ADVANCES IN KNOWLEDGE ORGANIZATION
Proceedings of International ISKO Conference - ISSN: 0938-5495 - 2026 - São Paulo

Indexing modelling from a frame-semantics perspective: challenges of extraction and interpretation by Artificial Intelligence

Juliana Oliveira Mauch
Federal University of Rio Grande do Sul, Brazil

Rita do Carmo Ferreira Laipelt
Federal University of Rio Grande do Sul, Brazil

Abstract

The paper discusses the terminological modelling of Information Science processes from the perspective of Frame Semantics, focusing on the term "indexing." Supported by Frame-based Terminology, the study takes Frame Semantics as a conceptual macro-structure and uses text typology to systematically extract associative relations from a corpus composed of a reference article on indexing published in Knowledge Organization. In parallel, the paper reports an experiment with a large language model (Perplexity Pro) in order to assess its ability to identify lexico-syntactic triggers and reconstruct the frame-based structure underlying the text, comparing the performance of the AI system with human analysis. The results show that the AI is particularly efficient in retrieving related terms and anchoring excerpts in the source text, although it exhibits limitations in formulating frame relations, which requires interpretive intervention by the researchers; nonetheless, the study highlights the potential of human–AI collaboration to refine frame-based terminological micro-networks for "indexing."

Keywords: Indexing; Frame semantics; Frame-based terminology; Large language models.

1.0 Introduction

The terminological modelling of processes in Information Science usually relies on genus–differentia definitions. Although useful, they tend to say little about the schematic scene structure, that is, the cognitive template that organises a typical situation into roles, objects, actions and relations. It is this structure that gives technical meaning to use: who acts, on what, with which instruments, according to which rules, to produce which results and for what purpose.

Frame Semantics (Fillmore 1982) offers the descriptive apparatus needed to make such components visible by understanding meaning as "enacted knowledge": each term activates a frame composed of Frame Elements (FEs) and interframe relations that

articulate functional chains. In the tradition of Frame-based Terminology (Faber 2012), analysis proceeds from the domain to micro-networks, specifying FEs, their valence patterns and the causal, instrumental and teleological connections that sustain informational practice.

Within this scope, Frame Semantics provides the conceptual macro-structure of the term "indexing" (FEs and prototypical scene), while text typology guides the systematic extraction of associative relations from the corpus. At the same time, the study proposes an experimental assessment of the interpretive capacity of Large Language Models (LLMs), specifically Perplexity Pro, with regard to their grasp of Frame Semantics. Rather than merely testing a passive understanding of concepts, we investigate the ability of these tools to identify the schematic structure underlying the corpus texts. Our aim is to analyse whether Artificial Intelligence (AI) can detect lexico-syntactic triggers and contextualise the application of the theory with the same level of analytical acuity required in human modelling.

Given this theoretical and methodological framework, the general objective of this paper is to model the indexing process from the perspective of Frame Semantics and Frame-based Terminology, taking a reference article in the field as corpus, and to assess the potential of a large language model, configured as a data analyst, in extracting and interpreting associative relations linked to the term "indexing."

Specifically, the study seeks to identify, in the selected article, the frame elements and associative relations that structure the concept of "indexing" within Knowledge Organization (KO); to reconstruct the associative network of "indexing" through manual frame-based analysis (mapping agents, processes, instruments, norms, products and goals); to apply an LLM (Perplexity Pro) to the same unit of analysis in order to extract related terms, textual evidence and proposed frame relations; to compare the granularity, consistency and limitations of the networks produced by human analysis and by academic AI, discussing to what extent human–AI collaboration can contribute to refining frame-based terminological micro-networks; and, finally, to reflect on the epistemological and methodological implications of this procedure for KO research and for the use of language models in indexing studies.

2.0 Indexing in Knowledge Organization

From the standpoint of domain analysis, Hjørland (2002, 2017) discusses indexing within specific discourse communities, emphasising that indexing decisions depend on the epistemological perspectives, research traditions and socio-cognitive practices of each domain. In this view, indexing is part of broader subject analysis processes, in which the indexer must consider not only "what" the document is about, but also "for whom" and "for what purposes" it will be useful, anticipating future uses and interpretations.

This emphasis on the epistemological grounding of indexing resonates directly with Dahlberg's (1978) Concept Theory, oriented towards the analytical decomposition of concepts and their structured representation. Dahlberg stresses that each concept is formed by essential characteristics and that terminological representation should systematically reflect this internal structure, ensuring coherence between the unit of knowledge and its designation.

In the context of terminological standardisation, this perspective becomes a fundamental pillar for both Terminology and KO, inasmuch as it seeks to ensure clarity, precision and interoperability in the representation of knowledge. The construction of Knowledge Organization Systems (KOS) — such as thesauri, subject heading lists and ontologies — therefore depends on making explicit the conceptual characteristics and semantic relations that will guide indexing decisions.

The increasing complexity of digital environments requires vocabulary systems that not only define isolated terms but are also capable of representing the networks of meaning that connect them. This requirement reinforces the importance of considering, beyond conceptual definitions, the semantic and epistemological relations that underpin representations in KOS.

From the perspective of Information Science, this shift has a direct impact on the work of information professionals. According to Krebs and Laipelt (2018, 81–82), KO "involves a cognitive effort on the part of the professional, to understand not only the content of the work to be indexed, but also the possible information needs of the users of the information unit." In other words, the work of indexers and librarians is not limited to identifying descriptive terms; it also requires understanding which conceptual schemes — or frames — are activated by user communities in their contexts of information production and use.

Bringing this framework closer to Mai's (2001) proposal of subject indexing as an interpretive process, Frame Semantics provides precisely the type of meta-language needed to make explicit the conceptual choices underlying indexing decisions. By modelling "indexing" as a frame that articulates agents, processes, instruments, norms and products, the study aligns itself with the KO tradition and, at the same time, introduces a cognitive-semantic apparatus that makes it possible to describe, with greater granularity (Faber 2012), the associative relations that sustain the quality of subject representation.

In this scenario, the use of LLMs offers fertile ground for deepening the understanding of indexing as an interpretive process, since such systems also need to make semantic choices when grouping terms, inferring relations and interpreting discursive contexts. By bringing the epistemology of KO closer to the internal operations of these models, it becomes possible to observe how different forms of cognitive mediation — human and algorithmic — impact the construction of subject descriptions and the configuration of conceptual networks. This dialogue between indexing professional practices and the functioning of LLMs constitutes the axis on which the experiment reported in this paper is

grounded, articulating semantic decisions, conceptual framings and uses of information in digital environments.

3.0 Frame Semantics

Frame Semantics is part of the Cognitive Linguistics movement which, since the late 1970s, has broken with the autonomy of language and has placed meaning and use back at the centre of linguistic description (Chishman 2016). Instead of treating meaning as a set of atomic features detached from practice, this approach adopts an experientialist stance: knowledge is encyclopaedic and distributed across the interaction between language and other cognitive abilities. Words do not mean in isolation; they activate frames and bring certain elements of the scene into the foreground (Fillmore 1982).

From an operational point of view, a frame is a system of inter-related concepts such that understanding one element presupposes understanding the structure that supports it (Fillmore 1982). Each frame organises Frame Elements (FEs) (Baker et al. 1998), that is, participant roles and parameters of the situation, for example: Agent, Patient/Artifact, Instrument, Standard/Norm, Process, Output, Goal, Conditions.

According to Ruppenhofer et al. (2010), the fundamental steps for constructing a frame involve collecting, in a corpus, sentences related to lexemes that are assumed to share common semantic traits. From this perspective, we propose employing an academic AI system to analyse the multiple nuances of a lexical item, drawing on empirical evidence so as to enable the establishment of a network of associated frames that is significantly broader than that normally achieved by conventional methodologies.

In practical terms, many contemporary AI systems operate with structures that are close to Minsky's (1974) frames, activating typical "scenes" with roles and parameters whenever they need to interpret an utterance or situation — as in the case of virtual assistants that recognise requests for future actions, or question-answering systems based on FrameNet and semantic role labelling, which identify who did what to whom and for what purpose in each clause. This logic of scene activation finds a direct parallel in the lexical Frame Semantics of FrameNet, which organises the lexicon into frames and frame elements, and it also resonates with Mai's (2001) conception of subject indexing as an interpretive process: when indexing, professionals bring the document into alignment with a set of conceptual "frames" that articulate agents, actions, objects and purposes, in a way that is analogous to how AI systems activate frames to interpret commands, questions or descriptions.

4.0 Methodology

This analytical-descriptive study is grounded in Frame Semantics applied to Terminology and takes the term "indexing" as its object. The corpus consists of one article published in the journal Knowledge Organization. We selected an article in which "indexing" appears as a theoretical-methodological problem for subject representation and which supports a frame-oriented reading.

The chosen article was "Folksonomy Indexing From the Assignment of Free Tags to Setup Subject: A Search Analysis into the Domain of Legal History," authored by José-Antonio Moreiro-González and Carmen Bolaños-Mejías and published in 2018. In this article, indexing is presented as an interpretive process of document analysis, in which an agent (professional or user) constructs, from their social and cognitive context, a condensed description of the subject and translates it into signs of an indexing language (controlled or free), with a view to future retrieval.

The manual analysis followed steps inspired by Frame-based Terminology, starting with the identification of the frame elements associated with "indexing" (Agent, Instrument, Process, Output, Goal, Conditions, among others) and then describing the causal, instrumental and teleological relations between these elements in the text by Moreiro-González and Bolaños-Mejías (2018). This structure guided both the specialised reading and the interpretation of the results produced by the AI, ensuring that the comparison between human and algorithmic procedures remained anchored in the same cognitive-semantic apparatus.

The analysis focuses on associative relations linked to the concept of indexing. This methodological decision is justified by the tendency of associative relations to make the functional interdependencies between processes, agents, instruments and products involved in subject representation more granular and explicit. The corpus analysis was conducted by the researchers and, in a complementary step, we consulted the Perplexity Pro system, an academic AI with access to models focused on complex reasoning.

To obtain an adequate response from the AI, we designed the following prompt to avoid hallucinations and ensure, through textual evidence, that the system would act as a data analyst, anchoring each associative relation in actual passages from the article: "Based on Information Science, Knowledge Organization, Frame Theory, and Frame-based Terminology, I want you to analyse this article and, assuming that concepts evoke knowledge frames that organise entities, roles and processes typical of a cognitive situation, extract the term 'indexing' and its associative relations (RT). I request the elaboration of an analytical table in order to guarantee traceability between the concept and the related term, the textual evidence (excerpt) and the theoretical modelling (frame)."

5.0 Results

In this section, we present the associative network of the term "indexing" as reconstructed from the corpus, first through manual frame analysis and then through the AI-mediated procedure. Table 1 summarises the relations identified by the researchers, linking each related term to its textual evidence and to the corresponding frame element, while Table 2 presents the relations returned by Perplexity Pro for the same article, including the proposed frame relations. Together, these tables make visible how the "indexing" frame is instantiated in the folksonomy context and allow us to compare the granularity and consistency achieved by human analysis and by the LLM.

Table 1 “Folksonomy indexing from the assignment of free tags to setup subject: a search analysis into the domain of legal history” – Corpus analysis carried out by the researchers.

Related term	Corpus excerpt	Elements of the frame
Folksonomy	“This is the context in which we find that folksonomies operate: for example, the practice of cooperative indexing, which is done by tagging with free keywords…”	CONDITIONS or STANDARD/NORM Occurs in (indexing → folksonomy)
Free tags	“cooperative indexing, which is done by tagging with free keywords… the difficulties caused by free tagging…” “We then examined the relevance, consistency and precision of the tags attached to the retrieved documents, in addition to their lexical composition.”	INSTRUMENT Is carried out from (cooperative indexing → free tags) PROCESS is a Process For (tags → retrieved documents / access)
Lexical quality	“The behaviour and lexical quality of the folksonomies is examined by comparing two online social networks…”	OUTPUT Is influenced by (indexing – folksonomies → user’s lexicon)
Search and retrieval	“these do nevertheless represent the most obvious problem to search and retrieval in folksonomies.”	GOAL Is influenced by (quality of indexing → quality of retrieval)

Table 2 “Folksonomy indexing from the assignment of free tags to setup subject: a search analysis into the domain of legal history” – Corpus analysis carried out by Perplexity Pro

Related term	Textual evidence (excerpt)	Frame relation
folksonomy / folksonomies	"This is the context in which we find that folksonomies operate: for example, the practice of cooperative indexing, which is done by tagging with free keywords..."	Folksonomy as indexing-environment frame; indexing realized as folksonomy tagging
cooperative indexing	"the practice of cooperative indexing, which is done by tagging with free keywords..."	Process subtype: INDEXING-AS-COOPERATIVE-ACTION
tagging / free tagging	"cooperative indexing, which is done by tagging with free keywords... the difficulties caused by free tagging..."	Lexical operation that instantiates indexing in the system
flat indexing system	"What we find in folksonomies is a flat indexing system, where all words have the same weight..."	Structural configuration of the INDEXING frame (no internal hierarchy between tags)
bottom-up indexing system	"A bottom-up indexing system provides everyone with user-oriented searchable computing."	Causal/organizational frame: origin of indexing from users rather than experts
retrieval systems	"projects and research and publication on thesauri, ontologies, indexing and retrieval systems..."	Complementary process frame: INDEXING–RETRIEVAL pair; indexing as input to retrieval
search and retrieval	"these do nevertheless represent the most obvious problem to search and retrieval in folksonomies."	Consequence relation: indexing quality → retrieval quality
related tags	"giving special attention to related tags within each retrieved document... tables were prepared, listing the related tags... to determine the indexation context."	Associative frame: INDEXING-AS-CONTEXTUAL-NETWORK
Tagmash	"This is called a 'tagmash,' and this function is what makes it possible to search a book by two or three tag combinations..."	Compositional frame: complex, conjunctive indexing expressions
synonym processing	"LibraryThing introduces all the synonyms that name a concept or fact... a conceptual and	Normalization frame: operations over the indexing vocabulary

Related term	Textual evidence (excerpt)	Frame relation
	formal synonym ring is established…"	(synonym ring)
Lemmatization	"The tags assigned reproduce, logically, the grammatical difficulties of any type of free indexing with key words… LibraryThing offers the asset that all these term variations are included when a search is carried out."	Morphological normalization within the INDEXING frame
controlled vocabularies	"controlled vocabularies are proving to be insufficient… another part is controlled at source… the folksonomies… employ a mix of free and controlled language."	Contrast/coordination frame: FREE-INDEXING vs CONTROLLED-INDEXING
corporate collections; controlled indexing	"another part is controlled at source. Many companies and organizations tag their photos themselves… subsequently, they are organized by a method of classification that usually reflects the same criteria that are used to organize the institution in question."	Institutional role: CORPORATE-INDEXER as specialized agent within indexing frame
subject indexing	"folksonomies can certainly be said to represent, nonetheless, a major improvement over traditional subject indexing."	Superordinate frame: traditional SUBJECT-INDEXING vs new FOLKSONOMY-INDEXING
hybrid system of indexing	"created by means of an increasingly hybrid system of indexing, that employs a mix of free and controlled language."	Integration frame: convergence between folksonomy tags and thesaural/ontological structures
messiness; inconsistency	"folksonomies can finally fall into 'messiness' and inconsistency… we can conclude that tag variations are the most prominent hindrance to search and retrieval."	Quality frame: noise, ambiguity as negatively valued states of the indexing system
legal history (as domain)	"the indexing procedure in the mental categorization of this legal frame… the terminology of legal history homogenizes the logical-semantic similarities."	Domain frame: LEGAL-HISTORY-FRAME providing semantic stability for indexing

6.0 Conclusions

The use of AI, configured as a data analyst, proved productive especially in the stage of identifying associative relations for the term "indexing," significantly expanding the set of related terms in comparison with the manual extraction carried out by the researchers. The excerpts returned remained strongly anchored in the text by Moreiro-González and Bolaños-Mejías (2018), which indicates that the prompt was effective in controlling hallucinations and in ensuring traceability between related term, textual evidence and proposed frame.

In dialogue with studies that employ semantic structures, such as Frame Semantics and FrameNet-inspired resources, in domain modelling and the organisation of specialised vocabularies, as well as research exploring the use of generative AI tools and LLMs in tasks of semantic analysis, categorisation and terminological harmonisation, the experiment reported here distinguishes itself by combining frame-based manual analysis with the use of a language model explicitly configured as a data analyst, ensuring traceability between related term, textual evidence and proposed frame. This methodological configuration highlights human–AI collaboration as a strategy for exploring, in a controlled manner, the complexity of the associative networks involved in "indexing" within folksonomy contexts.

On the other hand, the quality of the "frame relation" column fell short of expectations, since the AI tended to formulate relations that were more descriptive than genuinely analytical, often merely reformulating the excerpt in general terms instead of fully exploring the roles, processes and hierarchies envisaged in Frame Semantics and Frame-based Terminology. This suggests that the prompt was more precise in defining the behaviour for retrieving textual evidence than in guiding conceptual modelling, which required interpretive intervention by the researchers to refine the relations and validate the adequacy of the frames assigned.

Nevertheless, the procedure highlighted the potential of AI as an auxiliary tool in frame-oriented terminological studies, particularly for identifying patterns of co-occurrence, process subtypes and participant roles in the conceptual network of "indexing." The combination of specialised reading, grounded in Knowledge Organization and Frame Semantics, with the systematic exploration of the corpus mediated by AI made it possible to confirm and, in some cases, adjust the frame structure initially proposed for "indexing," showing that human–AI collaboration can be methodologically controlled and theoretically productive in research in this field.

Artificial Intelligence (AI) was utilized in the methodology of this study. Specifically, the Large Language Model (LLM) Perplexity Pro was configured as a data analyst to extract related terms, textual evidence, and proposed frame relations from the corpus. The AI served as a tool for an experimental assessment comparing human and algorithmic semantic analysis, rather than as a tool for authoring the manuscript text.

References

Baker, Collin F., Charles J. Fillmore, and John B. Lowe. 1998. “The Berkeley FrameNet Project.” In 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1, 86–90. Montreal, Quebec, Canada: Association for Computational Linguistics. doi: 10.3115/980845.980860.

Chishman, Rove. 2016. “Convergências entre semântica de frames e lexicografia.” Linguagem em (Dis)curso 16: 547–559. doi: 10.1590/1982-4017-160302-0316d.

Dahlberg, Ingetraut. 1978. “Teoria do conceito.” Ciência da Informação 7: 101–107. doi: 10.18225/ci.inf.v7i2.115.

Faber, Pamela, ed. 2012. A Cognitive Linguistics View of Terminology and Specialized Language. Berlin/Boston: De Gruyter Mouton. doi: 10.1515/9783110277203.

Fillmore, Charles J. 1982. “Frame Semantics.” In Linguistics in the Morning Calm: Selected Papers from SICOL-1981, edited by the Linguistic Society of Korea, 111–137. Seoul: Hanshin Publishing Company.

Hjørland, Birger. 2002. “Domain Analysis in Information Science: Eleven Approaches – Traditional as Well as Innovative.” Journal of Documentation 58: 422–462. doi: 10.1108/00220410210431136.

Hjørland, Birger. 2017. “Domain Analysis.” Knowledge Organization 44: 436–464. doi: 10.5771/0943-7444-2017-6-436.

Krebs, Luciana Monteiro, and Rita do Carmo Ferreira Laipelt. 2018. “Teorias da linguística cognitiva para pensar a categorização no âmbito da Ciência da Informação.” Transinformação 30: 81–93. doi: 10.1590/2318-08892018000100007.

Mai, Jens-Erik. 2001. “Semiotics and Indexing: An Analysis of the Subject Indexing Process.” Journal of Documentation 57: 591–622. doi: 10.1108/EUM0000000007095.

Minsky, Marvin. 1974. “A Framework for Representing Knowledge.” MIT Artificial Intelligence Laboratory Memo 306. Cambridge, MA: Massachusetts Institute of Technology.

Moreiro-González, José-Antonio, and Carmen Bolaños-Mejías. 2018. “Folksonomy Indexing from the Assignment of Free Tags to Setup Subject: A Search Analysis into the Domain of Legal History.” Knowledge Organization 45: 574–585. doi: 10.5771/0943-7444-2018-7-574.

Ruppenhofer, Josef, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, and Jan Scheffczyk. 2010. FrameNet II: Extended Theory and Practice. Berkeley, CA: International Computer Science Institute.