Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 05 Issue 01
2023
Research Article
https://doi.org/10.56979/501/2023

Generative Knowledge Graph Construction of In-Text Citation Data

Maryam Salam[1], and Imran Ihsan[1,*]

[1]Department of Creative Technologies, Air University, 44000, Pakistan.
[*]Corresponding Author: Imran Ihsan. Email: iimranihsan@gmail.com.

Received: April 22, 2023 **Accepted:** June 03, 2023 **Published:** June 05, 2023.

Abstract: Recent years have seen an increase in the use of the phrase ”knowledge graph” in academic and professional circles, frequently in conjunction with Semantic Web technologies, linked data, massive data analytics, and cloud computing. Though there has been a marked increase in the availability of scholarly information online in recent decades, all scholarly discourse continues to be conducted through the written word. Scholarly knowledge is difficult to process mechanically in this format. In this research, an extensive dataset is used which is composed of 8,700 academic scholar (research papers sentences). The proposed approach consist of multiple steps; data preprocessing, entity extraction, relationship extraction and knowledge graph construction. We propose a more efficient representation of a scalable knowledge graph by instantly extracting the information from corpus of ACL dataset, and we test whether a knowledge graph can be used as an effective application in analyzing and generating knowledge representation from the extracted corpus of research citations.

Keywords: Knowledge Graph; In-text Citation; Data Mining; Entity Recognition.

1. Introduction

There has been a massive and ongoing expansion of online data in recent years. Numerous formats for these statistics have been made public (e.g., CSV, JSON, and HTML, among others). Open data portals enable public institutions to provide a central location from which data may be accessible by a wide range of interested parties, and this trend toward increased data volume is partially attributable to open data laws and efforts (e.g., citizens, developers, and other public institutions). The modern era is characterized by the emergence of both massive data and increased processing capacity. There are so many bright prospects in the field of artificial intelligence right now. With the advent of new technologies, Computer Vision has advanced tremendously. By employing a sophisticated neural network, the computer can do object recognition. Similar efforts are being made to enhance the computer’s text comprehension. Wikipedia data and other large-structured data sets are largely responsible for the progress. But text comprehension and text extraction from the web remain challenging tasks. The need for expressive formalisms to combine factual knowledge dispersed across diverse data sources, and knowledge graphs have gained traction as the most relevant exponent of the usage of Semantic Web technology [1]. Graph analytics, question-and-answer systems, web search, etc. all rely heavily on knowledge graphs as a resource.

According to the authors [2-5], a knowledge graph is a multi-relational graph with entities as nodes and relations as different sorts of edges, and it is used to define and store facts as triples. Recent years have seen an increase in the number of initiatives (such as NELL, DBpedia, and Deep Dive) aimed at creating knowledge graphs that computers can understand and use. Large-scale knowledge bases (KBs) like DBpedia, Yago, and Freebase have been developed in recent years [6-8]; these KBs are frequently drawn from Wikipedia and are commonly referred to as KGs.

These KGs are completely machine-readable and include true knowledge about real-world entities, including their relations and properties. Recent years have witnessed a rebirth of information-rich techniques for many tasks in IR, particularly in Web Search, and Natural Language Processing because of the availability of these knowledge-based resources [9]. The entity boxes presented by most commercial search engines are a famous instance of IR. These compartments, which are typically displayed on the screen's right side, offer a more organized perspective on entities (such as people). Wikipedia serves as both an entity repository and a source for the information displayed there; this includes details such as a person's birth date and links to the person's offspring. The contribution of this research is the construction of KG using In-Text Citation Data. More formally, the research focuses on

- Utilizing the in-text citation sentences from ACL dataset [19, 21]
- Construct a Generative Knowledge Graph from the in-text citations
- Define optimal patterns to extract the relationship of entities.

Though the volume of available academic data continues to expand tremendously, finding high-quality sources can be time-consuming and difficult for researchers. Despite several attempts [10, 11, 12], extract high-quality article relationships from in-text citation to formulate knowledge graphs is still a challenging task. The ability to extract links between high-quality articles is becoming increasingly valuable, and it may help to regain cognitive connections. The implementation of a framework that facilitates the interchange of bibliographic and citation context data in a machine-readable format can be highly beneficial. A proper statement of research could read as follows, "The academic community lacks a unified way to depict Cognitive Relationships that would allow for the retrieval of high-quality research publications. It is necessary to extract and semantically describe such associations in a machine-understandable way with high accuracy and recall. Building a knowledge graph using the in-text citation dataset can create a link in academic research to retrieve quality research papers".

The structure of this research is presented here. The literature review of the work is shown in Section 2. The methodology of the proposed work is shown in the Section 3. The dataset used for generative knowledge graph is described in Section 4. Entity Extraction, Relation Extraction and Knowledge Graph building are detailed in Section 5, 6 and 7 respectively. The proposed work is concluded with future work Section 8.

2. Literature Review

For a generative knowledge graph, typically, there are four stages to the process: collecting relevant articles from online news sources, analyzing them, building a knowledge base, and finally a knowledge graph [10]. There are various examples of Scholarly Knowledge Graphs built using the 4-step approach. One example is the AceKG [11], a large-scale knowledge graph in the academic domain, that contains 3.13 billion triples. The information is based on a consistent ontology, such as the properties of publications, authors, fields of study, venues, institutes, and relations between them. The research provides three experimental assessments of AceKG and a comparison with several state-of-the-art methods. Another study [12] offers two strategies for improving the accessibility and re-usability of academic data, published by Springer Nature SciGraph through DBpedia. The author employs methods to improve identification resolution between the two sources using Linking Discovery for the data structure and uses entity recognition (NER) to add links to DBpedia items inside the SN SciGraph's unstructured text content. The procedure organized the content thematically, with links to further information in DBpedia and Wikipedia.

The Open Information Extraction (OpenIE) uses binary relations between the authors to generate KGs [13]. Instead of manual curation, the OpenIE automatically extracts all semantic relational tuples from source text [17]. To achieve this goal, they have developed methods for enhancing the extraction and connection of named items with KG persons, as well as their association with grammatical units that result in the generation of more cohesive facts. Furthermore, they have provided choices to construct potentially helpful RDF triples for the KG. Results have shown that using grammatical structures together with information extraction units improves comprehension of the proposition-based representations in the building of KGs.

Another study [14] employs a statistical method to identify the boundaries of the effect and the patterns of effect events in the graph, regardless of whether they occur in succession or simultaneously. All relevant causality events from texts for the graph building are extracted on many clauses/EDUs (Elementary Discourse Units) which assist in establishing effect-event structures using textual event sequences in documents. After using the verb pair rules to detect causality, the system deals with the effect-boundary determination challenges to locate the events from documents. The meteorological simulation knowledge is extracted from the vast literature using a deep learning technique called Bilateral Long Short-Term Memory Conditional Random Field (BiLSTM-CRF). The knowledge graph simulation is then built using the Neo4j graph database. It transforms huge literature into shareable and reusable information using the meteorological simulation knowledge graph (MSKG).

Another approach [15] extracts scholarly metadata to train a classifier on labelled data pairs to determine if a given pair belongs to the same author. This study provides an entity retrieval system for the scholarly domain by combining data from textual and structural embeddings for training on the KG IOS Press LD Connect. In this study, the authors employ two standard datasets to analyze the performance of low-dimensional representations of papers and entities (such as authors, organizations, etc.): 1) a standard dataset obtained from Semantic Scholar to assess the semantic relatedness of papers, and 2) a standard dataset derived from DBLP to assess the quality of co-authorship recommendations based on KGEs. Using Doc2Vec [16], the system learns to extract embedding of things from the SKG of IOS LD Connect and use those vectors to represent the content of the papers. The entity retrieval model uses Logistic Regression that features sentences vectors and hierarchical embedding as input. This model works on a corpus of Semantic Scholar publications with comparable characteristics.

We have examined and studied several studies as part of our investigation into the study of citation context, citation causes, and the extraction of cognitive links between articles. Not only that, but we have also investigated alternative methods of categorizing citation contexts [17-25]. Our research has shown that there are various ways of classifying sources used in academic papers. Numerous methods exist, but the two most fundamental involve counting how often something is referenced. In the past, researchers have focused on establishing connections between different types of scientific research papers by looking for things like shared authors or settings. As a result, it is urgently necessary to gather and organize the available citation context. We give citations with the most emotional weight priority because this is an emotional project. Finally, using the in-text citation method, we have created a framework to create a knowledge graph that shows the relationship between the study. Each researcher begins with a certain goal in mind, and they all develop unique strategies for achieving it. Table 1 will serve as a convenient summary of our research findings and survey replies.

Table 1. Literature review summary of recent research for KG.

Author	Dataset	Size	Key Points
Ahmed et. al. (2022) [26]	Arabic and English Text	English: 51,598,084 Arabic: 6,894,368	Comparison between AKG and EKG
Jiang et. al. (2022) [27]	Scholarly Abstracts	500 Abstracts	Bert Based
Oelen et. al. (2020) [28]	News Articles	2,626 Articles	NLP Based
Färber et. al. (2019) [29]	RDF Data Set	8,272,187,24 Triples	Ontology Based
Wang et. al. (2018) [30]	WN18, FV15k, AK18K	3.13 Billion Triples	Ontology Based

3. Methodology

The methodology to create a knowledge graph from In-Text citations is shown in Figure 1. The methodology outlines the steps involved in generating a knowledge graph from a given dataset using natural language processing techniques [31-37]. The dataset employed is a publicly available ACL Anthology dataset. To increase efficiency, the dataset is pre-processed to eliminate stop-words and is tokenized. The tokens are then subjected to entity extraction to identify the subjects and objects in the dataset. To improve the accuracy of the entity extraction process, the extracted tokens are then examined for compound terms and modifiers. An ontology with a hierarchical framework of concepts and their relationships is used to examine the extracted entities for alignment. The ontology matcher compares the retrieved entities to the ontology and maps them to the ontology's suitable concepts. The pattern matcher then uses specified patterns to extract the relationships between the elements. Now the extracted associations can be represented in a knowledge graph, where nodes represent entities and edges reflect their relationships. To ensure completeness, the knowledge graph is subjected to attribute rectification and entity alignment. Finally, the ontology is built to reflect the knowledge graph's links and concepts, allowing for additional analysis and querying. Overall, this methodology offers a thorough and disciplined approach to creating a knowledge network from any given dataset.

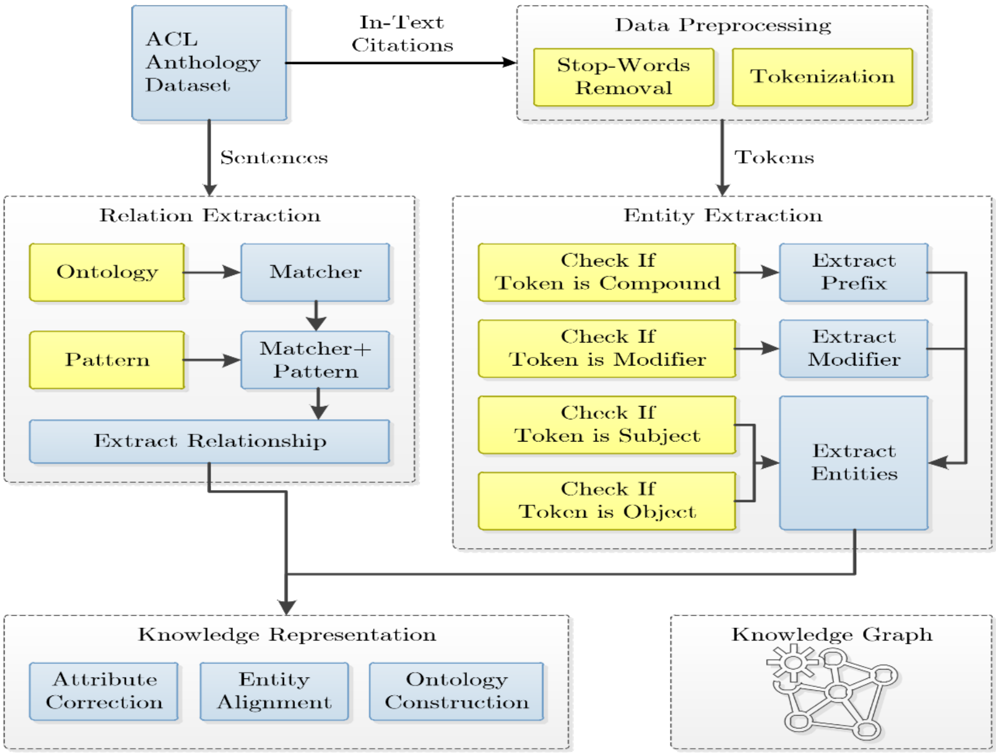


Figure 1. Proposed Methodology of In-Text Citation Knowledge Graph Generation.

4. Dataset

The ACL dataset is a collection of research papers related to natural language processing (NLP) and computational linguistics. It was created by the Association for Computational Linguistics (ACL) and is often used as a benchmark dataset for NLP research.

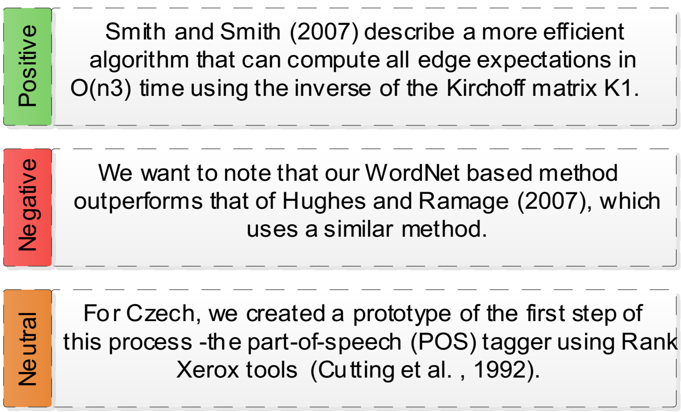


Figure 2. ACL Anthology Dataset Sentiment Polarity Sample.

The dataset contains over 22,000 research papers published between 1965 and 2021, including conference proceedings, journals, and workshop papers. The ACL dataset is organized into two main categories: Anthology Core and Anthology Non-Core. The Anthology Core contains papers from the ACL's flagship conferences and journals, while the Anthology Non-Core includes papers from other conferences and workshops that are relevant to NLP. To perform experiments, a variant of the ACL Anthology dataset is employed. The dataset is composed of more than 8,700 in-text citation sentences. The dataset contains four attributes: citing paper id, cited paper id, citation text, and sentiment polarity. The dataset is composed of three sentiment polarity classes (positive, negative, and neutral). Each in-text citation belongs to one of the polarity classes. A snapshot of the dataset in each sentiment polarity class is shown in Figure 2.

5. Entity Extraction

Entity extraction is a technique that can be used to expose the main subjects of interest rapidly and easily to obtain some comprehension of previously unknown data sets. The capacity to view all entity types (such as names, companies, and nations) in one location is one of the benefits that analysts receive from having access to such a structured corpus as a starting point for future research and analysis. A human can distinguish between different types of names (such as person, location, organization, product, etc.) with reasonable ease, however machines find it incredibly difficult to do so due to the ambiguities of language. Using word meanings and context, parts of speech analysis classify words in a sentence as nouns, verbs, adjectives, adverbs, and other types of words. This information can be used by machines to identify noun phrases, which gives hints about the identities of the important subjects being discussed. But for success, context is crucial. The different ways a word can be used, or its context cannot be accurately distinguished by a keyword-based NER algorithm. Extraction rules, which may be based on pattern matching, syntax, linguistics, semantics, or any combination of these, are powerful in entity extraction. Semantic entity extraction uses logic to clarify meaning or understand context, opening the door for a variety of subsequent actions that are advantageous to multiple business processes in numerous industries.

```
get_entities("Its applications range from sentence boundary disambiguation
"(Reynar and Ratnaparkhi, 1997) to part-of-speech tagging (Ratnaparkhi, 1996),
"parsing (Ratnaparkhi, 1997) and machine translation (Berger et al. , 1996).")

['applications', 'part  tagging']
```

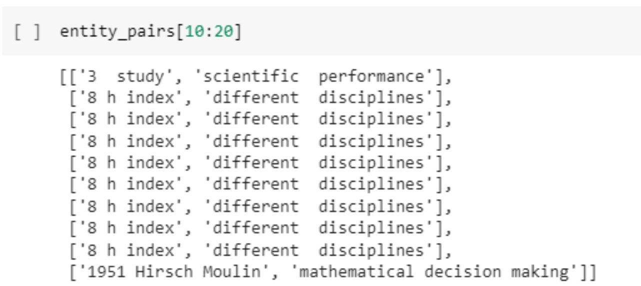
Now we can use this function to extract these entity pairs for all the sentences in our data:

```
[ ] entity_pairs = []

    for i in tqdm(candidate_sentences["sentence"]):
      entity_pairs.append(get_entities(i))

100%|██████████| 8736/8736 [02:36<00:00, 55.86it/s]
```

The list entity_pairs contains all the subject-object pairs from the sentences. Let's have a look

```
[ ] entity_pairs[10:20]

[['Jing', 'concise sentence reduction sentences'],
 ['work', 'many sentence ways'],
 ['based sentence results', 'based sentence combination'],
 ['such  components', 'such clause adverbs'],
 ['1996 compression Grefenstette', 'such Reizler et methods'],
 ['written  constituents', 'aligned  documents'],
 ['that', 'such  links'],
 ['reduced  clauses', 'reduced  example'],
 ['them', 'informative  errors'],
 ['how  professionals', 'summaries']]
```

Figure 3. Entity Extraction from ACL Dataset.



```
[ ] entity_pairs[10:20]

[['3  study', 'scientific  performance'],
 ['8 h index', 'different  disciplines'],
 ['8 h index', 'different  disciplines'],
 ['8 h index', 'different  disciplines'],
 ['8 h index', 'different  disciplines'],
 ['8 h index', 'different  disciplines'],
 ['8 h index', 'different  disciplines'],
 ['8 h index', 'different  disciplines'],
 ['8 h index', 'different  disciplines'],
 ['1951 Hirsch Moulin', 'mathematical decision making']]
```

Figure 4. Extracted Entity Sample.

In this study, we developed a technique to help extract the entity pairings from the imported data. The en-coreweb-sm model of the SPACY library was used to load the English language tokenizer, tagger, parser, NER, and word vectors into memory. The entity extraction function is intended to go in the sentence and contain the subject and object entities, respectively, if the tokens in the file are not punctuation, compound words, modifiers, or prefixes. The system's component that extracts significant entities is operating as planned. The subjects and entities of the data collection can now be added to the catalogues. The procedure adapted for entity extraction is shown in Figure 3, while some of the entities taken from the dataset are displayed in Figure 4. We start with a text data input and output the entities (application and part tagging) that we have got. After that, all the sentences in the dataset are extracted using this method as entity pairs. In the end, we were able to compile a list of entity pairs that contained every subject-object pair found in the phrases.

6. Relation Extraction

This feature explains not only straightforward relationships through direct connections, but also more intricate connections between diverse aspects. This makes it easier to summarize data quickly and effectively. Relationship extraction is the process of figuring out how words in a text are related informally. There are numerous semantic categories that can be used to categorize the extracted relationships, but they involve two or more entities of the same type (such as individuals, businesses, or actual places) (e.g., married to, employed by, resides in). There are five different approaches to perform Relation Extraction.

- Rule-based or Pattern-based
- Supervised
- Weakly Supervised
- Distantly Supervised
- Unsupervised

```
[5] def get_relation(sent):

      doc = nlp(sent)

      # Matcher class object
      matcher = Matcher(nlp.vocab)

      #define the pattern
      pattern = [{'DEP':'ROOT'},
                 {'DEP':'prep','OP':"?"},
                 {'DEP':'agent','OP':"?"},
                 {'POS':'ADJ','OP':"?"}]

      # matcher.add("matching_1", None, pattern)
      matcher.add("matching_1", [pattern])

      matches = matcher(doc)
      k = len(matches) - 1

      span = doc[matches[k][1]:matches[k][2]]

      return(span.text)

get_relation("Semantic classification programs
'(Brown et al. , 1992; Hatzivassiloglou and McKeown, 1993;
 'Pereira et al. , 1993) use statistical information based
 'on cooccurrence with appropriate marker words to partition
'a set of words into semantic groups or classes.")

'use statistical'
```

Figure 5. Various Relation Extracted from the Dataset.

The rule-based or pattern-based RE approach is used in this study. Hand-crafted patterns can be used in rule or pattern bases RE to locate multiple instances of relations by looking for triples (X, Y), where X are entities and Y are words connecting them. In relation extraction (RE), associations are taken out of the ACL Anthology dataset. This requires identifying and extracting connections between different elements present within the sentences. When two persons are mentioned in the text, the relationship extraction technique, for instance, will attempt to identify the nature of their relationship (friend, coworker, or family member). The most popular tool used to support this process is an ontology. This is a formal depiction of ideas and connections in a certain field, such natural language processing. The retrieved texts are then subjected to a matcher to look for patterns that might point to a relationship between various elements. For instance, recognizing the pattern "X is a student of Y" in a sentence could indicate a connection between X and Y.

To extract certain correlations from the dataset, the matcher and pattern (Matcher + Pattern) are finally combined. The relationships between the pieces of the dataset can then be depicted visually by representing these interactions in a knowledge graph. The complete procedure to extract the relations from the dataset is shown in Figure 5. The established pattern in Figure 6 searches for the ROOT word, which is typically the main verb. The pattern first looks for the ROOT and then determines if the words that follow it are agents or preps (for "preposition"). Finally, Figure 3.8 shows the most prevalent relationships found in the dataset.

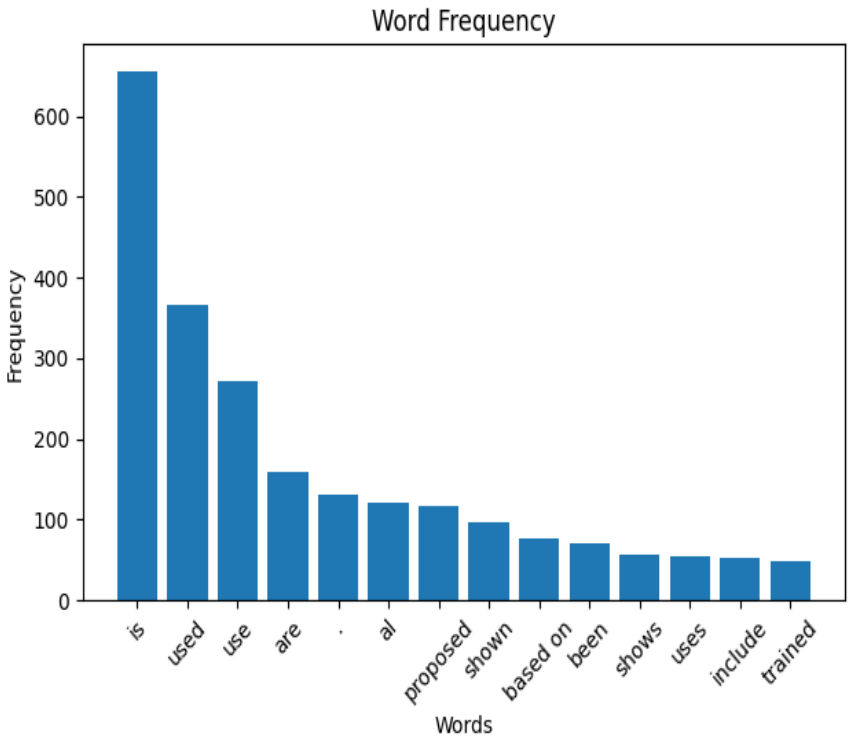


Figure 6. Relation Frequency Count.

7. Build Knowledge Graph

The creation of a knowledge graph utilizing the CSV files is the last step. We construct a knowledge graph using a sizable corpus of ACL dataset. Entities are represented by knowledge graph nodes. Edges are a representation of the connections between nodes. Each of a relation's edges is given a type when the relation is defined. A knowledge graph is a representation of data in a graph. A KG's nodes and edges both provide information about individual items and their connections, respectively. The ACL dataset contains many articles and papers; hence, this research initially chose a limited selection of articles to work with, as shown in Figure 7, to develop a knowledge graph utilizing this dataset.

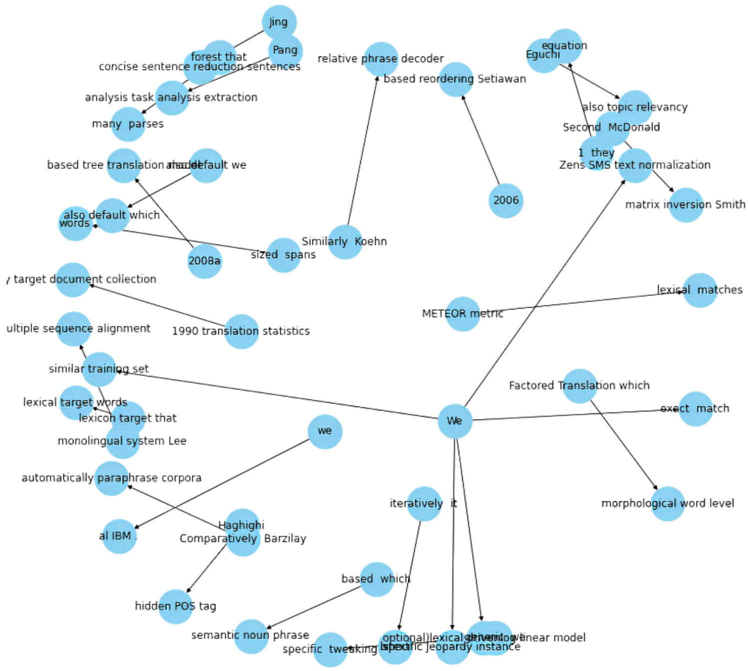


Figure 7. Knowledge Graph on Small Subset.

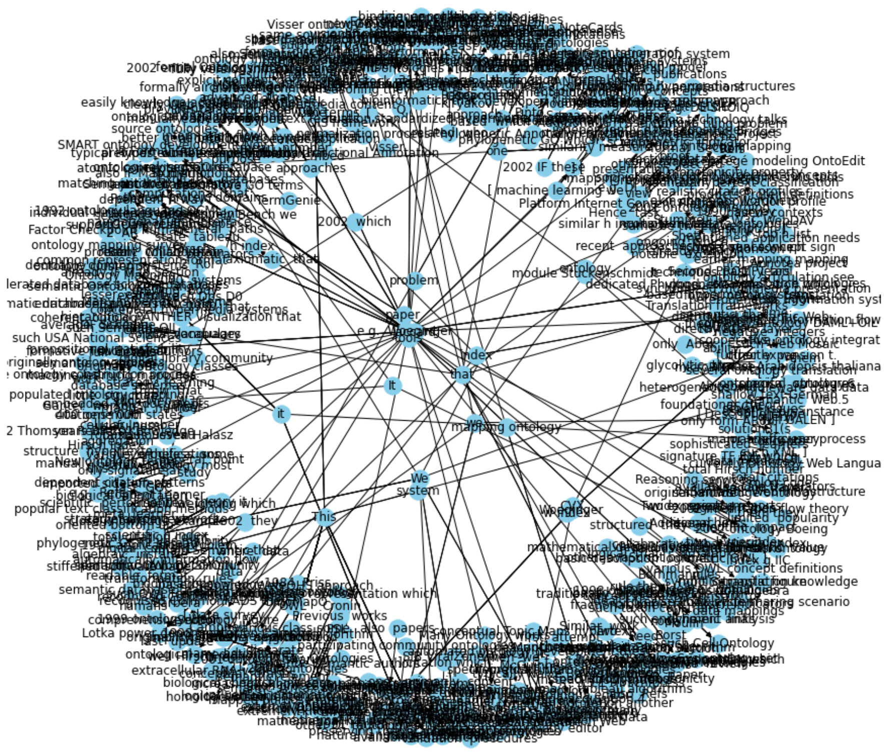


Figure 8. Knowledge Graph on Complete ACL Dataset.

The research then constructs a knowledge graph for wider comparison utilizing the entire ACL dataset. It took a significant number of computational resources and experience in NLP, data modelling, and graph visualization to build a comprehensive knowledge graph from an ACL dataset. The generated knowledge graph is used to investigate citation links between works, spot trends and patterns in the literature, and provide guidance for additional research and analysis. Figure 8 displays the knowledge graph created using the ACL dataset.

8. Conclusion and Future Work

The ACL dataset has the potential to yield insightful knowledge about the connections between publications, authors, and organizations in the field of computational linguistics. This knowledge can be derived from in-text citation data. Using information from in-text citations, this thesis investigated the process for creating a knowledge graph and gave a preliminary analysis of the resulting graph. The outcomes show that this strategy is workable and beneficial, and they also imply that it may be able to provide insight for future field research and analysis. This strategy does, however, come with several drawbacks and difficulties. In-text citation data correctness and completeness, the choice of suitable NLP techniques and graph analysis tools, and the approach's scalability to bigger datasets are a few of these. To overcome these issues and create more reliable and scalable approaches for building knowledge graphs from in-text citation data, more study is required.

The methodology for preprocessing and examining in-text citation data may be improved in future work. Additionally, various methods for representing and visualizing the resulting knowledge graph may be investigated. Future research can also concentrate on using this method on different datasets and subject areas to verify its generalizability and scalability.

Conflicts of Interest: "The authors declare no conflict of interest."

References

1. A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. d. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, et al., "Knowledge graphs," Synthesis Lectures on Data, Semantics, and Knowledge, vol. 12, no. 2, pp. 1–257, 2021.
2. J. Hoffart, F. M. Suchanek, K. Berberich, and G. Weikum, "Yago2: A spatially and temporally enhanced knowledge base from wikipedia," Artificial intelligence, vol. 194, pp. 28–61, 2013.
3. T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, B. Yang, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel, et al., "Never-ending learning," Communications of the ACM, vol. 61, no. 5, pp. 103–115, 2018.
4. J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. Van Kleef, S. Auer, et al., "Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia," Semantic web, vol. 6, no. 2, pp. 167–195, 2015.
5. C. Zhang, C. Re, M. Cafarella, C. De Sa, A. Ratner, J. Shin, F. Wang,´ and S. Wu, "Deepdive: Declarative knowledge base construction," Communications of the ACM, vol. 60, no. 5, pp. 93–102, 2017.
6. J. Volz, C. Bizer, M. Gaedke, and G. Kobilarov, "Silk-a link discovery framework for the web of data," in Ldow, 2009.
7. F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: A large ontology from wikipedia and wordnet," Journal of Web Semantics, vol. 6, no. 3, pp. 203–217, 2008.
8. K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: a collaboratively created graph database for structuring human knowledge," in Proceedings of the 2008 ACM SIGMOD international conference on Management of data, pp. 1247–1250, 2008.
9. R. Bhagat and E. Hovy, "What is a paraphrase?" Computational Linguistics, vol. 39, no. 3, pp. 463–472, 2013.
10. M. Lakshika and H. Caldera, "Knowledge graphs representation for event-related e-news articles," Machine Learning and Knowledge Extraction, vol. 3, no. 4, pp. 802–818, 2021.
11. R. Wang, Y. Yan, J. Wang, Y. Jia, Y. Zhang, W. Zhang, and X. Wang, "Acekg: A large-scale knowledge graph for academic data mining," in Proceedings of the 27th ACM international conference on information and knowledge management, pp. 1487–1490, 2018.
12. B. Yaman, M. Pasin, and M. Freudenberg, "Interlinking scigraph and dbpedia datasets using link discovery and named entity recognition techniques," in 2nd Conference on Language, Data and Knowledge (LDK 2019), Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.
13. J. L. Martinez-Rodriguez, I. Lopez-Ar´ evalo, and A. B. Rios-Alvarado,´ "Openie-based approach for knowledge graph construction from text," Expert Systems with Applications, vol. 113, pp. 339–355, 2018.
14. C. Pechsiri and R. Piriyakul, "Explanation knowledge graph construction through causality extraction from texts," Journal of computer science and technology, vol. 25, no. 5, pp. 1055–1070, 2010.
15. G. Mai, K. Janowicz, and B. Yan, "Combining text embedding and knowledge graph embedding techniques for academic search engines.," in Semdeep/NLIWoD@ ISWC, pp. 77–88, 2018.
16. A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," Advances in neural information processing systems, vol. 26, 2013.
17. I. Ihsan, Imran, S., Ahmed, O., Qadir, M.A., 2019. A Corpus-based Study of Reporting Verbs in Citation Texts Using Natural Language Processing. Corp.: J. Corpus Linguist. 2 (1), 25–35.
18. I. Ihsan, Mohib, U., Ullah, K.R., Irfan, U.M., Abdullah, A., Wael, A., 2022. SEAL: Semantically enriched authoring in LaTeX: a model for scientific discourse. IEEE Access 10, 13525–13535.
19. I. Ihsan and M. A. Qadir, "CCRO: Citation's Context & Reasons Ontology," in IEEE Access, vol. 7, pp. 30423-30436, 2019, doi: 10.1109/ACCESS.2019.2903450.
20. I. Ihsan, H. Rahman, A. Sheikh, A. Sulaiman, K. Rajab and A. Rajab, "Improving in-text citation reason extraction and classification using supervised machine learning techniques," in Computer Speech & Language, 82, 101526 (2023). https://doi.org/10.1016/j.csl.2023.10152.
21. I. Ihsan and M. A. Qadir, An NLP-based citation reason analysis using CCRO. Scientometrics 126, 4769–4791 (2021). https://doi.org/10.1007/s11192-021-03955-6.
22. Ijaz, S., Lali, M. I., Shahzad, B., Imran, A., & Tiwana, S. (2017, December). Biasness identification of talk show's host by using twitter data. In 2017 13th International Conference on Emerging Technologies (ICET) (pp. 1-6). IEEE.
23. Jubeen, M., Rahman, H., Rahman, A. U., Wahid, S. A., Imran, A., Yasin, A., & Ihsan, I. (2022). An Automatic Breast Cancer Diagnostic System Based on Mammographic Images Using Convolutional Neural Network Classifier. Journal of Computing & Biomedical Informatics, 4(01), 77-86.
24. Imran, A., Nasir, A., Bilal, M., Sun, G., Alzahrani, A., & Almuhaimeed, A. (2022). Skin Cancer Detection Using Combined Decision of Deep Learners. IEEE Access, 10, 118198-118212
25. Imran, A., Fahim, M., Alzahrani, A., Fahim, S., Alheeti, K. M. A., & Rehman, S. U. (2023, January). Twitter Sentimental Analysis using Machine Learning Approaches for SemeVal Dataset. In 2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC) (pp. 1-6). IEEE.
26. Ahmed, Ibrahim A, Fatima N AL-Aswadi, Khaled MG Noaman, et al. (2022). "Arabic Knowledge Graph Construction: a Close Look in the Present and into the Future". In: Journal of King Saud University-Computer and Information Sciences.
27. Jiang, Ming et al. (2022). "Evaluating BERT-based scientific relation classifiers for scholarly knowledge graph construction on digital library collections". In: International Journal on Digital Libraries 23.2, pp. 197–215.
28. Oelen, Allard, Markus Stocker, and Sören Auer (2020). "Creating a scholarly knowledge graph from survey article tables". In: International Conference on Asian Digital Libraries. Springer, pp. 373–389.

29. Färber, Michael (2019). "The Microsoft Academic Knowledge Graph: A linked data source with 8 billion triples of scholarly data". In: International semantic web conference. Springer, pp. 113–129.
30. Wang, Ruijie et al. (2018). "Acekg: A large-scale knowledge graph for academic data mining". In: Proceedings of the 27th ACM international conference on information and knowledge management, pp. 1487–1490.
31. Li, J., Hu, Q., Imran, A., Zhang, L., Yang, J. J., & Wang, Q. (2018, July). Vessel recognition of retinal fundus images based on fully convolutional network. In 2018 IEEE 42nd annual computer software and applications conference (COMPSAC) (Vol. 2, pp. 413-418). IEEE.
32. Akhtar, F., Li, J., Guan, Y., Imran, A., & Azeem, M. (2019). Monitoring bio-chemical indicators using machine learning techniques for an effective large for gestational age prediction model with reduced computational overhead. In Frontier Computing: Theory, Technologies and Applications (FC 2018) 7 (pp. 130-137). Springer Singapore.
33. Wajahat, A., Imran, A., Latif, J., Nazir, A., & Bilal, A. (2019, January). A Novel Approach of Unprivileged Keylogger Detection. In 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1-6). IEEE.
34. Bilal, A., Sun, G., Mazhar, S., Imran, A., & Latif, J. (2022). A Transfer Learning and U-Net-based automatic detection of diabetic retinopathy from fundus images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 10(6), 663-674.
35. Imran, A., Li, J., Pei, Y., Mokbal, F. M., Yang, J. J., & Wang, Q. (2020). Enhanced intelligence using collective data augmentation for CNN based cataract detection. In Frontier Computing: Theory, Technologies and Applications (FC 2019) 8 (pp. 148-160). Springer Singapore.
36. Ali, S., Rehman, S. U., Imran, A., Adeem, G., Iqbal, Z., & Kim, K. I. (2022). Comparative Evaluation of AI-Based Techniques for Zero-Day Attacks Detection. Electronics, 11(23), 3934.
37. Ashfaq, A., Imran, A., Ullah, I., Alzahrani, A., Alheeti, K. M. A., & Yasin, A. (2022, October). Multi-model Ensemble Based Approach for Heart Disease Diagnosis. In 2022 International Conference on Recent Advances in Electrical Engineering & Computer Sciences (RAEE & CS) (pp. 1-8). IEEE.