A Hybrid Dynamic Knowledge Graph Building Approach

Sana Ben Abdallah Ben Lamine[1,*], Redouane BOUHAMOUM[2] and Hajer BAAZAOUI[3]

[1]*RIADI Laboratory, ENSI, University of Manouba, Tunisia*
[2]*ETIS UMR 8051, CY University, ENSEA, CNRS, Cergy, France*
[3]*ETIS UMR 8051, CY University, ENSEA, CNRS, Cergy, France*

Abstract
Interest is increasingly focused on Knowledge Graphs (KGs) thanks to their ability to structure heterogeneous information into an organized and interconnected form, enabling better reasoning, decision-making, and supporting advanced AI applications. The construction method of these KGs is just as important, as it has a direct impact on automating their creation, guaranteeing their quality and their adaptation to constantly evolving needs and applications. In this paper, we propose a hybrid knowledge graph construction approach that combines the structured, ontology-driven top-down approach with the data-driven, iterative bottom-up approach, combining the rigor of top-down ontology design with the adaptability of bottom-up data integration. Our approach is three-phase : the first is a top-down ontology construction phase, the second is a bottom-up data integration and enrichment phase and the third is a hybrid integration and continuous evolution phase. Our aim is to create a more flexible construction method that takes into account structured, semi-structured and unstructured data, offering a comprehensive and dynamic KG. This approach is particularly effective for domains requiring both structured knowledge and dynamic data integration, from social media for example, like the healthcare domain. A detailed case study in this domain illustrates our proposal and demonstrates its feasibility and effectiveness.

Keywords
Hybrid Knowledge Graph construction, Ontology-driven, Data-driven

1. Introduction

Knowledge Graph Construction (KGC) consists mainly of extracting structural knowledge from unstructured texts, such as news articles, scientific literature, or web pages. This includes common subtasks like [1]: Named Entity Recognition (NER) which aims at identifying and classifying entities (e.g., persons, locations), Relation Extraction (RE) which aims at identifying relationships between pairs of entities, Event Extraction (EE) which aims at identifying event type, triggers and arguments, Entity Linking (EL) which aims at linking textual entity mentions to entries in a Knowledge Base (KB) like Wikidata or DBpedia, and Knowledge Graph Completion (KGComp), which can extend the KGC pipeline to include completion of missing links in the KG (i.e., inferring new knowledge).

For instance, given the sentence, ”Steve Jobs and Steve Wozniak co-founded Apple in 1977.”[2]: Named Entity Recognition : recognized entities, e.g., ‘Steve Job’, ‘Steve Wozniak’ classified as PERSON, and ‘Apple’ classified as ORG (ORGANIZATION); Relation Extraction : relation between the pairs ⟨Steve Job, Apple⟩ and ⟨Steve Wozniak, Apple⟩ is founded; Event Extraction : identifies the type of event as Business Start-Org where ‘co-founded’ triggers the event and the arguments are (Steve Jobs, Steve Wozniak) who are respectively participants in the event as AGENT and Apple as ORG. Entity Linking : links the mentions of Steve Job to Steven Jobs (Q19837) and Apple to Apple Inc. (Q312) on Wikidata. Knowledge Graph Completion : completes incomplete triples like ⟨Steve Job, create, ?⟩ by predicting blank entities giving for example NeXT Inc. and Pixar.

Many researchers have taken an interest in this topic, and several recent surveys and research papers cover different aspects of Knowledge Graph Construction (KGC), offering complementary points of

Cooperative Information Systems - Early Research Achievement and Demos, 2025.
*Corresponding author.
Sana.benabdallah@isamm.uma.tn (S. B. A. B. Lamine); redouane.bouhamoum@ensea.fr (R. BOUHAMOUM); hajer.baazaoui@ensea.fr (H. BAAZAOUI)
0000-0002-6018-2518 (S. B. A. B. Lamine); 0000-0002-5370-0401 (R. BOUHAMOUM); 0000-0002-2151-7397 (H. BAAZAOUI)

view. General KGC techniques are surveyed in [3], Multi-Modal KGC are discussed in the survey [4], generative KGC Methods are discussed in the survey of [2], automatic KG construction methods are detailed in [5] while integrating KG into Language Models is reviewed in [6] and integrating large language models (LLMs) across three key phases of KG construction is reviewed in [7].

Other surveys focus on specific applications of knowledge graphs like knowledge reasoning in [8], Biomedical KGC in [9] or semantic data integration [10].

Some other works classify KGC into top-down, bottom-up, and hybrid approaches [11], [12], [13] and [14].

Both top-down and bottom-up approaches present advantages and limitations.

In fact, top-down methods, which rely on predefined ontologies or schemas, ensure semantic consistency and facilitate reasoning, but they are often limited by their dependence on manually crafted knowledge structures and lack scalability in dynamic domains like social media or healthcare.

In the other hand, bottom-up approaches extract entities and relations directly from data leveraging statistical or machine learning techniques. In this way, they offer greater scalability and enable the discovery of new patterns. However, they often suffer from noise, a lack of semantic foundations and poor interpretability.

These complementary strengths and weaknesses motivate us to propose a hybrid method for knowledge graph construction, aiming to combine the semantic rigor of top-down approaches with the adaptability of bottom-up techniques, especially in an evolutionary field like healthcare, where both structured knowledge and data-driven insights are critical.

The rest of this paper is organized as follows. Section 2 presents related work, highlighting existing approaches to knowledge graph construction classified by construction type or method. In Section 3, we introduce our proposed hybrid method, which combines top-down and bottom-up strategies across three main phases : a top-down ontology construction phase, a bottom-up data integration and enrichment phase and a hybrid integration and continuous evolution phase. Section 4 describes a case study in the medical field, demonstrating the feasibility and relevance of our approach in a dynamic and complex domain. Finally, Section 5 concludes the paper and outlines directions for future work.

2. Related works

In this section, we present an overview of knowledge graph construction (KGC), classifying the methods based on their their construction techniques.

General KGC techniques are thoroughly discussed by Ji et al. [3], who present a comprehensive overview of methods for KGC, knowledge representation learning, and downstream applications. Their work offers a solid foundation for understanding traditional and modular approaches, where the task of construction of a knowledge graph is carried out as a pipeline of different subtasks.

Multi-Modal KGC is the focus of Zhu et al. in their survey [4], which proposes a systematic taxonomy for integrating text, images, and structured data into unified multi-modal knowledge graphs, while identifying key challenges and research directions.

Generative KGC Methods are reviewed by Ye et al. [2]. This paper presents a detailed taxonomy and theoretical framework for generative approaches such as prompting and fine-tuning with Large Language Models (LLMs).

Automatic knowledge graph construction methods are reviewed in [5]. The paper focuses on three central steps of KGC : acquisition, refinement, and evolution. It focuses on the transition from simple fact extraction to structured, conceptual organization of knowledge. It also analyzes methods in various scenarios and offers an overview of current challenges and future directions.

Incorporating KG into Language Models is reviewed by Safavi and Koutra [6], who examine how relational knowledge from KGs can be integrated into contextualized LMs to improve natural language understanding and reasoning.

In [7], the authors synthesize, survey and discuss advances in KG construction, since 2022. They focus on the use of Large Language Models (LLMs) mainly in three KGC phases: extraction, learning paradigm, and evaluation. Multimodal data extraction, advanced learning techniques like GNNs and Transformers, and evaluation methods are particularity addressed in this paper.

Other surveys target specific applications of knowledge graphs. For instance, knowledge reasoning over KGs is studied by Chen et al. [8], who survey symbolic, statistical, and neural reasoning methods. Biomedical KGC is reviewed by Nicholson and Greene [9]. This review discusses various approaches for constructing knowledge graphs in the biomedical domain, including manual curation and text mining systems. It also explores the applications of biomedical knowledge graphs in solving complex biomedical problems. In [10], authors provide a comprehensive survey on semantic data integration and querying, where KG and their construction seem to be a key method.

Some other works classify the technical architecture of KGC into two principal paradigms: top-down and bottom-up, along with the hybrid approach.

For instance, the survey authored by [11] offers a systematic review of approaches to building knowledge graphs using machine learning. In particular, it distinguishes between top-down (based on existing ontologies) and bottom-up (based on information extraction from raw data) methods, while exploring the integration of hybrid approaches.

In [12], Zuo et al. adopt a bottom-up approach to construct a journal knowledge graph along with an intelligent question answering system that integrates the proposed journal knowledge graph with LLMs. The proposed system enhances response accuracy through retrieval-augmented generation.

In [13] Tamašauskaitė et al., having carried out a systematic review on KGC, propose a general framework for the development of knowledge graphs, integrating top-down (conceptual modeling, schema elaboration) and bottom-up (data extraction, mapping) approaches to suggest a unified approach to knowledge graph development and provide guidance for both researchers and practitioners when constructing and managing knowledge graphs.

In [14], the StraKG was constructed leveraging a hybridization of top-down and bottom-up approaches. The upper schema layer leveraging the top-down approach is based on knowledge of the geological domain in existing geological dictionaries and ontologies, while the instance layer uses a bottom-up approach to extract instance triplets like strata, rocks and geologic ages from existing open data.

Our motivation is to cumulate the advantages of bottom-up and top-down approaches in a hybrid approach designing a knowledge graph construction method that offers both scalability and semantic consistency. Our proposal, detailed in the next section, aims to leverage data-driven patterns in the constructed KG, while ensuring consistency with domain ontologies. This is particularly interesting in complex and evolving domains such as healthcare.

3. A hybrid construction approach of dynamic Knowledge Graphs

In this section, we present our proposed dynamic knowledge graph hybrid construction approach (cf. Figure 1). The proposed hybrid approach leverages the strengths of both top-down and bottom-up methods. It starts with a well-defined ontology (top-down) to establish a clear semantic structure and then enriching this structure with automatically extracted data from heterogeneous sources (bottom-up). It aims at offering a flexible framework for constructing comprehensive dynamic knowledge graphs, and enhancing overall accuracy and adaptability to different data structures (structured, semi-structured and unstructured data). It reduces the rigidity and manual effort required by a solely used top-down approach, by allowing for dynamic incorporation of new information coming from new data (bottom-up), while ensuring consistency and reducing the introduction of irrelevant or redundant data by anchoring data to an established schema (the ontology). The following subsections detail our approach.

3.1. Phase 1: Top-down ontology construction

In this phase the domain ontology is defined. The goal is to establish a semantic foundation for the KG. A top-down process is leveraged, which is composed of two main steps:

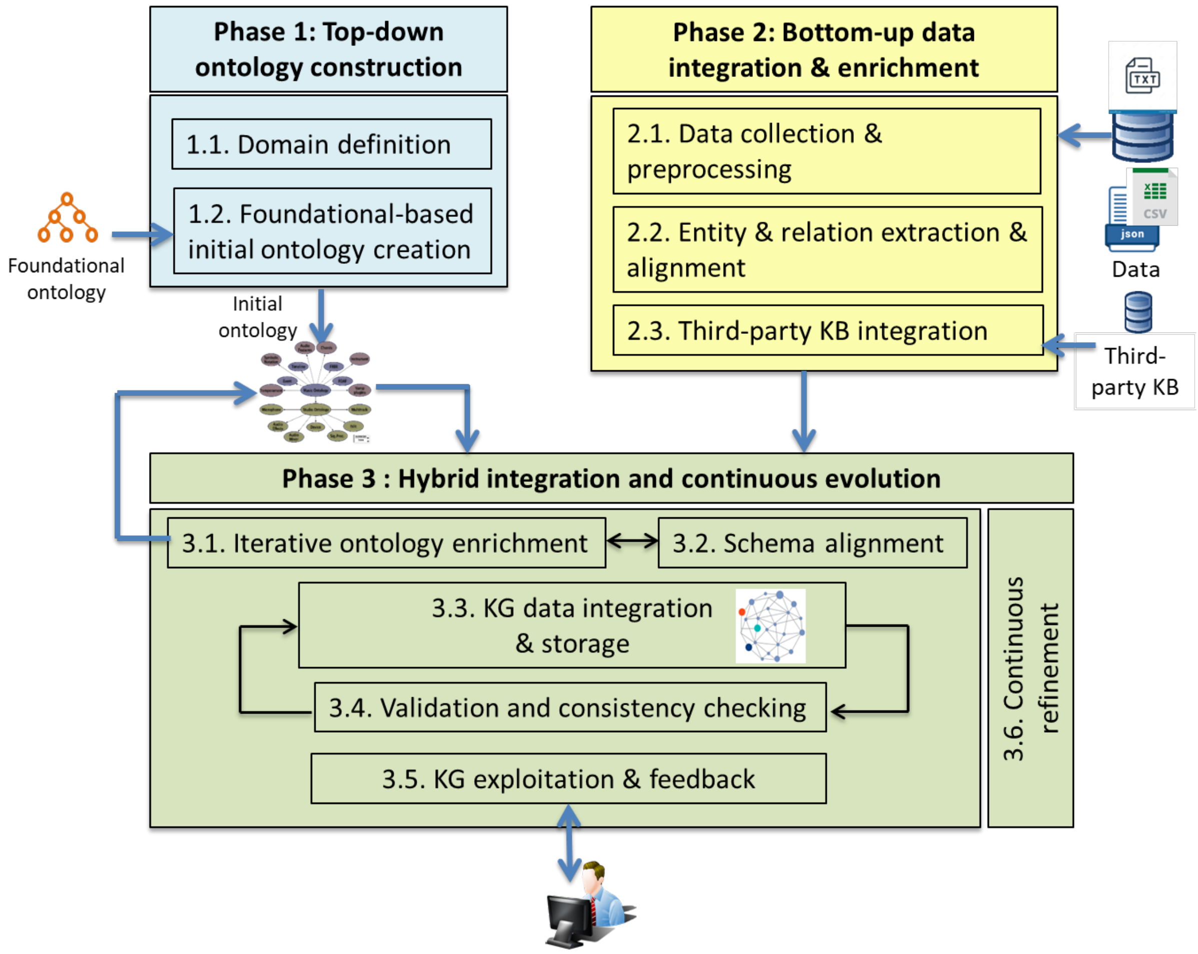

Figure 1: A hybrid construction approach of dynamic Knowledge Graphs

1. Domain definition :
    - Define the domain, objectives and scope of the knowledge graph.
    - Collaborate with domain experts to identify key concepts and relationships.
2. Creation of a foundational-based initial ontology :
    - Building an initial ontology using a top-down approach, based on a foundational ontology like BFO to ensure interoperability with existing ontologies and to allow different levels of domain abstraction/generalization and specialisation.
    - Define key concepts, classes, properties, and relationships.
    - Reuse well-known ontologies and standards to ensure interoperability.

3.2. Phase 2: Bottom-up data integration and enrichment

The second phase of our proposal leverages a bottom-up process. Its primary goal is to integrate structured, semi-structured, and unstructured data from heterogeneous sources. It is composed of three main steps:

1. Data collection and preprocessing :
    - Determine heterogeneous data sources, with structured, semi-structured, and unstructured data.
    - Process the raw data : cleaning, appropriate data transformation, feature engineering, outlier detection.

2. Entity and relation extraction and alignment
   - Extract entities and relationships using NLP, machine learning models, and deep learning models.
   - Resolve duplicates and align entities from multiple sources
   - Use feature matching, machine learning models to enhance alignment accuracy.
3. Third-party Knowledge Base (KB) integration :
   - Integrate external data sources such as DBpedia or Wikidata to enrich the dataset and improve alignment accuracy.

3.3. Phase 3 : Hybrid integration and continuous evolution

This third phase leverages a hybrid process. It merges the semantic schema obtained in the top-down phase (the ontology) with the bottom-up extracted data. It aligns the schema of the knowledge graph and enriches the ontology iteratively, seeking for consistency and continual evolution of the obtained KG. It is composed of six main steps :

1. Iterative Ontology Enrichment
   - The entities and relationships extracted in phase 2 are used to iteratively enrich the initial ontology.
   - The obtained ontology is regularly checked for consistency and accuracy.
2. Schema alignment
   - Align the extracted entities and relationships with the semantic schema (the ontology).
3. Knowledge graph data integration and storage
   - Construct the Knowledge Graph with the extracted and aligned entities and relationships.
   - Store the obtained KG using appropriate solutions (RDF-based stores, graph databases, etc.).
4. Validation and consistency checking :
   - Check the KG consistency
   - Infer new knowledge.
5. KG exploitation and feedback
   - Set up SPARQL endpoints and develop custom APIs for data retrieval.
   - Implement semantic retrieval techniques to enhance query results.
   - Implement interactive visualization tools to help users explore the Knowledge Graph.
   - Collect user feedback
6. Continuous refinement
   - Regularly update and refine the Knowledge Graph based on new data and business requirements and domain changes.
   - refine and validate iteratively the ontology and maintain data alignment.
   - Introduce a feedback collection mechanism to continually maintain and enhance the Knowledge Graph's accuracy and relevance.

4. Case Study: Constructing a Medical Knowledge Graph from Textual Input

To demonstrate the applicability of our proposed hybrid construction approach, we present a case study focused on constructing a medical knowledge graph from unstructured textual input. This case study illustrates the way our hybrid approach can be effectively used for the construction of a knowledge graph.

4.1. Phase 1: Top-down ontology construction

The top-down phase aims to develop an ontology that captures the structure of key concepts in an oncology. This will serve as the foundation for the knowledge graph and ensures that all data is consistently categorized. In this case study we have reused the ontology developed in [15]. It includes the following main concepts and relations:

- **Concepts:** Patient, Disease, Symptom, Treatment, Side Effect, Doctor.
- **Relationships:**
  - Patient suffers from Disease
  - Disease is associated with Symptom
  - Disease is treated by Treatment
  - Treatment may cause Side Effect
  - Patient is consulted by Doctor

This ontology provides a semantic framework that clearly defines how different healthcare entities are related and how they interact with each other, and it will be used for structuring the data extracted from unstructured sources.

4.2. Phase 2: Bottom-up data integration and enrichment

The following natural language sentence has been used as the unstructured input for bottom-up extraction:

> "Alice Dupont suffers from diabetes and takes Metformin, which may cause stomach pain.
> She consults Dr. Martin at Hospital Saint-Louis in Paris."

4.2.1. Entity and Relation Extraction

Using Natural Language Processing (NLP) techniques such as Named Entity Recognition (NER) and Relation Extraction [16], the entities and relations summarized in Table 1 were extracted and aligned with the ontology defined in phase 1:

Table 1
Extracted and Aligned Triples

Subject	**Predicate**	**Object**	**Ontology Mapping**
Alice Dupont	suffers from	diabetes	Patient → Disease
Alice Dupont	takes	Metformin	Patient → Treatment
Metformin	may cause	stomach pain	Treatment → Side Effect
Alice Dupont	consults	Dr. Martin	Patient → Doctor
Dr. Martin	works at (inferred)	Hospital Saint-Louis	(Out-of-ontology)
Hospital Saint-Louis	located in (inferred)	Paris	(Out-of-ontology)

4.3. Phase 3: Hybrid Integration and KG Construction

The extracted instances were aligned with the top-down ontology and integrated into a unified knowledge graph. During this phase, the system identified two out-of-scope concepts, *Hospital* and *Location (City)*, which are not initially defined in the ontology. These can be proposed for ontology enrichment in subsequent iterations.

The resulting RDF triples (abridged) are shown below:

```
:Alice_Dupont rdf:type :Patient .
:Diabetes rdf:type :Disease .
:Alice_Dupont :suffersFrom :Diabetes .
:Metformin rdf:type :Treatment .
:Alice_Dupont :treatedBy :Metformin .
:Metformin :mayCause :Stomach_Pain .
:Stomach_Pain rdf:type :SideEffect .
:Dr_Martin rdf:type :Doctor .
:Alice_Dupont :isConsultedBy :Dr_Martin .
:Dr_Martin :worksAt :Hospital_Saint_Louis .
:Hospital_Saint_Louis :locatedIn :Paris .
```

4.3.1. Validation and Reasoning

To ensure semantic integrity, reasoning tools were used to: 1) Verify class and property usage matches the ontology, 2) Infer additional facts (e.g., linking Metformin to Diabetes via treatment reasoning), 3) Detect inconsistencies (e.g., incorrect domains or disjoint classes, if any).

This case study demonstrates the effectiveness of the hybrid approach in dynamically building a knowledge graph from natural language statements. The top-down phase ensures that all extracted entities and relationships are consistent with the initial ontology. The bottom-up enrichment component allows for the dynamic integration of new information extracted from unstructured text, enhancing the graph's adaptability and evolution. Furthermore, the approach supports ontology evolution by identifying and incorporating missing domain concepts, such as *Hospital* and *City* in our use case, through iterative refinement. Lastly, the integration of external knowledge bases (e.g., DBpedia, UMLS) enhances entity linking and disambiguation, thereby improving the overall completeness and reliability of the constructed knowledge graph.

5. Conclusion

This paper presents a novel hybrid three-phase method for constructing knowledge graphs. Hybridizing top-down and bottom-up knowledge graph construction approaches aims to bring cumulate advantages of both methods, namely having a semantic anchor through the ontology built using a top-down approach (phase 1), and taking advantage of rich and evolving data through the entities and relations extraction from raw data leveraging a bottom-up process (phase 2). A third phase builds the knowledge graph, benefiting from the results of the two previous phases while taking care of the iterative schema alignment, the continuous initial ontology enrichment, and the validation and consistency checking.

The relevance of our approach is demonstrated through a case study in which the construction of a dynamic knowledge graph is of particular interest: the medical field.

Our future work aims to experiment with this approach on real use cases to demonstrate how it can improve graph completeness and the relevance of extracted or inferred information. We also plan to apply concrete evaluation methods based on performance metrics and user validation, along with scalability studies.

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] Schema-adaptable Knowledge Graph Construction, Association for Computational Linguistics, 2023. URL: https://aclanthology.org/2023.findings-emnlp.425.

[2] H. Ye, N. Zhang, H. Chen, H. Chen, Generative knowledge graph construction: A review, in: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2022, pp. 1–17. URL: https://aclanthology.org/2022.emnlp-main.1/. doi:10.18653/v1/2022.emnlp-main.1.

[3] S. Ji, S. Pan, E. Cambria, P. Marttinen, P. S. Yu, A survey on knowledge graphs: Representation, acquisition, and applications, IEEE Transactions on Neural Networks and Learning Systems 33 (2022) 494–514.

[4] X. Zhu, Y. Liu, Z. Wang, Z. Zhang, M. Sun, Multi-modal knowledge graph construction and application: A survey, arXiv preprint arXiv:2202.05786 (2022).

[5] L. Zhong, J. Wu, Q. Li, H. Peng, X. Wu, A comprehensive survey on automatic knowledge graph construction, ACM Computing Surveys 56 (2023). URL: https://doi.org/10.1145/3618295. doi:10.1145/3618295.

[6] T. Safavi, D. Koutra, Relational world knowledge representation in contextual language models: A review, arXiv preprint arXiv:2104.05837 (2021).

[7] S. Choi, Y. Jung, Knowledge graph construction: Extraction, learning, and evaluation, Applied Sciences 15 (2025) 3727. URL: https://doi.org/10.3390/app15073727. doi:10.3390/app15073727.

[8] X. Chen, Y. Chen, Z. Zhang, Y. Chen, Z. Wu, Knowledge reasoning over knowledge graph: A survey, Expert Systems with Applications 138 (2019) 112760.

[9] D. Nicholson, C. S. Greene, Constructing knowledge graphs and their biomedical applications, Computational and Structural Biotechnology Journal 18 (2020) 1414–1428.

[10] M. Masmoudi, S. Ben Abdallah Ben Lamine, M. H. Karray, B. Archimede, H. Baazaoui Zghal, Semantic data integration and querying: A survey and challenges 56 (2024).

[11] Z. Zhao, X. Luo, M. Chen, L. Ma, A survey of knowledge graph construction using machine learning, CMES - Computer Modeling in Engineering & Sciences 139 (2023) 225–257. URL: https://doi.org/10.32604/cmes.2023.031513. doi:10.32604/cmes.2023.031513.

[12] J. Zuo, J. Niu, Construction of journal knowledge graph based on deep learning and llm, Electronics 14 (2025) 1728. URL: https://doi.org/10.3390/electronics14091728. doi:10.3390/electronics14091728.

[13] G. Tamašauskaitė, P. Groth, Defining a knowledge graph development process through a systematic review, ACM Trans. Softw. Eng. Methodol. 32 (2023). URL: https://doi.org/10.1145/3522586. doi:10.1145/3522586.

[14] W. Li, X. Ma, X. Wang, L. Wu, S. Salati, Z. Xie, A hybrid knowledge graph for efficient exploration of lithostratigraphic information in open text data, Applied Computing and Geosciences 22 (2024) 100164. URL: https://www.sciencedirect.com/science/article/pii/S2590197424000119. doi:10.1016/j.acags.2024.100164.

[15] M. Radaoui, S. Ben Abdallah Ben Lamine, H. Baazaoui Zghal, C. Ghedira Guegan, N. Kabachi, Knowledge guided integration of structured and unstructured data in health decision process, in: A. Siarheyeva, C. Barry, M. Lang, H. Linger, C. Schneider (Eds.), Information Systems Development: Information Systems Beyond 2020 (ISD2019 Proceedings), ISEN Yncréa Méditerranée, Toulon, France, 2019. URL: https://aisel.aisnet.org/isd2014/proceedings2019/ManagingISD/7/.

[16] A. Hammami, S. Ben Abdallah Ben Lamine, H. Baazaoui, Bert-based semantic relations extraction from large-scale medical datasets, in: Proceedings of the 2024 IEEE International Conference on Big Data (BigData), IEEE, Washington, DC, USA, 2024, pp. 6460–6468. doi:10.1109/BigData62323.2024.10825162.