Methods for the Intellectual Properties Retrieval from Patents

Tianqing Xu[1], Zhibang Zhong[1], Lei Wang[1], Jing Ma[2], and Zihui Zhang[1(✉)]

[1] Institute of Automation, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
`{xutq,wanglei2,zhangzh}@sdas.org`
[2] Otago Polytechnic, Auckland International Campus, Auckland, New Zealand
`Julia.ma@op.ac.nz`

Abstract. With the increasing requirements to protect intellectual property, more and more attention is paid to the research of patent strategy. Intellectual property-oriented patent search is different from normal literature retrieval as its with professional field characteristics and diversified information retrieval. Therefore, this paper first summarizes the main research status of traditional information-based patent retrieval methods, semantic-based patent retrieval methods and knowledge graph-based patent retrieval methods, and then introduces representative patent retrieval applications. Finally, it suggested some ideas for further research on intellectual properties search from patents by discussing the application of deep learning in patent retrieval.

Keywords: Patent Search · Query Expansion · Semantic Retrieval · Knowledge Graph · Deep Learning

1 Introduction

The World Intellectual Property Organization (WIPO) defines a patent as "a legal document issued by a government for the owner of an invention, to protect the exclusive right of the inventor to the invention" [1]. Patent retrieval, as a method of retrieving patent literature, has the characteristics of complex data attributes, high technological sensitivity, and diverse text standards, and is often unable to effectively retrieve relevant patents using common information retrieval methods [2] such as mechanical semantic matching, keywords, and classification numbers. To address these issues, various retrieval methods have been developed to increase recall and precision, such as topic-based retrieval, reference-based retrieval, and dictionary-based expanded retrieval.

As artificial intelligence technology has developed, deep learning has been increasingly applied to patent retrieval technology and combined with natural language processing technology, providing new ideas and methods for patent retrieval. By using deep learning models, words can be represented in a semantic space as word vectors. From this, the semantic retrieval and knowledge graph method has been developed. In addition, patent retrieval methods based on deep learning also help to perform infringement analysis on patent holders and have potential for development in areas such as competitor proposals and market positioning suggestions [3].

2   Traditional Patent Search Methods

Traditional patent retrieval methods commonly include retrieval models, reference relationships, and query expansion. Retrieval models play an important role in various patent retrieval methods. The Vector Space Model (VSM) evolved from the Boolean model, which calculates the similarity between the query and the document by defining a keyword weighting function, and normalizes the Term Frequency (TF) and Inverse Document Frequency (IDF) to judge the importance of the keyword in the document [4]. Probabilistic models are based on probability theory to calculate the similarity between documents and queries and rank them by the size of the probability. The representative one is the BM25 model [5] and Language Model (LM) [6].

Query expansion is a technique used to supplement and refine the search results during retrieval by adding query terms, which mainly including extension-based methods and feedback-based methods [7]. Extension-based methods involve introducing similar terms or synonyms from patent literature or external resources or extracting new terms to expand or reformulate the query. Feedback-based methods, on the other hand, modify the query based on the results retrieved, using pseudo-relevance feedback or citation analysis [4]. Magdy and Jones et al. constructed a synonym set for terms using a bi-directional translation model for vocabulary to achieve query expansion [8].

The citation relationship refers to the reference of one object to another and observes its development context through the citation relationship [2]. By analyzing and quantifying the citation relationship, factors such as the degree of attention and value of patent literature by other patents can be obtained [9]. However, the referenced literature may be influenced by some motives or be limited by accessibility, so the number of references of patent literature is not equivalent to its importance [10].

3   Semantics-Based Patent Retrieval Method

In general, semantic retrieval refers to analyzing the user's retrieval requirements from the perspective of text understanding. The semantic retrieval method first requires semantic segmentation processing and extraction of keyword information for each patent document in the patent database and collects various concept information obtained from the analysis into a knowledge base. Then, the intelligent algorithm calculates the words and sentences of related technical solutions in the patent document and uses vector values to represent the calculation results. The cosine angle between each vector value is the degree of relevance of the information, and the smaller the angle, the higher the rank of the relevance. Finally, the retrieval information is sorted according to the relevance with the information in the patent database [11]. The principle of the semantic retrieval method is shown in Fig. 1.

Researchers have proposed various effective patent semantic retrieval algorithms by combining semantic retrieval methods with traditional patent retrieval algorithms [13, 14]. The semantic retrieval of patent literature usually processes the patent text and then matches the similarity through classification algorithms, which mainly includes the collection of patent text information, the preprocessing of patent text information, the classification of patent text, and the calculation of patent text semantic similarity [15].
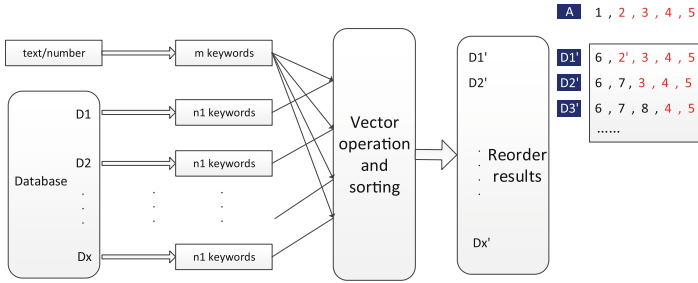
Fig. 1. Semantic Search Principle Diagram [12]
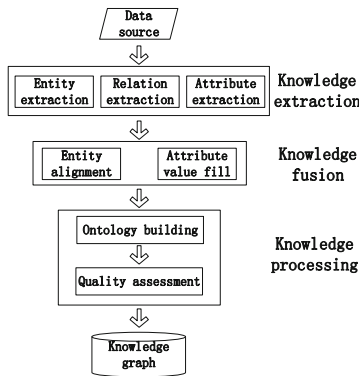

Fig. 2. Knowledge Graph Construction Process Diagram

A knowledge graph is a formal framework for describing semantic knowledge, which is a semantic network consisting of entities representing a certain type and their relationships [16]. In order to refine the patent from a document to a smaller structured knowledge granularity for finer knowledge management and application, Ma Guobing used a Bert-based knowledge extraction algorithm to extract specific entities from the patent text, recognize the relationships between entities [17]. Li et al. built a knowledge graph for the patent text in the new energy vehicle field, and proposed a method for measuring the similarity between composite metric terms [18]. Wei Wei et al. constructed a knowledge graph with the applicant, inventor and other information from the patent application as entity nodes [19]. Takes the Bert model as an example [17], and the knowledge graph construction flow chart is shown in Fig. 2.

4   Patent Retrieval Method Based on Deep Learning

Deep learning, as an important branch of machine learning, has been widely applied in patent retrieval today. Convolutional Neural Networks (CNNs) use word vector techniques to convert keywords into word vectors and then obtain their respective features through convolutional layers and pooling layers to form a complete feature vector. Compared to CNNs, recurrent neural networks (RNNs) mainly solve sequence problems and
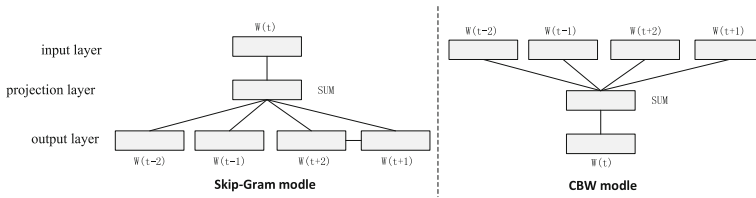
Fig. 3. Word2vec Model Diagram

long-term dependence issues and are more advantageous in preserving the globally optimal features [20].

Word2vec and GloVe models are commonly used word vector training models in deep learning. Word2vec is the most widely used word training model currently, and it can adopt two modes, one is CBOW mode, and the other is Skip-Gram mode. The principle diagram of the word2vec model is shown in Fig. 3. The GloVe word vector model utilizes statistical information to train the non-zero elements of the co-occurrence matrix of words in a very large corpus, thus capturing the linear relationships between words through word vectors.

Deep learning combined with various patent retrieval methods can effectively promote the development of patent retrieval methods. Khuda Bukhsh and others applied machine learning to patent query expansion and proposed a query classification method based on machine learning [21]. Liu Bin and others used deep learning-based patent retrieval methods to calculate the similarity between patents using Convolutional Neural Networks [22]. Hu Jie extracted entities and relationship types in patent texts through deep learning algorithms and proposed a knowledge acquisition and reuse prototype system for innovative design [23].

5   Conclusions

Considering the complex, technical sensitivity, and diversity of patent retrieval, this paper surveys on the major patent retrieval methods. With the development of artificial intelligence technology, various patent retrieval methods effectively handle problems such as multi-source information fusion by constantly optimizing the retrieval model and improving the patent retrieval task, which improves the effectiveness of patent retrieval. Especially the combination of deep learning with semantic resources, data graph construction, and query expansion further drives the development of patent retrieval methods towards intelligence and humanization.

References

1. Michael B, A., Daniel A., Francis, N., et al. (1991) Direct validation of citation counts as indicators of industrially important patents. Research policy, 20(3): 251-259. https://doi.org/10.1016/0048-7333(91)90055-U

2. Liu, M.L. (Wuhan University), 2018. Research on New Patent Retrieval Method under Deep Learning. https://kns.cnki.net/kcms2/article/abstract?v=jBOcPZekd6MCvSRjkX7a3Uv40TDEEPXOCE6wFoIZv4ba43RXsXdCnpc1vbkWFedcH5v_QlcMwCn_-oww49caXZXiKKAJqZO7k9exn4GO3m7Jwa2ki73o_A==&uniplatform=NZKPT&language=CHS

3. Jia, W.Z. (Jilin University), 2020. Patent Search Based On Deep Learning. https://doi.org/10.27162/d.cnki.gjlin.2020.006206

4. Chen, G.R. (Nanjing University of Science and Technology), 2021. Research on methods for patent retrieval and core patent identification. https://doi.org/10.27241/d.cnki.gnjgu.2021.000806

5. Stephen, R., Hugo, Z., Michael, T. (2004) Simple BM25 extension to multiple weighted fields. In: Proceedings of the thirteenth ACM international conference on Information and knowledge management. Washington. pp. 42–49. https://doi.org/10.1145/1031171.1031181.

6. Han, Y.N., Liu, J.W., Luo, X.L. (2021) A Survey on Probabilistic Topic Model. Chinese Journal of Computers, 44(06): 1095–139. https://doi.org/10.11897/SP.J.1016.2021.01095

7. Xu, K. (Dalian University of Technology), 2017. Research on Query Expansion for Patent Retrieval. https://d.wanfangdata.com.cn/thesis/ChNNaW5lclRoZXNpczIwMjMwMjEwEghZMzI4OTY2ORoIOG14c2ppa2o2o=

8. Magdy, W., Jones, G.J.F. (2011) A study on query expansion methods for patent retrieval. In: Proceedings of the 4th workshop on Patent information retrieval. New York. pp. 19–24. https://doi.org/10.1145/2064975.2064982.

9. Xie, P.L. (Hebei University), 2013. Research of Analysis and Quantization Algorithm for Reference Relationship of Scientific Literature. https://doi.org/10.7666/d.D349716

10. Liu, B.X. (2012) Research on optimizing literature retrieval based on citation relationships and clustering analysis. Information Studies: Theory & Application, 35(6): 5. https://doi.org/10.16353/j.cnki.1000-7490.2012.06.024

11. Wu, X.D., Huang, W.B. (2022) Semantic Search Strategies and Case Analyses in Intelligentialized Upgraded Retrieval System. China Invention & Patent, 19(05): 74-79. https://doi.org/10.3969/j.issn.1672-6081.2022.05.009

12. Hu, Y.Y., Chen, Y. (2019) The Strategy of Boolean Search Combined with Semantic. China Invention & Patent, 16(S2): 161–6+82. https://doi.org/10.3969/j.issn.1672-6081.2019.z2.030

13. Wang, F., Lin, L.F., Yang, S., Zhu, X.W. (2013) A semantic query expansion-based patent retrieval approach. In: 2013 10th International conference on fuzzy systems and knowledge discovery (FSKD). Shenyang. pp. 572–577. https://doi.org/10.1109/FSKD.2013.6816262.

14. Li, B. (Beijing Institute of Technology), 2015. Research of Patent Semantic Retrieval Based on Domain Ontology. https://kns.cnki.net/kcms2/article/abstract?v=jBOcPZekd6MUKk9X_-MftWGAr9x1W56E5Fcd3wDDO8quDD_zZ451lDw4ZbIAlMlPWJ1O2mU9lnmH_SdjDDEWK9zbjkvdsn8iZRoopH9f7IUQX3P5DjLGHA==&uniplatform=NZKPT&language=CHS

15. Wu, S.X. (Guangdong University of Technology), 2018. Patent Document Semantic Retrieval Based on Character Convolutional Neural Network. https://doi.org/10.7666/d.D01524502

16. Lv, T.T. (Beijing Jiaotong University), 2018. Construction and Application of Knowledge Graph Oriented to Professional Literature: Taking Machine Learning Literature as an Example. https://doi.org/10.26944/d.cnki.gbfju.2020.001085

17. Ma, G.B. (Harbin Institute of Technology), 2021. Research on Patent Knowledge Search Based on Knowledge Graph. https://doi.org/10.27061/d.cnki.ghgdu.2021.000794

18. Li, J.Q., Li, B.Q., You, X.D., Lv, X.Q. (2020) Computing Similarity of Patent Terms Based on Knowledge Graph. Data Analysis and Knowledge Discovery, 4(10): 104–12. https://doi.org/10.11925/infotech.2096-3467.2019.1157.

19. Wei, W., Li, X.J., Wang, J. (Institute of Computing Technology, Chinese Academy of Sciences), 2020. A knowledge graph-based intelligent patent retrieval method and system. CN111274332A [P/OL]. https://d.wanfangdata.com.cn/patent/ChJQYXRlbnROZXdTMjAyMjExMTkSEENOMjAyMDEwMDYxNDk0LlgaCGsxYXRiZW40.

20. Wang, R.Y. (Hebei University of Technology), 2018. The Research on Patent Classification Based on Deep Learning. https://doi.org/10.27105/d.cnki.ghbgu.2018.000051

21. KhudaBukhsh, A.R., Bennett, P.N., White, R.W. 2015 (2015) Building effective query classifiers: a case study in self-harm intent detection. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. New York. pp. 1735–1738. https://doi.org/10.1145/2806416.2806594.

22. Liu, B., Feng, L., Wang, F., Peng, Z.Y. (2016) Patent search and analysis supporting technology innovation. Journal on Communications, 37(03): 79–89. https://doi.org/10.11959/j.issn.1000-436x.2016055

23. Hu, J. (Guizhou University), 2018. Study on Knowledge Acquisition and Reuse Methods for Innovation Design of Mechatronic Products. https://d.wanfangdata.com.cn/thesis/ChJUaGVzaXNOZXdTMjAyMzAxMTISCUQwMTQ5Mzg4MxoIOHU1aThmaGg=