

# Leveraging Language Models for Generating Ontologies of Research Topics

Alessia Pisu<sup>1,\*</sup>, Livio Pompianu<sup>1</sup>, Angelo Salatino<sup>2</sup>, Francesco Osborne<sup>2,3</sup>,  
Daniele Riboni<sup>1</sup>, Enrico Motta<sup>2</sup> and Diego Reforgiato Recupero<sup>1</sup>

<sup>1</sup>Department of Mathematics and Computer Science, University of Cagliari, IT

<sup>2</sup>Knowledge Media Institute, The Open University, UK

<sup>3</sup>Department of Business and Law, University of Milano Bicocca, IT

## Abstract

The current generation of artificial intelligence technologies, such as smart search engines, recommendation systems, tools for systematic reviews, and question-answering applications, plays a crucial role in helping researchers manage and interpret scientific literature. Taxonomies and ontologies of research topics are a fundamental part of this environment as they allow intelligent systems and scientists to navigate the ever-growing number of research papers. However, creating these classifications manually is an expensive and time-consuming process, often resulting in outdated and coarse-grained representations. Consequently, researchers have been focusing on developing automated or semi-automated methods to create taxonomies of research topics. This paper studies the application of transformer-based language models for generating research topic ontologies. Specifically, we have developed a model leveraging SciBERT to identify four semantic relationships between research topics (*supertopic*, *subtopic*, *same-as*, and *other*) and conducted a comparative analysis against alternative solutions. The preliminary findings indicate that the transformer-based model significantly surpasses the performance of models reliant on traditional features.

## Keywords

research topics, ontology generation, language models, knowledge graph generation, SciBERT

## 1. Introduction

The current generation of artificial intelligence technologies, such as smart search engines, recommendation systems, tools for systematic reviews, and question-answering applications, plays a crucial role in helping researchers explore and interpret scientific literature [1]. However, managing the vast amount of scientific literature, which increases by approximately 2.5 million papers each year [2], still presents a significant challenge. Large language models have revolutionised the field of natural language processing [3, 4], but still struggle to process a large quantity of text. While they can answer questions about specific papers, they struggle to understand the broader context of a research area covering millions of papers.

To tackle this issue, it was proposed to develop structured and formal representations of the content of research publications, which could be more easily ingested by AI systems [5, 6]. We thus saw the release of several knowledge graphs (KG) [7] that describe the metadata of research publications (e.g., SemOpenAlex [8], AIDA-KG [9]) as well as KGs that focus on the content of

---

*Text2KG 2024: International Workshop on Knowledge Graph Generation from Text May 28, 2024, Creete, GR*

\*Corresponding author.



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0)

these publications and describe their key entities and concepts (e.g., ORKG [10], AI-KG [11], CS-KG [12], Nano-publications [13], SoftwareKG [14]). The community also produced various ontologies for annotating scholarly data [15, 16, 17].

The *research topic* is the most fundamental dimension for describing the concepts within a research paper and thus enabling a more comprehensive analysis of the literature [18]. Therefore, taxonomies and ontologies of research topics (e.g., MeSH, UMLS, CSO, NLM) are essential for organizing and querying academic information. They also provide a foundational structure that enables intelligent systems to navigate and interpret academic literature effectively [19, 20]. This includes search engines [20], conversational agents [21], analytics dashboards [22], academic recommender systems [19], and many other tools in this space. A solid representation of research topics is also the foundation for many AI-driven literature analyses [23, 24].

Manually constructing ontologies of research topics is an expensive and time-consuming process, often resulting in outdated and coarse-grained representations [25]. Consequently, researchers have been focusing on developing automated or semi-automated methods to create these taxonomies [25, 26, 27]. A notable example of this approach is Klink-2 [26], which has been used to produce the Computer Science Ontology (CSO) [16]. CSO is one of the largest resources in the field, including about 14K topics and 159K semantic relationships. It has been adopted by various organizations, including Springer Nature [28], to annotate research articles, course materials, software, and videos.

This paper initiates an investigation into the application of transformer-based language models [29] for generating research topic ontologies. Our primary objective, which we aim to pursue in future work, is to develop an innovative method for generating taxonomies of research topics that will effectively incorporate language model technology. The resulting approach will be used both to update CSO and to construct large-scale ontologies across various scientific disciplines. As a first step, we have developed a model leveraging SciBERT [30] to identify four semantic relationships between research topics (*supertopic*, *subtopic*, *same-as*, and *other*) and conducted a comparative analysis against traditional feature-based solutions [25, 26]. The models were trained and evaluated on a large section of CSO manually validated by domain experts. The preliminary findings indicate that the transformer-based model significantly surpasses the performance of models reliant on traditional features. To ensure reproducibility, we make available an (anonymous) repository with the gold standard and the codebase<sup>1</sup>.

The remainder of this paper is organised as follows. Section 2 provides a review of taxonomies in computer science, along with current approaches for their (semi-)automatic generation. Section 3 introduces the two main methodologies tested in this study for automatically generating ontologies of research topics. Section 4 reports the preliminary evaluation, and Section 5 outlines the future directions we intend to pursue.

## 2. Related Work

In this section, we delve into the literature concerning the evolution and utilization of research area ontologies, as well as the methodologies employed for their automated generation.

---

<sup>1</sup>Gold standard and code - <https://anonymous.4open.science/r/LeveragingLMforGeneratingOntologies-2107/>

## 2.1. Taxonomies in Computer Science

In the field of Computer Science, the ACM Computing Classification System<sup>2</sup> is a well-known taxonomy of research topics. It is developed and maintained by the Association for Computing Machinery (ACM), the world's largest educational and scientific computing society, and covers about 2K research topics. It is manually curated, which makes its update process laborious and costly. Consequently, this taxonomy undergoes infrequent updates, with the latest one occurring in 2012, and becomes quickly outdated.

The Computer Science Ontology (CSO), discussed in the introduction, is one of the largest topic classifications, covering 14K research areas [31]. It has been automatically generated using the Klink-2 algorithm [26] on a dataset of 16 million scientific articles. Different from alternative solutions, CSO offers two main advantages over alternative solutions: i) it provides a very fine-grained representation of the field, rendering all the nuances of the area, and ii) it can be easily updated by executing Klink-2 on recent corpora of publications. CSO serves as the backbone for several tools utilised by the editorial team at Springer Nature, contributing to diverse applications such as research publication classification, identification of research communities, and forecasting research trends [16].

The IEEE Taxonomy mainly covers the field of Engineering but also contains different concepts relevant to computer science. It was developed and maintained by the Institute of Electrical and Electronics Engineers<sup>3</sup> (IEEE). It supports the organisation of the Electrical and Electronics Engineering field, providing a standardised framework for classifying academic publications, research topics, and technical content within the IEEE's publications and databases. It contains around 5.6K topics and 24K relationships. The IEEE Taxonomy is also manually curated with minor updates released yearly.

In this paper, we will focus on CSO, as it represents the most extensive taxonomy in the field of computer science. Additionally, it includes sections that have undergone manual verification, making them suitable for use as a gold standard.

## 2.2. Ontology Generation

The review of existing literature reveals a variety of both semi-automatic and fully automatic approaches for the generation of ontologies and taxonomies. The initial step in formulating an ontology involves the identification of its underlying topics. In order to expedite this process, research is currently underway to develop automatic methods. For example, BERT [32] was used in [33] to solve the topic extraction task. Ontology extraction methods were traditionally based on natural language processing, clustering techniques, or statistical methods [34, 35]. For example, Text2Onto [34] is a framework designed to learn ontologies from a collection of documents. This method identifies synonyms, sub-/superclass hierarchies, and more through the application of natural language processing techniques on sentence structures, leveraging phrases such as "such as..." and "and other..." to imply hierarchies between terms.

Shan et al. [36] applied a variation of this technique to generate Fields of Study (FoS) for Microsoft Academic [36], incorporating both hand-crafted concepts (first two levels) and topics

---

<sup>2</sup>The ACM Computing Classification System – <http://www.acm.org/publications/class-2012>

<sup>3</sup>IEEE Taxonomy - <https://www.ieee.org/content/dam/ieee-org/ieee/web/org/pubs/ieee-taxonomy.pdf>

automatically derived from Wikidata. However, this taxonomy learning approach focuses on Wikidata and does not leverage metadata associated with research papers. The OpenAlex team adopted a similar strategy [37], by employing the ASJC structure in Scopus and augmenting it with topics drawn from the papers using citation analysis.

Other approaches included the combination of ontology learning and crowdsourcing strategies, integrating statistical measures and user opinions [38, 39]. For instance, Wohlgenannt et al. [38] merged human effort and machine computation by crowdsourcing the evaluation of an automatically generated ontology, aiming to dynamically validate the extracted relations.

Lately, the community has started to work towards leveraging LLMs for the creation of taxonomies, ontologies, and KGs [40]. For instance, Chen et al. [41] proposed an approach for taxonomy generation that consists of two modules: the first predicts parenthood relations and the other reconciles these predictions into trees. The parenthood prediction module generates likelihood scores for potential parent-child pairs, forming a graph of parent-child relation scores. The tree reconciliation module approaches the task as a graph optimisation problem, yielding the maximum spanning tree of this graph. The model is trained on subtrees sampled from Wordnet and tested on non-overlapping Wordnet subtrees.

To the best of our knowledge, specific methodologies employing language models for generating ontologies of research topics have not yet been established.

### 3. Methodology

This section outlines two main approaches for identifying the relationship between two research topics. As discussed in the introduction, this is the key component of a system for generating ontologies of research topics [26]. First, we describe and formalize the task (Section 3.1) and the dataset (Section 3.2). Then, we present a feature-based approach that uses a variety of traditional features adopted by the state-of-the-art methods (Section 3.3) and a transformer-based approach that employs the SciBERT model (Section 3.4).

#### 3.1. Task Definition

The addressed task is the identification of the relationship between two research topics. More formally, given a pair of topics  $(t_A, t_B)$ , we employ a single-label multi-class classification model to determine the specific semantic relationship between them. Naturally, various categories can be defined based on the specific predicates that need representation. For this paper, we have chosen three essential predicates from the CSO schema.

Therefore, we aim to classify the relationship between two topics according to four classes:

- *supertopic*:  $t_A$  is an ancestor of  $t_B$ , e.g., *semantic web* is a super area of *rdf*;
- *subtopic*:  $t_A$  is a descendant of  $t_B$ , e.g., *neural networks* is a sub-topic of *machine learning*;
- *same-as*:  $t_A$  and  $t_B$  are two alternative labels for the same topic, e.g., *haptic interface* and *haptic device*;
- *other*:  $t_A$  and  $t_B$  do not fit into any of the aforementioned relationships, e.g., *cryptocurrency* and *particle swarm optimizer*.

### 3.2. Datasets

To conduct the experiments, we relied on two datasets: the Computer Science Ontology (CSO) [16] (introduced in Section 2.1) and the AIDA Knowledge Graph (AIDA-KG) [9]. We used CSO to derive a gold standard and AIDA-KG to compute a set of features that require linking topics to relevant papers (e.g., co-occurrence between two topics).

CSO is made available on a website that allows domain experts to verify and modify the ontology. Therefore, different portions of the ontologies were manually verified and refined over time, often when conducting a specific analysis on certain topics (e.g., Software Engineering [42]). We thus take advantage of these manually verified portions to build a gold standard to train and evaluate the approaches. The CSO data model includes four main semantic relationships: *superTopicOf*: indicating that one topic is a sub-area of another (e.g., Artificial Intelligence is a super-area of Machine Learning); *relatedEquivalent*: denoting that two topics can be considered equivalent for the sake of exploring research data (e.g., Ontology Mapping and Ontology Matching); *contributesTo*: indicates that the research output of one topic contributes to another; *owl:sameAs*: it lists entities from other KGs (e.g., DBpedia, Wikidata) referring to the same concepts.

In order to build the gold standard, we selected 4,713 *superTopicOf* triples and mapped them as *superTopic*. We also selected 3,034 *relatedEquivalent* triples to represent equivalence through the *same-as* relation. Then, we derived 4,713 *subTopic* relationships by reversing the *superTopic* relationships. Finally, we randomly coupled topics to generate 5,151 *other* relationships, ensuring that none of these pairs shared any of the previously mentioned relationships according to the CSO framework.

The resulting gold standard counts 17,611 triples, which have been partitioned into 15,154 triples (~86%) for the training set, 2,166 triples (~12.3%) for the validation set, and 291 triples (~1.7%) for the test set. The test set is intentionally small for two main reasons. First, to prevent data leakage bias, we ensured that none of the couples of topics appearing in a triplet of one set appeared in a triple of another set. For instance, we avoided that a triple  $\langle t_A, \textit{superTopic}, t_B \rangle$  in the training set could appear as  $\langle t_B, \textit{subTopic}, t_A \rangle$  in the test set. Second, we generated the test set so that each triple contains at least one topic that is completely absent from the training set. It is important to note that these adjustments make this test set more challenging than the ones previously used to test Klink [25] and Klink-2 [26].

AIDA-KG [9] is a KG integrating 25 million publications linked to research topics in CSO, researcher profiles, and 66 industrial sectors. We employ this resource to derive the occurrence of the relevant topics across the paper abstracts as well as their co-occurrences. These metrics will be used for our feature-based methods.

### 3.3. Feature-based Method

The task defined in Section 3.1 has been usually tackled by leveraging a variety of numerical features, typically derived from the two topics frequency and common usage [26, 43]. These approaches typically involve combining these features in a mathematical function or with a classifier [26].

We implemented a feature-based classification method that, for each pair of topics ( $t_A, t_B$ ),

leverages four features:

- *occA*: number of times topic A appears in paper abstracts;
- *occB*: number of times topic B appears in paper abstracts;
- *cooccurrenceAB*: number of times both topic A and B simultaneously appear in abstracts;
- *subsumption*: it indicates the degree of overlap between the co-occurring topics, calculated using  $subsumption = \frac{cooccurrenceAB}{occA} - \frac{cooccurrenceAB}{occB}$ .

The initial two features reflect the popularity of a topic. The third feature quantifies how related two topics are, based on their frequency of co-occurrence in research papers. The fourth feature evaluates the presence of a hierarchical relationship between the two topics.

For each triple, we extracted these features by querying the AIDA KG. We normalised these features and then we trained two machine learning models: Gradient Boosting (GB) and Random Forest (RF). These approaches are widely employed and renowned for their strong performance across various domains [44], making them excellent candidates for our task. They are both ensemble models, combining multiple weak learners. We conducted several experiments with both models, varying the number of estimators, ranging from 10 to 3000.

### 3.4. Language Model-based Method

To devise a method leveraging language models we employed SciBERT [30], a model based on BERT [32]. BERT is a widely acclaimed model in natural language processing, renowned for its proficiency in understanding and processing human language. SciBERT extends BERT’s capabilities by specializing in scientific texts, making it an ideal choice for our objectives. Specifically, SciBERT was trained on a large corpus of scientific text, primarily from SemanticScholar. BERT and SciBERT excel in comprehending context and disambiguating polysemous words, demonstrating a human-like common sense in language parsing [32].

To adapt SciBERT for our specific classification task, we undertook a fine-tuning process using the training set described in Section 3.2. To this purpose, we leverage the *scibert-scivocab-uncased* with *Huggingface* [45]. We chose *AdamW* [46] as the optimiser, which is a weighted version of Adam [47] that helps prevent overfitting in large models.

The fine-tuning process involved providing the model with the surface forms of the two topics, separated by a semicolon, as well as the correct relationship class from the training set. In our experiments, we varied the number of epochs (from 1 to 10), while keeping 50 warm-up steps. Our best-performing model was obtained after training for five epochs.

## 4. Evaluation

We evaluated the three methods described in the previous section on the test set outlined in Section 3.2. Specifically, we compared: 1) the *feature-based method using Gradient Boosting*, 2) the *feature-based method using Random Forest*, 3) the *language model-based method leveraging SciBERT*. We assess and compare the performance of the three approaches employing standard metrics for text classification: accuracy, precision, recall, and F-score.

Table 1 reports the experimental results. The language model-based method significantly outperforms the feature-based methods across all metrics, yielding an impressive F1 of 0.9129,

**Table 1**  
Experimental results.

Classifier		Feature-based GB	Feature-based RF	Lang. Model-based
<b>Accuracy</b>		<b>0.5842</b>	<b>0.6426</b>	<b>0.9141</b>
<b>Precision</b>	supertopic	0.5424	0.5634	0.9143
	subtopic	0.4815	0.6200	0.9452
	same-as	0.5167	0.5804	0.9615
	other	0.8621	0.8793	0.8286
	<b>average</b>	<b>0.6007</b>	<b>0.6608</b>	<b>0.9124</b>
<b>Recall</b>	supertopic	0.4211	0.5263	0.8421
	subtopic	0.3421	0.4079	0.9079
	same-as	0.7750	0.8125	0.9375
	other	0.8475	0.8644	0.9831
	<b>average</b>	<b>0.5964</b>	<b>0.6528</b>	<b>0.9177</b>
<b>F-score</b>	supertopic	0.4740	0.5442	0.8767
	subtopic	0.4000	0.4921	0.9262
	same-as	0.6200	0.6771	0.9494
	other	0.8547	0.8718	0.8992
	<b>average</b>	<b>0.5872</b>	<b>0.6463</b>	<b>0.9129</b>

more than a 27% increase compared to the alternatives. Among the feature-based approaches, the Random Forest classifier yields better results across all metrics. The superiority of the language model-based method is especially marked when considering the *superTopic* and *subTopic* relations. Feature-based methods achieve rather poor results in recognizing these relations (i.e., F-score close to 0.5). This underperformance might stem from the presence of at least one unfamiliar topic in each pair within the test set.

Examining the precision/recall tradeoff, the language model-based approach obtains higher precision than recall for three relations, namely *superTopic*, *subTopic*, and *same-as*. On the other hand, in the case of the *other* relationship, the precision is considerably lower than the recall (i.e., 0.8286 vs 0.9831). This discrepancy suggests that the method is prone to overlooking some semantic connections between topic pairs, mistakenly classifying them as unrelated. We plan to further investigate this issue in future work.

## 5. Conclusions

In this paper, we presented a novel SciBERT-based method for identifying the relationship between research topics and conducted a comparative analysis against feature-based solutions. For this purpose, we fine-tuned a SciBERT model using a gold standard of triples derived from CSO. The SciBERT-based model attained an F1 score of 0.9129, marking an improvement of more than 27% compared to methods that utilize numerical features. These findings are significant considering the growing demand from the scholarly community for developing more fine-grained ontologies of research topics that can enhance the characterisation of content within scientific KGs.

In future work, we aim to develop an innovative method for generating taxonomies of research topics to enhance CSO and generate large-scale ontologies across various scientific fields. To

this end, we plan to integrate language models and numerical features by employing knowledge injection techniques [48]. We also intend to conduct experiments with recent large language models, such as Mistral [49] and LLaMa 2 [50]. This evaluation will take into account factors such as cost and environmental impact. Additionally, we intend to study the potential challenges that could arise when extending these techniques to other research domains, including fields like Engineering, Material Science, and Mathematics. Finally, we aim to explore whether a model trained in one discipline, such as Computer Science, can be effectively adapted and applied to a different field and assess the impact of such a cross-disciplinary application.

## Acknowledgments

Alessia Pisu and Livio Pompianu acknowledge MUR and EU-FSE for financial support of the PON Research and Innovation 2014-2020 (respectively D.M. 1061/2021 and D.M 1062/2021 programs). The work of Daniele Riboni was partially supported by the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.5—Project Code ECS0000038—Project Title eINS Ecosystem of Innovation for Next Generation Sardinia. Angelo Salatino, Francesco Osborne, and Enrico Motta gratefully acknowledge the financial support provided by Springer Nature.

## References

- [1] F. Bolanos, A. Salatino, F. Osborne, E. Motta, Artificial intelligence for literature reviews: Opportunities and challenges, arXiv preprint arXiv:2402.08565 (2024).
- [2] L. Bornmann, R. Mutz, Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references, *Journal of the Association for Information Science and Technology* 66 (2015) 2215–2222. URL: <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.23329>. doi:<https://doi.org/10.1002/asi.23329>. arXiv:<https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.23329>.
- [3] T. H. Kung, M. Cheatham, A. Medenilla, C. Sillos, L. De Leon, C. Elepaño, M. Madriaga, R. Aggabao, G. Diaz-Candido, J. Maningo, et al., Performance of chatgpt on usmle: Potential for ai-assisted medical education using large language models, *PLoS digital health* 2 (2023) e0000198.
- [4] OpenAI, Gpt-4 technical report, 2023. arXiv:2303.08774.
- [5] S. Auer, V. Kovtun, M. Prinz, A. Kasprzik, M. Stocker, M. E. Vidal, Towards a knowledge graph for science, in: *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics*, 2018, pp. 1–6.
- [6] T. Kuhn, M. Dumontier, Genuine semantic publishing, *Data Science* 1 (2017) 139–154.
- [7] C. Peng, F. Xia, M. Naseriparsa, F. Osborne, Knowledge graphs: Opportunities and challenges, *Artificial Intelligence Review* (2023) 1–32.
- [8] M. Färber, D. Lamprecht, J. Krause, L. Aung, P. Haase, Semopenalex: The scientific landscape in 26 billion rdf triples, in: *International Semantic Web Conference*, Springer, 2023, pp. 94–112.
- [9] S. Angioni, A. Salatino, F. Osborne, D. R. Recupero, E. Motta, Aida: A knowledge graph



- about research dynamics in academia and industry, *Quantitative Science Studies* 2 (2021) 1356–1398.
- [10] M. Y. Jaradeh, A. Oelen, K. E. Farfar, M. Prinz, J. D’Souza, G. Kismihók, M. Stocker, S. Auer, Open research knowledge graph: next generation infrastructure for semantic scholarly knowledge, in: *Proceedings of the 10th International Conference on Knowledge Capture*, 2019, pp. 243–246.
  - [11] D. Dessi, F. Osborne, D. Reforgiato Recupero, D. Buscaldi, E. Motta, H. Sack, Ai-kg: an automatically generated knowledge graph of artificial intelligence, in: *The Semantic Web–ISWC 2020: 19th International Semantic Web Conference*, Athens, Greece, November 2–6, 2020, *Proceedings, Part II* 19, Springer, 2020, pp. 127–143.
  - [12] D. Dessi, F. Osborne, D. Reforgiato Recupero, D. Buscaldi, E. Motta, Cs-kg: A large-scale knowledge graph of research entities and claims in computer science, in: *International Semantic Web Conference*, Springer, 2022, pp. 678–696.
  - [13] T. Kuhn, C. Chichester, M. Krauthammer, N. Queralt-Rosinach, R. Verborgh, G. Gianakopoulos, A.-C. N. Ngomo, R. Vigiante, M. Dumontier, Decentralized provenance-aware publishing with nanopublications, *PeerJ Computer Science* 2 (2016) e78.
  - [14] D. Schindler, B. Zopilko, F. Krüger, Investigating software usage in the social sciences: A knowledge graph approach, in: *European Semantic Web Conference*, Springer, 2020, pp. 271–286.
  - [15] S. Peroni, D. Shotton, The spar ontologies, in: *The Semantic Web–ISWC 2018: 17th International Semantic Web Conference*, Monterey, CA, USA, October 8–12, 2018, *Proceedings, Part II* 17, Springer, 2018, pp. 119–136.
  - [16] A. A. Salatino, T. Thanapalasingam, A. Mannocci, A. Birukou, F. Osborne, E. Motta, The Computer Science Ontology: A Comprehensive Automatically-Generated Taxonomy of Research Areas, *Data Intelligence* 2 (2020) 379–416. URL: [https://doi.org/10.1162/dint\\_a\\_00055](https://doi.org/10.1162/dint_a_00055). doi:10.1162/dint\_a\_00055. arXiv:[https://direct.mit.edu/dint/article-pdf/2/3/379/1857480/dint\\_a\\_00055.pdf](https://direct.mit.edu/dint/article-pdf/2/3/379/1857480/dint_a_00055.pdf).
  - [17] A. Salatino, F. Osborne, E. Motta, Cso classifier 3.0: a scalable unsupervised method for classifying documents in terms of research topics, *International Journal on Digital Libraries* (2022) 1–20.
  - [18] A. A. Salatino, Early detection of research trends, 2019. URL: <http://oro.open.ac.uk/67224/>. arXiv:1912.08928.
  - [19] J. Beel, B. Gipp, S. Langer, C. Breitinger, Paper recommender systems: a literature survey, *International Journal on Digital Libraries* 17 (2016) 305–338.
  - [20] M. Gusenbauer, N. R. Haddaway, Which academic search systems are suitable for systematic reviews or meta-analyses? evaluating retrieval qualities of google scholar, pubmed, and 26 other resources, *Research synthesis methods* 11 (2020) 181–217.
  - [21] A. Meloni, S. Angioni, A. Salatino, F. Osborne, D. R. Recupero, E. Motta, Integrating conversational agents and knowledge graphs within the scholarly domain, *Ieee Access* 11 (2023) 22468–22489.
  - [22] S. Angioni, A. Salatino, F. Osborne, D. R. Recupero, E. Motta, The aida dashboard: a web application for assessing and comparing scientific conferences, *IEEE Access* 10 (2022) 39471–39486.
  - [23] J. W. Goodell, S. Kumar, W. M. Lim, D. Pattnaik, Artificial intelligence and machine learn-

- ing in finance: Identifying foundations, themes, and research clusters from bibliometric analysis, *Journal of Behavioral and Experimental Finance* 32 (2021) 100577.
- [24] A. Salatino, S. Angioni, F. Osborne, D. R. Recupero, E. Motta, Diversity of expertise is key to scientific impact: a large-scale analysis in the field of computer science, *arXiv preprint arXiv:2306.15344* (2023).
- [25] F. Osborne, E. Motta, Mining semantic relations between research areas, in: *The Semantic Web–ISWC 2012: 11th International Semantic Web Conference*, Boston, MA, USA, November 11–15, 2012, Proceedings, Part I 11, Springer, 2012, pp. 410–426.
- [26] F. Osborne, E. Motta, Klink-2: Integrating multiple web sources to generate semantic topic networks, in: M. Arenas, O. Corcho, E. Simperl, M. Strohmaier, M. d’Aquin, K. Srinivas, P. Groth, M. Dumontier, J. Heflin, K. Thirunarayan, K. Thirunarayan, S. Staab (Eds.), *The Semantic Web - ISWC 2015*, Springer International Publishing, Cham, 2015, pp. 408–424.
- [27] K. Han, P. Yang, S. Mishra, J. Diesner, Wikicssh: extracting computer science subject headings from wikipedia, in: *ADBIS, TPD and EDA 2020 Common Workshops and Doctoral Consortium: International Workshops: DOING, MADEISD, SKG, BBIGAP, SIM-PDA, AIMinScience 2020 and Doctoral Consortium*, Lyon, France, August 25–27, 2020, Proceedings 24, Springer, 2020, pp. 207–218.
- [28] F. Osborne, A. Salatino, A. Birukou, E. Motta, Automatic classification of springer nature proceedings with smart topic miner, in: *The Semantic Web–ISWC 2016: 15th International Semantic Web Conference*, Kobe, Japan, October 17–21, 2016, Proceedings, Part II 15, Springer, 2016, pp. 383–399.
- [29] K. S. Kalyan, A. Rajasekharan, S. Sangeetha, Ammus: A survey of transformer-based pretrained models in natural language processing, *arXiv preprint arXiv:2108.05542* (2021).
- [30] I. Beltagy, K. Lo, A. Cohan, Scibert: A pretrained language model for scientific text, 2019. *arXiv:1903.10676*.
- [31] A. A. Salatino, T. Thanapalasingam, A. Mannocci, F. Osborne, E. Motta, The computer science ontology: a large-scale taxonomy of research areas, in: *The Semantic Web–ISWC 2018: 17th International Semantic Web Conference*, Monterey, CA, USA, October 8–12, 2018, Proceedings, Part II 17, Springer, 2018, pp. 187–205.
- [32] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2019. *arXiv:1810.04805*.
- [33] M. Grootendorst, Bertopic: Neural topic modeling with a class-based tf-idf procedure, 2022. *arXiv:2203.05794*.
- [34] P. Cimiano, J. Völker, Text2onto, in: A. Montoyo, R. Muñoz, E. Métais (Eds.), *Natural Language Processing and Information Systems*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 227–238.
- [35] M. Le, S. Roller, L. Papaxanthos, D. Kiela, M. Nickel, Inferring concept hierarchies from text corpora via hyperbolic embeddings, in: A. Korhonen, D. Traum, L. Màrquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Florence, Italy, 2019, pp. 3231–3241. URL: <https://aclanthology.org/P19-1313>. doi:10.18653/v1/P19-1313.
- [36] Z. Shen, H. Ma, K. Wang, A web-scale system for scientific knowledge exploration, in: F. Liu, T. Solorio (Eds.), *Proceedings of ACL 2018, System Demonstrations*, Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 87–92. URL:

<https://aclanthology.org/P18-4015>. doi:10.18653/v1/P18-4015.

- [37] OpenAlex, Openalex: End-to-end process for topic classification, ??? URL: <https://docs.google.com/document/d/1bDopkhuGieQ4F8gGNj7sEc8WSE8mvLZS/edit>.
- [38] G. Wohlgenannt, A. Weichselbraun, A. Scharl, M. Sabou, Dynamic integration of multiple evidence sources for ontology learning, *Journal of Information and Data Management* 3 (2012) 243–254.
- [39] J. Mortensen, M. Musen, N. Noy, Crowdsourcing the verification of relationships in biomedical ontologies, *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium 2013* (2013) 1020–9.
- [40] B. P. Allen, L. Stork, P. Groth, Knowledge engineering using large language models, *arXiv preprint arXiv:2310.00637* (2023).
- [41] C. Chen, K. Lin, D. Klein, Constructing taxonomies from pretrained language models, in: *North American Chapter of the Association for Computational Linguistics*, 2020. URL: <https://api.semanticscholar.org/CorpusID:233992529>.
- [42] F. Osborne, H. Muccini, P. Lago, E. Motta, Reducing the effort for systematic reviews in software engineering, *Data Science* 2 (2019) 311–340.
- [43] M. Sanderson, B. Croft, Deriving concept hierarchies from text, in: *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '99*, Association for Computing Machinery, New York, NY, USA, 1999, p. 206–213. URL: <https://doi.org/10.1145/312624.312679>. doi:10.1145/312624.312679.
- [44] A. Mohammed, R. Kora, A comprehensive review on ensemble deep learning: Opportunities and challenges, *Journal of King Saud University-Computer and Information Sciences* 35 (2023) 757–774.
- [45] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Brew, Huggingface’s transformers: State-of-the-art natural language processing, *CoRR abs/1910.03771* (2019). URL: <http://arxiv.org/abs/1910.03771>. arXiv:1910.03771.
- [46] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, 2019. arXiv:1711.05101.
- [47] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2017. arXiv:1412.6980.
- [48] A. Cadeddu, A. Chessa, V. De Leo, G. Fenu, E. Motta, F. Osborne, D. Reforgiato Recupero, A. Salatino, L. Secchi, A comparative analysis of knowledge injection strategies for large language models in the scholarly domain, *Engineering Applications of Artificial Intelligence* 133 (2024) 108166. URL: <https://www.sciencedirect.com/science/article/pii/S0952197624003245>. doi:<https://doi.org/10.1016/j.engappai.2024.108166>.
- [49] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. d. l. Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, et al., Mistral 7b, *arXiv preprint arXiv:2310.06825* (2023).
- [50] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, et al., Llama 2: Open foundation and fine-tuned chat models, *arXiv preprint arXiv:2307.09288* (2023).