

# Leveraging Knowledge Graphs for Orphan Entity Allocation in Resume Processing

Aagam Bakliwal<sup>1\*</sup>, Shubham Manish Gandhi<sup>1†\*</sup>, Yashodhara Haribhakta<sup>1</sup>

<sup>1</sup> Department of Computer Science and Engineering  
COEP Technological University  
Pune, Maharashtra, India

**Abstract**—Significant challenges are posed in talent acquisition and recruitment by processing and analyzing unstructured data, particularly resumes. This research presents a novel approach for orphan entity allocation in resume processing using knowledge graphs. Techniques of association mining, concept extraction, external knowledge linking, named entity recognition, and knowledge graph construction are integrated into our pipeline. By leveraging these techniques, the aim is to automate and enhance the efficiency of the job screening process by successfully bucketing orphan entities within resumes. This allows for more effective matching between candidates and job positions, streamlining the resume screening process, and enhancing the accuracy of candidate-job matching. The approach's exceptional effectiveness and resilience are highlighted through extensive experimentation and evaluation, ensuring that alternative measures can be relied upon for seamless processing and orphan entity allocation in case of any component failure. The capabilities of knowledge graphs in generating valuable insights through intelligent information extraction and representation, specifically in the domain of categorizing orphan entities, are highlighted by the results of our research.

**Index Terms**—Orphan Entity Allocation, Knowledge Graphs, Association Mining, Concept Extraction, External Knowledge Linking, Named Entity Recognition

## I. INTRODUCTION

### A. Problem Statement

A piece of information or data within a dataset or context that lacks sufficient surrounding context or clear relationships to other structured data is known as an orphan entity, making it difficult to categorize or determine its specific meaning. In the context of resume processing, the term 'Python' mentioned in the 'Skills' section of a job applicant's resume could serve as an example of an orphan entity. Whether 'Python' refers to a programming language or the reptile species remains unclear without additional context or explicit information.

The allocation of orphan entities, such as skills and qualifications, from a large number of resumes, is a time-consuming and error-prone task. The unstructured nature of resumes and the presence of ambiguous references create challenges for recruiters in allocating relevant entities. When manually identifying and linking orphan entities, valuable time can be consumed, and the possibility of human errors can be introduced. Therefore, a need arises for an automated and reliable solution to streamline the entity allocation process in talent acquisition and recruitment.

### B. Motivation

The immense potential of leveraging knowledge graphs as a promising solution is recognized, motivated by the persistent challenges encountered in orphan entity allocation. A structured representation of information necessitated by the complex nature of talent acquisition is precisely offered by knowledge graphs. The motivation stems from the desire to enhance the efficiency and accuracy of the talent acquisition process, bridging the gap between valuable but underutilized candidate information and the requirements of employers. Harnessing the capabilities of knowledge graphs, the handling of orphan entities is aimed to be revamped, ultimately streamlining talent acquisition and fostering improved matches between candidates and job opportunities.

### C. Objectives

A comprehensive pipeline, utilizing knowledge graphs to streamline orphan entity allocation in resume processing is aimed to be developed. The precision and reliability of assigning orphan entities in resume processing are aimed to be improved through concept extraction, association mining, named entity recognition (NER), external knowledge linking, and knowledge graph construction. The objective is to establish a streamlined and dependable pipeline that enables well-informed decision-making and enhances the alignment between candidate profiles and job requirements.

## II. RELATED WORK

Relevant information is extracted from resumes, which consist of structured and unstructured data. The identification and allocation of orphan entities are deemed essential for efficient talent acquisition and recruitment processes.

Three categories encompass traditional approaches for understanding orphan entities: sequence labeling-based, hypergraph-based, and span-based methods.

The sequence labeling approach [1] is used to predict labels for each token, but it struggles with nested NER. Some works such as [2] [3] have adapted the sequence labeling model for nested entity structures by designing a special tagging scheme. Span-based methods first extract spans through enumeration [4] or boundary identification [5], and then classify the spans. For example, [6] proposes a two-stage identifier that first locates entities and then labels them, treating NER as a joint task of boundary regression and span classification. However,

<sup>†</sup> Corresponding author

\* Indicates equal contribution

these methods mainly focus on boundary identification, which might not be necessary for short orphan entities.

A standard method for inferring word meanings by matching predefined keywords or dictionaries is keyword matching. While providing some level of disambiguation, it may have limited accuracy and fail to capture complete semantic context [7].

In the context of resume information extraction, a two-step algorithm for extracting information from resumes was proposed by [8]. The Writing Style feature is used to first identify resume blocks, and then factual information attributes are identified. Though promising results were obtained on real-world datasets, it falls short in extracting word meanings beyond section names. The effectiveness of cascaded processing and multiple passes through resumes was demonstrated by [9], outperforming the multi-pass flat model approach. A framework using text classifiers to build a structured resume repository, facilitating knowledge extraction and efficient data organization, was presented by another study [10].

To enhance orphan entity allocation, the use of ontologies [10] was explored. These systems leveraged ontology-defined hierarchies, concepts, and relationships to improve the extraction of contextual information from resumes. Furthermore, orphan entities are extracted through the combination of word embeddings and entity co-occurrence analysis by [11], capturing semantic similarity and contextual information and relationships within the resume.

Knowledge Graph-Augmented Abstractive Summarization with Semantic-Driven Cloze Reward [12] utilizes semantic knowledge from knowledge graphs to enhance entity allocation accuracy, however, it is resource-intensive and has a complex implementation.

A permutation-based bidirectional training approach is employed by [13] for the identification of named entities. This approach is incorporated into our NER module, alongside the foundational idea from [8], where a customized cascading model is introduced, specifically designed for resumes. The existing knowledge generated by [14] is leveraged, combined with Association Mining techniques, to tailor our algorithm for optimal processing of resumes.

### III. PROPOSED PIPELINE

A comprehensive pipeline for identifying orphan entities in resumes is presented, addressing the limitations of pre-trained language models like the Bidirectional Encoder Representations from Transformers (BERT), which capture general language representations but lack domain-specific knowledge [15]. The pipeline involves a combination of modules, namely, concept mining, association mining, NER, external knowledge linking, and knowledge graph creation.

For each module in the pipeline, an orphan entity and the corresponding resume are taken as input, and an output is generated. Through thorough investigation and experimentation using various thresholds for measuring the similarity between the output and the orphan, a predefined minimum threshold has been established for each module. If the distance

between the orphan entity and the identified output exceeds this threshold, the output is incorporated into the knowledge graph. Otherwise, the next module is executed.

The pipeline begins with a concept extraction module, where the orphan and the corresponding resume are analyzed to extract relevant concepts. If the distance between the concept and the orphan is below the predefined threshold, the identified entity is added to the knowledge graph.

If satisfactory results are not achieved by the concept extraction module, association mining is conducted using Apriori analysis and word similarity. This process discovers associations between the orphan entity and words present in resumes and identifies the most similar word to input to the knowledge graph. If the similarity of the identified word exceeds the predefined threshold for that module, that word is added to the knowledge graph.

If both the concept extraction and association mining modules fail to yield satisfactory results, NER is employed. NER identifies potential sources in the resume and finds the most similar word from the knowledge graph. If the similarity surpasses the predefined threshold, the closest word (source entity) is added to the knowledge graph.

Finally, even if an output is produced by the NER module whose distance is less than the orphan for the predefined threshold, the output of the external knowledge linking module is searched to check whether the orphan exists in it.

Throughout execution, the knowledge graph is updated with new entities and edges whenever a suitable source entity is identified.

#### A. Concept Mining

The concept extraction module extracts meaningful concepts from artifacts and utilizes embedding vector representations to analyze occurrence contexts, evaluate concept quality, and assess their suitability within local contexts, inspired by an innovative approach to concept mining [16].

Glove vectors [17], which are pre-trained word embeddings, are initialized. With the help of these vectors, similar or associated words can be found, thereby enhancing accuracy. A framework created by [14] is implemented to extract relevant related entities associated with orphan entities.

The Resume along with the orphan entities are taken as input, and the ConceptNet platform [14] is utilized to generate a set of potentially related words. The distance between context words extracted from artifacts and the related words is measured using Glove vectors, and the concept with the minimum distance below the pre-defined threshold is selected by the module. A tuple is returned by the module, pairing the orphan entity with its corresponding concept.

If the concept extraction module does not produce satisfactory outcomes, the subsequent module, namely Association Mining, is executed.

The workflow of this module is illustrated in Figure 1.

#### B. Association Mining

Inspiration is drawn from the challenges addressed by the Apriori algorithm in the study focused on e-commerce appli-

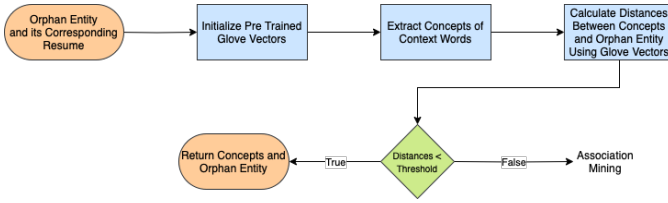


Fig. 1. Concept Mining Module Workflow

cations and the extraction of association rules from customer purchase transactions [18]. In our specific context of interest, we adapt the Apriori algorithm to uncover word associations.

A module has been devised that utilizes the Apriori algorithm. Initially, contextual words pertaining to the orphan are extracted from the resume, and transaction lists for uni-grams and bi-grams are generated. These transaction lists are created from the context words and encompass the entire corpus of resumes within the dataset. Subsequently, the Apriori algorithm is employed to process these transaction lists, and the results are filtered based on the support, confidence, and lift, as defined in [19].

If the distances between the results produced by the module and the orphan entity are above a specified threshold, the subsequent module, namely NER, is executed. Otherwise, the results are added to the knowledge graph.

The workflow of this module is illustrated in Figure 2.

### C. Named Entity Recognition

A pivotal role is assumed by the NER module, leveraging the XLNet model [13] to facilitate the identification and classification of named entities within textual data. By providing XLNet with a target word for identification and its encompassing context words, linguistic nuances are grasped in a dynamic, bidirectional manner. XLNet distinguishes itself from other models by implementing the "permutation-based" training approach [13]. This unique approach allows the limitations of previous models, such as BERT and roBERTa, to be overcome, with better capture of long-range dependencies and bidirectional context [13]. Given that the contextual words in a resume hold extreme importance in our task, the use of XLNet is proposed by the method. This model holds the promise of cultivating a deeper comprehension of the input word within its expansive linguistic context, ultimately leading to heightened accuracy and robustness in NER results.

The workflow of this module is illustrated in Figure 3.

### D. External Knowledge Linking

External knowledge linking is incorporated to enhance the understanding and context of the orphan entities. This process establishes connections between entities, concepts, or terms in the dataset and relevant external knowledge bases, specifically from Coursera and LinkedIn. This approach builds upon previous techniques that leverage external knowledge, such as Wikipedia, to enhance entity comprehension and connectivity in datasets [20].

By integrating external knowledge linking, we enrich our knowledge graph with valuable information, particularly regarding the relationships between entities. This step is not a one-time occurrence but rather a module that runs regularly to ensure the continuous update of our Knowledge Graph with emerging and up-to-date skills.

Moreover, we gain access to a wealth of additional insights and information through various external knowledge sources and the iterative nature enables us to keep our Knowledge Graph current and relevant, ensuring that it remains a comprehensive and dynamic resource for understanding the relationships and connections within our dataset.

The workflow of this module is illustrated in Figure 4.

### E. Knowledge Graph Building

Knowledge graphs serve as the backbone for our proposed methodology. A source-destination tuple seamlessly integrates newly discovered relationships into the knowledge graph through modules like Concept Mining, Association Mining, NER, or External Knowledge Linking. This approach allows us to visualize the connections between entities, facilitating rapid searches and accurate identification of relationships. The knowledge graph acts as a structured comprehensive representation of the underlying data, empowering us to efficiently navigate and explore the intricate network of relationships within the system.

The power of knowledge graphs becomes evident as they enable the efficient and rapid allocation of orphans that are already present in the graph. By leveraging the existing knowledge graph, we determine the context words of the orphan from the resume and return the word with the minimum cosine distance between their glove vectors and the neighboring words in the knowledge graph.

A subset of the knowledge graph is illustrated in Figure 6.

## IV. EXPERIMENTAL SETUP

### A. Dataset Description

The dataset utilized in this paper [21] encompasses a comprehensive assortment of resumes, with a substantial sample size exceeding 900 encompassing a wide array of professional backgrounds, including but not limited to testing, web design, HR (human resources). The diverse nature of this dataset, originating from different domains, facilitates a comprehensive assessment of the proposed pipeline. With its extensive collection of resumes, a robust and representative compilation of real-world instances is presented, capturing the intricacies and variabilities associated with distinct job roles and industries.

### B. Experimental Methodology

a) *Dataset Preparation:* The dataset preparation process involves several steps to preprocess the resumes for analysis. Firstly, the lowercase transformation was applied to all resumes for consistency and to avoid discrepancies in capitalization.

Next, the text of the resumes was tokenized, which involved splitting it into individual words or tokens. Tokenization

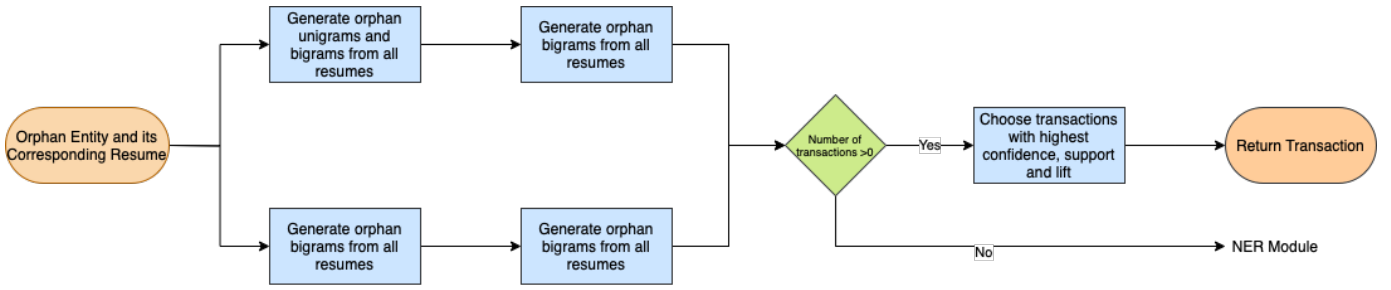


Fig. 2. Association Mining Module Workflow

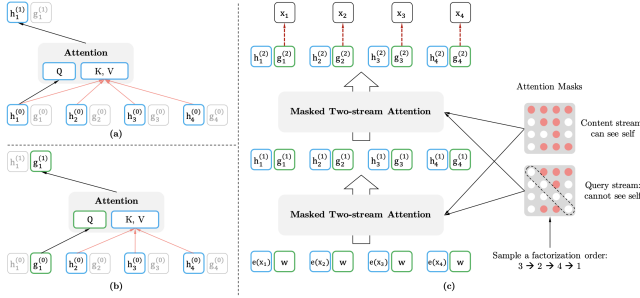


Fig. 3. XLNet model architecture as proposed in [13]

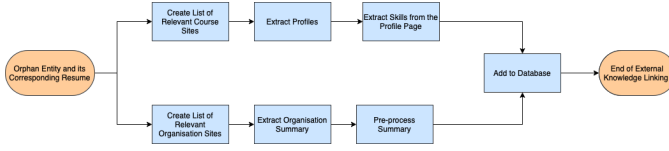


Fig. 4. External Knowledge Linking Module Workflow

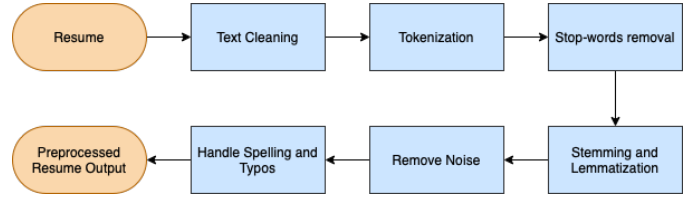


Fig. 5. Text-preprocessing Workflow

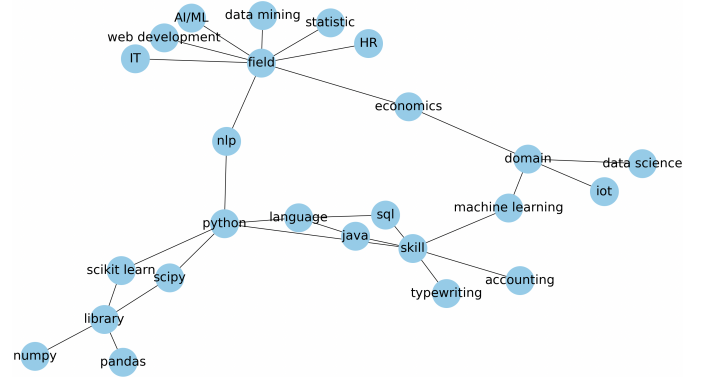


Fig. 6. Subset Graph within the Knowledge Graph

allowed for a more detailed analysis of the textual data by breaking it down into its constituent parts.

To enhance the quality and relevance of the textual data, stopwords, common words like "the" and "and" that carry little meaning, were removed. By eliminating stopwords, the focus shifted to more informative words and phrases in the resumes.

To further refine the textual data, stemming and lemmatization techniques were employed. Word suffixes were removed to obtain their root form through stemming, while lemmatization transformed words into their dictionary or lemma form. These techniques helped consolidate word variations, reducing redundancy, and improving the accuracy of the analysis by treating related words as the same.

The Text-Processing workflow is illustrated in Figure 5.

*b) Knowledge Graph Building:* The first module, concept mining, is where the orphan entity is initially presented along with the resume. If a relevant entity is not returned by the module (based on the predefined threshold), the system proceeds to the next module in the pipeline. The orphan entity is then attached to this node in the knowledge graph upon encountering a module that successfully returns a relevant node.

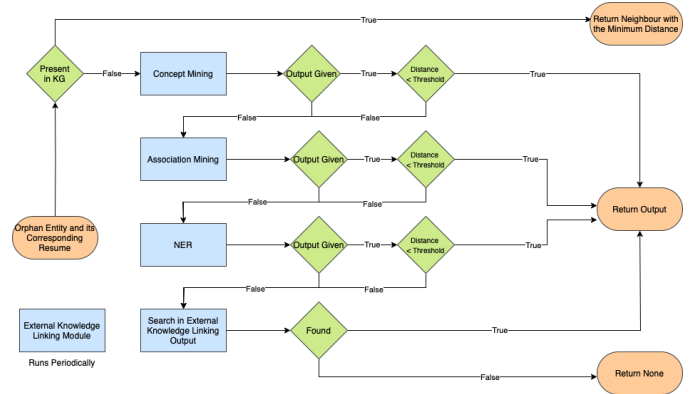


Fig. 7. Overall Program Workflow

*c) Integration:* Incorporating four distinct modules: Concept Mining, Association Mining, NER, and External Knowledge Linking, these modules have been synergistically merged within a hierarchical pipeline to harness their individual

capabilities collectively. The hierarchical model execution, as illustrated in Figure 7, allows for the seamless flow of information and knowledge extraction throughout the various stages of the analysis process.

## V. RESULTS AND ANALYSIS

### A. Algorithm Accuracy

After successive training on the dataset provided in [21], an algorithm was developed and remarkable precision was displayed in categorizing orphan entities. The corresponding accuracies are outlined in Table I, computed using the equation (1). Notably, due to the use of knowledge graphs, there was a consistent improvement in the algorithm's accuracy as the training dataset and training epochs expanded. These findings emphasize the potential for further refining and optimizing the algorithm's performance by incorporating supplementary resume data.

A specific formula is utilized to measure the accuracy of our methodology. The formula expressed mathematically is as follows:

$$\text{accuracy} = \frac{\text{Number of orphans correctly allocated}}{\text{Total number of orphans allocated}} \times 100 \quad (1)$$

where the accuracy is computed by dividing the number of correctly allocated orphan entities by the total number of orphan entities.

TABLE I  
ACCURACY ON DIFFERENT DATASETS

Dataset	Accuracy
Resume	86.32%
Non-resume specific text corpus	78.85%

### B. Noteworthy Observations

1) *Complementary Effect of Concept Mining and Association Mining*: The potential for enhanced accuracy and effectiveness in orphan entity allocation is demonstrated. Showcasing the complementary relationship between concept mining and association mining. When each resume is incorporated, it was observed that if significant results were not yielded by the concept mining algorithm, the association mining algorithm became progressively more effective. This suggests that the strengths of one method can mitigate the limitations of the other, highlighting the potential for leveraging this interdependence to achieve improved orphan entity allocation.

2) *Influence of Algorithm Sequence*: A noteworthy observation pertains to the sequencing of concept mining, association mining, and NER. Specifically, when results were not yielded within a specified threshold distance by the concept mining and association mining algorithms, high accuracy in predicting the type of entity was consistently achieved by NER. However, when the NER module was executed prior to the association mining algorithm, the results generated by the NER algorithm were seldom correct. This finding emphasizes the importance of executing the algorithms in a specific order to ensure optimal performance and accurate allocation.

### C. Contributions of the Research

The implications of our findings suggest that by utilizing structured representations and semantic connections, the potential for greatly enhancing the accuracy and efficiency of orphan entity allocation is demonstrated by knowledge graph-based approaches. This, in turn, can lead to the streamlining of the talent acquisition process, facilitating better matches between candidates and job opportunities.

Significant contributions to the advancement of knowledge-driven applications in talent acquisition and recruitment have been made by this research. The key contributions can be summarized as follows:

- 1) An integrated approach is taken by our developed solution to comprehensively address the challenge of orphan entity allocation in resume processing. Four distinct modules are utilized, which synergistically tackle each other's weaknesses, resulting in a robust and effective solution for automating the allocation of orphan entities.
- 2) The effectiveness and efficiency of knowledge graph-based approaches in automating orphan entity allocation are demonstrated by our research. By leveraging the power of knowledge graphs, complex relationships are captured, and accurate results are produced quickly and efficiently.
- 3) An optimization of the recruitment process is contributed by our research through intelligent information extraction and representation. By automating orphan entity allocation and improving the accuracy of entity mapping, better matches between candidates and job requirements are enabled by our pipeline.

Overall, our research provides valuable insights into the application of knowledge graphs in talent acquisition and recruitment. By bridging the gap between candidate information and job requirements, our approach enhances the recruitment process and contributes to the advancement of knowledge-driven applications in this domain.

### D. Limitations and Future Research Directions

A promising approach is presented for orphan entity allocation in resume processing. However, certain limitations warrant further investigation. The effectiveness of our pipeline is greatly influenced by the quality and completeness of the knowledge graph and external knowledge sources. Improving the accuracy of our pipeline using a limited dataset should be prioritized in future research. Furthermore, excellent results in other domains may be yielded by our pipeline, warranting further exploration. Additionally, entity classification, disambiguation, and linking accuracy can be enhanced by incorporating machine learning and deep learning techniques. Future research should focus on integrating advanced Artificial Intelligence (AI) models to further enhance the performance and effectiveness of the pipeline.

## VI. CONCLUSION

In conclusion, orphan entity allocation in resumes was emphasized in our research using a combination of concept

mining, association mining, and named entity recognition algorithms. Through experiments and analysis, several significant findings that contribute to the understanding and improvement of this domain have been made.

Firstly, promising accuracy was demonstrated by our algorithm, achieving approximately 86% when trained on a dataset. This indicates the potential for effectively allocating named entities in resumes using our approach. Moreover, the accuracy of the algorithm consistently improved with the inclusion of more resumes and iterations, highlighting the value of continuous training and refinement.

A complementary connection between concept mining and association mining algorithms was revealed by our research, where the weaknesses of one approach were counterbalanced by the strengths of the other. Moreover, the crucial role of algorithm sequencing was highlighted by our study, underscoring the significance of meticulously deliberating the execution order to enhance the allocation process.

It is important to note that our algorithm is optimized for resume text and does not achieve the same level of accuracy when applied to general text corpora. Therefore, customizing the approach to align with the unique characteristics and linguistic nuances of different text domains becomes essential.

Overall, valuable insights and methodologies for the effective allocation of orphan entities in resumes are provided by our research. The findings can be built upon in future studies to further enhance the accuracy and applicability of such algorithms in the field of resume analysis and information retrieval.

## REFERENCES

- [1] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," *CoRR*, vol. abs/1508.01991, 2015. [Online]. Available: <http://arxiv.org/abs/1508.01991>
- [2] B. Alex, B. Haddow, and C. Grover, "Recognising nested named entities in biomedical text," in *Biological, translational, and clinical language processing*. Prague, Czech Republic: Association for Computational Linguistics, Jun. 2007, pp. 65–72. [Online]. Available: <https://aclanthology.org/W07-1009>
- [3] J. Wang, L. Shou, K. Chen, and G. Chen, "Pyramid: A layered model for nested named entity recognition," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, Jul. 2020, pp. 5918–5928. [Online]. Available: <https://aclanthology.org/2020.acl-main.525>
- [4] Y. Luan, D. Wadden, L. He, A. Shah, M. Ostendorf, and H. Hajishirzi, "A general framework for information extraction using dynamic span graphs," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 3036–3046. [Online]. Available: <https://aclanthology.org/N19-1308>
- [5] C. Tan, W. Qiu, M. Chen, R. Wang, and F. Huang, "Boundary enhanced neural span classification for nested named entity recognition," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, pp. 9016–9023, Apr. 2020. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/6434>
- [6] Y. Shen, X. Ma, Z. Tan, S. Zhang, W. Wang, and W. Lu, "Locate and label: A two-stage identifier for nested named entity recognition," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Online: Association for Computational Linguistics, Aug. 2021, pp. 2782–2794. [Online]. Available: <https://aclanthology.org/2021.acl-long.216>
- [7] S. Sanyal, S. Hazra, N. Ghosh, and S. Adhikary, "Resume parser with natural language processing," Ph.D. dissertation, 03 2017.
- [8] T. Hanne, J. Chen, C. Zhang, and Z. Niu, "A two-step resume information extraction algorithm," *Mathematical Problems in Engineering*, vol. 2018, p. 5761287, 2018. [Online]. Available: <https://doi.org/10.1155/2018/5761287>
- [9] K. Yu, G. Guan, and M. Zhou, "Resume information extraction with cascaded hybrid model," in *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*. Ann Arbor, Michigan: Association for Computational Linguistics, Jun. 2005, pp. 499–506. [Online]. Available: <https://aclanthology.org/P05-1062>
- [10] J. Chen, Z. Niu, and H. Fu, "A novel knowledge extraction framework for resumes based on text classifier," in *Web-Age Information Management*, X. L. Dong, X. Yu, J. Li, and Y. Sun, Eds. Cham: Springer International Publishing, 2015, pp. 540–543.
- [11] S. Zu and X. Wang, "Resume information extraction with a novel text block segmentation algorithm," *International Journal on Natural Language Computing*, 2019.
- [12] L. Huang, L. Wu, and L. Wang, "Knowledge graph-augmented abstractive summarization with semantic-driven cloze reward," *CoRR*, vol. abs/2005.01159, 2020. [Online]. Available: <https://arxiv.org/abs/2005.01159>
- [13] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding," 2020.
- [14] R. Speer, J. Chin, and C. Havasi, "Conceptnet 5.5: An open multilingual graph of general knowledge," 2018.
- [15] W. Liu, P. Zhou, Z. Zhao, Z. Wang, Q. Ju, H. Deng, and P. Wang, "K-bert: Enabling language representation with knowledge graph," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 03, pp. 2901–2908, Apr. 2020. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/5681>
- [16] K. Li, H. Zha, Y. Su, and X. Yan, "Concept mining via embedding," in *2018 IEEE International Conference on Data Mining (ICDM)*, 2018, pp. 267–276.
- [17] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1532–1543. [Online]. Available: <https://aclanthology.org/D14-1162>
- [18] N. Riyadi, M. F. Mulki, and R. Susanto, "Analysis of customers purchase patterns of e-commerce transactions using apriori algorithm and sales forecasting analysis with weighted moving average (wma) method," *Scientific Research Journal*, vol. 7, no. 7, pp. 45–58, 2019.
- [19] H. Xie, "Research and case analysis of apriori algorithm based on mining frequent item-sets," *Open Journal of Social Sciences*, vol. 09, pp. 458–468, 01 2021.
- [20] K. Torisawa *et al.*, "Exploiting wikipedia as external knowledge for named entity recognition," in *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*, 2007, pp. 698–707.
- [21] G. Dutt, "Resume dataset," Year. [Online]. Available: <https://www.kaggle.com/datasets/gauravduttakiit/resume-dataset>