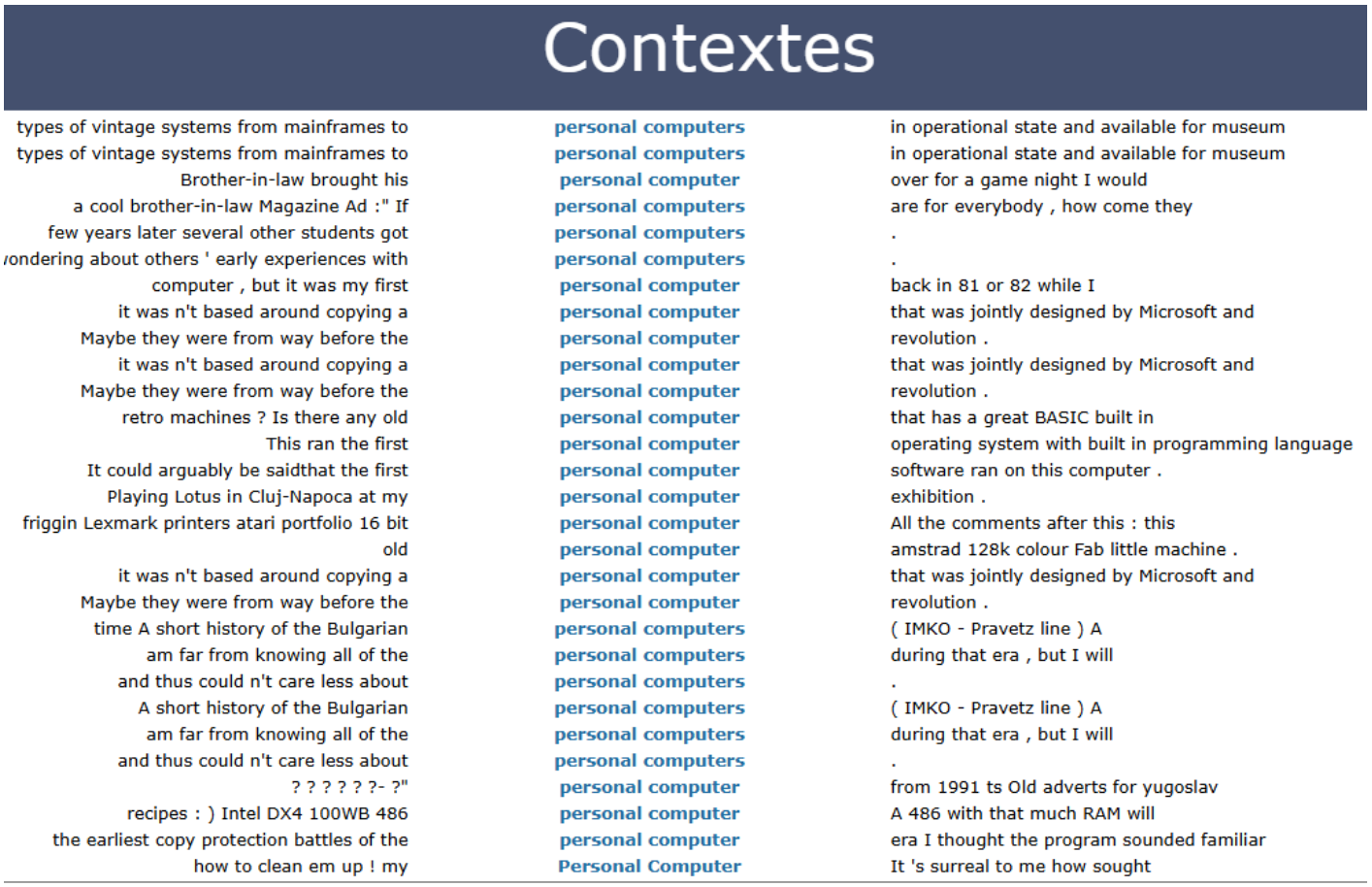
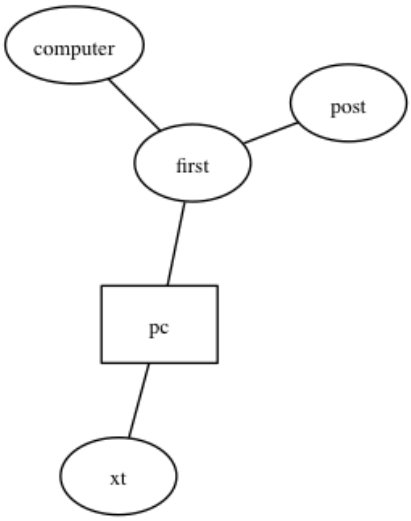


Figure 98 Concordanze del termine Personal Computer nel corpus Reddit

Per *personal computer*, invece, il contesto d'uso appare più ibrido: da un lato, compaiono riferimenti storici e museali (*available for museum, short history of the Bulgarian*…), dall'altro una dimensione tecnica *(jointly designed by Microsoft and revolution, built in programming language*). Rispetto a home computer, che attiva principalmente un frame domestico-familiare e affettivo, personal computer conserva una connotazione più istituzionale e tecnica

La forma abbreviata del personal computer in ***PC*** prevale, con 183 occorrenze e al quinto posto per frequenza:



Contextes		
of the entries are palmtops , pocket	PCs	or whatever you want to call them
Mac clone has a very typical cheap	PC	case with both 5.25 expansion slot covers
n't work any more , BUT the	PC	still works just fine ! Comes with
80s and early 90s	PC	manuals , books , and software If
and umpteen other 80s and 90s" IBM	PC	compatibles" .
Big Blue Week] Happy Birthday IBM	PC	! and special bonus sticker ## [
bonus sticker ## [Happy Birthday IBM	PC	!](http : .
Even the IBM	PC	seems bizarre when looking back on it
a Commodore ? To celebrate the IBM	PC	's birthday , anyone posting a picture
anyone posting a picture of their IBM	PC	5150 will get [one of these
will get [one of these IBM	PC	stickers](http : .
Happy Birthday	PC	! I have to ask : isthat
Mac clone has a very typical cheap	PC	case with both 5.25 expansion slot covers
n't work any more , BUT the	PC	still works just fine ! Comes with
Big Blue Week] Happy Birthday IBM	PC	! and special bonus sticker [Happy
special bonus sticker [Happy Birthday IBM	PC	!](http : .
Even the IBM	PC	seems bizarre when looking back on it
a Commodore ? To celebrate the IBM	PC	's birthday , anyone posting a picture
anyone posting a picture of their IBM	PC	5150 will get [one of these
will get [one of these IBM	PC	stickers](http : .
Happy Birthday	PC	! I have to ask : isthat
80s and early 90s	PC	manuals , books , and software If
and umpteen other 80s and 90s" IBM	PC	compatibles" .
drive setup built into an old Casio	PC	case A box full of Amiga and
Ever since then , inventing	PC	's just became joining newer hardware components
were just more interesting before the ibm	pc	became the standard .
drive setup built into an old Casio	PC	case A box full of Amiga and
IBM	PC	XT 5160 with custom built SD card
Matrix screen saver A picture of my	PC	running the matrix screen saver : SD
about connecting a C64 keyboard TO a	PC	via USB but not a PC keyboard
a PC via USB but not a	PC	keyboard to a C64 .
into packed mud compared to my WASD	PC	keyboard .
) by shokalion [Sanyo MBC-P100 Tablet	PC	- [Portables Week]](https : .
about connecting a C64 keyboard TO a	PC	via USB but not a PC keyboard
a PC via USB but not a	PC	keyboard to a C64 .
into packed mud compared to my WASD	PC	keyboard .
on the Texas InstrumentsNumber & the IBM	PC	XT , so there was a natural
Kinda reminds me of that Apple lunchbox	PC	in Hackers .
Hey , no shame about a Power	PC	G3 running OS9 ! That 's retro
Kinda reminds me of that Apple lunchbox	PC	in Hackers .

Figure 99 Concordanze del termine PC nel corpus Reddit

Workstation e ***Microcomputer*** hanno, invece, poche occorrenze (10 e 9).

L’analisi delle co-occorrenze e delle espansioni denominate conferma l’ipotesi di una oscillazione terminologica che accompagna l’evoluzione storica e culturale del concetto stesso di *personal computer*. Termini come *home computer*, *personal computer*, *PC*, *laptop*, *retro computer*, ma anche *console*, *device* e *machine*, si presentano in molti casi come quasi-sinonimi intercambiabili nei discorsi degli utenti, pur riflettendo di volta in volta specifici punti di vista (tecnico, nostalgico, commerciale, familiare).

Anche in assenza di una definizione univoca, queste etichette convivono e contribuiscono a costruire collettivamente il significato sociale e culturale del “computer” in contesto domestico. Tale fluttuazione è coerente con quanto rilevato anche tramite IRaMuTeQ e Tropes, e rafforza l’idea di una terminologia situata, contestuale e dinamica, sensibile alle trasformazioni dei contesti d’uso e delle comunità interpretative.

Questa pluralità non è solo lessicale, ma discorsiva e concettuale: l’informatica non si esprime attraverso un vocabolario fisso, bensì attraverso un repertorio concettualmente mobile, in cui la scelta terminologica risponde alle esigenze del discorso, ai valori della comunità e al tipo di memoria che si vuole attivare.

Il linguaggio tecnico è, per sua natura, stratificato e dinamico: i termini mutano nel tempo non solo sul piano lessicale, ma anche concettuale, poiché rispecchiano il modo in cui le comunità interpretano e organizzano la conoscenza. Nel caso dell’informatica, la continua innovazione tecnologica e la diffusione sociale dei dispositivi hanno generato una instabilità denominativa che rende difficile fissare confini semantici stabili.

La terminologia non evolve dunque come un sistema chiuso, ma come un ecosistema discorsivo, in cui i significati si ridefiniscono in funzione dei contesti d’uso, dei canali di comunicazione e degli attori coinvolti (utenti, produttori, divulgatori, istituzioni museali).

In questa prospettiva, la variazione terminologica osservata nel corpus Reddit, fortemente influenzata da pratiche comunitarie, esperienziali e culturali, trova un contrappunto nel linguaggio dei musei, dove la denominazione assume un valore documentario, classificatorio e conservativo.

Per completare il quadro e confrontare la prospettiva comunitaria con quella istituzionale, è stato quindi preso in esame anche il corpus museale, analizzato con il medesimo approccio tramite TermoStat. L’analisi del corpus museale tramite TermoStat mette in evidenza un lessico orientato alla catalogazione, classificazione e descrizione tecnica degli oggetti informatici.

Candidat de regroupement	Fréquence	Terme inclus
computer	2493	basic computer - wearable computer - luggable computer - first computer - powerful computers - personal computer - portable computer - top computer - programmable computer - disparate computers - early computers - mini computer - laptop computer - modern computer - electronic computer - hand-held computer - aquarius computer - digital computer
device	760	memory device - output device - created devices
game	737	educational game - electronic game - retro games - game cartridge - video game - arcade game - video game - game cassette - game console - computer game - types of games - game disk - intellivision games - cassette game - game compilation
make	481	
output	464	next output - output device
output device	440	
model	412	
use	397	
transfer	382	
â	381	worked â - abbreviation â - called â - accepted â - tool â
bit	304	bit development
personal	290	personal computer market - personal - personal computer - personal collection - personal record - personal organiser - personal computer model - personal computer industry - personal digital assistant
machine	279	cash machine - xt machine - desktop machine - test machine - machine intelligence - calculating machine
cassette	278	game cassette - cassette player - cassette recorder - audio cassette - cassette tape - cassette game - cassette interface
calculator	262	key-driven calculator - calculator model - pocket calculator - electronic calculator - scientific calculator - calculator chip - set calculator - desktop calculator - set calculator - printing calculator - programmable calculator - electric calculator
micro	257	micro computer - first micro
electronic	242	electronic desktop printing calculator - electronic game - electronic calculator - electronic - electronic components - electronic pocket calculator - electronic desktop calculator - electronic system - electronic payment terminal - electronic computer - electronic organiser - electronic desktop - electronic pocket

Figure 100 Termini in ordine di frequenza nel corpus museale

Il termine *computer* è di gran lunga il più frequente (2.493 occorrenze) e risulta al centro di una fitta rete di combinazioni terminologiche che ne articolano la varietà tipologica e storica.

Décomposition	
computer	
Tête	--
Expansion	--
Apposition gauche	--
Apposition droite	--
Adjectif	--
Termes en relation	personal (110.99) portable (51.13) electronic (47.98) powerful (10.91) disparate (10.56) early (9.24) hand-held (7.76) top (7.29) aquarius (6.82) laptop (5.47) luggable (4.09)
Inclus dans	aquarius computer basic computer digital computer disparate computers early computers electronic computer first computer hand-held computer laptop computer luggable computer mini computer modern computer personal computer portable computer powerful computers programmable computer top computer wearable computer

Figure 101 Scomposizione sintattica e lessicale del termine "computer" nel corpus museale

Le collocazioni più ricorrenti (*basic computer, wearable computer, luggable computer, portable computer, first computer, personal computer, mini computer, digital computer*) mostrano come il discorso museale privilegi un approccio tassonomico e genealogico, volto a tracciare la storia evolutiva delle forme del calcolo personale.

Gli aggettivi più frequenti associati a computer *(personal (110.99), portable (51.13), electronic (47.98), powerful (10.91), disparate (10.56), early (9.24), hand-held (7.76), top (7.29), aquarius (6.82), luggable (4.09)*) delineano un registro tecnico-descrittivo, che tende a definire l'oggetto

attraverso attributi funzionali, cronologici e classificatori. Rispetto al corpus Reddit, dove gli aggettivi connotano esperienze personali e affettive ("vintage", "old", "favorite"), nel corpus museale prevale una semantica della specificazione oggettuale, tipica dei linguaggi della museografia tecnica.

Le concordanze confermano questa tendenza. In corrispondenza di *personal computer*, le occorrenze ricorrono quasi esclusivamente in descrizioni di modelli, marche e date di produzione, come: "*Apple Macintosh personal computer", "IBM personal computer Model AT", "Commodore PET Series personal computer", "Casio FX 850P personal computer, 1989*".

Contextes

" A Ferranti	**personal computers**	table-top display sign .
drive , c Amstrad CPC 6128 colour	**personal computer**	.
iMac G3 ' Bondi Blue '	**personal computer**	, manufactured by Apple Inc .
standard and become the market leader in	**personal computers**	.
It was the first	**personal computer**	with a graphical user interface (GUI
hard drive by Rodime PLC Apple Macintoshse	**personal computer**	Apple Macintoshse personal computer .
PLC Apple Macintoshse personal computer Apple Macintoshse	**personal computer**	.
Apple StyleWriter printer , from Apple Macintoshse	**personal computer**	, c.1989 Apple TV digital media player
, the IBMps was closer to a	**personal computer**	than game console and worked â ?
Cables , associated with Wang PC	**personal computer**	, manufactured by Wang Laboratories , c.
box and case , Casio FX 7000G	**Personal Computer**	with manual , box and case ,
and case , 1989 Casio FX 730P	**Personal Computer**	with manual and case , 1989 Casio
750P with case , Casio FX 850P	**Personal Computer**	manual , 1989 Casio FX 850P Personal
Computer manual , 1989 Casio FX 850P	**Personal Computer**	with manual and case , 1989 Casio
For use with Commodore Pet 2001 Series	**personal computer**	.
ten floppy disks , from Apple Macintoshse	**personal computer**	, c.1989 Closed Intel 8008 microprocessor chip
Altair 8800 , one of the first	**personal computers**	, which was available to buy in
Commodore Pet 2001 Series	**personal computer**	, Serial No 1004030. Commodore PET Series
Serial No 1004030. Commodore PET Series 2001	**Personal Computer**	, made by Commodore Business Machines in
, the IBMps was closer to a	**personal computer**	than game console and worked â ?
Computer system box , from	**personal computer**	by IBM , Model AT , c.
by IBM , Model AT , c.	**Personal Computer**	by IBM , Model AT , c.
IBM dominated the	**personal computer**	market in the early to mid s
on wheels , associated with Wang PC	**personal computer**	Computerised Accounting System by Philips model P320
, the IBMps was closer to a	**personal computer**	than game console and worked â ?
, the IBMps was closer to a	**personal computer**	than game console and worked â ?
PDTNumber Programmable Data Terminal DEC Rainbow 100	**Personal computer**	System , c. DEC Rainbow 100 Processor
Documents , paperwork and manuals relating to	**personal computer**	kit by NASCOM , .
cassette interface cable for Texas Instruments TINumber	**personal computer**	, Texas Instruments Incorporated , , c.
disk , model CBM 8260 , from	**personal computer**	by Commodore Dynalogic Hyperion ' luggable '
with inexpensive materials including off the shelf	**personal computers**	and thin sheets of corkboard for insulation
Scientific , 1996 Graphics printer , from	**personal computer**	by IBM , Model 4860 Pc Jnr
, Model 4860 Pc Jnr , c.	**Personal Computer**	by IBM , Model 4860 PC Jnr
IBM introduced the first	**personal computer**	, the PC , in .
It was the catalyst for the	**personal computer**	industry .

Figure 102 Concordanze del termine Personal Computer nel corpus museale

In questi casi, il termine non denota un concetto astratto, ma un oggetto storico identificabile, e si inserisce in un discorso curatoriale di tipo archivistico o espositivo.

Décomposition	
personal	
Tête	--
Expansion	--
Apposition gauche	--
Apposition droite	--
Adjectif	personal
Termes en relation	collection (167.58) computer (119.71) digital assistant (40.15) computer model (22.97) organiser (14.94)
Inclus dans	personal personal collection personal computer personal computer industry personal computer market personal computer model personal digital assistant personal organiser personal record

Figure 103 Scomposizione sintattica e lessicale del termine "personal" nel corpus museale

L'aggettivo *personal* appare anche come testa di composti nominali (*personal collection, personal organiser, personal digital assistant*), indicando una dimensione possessiva e funzionale che collega la storia del personal computer a quella della cultura individuale e della mobilità tecnologica.

La presenza significativa di denominazioni come *portable computer, hand-held computer, wearable computer e luggable computer* segnala la progressiva miniaturizzazione e portabilità dei dispositivi, un tema centrale nella narrativa museale sull'evoluzione tecnologica.

Le concordanze mostrano l'uso sistematico di queste etichette per distinguere tipologie di macchine e per situarle cronologicamente, come in:

- "*Portable computer by Toshiba T1000LE";*
- *"Apple Newton eMate 300 laptop, made by Apple Inc., 1997";*
- *"Wearable computer with head-mounted display, University of Birmingham, 1995";*
- *"Russian hand-held computer in purple, ABBYY, 2000*".

Tali formulazioni combinano l'indicazione tecnica (modello, anno, produttore) con la descrizione morfologica o funzionale (*luggable, portable, wearable*), dando vita a un linguaggio fortemente nominale, informativo e classificatorio, ma anche capace di restituire una dimensione storica dell'innovazione.

Nel corpus museale, quindi, il termine computer agisce come nodo lessicale di una tassonomia storica, più che come elemento narrativo o valutativo. Le varianti terminologiche (*personal, portable, hand-held, wearable*) non si configurano come sinonimi, ma come etichette museotecniche che contribuiscono a costruire la genealogia materiale del personal computer e delle sue metamorfosi. L'assenza di marcatori soggettivi o espressivi (presenti invece nel corpus Reddit) rafforza l'idea che

il linguaggio museale sia orientato alla trasparenza informativa, alla precisione descrittiva e alla comunicazione scientifico-divulgativa.

In sintesi, la prospettiva museale rappresenta un passaggio discorsivo chiave: dal “computer vissuto” (emotivo, comunitario) al “computer esposto” (classificato, documentato, storicizzato). Questa polarità discorsiva, già emersa negli altri corpora, trova nel confronto tra Reddit e musei una delle sue manifestazioni più nette, rivelando la dimensione socio-cognitiva del lessico informatico come strumento di mediazione tra esperienza e memoria.

3.2.4 Sketch Engine

3.2.4.1 Analisi distribuzionale e contestuale delle denominazioni del personal computer

L'impiego di Sketch Engine (Kilgarriff *et al*., 2004), in sinergia con gli altri strumenti lessicometrici e semantici come IRaMuTeQ, Tropes e TermoStat, ha permesso di analizzare in profondità la rete terminologica che ruota attorno al concetto di ***personal computer***, rivelando una pluralità di etichette che, pur riferendosi a oggetti simili, non sono semanticamente equivalenti, né discorsivamente neutre.
Per esplorare ulteriormente la variabilità terminologica associata al concetto di *personal computer* nei testi tecnici del sottocorpus_computer1 (1961-1984), è stato predisposto un preprocessing finalizzato all'importazione in Sketch Engine. Il file testuale originale è stato trattato mediante uno script Python che ha rimosso punteggiatura, simboli non alfanumerici e stopword inglesi, mantenendo inalterate le maiuscole, così da non alterare i nomi propri e le sigle. Il file risultante, *sottocorpus_computer1_cleaned.txt*, superava il limite di 2 milioni di token previsto da Sketch Engine; si è quindi proceduto alla sua suddivisione in segmenti da 500.000 token ciascuno, raccolti in una cartella dedicata (*split_output*).
Per ottimizzare ulteriormente la dimensione complessiva e rientrare nei vincoli della piattaforma, è stato eseguito un ulteriore filtraggio automatico degli avverbi in *-ly*, con un approccio euristico, ottenendo una versione alleggerita del corpus salvata nella directory *split_output_no_adverbs*. È stata quindi caricata la prima porzione del corpus su Sketch Engine, abilitando le analisi su co-occorrenze, concordanze, frequenze e thesaurus distribuzionale.
L'analisi delle concordanze sui termini legati al concetto di *personal computer* nel sottocorpus 1961-1984 mostra una marcata variabilità terminologica, ma anche una significativa evoluzione temporale dei nomi.

- personal computer (327): forma formale, tecnico-commerciale/progettuale;
- PC (127): abbreviazione tecnica tipica di documentazione e procedure;
- home computer (134): espressione tipica degli anni '80, legata a consumo domestico, accessibilità, educazione;
- notebook (105): spesso in notebook computer, etichetta emergente;
- portable computer (10): raro, ma indice dell'avvio della mobilità;
- laptop / retro computer / mobile computer (0): assenti nel periodo;

- console (1.189): frequentissima ma con senso tecnico (terminale / I/O), non videogame.
- handheld (14): pochi casi, associati a tecnologia avanzata o militare. Segnala un uso emergente per piccoli dispositivi.

Se *personal computer* e *PC* compaiono con frequenze consistenti, altre etichette che diventeranno comuni in epoca successiva, come *laptop* o *retro computer*, sono del tutto assenti. In compenso, emergono forme precoci come *notebook computer* o *portable computer*, che riflettono una transizione verso la mobilità e l'individualizzazione dell'uso.
"Notebook computer way people use computers great notebook computers revolutionize computer usage[...]"

Il termine *console* è molto frequente, ma rivela un diverso campo semantico: non è sinonimo di computer domestico, bensì si riferisce a terminali o unità di controllo in ambito industriale o accademico.

Left context	KWIC	Right context
diting supported one requires aration illustrations video display terminal national computer conference system	console	crt screenoriented provides detailed information manage denning peter chairman undergraduate course ment
evented bookkeeping entries activating sys could reused new loans computer listings tem interrupts computer	console	invoices containing fake data describing collateral prepared system without recording used transactions comp
llected many free programs system convert terminal functional equiva offered computer centers lent computer	console	performed act returned france sold copies programs convince management computer facility collected caught
contains information computer incidents cases use computer translation machinelanguage editing controlled	console	time differential mans actions com puters responses buffered console used wide variety informationretrieval
machinelanguage editing controlled console time differential mans actions com puters responses buffered	console	used wide variety informationretrieval dataprocessing applica tions also included hardware requirements musl
tion gained actual function control read onto magnetic tape associated depends upon computer programming	console	restrictions user wishes place tape called format tape operator therefore involves change contains form allowa
re involves change contains form allowable inputs whatever pac circuits computer forms written change types	console	modes oper format tape soon computer programs ator identification requirements established process format
instruments simda pairs aircraft could statement written english changed business affilit ation indicates format	console	operator change required would labelling operator selects particular format computers key total systems contr
s outputs answer magnetic operator tape pac standard computer words versatile manmachine communication	console	comprising seven digits inparallel since theory operation pacis communications device may stream computer
l applied would altered outcome national computer conference incident two four safeguards measures analyst	console	pac computers key total systems control blocks shown figure next accepts routes data combination section de
dex figure typical format personnel information retrieval application pac versatile manmachine communication	console	rotate format memory entered computer taking memory routine controls reading important time processing pre
copying direct change safeguards software software mistakes identification final opera via keyboard tional	console	used acsimatic correct mistake routine reads system demonstrating potentiali zeros core memory replace eith
rapid inexpensive detection violations types computer system integrity violations br display ifam capabilities	console	display unit capable presenting full range alphabetic numerical graphical data test model description console
console display unit capable presenting full range alphabetic numerical graphical data test model description	console	allows direct communication operator computer display unit contains test activity order enhance technology ac
e domain degrees cdc computer bit word direct memory core access chahkel br goodyear display associative	console	memory light gui bit word program keyboard words alphahltcteric keyboard word parallel bit serial operation ca
s give greater emphasis compared efficiency standard done system master passes bit assembler mes control	consoles	telephone line terminal sage assembler computer rack power distribution monitoring check made pattern cont
aphics display processors cacm june pp fjcc pp cp cpu chip turns terminal standalone ni ninke satellite display	console	system chine electronics june pp multiaccess central computer proc ifip con cr crompton soane interactive gre
sockut stabler van dam sors cacm michel computer architecture instruction set design ninke satellite display	console	system multi ncce june access central computer proc ifips ee design assistant applicon corp news release rap
ut data complete national computer conference must know abling deflection servos rate response slides case	console	high lag associated rea slides case large screen display sonable human manipulation joystick computers appe
ication display devices manufactured thompson ramo wooldridge inc latter includes computer communication	console	online group display introduction performance doddac represents signifi cant advance large scale data handlir

Figure 104 Visualizzazione KWIC (Key Word In Context) che mostra le occorrenze del termine "console" all'interno di un corpus testuale analizzato tramite Sketch Engine.

"Console display unit capable of presenting full range alphabetic numerical graphical data..."
Questa frase indica chiaramente che la console è vista come unità di visualizzazione, capace di mostrare dati alfanumerici e grafici. Non è un dispositivo ludico o domestico, ma strumento tecnico per interagire con il sistema.

La parola “console” si colloca in un dominio semantico affine a “terminale”, “monitor”, “display unit”.
“Console used [for] a wide variety [of] information retrieval and data processing applications.”
La console viene descritta come interfaccia di input/output in contesti professionali (data processing, information retrieval). La console è parte di una infrastruttura di elaborazione dati complessa, non è un oggetto “personale” ma “collettivo”, da laboratorio, ufficio tecnico, o centro elaborazione.
“Time differential man’s actions computers responses buffered console...”
È una frase complessa, ma suggerisce che la console funge da interfaccia temporale tra l’azione umana e la risposta della macchina, probabilmente con buffer o ritardi.
La console è un mediatore tecnico, non un dispositivo indipendente, e soprattutto non ha connotazioni legate all’intrattenimento.

Questi dati confermano l’ipotesi che le etichette terminologiche siano storicamente situate, e che la rappresentazione del *personal computer* vari non solo per comunità di pratica, ma anche in relazione allo sviluppo tecnologico e ai discorsi che lo accompagnano.
Rispetto a *PC* o *personal computer*, che rimandano ancora a un vocabolario tecnico, *home computer* rappresenta un passaggio discorsivo verso la familiarizzazione sociale del concetto, prefigurando l’immaginario che dominerà la comunicazione degli anni ’80 e ’90, come emerge dagli esempi che seguono:

“Home computer paying bills... fulfilling familiar tasks.”
“Home computer pressing start button, faces data input techniques...”
“Shopping home... innovation consumers overlooked home computers...”
“Perhaps four models personal home computer...”
“Home computer bonanza remains a vision...”:
bonanza → indica un'opportunità, una prospettiva di grande successo commerciale o sociale.
remains a vision → segnala che, nel momento della scrittura, non si è ancora realizzata: è un’aspettativa, un’anticipazione.

Questa tendenza alla pluralizzazione e alla fluttuazione terminologica si accentua ulteriormente nel sottocorpus computer 2 (1985-2024). L’analisi del secondo sottocorpus, previa pulizia attraverso lo script Python per rispettare la memoria di Sketch Engine, conferma infatti una ridefinizione delle etichette legate al *personal computer*, in cui la terminologia non solo si amplia, ma assume una dimensione più flessibile, esperienziale e culturalmente situata.

Termini come *laptop* (61 occorrenze) e *notebook* (36 occorrenze), assenti o marginali nei testi più datati, si affermano nel sottocorpus recente all'interno di cornici discorsive nuove, incentrate sulla portabilità e all'individualizzazione dell'uso. Le occorrenze mostrano contesti legati alla progettazione tecnica, all'ingegneria del design e alla formazione universitaria, ma anche all'adattamento del dispositivo alle esigenze personali:

- *"**laptop** components design engineering technology cer jack son in said meaningful con duct reverse engineering corn";*
- *"we traced the global life cycle of a typical **laptop computer** or cell phone from its material origins…";*
- *"end-user devices such as their desktop or **laptop computers**, smartphones or tablets…"*

Analogamente, notebook compare in contesti che ricostruiscono la filiera produttiva globale e le dinamiche industriali:

- *"**notebook computer** development alliance wang case study research continuous participatory requirements notebook soft";*
- *"Quanta became one of the first companies in Taiwan to make a **notebook-sized computer**, mass-produced in 1990 and sold internationally.";*
- *"the calculator industry was the mother of the **notebook computer** industry in Taiwan."*

In queste frasi*, laptop* e *notebook* non vengono descritti come meri dispositivi di calcolo, ma come piattaforme evolute, oggetto di sperimentazione collettiva e strumento per applicazioni partecipative. L'enfasi si sposta quindi dalla macchina come prodotto finito alla macchina come ambiente flessibile, da costruire e adattare.

Inoltre, l'assenza di riferimenti diretti all'intrattenimento o alla domesticità suggerisce una transizione semantica rispetto a home computer: se quest'ultimo indicava l'ingresso del calcolatore nello spazio privato e famigliare, laptop e notebook ne segnano la mobilitazione, sia fisica (trasportabilità) sia sociale (integrazione nei ritmi quotidiani del lavoro, dello studio, della progettazione).

Nella seguente frase presente nel corpus: *"home computers videogame consoles modernday wars home computer culture always marked rivalry user groups typical confrontation"*, i due termini 'home computer' e 'videogame console' sono accostati come se fossero intercambiabili o affini, o comunque parte dello stesso spazio semantico e culturale.

Rispetto ai più generici *personal computer* o *PC*, che mantengono una connotazione tecnico-descrittiva, questi termini portano con sé un impianto discorsivo più individualizzato e situato, in

linea con la trasformazione culturale dell'informatica negli anni 2000: da strumento da laboratorio o ufficio, a oggetto personale, portatile, adattabile, e sempre più simbolico.

Infine, il numero ridotto di occorrenze di etichette come *mobile computer* (7) o *handheld* (23) conferma che, pur esistendo nel lessico tecnico, non si sono affermate nella pratica discorsiva comune, forse perché troppo generiche o perché inglobate da designazioni commerciali più efficaci come *laptop*.

I risultati del corpus Reddit, invece, rivelano un cambiamento profondo nella natura dei discorsi legati al personal computer, che si distaccano progressivamente dalla funzione classificatoria e progettuale osservata nei testi storici e tecnici. A differenza dei sottocorpora 1961-1984 e 1985-2024, dove le denominazioni rispecchiano l'evoluzione industriale e tecnologica dei dispositivi (*personal computer*, *notebook computer*, *laptop*, *portable computer*), nel corpus Reddit le etichette terminologiche vengono rinegoziate attraverso pratiche discorsive affettive, collettive e narrative.

Il contrasto più evidente riguarda il termine *console*, che nel primo sottocorpus indica sistemi di input/output specializzati, spesso in ambito militare o accademico ("*console display unit capable of presenting graphical data*"), mentre nei post Reddit è associato esclusivamente a dispositivi videoludici e alla cultura pop:

- *"My very first video game console was a Breakout/Basketball/Video Pinball combo made by Atari"*;
- *"Used for switching my console Gaming collection to my Second monitor"*.

L'oggetto non è più una "unità tecnica" ma un simbolo identitario.

Anche altri termini mostrano questo slittamento, come *retro computer*, che compare solo nel corpus Reddit, dove è marcato da una semantica di passione, collezionismo e comunità:

- *"People normally only care about my video game collection and ignore my love for retro computers"*;
- *"The NeXT Cube .This is one of the most beautiful and coolest on my collection"*;
- *"My best retro computer is my Amiga 500"* (Amiga 500 è nello specifico un home computer).

Nel corpus Reddit, la riflessione terminologica è spesso interna ai testi stessi: gli utenti non solo usano etichette differenti per descrivere dispositivi simili, ma manifestano esplicitamente la consapevolezza di questa pluralità, come nella frase*: "Portables, luggables, laptops, handhelds, whatever you call it, it was meant to travel."*

Nel corpus museale, la terminologia associata al concetto di *portable personal computer* mostra una stratificazione che riflette sia l'evoluzione tecnologica che la categorizzazione archivistica. Il termine *portable computer* assume un ruolo genealogico: compare frequentemente come etichetta ponte tra

le fasi precedenti dell'informatica (mainframe, desktop) e quelle successive, rappresentate da *laptop* e *notebook*. Alcune descrizioni lo collocano esplicitamente come tappa intermedia, sottolineando il passaggio dall'ingombro alla portabilità come svolta storica.

"*This is an early example of a portable computer. The 'portable' referred to the fact that the case contained the entire computer...*"

L'uso di *early example* segnala una posizione iniziale o fondativa, implicando un seguito evolutivo. L'attributo *portable* è spiegato nel contesto di un passaggio tecnico rilevante: il contenimento del sistema in un case unico, tipico della transizione tra desktop e laptop.

"*Weighing 10.7kg the Osborne 1 was the first system designed to be transported between the office and the home and was a precursor to the modern laptop."*

Questa frase esplicita il legame genealogico: il portable computer è presentato come un precursore diretto dei laptop.

Queste occorrenze giustificano l'interpretazione del termine *portable computer* come snodo genealogico, collocato in una linea evolutiva tra mainframe, desktop e laptop, che il corpus museale, per sua natura classificatoria, tende a rinforzare.

Al contrario, *laptop* si lega a marchi riconoscibili e a modelli emblematici (Grid Compass, Apple Newton eMate, IBM ThinkPad), e viene descritto in funzione delle sue caratteristiche tecniche, del design o dell'uso in contesti particolari (ad esempio missioni NASA), rafforzando così una visione museale e tecnico-documentaria. Infine, *notebook* risulta marginale nel corpus, usato quasi esclusivamente in chiave strumentale, per accompagnare manuali, dispositivi o software, senza valore narrativo autonomo.

Questo quadro si distingue nettamente da quanto osservato nel corpus Reddit, dove gli stessi termini subiscono una marcata ridefinizione discorsiva: *laptop* e *notebook* vengono spesso usati in modo intercambiabile, con toni affettivi e valoriali (*"whatever you call it, it was meant to travel"*), mentre etichette come *retro computer* o *console* emergono in associazione a pratiche nostalgiche e a memorie individuali. In una discussione si legge ad esempio:

"Interesting how many of the entries are palmtops, pocket PCs or whatever you want to call them. I don't think they sold much, but must have generated some geek envy to save or collect."

Questo tipo di commento mette in luce un uso linguistico meno normativo e più orientato all'esperienza personale e al valore simbolico dell'oggetto tecnologico. A differenza dei corpora tecnici e storici, in cui domina la classificazione funzionale, il corpus museale costruisce una rappresentazione ibrida, a cavallo tra catalogazione oggettiva e narrazione evolutiva. E proprio questa tensione tra linearità tecnologica e molteplicità discorsiva diventa uno snodo chiave nell'analisi comparativa dei corpora. L'analisi condotta sui quattro sottocorpora conferma che la terminologia

informatica non costituisce un sistema stabile e uniforme, bensì una rete dinamica, stratificata e fortemente situata. Le forme lessicali e discorsive emergenti nei corpora analizzati – storici, tecnici, museali e comunitari – mostrano come il linguaggio dell'informatica si adatti costantemente ai contesti d'uso, agli obiettivi comunicativi e alle comunità interpretative che lo plasmano.

A livello semantico e morfosintattico, le differenze tra le parti del discorso, le distribuzioni verbali e gli aggettivi dominanti rivelano traiettorie retoriche differenti: prescrittive nei testi tecnici, narrative nei post comunitari, classificatorie nei documenti museali. Gli strumenti impiegati, IRaMuTeQ, Tropes, TermoStat e Sketch Engine, hanno permesso di evidenziare la ricchezza di questo repertorio.

I dati mostrano che termini apparentemente equivalenti - *home computer*, *personal computer*, *PC*, *laptop*, *retro computer* - veicolano sfumature concettuali diverse a seconda del contesto d'uso, del periodo storico e della comunità interpretativa.

In tale prospettiva si esprime un principio fondamentale dell'uso terminologico del grande pubblico: non è tanto il nome a definire l'oggetto, quanto la funzione condivisa e il contesto d'uso. L'oggetto viene riconosciuto come appartenente a una categoria mobile e personale, indipendentemente dalla sua forma o dalla denominazione precisa.

Tale atteggiamento metalinguistico si discosta profondamente dal rigore classificatorio dei corpora tecnici e museali. Mentre i primi mirano a definizioni stabili e univoche, Reddit assume un paradigma pragmatico: la terminologia si adatta, si reinventa e si piega alle esigenze della memoria, della narrazione e dell'identità.

La varietà di etichette individuate nei corpora: *home computer*, *personal computer*, *PC*, *laptop*, *retro-computer*, non rappresenta semplicemente un'evoluzione lessicale, ma contraddice apertamente l'idea che il computer si sia sviluppato secondo una traiettoria unica, culminata nel PC moderno.

Come osserva Mahoney (2005), la narrazione dominante nella storia dell'informatica tende a presentare il PC come forma finale e naturale del computer, oscurando invece la varietà di configurazioni coesistenti (mainframe, minicomputer, console, home computer).

"Una volta inventato, il computer si evolve naturalmente nel PC come sua forma attuale più visibile, piuttosto che in una varietà di forme coesistenti e reciprocamente complementari"[33] (Mahoney, 2005).

Questa visione deterministica, che Mahoney chiama *teoria dell'impatto*, ignora il fatto che la tecnologia è anche appropriazione culturale, molteplicità d'uso, nomi e discorsi. La terminologia rilevata nei corpora Reddit e museale conferma che, anche nel presente, convivono denominazioni multiple e talvolta nostalgiche, che raccontano altri percorsi storici e affettivi del concetto stesso di "computer".

[33] Traduzione mia

Come sottolineato da Haigh, è impossibile comprendere pienamente l'informatica limitandosi alla sola dimensione tecnico-funzionale:

«L'uso della tecnologia informatica in uno specifico contesto sociale (come il laboratorio, l'ufficio o la fabbrica) non può essere compreso senza studiarne anche la storia precedente, le persone coinvolte e gli obiettivi per cui la macchina è stata adottata.» (Haigh, 2001)

Questa riflessione è formulata in un articolo pubblicato negli *IEEE Annals of the History of Computing*, una delle fonti che compongono il secondo corpus analizzato.

L'evidenza testuale raccolta nei corpora mostra proprio questo: le parole dell'informatica raccontano molto più della macchina in sé. Riflettono valori, nostalgie, usi, soggettività e appartenenze. In questo senso, la terminologia informatica non è solo uno strumento di descrizione, ma un dispositivo culturale che contribuisce a costruire la memoria condivisa della tecnologia.

Questa consapevolezza costituisce il punto di partenza per il capitolo successivo, dedicato al progetto **ITinHeritage**. Qui l'attenzione si sposterà dal piano descrittivo e analitico a quello applicativo e patrimoniale: la sfida sarà tradurre la variabilità linguistica e concettuale osservata in strumenti di rappresentazione terminologica e ontologica, capaci di valorizzare il patrimonio informatico in una prospettiva multilingue, comparativa e interoperabile.

4. CASO DI STUDIO: il progetto di ricerca ITinHeritage

4.1 Introduzione al progetto

Il patrimonio del calcolo e dell'informatica costituisce una categoria peculiare di beni culturali: esso è materiale (macchine, componenti, supporti), ma al tempo stesso immateriale (software, modelli di calcolo, architetture, paradigmi). La sua descrizione non può limitarsi né al livello documentario né a quello puramente tecnico, poiché gli oggetti informatici sono inscritti in storie di uso, appropriazione, obsolescenza e risemantizzazione.

Come mostrato nel capitolo precedente, questa complessità emerge anche a livello terminologico: le denominazioni adottate per riferirsi agli stessi oggetti o concetti possono variare in base al contesto d'uso, alla comunità discorsiva e al periodo storico. Termini come *home computer, personal computer, PC, laptop* o *retro computer* non sono equivalenti né neutri, ma riflettono differenti visioni culturali, pratiche d'uso e processi di ricategorizzazione. Ciò significa che la rappresentazione del patrimonio informatico non può essere pensata come una semplice raccolta di nomi, ma richiede la ricostruzione dei concetti sottostanti e dei criteri che ne motivano la differenziazione.

La crescente complessità del patrimonio culturale digitale, unita alla molteplicità delle fonti e delle pratiche descrittive nei contesti museali, impone una riflessione sulla necessità di strumenti semantici capaci di rappresentare la conoscenza in modo interoperabile, aggiornabile e sensibile alle specificità storiche e culturali. In questo quadro, il progetto ITinHeritage si propone di affrontare la variabilità terminologica non come un ostacolo, ma come una risorsa epistemica, adottando un approccio ontoterminologico capace di mettere in relazione designazioni linguistiche, modelli concettuali e oggetti materiali, in modo coerente e interoperabile.

Il progetto ITinHeritage è coordinato da Caroline Djambian (Università di Grenoble Alpes - UGA) e si sviluppa all'interno di un consortium interdisciplinare. Tra le figure scientifiche coinvolte nel progetto figurano Christophe Roche (Università di Creta), ERA Chair del progetto ITinHeritage, Micaela Rossi (Università di Genova) e Marie Gevers (Università di Namur)[34]. Il progetto nasce da una prospettiva patrimoniale e interdisciplinare e coinvolge diversi musei europei, tra cui il Musée des Arts et Métiers (MAM, Parigi), ACONIT – Association pour un Conservatoire de l'Informatique et de la Télématique (Grenoble), il Science Museum di Londra, il NAM-IP – Namur Institute for Digital Heritage (Namur), il Museo degli Strumenti per il Calcolo (Università di Pisa), l'HomeComputerMuseum (HCM, Helmond) e l'HNF – Heinz Nixdorf MuseumsForum (Paderborn).

[34] L'elenco completo e aggiornato dei membri del consortium, dei partner e delle istituzioni coinvolte è disponibile sulla piattaforma ITinHeritage, sezione Discover (*http://95.179.216.242/discover*)

Il settore dell'informatica, definito da Djambian come *"the emblem of contemporary scientific and technical heritage"*, rappresenta un dominio particolarmente complesso per la musealizzazione e la mediazione, poiché i suoi artefatti non offrono una immediata intelligibilità senza un adeguato apparato concettuale e terminologico di supporto (Djambian, 2023). A differenza di altri oggetti tecnico-scientifici, molti artefatti informatici non sono "mediatori" in sé: non comunicano visivamente la propria funzione, non rendono trasparenti i principi logici o computazionali che li governano e non permettono di inferire, senza conoscenze pregresse, la loro collocazione storica o tecnologica. Per questo motivo, la loro interpretazione e valorizzazione richiedono strumenti di mediazione museale e comunicativa capaci di rendere esplicite relazioni, processi e modelli che non sono direttamente leggibili sull'oggetto, ma che costituiscono il nucleo della conoscenza che esso incarna.

A partire da questa constatazione, il progetto assume come centrale la necessità di una rappresentazione concettuale e linguistica del patrimonio informatico che integri approcci terminologici, ontologici e corpus-based, con l'obiettivo di rendere tali conoscenze trasmissibili, comparabili e riutilizzabili al di là della sola esposizione museale, in un'ottica multilivello, multilingue (italiano, inglese e francese) e multidimensionale.

La domanda di ricerca che guida il lavoro può essere così formulata:

È possibile rappresentare, trasmettere e mediare in modo coerente, interoperabile e dinamico il patrimonio informatico presente nei musei, nonché quello tramandato attraverso la letteratura storica sull'informatica?

Questa riflessione riprende il nodo teorico evidenziato da Djambian: la necessità di strumenti capaci di rendere esplicite le relazioni concettuali e linguistiche che sostengono la comprensione degli artefatti informatici. In questa prospettiva, il progetto si interroga su come concetti scientifici e tecnici complessi possano essere trasmessi, spiegati e collegati a contesti più ampi attraverso strumenti terminologici e semantici (Djambian *et al.*, 2024).

4.2 Obiettivi, caratteristiche e fasi metodologiche della risorsa ITinHeritage

Questa sezione presenta gli obiettivi le caratteristiche e l'impianto metodologico della risorsa ITinHeritage, concepita come una piattaforma[35] web semantica per la rappresentazione e diffusione delle conoscenze relative al patrimonio del calcolo e dell'informatica (Djambian *et al.*, 2024). Il progetto mira a costruire una risorsa terminologica e ontologica dedicata, strutturata per essere:

- multilingue (italiano, inglese e francese), in modo da rappresentare la diversità linguistica dei musei e delle comunità di riferimento;
- multimodale, includendo componenti visive, descrittive e semantiche;
- interoperabile, secondo gli standard del Web Semantico (RDF, OWL, SKOS);
- aperta e aggiornabile, per accogliere nel tempo nuove acquisizioni museali e nuove interpretazioni storiche.

Il progetto si articola lungo più assi complementari, tra i quali il lavoro terminologico rappresenta uno degli elementi fondamentali, ma non esclusivi:

1. Costruzione di un corpus tematico per l'identificazione e classificazione dei termini e concetti rilevanti.
2. Creazione di un dizionario multilingue e multimediale (italiano, francese, inglese), accessibile anche a utenti non specialisti.
3. Analisi delle denominazioni e delle pratiche linguistiche nei contesti museali e storici, per favorire mediazione linguistica e traduzione.
4. Definire e formalizzare una rete concettuale, organizzata secondo assi di analisi (tecnici, funzionali, storici, materiali), capace di distinguere i concetti in base ai caratteri essenziali che li definiscono;
5. Formalizzazione ontoterminologica interoperabile, in cui la struttura concettuale è inizialmente modellata in Protégé e le designazioni linguistiche sono esplicitate e controllate in TEDI[36], secondo l'approccio ontoterminologico.

L'attenzione alla distinzione tra concetto e designazione linguistica, elemento centrale dell'approccio ontoterminologico, consente di rappresentare la variabilità terminologica senza perderne la coerenza concettuale, garantendo così una modellizzazione del patrimonio informatico accurata, trasparente e

[35] Piattaforma prototipale sviluppata nell'ambito del progetto ITinHeritage. Visualizzazione disponibile al link: http://95.179.216.242/

[36] TEDI (ontoTerminology EDItor) è un ambiente software sviluppato da Christophe Roche e dal gruppo di ricerca Condillac per la costruzione di risorse ontoterminologiche

riutilizzabile. Alla base, è fondamentale definire con chiarezza gli obiettivi e le caratteristiche della risorsa, rispondendo a domande fondamentali relative a destinatari, finalità e modalità di integrazione con sistemi esistenti.

La piattaforma risultante dovrà essere destinata a diversi tipi di pubblico, comprendendo sia ricercatori e professionisti del patrimonio culturale, sia utenti non specialisti, con lo scopo di supportare la ricerca accademica, la valorizzazione museale, la didattica e la conservazione digitale. La struttura della risorsa deve essere compatibile con formati strutturati standard (come XML, JSON e CSV), al fine di favorire l'apertura, la diffusione e il riutilizzo dei dati.

La costruzione del progetto ITinHeritage si sviluppa attraverso un insieme di fasi metodologiche integrate, che vanno dall'osservazione delle pratiche linguistiche alla formalizzazione ontoterminologica. Queste fasi non costituiscono una sequenza lineare, ma un processo iterativo e circolare: i risultati di ciascuna fase influenzano e ridefiniscono le fasi successive, permettendo un affinamento progressivo dei dati, dei modelli e delle rappresentazioni concettuali. La metodologia adottata combina analisi corpus-based, modellizzazione concettuale e formalizzazione semantica mediante strumenti di rappresentazione formale, nonché verifiche successive di coerenza interna ed esterna, con l'obiettivo di produrre una risorsa interoperabile, stabile e riutilizzabile.

4.2.1 Costruzione del corpus

La costruzione del corpus ha seguito un percorso in tre passi, pensato per coniugare copertura diacronica e aggiornamento rispetto a tecnologie in rapida evoluzione.

Il primo passo ha previsto una ricognizione strutturata delle risorse terminologiche e storico-documentarie (dizionari, glossari, enciclopedie, repertori museali, siti istituzionali). Questa fase ha avuto una doppia funzione: (i) valutare la qualità e l'adeguatezza delle fonti (lingue coperte, approccio semasiologico/onomasiologico, presenza di definizioni, aggiornamento editoriale); (ii) preparare un corpus complementare selezionato e giustificato. Il criterio guida è stato combinare opere che costituiscono dei "classici" della disciplina - utili a descrivere l'"evoluzione linguistica" nel tempo - con risorse recenti/aggiornate, in modo da intercettare nuovi usi e termini emergenti.

URL	NOME	AUTORE	ENTE	LUOGO,ANNO	LINGUE	TIPO RISORSA	TOPIC	NOTE
https://www.aconit.org/spip/spip.php?article234	Catalogue informatique de Henri Boucher	Henri Boucher	ACONIT		inglese->francese	glossario	Architetture, costruttori, software, macchinari	ONLINE, TERMINE INGLESE, DEFINIZIONE FRANCESE
https://studfile.net/preview/393947/	Dictionary of computing (5th edition)	S. M. H. Collin	Bloomsbury Publishing Plc	2004	inglese	dizionario		Online
https://www.isaca.org/-/media/files/isacadp/project/isaca/resources/glossary/isaca-glossary-english-italian_0320.pdf?la=en&hash=8309221CC8C13CD1CFD5BDF3273484F89F33CA14	Glossary of Terms	ISACA®		2015	inglese->italiano	glossario	TERMINI E DEFINIZIONI IN INGLESE E ITALIANO	pdf
https://www.isaca.org/-/media/files/isacadp/project/isaca/resources/glossary/isaca-glossary-english-french_mis_fra_0615.pdf?la=en&hash=FAD3B40FDBC0C70560088A9287FD7A038	//	//		//	inglese->francese	//	TERMINI E DEFINIZIONI IN INGLESE E FRANCESE	pdf
http://www.labinfca.unipr.it/glossario/gloss.htm	GLOSSARIO di terminologia informatica	Claudio Gucchierato Andrea Leotardi	Università di Parma		italiano	glossario		Online
https://isaaccomputerscience.org/glossary?examBoard=all&stage=all	Glossario	Isaac Computer Science	Università di Cambridge		inglese	glossario		Motore di ricerca ma scaricabile in pdf
https://www.oxfordreference.com/view/10.1093/acref/9780199688975.001.0001/acref-9780199688975	A Dictionary of Computer Science	Andrew Butterfield , Gerard Ekembe Ngondi , and Anne Kerr	Oxford University Press	2016	inglese	Dizionario		Cartaceo e online
http://www.fondazionegalileogalilei.it/museo/collezioni/strumenti_scientifici/str_scientifici.html	Strumenti Scientifici	MUSEO DEGLI STRUMENTI PER IL CALCOLO	Università di Pisa		italiano	glossario	Strumenti Scientifici	Online
http://www.museoscienza.it/dipartimenti/catalogo_collezioni/lista.asp?arg=Calcolo%20e%20Informatica&c=10	Calcolo e Informatica	MUSEO NAZIONALE SCIENZA E TECNOLOGIA LEONARDO DA VINCI		Milano	italiano	Dizionario della collezione	Strumenti e macchine	Online
http://www.museodelcomputer.org/index.php/nav=Catalogo.20	Elenco macchine diviso per marche	MUSEO DEL COMPUTER		Novara / Biella	italiano	Catalogo	Macchine divise per marca	Online
http://www.museotecnologicamente.it/category/collezione/	La collezione: OLIVETTI	Laboratorio-Museo Tecnologic@mente		Ivrea	italiano	Dizionario della collezione	Storia industriale della Olivetti	Online
http://histoire.info.online.fr/	Histoire de l'Informatique	Serge Rossi		2004	francese	glossario	Storia dell'informatica	Online
https://doc.lagout.org/Others/Encyclope	Computer Science and technology	Harry Henderson		2002/2009	inglese	enciclopedia	enciclopedia ILLUSTRATA sull'informatica e la tecnologia	
https://www.oxfordreference.com/display/10.1093/acref/9780199234004.001.0001/acref-9780199234004;jsessionid=25B5F5B5A833F4689B1C705CE603C842	The Oxford Dictionary of Computing	John Daintith and Edmund Wright	Oxford University Press	2008	inglese	dizionario	termini su argomenti di informatica, tra cui intelligenza artificiale, computer grafica e sistemi di database.	
http://www.pc-	Glossario Informatico	pc-facile (blog con news, info			italiano	glossario		

Figure 105 Tabella delle risorse terminologiche identificate nella fase di mappatura dello stato dell'arte.

Le risorse individuate sono state poi **schedate** mediante un template uniforme, progettato per rendere espliciti e confrontabili criteri, contenuti e limiti delle diverse fonti. Ogni voce riporta: dati anagrafici (autore, editore, anno, destinatari), approccio (semasiologico/onomasiologico), sottodomini coperti, lemmario (ampiezza, struttura, esempi), uso dell'iconografia e note critiche (licenze, limiti, paywall). La schedatura copre terminologie in francese, inglese e italiano e una sezione musei, con esempi quali Boucher/ACONIT, ISACA (EN→IT/FR), Oxford e Bloomsbury, repertori SUPSI, e i portali museali a tema calcolo/informatica. Questa base rende trasparenti criteri e scelte, consente confronti trasversali (ad es. tra dizionari generalisti e glossari di settore) e fornisce metadati metodologici per citazioni e riproducibilità.

Le opere di riferimento nel campo della storia dell'informatica e del calcolo hanno svolto un ruolo fondamentale per l'acquisizione di una cultura di dominio, per l'orientamento concettuale e per la comprensione delle principali tradizioni narrative e interpretative del settore. Tuttavia, tali opere non sono state utilizzate come fonti testuali per l'analisi terminologica o computazionale, né sottoposte a operazioni di text mining.

Le attività di estrazione e analisi linguistica sono state limitate ai dati testuali provenienti dalle collezioni museali, nell'ambito di un progetto di ricerca, nel rispetto delle condizioni di accesso e delle eccezioni previste per la ricerca scientifica. In questo senso, la schedatura non ha una funzione meramente descrittiva, ma costituisce uno strumento metodologico di controllo: consente di distinguere tra fonti analizzate e fonti di riferimento, chiarisce i diritti di utilizzo dei materiali e documenta le condizioni entro cui i dati sono stati impiegati come supporto all'analisi terminologica e concettuale.

Schedatura Risorse sull'Informatica

Sommario

Figure 106 Sommario delle risorse terminologiche e documentarie schedate (francese, inglese, italiano e musei

TERMINOLOGIE IN LINGUA FRANCESE

1. Catalogue informatique de Henri Boucher

Dati anagrafici
Data pubblicazione: Documentazione raccolta dal 1955 fino al 2000.
Editore: ACONIT (Associazione per la conservazione dell'informatica e della telematica)
Autori: Henri Boucher
Lille 1925 – 2021. Pioniere dei computer nella marina francese e nella difesa in generale.
Destinatari: Esperti / ricercatori

Lemmario
Numero dei concetti/lemmi analizzati: 300 (Volume A); 301-338 (Volume B); 339-548 (Volume C); 549-696 (Volume D); 700-745 (Volume E); 746-773 (Volume F). Totale = 769
Approccio: onomasiologico. Concetti enumerati e raccolti in 6 volumi compilati non in ordine alfabetico. La ricerca, dunque, viene facilitata dall'indice che divide i termini per argomento e in ordine alfabetico. Si tratta di una raccolta di informazioni a partire dallo studio di documentazioni specifiche del campo, in un dato territorio. Non è un caso che il volume B si presenta come un'autobiografia;

Sottodomini: 6 volumi (a-b-c-d di informatica americana, e-f informatica di altri Paesi)
Parte nel volume A con la storia dei costruttori; le altre categorie sono: Amministrazioni, Educazione, Aziende, Architetture Software, Macchinari, Periferiche, Personalità, Sistemi.
Lemmi: Termini (in alcuni volumi sono in lingua inglese ed altri in francese) con definizione dettagliata in lingua francese.

VOLUMI A-B-C-D
Contesto: Informazioni tratte dalla documentazione americana
Dominio: Storia delle macchine e dei suoi costruttori. (non in ordine logico o alfabetico)
Sottodomini
aziende: appaltatori; ricercatori; riferimenti temporali, di luogo, macchine prodotte, vendite riferimenti letterari;
macchine: descrizione (info generali), logica, software, hardware modelli, aziende produttrici, varianti linguistiche, componenti, prezzo.
Ogni informazione riguardante la descrizione o i domini sono variabili da un concetto all'altro. Non viene seguito uno schema preciso per caratterizzare i singoli termini.

Volume A
Ogni concetto viene caratterizzato da una descrizione di lunghezza variabile, con lo scopo di contestualizzarlo.
Volume B

Per ogni concetto viene descritta la storia, il contesto, l'evoluzione, in uno stile più narrativo e autobiografico (es. vengono esposte delle riflessioni/opinioni da parte dell'autore)
La descrizione dei concetti è meno strutturata. Per alcuni troviamo sempre variabili temporali, riferite al contesto, ai soggetti.
Volume C
Ogni concetto viene caratterizzato da una descrizione di lunghezza variabile, con lo scopo di contestualizzarlo.
Volume D
Ogni concetto viene caratterizzato da una descrizione di lunghezza variabile, con lo scopo di contestualizzarlo. Per alcuni concetti la descrizione è molto lunga, *vedi. 549*, il cui termine, inoltre, non è solo il nome dell'azienda (di cui viene raccontata la storia, i calcolatori, le annate ecc) ma più un titolo '*La sécurité avec Tandem*'.

VOLUMI E-F
Contesto: Informazioni tratte da una documentazione mondiale, esclusa quella americana.
Dominio: Storia dell'informatica con cenni alle macchine, ai costruttori, agli ideatori, mostrando grafici e percentuali.
Volume E
Storia dell'informatica in Germania, Inghilterra (più approfondite), Australia, Austria, Belgio, Canada, Cina, Corea del Sud, Cuba, Danimarca, Spagna, informatica europea, Finlandia. (non in ordine logico o alfabetico)
Volume F
Storia dell'informatica in Grecia, Olanda, Israele, Italia, Asia sud-est, Norvegia, Nuova Zelanda, Svezia, Svizzera, Irlanda, Taiwan, Paesi dell'est (URSS, Ungheria, Polonia, Romania, Cecoslovacchia, Germania dell'est

Figure 107 Esempio di scheda analitica delle risorse terminologiche, con dati anagrafici, approccio, sottodomini e struttura del lemmario.

La sequenza metodologica adottata per la costruzione del corpus (stato dell'arte → schedatura → corpus museale) risponde a tre esigenze principali:

- Il foglio stato dell'arte consente di giustificare la scelta delle fonti e di aggiornare rapidamente la bibliografia di lavoro.
- La schedatura rende confrontabili risorse molto eterogenee (dizionari wüsteriani con definizioni + schemi/iconografia; glossari didattici; repertori museali con immagini e schede) e prepara il terreno per definire quali campi sfruttare in estrazione.
- Il corpus museale fornisce l'ancoraggio agli oggetti reali e alle pratiche curatoriali, requisito centrale del caso di studio ITinHeritage.

A valle della ricognizione, il corpus museale (oltre 25.500 record) è stato acquisito dai partner e normalizzato per l'analisi: schede oggetto, descrizioni, note curatoriali, campi tecnici (materiali, misure, produttore, inventore), periodizzazioni e asset multimediali. In parallelo sono stati analizzati dataset già strutturati (es. ACONIT, Londonc Science Museum, HNF) per calibrare tipologie (macchine, documenti, software, iconografia) e quantità, e per definire regole di bonifica (gestione di caratteri speciali, virgolette spurie, duplicazioni/righe spezzate; interpretazione dei "pallini" come meronimia). Questo garantisce che la variabilità linguistica sia conservata come dato e non eliminata come "rumore", a beneficio delle fasi terminologiche e concettuali successive.

La scelta di affiancare fonti testuali e iconografiche riprende l'impostazione wüsteriana dei dizionari tecnico-scientifici (termini + schemi/immagini) e si riflette nelle pratiche museali, in cui le schede combinano testo, metadati strutturati e media. In continuità con il quadro normativo delineato nel Capitolo 1 (ISO 704; ISO 1087), il lavoro di selezione delle fonti, di costruzione del corpus e di estrazione linguistica si colloca nel quadro definito dalla norma ISO/DIS 5078:2023, che fornisce i criteri per la costituzione di corpora terminologici a partire da dati testuali. Tali criteri sono stati declinati nel progetto ITinHeritage in relazione alla specificità dei metadati museali e del patrimonio informatico, come discusso in Djambian *et al.* (2024).

La costruzione del corpus ha seguito inizialmente una logica ibrida semasio-onomasiologica: partire dai testi per osservare le designazioni in uso e, in parallelo, orientare l'estrazione alla ricostruzione dei concetti. L'esigenza (emersa in fase di progettazione) è duplice: coprire l'evoluzione diacronica della terminologia e allo stesso tempo aggiornare il lessico secondo le tecnologie emergenti. Ne deriva un corpus stratificato, diacronico e aggiornabile, che combina fonti museali e fonti storico-terminologiche. La costruzione di un corpus tematico viene così concepita come base empirica per l'analisi terminologica e per la successiva modellizzazione concettuale.

Il corpus è stato quindi articolato in due componenti principali:

- Il **corpus principale** che raccoglie i metadati provenienti dalle collezioni dei musei partner (oltre 25.500 voci), che documentano non soltanto gli oggetti ma anche le loro descrizioni, le note curatoriali, le provenienze e i contesti d'uso. Si tratta della sezione più ampia e strutturata del corpus, che verrà presentata nel dettaglio nelle pagine successive.
- Il **corpus complementare** che è composto da testi storici e opere specialistiche (es. *Histoire illustrée de l'informatique* di Lazard e Mounier-Kuhn), selezionati con il supporto degli esperti museali al fine di rappresentare le principali tradizioni narrative e interpretative del dominio.

La fase iniziale ha previsto la consultazione di bibliografie specialistiche, la selezione di consulenti esperti e la costruzione di un corpus tematico che, secondo Cabré (1999), deve essere pertinente, completo e aggiornato. La scelta di integrare cataloghi museali e testi di riferimento non ha una funzione meramente quantitativa, ma risponde all'obiettivo di ricostruire e documentare la variabilità delle denominazioni e delle rappresentazioni concettuali del patrimonio informatico, così come essa si manifesta nelle pratiche di descrizione, conservazione e mediazione adottate in ambito tecnico, curatoriale e divulgativo.

Per garantire qualità e rappresentatività dei dati, sono stati definiti criteri relativi a: dimensione (equilibrio tra estensione e gestibilità), copertura tematica, tipologia testuale, livello di specializzazione e arco cronologico. Tali criteri assicurano che il corpus sia effettivamente in grado di riflettere la varietà del patrimonio informatico e delle comunità che lo hanno prodotto, utilizzato e interpretato - varietà e dinamiche discorsive discusse nel capitolo 3.

4.2.2 Estrazione terminologica e creazione di un dizionario

L'estrazione terminologica è stata condotta secondo i principi della norma ISO/DIS 5078:2023, adottando un approccio multilivello che combina in modo complementare procedure manuali, semi-automatiche e automatiche.

- Estrazione manuale: svolta da esperti del dominio, che identificano le designazioni rilevanti sulla base della loro conoscenza specialistica e delle pratiche curatoriali consolidate.
- Estrazione semi-automatica: ottenuta mediante strumenti linguistici e statistici (ad es. TF-IDF), in cui la capacità computazionale è integrata dal giudizio umano, che valida, filtra e interpreta i risultati.
- Estrazione automatica: affidata a software specializzati in grado di individuare termini sulla base di metriche frequenziali, modelli distribuzionali e pattern ricorrenti all'interno dei testi.

Secondo ISO/DIS 5078:2023(F), l'estrazione terminologica consiste nell'identificazione e nel trattamento dei dati terminologici e può essere effettuata attraverso modalità manuali, automatiche o semi-automatiche. La norma riconosce inoltre l'estrazione basata su regole, che utilizza schemi testuali, configurazioni morfosintattiche o metadati (tag) per individuare unità terminologiche che non emergono da approcci puramente frequenziali (ISO/DIS 5078:2023(F), §4.4.3.3).

Gli strumenti utilizzati per l'estrazione terminologica includono **Sketch Engine**, **TermoStat** e **Python** (libreria NLTK), impiegati in modo complementare lungo fasi successive di raffinamento.
L'impiego di Sketch Engine e TermoStat si colloca nel quadro delle procedure già descritte nel Capitolo 3, in particolare nella sezione dedicata agli strumenti di analisi lessicale e terminologica nei diversi sottocorpora (*cfr.* 3.2, pag. 116/126), con attenzione alla sezione dedicata all'analisi distribuzionale e contestuale delle denominazioni del personal computer.
Con NLTK sono state estratte le prime keyword**,** i bigrammi e i trigrammi, applicando una pulizia preliminare del lessico (rimozione della punteggiatura e delle stopwords: congiunzioni, avverbi, preposizioni). Mediante espressioni regolari (regex) è stato possibile estrarre grammatiche specifiche, come sequenze di acronimi o termini che iniziano con lettera maiuscola seguita da una serie di numeri. Tali pattern — spesso denominazioni di macchine, modelli o produttori (ad es. Gamma 30**,** IBM 1130**,** IBM 608) — tendono a essere separati dagli estrattori automatici, ma per noi rappresentano un'importante fonte di informazione.
I dati estratti sono stati consolidati in un file Excel unico, ottenendo un primo livello di lessico controllato, poiché i metadati museali di origine presentano già un grado di normalizzazione interna. Le prime fasi di estrazione (con Sketch Engine e TermoStat) hanno evidenziato rumore su keyword

puramente frequenziali; per questo si sono impiegate regex mirate per acronimi e forme ibride e, con NLTK, si sono estratti bigrammi/trigrammi e sintagmi nominali, escludendo deliberatamente POS verbali/aggettivali e articoli e impostando la lingua di lavoro nel codice.

Le strutture complesse (es. acronimi seguiti da numeri) richiedono una gestione ad hoc: sono stati sviluppati script personalizzati per mantenerne l'integrità.

Il lessico derivato dai metadati museali costituisce già una base parzialmente controllata; resta comunque indispensabile l'intervento umano. L'estrazione di strutture grammaticali complesse, come sintagmi nominali o acronimi seguiti da numeri, ha presentato sfide significative: in più casi è stata necessaria una gestione ad hoc tramite script personalizzati, per mantenere l'integrità delle designazioni ed escludere categorie non pertinenti (ad es. POS verbali o aggettivali, articoli), anche specificando la lingua di lavoro nel codice. Nel nostro caso l'attenzione specifica agli artefatti — usati come "specchio" della disciplina — costituisce una difficoltà ma anche un valore aggiunto: vincola la modellizzazione ai vincoli materiali e storici degli oggetti, rafforzando l'aderenza tra concetti e realtà curatoriali.

Il dizionario non sarà soltanto multilingue, ma anche multimediale, integrando componenti visive e audio. Sarà concepito come Knowledge Organization System (KOS), una categoria che l'enciclopedia della ISKO definisce secondo una duplice prospettiva: in un senso ristretto (*narrower*), esso si riferisce a strumenti funzionali come sistemi di classificazione, thesauri e ontologie intesi come una selezione di concetti legati da specifiche relazioni semantiche; in un senso ampio (*broader*), il concetto di KOS può estendersi fino a includere biblioteche, database e modelli di attività sociali. Adottando questa struttura, il dizionario sarà capace di analizzare le varietà linguistiche a livello diacronico e sincronico.

In particolare, l'integrazione di immagini, schemi e materiali iconografici consente di supportare la comprensione concettuale di termini fortemente legati alla materialità degli oggetti, come nel caso di componenti hardware (ad es. scheda perforata, microprocessore, unità di memoria) o di dispositivi storici, la cui funzione e collocazione tecnica risultano difficilmente intelligibili attraverso la sola definizione verbale. Analogamente, la dimensione multilingue permette di evidenziare variazioni terminologiche e concettuali tra lingue e comunità di pratica, ad esempio nel confronto tra denominazioni come home computer, personal computer e micro-ordinateur, o tra termini stabilizzati a livello tecnico e forme più divulgative o museali.

L'analisi multilingue introduce sfide aggiuntive: come ricorda Vezzani (2023), «quante più lingue vengono aggiunte a una banca terminologica, maggiore è la complessità concettuale del compito». Questa molteplicità, tuttavia, costituisce anche una risorsa, poiché consente di valorizzare le specificità linguistiche e culturali ed evitare una standardizzazione eccessiva. Il dizionario, ispirato

all'approccio wüsteriano e costruito secondo gli standard ISO 704 e ISO 1087, si fonda sulla dottrina dei padri fondatori viennesi, Wüster e Felber ed è concepito non solo in più lingue ma anche con componenti visive e audio, per favorire l'accessibilità e la comprensione da parte di pubblici differenti.

Tale modello è caratterizzato da un'elaborazione "sistematica" del vocabolario, mirata a ridurre l'ambiguità attraverso il superamento di polisemia e sinonimia. Come sottolineato da Humbley (1997), questo metodo permette di rappresentare il concetto come una unità inserita in un sistema di relazioni, tipicamente visualizzato attraverso diagrammi a albero (*tree diagrams*) che mettono i termini in relazione reciproca. Questo approccio risulta particolarmente efficace nei "micro-domini" tecnici, dove la strutturazione omogenea dei concetti facilita la precisione.

Il processo di raccolta dei termini rappresenta una fase cruciale nella costruzione del dizionario e si basa su due linee principali: l'analisi di risorse terminologiche preesistenti (glossari, vocabolari e thesauri specifici del dominio), che forniscono termini già consolidati e la costituzione di un corpus tematico capace di rappresentare la diversità di fonti e pratiche nel settore.

Il corpus è progettato per includere testi storici, documentazione tecnica e risorse provenienti dai musei partner, assicurando così una rappresentatività adeguata della terminologia utilizzata nel patrimonio informatico. I criteri di selezione (campione–popolazione) garantiscono che l'insieme dei termini raccolti sia rappresentativo dell'intero dominio del patrimonio informatico; la scelta dei testi tiene conto di tipologia, livello di specializzazione, arco cronologico e tematiche trattate.

Uno degli aspetti terminologici più innovativi del progetto è proprio la creazione di un dizionario multilingue e multimediale che rappresenti non solo la terminologia specifica legata al patrimonio informatico, ma anche le sue dimensioni storiche, culturali e linguistiche. Sebbene la terminologia per la rappresentazione del patrimonio culturale esista da tempo, la specifica applicazione alla sfera dell'informatica, soprattutto in una prospettiva diacronica, è ancora limitata. Questa impostazione consente di trarre spunti di riflessione dai risultati delle analisi terminologiche (lessico–semantiche) e concettuali, con particolare attenzione alle variazioni diacroniche dei concetti informatici e alle dinamiche linguistiche che emergono dall'interazione tra diverse lingue. Il dizionario è concepito come strumento di mediazione terminologica, in grado di rendere esplicite tali variazioni senza ridurle a un'unica forma standardizzata.

Per le estrazioni lessicali sono stati individuati metadati specifici, secondo criteri di ricchezza informativa e utilità per la modellizzazione.

Per il corpus in lingua francese ACONIT: *Name + Model, Description* e *Use*; inoltre, i campi relativi a *costruttori* e *inventori* sono risultati utili per l'identificazione e la documentazione delle designazioni terminologiche.

Tali campi sono particolarmente ricchi dal punto di vista argomentativo pur essendo metadati già parzialmente "vocabolarizzati". Con *vocabolarizzati* si intende che questi campi sono stati prodotti all'interno di pratiche descrittive museali consolidate, in cui la terminologia è controllata almeno in parte da standard interni (linee guida compilative, schemi catalografici, thesauri adottati dall'istituzione, convenzioni redazionali). Di conseguenza, il lessico impiegato tende a essere relativamente stabile, omogeneo e coerente.

Allo stesso tempo, però, questi campi come *Description* e *Use*, non si limitano a elencare attributi formali dell'oggetto, ma includono elementi narrativi e interpretativi: spiegano il contesto d'uso, chiariscono la funzione tecnica, forniscono riferimenti storici, specificano trasformazioni, ricostruiscono filiazioni progettuali o tecnologiche. È in questa componente discorsiva che si concentra la ricchezza argomentativa utile per l'analisi terminologica e concettuale.

Number of results: 1222

Selection: machine document logiciel objet

Tick all / none Select and display

Inventory ref.	Type	Name	Nbr of medias	Vignette
ACONIT 11001SE	Hardware	Set micro-ordinateur T1600 - Laptop (Toshiba Corporation)	4	
ACONIT 11002	Hardware	Power supply Adaptateur courant continu pour T1600 - AC Battery pack (Toshiba Corporation)	1	
ACONIT 11004	Document	Manual : Guide de l'utilisateur du Toshiba 1600 (Toshiba Corporation - Toshiba Corporation)	1	
ACONIT 11005	Document	Manual : Manuel de référence du Toshiba 1600 (Toshiba Corporation - Toshiba Corporation)	1	
ACONIT 11007	Hardware	Microcomputer portable T2000 SXE (Toshiba Corporation)	6	
ACONIT 11013	Hardware	Electronic board Printed board P880 (Philips)	2	

Figure 108 A titolo illustrativo, la figura mostra l'interfaccia pubblica di ricerca del database ACONIT, attraverso la quale sono resi accessibili alcuni dei campi descrittivi impiegati nel processo di analisi terminologica (Type,Name..)

title	type	ident	domain	constructor	city	country	material	measurer	museum of	created	image	description	copyrights	language	type
di Thomas d	aritmometro	1395	FISICA - INF	Payen, Louis	Toscai	ITALIA	materiali vai	heightxle	Museo degli	1865-1887	MSC_00	Realizzato come prototipo già nel 1822 da Charles Xavier deColmar, fu regolarmente prodotto in serie a partire dal 1850circa, iniziando cosà¬ l'era degli strumenti di calcolo commerciali.Sui cursori si impostano le cifre degli addendi (o delmoltiplicando), i cursori spostano delle ruote dentate che ingranando su altrettanti cilindri di Leibniz trasmettono il valore all'accumulatore posto sul carrello scorrevole. Su questo modello le cifre del risultato e del contatore possono essere modificate a mano.	https://crea	Italian	machine
ANITA Mk 9	calcolatrice	1395	FISICA - INF	Kitz, Norber	Toscai	ITALIA	materiali vai	heightxle	Museo degli	1961-1966	MSC_00	L'Anita fu la prima calcolatrice elettronica. Il principio di funzionamentoè basato sull'aritmetica in base 10. Il modello esposto introduceva la possibilità di impostare l'operando con il valore del risultato.	https://crea	Italian	machine
Brunsviga B	calcolatrice	1395	FISICA - INF	Trinks, Franz	Toscai	ITALIA	materiali vai	heightxle	Museo degli	1892-1927	MSC_00	La calcolatrice meccanicaè a ruote a denti mobili rendendo la macchina più compatta e veloce da usare. Come in molte macchine che nel contatore non hanno i riporti, il numero 9è in rosso per segnalare all'operatore di spostare il carrello.	https://crea	Italian	machine
Burroughs C	calcolatrice	1395	FISICA - INF	Burroughs, \	Toscai	ITALIA	materiali vai	heightxle	Museo degli	1892-1920	MSC_00	Burroughs fu il marchio più noto nella produzione di macchine contabili a tastiera estesa e scriventi. La Class 1è dotata sul retro di stampante a rullo. Il vetro frontaleè necessario per leggere il risultato sull'accumulatore, gli altri sono un omaggio all'estetica della meccanica.	https://crea	Italian	machine
Comptomet	calcolatrice	1395	FISICA - INF	Felt, Dorr	Toscai	ITALIA	materiali vai	heightxle	Museo degli	1911-1913	MSC_00	Calcolatrice meccanica a pressione di tasti. Per favorire l'identificazione dei tasti, le colonne sono state colorate diversamente e, per le righe, i tasti sono alternativamente piatti e concavi. Benché sia essenzialmente un'addizionatrice, può eseguire anche sottrazioni, moltiplicazioni e divisioni. I tasti sopra il visualizzatore inibiscono il riporto; servono per fare la sottrazione in complemento, nella quale il riporto sulla cifra più significativa non deve essere considerato. I tastini sopra il visualizzatore inibiscono il riporto; introdotti anch'essi con il modello A, servono per fare la sottrazione in complemento, nella quale il riporto sulla cifra più significativa non deve essere considerato. Per esempio, per fare 100 - 66 si deve impostare 100, poi tenendo premuto il tastino corrispondente alle centinaia, si deve battere 066 considerando i complementi ovvero i numeri piccoli (cioà¨ 933) e aggiungere 1. Il risultato sarà l'atteso 34.	https://crea	Italian	machine
Everest M4	calcolatrice	1395	FISICA - INF	Serio spa	Toscai	ITALIA	materiali vai	heightxle	Museo degli	1958-1962	MSC_00	Calcolatrice elettrica molto particolare per l'adozione di un disco combinatore telefonico per la composizione del moltiplicatore. Per moltiplicare bisogna prima "dichiarare" il numero di cifre del moltiplicatore premendo i tasti degli zeri. Il limiteè implicito nella capacità del risultato: avendo 11 cifre, al massimo si possono moltiplicare due numeri di cinque cifre.	https://crea	Italian	machine
Facit LX	calcolatrice	1395	FISICA - INF	Rudin, Karl	Toscai	ITALIA	materiali vai	heightxle	Museo degli	1938-1954	MSC_00	Il progettista propose un cambiamento radicale all'architettura di questa macchina: l'accumulatore, normalmente su un carrello mobile, divenne fisso ed il tamburo divenne mobile. La velocità di utilizzo era comparabile alle macchine a tastiera estesa e l'ingombro alle macchine a cursori. L'esemplare espostoè un modello con carrello tradizionale.	https://crea	Italian	machine
Marchant P	calcolatrice	1395	FISICA - INF	Marchant, R	Toscai	ITALIA	materiali vai	heightxle	Museo degli	1913-1921	MSC_00	Questo modello di calcolatrice meccanica ha il ripetitore per visualizzare il valore impostato sui cursori e un meccanismo di scorrimento del carrello mediante due tasti.	https://crea	Italian	machine
												Calcolatrice elettromeccanica a tastiera estesa prodotta negli Stati Uniti. Il meccanismo a ruote proporzionali si basava su ingranaggi che giravano a velocità			

Figure 109 Campi selezionati dal corpus del Museo degli Strumenti per il Calcolo di Pisa (title, type, description) per l'analisi lessicale e concettuale.

A valle della prima estrazione terminologica, si è proceduto a una pulizia mirata del corpus, volta a eliminare elementi che avrebbero potuto interferire con il riconoscimento automatico delle unità lessicali significative. Tale normalizzazione è stata effettuata tramite script Python e operazioni mirate in Excel, con l'obiettivo di garantire coerenza e stabilità nella forma delle designazioni. Sono stati rimossi:

- segni tipografici (*, «», ", (), …),
- note descrittive o editoriali non rilevanti,
- accenti e doppie spaziature non codificate

Per verificare l'effettivo impatto di tali anomalie sugli strumenti di estrazione, è stato predisposto un file di test contenente deliberatamente casi-limite (elenchi, virgolette eterogenee, spazi irregolari, sequenze testuali complesse o frammentate). L'analisi condotta con Sketch Engine ha confermato che il file .txt viene interpretato come blocco continuo di testo, senza perdita di segmentazione, e che le unità terminologiche complesse vengono correttamente riconosciute (ad es. *ordinateur portable, central processing unit*). Nel caso di designazioni composite (come *ordinateur portable Toshiba T3100/20*), si è osservata la separazione parziale in componenti (Toshiba + T3100), pur mantenendo la recuperabilità della multiword principale: una limitazione nota ma gestibile nella fase di normalizzazione.

In questo contesto si colloca l'uso di CQL in Sketch Engine. Per individuare in modo sistematico forme ibride tipiche delle denominazioni museali - come Gamma 30, Olivetti Programma 101, IBM 1130 - è stata condotta una scansione morfosintattica sul corpus ACONIT tramite CQL (Corpus Query Language).

Esempio di query utilizzata:

[tag="N.*"] [tag="Z.*"]

La ricerca identifica un nome (POS N) seguito da una cifra (POS Z), restituendo oltre 4.000 occorrenze nel volume A di ACONIT. Questo metodo ha permesso di isolare automaticamente pattern che gli estrattori standard tendevano a frammentare, fornendo un livello intermedio di estrazione fondamentale per riconoscere modelli, serie produttive e altre designazioni tecnico–numeriche tipiche dei metadati museali.

Per l'estrazione terminologica sono stati selezionati i campi che, all'interno delle schede museali, presentano il più alto grado di densità semantica e variabilità discorsiva. In particolare, sono stati utilizzati in *Sketch Engine* i metadati:

- ACONIT_description
- ACONIT_use

- Pisa_description
- LONDON_description

Questi campi contengono descrizioni estese, riferimenti funzionali e informazioni contestuali sull'uso degli oggetti, e dunque costituiscono una fonte privilegiata per l'identificazione delle unità terminologiche e dei potenziali nuclei concettuali. Queste osservazioni sono state utilizzate per definire le regole di normalizzazione per la fase successiva di costruzione del lessico controllato.
Sono stati testati diversi strumenti (Sketch Engine, TermoStat e Python NLTK), con la finalità di confrontarne:

- capacità di riconoscimento delle multiwords,
- sensibilità ai contesti museali,
- gestione dei modelli e delle sequenze alfanumeriche,
- robustezza nei confronti di rumore e irregolarità formali della codifica e della rappresentazione grafica.

Tra le soluzioni sperimentate, **Sketch Engine** ha restituito i risultati più coerenti e affidabili, sia in termini di specificità terminologica, sia nella capacità di mantenere la struttura delle multiword necessaria alla fase successiva di costruzione del lessico controllato. Per tale ragione è stato adottato come strumento principale nella fase di estrazione.
È stata successivamente effettuata un'ulteriore pulizia del corpus, con l'eliminazione di segni di punteggiatura potenzialmente rumorosi (ad es. asterischi, doppie virgolette, parentesi graffe, puntini di sospensione), la rimozione di note marginali e la correzione di alcuni accenti irregolari. Queste operazioni hanno contribuito a migliorare la stabilità grafica delle designazioni e a ridurre interferenze nei processi di tokenizzazione.
Un caso emblematico da citare nuovamente è costituito da designazioni composite come *ordinateur portable Toshiba T3100/20*, che in Sketch Engine vengono segmentate in due single words separate (*Toshiba* e *T3100*, con omissione del "20") e in una multiword autonoma (*ordinateur portable*). Per questo motivo, i termini estratti da Sketch Engine dai campi description e use sono risultati perlopiù complessi e sono stati integrati con le designazioni presenti nel campo name, inserite e corrette manualmente. In fase di normalizzazione, le liste così ottenute sono state sottoposte a operazioni di pulizia sistematica: eliminazione di duplicati, riconduzione al singolare, rimozione di forme fuori dominio, in coerenza con le raccomandazioni della norma ISO 5078.
Ad esempio, a partire dal campo *description* del corpus ACONIT (circa 1500 righe, caratterizzate da frasi complesse e variabili in lunghezza), è stata condotta in Sketch Engine un'estrazione delle

multiword terminologiche tramite il modulo *Keywords (Advanced)*, selezionando l'opzione *Identify terms* (multi-word terms) e confrontando il corpus con il corpus di riferimento French Web 2023 (frTenTen23). Il confronto con un corpus web generale ha permesso di evidenziare le espressioni salienti del dominio, filtrando automaticamente le occorrenze non specifiche.

L'estrazione ha prodotto una lista iniziale di 5000 multiword candidate:

1	method name: extract_keywords					
2	corpus: user/giada.dippolito/aconit_description_3					
3	subcorpus: -					
4	Item	Frequency (focus)	Frequency (reference)	Relative frequency (focus)	Relative frequency (reference)	Score
5	unité centrale	192	13475	1389,19031	0,48335	937,2
6	lecteur de disquettes	99	3286	716,30127	0,11787	641,67
7	face arrière	83	13667	600,5354	0,49024	403,65
8	lecteur de disquette	51	3781	369,00369	0,13562	325,82
9	station de travail	63	14360	455,82809	0,51509	301,52
10	type azerty	34	75	246,00246	0,00269	246,34
11	micro-ordinateur portable	30	253	217,06099	0,00908	216,1
12	touche de fonction	31	2990	224,29636	0,10725	203,47
13	pavé numérique	39	10961	282,17929	0,39317	203,26
14	tube cathodique	33	5064	238,76709	0,18165	202,91
15	disquette vierge	28	243	202,59026	0,00872	201,83
16	vierge enregistrée	27	0	195,35489	0	196,36
17	clavier alphanumérique	27	560	195,35489	0,02009	192,49
18	circuit intégré	40	14631	289,41467	0,52481	190,46
19	disquette vierge enregistrée	26	0	188,11952	0	189,12
20	carte perforée	29	3527	209,82562	0,12651	187,15
21	clavier azerty	30	4908	217,06099	0,17605	185,42
22	fond de panier	26	1073	188,11952	0,03849	182,11
23	touche de fonctions	25	1042	180,88416	0,03738	175,33
24	prise d" alimentation	26	2412	188,11952	0,08652	174,06
25	plastique beige	24	140	173,64879	0,00502	173,78
26	crayon optique	24	487	173,64879	0,01747	171,65
27	bande magnétique	36	14744	260,47321	0,52887	171,02
28	plastique noir	28	9868	202,59026	0,35397	150,37
29	deux connecteurs	22	3640	159,17805	0,13057	141,68
30	clavier intégré	20	812	144,70734	0,02913	141,58
31	deux lecteurs	20	2901	144,70734	0,10406	131,97
32	bouton marche	21	5241	151,9427	0,188	128,74
33	plastique gris	18	1342	130,2366	0,04814	125,21
34	arrière de la machine	18	1347	130,2366	0,04832	125,19
35	machine à écrire	29	19365	209,82562	0,69462	124,41
36	numéro série	17	428	123,00123	0,01535	122,13
37	écran cathodique	18	2417	130,2366	0,0867	120,77

extract keywords

Figure 110 Output del modulo Keywords (Advanced) di Sketch Engine per l'estrazione delle multiword dal campo description del corpus ACONIT.

Successivamente i termini candidati sono stati sottoposti alle seguenti operazioni di normalizzazione:

1. ordinamento per frequenza e salienza terminologica (keyness[37]),
2. riorganizzazione alfabetica,
3. deduplicazione delle forme ricorrenti,
4. riconduzione delle occorrenze al singolare,
5. filtraggio in base alla pertinenza concettuale rispetto al dominio.

[37] Il parametro di *keyness* misura il grado di salienza terminologica di una forma confrontandone la frequenza nel corpus analizzato con quella in un corpus di riferimento; valori elevati indicano termini statisticamente distintivi del dominio (https://www.sketchengine.eu/documentation/simple-maths/)

I termini ridondanti e le varianti formali sono stati annotati in una colonna dedicata, così da selezionare successivamente il termine preferito.

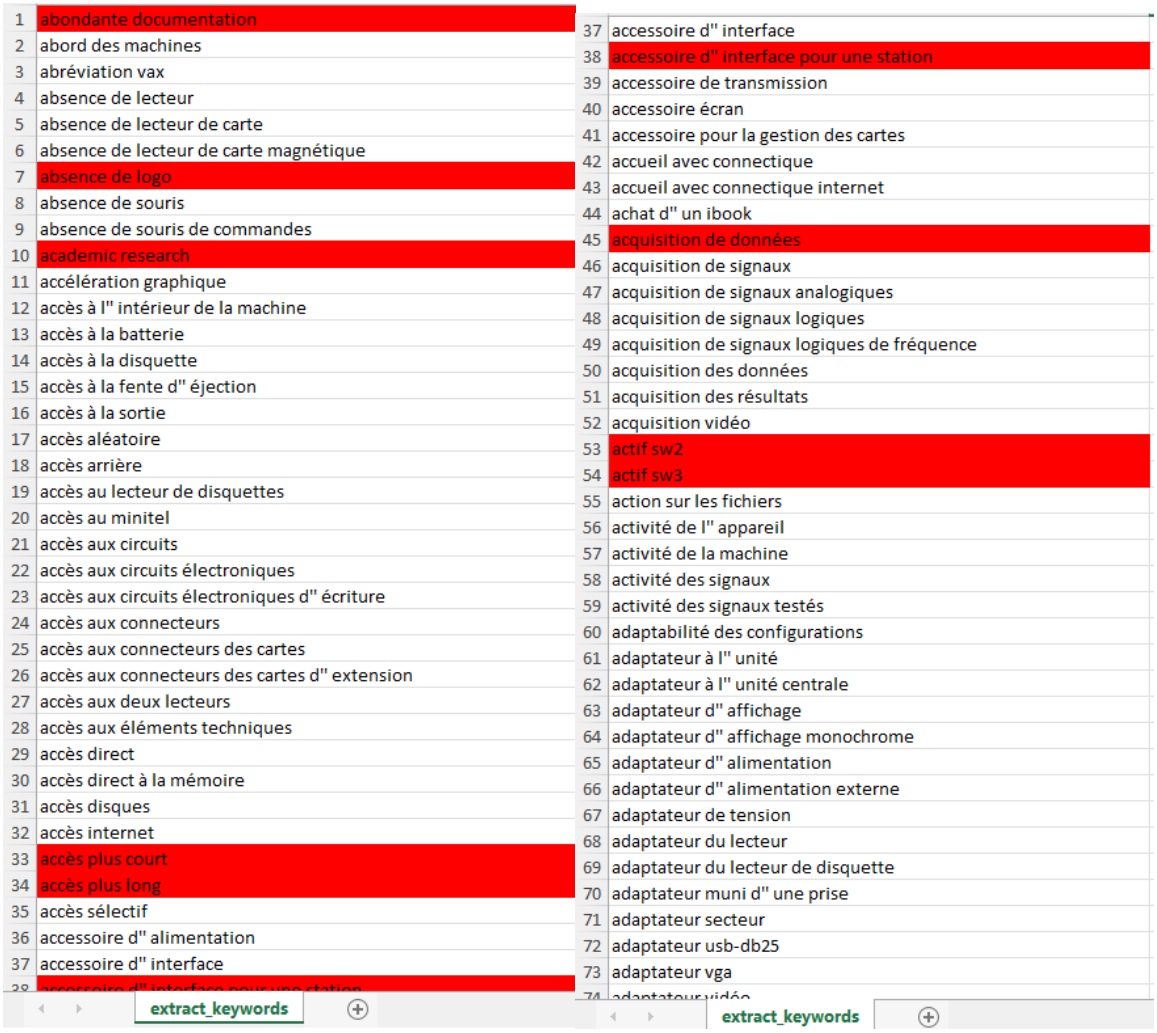


Figure 111 Lista delle multiword estratte dal corpus ACONIT, con i termini evidenziati in rosso corrispondenti alle forme eliminate nella fase di pulizia.

In questa fase è stato deliberatamente privilegiato il mantenimento delle designazioni complesse, in accordo con quanto indicato in ISO 5078 *(§ 4.4.6.2.1 Généralités*), al fine di favorire la loro successiva scomposizione e definizione nella fase di modellizzazione concettuale.

L'applicazione di tali procedure ha permesso di ridurre l'insieme iniziale delle 5000 multiword a 765 termini selezionati. A questi sono stati successivamente integrati i risultati ottenuti con la stessa metodologia applicata al campo *use*, nonché i glossari relativi a costruttori, inventori, nomi commerciali e modelli delle macchine (campi *name* e *model*), consolidando così un lessico ampio ma controllato, rappresentativo tanto delle funzioni tecniche quanto delle pratiche discorsive del patrimonio informatico.

L'estrazione terminologica svolta sui due sottocorpora (corpus museale e corpus storico-specialistico) ha permesso di evidenziare una prima serie di designazioni significative. Tuttavia, già in questa fase

è emerso con chiarezza il limite di un approccio puramente **semasiologico** (cfr. cap. 1): in un dominio ampio, storico e fortemente innovativo come quello informatico, i termini:

- variano nel tempo (diacronia),
- cambiano significato secondo il contesto d'uso e la comunità (es. museale, tecnica, amatoriale),
- possono essere sinonimici o polisemici (es. *laptop, notebook, personal computer*),
- dipendere dalla marca, serie o modello (*IBM 1130, Gamma 30, HP-35*), elementi che gli algoritmi di estrazione tendono a scartare.

L'estrazione terminologica mostra, inoltre, che il dominio non presenta un livello uniforme di stabilizzazione concettuale. Alcune aree risultano infatti altamente normate (come l'architettura hardware), mentre altre, più direttamente esposte alla corsa tecnologica e commerciale, appaiono più fluide e instabili, come quelle legate al software, ai videogiochi o alle interfacce. Ciò conferma quanto osservato da Selosse (2023): la stabilizzazione terminologica riflette l'emersione disciplinare e l'evoluzione cognitiva dei saperi, ma non procede allo stesso ritmo in tutti i sottodomini.
Da ciò deriva una conseguenza metodologica fondamentale: la lista dei termini, da sola, non basta. L'organizzazione semasiologica iniziale ha mostrato limiti: per distinguere, ad esempio, un abaco da un Gamma 30, non è sufficiente l'estrazione automatica, ma occorre una classificazione onomasiologica fondata su criteri fisici, logici, cronologici e storici.
Queste difficoltà non sono peculiari del corpus museale, ma riflettono problemi strutturali ben noti nella rappresentazione del dominio informatico. Come mostrato da L'Homme (2008), anche l'applicazione di TermoStat a corpora specializzati produce una lista di candidati terminologici che le risorse esistenti - dizionari online, ontologie di settore, repertori generalisti — coprono solo parzialmente e in modo disomogeneo. Nel suo studio, 75 termini estratti automaticamente sono stati confrontati con sei risorse: nessuna restituisce una copertura completa e coerente, e la risorsa paradossalmente più ricca risulta essere WordNet, un database lessicale generale. Lo studio mette inoltre in evidenza problemi ricorrenti quali polisemia diffusa, variazione tra parti del discorso, presenza di unità che compaiono solo in espressioni lunghe, regimi sintattici diversi e trattamenti eterogenei nelle risorse consultate.
Questi risultati convergono con quanto emerso nella nostra analisi: l'estrazione terminologica automatica identifica unità rilevanti, ma non è sufficiente a ricostruire un sistema terminologico stabile o coerente. Da ciò deriva la necessità di integrare i risultati con un approccio onomasiologico fondato su criteri concettuali e sulla modellizzazione formale del dominio.

Il sistema terminologico non può essere ricostruito partendo dalle parole, ma dalle strutture concettuali che le parole contribuiscono a designare. La questione emerge chiaramente nel confronto tra oggetti molto distanti — come un abaco e un calcolatore elettromeccanico — o tra oggetti molto vicini ma nominati diversamente — come home computer e PC. Per includerli nello stesso spazio descrittivo, è necessario definire preliminarmente le caratteristiche che li distinguono e li connettono (funzione, principio di operazione, grado di automatismo, periodo storico, materiali, contesto d'uso), e solo successivamente associare le designazioni linguistiche.
In altre parole, la lingua non costituisce né un punto di partenza né un punto di arrivo statico, ma un elemento che si ridefinisce progressivamente all'interno di un processo iterativo tra osservazione dei dati, ricostruzione concettuale e formalizzazione.
Per costruire una risorsa coerente è necessario passare da un approccio semasiologico a un approccio onomasiologico.
Questo passaggio segna l'inizio della fase successiva: la costruzione della rete concettuale, che funge da livello intermedio tra corpus e ontologia, consentendo di:

- identificare i concetti del dominio,
- organizzare classi e sottoclassi,
- esplicitare relazioni tassonomiche e meronimiche,
- preparare la strutturazione ontologica in TEDI.

L'estrazione manuale ha avuto un ruolo fondamentale nella selezione delle designazioni più significative: gli esperti del dominio (curatori, storici dell'informatica, e membri dell'équipe museale) hanno contribuito a identificare i termini rilevanti, a chiarire ambiguità, a distinguere varianti nominali locali e a riconoscere nomi storici o vernacolari non immediatamente rilevabili tramite strumenti computazionali. Questa fase è stata indispensabile per verificare che le occorrenze individuate dal corpus corrispondessero effettivamente a concetti rilevanti nel dominio patrimoniale, evitando una dispersione terminologica.

4.2.3 Costruzione della rete semantica

L'analisi terminologica, condotta in una prospettiva inizialmente semasiologica (incentrata sulle designazioni e sulle loro variazioni d'uso), non è sufficiente da sola a ricostruire l'organizzazione concettuale del dominio. Per passare da un elenco di termini a un sistema di conoscenze strutturato è necessario individuare i concetti, le loro proprietà distinte e le relazioni che li collegano.

Questo passaggio costituisce uno strato intermedio di modellizzazione, in cui il dominio viene organizzato in concetti, relazioni e campi semantici, senza introdurre ancora le strutture formali proprie di un'ontologia, che saranno definite nella fase successiva.

La costruzione della rete semantica e della relativa ontologia risponde precisamente a questa esigenza, consentendo di esplicitare le relazioni iperonimiche (classe/sottoclasse) e meronimiche (parte/tutto) che definiscono l'evoluzione del patrimonio informatico e delle sue rappresentazioni discorsive, intese come le modalità attraverso cui le tecnologie e i concetti informatici sono stati descritti, narrati e concettualizzati nel tempo nei testi, nei musei e nelle comunità digitali.

Se il corpus permette di osservare la variabilità delle designazioni, la formalizzazione ontologica stabilizza tale struttura definendo i concetti come entità formali (livello meta-linguistico), tipizzando le relazioni e ancorandole a un modello concettuale condiviso.

L'ontologia risultante costituisce l'esito della formalizzazione e non la base della rete semantica. In tale fase, il modello CIDOC-CRM (cfr. cap. 2, p. 56) è adottato per quei concetti pertinenti alla descrizione patrimoniale degli oggetti informatici, grazie alla sua capacità di rappresentare oggetti, eventi, agenti e relazioni spazio-temporali (ad es. *E52 Time-Span*), garantendo coerenza nella rappresentazione di collezioni complesse. Questa struttura costituisce la base dell'ontologia, successivamente raffinata e ampliata in TEDI.

Infine, occorre distinguere tale processo di strutturazione concettuale dalla sua successiva implementazione informatica. La creazione di un Knowledge Graph su piattaforma Wikibase e la rappresentazione dei dati secondo il paradigma del Web Semantico (RDF) e i principi FAIR rappresentano l'esito tecnico del lavoro. Questa fase di armonizzazione informatica garantisce l'effettiva circolazione e l'interoperabilità dei dati tra diversi sistemi museali e informativi, separando nettamente la gestione digitale del patrimonio dalla sua analisi terminologica e concettuale.

A partire dal lessico estratto, si è proceduto in primo luogo a organizzare i termini in campi semantici, distinguendo insiemi relativi a macchine, teorie, produttori, acronimi, componenti ecc. Questa fase ha consentito di individuare quindi relazioni concettuali di tipo iperonimico, meronimico, sinonimico e coiponimico, oltre a rendere osservabili dinamiche di ridenominazione e riorganizzazione categoriale, come nel passaggio da *home computer* a *personal computer* fino a *laptop*.

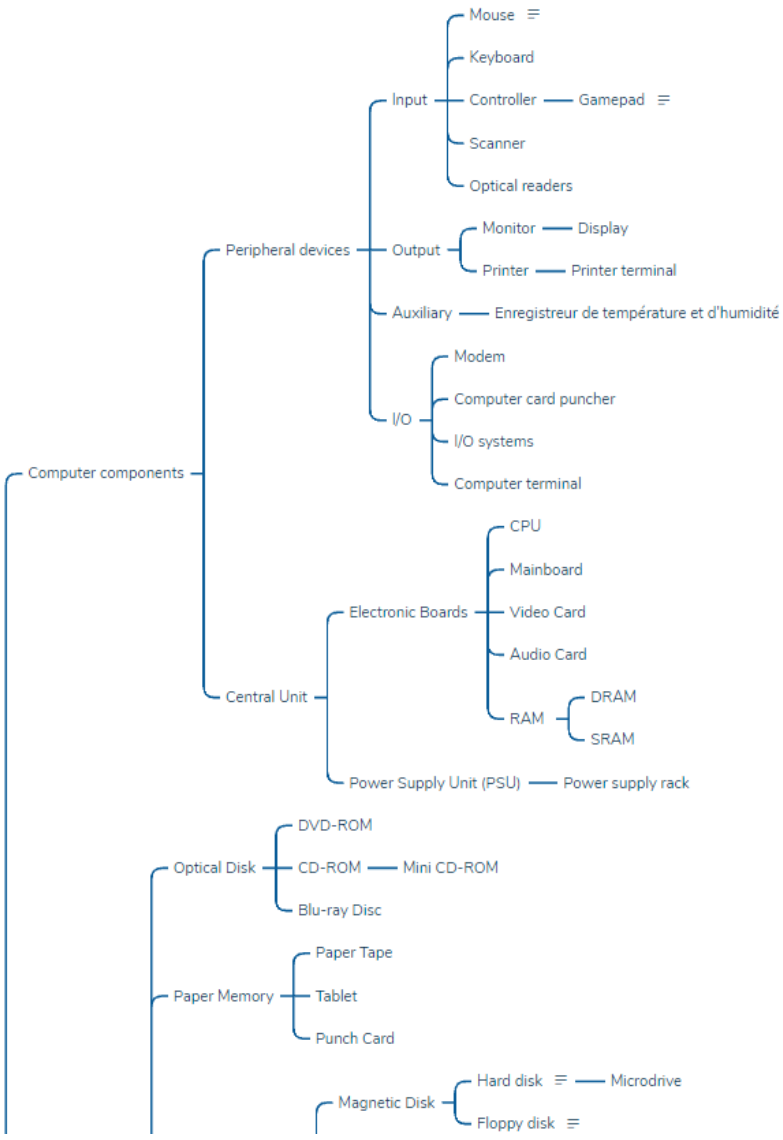
Ad esempio, per il sottodominio relativo ai transistor e ai sistemi batch (seconda generazione di computer), la rete semantica ha permesso di identificare:

- **Dominio:** *Transistors and batch systems, second generation computer*
- **Sottodominio:** *Mainframes*, *Macro-computer*
- **Campi semantici (esempi):** circuiti elettronici (diodi, resistori, transistor), schede logiche, flip-flop, memorie toroidali, circuiti di termoregolazione, alimentazione, periferiche pesanti, lettori e perforatori di schede, unità a nastro magnetico, stampanti a 132 colonne, sistemi elettromeccanografici, calcolatrici elettroniche, unità centrale di elaborazione, armadi dei moduli di collegamento, console di controllo danni.

In una fase preliminare è stato utilizzato XMind[38], un software per la creazione di mappe mentali e diagrammi gerarchici, al fine di impostare una prima articolazione concettuale del dominio. Questo strumento ha consentito di organizzare visivamente una struttura iniziale di classi e sottoclassi, individuando domini principali (ad es. Informatica, Computer, Teorie, Produttori) e sottodomini correlati (ad es. Periodizzazioni, Inventori, Periferiche). Tale rappresentazione ha fornito un supporto utile per delineare una gerarchia preliminare, da affinare successivamente nella rete semantica e nella formalizzazione ontologica.

[38] XMind (versione 2024), disponibile su: https://xmind.app

IT heritage


Computer components
Peripheral devices
Input
Mouse
Keyboard
Controller
Gamepad
Scanner
Optical readers
Output
Monitor
Display
Printer
Printer terminal
Auxiliary
Enregistreur de température et d'humidité
I/O
Modem
Computer card puncher
I/O systems
Computer terminal
Central Unit
Electronic Boards
CPU
Mainboard
Video Card
Audio Card
RAM
DRAM
SRAM
Power Supply Unit (PSU)
Power supply rack
Optical Disk
DVD-ROM
CD-ROM
Mini CD-ROM
Blu-ray Disc
Paper Memory
Paper Tape
Tablet
Punch Card
Magnetic Disk
Hard disk
Microdrive
Floppy disk

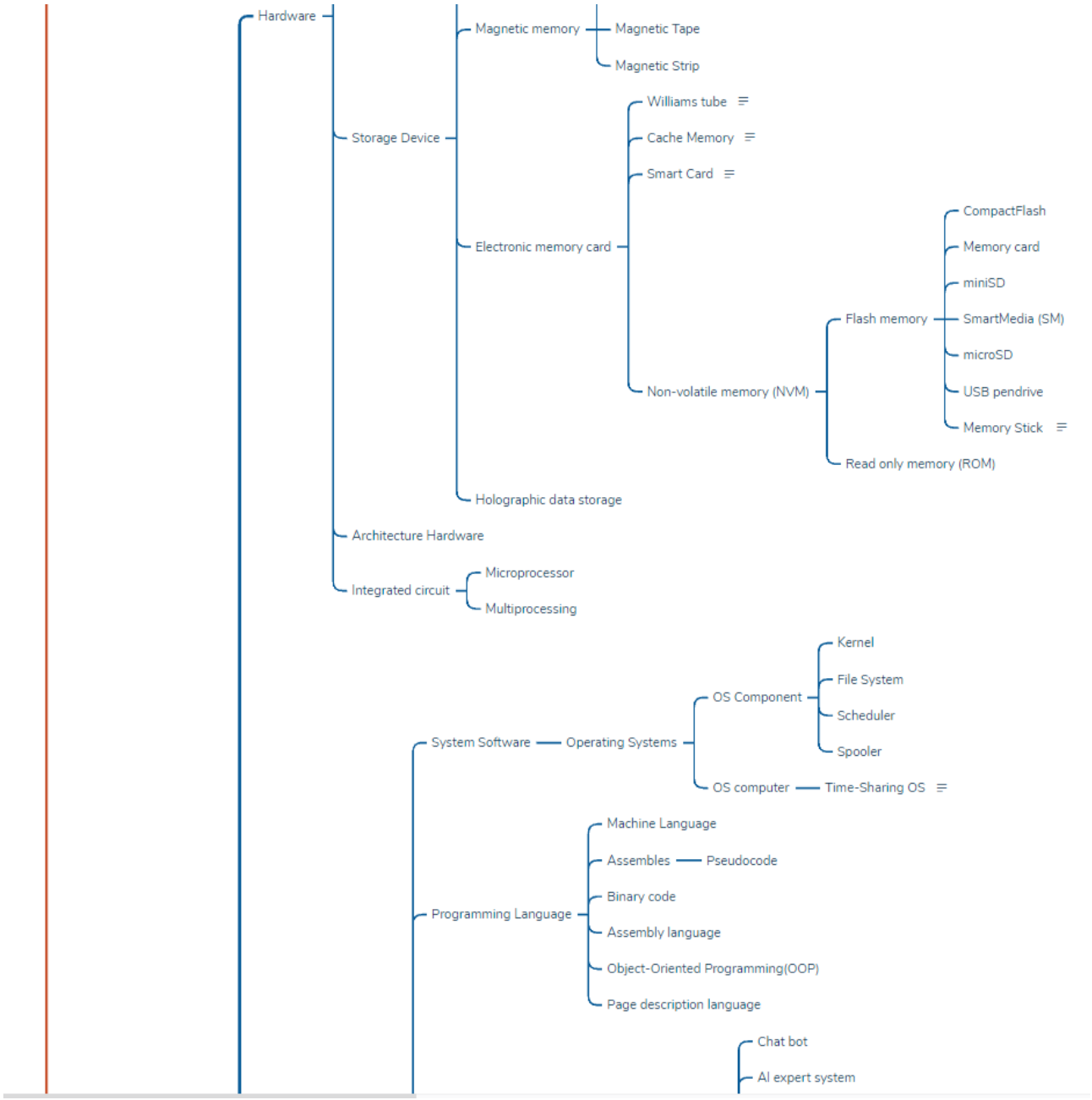


Figure 112 Porzione della mappa concettuale preliminare elaborata in XMind, rappresentativa dell'articolazione iniziale di classi e sottodomini.

A seguire è riportata la mappa semantica nella sua interezza[39]. Sebbene non sia possibile leggerne nel dettaglio i singoli elementi, la visualizzazione permette di apprezzare la complessità del dominio e la presenza di numerosi termini ancora privi di istanze, evidenziando la fase di astrazione preliminare necessaria alla successiva modellizzazione ontologica.

IT heritage.pdf

39

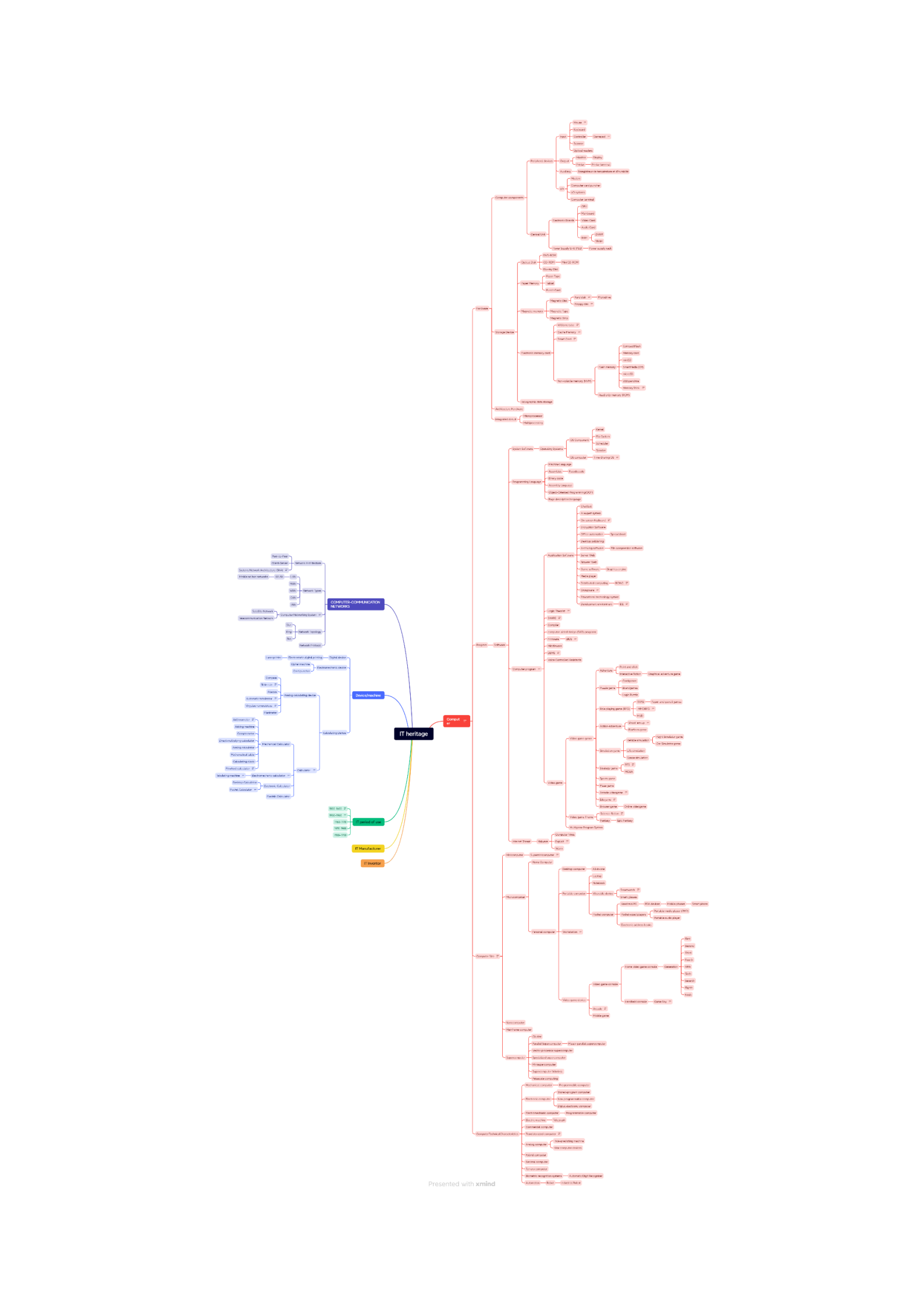

IT heritage
COMPUTER-COMMUNICATION NETWORKS
Device/machine
IT period of use
IT Manufacturer
IT inventor
Comput er
Presented with xmind

Figure 113 Mappa concettuale completa del dominio IT elaborata in XMind

La rappresentazione grafica si è rivelata uno strumento fondamentale nelle fasi iniziali del lavoro, poiché ha permesso di evitare dispersioni, evidenziare sovrapposizioni concettuali e rendere esplicite le relazioni tra i vari livelli della rete semantica. Il modello è stato oggetto di numerose modifiche e rivisitazioni nel tempo, soprattutto a seguito di consultazioni con esperti del settore, curatori museali e studiosi di storia dell'informatica. Tali revisioni sono state condotte all'interno del lavoro di équipe del progetto e recepite nel presente studio attraverso i materiali e le versioni consolidate messi a disposizione dal team che hanno contribuito a consolidarne la coerenza terminologica e concettuale chiarendo ambiguità, valutare ipotesi di sinonimia e riconoscere denominazioni locali o vernacolari considerate rilevanti nel contesto patrimoniale.

In questa fase, il lavoro si colloca ancora in una prospettiva semasiologica, orientata all'analisi e all'organizzazione delle designazioni e dei loro usi, senza implicare una costruzione concettuale ontologica validata da esperti del dominio.

In una fase successiva, la struttura semantica esplorativa è stata trasferita in tabelle strutturate (Excel), che hanno consentito di organizzare in modo sistematico le informazioni relative alle linee temporali di origine ed evoluzione dei concetti, ai contesti geografici di sviluppo e diffusione, agli inventori e produttori associati, alle caratteristiche essenziali degli oggetti e alle relazioni tra designazioni linguistiche e artefatti materiali.

Questa organizzazione ha segnato la transizione dalla prospettiva semasiologica alla onomasiologica, in cui le designazioni vengono riorganizzate a partire dai concetti, dando forma a una rappresentazione coerente e sistematica del dominio.

In questa fase, il lavoro ha assunto una dimensione sistematica, attraverso una prima organizzazione tabellare delle denominazioni emerse dall'analisi terminologica. I termini sono stati distribuiti lungo assi cronologici e geografici, e messi in relazione attraverso campi come *Generic relation, Associated term* e *Time Span*, che hanno permesso di descrivere in modo coerente la genealogia dei dispositivi di calcolo e dei concetti informatici.

La rete semantica è stata articolata in più fogli di lavoro tematici (*Hardware, Software, IT heritage artefact, Machines ontology, Videogame groupBy_Company*) che riflettono i diversi ambiti del patrimonio informatico e del calcolo. Tale organizzazione, tuttavia, non ha costituito una soluzione definitiva: la strutturazione diacronica proposta in questa fase è stata successivamente discussa all'interno del progetto e valutata come non pienamente funzionale dagli esperti coinvolti.

Questa fase ha pertanto assunto il ruolo di livello intermedio di lavoro, utile a mettere in evidenza criticità e limiti dell'organizzazione semantica iniziale e a orientare progressivamente il lavoro terminologico verso soluzioni più adeguate.

Poiché il lavoro si colloca ancora in un contesto semantico e descrittivo, l'attenzione si è concentrata sui termini generici, sulle loro relazioni e sugli ambiti d'uso, senza introdurre ancora una formalizzazione concettuale in senso ontologico. Solo nel successivo passaggio all'ontoterminologia (sviluppata in TEDI), tale materiale descrittivo è stato rielaborato per una formalizzazione rigorosa. In questa sede, le distinzioni emerse dall'analisi dei termini sono state tradotte in caratteristiche essenziali dei concetti, permettendo di definire le entità come classi e di strutturare le relazioni gerarchiche e associative in modo formale.

Nell'organizzazione dei dati terminologici, l'analisi del dominio è stata condotta dal team secondo un'impostazione multidimensionale, fondata sull'utilizzo di più assi di analisi complementari, ciascuno dei quali descrive una specifica caratteristica dei concetti e delle designazioni.

Tra questi, l'asse tecnico rappresenta uno dei criteri utilizzati per distinguere gli artefatti in base alla tecnologia di funzionamento, consentendo di differenziare, ad esempio, tra dispositivi meccanici, elettromeccanici ed elettronici. Tale distinzione risulta essenziale per una corretta classificazione semantica: l'inserimento di oggetti come l'abaco o il Gamma 30 richiede infatti la comprensione delle differenze tecnologiche e storiche di base che ne definiscono l'appartenenza a specifiche tipologie e generazioni di strumenti di calcolo.

L'asse tecnico opera, tuttavia, in relazione con altri assi di analisi, tra cui l'asse funzionale (*function of calculation*), che descrive le modalità di esecuzione del calcolo, e assi di natura teorica e programmativa, utilizzati per distinguere, ad esempio, tra dispositivi analogici e digitali, o tra macchine programmabili e non programmabili. L'integrazione di questi diversi assi consente di articolare il dominio in modo più fine, evitando classificazioni unidimensionali e rendendo esplicite le molteplici proprietà rilevanti degli artefatti.

Questo lavoro di definizione e combinazione degli assi è stato svolto in una prospettiva prevalentemente linguistica e terminologica, finalizzata a organizzare le designazioni e a chiarire le relazioni semantiche tra i concetti, senza introdurre una formalizzazione ontologica.

4.2.4 Rappresentazione spaziale e documentazione del patrimonio: la timeline georeferenziata e l'esperienza SIGEC-Web

Parallelamente alla costruzione della rete semantica, è stata realizzata una mappa cronotopica del patrimonio informatico tramite TimeMapper[40] e integrata all'interno della piattaforma ITinHeritage. TimeMapper è uno strumento open-source sviluppato dall'Open Knowledge Foundation Labs, progettato per unire in un'unica interfaccia timeline e mappa geografica, integrando elementi testuali, visivi e spaziali. Il workflow prevede tre passaggi principali:

1. Creazione di un foglio di calcolo (Google Sheets) contenente eventi, luoghi, coordinate geografiche, immagini, descrizioni e metadati;
2. Connessione e personalizzazione, attraverso la quale TimeMapper genera automaticamente la timeline e la mappa interattiva;
3. Pubblicazione ed embedding, che consente di condividere il risultato tramite un URL dedicato o integrarlo in siti web e repository digitali.

Dal punto di vista tecnologico, come indicato nella documentazione ufficiale dello strumento, TimeMapper si basa su componenti open-source quali TimelineJS, ReclineJS, Leaflet, Backbone e Bootstrap, che ne garantiscono trasparenza, interoperabilità e replicabilità.

La mappa è stata costruita a partire da un foglio Google strutturato manualmente, in cui ogni riga rappresenta un concetto, un artefatto o un evento rilevante nella storia dell'informatica. Il dataset include i campi richiesti da TimeMapper:

- Title (es. Antikythera mechanism, Greek logic, Automata of Hero);
- Start (datazione, anche negativa per gli anni a.C.);
- Description;
- Image (URL e licenza);
- lat / lon (coordinate geografiche);
- Place (località di produzione o conservazione);
- Source e Source URL (provenienza dell'immagine o della descrizione).

[40] Disponibile su: https://timemapper.okfnlabs.org/

Title	Start	Description	lat	lon	Place	Image	Source	Source URL
Clay balls and tokens	-7500	Saint Peter Damian, O.S.B. (Petrus Damiani, also Pietro Damiani or Pier Damiani; c. 1007 – February 21/22, 1072) was a reforming monk in the circle of Pope Gregory VII and a cardinal. In 1823, he was declared a Doctor of	33.12990841389535	43.29759554701045	Uruk, Iraq (today)	https://upload.wikimedia.org/wikipedia/commons/d/da/Unknown%2C_gaming_pieces_%28FindID_457786%29.jpg	Oxfordshire County Council, Anni Byard	Creative Commons Attribution-Share Alike 2.0
Greek logic	-330	Anselm of Canterbury (c. 1033 – 21 April 1109), also called of Aosta for his birthplace, and of Bec for his	38.9953683	21.9877132	Greece		Lysippos	Public domain
Greek algorithm	-300	Guillaume de Champeaux (c. 1070 – 18 January 1121 in Châlons-en-Champagne), known in English as William of Champeaux and Latinised to Gulielmus de Campellis , was a French philosopher and theologian.	38.9953683	21.9877132	Alexandria, Egypt	https://upload.wikimedia.org/wikipedia/commons/7/7a/Euclid%27s_algorithm_Book_VII_Proposition_2_1.png	Wvbailey	Public domain
Counting boards	-300	Peter Abelard (/ˈæb.əˌlɑːrd/; Latin: Petrus Abaelardus or Abailardus; French: Pierre Abélard, pronounced: [a.beˈlaːʁ]; 1079 – 21 April 1142) was a medieval French scholastic philosopher, theologian and preeminent logician. The story of his affair with and love for Heloise has become	39.64371316287571	21.759789218259023	Salamis		Jacobolus	Creative Commons Attribution-Share Alike 4.0
Antikythera mechanism	-100	Robert Grosseteste (/ˈgroʊstɛst/ grohs-test) or Grossetete (/ˈgroʊsteɪt/ grohs-tayt; c. 1175 – 9 October 1253) was an English statesman, scholastic philosopher, theologian, scientist and Bishop of Lincoln. He was born of humble parents at Stradbroke in Suffolk. A C	39.64371316287571	21.759789218259023	Antikythera	https://upload.wikimedia.org/wikipedia/commons/5/5c/Antikythera_Mechanissem_w.jpg	Marsyas	Creative Commons Attribution 2.5
Automata of Hero of Alexandria	100	Albertus Magnus, O.P. (1193/1206 – November 15, 1280), also known as Albert the Great and Albert of Cologne, is a Catholic saint. He was a German Dominican friar and a bishop who achieved fame for his comprehensive knowledge of and advocacy for the peaceful coexistence of	26.429631872192395	29.264976765891028	Alexandria, Egypt	https://upload.wikimedia.org/wikipedia/commons/1/1b/Hero%27s_mobile_automaton%2C_1st_century_AD%2C_Alexandria_%28model%29.jpg	Gts-tg	Creative Commons Attribution-Share Alike 4.0 International

Figure 114 Estratto del dataset (Google Sheets) impiegato per alimentare TimeMapper: ogni riga rappresenta un concetto o artefatto con titolo, datazione, descrizione, coordinate, luogo, immagine e riferimenti di provenienza.

Il risultato è una timeline georeferenziata della storia dell'informatica. La traduzione della rete semantica in una rappresentazione dinamica ha consentito di associare ogni concetto a coordinate spazio-temporali, a una documentazione visiva e a insiemi di metadati relativi alla provenienza, alla produzione e alla conservazione degli artefatti. Le attività preliminari hanno incluso la raccolta e la verifica delle risorse iconografiche e descrittive, con particolare attenzione agli aspetti di licenza, e la geolocalizzazione dei manufatti e dei concetti mediante l'integrazione di coordinate validate.

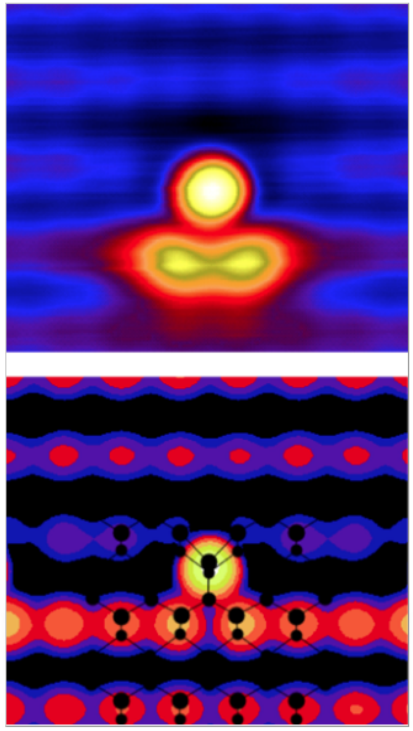
Una volta sincronizzato il foglio con TimeMapper, il sistema ha generato una rappresentazione integrata "timeline + mappa" che consente di:

• visualizzare l'evoluzione temporale dei concetti informatici, dai calcoli proto-cuneiformi agli algoritmi dell'antichità e alle prime macchine meccaniche;

• riconoscere la diffusione geografica dei principali poli di innovazione, come Uruk, Alessandria, Salamina, Anticitera o Cusco;

• navigare tra immagini, descrizioni e metadati, facilitando un'interpretazione storica, visuale e geolocalizzata degli artefatti.

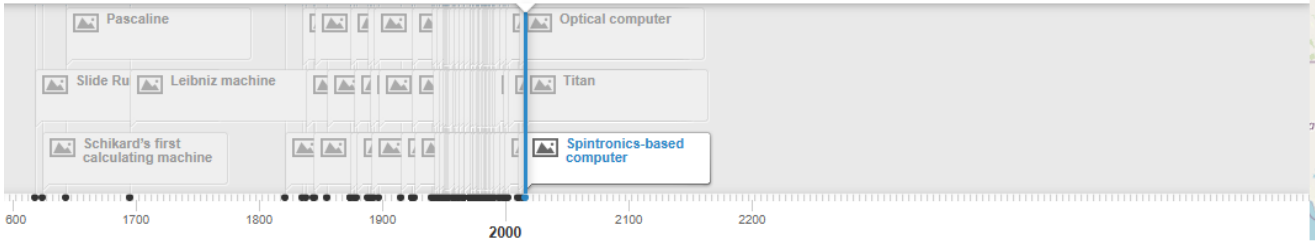
Figure 115 Vista della timeline georeferenziata generata con TimeMapper, che integra immagini, descrizioni e geolocalizzazione dei concetti e degli artefatti della storia dell'informatica. A sinistra la visualizzazione temporale (timeline), a destra la mappa interattiva con i punti di provenienza e conservazione. L'esempio mostrato in figura si riferisce al nodo Greek logic (330 a.C.).



2015

Spintronics-based computer

Source: National Institute of Standards and Technology



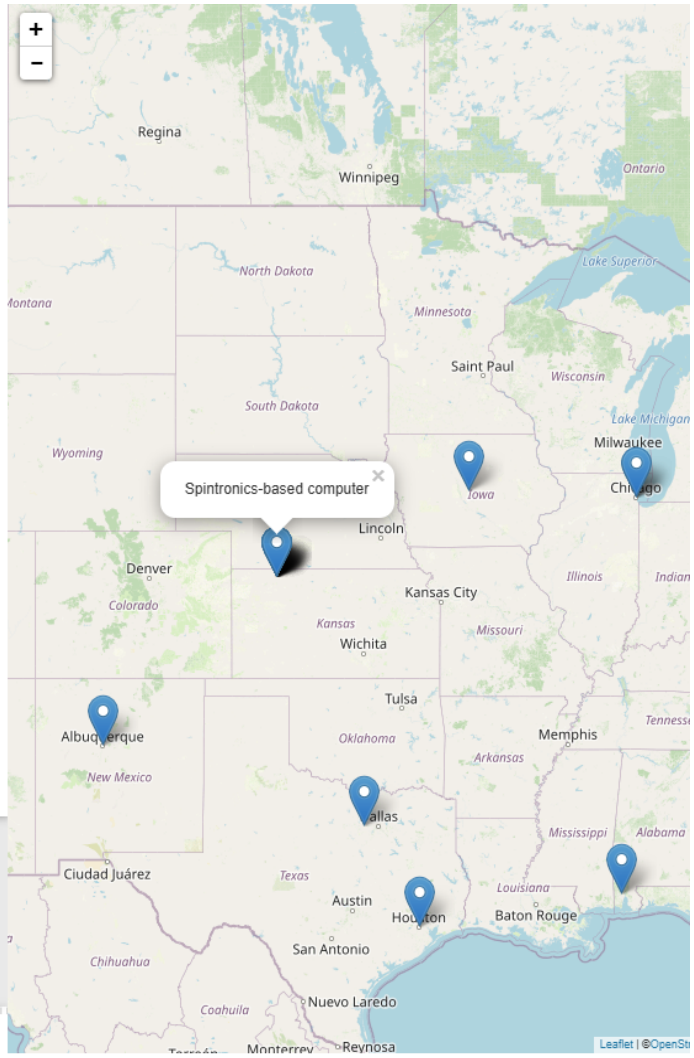


Figure 116 L'esempio mostrato in figura si riferisce al nodo Spintronics-based computer (2015), rappresentativo delle tecnologie computazionali più recenti.

La timeline permette di osservare in modo immediato la profondità storica del patrimonio informatico, visualizzando la progressione dei concetti e degli artefatti dalle prime forme di registrazione numerica fino alle tecnologie computazionali più avanzate.

Nelle prime sezioni della timeline compaiono infatti i *Clay balls and tokens* di Uruk (ca. 7500 a.C.), considerati forme protostoriche di contabilità e rappresentazione numerica, seguiti dai contributi dell'antichità greca come *Greek logic* e *Greek algorithmics*, che costituiscono la base teorica del pensiero computazionale occidentale.

Procedendo lungo la rappresentazione cronologica, la timeline mostra l'emergere dei primi strumenti meccanici (abachi, macchine di Anticitera, automata ellenistici) e successivamente delle macchine matematiche rinascimentali e moderne (*Pascaline, Leibniz, Schickard*). Avanzando verso l'età contemporanea si osserva un aumento significativo di eventi, invenzioni e tecnologie, fino a raggiungere le innovazioni del XXI secolo, come i prototipi di calcolo basati sulla spintronica (*Spintronics-based computer*, 2015).

L'intero arco temporale evidenzia quindi non solo la continuità storica del pensiero computazionale, ma anche i passaggi di trasformazione, accelerazione e diversificazione tecnologica che caratterizzano l'evoluzione dell'informatica.

Figure 117 Estratto della mappa georeferenziata generata con TimeMapper, che visualizza la distribuzione mondiale dei principali concetti, luoghi e artefatti storici rilevanti per il dominio dell'informatica

Questa mappa costituisce un primo livello di organizzazione visuale di alcuni artefatti del dominio, funzionale alla successiva formalizzazione. La visualizzazione risultante permette di collegare la storia delle tecnologie computazionali ai luoghi e ai tempi della loro diffusione, rendendo immediatamente percepibile l'evoluzione delle denominazioni, dei dispositivi e delle pratiche d'uso. La rete semantica e la timeline operano pertanto su piani distinti e non sovrapponibili. La prima è finalizzata all'analisi e all'organizzazione delle designazioni e delle loro relazioni all'interno del dominio, mentre la seconda risponde a obiettivi di mediazione e divulgazione scientifica, occupandosi della rappresentazione di istanze, eventi e traiettorie storiche nello spazio e nel tempo.

In questa prospettiva, il trattamento delle istanze e delle rappresentazioni visive adottato nel progetto ITinHeritage si inserisce in continuità con precedenti attività di catalogazione museale svolte durante il triennio del dottorato presso il Museo degli Strumenti per il Calcolo dell'Università di Pisa. In tale contesto, le macchine di calcolo storiche sono state catalogate secondo gli standard ICCD, attraverso la piattaforma nazionale SIGEC-Web – Sistema Informativo Generale del Catalogo, sviluppata e gestita dall'Istituto Centrale per il Catalogo e la Documentazione (ICCD) del Ministero della Cultura[41].
La catalogazione ha comportato la redazione di schede descrittive strutturate, comprensive di dati tecnici, cronologici, autoriali, localizzativi e iconografici, secondo una distinzione rigorosa tra il livello dell'oggetto materiale individuale e quello delle informazioni descrittive e documentarie a esso associate.

L'esperienza maturata in ambito catalografico ha influito direttamente sulle scelte metodologiche adottate nel progetto ITinHeritage, in particolare per quanto riguarda l'attenzione alla tracciabilità delle fonti, alla distinzione tra concetti generali e oggetti materiali individuali, e all'uso delle immagini come rappresentazioni visive documentarie degli artefatti, corredate da indicazioni di provenienza e di licenza. Questi principi trovano una traduzione operativa nell'uso di TEDI (*ontoTerminology Editor*), l'ambiente software che vedremo nel dettaglio nei paragrafi successivi e che costituisce il supporto tecnologico per la costruzione della risorsa ontoterminologica.
In TEDI, le istanze associate ai concetti riprendono infatti una logica descrittiva analoga a quella delle schede museali, pur collocandosi all'interno di un quadro terminologico e ontoterminologico formale, distinto sia dalla semplice documentazione iconografica sia dalla mediazione divulgativa.

[41] https://www.iccd.beniculturali.it/it/sigec-web

Immagine	Oggetto	Localizzazione	Tipo scheda	Codice univoco	Stato
	compasso	Pisa (PI), Via Bonanno Pisano, 2/B	PST 4.00	09 01411549	Inviata in verifica scientifica
	personal computer, IBM PS/2	Pisa (PI), Via Bonanno Pisano, 2/B	PST 4.00	09 01411525	Inviata in verifica scientifica
	calcolatrice, Totalia 2361	Pisa (PI), Via Bonanno Pisano, 2/B	PST 4.00	09 01411558	Inviata in verifica scientifica
	terminale, Texas Instruments Silent 700 ASR	Pisa (PI), Via Bonanno Pisano, 2/B	PST 4.00	09 01411539	Inviata in verifica scientifica
	calcolatrice meccanica, Lagomarsino Numeria mod. 9200	Pisa (PI), Via Bonanno Pisano, 2/B	PST 4.00	09 01411552	Inviata in verifica scientifica
	computer palmare (PDA), Apple Newton MessagePad 130	Pisa (PI), Via Bonanno Pisano, 2/B	PST 4.00	09 01411550	Inviata in verifica scientifica
	calcolatrice elettromeccanica, Diehl	Pisa (PI), Via Bonanno Pisano, 2/B	PST 4.00	09 01411554	Inviata in verifica scientifica
	calcolatrice elettronica, Precisa 164-12	Pisa (PI), Via Bonanno Pisano, 2/B	PST 4.00	09 01411565	Inviata in verifica scientifica

Figure 118 Esempio di scheda di catalogazione ICCD (PST 4.00) compilata in SIGEC-Web per una macchina di calcolo del Museo degli Strumenti per il Calcolo (Università di Pisa).

La struttura a campi evidenzia il livello istanziale e descrittivo della catalogazione museale (identificazione, dati tecnici, cronologia, localizzazione, documentazione)

In questo senso, il lavoro svolto in ambito ICCD e SIGEC-Web costituisce un precedente metodologico diretto per la gestione delle istanze in TEDI, pur nel passaggio da un contesto catalografico a uno terminologico e ontoterminologico formalizzato.

4.2.5 Strutturazione ontologica e formalizzazione ontoterminologica in TEDI

La formalizzazione del dominio informatico nel progetto ITinHeritage si sviluppa lungo un percorso continuo che integra modellizzazione ontologica, riflessione teorica e formalizzazione terminologica. In questa fase è pertanto necessario distinguere i diversi livelli di intervento e i relativi strumenti.

Protégé costituisce l'ambiente di modellizzazione ontologica globale adottato dal progetto per la definizione delle classi, delle gerarchie concettuali, delle proprietà e dei vincoli formali che strutturano logicamente il dominio, secondo un approccio conforme agli standard OWL. Questa attività è stata sviluppata a livello di team e riguarda la costruzione e il mantenimento dell'ontologia complessiva del patrimonio informatico.

In parallelo, la formalizzazione ontoterminologica è stata sviluppata in ambiente TEDI, a partire dai risultati dell'analisi linguistica e terminologica. In questo contesto, il contributo qui presentato si concentra sulla definizione e sull'organizzazione di campi descrittivi, assi di analisi e criteri terminologici, conformi ai principi metodologici degli standard ISO 704 e ISO 1087, nonché sull'inserimento e sulla strutturazione delle designazioni e delle informazioni concettuali relative a una porzione dell'ontologia.

TEDI si configura così come uno spazio di raccordo tra l'analisi terminologica e la modellizzazione concettuale, consentendo di formalizzare i concetti a partire dai dati linguistici e dalle pratiche descrittive museali, senza sovrapporsi alle attività di progettazione ontologica globale svolte in Protégé.

TEDI (*ontoTerminology EDItor*), sviluppato da Roche e dal gruppo Condillac, consente invece di costruire l'apparato concettuale e terminologico. TEDI si fonda su una teoria del concetto basata sullo schema aristotelico *genus* + *differentiae*, in cui un concetto è definito come combinazione unica di caratteristiche essenziali e un termine ne rappresenta la designazione linguistica (Desprès *et al.*, 2020). TEDI permette quindi di precisare i concetti attraverso le loro caratteristiche distintive e di collegarli alle designazioni multilingui effettivamente attestate, garantendo coerenza tra piano concettuale e piano linguistico.

Come evidenziato da Desprès, Roche e Papadopoulou nell'*Étude comparative de deux méthodes de modélisation*, i due strumenti adottano prospettive diverse ma perfettamente complementari. Protégé si colloca nel quadro della logica descrittiva e prende avvio dalla definizione di classi, attributi e relazioni ontologiche, con l'obiettivo di garantire interoperabilità, coerenza formale e possibilità d'inferenza all'interno del modello OWL. Questa impostazione presuppone competenze di ingegneria della conoscenza e riflette un approccio orientato alla struttura formale del dominio: come osservano

gli autori, «l'uso di un simile ambiente rimane difficile senza l'aiuto di un ingegnere della conoscenza»[42].

TEDI, al contrario, è uno strumento ontoterminologico che parte dai termini e dai concetti esperti, modellizzati attraverso la selezione e l'organizzazione delle caratteristiche essenziali. Il suo obiettivo non è la costruzione né la gestione di un knowledge graph globale, ma la creazione di risorse ontoterminologiche e dizionari specializzati relativi a porzioni circoscritte del dominio, mantenendo al contempo la coerenza con gli standard ISO e consentendo agli esperti del dominio di costruire ontoterminologie multilingui (caratteristiche essenziali, criteri di distinzione, definizioni) senza dover padroneggiare la logica descrittiva né i formalismi propri dell'ingegneria della conoscenza. TEDI integra, inoltre, un motore di inferenza che verifica in tempo reale la compatibilità logica delle caratteristiche selezionate, suggerisce i possibili concetti iperonimi e identifica automaticamente sinonimi terminologici, iponimi ed equivalenti multilingui.

Uno degli obiettivi espliciti di TEDI è proprio quello di «ridurre il divario che può esistere tra gli esperti del dominio e questo tipo di strumenti»[43], e ciò perché la sua teoria del concetto è «vicina alla comprensione che ne hanno gli esperti», offrendo «un'impostazione che permette all'esperto di rimanere autonomo»[44]. Per questo TEDI risulta più accessibile agli esperti del dominio – storici dell'informatica, curatori museali, archivisti, linguisti – che possono contribuire alla modellazione senza dover padroneggiare la complessità tecnica degli strumenti ontologici formali. Questa idea di *divario* (gap) fra strumenti formali del Web Semantico e pratiche concettuali degli esperti è al centro anche della riflessione di Roche e Papadopoulou in *Mind the Gap* (2019), che evidenzia come la distanza tra logica descrittiva e modellazione concettuale umanistica richieda strumenti intermedi pensati per gli specialisti.

Nel progetto ITinHeritage, la complementarità tra Protégé e TEDI risponde a un'esigenza metodologica di coordinamento tra livelli diversi di modellizzazione. Protégé è utilizzato per la costruzione e il mantenimento dell'ontologia globale del patrimonio informatico, mentre TEDI supporta lo sviluppo di risorse ontoterminologiche mirate, funzionali alla formalizzazione dei concetti a partire dai dati linguistici e dalle pratiche descrittive museali, senza sovrapporsi alla gestione del knowledge graph complessivo.

A titolo esemplificativo, mentre la modellazione ontologica in Protégé consente di rappresentare classi, attributi *(Data Properties*) e relazioni formali *(Object Properties)* – ad esempio definire

[42] "L'utilisation d'un tel environnement reste difficile sans l'aide d'un ingénieur de la connaissance." (Desprès et al., 2020, p. 45). Traduzione mia.

[43] "Tedi vise à réduire le "fossé" qui peut exister entre les experts du domaine et ce type d'outils." (Desprès et al., 2020, p. 45). Traduzione mia.

[44] "Les avantages de Tedi reposent […] sur une théorie du concept et du terme proche de la compréhension qu'en ont les experts, avantages qui permettent à l'expert d'être autonome." (Desprès et al., 2020, p. 52). Traduzione nostra.

PersonalComputer come sottoclasse di *ElectronicComputingDevice* e specificare che *haParte* una CPU o che *èProdottoDa* un certo produttore, la formalizzazione in TEDI permette di esplicitare i criteri semantici che distinguono realmente un concetto dall'altro. Attraverso TEDI è infatti possibile creare le definizioni dei concetti, specificandone le dimensioni distintive: un *personal computer* viene così definito come dispositivo autonomo, destinato all'uso individuale, caratterizzato da interazione diretta e da un certo grado di portabilità; allo stesso tempo, TEDI consente di differenziarlo da concetti affini ma non equivalenti, come *home computer* (connotazione domestica) o *laptop* (criterio della portabilità come proprietà definitoria).

La struttura dell'ontologia costruita con gli esperti e la modellizzazione ontologica in Protégé è stata sviluppata da Caroline Djambian, coordinatrice del progetto ITinHeritage. Il presente lavoro si concentra invece sulle fasi di analisi terminologica, ricostruzione concettuale e formalizzazione ontoterminologica condotta in TEDI.

La transizione dalla rete semantica alla formalizzazione richiede un quadro teorico specifico. Tale quadro è fornito dall'**approccio ontoterminologico** (Roche, 2019), che si colloca all'intersezione tra studi terminologici, modellizzazione concettuale e ingegneria della conoscenza (cfr. cap. 1, p. 46).

Esso integra:

- onomasiologia, dedicata alla costruzione della rete concettuale;
- semasiologia, necessaria nelle fasi corpus-based per l'estrazione dei fenomeni lessicali;
- terminologia cognitiva (Geeraerts, Temmerman), attenta alla categorizzazione prototipica.

All'interno di questo quadro, il lavoro ha fatto ricorso a una serie di assi di analisi intesi — secondo la terminologia dello standard ISO 704 — come criteri di suddivisione (*criteria of subdivision*). Tali assi non sono categorie teoriche astratte, ma strumenti descrittivi e operativi finalizzati all'organizzazione e alla differenziazione dei termini e degli oggetti del dominio. Essi rappresentano prospettive logiche che raggruppano caratteristiche essenziali esclusive, definendo ciò che la norma ISO 1087 identifica come la combinazione unica di proprietà che costituisce un concetto. Questi assi consentono di esplicitare proprietà rilevanti (tecnologiche, funzionali, cronologiche, teoriche) necessarie alla formalizzazione del dominio.

Questa riflessione teorica si colloca nel dibattito classico tra termini e concetti, e tra semasiologia e onomasiologia. Come ricorda Temmerman (2000), «*concepts evolve over time, as do their designations*», e tale evoluzione rende necessaria una riflessione sulle modalità di rappresentazione. Nel contesto della terminologia specialistica, l'onomasiologia risulta particolarmente adeguata, soprattutto in una prospettiva ontoterminologica, poiché la costruzione della rete concettuale precede la definizione terminologica. Tuttavia, il lavoro su corpora, specialmente nelle fasi iniziali, richiede un passaggio semasiologico per la raccolta dei dati lessicali. Come osserva Sager (1990), l'estrazione

terminologica è spesso corpus-based, anche quando l'obiettivo finale è concettuale, e questo giustifica il ruolo della semasiologia nelle prime fasi.

In un'ottica pragmatica, la terminologia concentra l'attenzione sui concetti, che esistono indipendentemente dalle denominazioni linguistiche (Cabré, 1999; L'Homme, 2004). Ne deriva un continuo passaggio tra dimensione linguistica (sinonimia, polisemia, co-occorrenze) e dimensione concettuale (relazioni tassonomiche, meronimiche, temporali).

Tuttavia, in una prospettiva ontoterminologica, è l'**onomasiologia** a garantire la stabilità concettuale lungo il processo di standardizzazione, come mostrano anche gli studi di Selosse (2023) sulla maturazione terminologica delle discipline scientifiche. Selosse mette in luce come la stabilizzazione terminologica sia segno della costituzione di una disciplina: dal descrittivo alla standardizzazione, fino all'eliminazione delle variazioni e all'aggettivazione. Questo fenomeno è evidente in sottodomini come hardware e architettura dei computer, in cui la precisione terminologica accompagna l'affinamento delle conoscenze.

In questo quadro, un aspetto centrale del lavoro consiste nel mantenere coordinati, ma distinti, il livello concettuale e quello linguistico. Le varianti terminologiche attestate nei corpora e nei cataloghi museali non vengono trattate come concetti separati, ma ricondotte alla stessa entità concettuale quando condividono la medesima struttura di caratteristiche essenziali. Ciò evita che sinonimie storiche, allonimi o localismi museali generino ambiguità concettuali e permette agli esperti del dominio di intervenire direttamente nella definizione dei concetti senza dover delegare tali scelte a specialisti della modellizzazione formale.

È in questa fase che interviene TEDI, utilizzato per:

- integrare e controllare gli assi di analisi (es. asse tecnico, asse tipologico, asse temporale) e le caratteristiche essenziali dei concetti (caratteri distintivi e non accidentali);
- collegare ogni concetto alle designazioni multilingue, specificando varianti, stati (preferito/alternativo/obsoleto) e contesti d'uso;
- generare definizioni in linguaggio naturale coerenti e verificabili a partire dalla struttura concettuale.
- Gestire le designazioni linguistiche (termini preferiti, alternativi, allonimi, equivalenti multilingue);

L'attività su TEDI ha richiesto un intenso lavoro di armonizzazione tra logica concettuale, terminologia e rappresentazione visiva: ogni passo è stato verificato sulla coerenza tra le categorie concettuali (ontologia) e le designazioni linguistiche (terminologia). Il processo ha permesso di

passare da una rete semantica descrittiva a una ontoterminologia formalizzata, capace di rappresentare l'evoluzione storica, tecnologica e linguistica del patrimonio informatico.

In termini metodologici, la complessità del lavoro ha confermato il valore dell'approccio onomasiologico e multiasse, che consente di mettere in relazione variabili tecnologiche, storiche e linguistiche in un'unica risorsa interoperabile.

In TEDI ogni elemento è formalizzato su tre livelli interdipendenti:

1. Concetto → inteso come unità di conoscenza creata da una combinazione unica di caratteristiche. Secondo la norma ISO 1087, esso rappresenta l'entità intellettuale definita tramite caratteristiche essenziali e relazioni (ad es. «Adding machine with cylindrical inscriber»);

2. Termine → definito come designazione verbale di un concetto generale in un ambito specialistico (ad es. *Adding machine* / *Machine à additionner* / *Macchina addizionatrice*);

3. Oggetto o nome proprio → istanza concreta del concetto (ad es. *Schickard Adding Machine*, *Pascaline*, *UNIVAC I*).

L'impiego di TEDI si articola quindi attraverso tre livelli operativi che lavorano in modo coordinato: il Concept Editor, il Term/Proper Name Editor e l'Object Editor.

Ogni entità è stata creata manualmente nel Concept Editor, quindi arricchita nel Term Editor e collegata agli oggetti concreti e ai propri nomi nel Proper Name Editor.

Queste classi, come ad esempio *Mechanical instrument and calculating machine*, *Analogue instrument and machine*, *Algebraic machine*, *Astronomical calendar*, sono state importate in TEDI e ricreate manualmente nel Concept Editor, costituendo la gerarchia di base del dominio "Calculation and computing device".

Nel **Concept Editor** il concetto *Calculation and computing device* non costituisce una categoria generica astratta, ma il nodo concettuale di livello superiore che organizza l'intero dominio del calcolo. La sua funzione non è quella di essere definito tramite caratteristiche essenziali proprie, bensì quella di fornire il quadro classificatorio entro cui gli assi di analisi (tecnico, funzionale, storico, materiale) si applicano ai concetti sottostanti. È nei livelli inferiori che la struttura *genus–differentiae* opera in modo pieno, permettendo di distinguere in modo rigoroso concetti contigui: ad esempio *Adding machine*, *Multiplying machine*, *Difference arithmetical machine* non sono varianti linguistiche, ma tipi concettuali differenziati da funzioni e meccanismi (es. trasmissione a ingranaggi, cilindro scanalato, sistema di tabulazione).

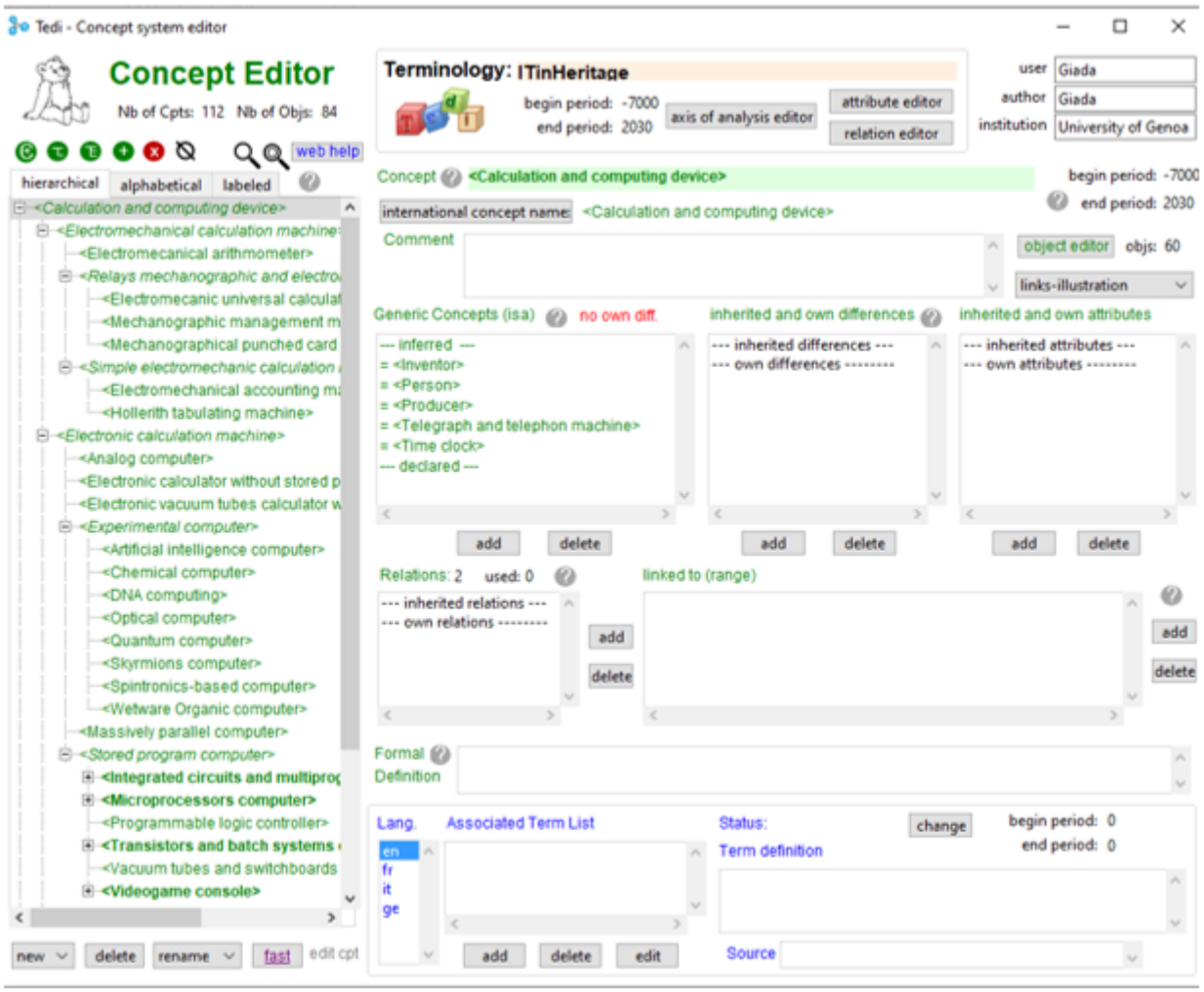


Figure 119 Concept Editor di TEDI

Un caso emblematico è dato dalle varianti storiche dell'abaco: *Abacus japonais*, *Soroban*, *Abaque japonais* sono trattati separatamente come nomi propri riferiti a oggetti specifici, ma collegati al concetto comune *Digital tablet*, evitando che la variazione denominativa produca confusione concettuale.

L'attività di ricostruzione concettuale, intesa come il processo di adattamento e trasposizione del lavoro precedentemente svolto su Protégé e con gli esperti di dominio nei formati strutturati di TEDI, è stata lunga e complessa. Ciascun concetto doveva essere collegato non solo al proprio iperonimo e ai relativi iponimi, ma anche ai caratteri essenziali (*essential characteristics*) che ne definiscono l'identità secondo la logica aristotelica.

Per ogni concetto sono stati individuati e formalizzati:

- il concetto padre/genus ("a kind of"),
- gli assi di analisi applicabili (technical, theoretical, production, generation, ecc.),
- le relazioni di ruolo (ad esempio *hasInventor*, *isInventorOf*, *hasComponent*, *isPartOf*),

- le corrispondenze terminologiche multilingue (en, fr, it),
- le eventuali istanze oggettuali (object editor).

Il livello concettuale è poi collegato, nel **Term / Proper Name Editor**, alle designazioni linguistiche reali utilizzate nei diversi contesti (documentazione museale, letteratura storica, discorso comunitario). Nel **Term Editor** le denominazioni sono organizzate secondo lo schema terminologico adottato da TEDI, che distingue tra diversi status *(preferred, alternative, tolerated, not recommended, obsolete*), consentendo di rendere esplicita la variabilità linguistica interna a una stessa lingua e la sua evoluzione nel tempo.

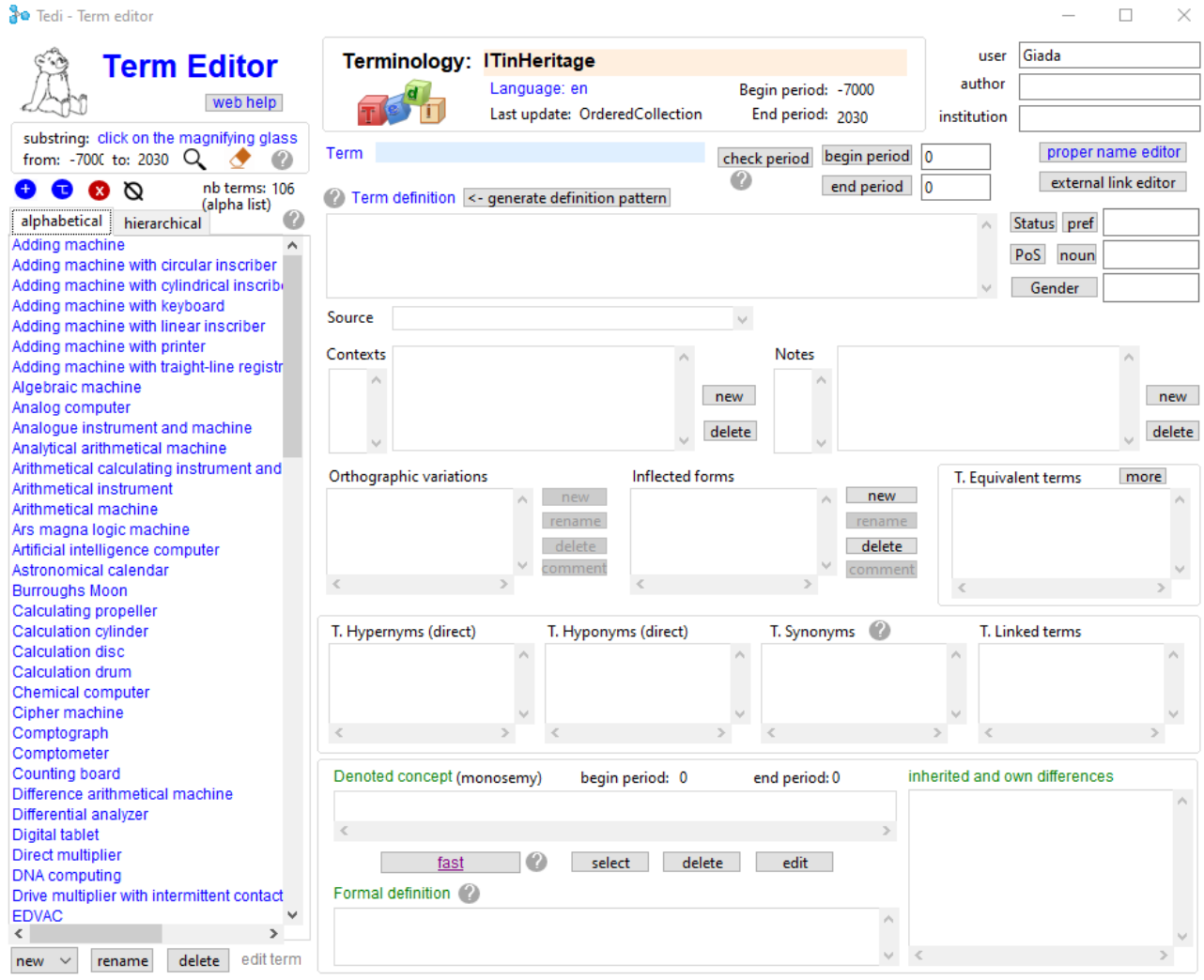


Figure 120 Term Editor di TEDI

TEDI permette inoltre di inserire termini equivalenti in lingue diverse, rendendo evidente la dimensione multilingue della terminologia. A ciascun concetto vengono quindi associati:

- preferred term: la designazione principale nella lingua di lavoro;
- alternative term: variante accettata nella stessa lingua;
- equivalent term: traduzione in un'altra lingua, quando attestata e valida;

- status terminologici che indicano il grado di preferenza o obsolescenza.

Esempio (Concept: Antikythera mechanism):

en: *Antikythera mechanism* (preferred), *Antikythera machine* (alternative)

fr: *Mécanisme d'Anticythère* (equivalent), *Anticythère analog calculator* (alternative)

it: *Meccanismo di Anticitera* (equivalent)

Figure 121 Proper Name Editor di TEDI

Questa fase ha permesso di armonizzare le tre lingue di lavoro del progetto, mantenendo però allineate le forme realmente attestate nei corpora e nelle schede museali.

La sezione **Proper Name Editor** è stata costruita per gestire le designazioni uniche di oggetti individuali, con relazioni di sinonimia e polireferenzialità interna.

TEDI distingue infatti le designazioni dei concetti dai nomi propri degli artefatti, introducendo qui la nozione di allonimo oggettuale (*terminological allonym*). Il software definisce esplicitamente:

«Proper names are terminological allonyms when they refer to the same object»[45].
L'allonimo è pertanto una variante denominativa riferita allo stesso oggetto individuale, non a un concetto. Questa funzione è cruciale nei contesti museali, in cui un oggetto può essere citato con nomi diversi a seconda delle fonti, della lingua, della tradizione curatoriale o della storia del reperimento.
Ogni voce è collegata:

- a un concetto di riferimento (es. *Astronomical calendar* per *Antikythera analog calculator*);
- a uno o più allonimi (es. *IBM ASCC* ↔ *Aiken Relay Calculator*);
- a equivalenti multilingue (es. *Macchina di Anticitera*, *Machine d'Anticythère*);
- a immagini e link semantici (*foaf:depiction*).

L'esempio dell'**Anticythere analog calculator** è emblematico:
«Proper name designating a unique historical first analog computer, discovered ca. 200 BCE in Antikythera (Greece). »
È formalizzato come istanza del concetto *Astronomical calendar*, e dotato di equivalenti linguistici coerenti nelle tre lingue di lavoro (en/fr/it), testimoniando la dimensione storica e terminologica del primo dispositivo calcolante noto.
Questa ricostruzione ha richiesto un costante passaggio tra livello concettuale e dimensione linguistica, al fine di mantenere coerenza tra la logica classificatoria e le denominazioni attestate nei corpora, nelle fonti storiche e nei cataloghi museali.

Uno dei passaggi più impegnativi ha riguardato la definizione e l'applicazione degli assi di analisi (*axes of analysis*), che permettono di specificare le caratteristiche distintive dei concetti, fornendo criteri coerenti di differenziazione interna.
Sono stati integrati più di dieci assi, tra cui:

- Technical axis → *mechanical*, *electronic*, *analogue*, *digital*;
- Theoretical axis → *analog computation*, *digital computation*;
- Architecture axis → *Von Neumann*, *Harvard*, *MIMD architecture*;
- Production axis → *industrial production*, *experimental prototype*, *no longer in production*;

[45] TEDI, Proper Name Editor

- Generations axis → con doppia funzione: concettuale (classificazione storica) e linguistica (attribuzione temporale e denominativa a modelli come *PlayStation 3*, *PlayStation 5*);
- Function of calculation → *addition and subtraction*, *multiplication and division*, *statistical tabulation*, *accounting operations*;
- Type of drive mechanism → *gear-based drive*, *relay mechanism*, *variable-tooth gear drive*, *grooved cylinder drive*;
- Technology of electronic components → *vacuum tube*, *transistor*, *integrated circuit*, *microprocessor*;
- Material or medium of representation (per gli strumenti primitivi) → *clay token*, *rope and cord*, *tablet-based notation*, *board-based counting*;
- Form of logarithmic representation → *disc-based*, *cylindrical*, *slide-based*, *propeller-based*, ecc.

Ogni asse è stato creato manualmente, spesso partendo da gruppi concettuali diversi, e poi collegato ai concetti pertinenti. Questo lavoro, svolto progressivamente nel **Concept Editor**, ha permesso di formalizzare oltre cento concetti e più di ottanta oggetti-istanza, ognuno con il proprio set di caratteristiche essenziali, coerenti con la definizione ISO di *concept system*.
Parallelamente, è possibile introdurre le relazioni di ruolo che collegano entità di tipo diverso come:

- *Inventor* ↔ *hasInvented / isInventorOf* ↔ *device*
- *Device* ↔ *hasComponent / isPartOf* ↔ *subdevice*
- *Device* ↔ *hasProducer* ↔ *Organization*

Le relazioni hanno richiesto la definizione precisa di *Domain* e *Range* per garantire coerenza logica, seguendo il modello ontologico importato da Protégé. Ogni inventore — ad esempio *Blaise Pascal*, *Charles Babbage*, *Torres Quevedo* — è stato inserito come istanza della classe *Inventor*, collegata al rispettivo artefatto (*Pascaline*, *Analytical Machine*, *Algebraic Machine*).
Infine, nel **Object Editor**, ogni concetto può essere associato alle relative istanze-oggetto, ovvero agli artefatti reali provenienti dai musei partner o da collezioni pubblicamente documentate. Questa sezione consente di integrare il livello concettuale con la materialità degli oggetti, rafforzando l'allineamento tra modellizzazione teorica e patrimonio culturale effettivamente conservato. Per ogni oggetto sono gestiti:

- identificatore e periodo,

- contesto di provenienza,
- relazioni con produttori e inventori,
- immagini e fonti (con attenzione alle licenze).

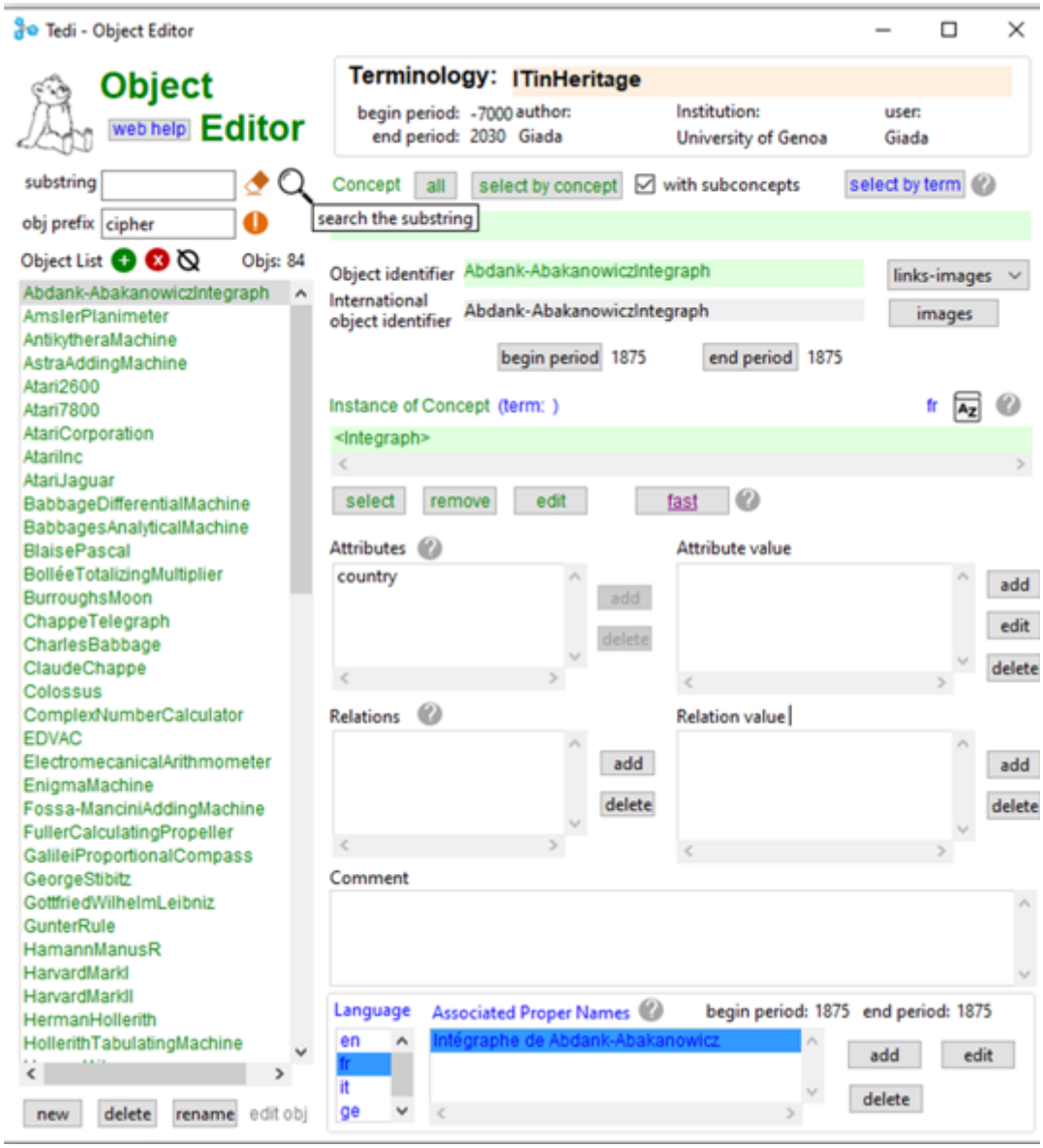


Figure 122 Object editor di TEDI

Ogni concetto è stato arricchito con uno o più oggetti/istanze, definiti nel **Object Editor**, ciascuno corredato da attributi descrittivi (materiale, periodo, produttore, luogo) e da immagini collegate via link esterno (URL), in conformità alle licenze dei musei partner o di Wikimedia Commons.

L'integrazione delle risorse iconografiche all'interno della risorsa ontoterminologica tramite TEDI non risponde a una semplice esigenza estetica, ma a una precisa funzione documentaria. Sebbene nell'attuale fase di implementazione le immagini siano gestite come attributi diretti delle istanze, l'architettura dei dati basata su standard RDF/OWL è intrinsecamente aperta all'integrazione di protocolli internazionali come IIIF (International Image Interoperability Framework).

L'adozione di tali standard, in linea con i principi FAIR e le pratiche più avanzate delle Digital Humanities, permetterebbe di trasformare la riproduzione digitale dell'artefatto in un dato pienamente interoperabile, collegandolo in modo univoco e permanente alla sua definizione formale e garantendone la massima fruibilità all'interno di ecosistemi digitali distribuiti.

Così, ad esempio, la **Schickard Adding Machine** viene modellata come istanza del concetto *Adding machine with cylindrical inscriber*, collegata al suo inventore Wilhelm Schickard e accompagnata da un'immagine con provenienza verificata (ad es. Wikimedia Commons), in modo che la rappresentazione semantica sia riproducibile, verificabile e tracciabile:

Schickard Adding Machine

• Concetto: *Adding machine with cylindrical inscriber*, prima realizzazione nota.

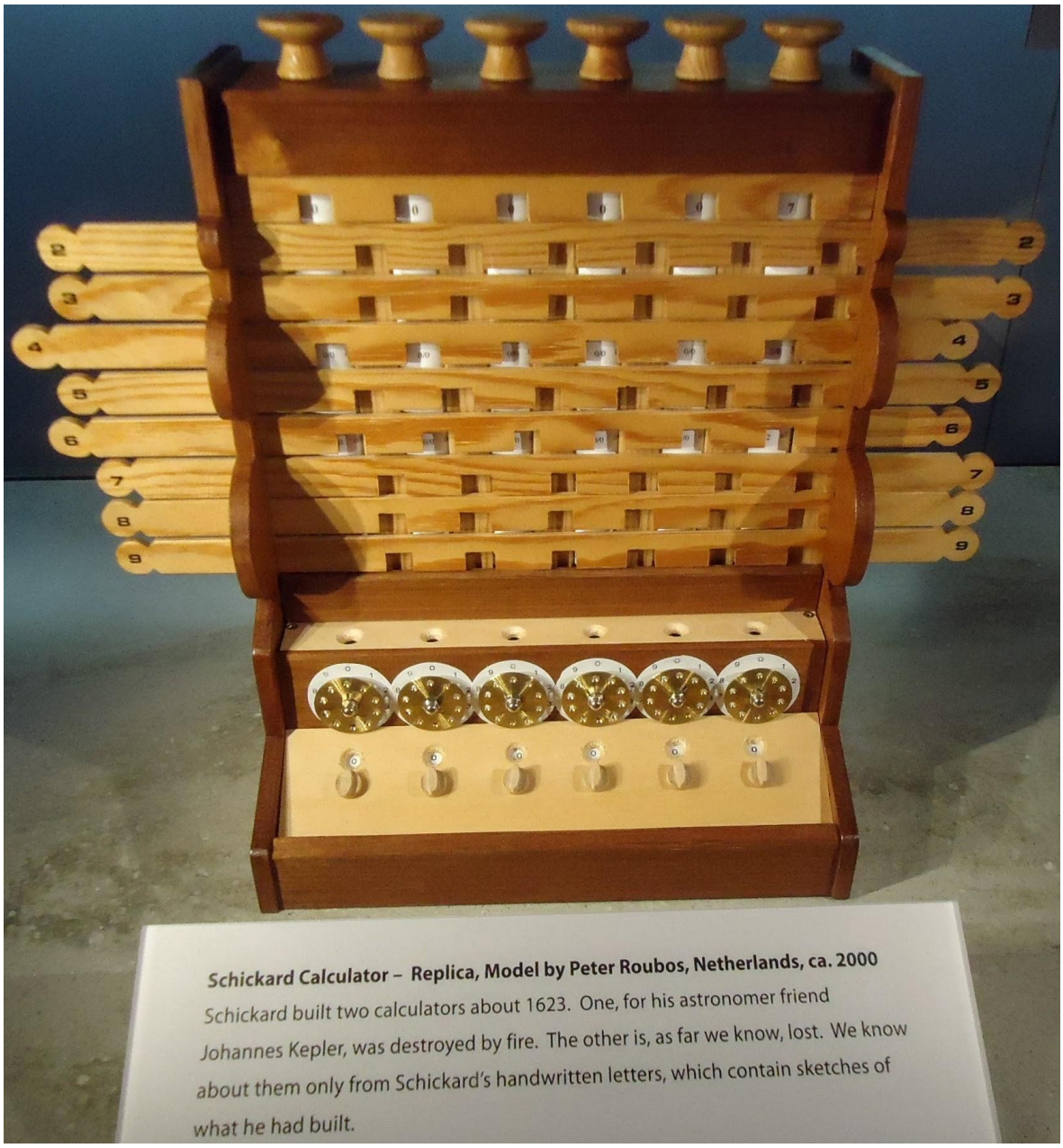
• Inventore: Wilhelm Schickard

• Equivalenti: fr. *Machine à additionner de Schickard*, it. *Addizionatrice di Schickard*

• Asse tecnico: mechanical, gear-based drive

• Nota concettuale: la definizione evidenzia l'uso di cilindri dentati per la trasmissione del calcolo, ponendo il dispositivo in continuità con le successive arithmetical machines.

• Immagine: Schickard Calculator replica (Wikimedia Commons)



Per rendere visibile il funzionamento del modello e la sua applicazione concreta ai materiali museali, si riportano altri esempi organizzati per categorie storiche e funzionali.

Leibniz Multiplier

• Concetto: *Multiplier with grooved cylinder drive*

• Inventore: Gottfried Wilhelm Leibniz

• Equivalenti: fr. *Machine à calcul de Leibniz*, it. *Macchina di Leibniz*

• Asse tecnico: mechanical analog computation

• Definizione: macchina basata sul cilindro scanalato (stepped reckoner), utilizzata per moltiplicazione e divisione.

• Immagine: Leibniz-Rechenmaschine (Wikimedia Commons)

Babbage Difference Machine

• Concetto: *Difference arithmetical* machine, derivato da *Analogue equation-solving machine*

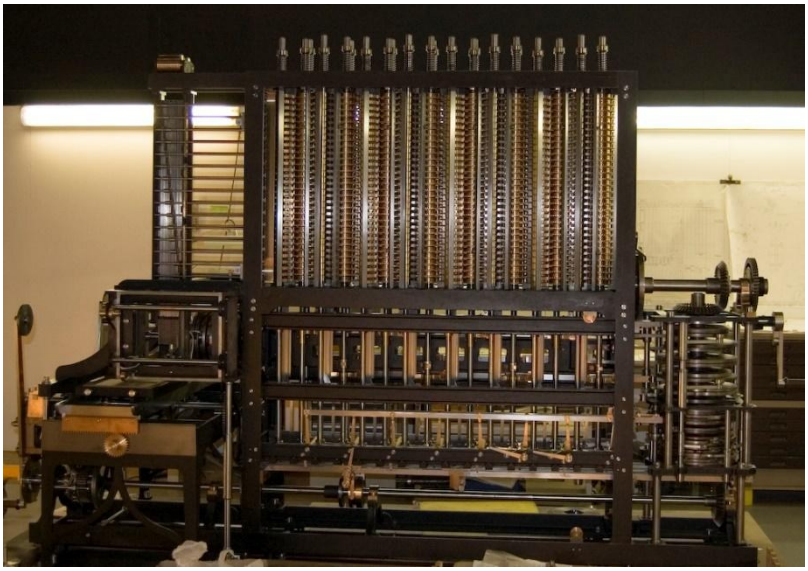
• Inventore: Charles Babbage

• Equivalenti: fr. *Machine à différences de Babbage*, it. *Macchina Differenziale di Babbage*

• Relazione: hasInventor → Babbage

• Asse teorico: mechanical computing of polynomials

• Immagine: Difference Engine replica



Strumenti analogici pre-meccanici e dispositivi di calcolo manuale

Napier's Bones

•Concetto: *Neper stick and derivative*

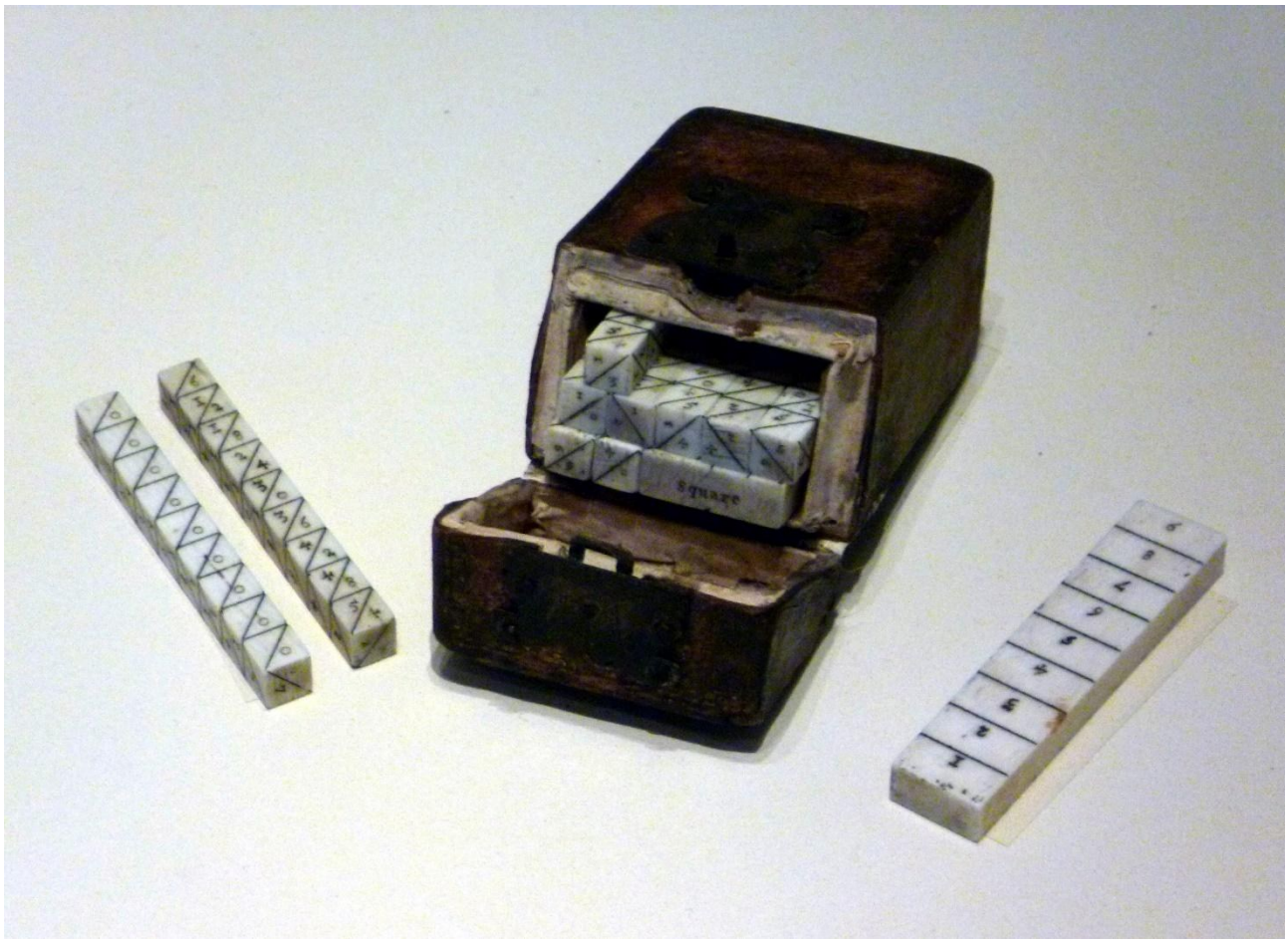
• Inventore: John Napier

• Allonimi: *Napier's rods, Neper sticks*

• Equivalenti: fr. *Bâtons de Neper*, it. *Bastoncini di Nepero*

• Asse tecnico: manual analog device

• Immagine: Napier's Bones (Wikimedia Commons)



Macchine elettromeccaniche e primi calcolatori digitali (XX secolo)

Aiken Relay Calculator

• Concetto: *Electromechanical universal calculator*

• Inventore: Howard Aiken

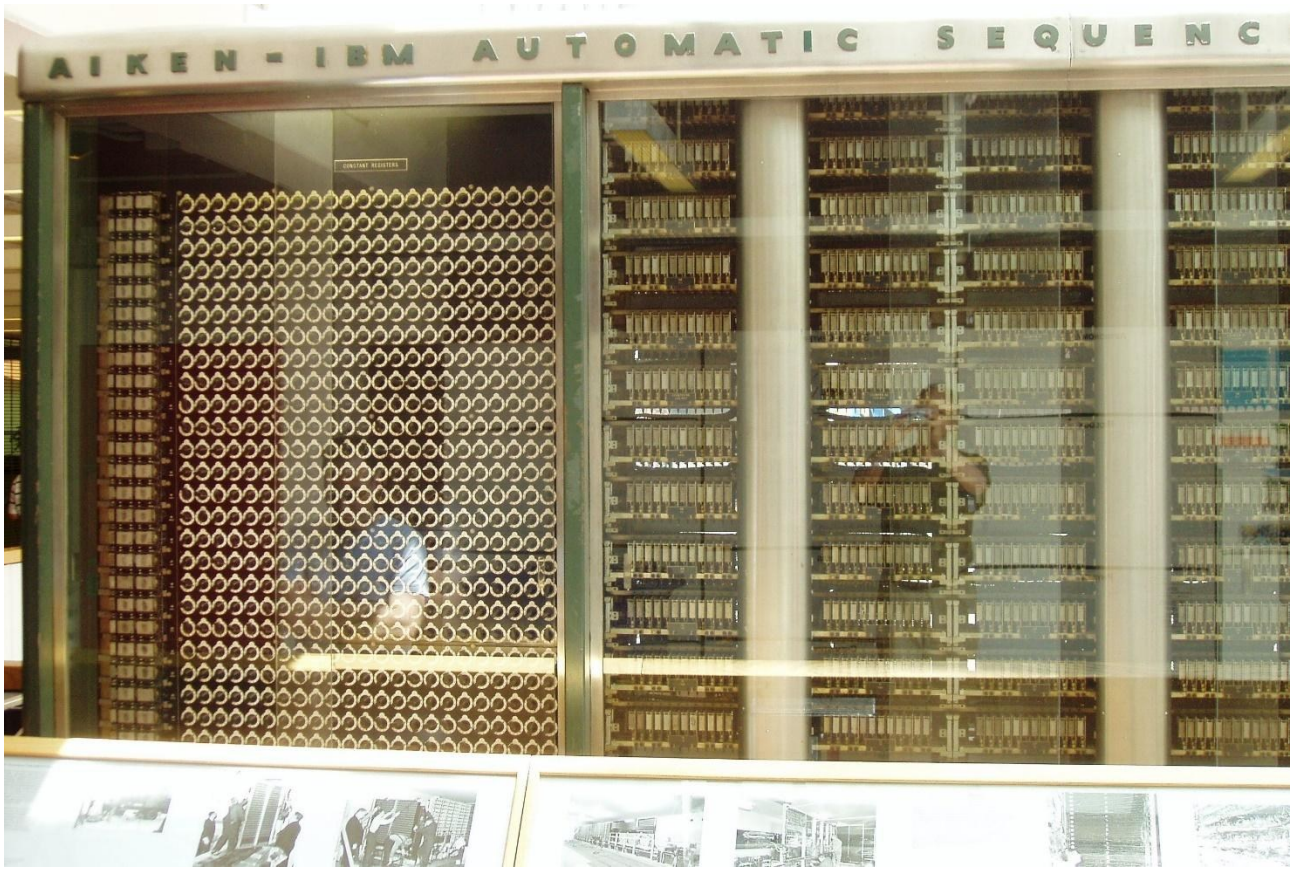
• Relazione: hasInventor → Howard Aiken

•Allonimo*: IBM Automatic Sequence-Controlled Calculator* (ASCC)

• Oggetto: Harvard Mark I

• Nota: mostra il passaggio lessicale da *relay calculator* a *automatic sequence-controlled calculator,* anticipando la terminologia del “programma”.

• Immagine: Harvard Mark I (Wikimedia Commons)



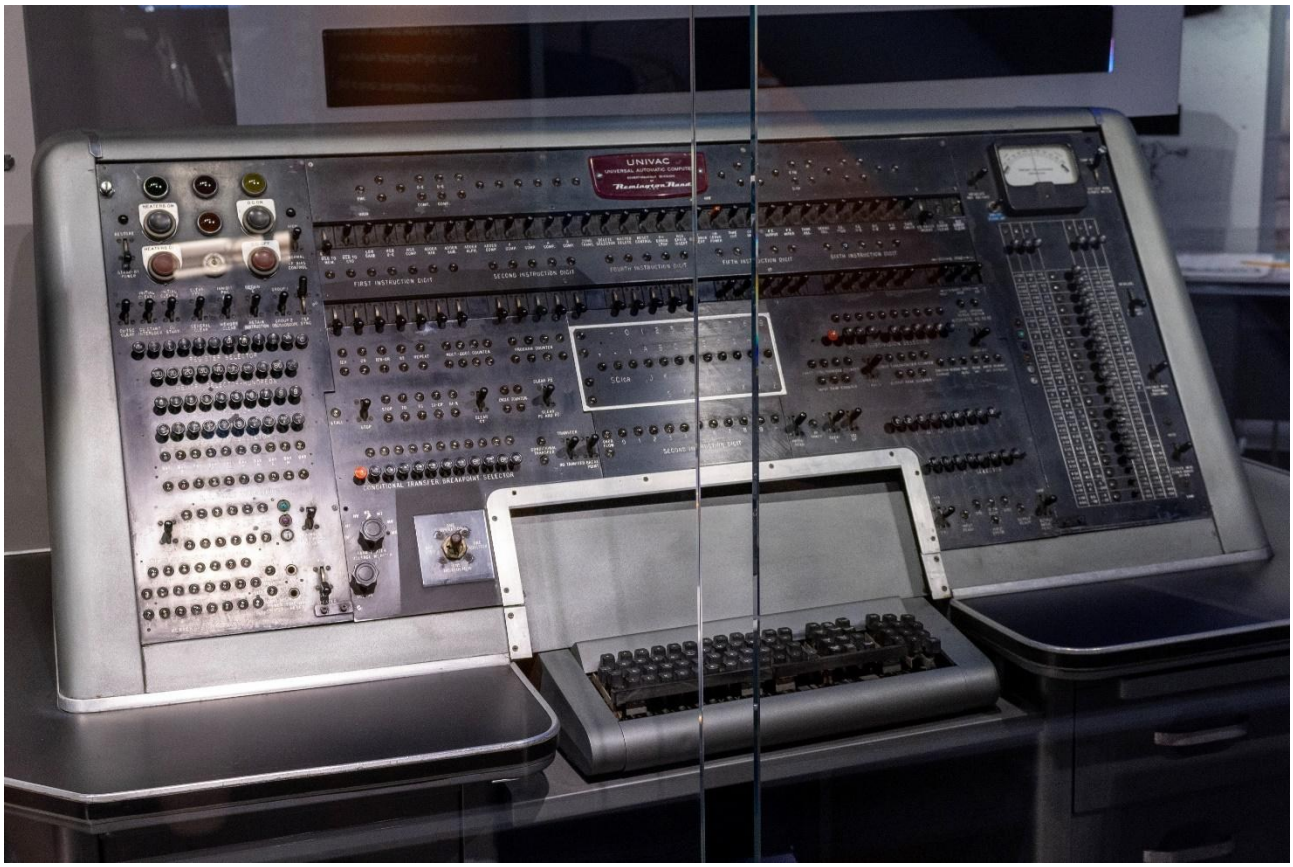
UNIVAC I

• Concetto: *Mainframe*

• Produttore: Remington Rand

• Nota: Il nome proprio incarna la transizione dal linguaggio descrittivo (*automatic computer*) a un nome commerciale, evidenziando la fusione fra terminologia tecnica e marchio.

•Immagine: UNIVAC I (Wikimedia Commons)



Console di gioco elettroniche (1980–2020): l'asse generazionale

Ogni console è modellata come istanza di un concetto specifico lungo l'asse *Generation of videogame console*:

- *Atari VCS → Second generation videogame console*
- *Atari 7800 ProSystem → Third generation*
- *Super Nintendo / Super NES → Fourth generation*
- *PlayStation 1 (PS1) → Fifth generation*
- *PlayStation 2 (PS2) → Sixth generation*
- *Xbox 360 → Seventh generation*
- *Xbox Series X/S → Eighth generation*
- *PlayStation 5 (PS5) → Ninth generation*
- Relazioni: *isProducerOf → Sony Computer Entertainment*, *Nintendo*, *Microsoft*
- Allonimi ed equivalenti: *Super NES ↔ Super Nintendo*; *PS1 ↔ Sony PlayStation 1*; *Xbox Series S ↔ Xbox Series X*.

L'asse generazionale, introdotto nel progetto, permette di collocare i dispositivi in una sequenza cronologica che riflette sia l'evoluzione tecnologica (dal CD al Blu-ray e all'online gaming) sia la variazione terminologica (da *console* a *system* o *series*).

TEDI consente l'esportazione dei dati in formati interoperabili (OWL, RDF, JSON, CSV e dizionari HTML), facilitando l'integrazione con sistemi museali, ontologie culturali come CIDOC CRM e architetture LOD. Queste esportazioni alimentano i dizionari HTML *(SeatTermDictionary, SeatObjectCollection*), la collezione di oggetti e il file RDF finale utilizzati nel progetto ITinHeritage.

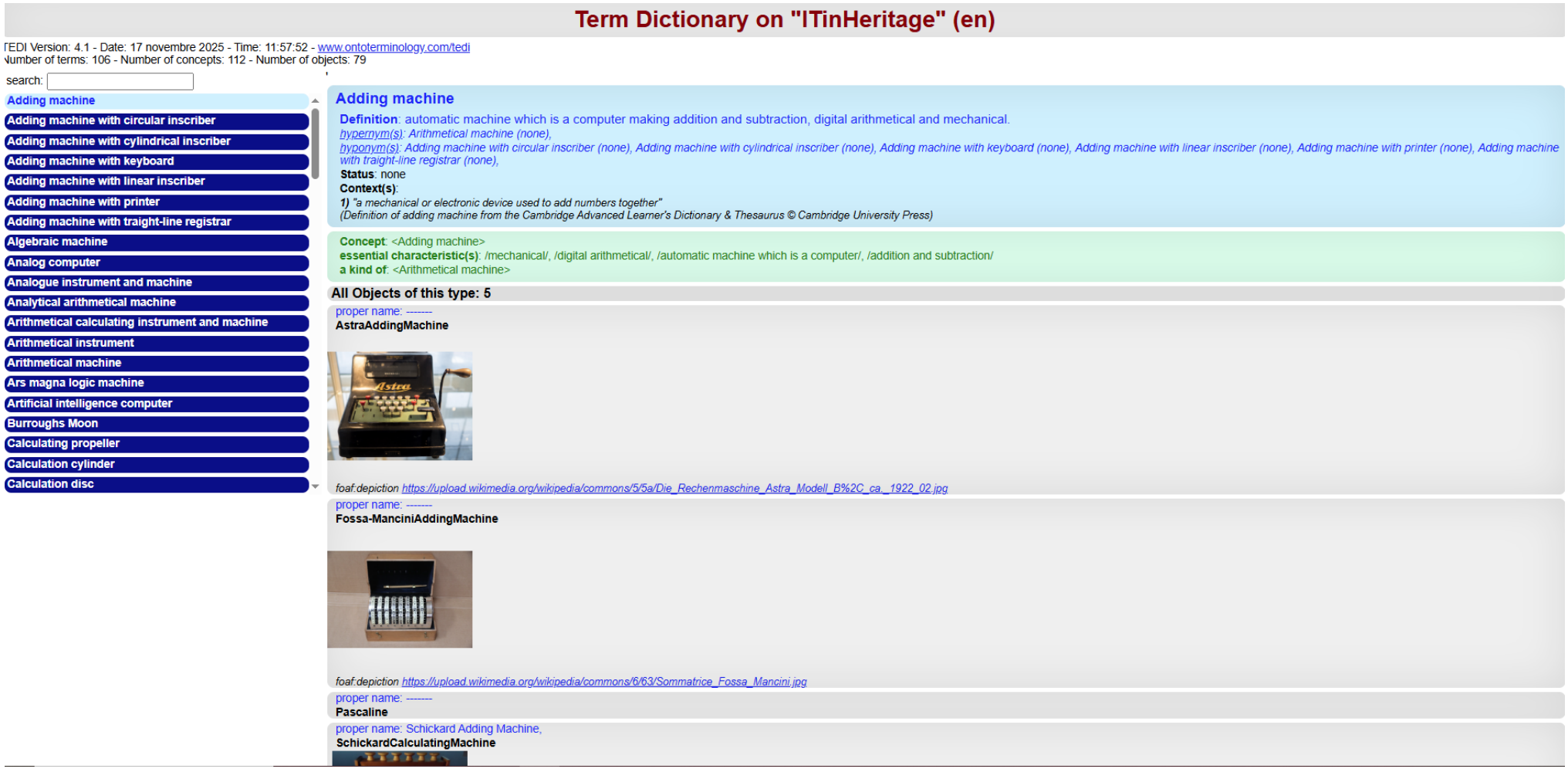


Figure 123 Dizionario terminologico generato da TEDI per il dominio ITinHeritage. Per ogni concetto (es. Adding machine, Electronic calculation machine) il sistema mostra definizione, relazioni gerarchiche, caratteristiche essenziali e istanze concrete associate,

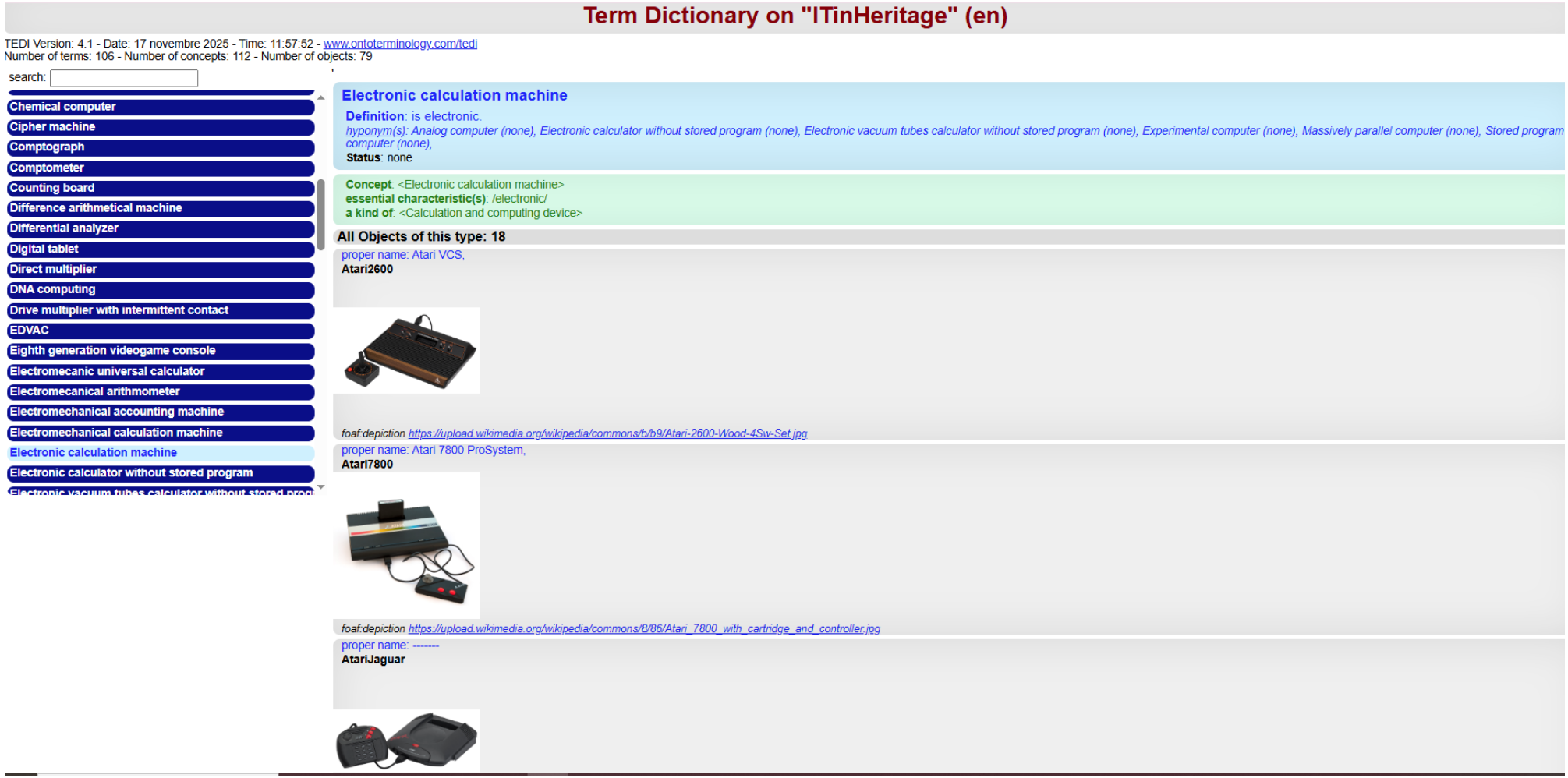


Figure 124 Esempio relativo al concetto Electronic calculation machine: il dizionario HTML generato da TEDI mostra la definizione, le caratteristiche essenziali e l'elenco delle istanze associate (Atari 2600, Atari 7800, Atari Jaguar, ecc.), ciascuna accompagnata

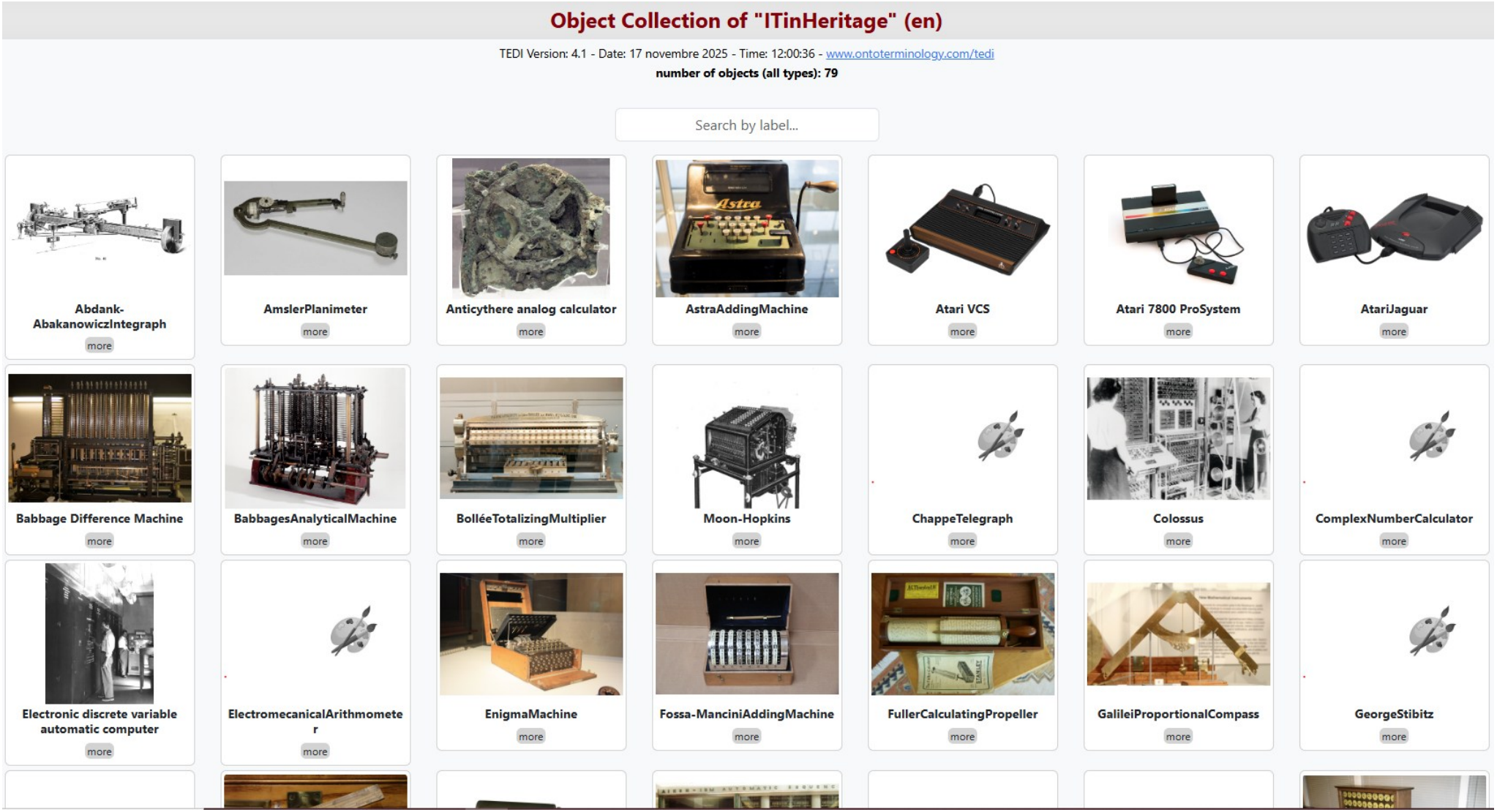


Figure 125 Schermata della Object Collection generata da TEDI, che raccoglie tutte le istanze modellate nel progetto ITinHeritage.

La **Object Collection** mostra l'insieme delle istanze-oggetto modellate nel progetto ITinHeritage. Ogni riquadro rappresenta un artefatto concreto (es. *Pascaline, Astra Adding Machine, Atari VCS*), collegato al rispettivo concetto di appartenenza nella gerarchia "*Calculation and computing device*". Questa visualizzazione consente di esplorare l'intera collezione attraversando i livelli concetto → oggetto e rende evidente la varietà tipologica e cronologica del patrimonio informatico modellato in TEDI.

L'intera risorsa è stata esportata in formato RDF nel file *owlitinheritage.rdf*, consultabile tramite l'OTV Ontology Viewer[46], un'applicazione web interattiva progettata per visualizzare ontologie espresse in RDF/OWL. L'OTV costruisce automaticamente un grafo gerarchico dei concetti e delle relative istanze, distinguendo le relazioni *isA* (tra concetti) e *instanceOf* (tra concetti e oggetti). L'interfaccia offre funzionalità di zoom, pan, rotazione, espansione e collasso dei rami, ricerca per parola chiave, apertura delle schede dettagliate e, quando disponibili, visualizzazione geografica basata sulle coordinate degli oggetti. In questo modo, l'OTV Viewer permette di esplorare il modello ontologico come un grafo dinamico, evidenziando la struttura concettuale del dominio e i legami con gli artefatti materiali. TEDI garantisce la compatibilità con gli standard del Web Semantico (RDF, OWL, SKOS, OntoLex) e genera automaticamente dizionari navigabili e grafi interconnessi.

[46] http://talos-ai4ssh.eu/OTV_Viewer/

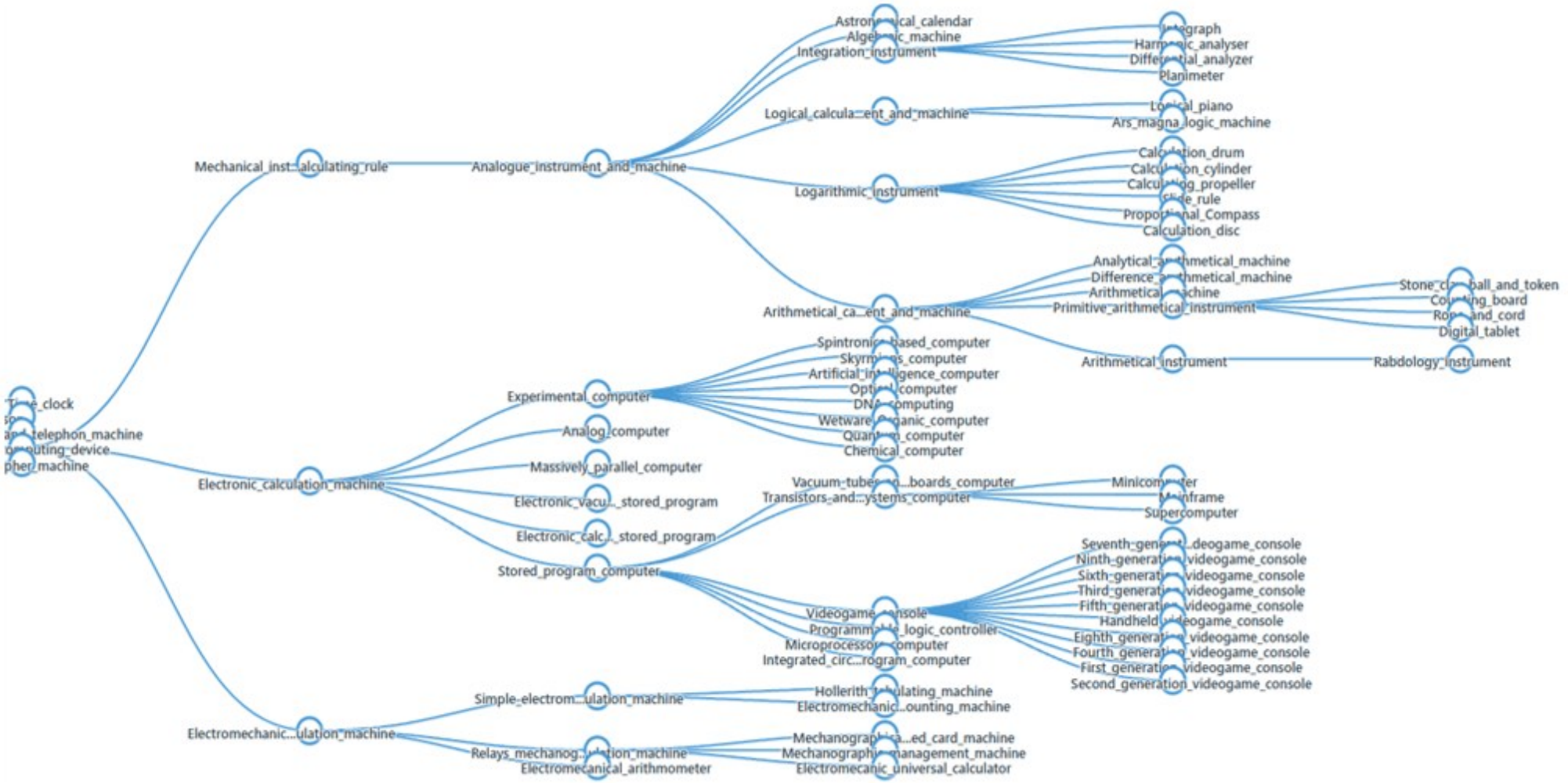


Figure 126 Rete semantica esportata da TEDI e visualizzata nell'OTV Viewer, con le principali classi del dominio Calculation and computing device e le loro relazioni gerarchiche e associative.

Questo approccio garantisce la distinzione formale tra concetti generali e oggetti individuali.
In accordo con la norma ISO 1087, le designazioni terminologiche (termini) sono utilizzate per rappresentare concetti generali del dominio, mentre le entità individuali sono identificate come oggetti o istanze specifiche, eventualmente denominate tramite nomi propri.
Tale distinzione permette di separare il livello concettuale e terminologico da quello degli artefatti materiali, mantenendo coerenza tra analisi linguistica, rappresentazione semantica e modellizzazione.
La rete terminologico-concettuale ottenuta rappresenta oltre 80 oggetti storici, oltre 100 concetti e centinaia di denominazioni multilingue, costituendo una vera ontoterminologia del calcolo e dell'informatica materiale. A questa struttura concettuale si affianca un insieme di istanze documentate, derivate da fonti museali e catalografiche, che fungono da esempi concreti e verificabili dell'applicazione dei concetti. Tali istanze non coincidono con gli oggetti materiali in sé, ma con le loro rappresentazioni documentarie (schede descrittive, immagini, record museali), che rendono osservabile e tracciabile il legame tra concetti, designazioni linguistiche e artefatti storici.
Il passaggio a tre livelli (concetto → termine → rappresentazione visiva) consente a ITinHeritage di coniugare astrazione teorica e materialità museale: la terminologia non rimane mai puramente astratta e, al contempo, gli oggetti non sono trattati come semplici "pannelli fotografici", ma come testimoni storici di categorie concettuali definite. Il risultato è una risorsa che non solo descrive, ma spiega, perché rende esplicito il motivo per cui un artefatto appartiene a una classe e non a un'altra.
Questa è esattamente la differenza tra un dizionario illustrato e una ontoterminologia operativa.

4.3 Discussione critica e prospettive di sviluppo del modello ITinHeritage

Il caso di studio ha mostrato come l'integrazione tra analisi corpus-based, la sistematizzazione della conoscenza esperta (attuata attraverso la ricostruzione concettuale in ambiente TEDI) e formalizzazione ontoterminologica consenta di modellare in modo coerente un dominio complesso come quello del patrimonio informatico e del calcolo. Il passaggio dalla variabilità linguistica osservata nei corpora alla costruzione di un sistema concettuale stabile, fondato su assi di analisi e caratteristiche essenziali, ha permesso di ridurre l'ambiguità terminologica e di rendere esplicite le relazioni tra concetti, designazioni linguistiche e oggetti materiali.

L'uso di TEDI si è rivelato particolarmente efficace nel mantenere separati - ma coordinati - i livelli concettuale e linguistico, rendendo modellabile la conoscenza esperta emersa dalle pratiche curatoriali. Il lavoro ha inoltre evidenziato l'importanza di inserire gli artefatti materiali all'interno di un sistema concettuale che spiega il perché un oggetto appartenga a una classe e non a un'altra, superando l'approccio descrittivo dei cataloghi museali tradizionali.

Un confronto significativo può essere stabilito con progetti recenti come CHAD-KG (Barzaghi *et al.*, 2025), che dimostra come un knowledge graph strutturato su standard come CIDOC-CRM e vocabolari controllati (come AAT) possa efficacemente rappresentare oggetti di patrimonio culturale e i relativi metadati di digitalizzazione. ITinHeritage condivide lo stesso orientamento verso l'interoperabilità e la formalizzazione ontologica, ma si distingue per due aspetti chiave: l'adozione sistematica dell'ontoterminologia, che consente di precisare la struttura intensiva dei concetti attraverso assi, caratteri essenziali e distinzioni *genus–differentiae* e la specializzazione di dominio sugli artefatti informatici, che richiede criteri classificatori specifici (es. tecnologie elettroniche, generazioni di console, funzioni di calcolo, architetture).

Il modello attuale pone quindi basi solide per successive attività di allineamento semantico verso thesauri e ontologie esterne (AAT, Wikidata, CCS), con la possibilità di sfruttare metodologie basate su intelligenza artificiale per suggerire corrispondenze concettuali e incrementare la coerenza e la scalabilità della risorsa.

Il nostro lavoro mostra come il percorso corpus, rete semantica e ontoterminologia non sia soltanto un processo tecnico, ma una strategia epistemologica che consente di connettere linguaggi, concetti e oggetti in un modello computabile, trasparente e riusabile.

Il caso di studio rappresenta un primo passo verso una piattaforma ontoterminologica più ampia, capace di estendersi e adattarsi nel tempo, in dialogo con gli standard museali e le comunità di ricerca.

Conclusioni

1. Sintesi dei risultati: la stratificazione del sapere informatico

La presente tesi ha affrontato il problema della rappresentazione del patrimonio informatico a partire da una constatazione di fondo: gli oggetti dell'informatica, pur essendo centrali nella storia tecnologica e culturale contemporanea, risultano spesso difficili da interpretare, descrivere e mediare. Questa difficoltà non riguarda unicamente la conservazione materiale degli artefatti, ma investe più profondamente la possibilità di strutturare e trasmettere la conoscenza che essi incorporano. Al termine di questo percorso di ricerca, emerge con chiarezza come il patrimonio informatico non sia un blocco monolitico, ma una realtà polifonica e stratificata.

L'analisi condotta su corpora eterogenei — storici, accademici, comunitari e museali — ha confermato l'esistenza di una marcata variabilità semantica tra le diverse comunità di riferimento. Abbiamo osservato come le designazioni informatiche non siano entità stabili, ma riflettano dinamiche di specializzazione, risemantizzazione e mediazione. Mentre il discorso accademico-tecnico si concentra su una visione "internalista", focalizzata sull'architettura e sulla performance (dove termini come *system, data, processor*, mantengono un rigore funzionale), il discorso delle comunità online, analizzato attraverso piattaforme come Reddit, carica l'oggetto informatico di valori affettivi, nostalgici e collezionistici. In questo contesto, il computer cessa di essere un mero strumento di calcolo per diventare un oggetto identitario, una *retromachine* che evoca memorie personali e collettive.

Il museo, situato al centro di queste tensioni discorsive, ha il compito fondamentale di sintetizzare queste visioni. La tesi ha dimostrato che una gestione consapevole della terminologia può fungere da luogo di intersezione tra linguaggio, concetti e oggetti, colmando lo scarto esistente e rendendo i concetti specialistici accessibili al grande pubblico senza tuttavia sacrificarne il rigore scientifico.

Parlare di stratificazione del sapere informatico significa riconoscere che ogni artefatto incorpora livelli molteplici di conoscenza che non coincidono né temporalmente né concettualmente.

A uno stesso oggetto materiale si sovrappongono infatti strati linguistici (le denominazioni che lo hanno accompagnato nel tempo), strati concettuali (i modelli teorici e computazionali che ne hanno guidato la progettazione), strati discorsivi (le narrazioni tecniche, divulgative e comunitarie) e strati culturali (le pratiche d'uso e le rappresentazioni sociali).

Questa complessità rende evidente come la conservazione del solo oggetto fisico non sia sufficiente a preservarne il valore conoscitivo, se non è accompagnata da una ricostruzione semantica di tali livelli.

Un ulteriore elemento che emerge con forza è la dimensione temporale del sapere informatico.

A differenza di altri patrimoni tecnico-scientifici, l'informatica è caratterizzata da cicli di innovazione estremamente rapidi, che producono una continua sovrapposizione di fasi di emergenza, stabilizzazione e obsolescenza. I termini non accompagnano semplicemente questi cambiamenti, ma li rendono visibili: alcune denominazioni si fossilizzano come tracce storiche, altre si generalizzano fino a perdere la loro specificità tecnica, altre ancora scompaiono insieme alle tecnologie che designavano. In questo senso, la terminologia si configura come uno strumento privilegiato per osservare il tempo lungo dell'informatica, rendendo leggibili le trasformazioni concettuali che altrimenti rimarrebbero implicite o invisibili.

2. Valutazione metodologica: l'efficacia dell'approccio corpus-based

Dal punto di vista metodologico, il lavoro ha proposto un percorso innovativo che integra approcci corpus-based, analisi statistico-lessicali e strumenti di estrazione terminologica. L'impiego sinergico di strumenti lessicometrici e distribuzionali quali IRaMuTeQ, TermoStat, Sketch Engine e Tropes su un campione testuale eterogeneo ha consentito di mettere in luce fenomeni di variazione terminologica che altrimenti sarebbero rimasti invisibili a un'analisi puramente manuale, come, ad esempio, il mutamento di statuto del calcolatore nella cultura contemporanea.

L'identificazione di unità polilessicali e il confronto tra registri discorsivi diversi hanno mostrato come il patrimonio informatico venga narrato in modi differenti a seconda del contesto. Questa combinazione di fonti e tecniche ha evidenziato il valore euristico della linguistica computazionale applicata ai beni culturali: non si è trattato solo di contare occorrenze, ma di mappare la densità concettuale di un dominio in continua trasformazione. La terminologia emerge così come uno strumento privilegiato per osservare l'evoluzione delle rappresentazioni dell'informatica nel tempo, permettendo di tracciare la transizione dei termini da neologismi tecnici a parole del lessico comune o, al contrario, a relitti linguistici di tecnologie obsolete.

Nel passaggio dal macro-blocco accademico delle prime decadi (1961–1984) alla produzione scientifica successiva (1985–2024), l'Analisi Fattoriale delle Corrispondenze (AFC) e il calcolo dei valori di specificità hanno registrato una radicale riconfigurazione semantica del lemma *computer*.

Nello specifico, mentre nella prima fase la parola si colloca all'interno di una costellazione terminologica strettamente orientata al funzionamento e alla progettazione dell'architettura computazionale (co-occorrendo stabilmente con *processor, memory, input/output* e *mainframe*), nella seconda fase essa si ridefinisce come fulcro di una narrazione autoriflessiva e storiografica (legandosi a *history, development, research* e *museum*). Questo slittamento distribuzionale non documenta

semplicemente l'evoluzione tecnologica, ma segnala la transizione del calcolatore da macchina operativa a oggetto storico e patrimoniale, inserito all'interno di dinamiche sociali e istituzionali.
Inoltre, l'indagine comparativa estesa al registro spontaneo delle comunità online (Reddit) e a quello controllato dei contesti curatoriali (Science Museum di Londra e HomeComputerMuseum) ha evidenziato profonde asimmetrie discorsive che mettono in crisi l'ideale wüsteriano di corrispondenza biunivoca tra concetto e termine. Se l'universo museale si focalizza sulla materialità oggettuale e sulla nomenclatura dei marchi commerciali (come testimoniano le alte specificità di *device, component, Commodore* o *Apple*), il discorso comunitario reintroduce nel dominio IT una forte dimensione esperienziale, affettiva e nostalgica *(retro, love, cool, keyboard, floppy*). La fluidità e la concorrenza tra varianti apparentemente intercambiabili, quali *personal computer, home computer, PC* e *laptop*, riflettono pertanto prospettive d'uso e stratificazioni culturali divergenti, che richiedono una costante mediazione semantica per non generare opacità nei sistemi informativi digitali.Dal punto di vista epistemologico, l'approccio adottato ha consentito di spostare l'attenzione dalla ricerca di definizioni statiche alla ricostruzione di processi dinamici di concettualizzazione. L'uso dei corpora come osservatori privilegiati delle pratiche discorsive ha permesso di cogliere non soltanto cosa viene detto sull'informatica, ma *come* e *da chi*. In questo senso, la metodologia corpus-based non si limita a supportare l'analisi terminologica, ma contribuisce a ridefinire il modo stesso in cui il patrimonio informatico può essere studiato all'interno delle Digital Humanities, come fenomeno linguistico, culturale e sociale oltre che tecnico.

3. Il valore del caso di studio ITinHeritage: verso l'ontoterminologia

Il progetto europeo ITinHeritage ha rappresentato il banco di prova fondamentale in cui le analisi teoriche sono state tradotte in scelte operative concrete. In questo ambito, la tesi ha documentato il passaggio cruciale dall'analisi semasiologica dei testi (lo studio dei termini nei loro contesti d'uso) a una progressiva strutturazione ontoterminologica del dominio.

La creazione dell'ontologia non è stata concepita come un mero esercizio di catalogazione, ma come un vero atto di "traduzione culturale". Attraverso l'organizzazione multidimensionale delle designazioni mediante assi di analisi (funzione, architettura, periodo storico), è stato possibile gestire la sinonimia e la polisemia in modo formale e rigoroso. Un esempio emblematico è rappresentato dalla distinzione tra Personal computer e Home Computer: grazie alla modellizzazione in ambiente TEDI, tale distinzione cessa di essere una fonte di confusione terminologica per diventare una differenziazione logica basata sul target di mercato, sull'uso e sulla configurazione hardware.

L'ontoterminologia emerge dunque come il vero nodo metodologico del lavoro. Essa non si configura come la semplice applicazione di un modello astratto, ma come un processo che consente di articolare concetti e denominazioni tenendo conto della dimensione storica e patrimoniale. Il confronto con gli standard di catalogazione (come il CIDOC-CRM) ha inoltre evidenziato l'importanza di distinguere tra concetti generali e oggetti materiali individuali, trattando le rappresentazioni visive e i documenti come elementi di supporto che integrano, ma non sostituiscono, la descrizione formale.

Al tempo stesso, la ricerca ha messo in evidenza una tensione strutturale tra esigenza di formalizzazione e complessità semantica. Ogni processo ontologico comporta inevitabilmente una riduzione: alcune sfumature discorsive, ambiguità o usi marginali vengono ricondotti a categorie più stabili. Tuttavia, l'approccio ontoterminologico adottato consente di rendere questa riduzione esplicita e controllata, evitando che essa avvenga in modo implicito o arbitrario. La formalizzazione non elimina la complessità, ma la governa, offrendo strumenti per documentarla, storicizzarla e, se necessario, riarticolarla nel tempo.

In questo quadro, il museo non si configura come uno spazio neutrale di esposizione, ma come un attore epistemico che opera scelte interpretative precise. Decidere quali termini adottare, quali concetti esplicitare e quali relazioni rendere visibili significa orientare la comprensione del pubblico e influenzare la memoria collettiva del passato tecnologico. La terminologia diventa così uno strumento di responsabilità culturale, chiamato a bilanciare esigenze di semplificazione e accuratezza, evitando tanto la banalizzazione quanto l'esclusione cognitiva dei non specialisti.

L'applicazione dell'approccio ontoterminologico al patrimonio informatico attraverso il progetto *ITinHeritage* permette di formulare un bilancio critico che supera la dimensione puramente tecnica della modellizzazione. La formalizzazione condotta in ambiente TEDI ha dimostrato che la riduzione

semantica – inevitabile in ogni operazione di strutturazione dei dati – può essere governata senza tradursi in un appiattimento della ricchezza linguistica. Se i database tradizionali o i sistemi catalografici standardizzati tendono a sacrificare la variazione in nome di una normalizzazione sincronica, l'architettura qui difesa accoglie le oscillazioni denominative emersi dai corpora (come la fitta rete polisemica intorno a *personal computer* e *home computer*) e le stabilizza descrivendone le relazioni concettuali sottostanti.

Il contributo originale di questa ricerca risiede nell'aver dimostrato che il terminologo non si configura come un mero esecutore tecnico al servizio del sistema informativo, ma come un mediatore epistemico indispensabile nell'alveo delle Digital Humanities. In un panorama scientifico segnato dalla proliferazione di dati non strutturati, la costruzione di basi di conoscenza (Knowledge Bases) semanticamente fondate e culturalmente situate rappresenta l'unico strumento capace di garantire l'interoperabilità dei dati e, al contempo, la salvaguardia della memoria storica della tecnologia.

4. Contributi alle Digital Humanities e riflessioni sul ruolo del mediatore

Il lavoro si propone di contribuire al campo delle Digital Humanities, riaffermando la centralità del terminologo come figura di mediazione tra esperti di dominio, curatori museali e sviluppatori di sistemi informativi. In un'epoca dominata dall'intelligenza artificiale e dalla proliferazione di dati non strutturati, la necessità di basi di conoscenza strutturate (Knowledge Bases) e interoperabili è più viva che mai.

L’approccio qui proposto garantisce che la verità storica e tecnica sia preservata e resa fruibile attraverso il Web Semantico, evitando il rischio che il patrimonio informatico diventi una collezione di oggetti muti confinati in database isolati. La tesi dimostra che l'integrazione tra rigore formale e descrizione linguistica è l'unico modo per garantire una valorizzazione del patrimonio che sia al contempo scientificamente accurata e socialmente inclusiva.

5. Limiti della ricerca e prospettive future

Nonostante i risultati raggiunti, la presente ricerca ha fatto emergere alcuni nodi critici che ritengo fondamentale evidenziare. L’ampiezza e la complessità del dominio informatico hanno imposto scelte selettive, sia nella costruzione dei corpora sia nella porzione di concetti effettivamente formalizzati. Tuttavia, questa selettività non va intesa come una parzialità dei dati, ma come la necessità — da me difesa in sede metodologica — di una modellizzazione granulare e situata, che preferisce la profondità storica alla vastità superficiale. Inoltre, la progettazione dell’ontologia globale, inserendosi in un processo collaborativo di ampio respiro come ITinHeritage, ha dovuto mediare tra esigenze diverse,

talvolta esterne agli obiettivi puramente individuali della ricerca. Laddove gli sviluppi applicativi possono divergere verso necessità gestionali, la mia posizione scientifica rimane ancorata alla difesa della complessità semantica: il terminologo non deve "appiattire" le variazioni per favorire il database, ma documentarle per favorire la conoscenza.

Tuttavia, tali limiti aprono la strada a promettenti linee di sviluppo futuro:

- Estensione multilingue: L'analisi potrebbe essere estesa ad ulteriori lingue e corpora per approfondire la dimensione interculturale del patrimonio informatico. Sebbene l'estrazione empirica di questa ricerca si sia concentrata sulla lingua inglese come vettore principale del discorso IT, la struttura ontologica di *ITinHeritage* è nativamente predisposta per l'espansione multilingue. Il passo successivo consisterà nell'integrazione dei sistemi concettuali in lingua italiana e francese, mappando le asimmetrie di prestito e calco (si pensi alla dialettica tra *ordinateur* e *computer*) e le relative implicazioni nei sistemi di catalogazione nazionale (come i tracciati ICCD in Italia).
- Popolamento automatico e interconnessione: Un'estensione naturale riguarderebbe l'uso di tecniche di Natural Language Processing più avanzate per l'estrazione automatica di termini da manuali storici scansionati tramite OCR, velocizzando la fase di popolamento delle ontologie. L'ontologia sviluppata in ambiente Protégé pone le basi per una transizione dell'intero dataset verso il Web Semantico. La pubblicazione della risorsa in formato RDF/OWL e l'allineamento con schemi di livello superiore (*top-level ontologies*) come CIDOC-CRM consentiranno di trasformare *ITinHeritage* in un hub di Linked Open Data, in grado di connettere virtualmente collezioni museali informatiche fisicamente distanti.
- Interfacce di accesso per pubblici differenziati: Infine, sul piano museologico, la risorsa ontoterminologica si presta a essere applicata per la progettazione di interfacce utente dinamiche. La separazione tra il livello concettuale (stabile) e il livello delle designazioni (variabile) permetterà di tarare i percorsi di ricerca all'interno del museo digitale: un utente generico o un appassionato di retrocomputing potrà interrogare il sistema usando lo slang o la terminologia commerciale emersa dal corpus Reddit, mentre un curatore o un ricercatore potrà muoversi attraverso i descrittori controllati della letteratura scientifica o degli standard istituzionali.
- Software Heritage[47]: La metodologia potrebbe essere applicata al patrimonio del software, un ambito ancora più complesso dove il linguaggio naturale del programmatore si fonde con il codice eseguibile.

[47] Disponibile su: https://www.softwareheritage.org/

In un contesto sempre più orientato all'automazione della conoscenza e all'uso di sistemi di intelligenza artificiale, questa ricerca sottolinea l'importanza di una mediazione umana consapevole nella strutturazione del sapere. Le ontologie e le basi di conoscenza non sono semplici infrastrutture tecniche, ma dispositivi che riflettono scelte interpretative, priorità culturali e modelli di comprensione del mondo. La terminologia, in quanto punto di contatto tra linguaggio naturale e rappresentazione formale, occupa una posizione strategica in questo equilibrio, garantendo che i sistemi computazionali non si limitino a manipolare dati, ma operino su conoscenze semanticamente fondate.

In conclusione, questa tesi ribadisce con forza che preservare l'informatica significa, prima di tutto, preservare il linguaggio con cui è stata pensata, progettata e raccontata. Senza una strutturazione semantica consapevole, il rischio è che la rivoluzione digitale venga ricordata attraverso frammenti isolati, privi di connessione e di profondità storica. L'ontoterminologia, in questa prospettiva, non rappresenta soltanto una tecnica di organizzazione della conoscenza, ma un dispositivo critico per mantenere viva la memoria di una trasformazione che ha ridefinito il rapporto dell'umanità con il sapere, il lavoro e la cultura. Comprendere il passato tecnologico attraverso la sua terminologia diventa così una condizione necessaria per orientare in modo consapevole il futuro digitale.

Bibliografia

Adamo, G. (1999). *Terminologia vs. Lessicologia. Istituto per il Lessico Intellettuale Europeo e Storia delle Idee* (ILIESI).

a16z Crypto. (2021). *Why Web3 matters*. https://a16zcrypto.com/posts/article/why-web3-matters/

Barzaghi, A., Frontoni, E., Malavolta, I., *et al.* (2025). *CHAD-KG: a Knowledge Graph for Cultural Heritage Audiovisual Data*. arXiv. https://arxiv.org/html/2505.13276v1

Berners-Lee, T. (2012). *Linked data, web science and the semantic web*. In J. Brockman (Ed.), Edge. https://www.edge.org/conversation/tim_berners_lee-linked-data-web-science-and-the-semantic-web

Berners-Lee, T., Hendler, J., Lassila, O. (2001). *The Semantic Web.* Scientific American, 284(5), 34–43.

Biagetti, M. T. (2020). *Ontologies (as Knowledge Organization Systems).* In ISKO Encyclopedia of Knowledge Organization.

Bonomi, F. (2004–2008). *Vocabolario Etimologico della Lingua Italiana.* https://www.etimo.it/?term=ontologia&find=Cerca

Borchman, R., Levesque, H. (2004). *Knowledge Representation and Reasoning. Morgan Kaufmann.*

Boulanger, J.-C. (1995). *Présentation: images et parcours de la socioterminologie*. Meta, 40(2), 194–205. https://doi.org/10.7202/002117ar

Bourigault D., Slodzian M., 1999, *Pour une terminologie textuelle*. Terminologies nouvelles 19, pp. 29-32.

Bowker, L. (1995). *A multidimensional approach to classification in Terminology*. PhD Thesis, University of Manchester.

Brachman, R., Levesque, H. (2004). *Knowledge Representation and Reasoning*. Morgan Kaufmann.

Cabré, M. T. (1999). *Terminology. Theory, Methods and Applications*. John Benjamins.

Cabré, M. T. (2023). *Terminology: Cognition, language and communication*. John Benjamins.

Capuano, N. (2005). *Ontologie OWL: Teoria e Pratica.* COMPUTER PROGRAMMING, 148, 59–64.

CIDOC CRM Special Interest Group (2021). *Definition of the CIDOC Conceptual Reference Model*. Versione 7.1.1. https://cidoc-crm.org/

Condamines, A. (2018). T*erminological knowledge bases: From Texts to Terms, from Terms to Texts*. In The Routledge Handbook of Lexicography. Routledge.

Condamines, A. (2022). *Terminologie, Intelligence Artificielle et Psychologie cognitive*. De Europa, numero speciale, 131–148.

Computer Science Ontology (CSO). (2024). Sito ufficiale.
https://cso.kmi.open.ac.uk/

Crofts, N., Doerr, M., Gill, T., Stead, S., Stiff, M. (2009). *CIDOC Conceptual Reference Model* (v. 5.0.1). ICOM/CIDOC.
https://cidoc-crm.org/

de Boer, V., Wielemaker, J., van Gent, J., Oosterbroek, M., Hildebrand, M., Isaac, A., van Ossenbruggen, J., Schreiber, G. (2013). *Amsterdam museum linked open data*. Semantic Web, 4(3), 237–243.

Desprès, S., Roche, C., Papadopoulou, M. (2020). *Étude comparative de deux méthodes outillées*. In TOTh 2020.

Djambian, C. (2025). *The ITinHeritage project*. AIDAinformazioni, 1–2.

Djambian, C., Rossi, M., Frérot, C., Poncet, E., D'Ippolito, G. (2024). *Des représentations du monde des Technologies de l'Information.* In Actes TOTh 2024.

Doerr, M., Gradmann, S., Hennicke, S., Isaac, A., Meghini, C., Van de Sompel, H. (2010). *The Europeana Data Model.* IFLA.

Drouin, P. (2003). *Term extraction using non-technical corpora as a point of leverage*, In Terminology, vol. 9, no 1, p. 99-117.
https://termostat.ling.umontreal.ca/ [Software]

Europeana Foundation. (2008).

Faber, P. (2002). *A cognitive linguistics view of terminology and specialized language*. https://doi.org/10.1515/9783110277203

Faber, P. (2015). *Frames as a framework for terminology*. In Handbook of Terminology.

Faber, P., L'Homme, M. C. (2022). *Theoretical Perspectives on Terminology*. John Benjamins.

Felber, H. (1984). *Terminology Manual*.

Fóris, Á. (2008). *Terminologia, modelli terminologici e reti*. AIDAinformazioni, 26(1–2), 37–46.

Gaudin, F. (1993). *Pour une socioterminologie: des problèmes sémantiques aux pratiques institutionnelles*. Université de Rouen.

Getty Research Institute (2024). *Art & Architecture Thesaurus Online.*
https://www.getty.edu/research/tools/vocabularies/aat/help.html

Giunchiglia, F., Dutta, B., Maltese, V. (2014). *From Knowledge Organization to Knowledge Representation.* Knowledge Organization, 41(1).

Gnoli, C., Marino, V., Rosati, L. (2006). *Organizzare la conoscenza*. HOPS.

Gruber, T. R. (1993). *A translation approach to portable ontology specifications. Knowledge Acquisition*, 5(2), 199–220.

Grüninger, M., Fox, M. S. (2005). *Methodology for the Design and Evaluation of Ontologies*.

Haigh, T. (2001). *The Chromium-Plated Tabulator*. IEEE Annals of the History of Computing, 23(4).

Hjørland, B. (2008). *What is Knowledge Organization? Knowledge Organization*, 35(2).

Hjørland, B. (2016). *Knowledge organization. Knowledge Organization*, 43(6).

Hodge, G. (2000). *Systems of Knowledge Organization for Digital Libraries*. CLIR.

Humbley, J. (1997). *Is terminology specialized lexicography? The experience of French-speaking countries*. Hermes, Journal of Linguistics, (18).

Hyvönen, E., Mäkelä, M., Salminen, M., *et al.* (2005). *MuseumFinland. Journal of Web Semantics*, 3(2).

IIIF Consortium. (2024). *International Image Interoperability Framework (IIIF) Specifications*. Disponibile su: https://iiif.io/technical-details/

Intorre, S. (2008). *La Tecnologia dell'Informazione per musei e collezioni*. IMAFRONTE, 19–20.

IRaMuTeQ. (2014). *Documentation officielle*. http://iramuteq.org/documentation/fichiers/documentation_19_02_2014.pdf

ISO 1087-1. (2000). *Terminology work — Vocabulary*.

ISO 704. (2009/2022). *Terminology work — Principles and methods*.

ISO 5078. (2023). *Terminology work — Definitions.*

Hjørland, B. (2023). *Knowledge Organization Systems (KOS)*. In Encyclopedia of Knowledge Organization (IEKO). ISKO.

Kageura, K. (1997). *Multifaceted/multidimensional concept systems*. In Handbook of terminology management.

Kageura, K. (1999). *On the Study of Dynamics of Terminology*. NACSIS.

Kageura, K. (2002). *The Dynamics of Terminology: A Descriptive Theory of Terminology in Relation to Thesaurus Construction.* John Benjamins Publishing Company. https://books.google.it/books/about/The_Dynamics_of_Terminology.html?id=S8VdJoDe8YwC

Kerremans, K., Temmerman, R., Tummers, J. (2003). *Termontography*. In OTM Conferences. Springer.

Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, D. (2004). *The Sketch Engine*. In Proceedings of the 11th EURALEX International Congress, Lorient, Francia, pp. 105-116. https://www.sketchengine.eu [Software].

LARHRA. *OntoMe: Ontology Management Environment*. Laboratoire de Recherche Historique Rhône-Alpes. https://ontome.net/

Lazard, E., Mounier-Kuhn, P. (2016). *Histoire illustrée de l'informatique*. EDP Sciences.

L'Homme, M. C. (2004). *La terminologie : principes et techniques*. PUM.

L'Homme, M. C. (2008). *Ressources lexicales, terminologiques et ontologiques*. RFLA, XIII.

L'Homme, M. C. (2012). *Le verbe terminologique*. SHS Web of Conferences.

L'Homme, M. C. (2016). *Terminologie de l'environnement et sémantique des cadres*. CMLF.

Longhi, J., Marinica, C., Després, Z., Plancq, C. (2018*). CMC corpora analysis*. HAL.

Mahoney, M. S. (1988). *The History of Computing in the History of Technology.* Annals of the History of Computing.

Mahoney, M. S. (2005). *The histories of computing(s).* Interdisciplinary Science Reviews.

Miarka, R., Zacek, M. (2011). *Knowledge patterns for RDF*. FedCSIS.

Molette, P., Landre, A. (2021). *Tropes* (version 8). http://www.tropes.fr/ [Software]

Noy, N. F., McGuinness, D. (2001). *Ontology Development 101*.

Øhrstrøm, P., Schärfe, H., Uckelman, S. (2008). *Jacob Lorhard's Ontology*. In ICCS 2008. Springer.

Ontotext. (n.d.). *Five-Star Linked Open Data*.
https://www.ontotext.com/knowledgehub/fundamentals/five-star-linked-open-data/

Perry, S. (2024). *The Museum Data Service: unlocking 80 million object records*. Art UK.

Pontrandolfo, G. (2010). *La fase preliminare del processo penale. Il contributo dell'approccio sociocognitivo a un'indagine terminografica spagnolo-italiano*, Università di Trieste.

Ratinaud, P. (2014). *IRAMUTEQ: Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires [Computer software]*.
Testo disponibile al sito: http://www.iramuteq.org

Reinert, M. (1983). *Méthode de classification descendante hiérarchique*. CAD, 8(2).

Riediger, H. (2023). *Cos'è la terminologia e come si fa un glossario*.

Roche, C. (2005). *Terminologie et ontologie*. Langages, 157.

Roche, C. (2007). *Le terme et le concept: fondements d'une ontoterminologie*. In Proceedings of TOTh 2007 (Terminologie & Ontologie: Théories et Applications), Annecy.

Roche, C., Calberg-Challot, M., Damas, L., Rouard, P. (2009) *Ontoterminology: A new paradigm for terminology*. International Conference on Knowledge Engineering and Ontology Development, Madeira, Portugal. pp.321-326. ffhal-00622132

Roche, C. (2011). *TOTh Proceedings - Terminology & Ontology: Theories and applications*. Annecy, Institut Porphyre, Savoir et Connaissance, hal-01354937.

Roche, C., Papadopoulou, M. (2019). *Mind the Gap: Ontology Authoring for Humanists.* Joint Ontology Workshops, JOWO. https://api.semanticscholar.org

Roche, C., Papadopoulou, M. (2020). *Terminologie et ontologie pour les humanités numériques. Humanités numériques*.

Rosch, E. (1978). *Principles of Categorization*. University of California.

Sager, J. C. (1990). *A Practical Course in Terminology Processing*. John Benjamins.

Sakr, A. B. (2022). *Sulle caratteristiche delle collocazioni*. Sapienza Università Editrice.

Salatino, A. A., Thanapalasingam, T., Mannocci, A., Birukou, A., Osborne, F., Motta, E. (2020). *Computer Science Ontology*. Data Intelligence, 2(3).

Salatino, A. A., Thanapalasingam, T., Mannocci, A., Osborne, F., Motta, E. (2018). *Computer Science Ontology*. ISWC 2018.

Salem, A. (1991). *Les séries textuelles chronologiques*. Histoire & Mesure, 6(1/2), 149–175.

Sanfilippo, E., Markhoff, B., Pittet, P. (2020). *Ontological Analysis of CIDOC-CRM*. FOIS.

Santos, R. C., Famaey, J., Schönwälder, J., Granville, L. Z., Pras, A., Turck, F. (2016). *Taxonomy for the network and service management research field*. Journal of Network and Systems Management, 24, 764–787. https://doi.org/10.1007/s10922-015-9363-7

Silva, A. L., Terra, A. L. (2024). *Cultural heritage on the Semantic Web*. IFLA Journal, 50(1).

Soergel, D. (2009*). Knowledge Organization Systems*. Overview.

Studer, R., Benjamins, V. R., Fensel, D. (1998). *Knowledge Engineering. Data & Knowledge Engineering.*

Temmerman, R. (2000). *Towards New Ways of Terminology Description*. John Benjamins.

Torre, I., Albertoni, R. (2023/2024). *Semantic Web e Linked Data*. Dispense, Università di Genova.

Vezzani, F. (2022). *Terminologie numérique.* Linguistic Insights.

W3C. (2026) *RDF 1.2 Concepts and Abstract Syntax* [online]. W3C Working Draft. Disponibile su: https://www.w3.org/TR/rdf12-concepts/

W3C. (2009). *SKOS Simple Knowledge Organization System Reference*. W3C Recommendation. Disponibile su: https://www.w3.org/TR/skos-reference/

Warburton, T., Humbley, J. (Eds.). (2025). *Terminology throughout History. A Discipline in the Making*. John Benjamins.

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al*. (2016). *The FAIR Guiding Principles for scientific data management and stewardship*. Sci Data 3, 160018 https://doi.org/10.1038/sdata.2016.18

Wüster, E. (1979) *Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie*. Wien & New York: Springer.

Zeng, M. L. (2008). *Knowledge Organization Systems* (KOS). Knowledge Organization, 35(2–3).