

Research on the Application of Natural Language Processing Technology in the Field of Information Retrieval

Jiayang Dai, Dong Zhou*

Hunan University of Science and Technology & School of Computer Science and Engineering,
Hunan, China

Abstract

In recent years, with the rapid development of China's Internet information technology, all walks of life in the field of information retrieval through the use of natural language processing technology, can effectively screen information and improve work efficiency and accuracy, but in the process of application, there are still some problems. This paper mainly analyzes and studies the application effect of natural language processing technology in the field of information retrieval, and analyzes and studies the basic principle of information retrieval system, the level of natural language processing and the application of natural language processing technology in the field of information retrieval.

Keywords

Natural Language Processing Technology, Information Retrieval Field, Using.

1. Basic Principles of Information Retrieval System

1.1 Information Indexing and Storage

The main purpose of information retrieval system is to help users quickly screen, query relevant information, so as to improve work efficiency and accuracy. The information indexing technology is in the information retrieval system, after obtaining the relevant information, through the way of information indexing, to mark and record the information, so as to facilitate users to search in the future. After the user tags the relevant data information, the data information is stored in the information retrieval system, and the screening data is arranged and combined, so as to facilitate the user to organize in the future.[1.2.3].

1.2 Requirement Analysis and Retrieval

The information retrieval system can archive and store the searched and screened content, and then analyze the content, so as to facilitate users to find relevant content in the information retrieval system in the future. After the user analyzes the retrieval data, it is necessary to store the filtered and indexed contents of the information retrieval system, so as to avoid the system load and waste of system resources caused by too much filtered data when users use the information retrieval system in the future.

1.3 Content Judgment and Feedback

When information retrieval system retrieves and screens relevant data, users need to judge according to the retrieved content, so as to screen out the content they want to find. When users enter the content they want to find in the information retrieval system, the information retrieval system will automatically screen out the content they want to make intuitive judgment for users.

2. Levels of Natural Language Processing

2.1 Speech Hierarchy

Voice level as the name suggests is the way users can through the voice to the information retrieval, this way of oral find needed information retrieval on the speech recognition system especially seriously, if oral language or voice recognition system cannot receive timely information retrieval by users, so will result in speech recognition, failure to retrieve other content. Professional speech analysis and processing systems will directly recognize the language, and even form oral text content according to the speech.

2.2 Part of Speech Hierarchy

Speech level mainly have a prefix, suffix, words of stems, in the process of natural language processing can be according to the content of stems simple expansion of new words, even a word plural, singular, and so on, this will be conducive to automatic identification, information retrieval system based on user selected words, automatically provides a few other expansion of words.

2.3 Semantic Level

Semantic level can automatically analyze the meaning of the word through the content of the context in the document, and even analyze the number of times the word appears in this paper and the meaning of the number of times. There is not only one kind of word semantics, users can understand the meaning of the word in the context of the text. Some words even have more meanings, which need to be judged according to different contexts. This type of word processing is called semantic hierarchy. In the process of semantic level processing, in addition to searching the synonym meaning, we also need to search the implied meaning of the word in the word, so that users can fully understand the meaning of the word.[4.5.6].

2.4 Context Level

Context level mainly refers to that the information retrieval system will automatically supplement and expand the relevant content according to the context of the text in addition to searching relevant content through the information retrieval system, so that users can understand other meanings of the word, so as to meet the needs of users from multiple angles. This way of information retrieval and communication can effectively express the meanings of various meanings of words in different contexts, which can enrich the user's knowledge of the words, thus forming a perfect knowledge base of the retrieval system and facilitating the exchange and learning between users and the information retrieval system.

3. Application of Natural Language Processing in Information Retrieval

3.1 Unified Model of Natural Language Processing and Information Retrieval

Natural language processing technology has been widely used in information retrieval system, the data is retrieved by the researchers in order to make the information retrieval system has more accuracy, of information retrieval systems need to continue to reform and innovation, continuous improvement on the existing problems, the natural language processing technology combined with information retrieval system, so as to build a unified retrieval model. China's main technical staff by adopting direct method, extension methods, extract method, method and unified method to transform to deal with different situations on behalf of the meaning of language, which can effectively improve the usage of words, to reduce the gap between natural language and information retrieval, make both supplement each other, harmonious and unified. At present NLP technology is a common application system of natural language processing technology in information retrieval in China. The specific framework is shown in Figure 1 below.

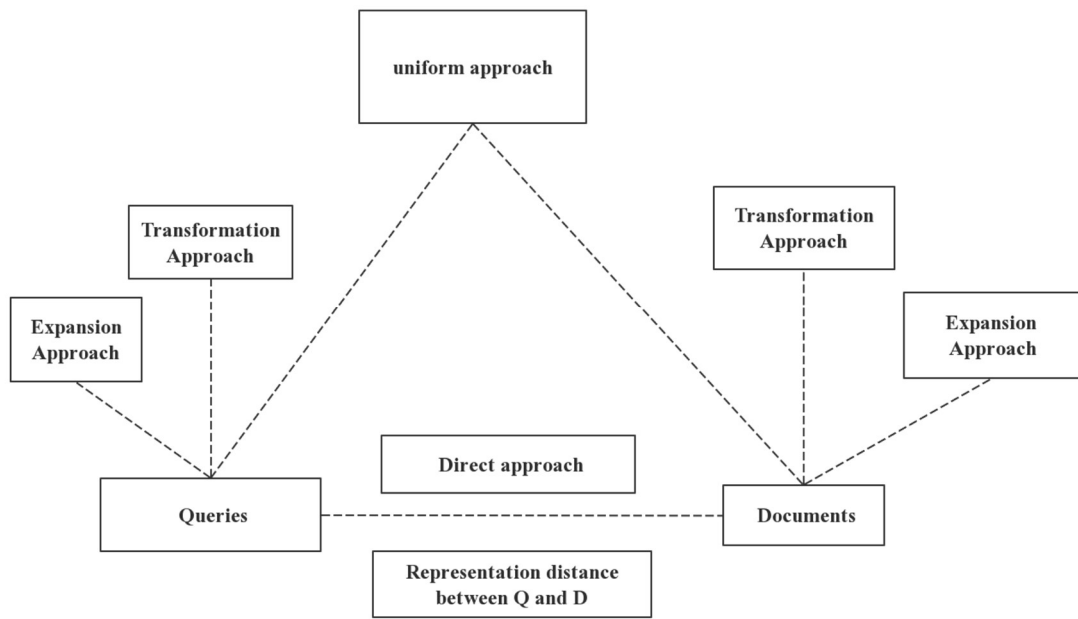


Figure 1. NLP IR framework

3.2 Application of Natural Language Processing Technology in Information Retrieval in a Broad Sense

The application of natural language processing technology mentioned above in information retrieval generally refers to document retrieval and so on, and the following is mainly through abstract retrieval, question answering system, information extraction and other angles to the application of natural language processing technology in information retrieval. This paper mainly uses TREC technology to screen the information of different length paragraphs in the information retrieval system. Figure 2 shows the technical staff's screening of query length and result length. It can be seen from Figure 2 that query length with longer length is more advantageous in natural language processing technology, which can quickly screen literature content and improve retrieval efficiency and accuracy. When retrieving short documents, the information retrieval system not only needs to ensure the accuracy of retrieval but also needs to process the language simply. Therefore, it can be seen that the question-and-answer system is more suitable for adopting natural language processing technology.

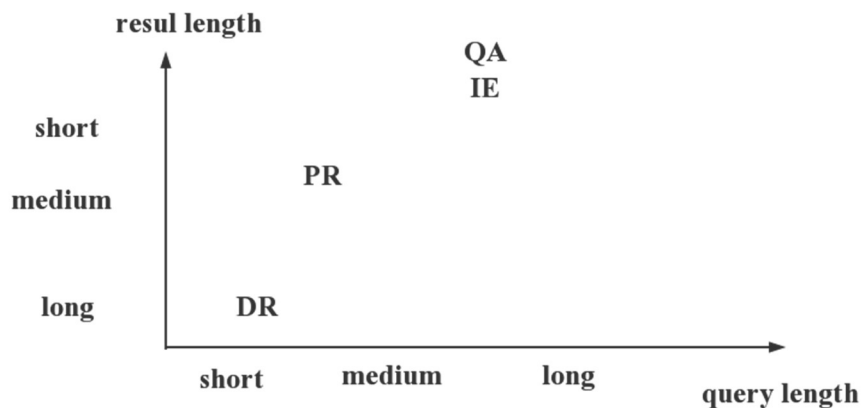


Figure 2. retrieves documents, paragraphs, questions and answers, and information extraction systems according to query length and result length

4. Conclusion

In recent years, with the continuous efforts of Chinese technicians to develop and innovate, natural language processing technology has been better applied in the information retrieval system, which can provide content quickly, efficiently and accurately when users are sifting literature. It can be seen that natural language processing technology and information retrieval system are inseparable, both are indispensable. Technical personnel need to continue to improve the two systems, optimize the system functions, so as to improve the efficiency of the system, so that the system can better serve users.

References

- [1] Juan Li ,Chen CAO . Development of information retrieval system in the era of big data [J]. China Science and Technology Information,2020(9):105-106.
- [2] yuan zhong. Research and application of natural language processing in information retrieval [D]. Zhenjiang: jiangsu university of science and technology,2019.
- [3] Han Zhang ,Zhichang ZHANG . Application analysis of Natural Language processing in information retrieval [J]. Computer Fans,2018(2):199.
- [4] Chandrasekaran M K , Mayr P . Report on the 4th Joint Workshop on Bibliometric-Enhanced Information Retrieval and Natural Language Processing for Digital Libraries at SIGIR 2019[J]. ACM SIGIR Forum, 2019, 53(2):3-10.
- [5] Meesad P. Thai Fake News Detection Based on Information Retrieval, Natural Language Processing and Machine Learning[J]. SN Computer Science, 2021, 2(6):1-17.
- [6] Modi S, Pandya D. VQAR: Review on Information Retrieval Techniques based on Computer Vision and Natural Language Processing[C]// 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2019.
- [7] Zhang K. Research on the Optimizing Method of Question Answering System in Natural Language Processing[C]// 2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS). IEEE, 2019.
- [8] Zhang Y, Pei Y . Research on the Application of Artificial Intelligence Technology in Island Tourism[J]. Journal of Physics: Conference Series, 2021, 1852(3):032015 (6pp).