

Expanding the User's Query to Enhance Semantic Information Retrieval Using the Reasoning Mechanism Based on Homomorphism Between Semantic Annotations



Djamel Nessah^{1*}, Sofiane Mounine Hemam², Mohamed Lamin Laouadi³

¹ ICOSI Laboratory, Computer Science Department, University Abbes Laghrour-Khenchela, Khenchela 40002, Algeria

² Computer Science Department, University Abbes Laghrour-Khenchela, Khenchela 40002, Algeria

³ Computer Science Department, University Setif 1, Setif 19000, Algeria

Corresponding Author Email: Nessah_djamel@univ-khenchela.dz

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.290518>

ABSTRACT

Received: 20 May 2024

Revised: 3 September 2024

Accepted: 21 September 2024

Available online: 24 October 2024

Keywords:

conceptual graph, semantic annotation, semantic search, projection operation, homomorphism operation, ontology

Semantic search encompasses advanced technological approaches to information discovery and retrieval, employing semantic techniques to extract information from intricately structured data sources. An effective search engine must have the ability of accessing and retrieving information of interest by employing reasoning with conceptual models. However, the structural and semantic information intrinsic in conceptual models is not readily amenable to reasoning and AI-enhanced semantic processing. Therefore, the chosen model should enable understanding the meaning of concepts and the relationships between them. It should also carefully consider the context of the search, ultimately enhancing the accuracy of the returned results. Among the models that fulfill these objectives, conceptual graphs stand out as particularly interesting. They are built upon a robust theoretical framework that spans multiple domains, including philosophical, psychological, linguistic, and artificial intelligence disciplines. In this paper, we describe a method for semantic search driven by conceptual graph-based representation, and a powerful matching reasoning supported by a projection operation between the semantic annotations associated with the document and the submitted query.

1. INTRODUCTION

In keyword-based search systems, a document's content is summarized into keywords to facilitate efficient and straightforward search matching. However, keywords often fail to effectively represent the document's contents as they disregard its semantic information. Consequently, search results returned from the web may not always align with the user's query requirements [1].

Semantic search, supported by ontological frameworks, is increasingly utilized in information retrieval systems. Ontology-driven systems are instrumental in improving semantic comprehension and search accuracy by anchoring search processes in the meanings of terms rather than their syntactic structures [2]. Moreover, by strengthening the connection between information on web pages and background knowledge, these systems efficiently reduce the semantic gap between the keywords in documents and user queries, thereby enhancing the matching process [3, 4].

The initial focus lies on capturing the comprehensive knowledge embedded within a diverse document collection. This encompasses not only the domain-specific content but also structural elements like metadata and their dependencies. Meanwhile, the subsequent emphasis is on delivering users novel search outcomes drawn from this varied document pool. These outcomes aid users in understanding pertinent information related to their queries and in tracing

dependencies across multiple documents [5].

The main challenges faced by information retrieval techniques today are especially prominent in areas such as Natural Language Processing (NLP), which includes techniques such as text mining, sentiment analysis, and entity recognition, faces some challenges such as handling the ambiguities of human language and the need for continuous learning from new data.

Other modern semantic search systems use vector search engines to represent documents and queries as vectors in a high-dimensional space. These systems calculate the similarity between vectors to find relevant results based on semantic proximity. However, the formalization of semantic similarity remains challenging. Machine learning models, such as decision trees and deep neural networks, are used to analyze large textual datasets to improve response accuracy, relevance, and reduce response times. However, the data must be well-prepared for efficient retrieval and linkage, the goal is to uncover patterns that enhance search quality.

Graph-based Models, map the relationships between entities and concepts. These can be particularly useful for identifying and leveraging dependencies and complex relational structures within the data.

This contribution aims to apply a powerful formalism for representing complex knowledge. Indeed, the conceptual graph used to annotate queries and documents is a formalism that has its equivalent in logic, to apply logical inferences that

reduce the potential for varying interpretations of the text's meaning.

The conceptual graph, a distinctive form of knowledge graphs, establishes semantic connections among concepts and has demonstrated its value in various applications, including short text understanding, word sense disambiguation, semantic query extension and entity linking [6, 7]. This formalism offers a promising solution, as it can represent concepts as entities linked by various relationships.

Within semantic search, it is classified as a distinct subset of knowledge graphs providing a powerful knowledge representation and inference environment [8, 9].

In this perspective, this contribution aims to characterize document content through metadata and semantic annotations [10] that support content semantics and enable reasoning and inference production. To achieve this goal, the graph projection operation, as a powerful reasoning tool, will be extensively used to perform matching between a document and user's query annotations.

The search for such a matching is a reasoning process based on specialization/generalization operations much like the connections between these entities. Indeed, if a graph projects into another, it may reveal a discernible structure.

The structure of this manuscript includes in the following section a review of related work, followed by the description of the approach. Finally, we have a conclusion and future perspectives.

2. BACKGROUND

2.1 Conceptual graphs

The conceptual graphs (CGs) model, introduced by Sowa in 1984, serves as a formal and expressive framework for knowledge representation and reasoning mechanisms. This model is built on labeled graphs and comprises two main components:

- * Terminological Support: This component represents fundamental ontological knowledge.
- * Fact Representation: carries factual knowledge by representing data [11].

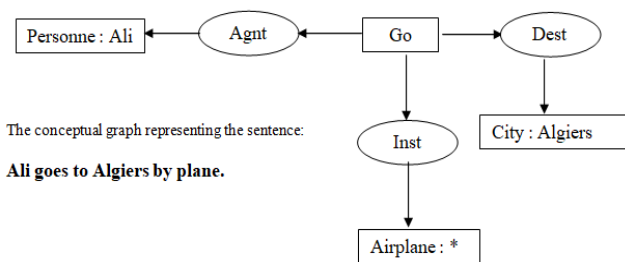


Figure 1. Representation of factual knowledge by CGs

Conceptual graphs enable the formalization of domain knowledge using a structure of concepts and relations. As illustrated in Figure 1, a concept represents a term or class of objects characterized by a set of attributes. A relation denotes the connection between concepts within the model. Both concepts and relations are organized through hierarchical taxonomies. This results in a root concept, which serves as the most general subsume, and various subsumed concepts that are progressively more specific [12, 13].

Taxonomic relations, such as Is-A, Sort-Of, and Part-Of, are represented by forward-directed arcs that illustrate how specialized terms derive from a root concept. A sub-concept inherits attributes from its parent concept, thereby enabling reasoning grounded in graph theory.

This model is especially effective for natural language processing (NLP) and semantic information retrieval, owing to its expressive capability and the ability to translate the conceptual graph into first-order logic formalism using a set of transformation rules.

2.2 Related works

In this section we survey related work carried out within the context of this contribution.

The paper presented by Elangovan and Nirmala [1] evaluates various ontology-based search techniques developed in recent years; examine the diverse approaches utilized in domain ontology to handle search requests. The principal objective of employing semantic search with ontology-based systems is to optimize three crucial parameters: precision, recall, and F-Measure. An important criterion in the selection of ontology for the semantic Web involves the underlying technology upon which it is based. This includes various components such as inference engines, annotation tools, and mining tools. Moreover, the second criterion for semantic annotation should enable the linking of entities in text to their respective semantic descriptions. Generally, there are three types of semantic annotation: manual, semi-automatic, and automatic.

The third criterion pertains to the indexing process which is the method of storing information to advance retrieval based on search queries. It involves a search engine preserving the content encountered during crawling, organizing it into indexes for swift retrieval in the future. Indexing streamlines the matching process; without it, retrieving information would necessitate selecting through collected web pages. The types of indexing are:

- Forward Indexing: Records a list of words for each document.
- Inverted Indexing: Records a list of documents for each word.

Graph Indexing: Utilizes a query graph to retrieve a set of answers from an index, verifying those graphs containing the query graph, and returning the query results.

At the end, the ranking task determines the order of results returned during a search query.

Additionally, the ontology description language, such as RDF and OWL, should be considered for manipulating and storing the RDF data.

Kong et al. [14] investigated how conceptual graphs (CGs), which are grounded in logic and support visual reasoning, play an integral role in advancing artificial intelligence. Also, CGs are applicable for semantic matching [15] and find extensive use across various applications.

The homomorphism matching, a fundamental operation for logical deduction using CGs, poses a challenge due to its NP-complete nature. As the size of CG databases grows, the efficient execution of homomorphism matching becomes pivotal in the application of graph rules. In this background, the authors aim to address the challenge of efficiently performing homomorphism matching for graph rules. They propose a novel hybrid approach that involves transforming Conceptual Graphs (CGs) into labeled undirected graphs

devoid of multiple edges

Conceptual graphs are seen as formalism and an extension of semantic networks. They offer a powerful representation method for expressing, describing, and manipulating knowledge. Conceptual graphs are labeled graphs composed of two types of nodes:

- Concept nodes: which represent objects, entities, or ideas related to domain knowledge.
- Relation nodes: which represent the links between concepts.

The arcs connecting relation nodes to concept nodes, and concept nodes to relation nodes, are directed.

Additionally, every concept and relation are typed, corresponding to a type within the hierarchy of types relevant to the application domain.

The hierarchy of types for concepts and relations is ordered according to the specificity of each type. A type "t" is more specific than a type "s" if it inherits all the information contained in "s". Type inheritance is thus a relation between a type and a set of subtypes. It is expressed as "a kind of" or "is a" relation. It establishes a hierarchy between the types.

The projection operation (homomorphism) constitutes the essential inference mechanism in CGs and allows for computing specialization operations between graphs. It has been demonstrated that for two given CGs, U and V, where $U \leq V$ (U specializes V), the following holds:

There exists a mapping $\pi: V \rightarrow U$ where $\pi(V)$ is a sub graph of U, called the projection of V into U, if and only if there exists a projection of U into V.

The projection operation satisfies the following properties:

- a) $\forall c \in V$, $\pi(c)$ is a concept of $\pi(V)$. The type of $\pi(c)$ is either the same as that of c or a subtype of it.
- b) $\forall r \in V$, $\pi(r)$ is a conceptual relation of $\pi(V)$. The type of $\pi(r)$ is either the same as that of r or a subtype of it.
- c) If the i^{th} arc of r is connected to a concept c in graph V, then the i^{th} arc of $\pi(r)$ is connected to $\pi(c)$ in the sub graph $\pi(V)$.

Zhang et al. [16] have developed a comprehensive conceptual graph, ALiCG, which includes over 5,000,000 finely detailed concepts. This graph is rapidly expanding, with concepts being automatically extracted from search logs despite various inaccuracies and inconsistencies. Echoing the methodology presented in the study by Ni et al. [17], this approach represents a document as a compact concept graph, where nodes represent concepts extracted from the text and linked to entities in a knowledge base. The edges within the graph reflect the semantic and structural relationships between these concepts. A range of techniques are employed to measure and assess the strength of these relationships. Within the concept graph, concepts are weighted using the closeness centrality measure, reflecting their importance with regard to the document's aspects. The framework ALiCG leverages findings from previous studies, such as YAGO and Probase, as referenced by researchers [18, 19], to extract knowledge from formal texts (e.g., Wikipedia) and concepts from semi-structured web documents.

The framework is structured into levels:

Level-1 includes concepts that represent the domain of the instances.

Level-2 consists of concepts that denote the type or subclass of these instances.

Level-3 involves concepts that provide a detailed conceptualization of instances, reflecting implicit user intentions.

The instance layer comprises all instances, including entities and non-entity phrases. ALiCG is currently employed to support various business services, including intent classification, named entity recognition, query rewriting, and more.

The methodology comprises three main steps:

1) Fine-grained Concept Acquisition: This step concentrates on extracting prevalent, finely detailed concepts from noisy search logs.

2) Long-Tail Concept Mining: Although iterative pattern matching can identify numerous high-frequency concepts, extracting long-tail concepts presents challenges, primarily due to limited pattern generalization and sparse co-occurrence samples.

3) Taxonomy Evolution: updating existing taxonomies and incorporating new emerging concepts are essential aspects of this work.

In the study by Maksimov et al. [20], the authors propose an approach to develop a cognitive-process-oriented model that enhances data retrieval and associated human interaction tools. This model conceptualizes cognitive search as the process of constructing ontology for the subject area or target object. The cognitive search process is described as the formation of a subject area ontology, which consists of three interrelated systems: functional, conceptual, and terminological. Consequently, the results and pathways of information retrieval are designed to reflect and enhance the cognitive process. To align the graph's dimensions with perceptual capabilities, aspect projection operations based on the taxonomy of relationships and entities are utilized.

The cognitive information retrieval model performs the following tasks:

- * Selection of documents from information resources.
- * Construction of information components is based on a set of distinct features of clusters, along with an evaluation of their integrity.
- * Ordering these clusters according to their "value" to reduce the sample size that the user needs to review.

In the study by Tuteja and Kumar [21], the aim is to develop and analyze query-driven graph models that incorporate additional nodes and edges to enhance query processing capabilities. Graph models, noted for their flexibility, have significantly advanced the development of various artificial intelligence techniques for mapping data into a graph format. The research details three distinct graph models, with a foundational model specifically designed for e-commerce applications. The first model involves the transformation of attributes intended for search queries into separate nodes. The second model enhances the baseline graph by introducing new relationships, focusing on minimizing the query path length. The third model integrates the nodes and relationships from both the first and second models to unify their features.

Along similar lines, Sequeda et al. [22] introduced a method for transforming a relational model into an RDF graph, wherein every property of an object is transformed into a node within the graph structure. Additionally, the authors validated the preservation of data and queries in the proposed RDF graph.

The research paper cited by Devezas and Nunes [8] focuses on entity-oriented searches. It notes that whereas traditional search systems primarily retrieved documents, contemporary search engines have advanced to retrieve entities and provide direct answers to users' information needs. This evolution highlights the importance of cross-referencing information

from diverse sources.

In entity-oriented search, the content may be either unstructured, structured, or a combination of both, often resulting in semi-structured data or interactions between unstructured and structured data. Bast et al. [23] describe combined data in semantic search as text enhanced with entities from a knowledge base, or as an integration of multiple knowledge bases with varied naming conventions. This integrated data is central to entity-oriented search.

The approach described by Ali [4] undertakes a comprehensive examination of the advantages stemming from the application of AI-driven techniques, such as graph-based machine learning, to represent conceptual models within Knowledge Graphs (KGs). The approach aims to address the gap in conceptual model search by first devising a versatile KG transformation tool. This tool is designed to construct a representation that effectively joins heterogeneous models using domain-specific, core, and foundational ontologies as a semantic support.

This representation establishes semantic connections among diverse knowledge assets and, during model retrieval, serves as the contextual framework for the model and its components, thereby delivering highly pertinent results. Leveraging the structural and semantic knowledge encapsulated within the KG representation of a conceptual model, this representation is subsequently utilized as input for the search engine workflow. This workflow facilitates the indexing, storage, and retrieval of conceptual models based on varied search criteria.

The research paper by Muniyappa and Kim [24] utilizes the Universal Sentence Encoder (USE) to assess the semantic similarity of text. It also applies transfer learning techniques to implement Genetic Algorithm (GA) and Differential Evolution (DE) algorithms to search and retrieve the N relevant documents regarding the user's query. This methodology is demonstrated using the Stanford Question and Answer (SQuAD) Dataset to identify user queries.

3. THE PROPOSED APPROACH

Information retrieval on the web is an active research field where researchers strive to address major challenges. It is true that search engines such as Amazon, Yahoo, Bing, and notably Google partially meet users' queries regarding the relevance of results to their information retrieval needs, particularly on the web.

Indeed, some responses are irrelevant to the search context, and others are missing despite their relevance because keyword comparisons are based on syntax and overlook their semantics. Terms are thus treated independently, and the generated results will require additional processing to select only relevant documents.

On the other hand, we often need to reformulate the query to regenerate missing documents. These two aspects are acknowledged by recall, precision and F-measure metrics.

3.1 Syntactic search and semantic search

Traditional information retrieval models, including the Boolean and vector space models, though practical and straightforward, lack any integration of semantic understanding and do not differentiate between documents that may contain similar terms but provide different semantics

when combined. For example, the terms "Library" and "Book" combined as "library of books" and "book of libraries" provide different semantics.

Therefore, the context of term occurrences must be considered to define the relationships between these terms and thus establish the resulting semantics. Today, the best model for considering such constraints is the concept of ontology, a formalized framework for modeling entities and relationships in a knowledge domain and subsequently reasoning to derive semantics related to implicit knowledge.

3.2 Performance measures of an information retrieval system

Given a collection of documents and a user's query, the performance measures of an information retrieval system on the collection and with respect to this query are:

3.2.1 Recall

Recall is measured as the fraction of relevant documents successfully retrieved out of the total number of relevant documents in the collection.

3.2.2 Precision

Precision represents the percentage of relevant documents among those that were retrieved.

Low precision means that the user must spend time reading irrelevant information, which is a consequence of a high presence of polysemy; whereas low recall means that the user will not have access to a set of relevant and desirable information, which the effect is caused by synonymy.

Additionally, two other measures are defined: noise and silence, as complementary notions of precision and recall, thus we have: $\text{Noise} = 1 - \text{Precision}$ and $\text{Silence} = 1 - \text{Recall}$.

The ideal would be to have a precision and recall of "1", but these two requirements are often contradictory and inversely proportional. Very high precision can only be achieved at the expense of low recall, and vice versa.

3.3 The ontology model for semantic search

Ontology is a declarative model that associates classes, individuals, relations, functions, and assertions, preventing various semantic interpretations and ensuring the proper use of these ontological terms. Ontology also addresses interoperability, sharing, and reuse of knowledge issues.

In this sense, ontology allows:

- Sharing knowledge among agents/ services on the web.
- Performing semantic indexing (annotation) of documents.
- Improving information retrieval processes.

Through this controlled vocabulary, we can classify documentary content; extend user queries based on hierarchies of classes, relations, and rules to: Figure 2.

- Process and index large amounts of information available on the web in various form, structured, semi-structured and unstructured.

- Maintain semantic coherence, as the semantics of a document are not equal to the sum of the semantics of its constituent fragments.

- Translate the ontology model formally compared to procedural models.

- Facilitate the consideration of various changes related to the semi-structured or unstructured nature of some documentary resources.

- Ontology is relatively expressive in describing rich and complex knowledge domains.
- Facilitate communication between agents, such as cognitive MAS.

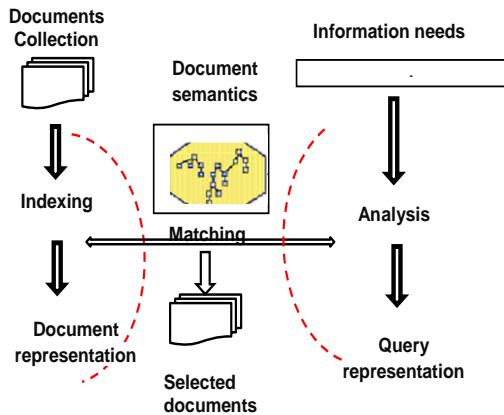


Figure 2. Conceptual resources

3.4 Methodology

In this work, we will use the following concepts:

3.4.1 Semantic annotation

It is a process of associating semantic information with data documents or digital resources, to make them understandable by machines. Semantic annotation helps structure and gives meaning to data by linking them to specific concepts defined in ontologies or taxonomies.

The logical model of semantic information retrieval defines this search as the extraction of a set of documents "d" such that for the query "r", these documents validate the proposition:

$$K \vdash d \rightarrow r$$

K: domain of knowledge

d, r: logical formulas

The common view is that it is necessary to annotate both document contents and user queries with terms defined in ontology.

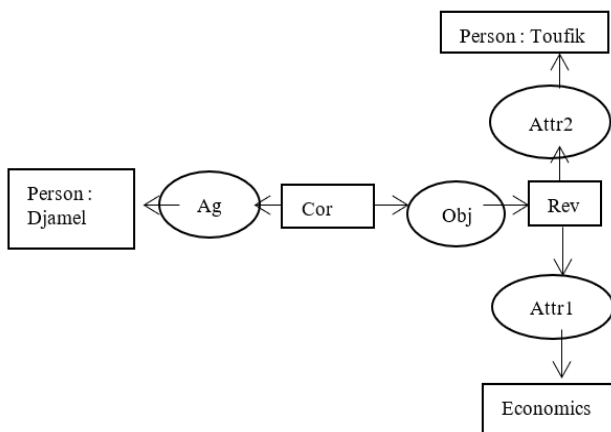


Figure 3. The related conceptual graph

Figures 3-4 provide an overview of the semantic annotation process related to the textual fragment:

"Djamel corrects the review of economic of Toufik".

Our approach utilizes semantic annotation formalized by a conceptual graph, which is a widely utilized expressive

framework for representing semantics in natural language applications. It carries out projection operations, which serve as a key component of our research. Additionally, conceptual graphs can be translated into logical formulas, offering a foundation for constructing a rigorous reasoning mechanism grounded in this formal logic framework [25].

Figure 4 illustrates the XML/RDF syntax of the textual fragment above.

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-
rdf-syntax-ns#"
xmlns:Econom="http://www.Econom-
indic.com/onto#Econom">
<rdf:Description
Rdf:about="http://.../Djamel">
<Econom :cor rdf:resource="http://.../Review"/>
</rdf:Description>
<rdf:Description
Rdf:about="http://.../Review">
<Econom :Pers>Toufik</Econom :Pers>
<Econom:type> rdf:resource="http://.../Economics"/>
</rdf:Description>
</rdf:RDF>
```

Figure 4. XML/RDF schema associated to CG in Figure 3

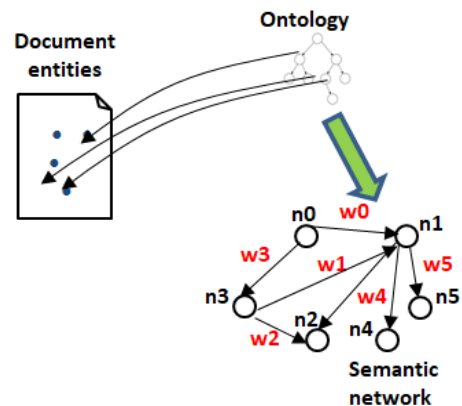


Figure 5. Ontology-based document annotation

3.4.2 Homomorphism to query-document mapping

Based on the operation of projection (homomorphism) between ontologies (conceptual graphs), this process involves using a mapping or correspondence between the terms in the search query and those in annotated documents, usually based on a representation in the form of a conceptual graph. The projection operation consists of transferring relevant elements from the search query to corresponding documents in the conceptual space defined by the ontologies. This improves the relevance of results by considering semantic relationships between concepts rather than just searching for lexical matches (Figure 5).

This model provides the essential inference mechanism and allows for computing specialization operations between graphs. To apply this operation, especially for mapping two conceptual graphs CGd and CGq, annotating a document and

a query respectively, we have:

Concepts. We should build

A) Intentional definition: Hierarchy of concept types E_C (Figure 6).

B) Two special maximal elements: The "universal" type: denoted \top and the "absurd" type: denoted \perp . For a pair of concepts, we can define a minimal common super type and a minimal common subtype.

C) Extensional definition:

- For each type t of E_C , we associate a set of objects $[(t)]$: the possible referents of t .

- A concept is represented by a pair $[type, t]$

referent= $*$: generic concept (by default)

referent= $\#i$: individual concept

referent= $@$: measure

Relations. We should have

A) Hierarchy of relation types: A relation is defined by two elements (R, A)

R: name of the relation

A: arity (number of arguments) of the relation = integer n indicating the number of arcs.

B) Signature of a relation: set of n types of concepts.

C) Nature of relation

Function with one or more arguments, where the value domain is $\{\text{true}, \text{false}\}$.

Binary predicate or more.

Property: unary predicate.

It is verified that for two given conceptual graphs U and V such that $U \leq V$ (U specializes V), then:

There exists a function $\Pi: V \rightarrow U$ where $\Pi(V)$ is a sub-graph of U called the projection of V into U, if and only if:

1) $\forall C \in V, \Pi(C)$ is a concept in $\Pi(V)$. The type of $\Pi(C)$ is the same as that of C or it is a subtype of it.

2) $\forall R \in V, \Pi(R)$ is a conceptual relation in $\Pi(V)$. The type of $\Pi(R)$ is the same as that of r or it is a subtype of it.

3) If the I^{th} arc of R is connected to a concept in the graph V, then the I^{th} arc of $\Pi(R)$ is connected to $\Pi(C)$ in the sub-graph $\Pi(V)$.

Mapping query-document algorithm. For this work, given two conceptual graphs:

U: represents the semantic annotation of a document D, it is the conceptual graph denoted as U.

V: represents the semantic annotation of the query R, it is the conceptual graph denoted as V (Figure 6).

a) If U specializes $V(U \leq V)$ and there exists a projection operation $\Pi: V \rightarrow U$ where $\Pi(V)$ is a sub-graph of U, then in this case, we can infer that the document D in question answers the query with a high precision measure.

b) If V specializes $U(V \leq U)$ and there exists a homomorphism $\Pi: U \rightarrow V$ where $\Pi(U)$ is a sub-graph of V, then in this case, we can infer that the document D in question answers the query with significant recall.

c) If $U=V$ then the document in question exactly answers the semantics of the query, which is the ideal case, but rarely achieved.

Figure 6 illustrates the projection of a graph G onto the graph $\Pi(G)$. The upper section of the figure shows G, while the lower section presents its corresponding $\Pi(G)$.

Experimentations and discussion. To test the mapping approach and analyze the results, we have chosen the example of annotations provided in Figure 7 and Figure 8.

Figure 7 shows an example of semantic annotation of a document D, which deals with analysis methods used in Business Intelligence.

Figure 8 illustrates an example of annotation for a query to search documents describing descriptive methods in Business Intelligence. For the document in question, the colored part constitutes a sub-graph of the query graph, satisfying the conditions for applying a projection operation. Thus, the mapping is performed and the document is considered as a relevant response to this search.

The first annotation (A) describes the analysis methods used in business intelligence, with a list of reporting tools included in the descriptive class of analysis methods. Suppose this annotation is a semantic indexing linked to a document describing business intelligence practices.

The second annotation (B) indexes a specific query, focusing on descriptive analysis methods in the context of business intelligence.

In this context:

A projection operation is possible from B to A. Indeed, the conceptual graph A specializes the conceptual graph B, and a projection $\pi: B \rightarrow A$ is possible as given in Figure 9. The information about reporting tools present in the document linked to the first annotation A can be relevant responses to the query modelled by the second annotation B.

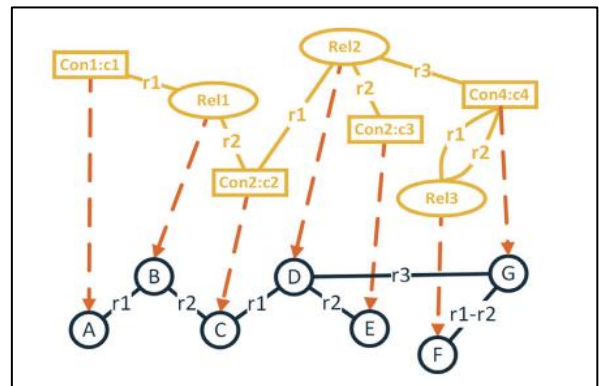


Figure 6. Homomorphism mapping between semantic annotations

A) Annotation of a document dealing with analysis methods in business intelligence.

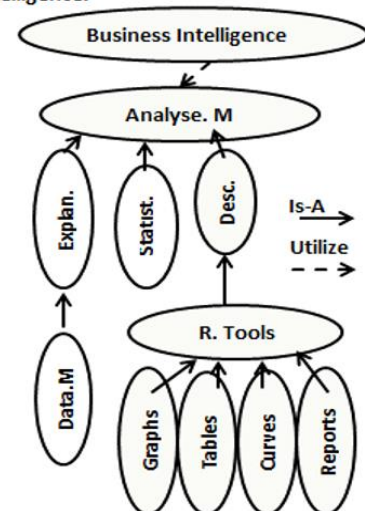


Figure 7. Document annotation

B) Annotation of a query to search descriptive methods used in Business Intelligence

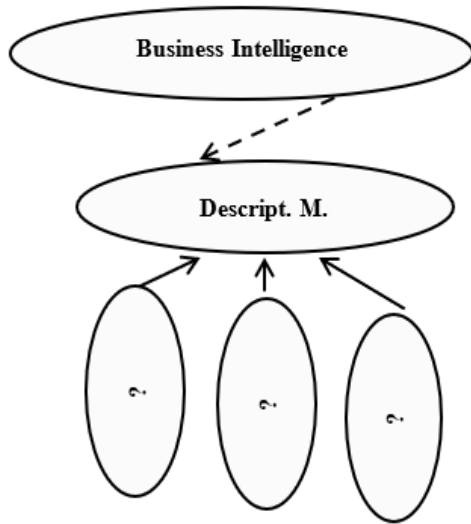


Figure 8. Query annotation

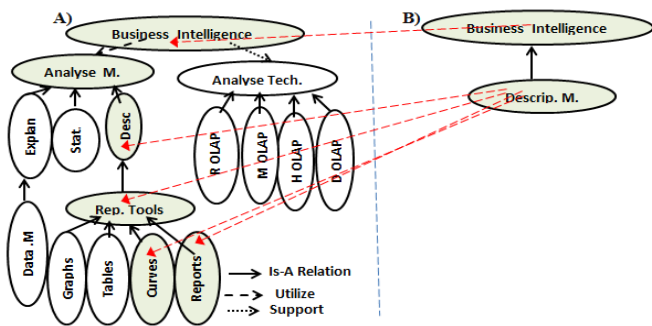


Figure 9. Projection operation from B) in A)

In annotation A, we have multiple methods and technologies associated with business intelligence, including explanatory, statistical, and descriptive methods, as well as OLAP technologies. Additionally, we have a list of specific reporting tools in the descriptive section.

In annotation B, we simply have a mention of descriptive methods in business intelligence.

In summary the information about reporting tools in annotation A can be projected onto annotation B, which focuses on the "descriptive" method. These can be sought as a result of an inference mechanism.

4. CONCLUSION AND FUTURE WORK

The use of knowledge representation through conceptual graphs is a widely adopted artificial intelligence technique, enabling the explicit representation of semantics through concepts, relations, and functions.

For semantic web applications, this facilitates providing higher-quality services and relieving users of additional selection tasks concerning their service needs.

In this approach, we chose to use conceptual graphs as a model for semantically annotating documents on the web and user-submitted queries. This approach is further explored for

conducting semantic searches using a powerful reasoning mode, which involves the search and application of projection operations between annotations.

This operation enables considering subtypes, whether within concept hierarchies or relation hierarchies. Thus, the semantics of implicit knowledge are provided through the execution of this powerful reasoning mechanism.

The long-term impact and future challenges of this approach are varied. First, there is the difficulty in accurately understanding the contextual meaning of words in a search query without a generic knowledge base. Additionally, the exponential growth of data, where the daily volume of data has rapidly increased, renders traditional data management systems inadequate, especially regarding the evolution of semantic web tools to ensure the most descriptive possible annotation of documents. Finally, there is a growing need for precise searches. As the volume of data continues to increase, users require more specific and accurate search results, making it crucial to maintain the accuracy of the precision metric. For future perspectives, it is planned to incorporate tools for annotating complex knowledge, where relationships between concepts may necessitate modelling with more suitable ontology notions, tools, and appropriate languages.

REFERENCES

- [1] Elangovan, D., Nirmala, K. (2016). Semantic search of e-learning documents using ontology-based system. *International Journal of Business Intelligents*, 6(2): 52-54. <https://doi.org/10.20894/IJBI.105.006.002.006>
- [2] Remi, S., Varghese, S.C. (2015). Domain ontology driven fuzzy semantic information retrieval. *Procedia Computer Science*, 46: 676-681. <https://doi.org/10.1016/j.procs.2015.02.122>
- [3] Wang, Z., Fu, Z., Sun, X. (2018). Semantic contextual search based on conceptual graphs over encrypted cloud. *Security and Communication Networks, Security and Communication Networks*, 2018(1): 1420930. <https://doi.org/10.1155/2018/1420930>
- [4] Ali, S.J. (2022). Knowledge Graph-based Conceptual Models Search. In *ER Forum/PhD Symposium.*, Hyderabad, India.
- [5] Charbel, N. (2018). Semantic representation of a heterogeneous document corpus for an innovative information retrieval model: Application to the construction industry. Doctoral dissertation, LIUPPA-Laboratoire Informatique de l'Université de Pau et des Pays de l'Adour. <https://theses.hal.science/tel-02074440/>.
- [6] Luan, Y., Hauer, B., Mou, L., Kondrak, G. (2020). Improving word sense disambiguation with translations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 4055-4065. <http://doi.org/10.18653/v1/2020.emnlp-main.332>
- [7] Chen, L., Liang, J., Xie, C., Xiao, Y. (2018). Short text entity linking with fine-grained topics. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy*, pp. 457-466. <https://doi.org/10.1145/3269206.3271809>
- [8] Devezas, J., Nunes, S. (2021). A review of graph-based models for entity-oriented search. *SN Computer Science*, 2(6): 437. <https://doi.org/10.1007/s42979-021-00828-w>
- [9] Polovina, U.P.S., Hill, R. (2007). Conceptual structures:

- Knowledge architectures for smart applications. In 15th International Conference on Conceptual Structures, ICCS, Sheffield, UK, pp. 22-27. <https://doi.org/10.1007/978-3-540-73681-3>
- [10] Sharma, A., Kumar, S. (2023). Machine learning and ontology-based novel semantic document indexing for information retrieval. *Computers & Industrial Engineering*, 176: 108940. <https://doi.org/10.1016/j.cie.2022.108940>
- [11] Okina, S.N., Taillandier, F., Ahouet, L., Hoang, Q.A., Breyse, D., Louzolo-Kimbembe, P. (2023). Using Conceptual Graph modeling and inference to support the assessment and monitoring of bridge structural health. *Engineering Applications of Artificial Intelligence*, Elsevier, 125: 106665. <https://doi.org/10.1016/j.engappai.2023.106665>
- [12] Chen, X., Jia, S., Xiang, Y. (2020). A review: Knowledge reasoning over knowledge graph. *Expert Systems with Applications*, 141: 112948. <https://doi.org/10.1016/j.eswa.2019.112948>
- [13] Sowa, J.F. (1984). *Conceptual structures: Information processing in mind and machine*. Addison-Wesley Longman Publishing Co., Inc.
- [14] Kong, J., Huang, J., Chen, H., Gong, J. (2019). A new hybrid approach to improve the efficiency of homomorphic matching for graph rules. *Knowledge-Based Systems*, Elsevier, 182: 104852. <https://doi.org/10.1016/j.knosys.2019.07.023>
- [15] Peiqi, L.I.U., Miao, H.U.A.N.G., Hao, F.E.N.G., Wei, Z.H.O.U. (2017). Research on pragmatic inference of fuzzy conceptual graph matching. *Journal of Frontiers of Computer Science & Technology*, 11(9): 1513. <https://doi.org/10.3778/j.issn.1673-9418.1606027>
- [16] Zhang, N., Jia, Q., Deng, S., Chen, X., Ye, H., Chen, H., Tou, H., Huang, G., Wang, Z., Hua, N., Chen, H. (2021). Alic: Fine-grained and evolvable conceptual graph construction for semantic search at Alibaba. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. Virtual Event, Singapore. ACM, New York, NY, USA, pp. 3895-3905. <https://doi.org/10.1145/3447548.3467057>
- [17] Ni, Y., Xu, Q.K., Cao, F., Mass, Y., Sheinwald, D., Zhu, H.J., Cao, S.S. (2016). Semantic documents relatedness using concept graph representation. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, San Francisco, California, USA, pp. 635-644. <https://doi.org.sndll.arn.dz/10.1145/2835776.2835801>
- [18] Rebele, T., Suchanek, F., Hoffart, J., Biega, J., Kuzey, E., Weikum, G. (2016). YAGO: A multilingual knowledge base from wikipedia, wordnet, and geonames. In *The Semantic Web-ISWC 2016: 15th International Semantic Web Conference*, Kobe, Japan, Proceedings, Part II. Springer International Publishing. Springer, Cham, 15: 177-185. https://doi.org/10.1007/978-3-319-46547-0_19
- [19] Wu, W., Li, H., Wang, H., Zhu, K.Q. (2012). Probase: A probabilistic taxonomy for text understanding. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, Scottsdale Arizona, USA, pp. 481-492. <https://doi.org/10.1145/2213836.2213891>
- [20] Maksimov, N., Golitsina, O., Monankov, K., Gavrilkina, A. (2020). Knowledge representation models and cognitive search support tools. *Procedia Computer Science*, 169: 81-89. <https://doi.org/10.1016/j.procs.2020.02.118>
- [21] Tuteja, S., Kumar, R. (2023). Query-driven graph models in e-commerce. *Innovations in Systems and Software Engineering*, Springer, 19(2): 177-195. <https://doi.org/10.1007/s11334-021-00421-7>
- [22] Sequeda, J.F., Arenas, M., Miranker, D.P. (2012). On directly mapping relational databases to RDF and OWL. In *Proceedings of the 21st International Conference on World Wide Web*. Association for Computing Machinery, New York, NY, USA, 12: 649-658. <https://doi.org/10.1145/2187836.2187924>
- [23] Bast, H., Buchhold, B., Haussmann, E. (2016). Semantic search on text and knowledge bases. *Foundations and Trends® in Information Retrieval*, 10(2-3): 119-271. <https://doi.org/10.1561/15000000032>
- [24] Muniyappa, C., Kim, E. (2023). Evolutionary algorithms approach for search based on semantic document similarity. In *Proceedings of the 2023 11th International Conference on Computer and Communications Management*, Nagoya, Japan, pp. 123-128. <https://doi.org.sndll.arn.dz/10.1145/3617733.3617753>
- [25] Nessah, D., Kazar, O. (2012). Document analysis to provide semantic metadata based ontologies. In *2012 International Conference on Multimedia Computing and Systems*, Tangiers, Morocco, IEEE Xplore, Tangiers, Morocco, pp. 725-731. <https://doi.org/10.1109/ICMCS.2012.6320115>