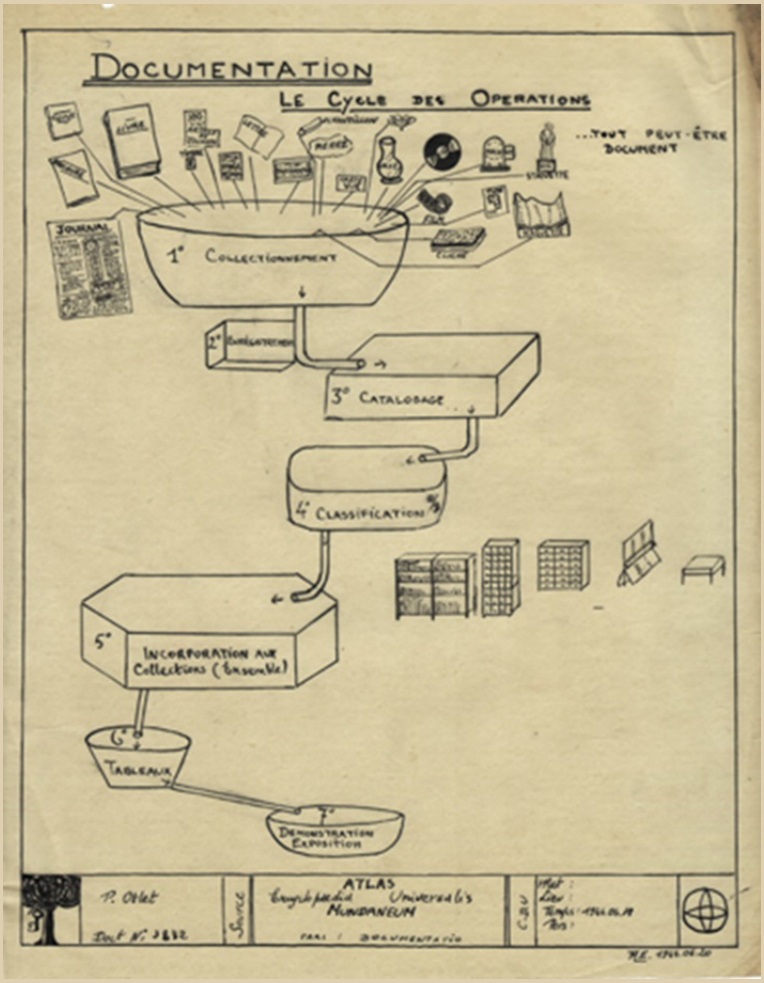
AIDAinformazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

NUMERO 3-
ANNO 43
LUGLIO-DICEMBRE 2025



CACUCCI EDITORE
BARI

AIDAinformazioni

Rivista semestrale di Scienze dell'Informazione

Fondata nel 1983 da Paolo Bisogno

Proprietario della rivista:
Università della Calabria

Direttore Scientifico:
Roberto Guarasci, *Università della Calabria*

Direttore Responsabile:
Fabrizia Flavia Sernia

Comitato scientifico:
Anna Rovella, *Università della Calabria*;
Maria Guercio, *Sapienza Università di Roma*;
Giovanni Adamo, *Consiglio Nazionale delle Ricerche* †;
Claudio Gnoli, *Università degli Studi di Pavia*;
Ferruccio Diozzi, *Centro Italiano Ricerche Aerospaziali*;
Gino Roncaglia, *Università della Tuscia*;
Laurence Favier, *Université Charles-de-Gaulle Lille 3*;
Madjid Ihadjadene, *Université Vincennes-Saint-Dénis Paris 8*;
Maria Mirabelli, *Università della Calabria*;
Agustín Vivas Moreno, *Universidad de Extremadura*;
Douglas Tudhope, *University of South Wales*;
Christian Galinski, *International Information Centre for Terminology*;
Béatrice Daille, *Université de Nantes*;
Alexander Murzaku, *College of Saint Elizabeth*, USA;
Federico Valacchi, *Università di Macerata.*

Comitato di redazione:
Antonietta Folino, *Università della Calabria*;
Erika Pasceri, *Università della Calabria*;
Maria Taverniti, *Consiglio Nazionale delle Ricerche*;
Maria Teresa Chiaravalloti, *Consiglio Nazionale delle Ricerche*;
Assunta Caruso, *Università della Calabria*;
Claudia Lanza, *Università della Calabria.*

Segreteria di Redazione:
Valeria Rovella, *Università della Calabria*

Editrice: Cacucci Editore S.a.s.
Via D. Nicolai, 39 – 70122 Bari (BA)
www.cacucciditore.it
e-mail: riviste@cacucciditore.it
Telefono 080/5214220

AIDAinformazioni

Rivista semestrale di Scienze dell'Informazione

«AIDAinformazioni» è una rivista scientifica che pubblica articoli inerenti alle Scienze dell'Informazione, alla Documentazione, all'Archivistica, alla Gestione Documentale e all'Organizzazione della Conoscenza ma amplia i suoi confini in ulteriori campi di ricerca affini quali la Terminologia, la Linguistica Computazionale, la Statistica Testuale, ecc. È stata fondata nel 1983 quale rivista ufficiale dell'Associazione Italiana di Documentazione Avanzata e nel febbraio 2014 è stata acquisita dal Laboratorio di Documentazione dell'Università della Calabria. La rivista si propone di promuovere studi interdisciplinari oltre che la cooperazione e il dialogo tra profili professionali aventi competenze diverse, ma interdipendenti. I contributi pubblicati affrontano questioni teoriche, metodologie adottate e risultati ottenuti in attività di ricerca o progettuali, definizione di approcci metodologici originali e innovativi, analisi dello stato dell'arte, ecc.

«AIDAinformazioni» è riconosciuta dall'ANVUR come rivista di Classe A per l'Area 11 – Gruppo Scientifico Disciplinare 11/HIST-04 – Scienze del libro, del documento e storico-religiose e come rivista scientifica per le Aree 10 – Scienze dell'antichità, filologico-letterarie e storico-artistiche; 11 – Scienze storiche, filosofiche, pedagogiche e psicologiche; 12 – Scienze giuridiche; 14 – Scienze politiche e sociali. È anche annoverata dall'ARES (Agence d'évaluation de la recherche et de l'enseignement supérieur) tra le riviste scientifiche dell'ambito delle Scienze dell'Informazione e della Comunicazione. La rivista è, inoltre, indicizzata in: ACNP – Catalogo Italiano dei Periodici; BASE –Bielefeld Academic Search Engine; ERIH PLUS – European Reference Index for the Humanities and Social Sciences – EZB – Elektronische Zeitschriftenbibliothek – Universitätsbibliothek Regensburg; Gateway Bayern; KVK – Karlsruhe Virtual Catalog; The Library Catalog of Georgetown University; SBN – Italian union catalogue; Ulrich's; Union Catalog of Canada; LIBRIS – Union Catalogue of Swedish Libraries; Worldcat.

I contributi sono valutati seguendo il sistema del *double blind peer review*: gli articoli ricevuti sono inviati in forma anonima a due referee, selezionati sulla base della loro comprovata esperienza nei topics specifici del contributo in valutazione.

AIDAinformazioni

Anno 43

N. 3-4 – luglio-dicembre 2025

CACUCCI EDITORE

BARI

PROPRIETÀ LETTERARIA RISERVATA

© 2025 Cacucci Editore – Bari
Via Nicolai, 39 – 70122 Bari – Tel. 080/5214220
http://www.cacucciedit ore.it e-mail: info@cacucci.it

Ai sensi della legge sui diritti d'Autore e del codice civile è vietata la riproduzione di questo libro o di parte di esso con qualsiasi mezzo, elettronico, meccanico, per mezzo di fotocopie, microfilms, registrazioni o altro, senza il consenso dell'autore e dell'editore.

Sommario

Giovanna Aracri, Antonietta Folino, Maria Teresa Guaglianone, Marco Manna, Identification of thesaurus relationships with the support of ChatGPT and Gemini 9

Alessandra Bonsignori, Terminologia giuridica e organizzazione della conoscenza 51

Maria Teresa Chiaravalloti, Maria Teresa Guaglianone, Elena Cardillo, Metadatazione delle tipologie di documenti clinici: un approccio alla standardizzazione mediante LOINC 73

Carla Beatriz Marques Felipe, Carla Maria Martellote Viola, Raimunda Fernanda dos Santos, Domain Analysis on Data Management 91

Claudio Gnoli, Emanuela Valenzano, Che faccetta sei? 111

Manuela Grillo, Strumenti di convergenza tra archivi e biblioteche: i thesauri 121

Bianca La Manna, Rappresentare gli atti notarili medievali in Linked Open Data (LOD) 141

Claudia Lanza, L'univocità e la variabilità terminologica a supporto della rappresentatività delle risorse semantiche di organizzazione e rappresentazione della conoscenza specialistica 163

Larissa Moraes Martins, Carlos Eduardo da Silva Carvalho, Ana Cristina de Albuquerque, Mario Barité, Possibilities for using cultural warrant to complement literary work 181

Bianca Therezinha Carvalho Panisset, Lucia Maria Velloso de Oliveira, Leticia Souza da Costa Sampaio, Camilla Campoi de Sobral, Identification of types of documents and knowledge organization: transparency and access in records management 197

SoYoung Yoon, Claudio Gnoli, Ziyoung Park, Applying Integrative Levels Classification to Korean KOS Registry 211

Emanuela Nicole Donato, Grazia Serratore, Maria Taverniti, Maria Teresa Chiaravalloti, L'evoluzione della struttura documentale nelle cartelle cliniche psichiatriche dell'archivio storico del manicomio di Girifalco (1881-1978) 237

Lorenza Talarico, Manuscripts in the Digital Age 263

Maria Lígia Triques, Ananda Fernanda de Jesus, Mario Barité, José Eduardo Santarem Segundo, Ana Cristina de Albuquerque, Vocabularies in the digital context 275

Marcin Trzmielewski, Context-Driven Use of Automatic Tools in Medical Ontology Design: A Case Study in Allergology 291

Carla Maria Martellote Viola, Carla Beatriz Marques Felipe, Exploring the relationships between the Cross-Cutting Women's Agenda - Multi-Year Plan 2024-2027 and the Knowledge Organization of legislative proposals under processing on women in Brazil 313

Elena Cencetti, Camillo Carlo Pellizzari di San Girolamo, Elisabetta Viti, Modellizzazione delle conoscenze: Thesaurus Nuovo soggettario e Wikidata, due strumenti a confronto 325

Pontish Yeramyan, Alexander Murzaku, Nick Uva, Curt Anderson, The transferability of brilliance 353

Ana Flávia Dias Zammataro, Ana Cristina de Albuquerque, Giulia Crippa, Lucia Sardo, Carlo Bianchini, Maria Lígia Triques, ISKO Italy in perspective 375

Contributi

AIDAinformazioni
ISSN 1121–0095
ISBN 979-12-5965-639-1
DOI 10.57574/596563918
pag. 163-180 (luglio-dicembre 2025)
Open access article licensed under CC-BY-NC-SA

L'univocità e la variabilità terminologica a supporto della rappresentatività delle risorse semantiche di organizzazione e rappresentazione della conoscenza specialistica

Claudia Lanza*

Abstract: This paper aims at presenting a reflection upon some of the main mechanisms with which the semantic representativeness can be guaranteed by organization and representation knowledge systems with reference to a specialized domain of study. The analysis will dive into the concept of term uniqueness and update within these semantic resources. The perspective is to offer a methodological approach that could be employed over other domains characterized by a specialized terminological asset extracted from a source documentation strictly compliant with the sector of interest.

Keywords: Knowledge Organization Systems, Corpus-based representativeness, Information retrieval, Thesauri, Semantic coverage.

1. Introduzione

L'analisi verterà dapprima sull'importanza rivestita dal livello di specificità terminologica all'interno dei sistemi di organizzazione e gestione della conoscenza (Knowledge Organization Systems - KOSs). Verranno approfondite le fasi proprie dello sviluppo degli strumenti semantici, i.e., (i) la costruzione del corpus di partenza specialistico, connotato dai concetti chiave dei settori oggetto di studio selezionati sulla base di un'accurata estrazione all'interno della documentazione di riferimento, (ii) la trasformazione dei concetti chiave in dati strutturati, termini candidati per far parte delle risorse di organizzazione della conoscenza, (iii) la validazione di determinati asset terminologici nella prospettiva di garantire una copertura semantica rispetto al dominio di riferimento. Successivamente si presenteranno alcune tecniche qualitative e statisti-

* Dipartimento di Culture, Educazione e Società, Università della Calabria, Rende (CS), Italia. claudia.lanza@unical.it. ORCID: 0000-0002-3018-1987.

Il lavoro presenta un approfondimento in italiano di parte delle tecniche illustrate nella monografia pubblicata con l'editore CRC Press: Lanza, Claudia. 2022. *Semantic control for the cybersecurity domain: investigation on the representativeness of a domain-specific terminology referring to lexical variation*. CRC Press.

2025, The Author(s). This is an open access article, free of all copyright, that anyone can freely read, download, copy, distribute, print, search, or link to the full texts or use them for any other lawful purpose. This article is made available under a Creative Commons Attribution Share-Alike 4.0 International License. AIDAinformazioni is a journal of the Department of Culture, Education and Society, University of Calabria, Rende (CS), Italy, published by Cacucci Editore, Bari, Italy.

che per la determinazione della soglia di rappresentatività di uno strumento di controllo semantico. Queste analisi permetteranno di valutare la qualità delle risorse semantiche considerate come sistemi per guidare verso la comprensione del dominio specialistico (Albertoni et. al. 2016). Per quanto riguarda la prima tipologia di valutazione per la misurazione della rappresentatività semantica verranno presi in considerazione indici quali il grado di specificità dei termini, l'aderenza al dominio di studio, e la competenza degli esperti del campo di studi oggetto d'analisi. Il secondo gruppo di tecniche riguarda, invece, il rilevamento della rappresentatività in rapporto alla copertura semantica rispetto al corpus di partenza attraverso tecniche di mappatura con i sistemi di disseminazione informativa del settore, i.e., i *gold standard*, misure di frequenza, di richiamo e precisione, Type/token ratio e hapax, e infine la variabilità semantica.

2. Strutturare la conoscenza specialistica

I sistemi organizzazione della conoscenza specialistica hanno come presupposto fondante quello di consentire una gestione ottimale e strutturata di un determinato dominio di studio per finalità di recupero delle informazioni, fungendo da strumenti di mediazione tra i bisogni informativi espressi da un utente e la base di conoscenza del settore di riferimento (Hjørland 2008). La maggior parte degli studi presenti in letteratura che hanno descritto e approfondito le caratteristiche dei KOSs pone l'attenzione sulla loro funzione di esplicitazione delle informazioni dei settori specialistici in forma strutturata, in riferimento all'organizzazione della conoscenza in linearità con il dominio a cui quest'ultima appartiene (per citarne alcuni Hjørland 2003;2008;2016; Smiraglia 2014; Giunchiglia et al. 2014; Pieterse e Kourie 2014; Zeng 2008; Mazzocchi 2018; Souza et al. 2012; Alavi e Leidner 2001; Soergel 2009). La specificità terminologica è un elemento che caratterizza comunemente le analisi dei lavori presenti sui sistemi di organizzazione della conoscenza in quanto tale caratteristica conferisce aderenza al dominio. Un ulteriore tema, centrale nei lavori dei suddetti autori, è relativo alla diversificazione degli strumenti di controllo semantico che possono essere sviluppati per strutturare la conoscenza di un dato dominio di studio. Sigel (2016) passa in rassegna, così come altri contributi in letteratura, (tra cui Amirhosseini 2021; Souza et al. 2012; Zeng 2008), le diverse tipologie di KOSs partendo dalle liste di autorità, fino ad arrivare ai thesauri e alle ontologie come forme di massima espandibilità conoscitiva semantica per guidare l'utente verso un corretto recupero dell'informazione specialistica. Il processo di costruzione di una risorsa semantica parte generalmente da esigenze specifiche di determinati gruppi di utenti, per cui è necessario fornire un mezzo attraverso cui questi ultimi possano identificare il quadro informativo aggiornato del dominio. Pavel e Nolet (2002) evidenzia-

va come il principale punto d'analisi nell'ambito della ricerca terminologica dovesse essere orientato allo sviluppo di un modello di raggruppamento di termini che possano riflettere i concetti di domini strutturati della conoscenza specialistica, intesa come veicolo di diffusione dell'informazione tecnica verso determinati utenti.

2.1. Compilazione del corpus di partenza specialistico

La costruzione delle risorse di organizzazione e rappresentazione della conoscenza prevede una fase preliminare di costruzione del corpus di partenza specialistico (Broughton 2006), connotato dai concetti chiave dei settori oggetto di studio selezionati sulla base di un'accurata estrazione all'interno della documentazione di riferimento.

Questo primo passaggio implica sia una fase di trasformazione dei concetti chiave in dati strutturati, termini candidati per far parte delle risorse di organizzazione della conoscenza, che di validazione da parte di esperti del campo di studi oggetto d'analisi di determinati asset terminologici nella prospettiva di garantire una copertura semantica rispetto al dominio di riferimento.

Più specificatamente, la compilazione del corpus di partenza prevede una fase inziale di raggruppamento della documentazione del settore oggetto d'analisi, e da questo insieme l'avvio delle successive fasi approfondite da ampi studi presenti nella letteratura (tra cui Kageura e Marshman 2019; Vivaldi e Rodríguez 2007; Pazienza et al. 2005; Jacquemin et al. 2005; Peñas et al. 2001) che coinvolgono l'estrazione dei termini di dominio per essere inclusi nel sistema di organizzazione e gestione della conoscenza prescelto e la validazione dell'estrazione terminologica per garantire la copertura semantica rispetto al dominio di riferimento.

La compilazione di un corpus documentale di partenza è un passaggio fondamentale per consentire la realizzazione di strumenti di organizzazione e rappresentazione della conoscenza al fine di classificare e sistematizzare le informazioni tecniche interne.

2.2. La costruzione di un sistema terminologico

La costruzione di un sistema terminologico efficiente implica solitamente l'acquisizione di informazioni orientate al dominio a partire dai testi del corpus, in particolare quelli in grado di fornire densità conoscitiva semantica e granularità sul lessico che si intende rappresentare (Barrière 2006). Queste strutture sono conosciute in letteratura come Basi di Conoscenza Terminologica (Terminological Knowledge Bases - TKB) (Condamines 2018). Nel dettaglio, queste ultime supportano i processi di strutturazione delle conoscenze

specialistiche fondendo le competenze proprie della linguistica computazionale. Lo scopo di creare risorse terminologiche solide, attraverso una selezione di termini candidati rappresentativi, è legato al tentativo di fornire un'aderenza alle conoscenze specialistiche incluse nel corpus di fonti consultate (Bowker e Pearson 2002).

I documenti selezionati per far parte del corpus di partenza, destinato a fornire un dataset informativo in input per le attività di pre/post-elaborazione linguistica, sono scelti nella prospettiva di recupero delle informazioni riferite a domini specifici e popolamento terminologico nelle risorse semantiche. Il presupposto da cui muove la raccolta dei documenti di un corpus di partenza è fondato sul principio di autorevolezza delle fonti al fine di portare a termine il processo di costruzione di uno strumento semantico secondo un principio di affidabilità nelle informazioni elaborate e condivise. Infatti, la scelta dei documenti che rientrano all'interno di un corpus specialistico è caratterizzata da principi relativi (i) all'affidabilità degli editori, (ii) al rispecchiamento delle tematiche del dominio, (iii) all'osservazione di un preciso arco temporale, (iv) o allo status ufficiale dei documenti riconosciuto da esperti del settore.

Nello sviluppo di tali strumenti di organizzazione e gestione della conoscenza per determinati domini di studio specialistici, le unità terminologiche generalmente provengono dai risultati di un'estrazione automatizzata dai documenti di partenza I termini inclusi nelle risorse semantiche devono mantenere un valore univoco, come affermato nello standard *NISO TR-06-2017, Issues in Vocabulary Management* in riferimento alla definizione dei thesauri: «Vocabolario controllato: un elenco di termini che sono stati enumerati esplicitamente. Questo elenco è controllato ed è disponibile presso un'autorità di registrazione del vocabolario controllata. Tutti i termini di un vocabolario controllato devono avere una definizione univoca e non ridondante».

In questo contributo verrà presentata l'analisi sulla rappresentatività di una risorsa tesaurale, poi reingenirizzata in ontologia (Damerchiloo e Beheshti 2022; Kless et al. 2012) sul dominio della Cybersecurity (Lanza 2022; Lanza e Daille 2019)[1]. Il thesaurus contiene attualmente 243 termini di dominio strutturati secondo le relazioni di gerarchia, associazione ed equivalenza seguendo le procedure di costruzione dei rapporti semantici stabiliti nella norma ISO 25964 del 2011 e del 2013.

[1] Si vedano i lavori pubblicati per il progetto dell'Osservatorio di Cybersecurity (OCS) dell'Istituto di Informatica e Telematica (IIT) – Consiglio Nazionale delle Ricerche (CNR): (Lanza 2022; Lanza e Daille 2019).

3. Misure qualitative della rappresentatività delle risorse di organizzazione e gestione della conoscenza

3.1. Specificità e univocità terminologica

Condamines e Picton (2022) hanno approfondito dettagliatamente i presupposti di base della cosiddetta *terminologia testuale* che pone come prospettiva quella di accedere al corpus tramite i termini specialistici del dominio oggetto di studi. Per concretizzare questa forma di accesso gli autori propongono tre passaggi da seguire: (i) la redazione di un elenco dei termini candidati, studiando il significato in base alla distribuzione dei termini nel corpus (secondo un approccio di tipo bottom-up); (ii) l'analisi dei contesti in cui compaiono i termini, ricercando quelli in cui sono presenti elementi informativi rilevanti per lo studio (secondo un approccio di tipo top-down); e (iii) la convalida da parte di esperti del settore.

Il primo passaggio implica la categorizzazione dei contesti dei termini e la loro successiva interpretazione e prende le basi dalla sociolinguistica (Firth 1957) e dalla visione matematica del linguaggio (Harris 1968). L'idea principale alla base di questo approccio è che esista una correlazione tra similarità distributiva e similarità di significato. È quindi possibile comprendere il significato di una parola o di un gruppo di parole considerando tutti i suoi contesti. All'interno del secondo passaggio, invece, vengono definiti e specificati gli indicatori di relazione concettuale o elementi lessico-sintattici a cui è associata un'interpretazione preliminare, ovverosia gli indicatori associati a contesti ricchi di conoscenza hanno meno probabilità di variare da un corpus specializzato all'altro. Per ogni analisi, gli elementi direttamente correlati all'obiettivo dello studio possono essere definiti a priori, ovvero la presenza di elementi che indicano uno spostamento di significato, una variazione, ecc.

Pertanto, la specificità del dominio dovrebbe essere considerata una delle caratteristiche principali da tenere in considerazione per calcolare la soglia di copertura degli strumenti semantici e la variazione della terminologia nel tempo.

Sulla stessa linea del contributo precedentemente menzionato, Pulizzotto et al. (2022) nel framework dell'analisi concettuale prendono in considerazione il concetto lessicale. Nel contributo degli autori si fa specifico riferimento al campo della *Conceptual analysis (CA)* per cui il significato di una parola veicola sempre un concetto e un concetto ha sempre la sua espressione lessicale, la forma lessicale standard di un concetto. Non sempre, però, seguendo le riflessioni degli autori, i concetti hanno una chiara corrispondenza nelle lingue naturali e possono essere espressi anche da un gruppo di parole. La visione offerta dagli autori presuppone la considerazione della terminologia basata sulla produzio-

ne del testo in stretto rapporto con il suo contesto, il corpus documentale di partenza. I testi vengono raccolti per formare un corpus perché appartengono a generi pertinenti allo scopo dello studio e l'interpretazione dei dati linguistici deve tenere in considerazione la natura dei testi e l'oggetto dello studio.

La maggior parte delle considerazioni esposte nell'ambito della terminologia testuale si concentra sull'analisi dei corpora di partenza e sul confronto tra sub-corpora, dove le differenze di distribuzione sono associate a differenze ravvisate nelle situazioni comunicative che precedono la costruzione del sub-corpus. Nello specifico, nei casi in cui vi è la necessità di identificare l'utente finale, un gruppo di utenti esperti o non esperti del settore per i quali si procede alla realizzazione di sistemi di rappresentazione semantica, possono emergere differenze nella distribuzione dei termini, ad esempio quando vi è un utilizzo di alcuni di essi nel linguaggio generale e può essere necessario avviare un processo di normalizzazione terminologica per rappresentare univocamente il concetto veicolato. Tutto ciò viene osservato nella prospettiva di garantire una rappresentatività adeguata alle risorse semantiche proprie di un determinato campo di studi specialistico che è connotato da documenti che presentano lessici settoriali.

Condamines (2018) sottolinea l'importanza per i TKB di essere allineati alla specificità del corpus, nel modo in cui dovrebbero aderire alla natura inequivocabile dei domini specializzati e ai contesti applicativi in cui sono destinati ad essere utilizzati (Dancette e L'Homme 2004).

L'accuratezza e l'appropriatezza dei termini rispetto alla conoscenza di un dominio specialistico sono temi ampiamente approfonditi negli studi di Cabré et al.(2001) quando esplora i molteplici strumenti di estrazione terminologica. L'autrice fa specifico riferimento all'applicazione pratica delle risorse semantiche, siano esse appartenenti ad ambiti generali di studio o specialistici. Ciò che viene estratto è infatti un elenco di termini singoli o composti, che devono presentare un elevato livello di unità, concetto introdotto da Kageura e Umino (1996) che indica il grado di forza delle collocazioni sintagmatiche.

3.2. Copertura semantica rispetto alla documentazione del dominio di studi

Nel rapporto che si instaura tra la specificità e univocità e i termini di un determinato settore di studio gioca un ruolo fondamentale l'analisi del dominio di studi specialistico. Per Santini (2018) nell'analisi della qualità dei corpora nell'ambito del web possono essere individuate come dimensioni della variazione linguistica variabili quali dominio, genere, registro, stile, mezzo. Seguendo le sue affermazioni il dominio è il "campo tematico", "area" in cui viene utilizzato il documento.

Kageura e Umino (1996) esplorano il concetto di univocità in termini di connessione lessicale stabile che si instaura rispetto ai concetti specifici del dominio. Il grado di specificità può essere calcolato confrontando il corpus specializzato con uno generico, ciò può evidenziare il peso dei termini in contesti diversi e aiutare la selezione di quelli più tecnici per il dominio ai fini di garantire una maggiore rappresentatività rispetto al campo della conoscenza oggetto d'interesse. Nella riflessione portata avanti da Faber e Rodriguez (2012) la specificità terminologica fa parte di domini settoriali in contesti in cui la conoscenza specializzata è permeata da una presentazione di concetti che hanno la necessità di acquisire una forma standardizzata terminologica per la loro designazione univoca, diventando così punti di accesso a strutture informative più scientifiche e tecniche. Per ottenere questa forma di sistematizzazione della conoscenza gli autori evidenziano l'importanza di affidarsi a risorse documentali aderenti al grado di settorialità del dominio, cioè che possano allineare il lessico del settore con le attività di normalizzazione terminologica e stabilire una una connessione interlinguistica con le strutture della conoscenza. Un recente studio che può essere incluso nelle riflessioni sul rapporto tra univocità terminologica e aderenza al dominio di studi è quello di Barité e Rauch (2022) che presentano un approfondimento agli studi sulla terminologia riferendosi all'analisi del dominio, nella considerazione macrostrutturale della mappatura dei termini su un dominio di studi.

La specificità terminologica in rapporto al dominio di studi è in larga parte osservata da studi che analizzano la documentazione da cui i termini provengono, il corpus di partenza, quale « Raccolta di testi (scritti, orali o multimediali) o parti di essi in numero finito in formato elettronico trattati in modo uniforme (ossia tokenizzati ed addizionati di markup adeguato) così da essere gestibili ed interrogabili informaticamente; se (come spesso) le finalità sono linguistiche (descrizione di lingue naturali o loro varietà), i testi sono perlopiù scelti in modo da essere autentici e rappresentativi.» (Barbera et al. 2007, 26). Nello specifico, il corpus documentale creato per la realizzazione di KOSs, e composto in massima parte da testi specialistici, deve seguire delle fasi di compilazione che prendano in considerazione l'allineamento della terminologia utilizzata nella comunità di riferimento circoscritta a un periodo di tempo delimitato a seconda delle esigenze degli utenti finali (Pearson 1998).

Sulla base dell'aderenza terminologica al dominio di studi e all'univocità che i termini riescono a mostrare in linearità con la documentazione di partenza, diversi studi in letteratura hanno posto l'attenzione ai principi di rappresentatività del corpus documentale rispetto all'ambito della conoscenza da strutturare. Sinclair (2005), Leech (1991), Biber (1993) condividono l'idea che il valore di un corpus possa essere attribuito alla sua proporzione numerica dei testi posizionando il processo di compilazione del sistema documentale di partenza in un determinato arco temporale che possa delimitare l'osservazione

della variabilità semantica all'interno della documentazione collazionata. Studi sulla rappresentatività dei corpora, tra cui Biber (1993), Condamines (2007), Hunston (2002) suggeriscono che la copertura dei corpora possa essere correlata al contesto applicativo al quale le risorse semantiche saranno applicate in modo efficace.

Per Indurkhya (2010) il numero di campioni nelle diverse categorie di testo dovrebbe essere proporzionale alle loro frequenze e/o pesi nella popolazione target affinché il corpus risultante sia considerato rappresentativo, «The proportion and number of samples for each text category. The num-bers of samples across text categories should be proportional to their fre-quencies and/or weights in the target population in order for the resultingcorpus to be considered as representative» (Indurkhya 2010, 153). Per Condamines e Picton (2022) la rappresentatività del corpus è correlata all'obiettivo dello studio, pertanto la tipologia di testi attraverso cui i termini vengono diffusi gioca un ruolo importante. Occorre inoltre prestare attenzione alla presenza di sottocorpora e ai livelli di difficoltà di accesso ai dati. Per l'autrice la rappresentatività del corpus è legata alla nozione di dominio e sembra più facile da raggiungere in corpora specializzati, ma la conoscenza non è organizzata a priori in domini in cui si presume che discorso e terminologia siano strutturati; la conoscenza, infatti, non è naturalmente discontinua (Rogers 2013).

Hunston (2002) suggerisce l'ipotesi di considerare il corpus come un riferimento statico o dinamico. Nel primo caso possiamo riferirci a un corpus come a un campione di documenti che possono rappresentare un dominio, nel secondo la compilazione del corpus si riferisce al monitoraggio del cambiamento semantico. Come chiarito nei lavori pubblicati dalla sezione dedicata allo studio dei corpora dell'Università di Lancaster (McEnery e Hardie 2012), i modelli statici possono presentare diverse opportunità per supportare un'analisi molto efficiente della variazione linguistica interna ai documenti utilizzati, trasformandoli in frame di campionamento, che a loro volta divengono strumenti essenziali per comprendere la granularità di un corpus specialistico: i corpora basati su una documentazione non specialistica tendono a calcolare la loro rappresentatività sulla base dell'incorporazione di diversi tipi di testi, mentre per quanto riguarda quelli specializzati, l'attenzione dovrebbe essere mantenuta sull'intervallo di "chiusura" o "saturazione" del corpus.

Barbera et al. (2007, 50) chiariscono che quando si analizza la rappresentatività delle risorse semantiche si pone l'attenzione alla nozione di "bilanciamento":

> Una discussione particolarmente efficace, che mette in luce i vari problemi della nozione, è data da Tognini-Bonelli 2001, p. 55-57, che si pone direttamente il problema di come e cosa campionare ("sampling") per conseguire tale rappresentatività. Alla rappresentatività è strettamente legato anche il dibattito sulle dimensioni di un corpus (più il corpus è ampio, più lo si potrebbe rite-

nere rappresentativo di una lingua. E se rappresentatività è bilanciamento, si tratta anche, per dirla alla Engwall (1994), di una questione di "choice e non di chance", in cui la fase di progettazione e finalizzazione non è dunque secondaria: l'ipotesi della scelta casuale implica infatti che la varietà dei testi, qualunque sia la sua estensione, non sarà rappresentativa dell'intera popolazione, né di uno specifico gruppo; la selezione programmata ed organizzata dei materiali, invece, deve aderire alla definizione della lingua che il corpus si propone di rappresentare ed i confini entro i quali le scelte avvengono devono essere ben tracciati. Rappresentatività, finalizzazione e dimension sono pertanto parametri interrelati: è infatti chiaro però che l'attenzione per la rappresentatività di un corpus significa attenzione per la funzione cui il corpus stesso è destinato; ed ai fini di ricerche specifiche, una raccolta di dimensioni limitate ma in sé equilibrata, ben etichettata ed accuratamente verificata, comprendente in sè la dimensione linguistica che interessa studiare, risulterà più efficace di una raccolta ampia ma non proporzionata al suo interno.

4. Misure quantitative della rappresentatività delle risorse di organizzazione e gestione della conoscenza

Nella seguente sezione verranno presentate le principali tecniche di misurazione adottate per l'attività di ricerca eseguita sul domino di studi della sicurezza informatica[2].

4.1. Frequenza terminologica

Facendo riferimento ai lavori di Rapp (2014) e Francom (2010) la rappresentatività garantita dalle risorse semantiche viene calcolata dalla frequenza delle parole in un corpus, dalla loro co-occorrenza e dalle parole comuni nei testi. «A corpus whose word frequencies are highly correlated to the familiarity norms is more likely to be a good surrogate for everyday language» (Rapp, 2014, 2118).

La determinazione dei valori di precisione dei concetti rappresentati dai termini in una risorsa semantica può essere calcolata attraverso diverse tecniche tra cui la misura della frequenza, che indica la rappresentatività di un termine rispetto a un concetto specializzato di un dominio.

Questa misurazione statistica assistita a livello computazionale si ottiene mettendo in relazione il rapporto tra due liste terminologiche (una specifica per il dominio e una per la documentazione generale), ottenute calcolando le frequenze relative per ciascuna di esse. I punteggi di diversi termini conte-

[2] Il corpus creato per la realizzazione del thesaurus sulla Cybersecurity è composto da 560 documenti che coprono un arco temporale dal 2002 al 2017 e include sia documentazione legislativa che divulgativa (riviste di settore specialistiche) in lingua italiana, così come report e documentazione tecnici.

nuti nella lista ottenuta dalla fase di estrazione terminologica[3] possono essere messi a confronto con un elenco più generico di termini estratti da un corpus linguistico generale per evidenziare come alcuni di essi siano caratterizzati da un livello di specificità più elevato. Questa comparazione ha permesso di selezionare i termini ad altri come termini di ingresso nel thesaurus. I risultati hanno dimostrato che alcuni dei termini candidati rappresentativi, selezionati in base alla validazione dei gruppi di esperti, presentavano un punteggio più elevato nel calcolo della specificità, come ad esempio *spam* o *malware*, mentre altri venivano recuperati molto frequentemente in contesti generici e influenzavano da un punto di vista terminologico alcune scelte terminologiche da fare nelle fasi di validazione verso l'inclusione in modo manuale, come nel caso, ad esempio, di *virus*, *sicurezza*, *accesso*, *protezione*.

4.2. Richiamo e Precisione

Le metriche di Richiamo e Precisione rilevano la specificità dei concetti come valore discriminante per garantire la copertura di uno strumento semantico rispetto a determinati documenti di partenza. Nell'analisi portata avanti sul dominio della sicurezza informatica il risultato è stato ottenuto stabilendo un confronto con corpora esistenti sul settore di studio, ovvero i *gold standard* (Shadbolt et al. 1999) attraverso una procedura di corrispondenza nell'allineare termini rappresentativi scelti da quelli estratti con le unità semantiche ufficialmente accettate dai gruppi di esperti. Nel dettaglio, sono state messe a confronto le liste di termini estratti con i software impiegati per questa attività (T2K e Termsuite) con quelle estratte dai principali asset terminologici presenti sul dominio della Cybersecurity. In questo modo è stato possibile identificare il livello della copertura che le fonti selezionate per compilare il corpus di partenza e, di conseguenza, la risorsa semantica possono garantire rispetto alle informazioni condivise all'interno di comunità di esperti. È anche possibile stabilire così la presenza di discrepanze e procedere, seguendo la supervisione degli utenti esperti del dominio di studi, con una diversa sistematizzazione delle unità terminologiche all'interno delle risorse semantiche al fine di garantire una forma di rappresentatività sempre più alta.

4.3. Type/token Ratio e Hapaxes

Tra le altre misurazioni statistiche a supporto della definizione di un modello di rappresentatività semantica rientrano quelle incentrate sul calcolo della

[3] L'estrazione terminologica da cui è stata avviata la fase di scelta dei termini candidati a far parte della risorsa semantica è stata eseguita con l'ausilio di software automatici Text2Knowledge (Dell'Orletta et al. 2014) e Termsuite (Daille 2016).

ricchezza lessicale e della diversità dei vocabolari controllati in determinati corpora, cioè il rapporto *type/token ratio* (Kettunen 2014), e quelle riferite all'individuazione della proporzione delle parole che ricorrono molto raramente e una sola volta in un corpus di partenza, e che possono evidenziare quanto possa essere valutato specifico un elenco di termini candidati da includere in una risorsa semantica, gli *hapax*. Il risultato ottenuto dal rapporto di tutte le parole presenti in un corpus (token, unità minime in cui è diviso il testo (Cfr. Piscopiello e Beraccini 2009) e quelle uniche (type) indica

> in valori percentuali, il numero di parole diverse all'interno di un determinato testo rispetto al numero di parole complessive dello stesso. A titolo d'esempio, affermare che, a parità di tokens, TTR1 > TTR2 (dove TTR1 e TTR2 rappresentano rispettivamente la Type/Token Ratio di un ipotetico testo 1 e di un ipotetico testo 2) equivale a dire che il numero di parole diverse del testo 1 è superiore rispetto a quello del testo 2. Pertanto, applicando tale parametro ai concetti di densità e ricchezza terminologica, si potrebbe affermare che la TTR è direttamente proporzionale al carattere divulgativo di un testo: ciò vuol dire che quanto più elevato è il valore della TTR (e, quindi, quanto più grande è il numero di types rispetto ai tokens), tanto più il testo sarà ricco a livello lessicale/terminologico e, di conseguenza, tante più probabilità si avranno che lo stesso abbia un carattere divulgativo piuttosto che scientifico. In definitiva, tale ricchezza si traduce generalmente in un numero significativo di varianti e sinonimi, condizione questa – come si è visto – tipica dei testi divulgativi", la formula che invece fornisce l'elenco degli hapax può essere considerata una misura efficiente per calcolare il grado di specificità e accuratezza terminologica in rapporto ai documenti di partenza (Piscopiello e Bertaccini 2009, 5).

Per quanto riguarda il corpus costruito per la Cybersecurity, quasi il 70% delle parole ricorre una volta, le più frequenti sono, ad esempio: *rischi, controllo, tecnologie, Internet, dispositivi,* e nelle prime centinaia è incluso anche il termine *Cybersecurity*. Per quanto riguarda le parole uniche recuperate con il calcolo 1 e 2 gram, le più frequenti sono - ad eccezione delle stopword che in questa fase vengono fornite in output e poi saranno rimosse utilizzando tecniche di estrazione dei termini più sofisticate, rispettivamente: 1-gram: *sicurezza, servizi, protezione, controllo, software*; 2-grams: *Parlamento europeo, decreto legislativo, dati personali, accesso, comunicazioni elettroniche*.

4.4. Variabilità semantica

L'evoluzione linguistica che interessa i diversi campi del sapere specialistico va di pari passo con gli aggiornamenti dal punto di vista terminologico, che, caratterizzati da fattori di continuità, puntano a rappresentare in modo onnicomprensivo lo spettro informativo di un determinato dominio. Come affermato da L'Homme (2015) i termini sono strettamente dipendenti dai conte-

sti in cui si suppone essi siano impiegati. Le risorse semantiche contenenti i termini rappresentativi che riflettono il quadro concettuale di specifici ambiti di studio sono destinate ad essere aggiornate nella prospettiva di seguirne la variabilità semantica e, di conseguenza, agevolare il miglioramento della specificità e dell'univocità terminologica nella rappresentazione della conoscenza specialistica (Cabré 2009).

La variabilità dei termini in determinati strumenti di controllo semantico (Daille 2005) è una caratteristica essenziale da tenere in conto per lo sviluppo di risorse volte a rappresentare in forma standardizzata uno specifico campo della conoscenza. L'aggiornamento rispetto ai cambiamenti presenti nei contenuti pubblicati da una serie di fonti prestabilite negli ambiti specialistici d'indagine può garantire una versione attuale, e di conseguenza più affidabile, di uno strumento semantico se collegato a flussi di nuove informazioni. La misurazione che può essere effettuata sulla lista dei termini inclusi all'interno di una risorsa semantica per garantirne la sua rappresentatività rispetto al dominio d'interesse specialistico prevede una mappatura temporale con i sistemi scientificamente utilizzati per la disseminazione della conoscenza settoriale.

5. Conclusioni

La prospettiva di questo contributo è di fornire un approfondimento teorico sui principi di accuratezza e appropriatezza dei termini rispetto alla conoscenza di un dominio specialistico, tema ampiamente approfondito negli studi di Cabré et al. (2001) quando esplora i molteplici strumenti di estrazione terminologica. Le risorse semantiche di organizzazione della conoscenza specialistica contengono al loro interno un elenco di termini singoli o composti in forma strutturata che devono presentare un elevato livello di *unità*, concetto introdotto da Kageura e Umino (1996). Per garantire una forma di copertura semantica di tali risorse rispetto al campo del sapere che queste ultime sistematizzano è stata presentata una panoramica di alcune tecniche adoperabili in differenti contesti applicativi. Le valutazioni qualitative relative alla specificità e univocità terminologica si sommano alle misurazioni statistiche sulla copertura semantica rispetto all'informazione contenuta nei *gold standard* del dominio di studi in base a criteri di frequenza terminologica e ricchezza lessicale. La possibilità di disporre delle competenze nel dominio di studio, caratteristica spesso resa possibile dalla presenza della figura del gruppo di esperti del dominio (Schultz et al.1968), rappresenta un ulteriore supporto essenziale per le fasi di validazione verso l'inclusione dei termini candidati tali da garantirne una rappresentatività aderente ai contesti d'uso. In ultima istanza, risulta fondamentale porre l'attenzione sui processi di aggiornamento dei KOSs nell'asset terminologico che offrono prendendo in considerazione il principio della

variabilità semantica per poter assicurare una forma di allineamento accurato rispetto alle variazioni avvenute nello spettro conoscitivo del settore di studio.

Riferimenti bibliografici

Alavi, Maryam, and Dorothy E Leidner. 2001. "Knowledge management and knowledge management systems: Conceptual foundations and research issues." *MIS quarterly*, 25:107-36. https://doi.org/10.2307/3250961.

Albertoni, Riccardo, Monica De Martino, and Alfonso Quarati. 2016. "Integrated Quality Assessment of Linked Thesauri for the Environment." In *Electronic Government and the Information Systems Perspective. EGOVIS 2016*, edited by A. Kö and E. Francesconi, *Lecture Notes in Computer Science* 9831:221-35. Springer, Cham. https://doi.org/10.1007/978-3-319-44159-7_16.

Amirhosseini, Maziar. 2021. "A Dialectic Perspective on the Evolution of Thesauri and Ontologies." *Knowledge organization* 48(6): 403-29. https://doi.org/10.5771/0943-7444-2021-6-403.

Barbera, Manuel, Elisa Corino, e Cristina Onesti. 2007. *Corpora e linguistica in rete.* Guerra Edizioni.

Barité, Mario, and Mirtha Rauch. 2022. "Terminological studies as domain analysis: a critical exploration." *Brazilian Journal of Information Science: Research trends* 16. https://doi.org/10.36311/1981-1640.2022.v16.e02140.

Barrière, Caroline. 2006. "Semi-automatic corpus construction from informative texts." In *Lexicography, Terminology and Translation, Text-Based Studies in honour of Ingrid Meyer* , edited by Lynne Bowker. University of Ottawa Press. https://doi.org/10.2307/j.ctt1ckpgs3.9.

Biber, Douglas. 1993. "Representativeness in corpus design." *Literary and linguistic computing* 8 (4): 243-57. https://doi.org/10.1093/llc/8.4.243.

Bowker, Lynne, and Jennifer Pearson. 2002. *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge.

Broughton, Vanda. 2006. *Essential thesaurus construction*. Facet publishing.

Cabré, María Teresa. 2009. "La Teoría Comunicativa de la Terminología, una aproximación lingüística a los términos." *Revue Francaise De Linguistique Appliquee* 14 (2009): 9-15.

Cabrè, Maria Teresa, Rosa Estopà, and Jordi Vivaldi Palatresi. 2001. "Automatic term detection: A review of current systems."*In Natural Language Processing 2 - Recent Advances in Computational Terminology*, edited by Didier Bourigault, Christian Jacquemin and Marie-Claude L'Homme. John Benjamins Publishing Company.

Condamines, Anne and Aurélie Picton. 2022. "Textual Terminology: Origins, principles and new challenges." In *Theoretical Perspectives on Terminology*, John Benjamins Publishing Company.

Condamines, Anne. 2007. "L'interprétation en sémantique de corpus : le cas de la construction de terminologies. " *Revue Française de Linguistique Appliquée* XII (1):39-52. https://doi.org/10.3917/rfla.121.0039.

Condamines, Anne. 2018. "Terminological knowledge bases From Texts to Terms, from Terms to Texts." In *The Routledge Handbook of Lexicography*, edited by Pedro A. Fuertes-Olivera. Routledge.

Cram, Damien, and Béatrice Daille. 2016. "Terminology extraction with term variant detection." In *Proceedings of ACL-2016 System Demonstrations, August 2016 Berlin, Germany*, edited by Sameer Pradhan and Marianna Apidianaki. Association for Computational Linguistics. Kerrville (Texas), 13-18.

Daille, Béatrice. 2005 "Variations and application-oriented terminology engineering." *Terminology International Journal of Theoretical and Applied Issues in Specialized Communication*, 11(1): 181-97. https://doi.org/10.1075/term.11.1.08dai.

Damerchiloo, Mansoureh, and M. Hosseini Beheshti. 2022 "Converting thesaurus to ontology (a systematic review)." *Library and Information Science Research* 11 (2):105-27.

Dancette, Jeanne Eugenie, and Marie-Claude L'Homme. 2004. "Building specialized dictionaries using lexical functions." *Linguistica Antverpiensia*.

Dell'Orletta, Felice, Giulia Venturi, Andrea Cimino, and Simonetta Montemagni. 2014. "T2K^2: a system for automatically extracting and organizing knowledge from texts." In*Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, edited by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk and Stelios Piperidis, Reykjavik, Iceland, 2014, 2062-70.

Faber, Pamela, and Clara Inés López Rodrıguez. 2012. "2.1 terminology and specialized language." *A cognitive linguistics view of terminology and specialized language* 20.9.

Firth, John Rupert.1957. *Applications of General Linguistics*. https://doi.org/10.1111/J.1467-968X.1957.TB00568.X.

Francom, Jerid, Amy LaCross, and Adam Ussishkin. 2010. "How Specialized are Specialized Corpora? Behavioral Evaluation of Corpus Representativeness for Maltese." In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta,* edited by Nicoletta Calzolari, Khalid Choukr, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner and Daniel Tapias European Language Resources Association (ELRA).

Giunchiglia, Fausto, Biswanath Dutta, and Vincenzo Maltese 2014. "From knowledge organization to knowledge representation." *Knowledge Organization* 41 (1):44-56. https://doi.org/10.5771/0943-7444-2014-1-44.

Harris, Zellig. 1968. *Mathematical Structures of Language*. *Interscience Tracts in Pure and Applied Mathematics* 21. New Yor: Interscience Publishers.

Hjørland, Birger. 2003. "Fundamentals of knowledge organization." *Knowledge organization* 30 (2):87-111. https://doi.org/10.5771/0943-7444-2003-2-87.

Hjørland, Birger. 2008. "What is knowledge organization (KO)?" *Knowledge Organization*, 35(2-3): 86-101. https://doi.org/10.5771/0943-7444-2008-2-3-86.

Hjørland, Birger. 2016. "Knowledge organization (KO)." *Knowledge Organization*, 43(6):475-84. https://doi.org/10.5771/0943-7444-2016-6-475.

Hunston, Susan. 2002. *Corpora in Applied Linguistics.* Cambridge University Press, https://doi.org/10.1017/CBO9781139524773.

Kageura, Kyo and Bin Umino. 1996. "Methods of automatic term recognition: A review." *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 3(2):259-89.https://doi.org/10.1075/term.3.2.03kag.

Kageura, Kyo, and Elizabeth Marshman. 2019. "Terminology extraction and management." In *The Routledge handbook of translation and technology*, edited by Minako O'Hagan. Routledge. 61-77.

Kettunen, Kimmo. 2014. "Can Type-Token Ratio Be Used to Show Morphological Complexity of Languages?" *Journal of Quantitative Linguistics*, 21(3): 223–45. https://doi.org/i:10.1080/09296174.2014.911506.

Kless, Daniel, Ludger Jansen, Jutta Lindenthal, and Jens Wiebensohn. 2012. "A method for re-engineering a thesaurus into an ontology". In *Formal Ontology in Information Systems. Proceedings of the Seventh International Conference (FOIS 2012)*, IOS Press.133-46.

Jacquemin, Christian, and Didier Bourigault. 2005. "Term extraction and automatic indexing." In *The Oxford Handbook of Computational Linguistics*, edited by Ruslan Mitkov. Oxford University Press. 599-615.

Indurkhya, Nitin, and Fred J. Damerau (Eds.). 2010. *Handbook of Natural Language Processing* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781420085938.

ISO (International Organization for Standardization). 2011. *Information and documentation -Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval*. ISO 25964-1:2011.

ISO (International Organization for Standardization). 2013. *Information and documentation – Thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies*. ISO 25964- 2:2013.

Lanza, Claudia. 2022. *Semantic control for the cybersecurity domain: investigation on the representativeness of a domain-specific terminology referring to lexical variation*. CRC Press.

Lanza, Claudia, and Béatrice Daille. 2019. "Terminology systematization for Cybersecurity domain in Italian Language." In *Actes de la Conférence sur le Traitement Automatique des Langues Naturelles (TALN) PFIA* 2019, *Terminologie et Intelligence Artificielle (atelier TALN-RECITAL\& IC) Toulouse, France*, edited by Emmanuel Morin, Sophie Rosset and Pierre Zweigenbaum. ATALA, 7-18.

Leech, Geoffrey. 1991."The state of the art in corpus linguistics." In *English Corpus Linguistics*, Routledge.

L'Homme, Marie-Claude. 2015. "Terminologies and Taxonomies." In *The Oxford Handbook of the Word*, edited by John R Taylor, Oxford University Press.

Mazzocchi, Fulvio 2018. "Knowledge organization system (KOS): an introductory critical account." *Knowledge Organization* 45(1): 54-78.

McEnery, Tony, and Andrew Hardie. 2012. "Answers to exercises: Chapter One discussion questions." In *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.

NISO (National Information Standards Organization). 2017. *TR-06-2017Issues in Vocabulary Management.*

Pavel, Silvia, et Diane Nolet. 2002. *Précis de Terminologie*. Terminologie et normalisation, Bureau de la traduction.

Pazienza, Maria Teresa, Marco Pennacchiotti, and Fabio Massimo Zanzotto. 2005. "Terminology extraction: an analysis of linguistic and statistical approaches." In *Knowledge mining: Proceedings of the NEMIS 2004 final conference*. Berlin, Springer-Verlag Berlin and Heidelberg GmbH & Co. KG., 185:255-79.

Pearson, Jennifer. 1998. *Terms in Context*. Amsterdam: John Benjamins Publishing Company.

Peñas, Anselmo, Felisa Verdejo, and Julio Gonzalo. 2001. "Corpus-based terminology extraction applied to information access." *Proceedings of corpus linguistics*.

Pieterse, Vreda, and Derrick G. Kourie. 2014. "Lists, taxonomies, lattices, thesauri and ontologies: paving a pathway through a terminological jungle." *Knowledge Organization* 41 (3): 217-29. https://doi.org/10.5771/0943-7444-2014-3-217.

Piscopiello, Maria Grazia, e Franco Bertaccini. 2009. "Densità e ricchezza terminologica: generi testuali a confronto." *mediazioni* 7. https://mediazioni.sitlec.unibo.it/index.php/no7-anno2009.html.

Pulizzotto, Davide, Jean-François Chartier, Francis Lareau, Jean-Guy Meunier, and Louis Chartrand. 2018. "Conceptual Analysis in a Computer-Assisted Framework: Mind in Peirce." *Umanistica Digitale* 2 (2). https://doi.org/10.6092/issn.2532-8816/7305.

Reinhard Rapp. 2014. "Using Collections of Human Language Intuitions to Measure Corpus Representativeness." In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2117–2128, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.

Rogers, Margaret. 2013. "What is a 'Domain' and is this a Useful Question?" *Asp, revue du GERAS* 64: 5-16. https://doi.org/10.4000/asp.3810.

Santini, Marina, Wiktor Strandqvist, Mikael Nyström, Marjan Alirezai and Arne Jönsson. 2018. "Can We Quantify Domainhood? Exploring Measures to Assess Domain-Specificity in Web Corpora." In *Database and Expert Systems Applications. DEXA 2018. Communications in Computer and Information Science*, edited by M. Elloumi et al. 903. Springer, Cham. https://doi.org/10.1007/978-3-319-99133-7_17.

Sinclair, John. 2005. "Corpus and Text - Basic Principles". In *Developing Linguistic Corpora: A Guide to Good Practice*, edited by Martin Wynne.

Schultz, Claire K. *et al*. 1968. *Evaluation of Indexing by Group Consensus. Final Report*. ED025288.

Shadbolt, Nigel, Kieron O'hara, and Louise Crow. 1999. "The experimental evaluation of knowledge acquisition techniques and methods: history, problems and new directions." *International Journal of Human-Computer Studies* 51(4):729-55, https://doi.org/10.1006/ijhc.1999.0327.

Sigel, Alexander. 2006. "From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web." in *Introducing Terminology-based Ontologiesedit* edited by Winfried Schmitz-Esser & Alexander Sigel.

Smiraglia, Richard P. 2014. *The elements of knowledge organization*. Springer Cham Heidelberg.

Soergel, Dagobert. 2009. *Knowledge organization systems: overview*. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.dsoergel.com/SoergelKOSOverview.pdf

Souza, Renato R., Douglas Tudhope, and Mauricio B. Almeida. 2012. "Towards a Taxonomy of KOS: Dimensions for Classifying Knowledge Organization Systems." *Knowledge organization* 39(3):179-92. https://doi.org/10.5771/0943-7444-2012-3-179.

Vivaldi, Jorge, and Horacio Rodríguez. 2007. "Evaluation of terms and term extraction systems: A practical approach. Terminology". *International Journal of Theoretical and Applied Issues in Specialized Communication* 13 (2): 225-48. https://doi.org/10.1075/term.13.2.06viv.

Zeng, Marcia. 2008. "Knowledge organization systems (KOS)." *Knowledge Organization* 35 (2-3):160-82. https://doi.org/10.5771/0943-7444-2008-2-3-160.

AIDAinformazioni

Rivista semestrale di Scienze dell'Informazione

Anno 43

N. 3-4 – luglio-dicembre 2025

Contributi

Giovanna Aracri, Antonietta Folino, Maria Teresa Guaglianone, Marco Manna, Identification of thesaurus relationships with the support of ChatGPT and Gemini

Alessandra Bonsignori, Terminologia giuridica e organizzazione della conoscenza

Maria Teresa Chiaravalloti, Maria Teresa Guaglianone, Elena Cardillo, Metadatazione delle tipologie di documenti clinici: un approccio alla standardizzazione mediante LOINC

Carla Beatriz Marques Felipe, Carla Maria Martellote Viola, Raimunda Fernanda dos Santos, Domain Analysis on Data Management

Claudio Gnoli, Emanuela Valenzano, Che faccetta sei?

Manuela Grillo, Strumenti di convergenza tra archivi e biblioteche: i thesauri

Bianca La Manna, Rappresentare gli atti notarili medievali in Linked Open Data (LOD)

Claudia Lanza, L'univocità e la variabilità terminologica a supporto della rappresentatività delle risorse semantiche di organizzazione e rappresentazione della conoscenza specialistica

Larissa Moraes Martins, Carlos Eduardo da Silva Carvalho, Ana Cristina de Albuquerque, Mario Barité, Possibilities for using cultural warrant to complement literary work

Bianca Therezinha Carvalho Panisset, Lucia Maria Velloso de Oliveira, Leticia Souza da Costa Sampaio, Camilla Campoi de Sobral, Identification of types of documents and knowledge organization: transparency and access in records management

SoYoung Yoon, Claudio Gnoli, Ziyoung Park, Applying Integrative Levels Classification to Korean KOS Registry

Emanuela Nicole Donato, Grazia Serratore, Maria Taverniti, Maria Teresa Chiaravalloti, L'evoluzione della struttura documentale nelle cartelle cliniche psichiatriche dell'archivio storico del manicomio di Girifalco (1881-1978)

Lorenza Talarico, Manuscripts in the Digital Age

Maria Lígia Triques, Ananda Fernanda de Jesus, Mario Barité, José Eduardo Santarem Segundo, Ana Cristina de Albuquerque, Vocabularies in the digital context

Marcin Trzmielewski, Context-Driven Use of Automatic Tools in Medical Ontology Design: A Case Study in Allergology

Carla Maria Martellote Viola, Carla Beatriz Marques Felipe, Exploring the relationships between the Cross-Cutting Women's Agenda - Multi-Year Plan 2024-2027 and the Knowledge Organization of legislative proposals under processing on women in Brazil

Elena Cencetti, Camillo Carlo Pellizzari di San Girolamo, Elisabetta Viti, Modellizzazione delle conoscenze: Thesaurus Nuovo soggettario e Wikidata, due strumenti a confronto

Pontish Yeramyan, Alexander Murzaku, Nick Uva, Curt Anderson, The transferability of brilliance

Ana Flávia Dias Zammataro, Ana Cristina de Albuquerque, Giulia Crippa, Lucia Sardo, Carlo Bianchini, Maria Lígia Triques, ISKO Italy in perspective

mundaneum

In copertina
Disegno di Paul Otlet, Collections Mundaneum, centre d'Archives, Mons (Belgique).

ISBN 979-12-5965-639-1

ISSN 1121-0095