ISSN 2663-8169 Vol. 2, Issue 1: 29-32, DOI: 10.25236/IJNDE.2020.020107

The Natural Language Processing Based on Internet Natural Annotation Resources

Liu Chunhua

Yunnan Open University, kunming Yunnan 650223, China

ABSTRACT. *Natural language processing has its own characteristics since its birth, and gradually integrate rationalism and empirical attention into the corresponding application field. With the growing maturity of Internet technology, natural language processing shows its own advantages based on the application of Internet natural annotation resources, and also plays a role in promoting Internet applications. This paper analyzes the natural language processing based on Internet natural annotation resources, hoping to provide reference for the integration and development of Internet and natural language.*

KEYWORDS: *Internet Natural annotation; Natural language processing*

1. Introduction

With the development of Internet technology, Internet application has been an indispensable part of people's daily life, which is very convenient for people. The requirements for natural language processing of natural annotation resources are becoming more and more high, and the relevant personnel should clarify the concept of natural language processing of natural annotation resources, analyze and expound it academically, so as to increase the efficiency of computer use and make it better to provide services for the production and life of human beings. At present, natural language processing has two more mainstream research methods, one is rationalism method, the other is empiricism method. However, the application of rationalism method is relatively few. With the development of the times, the research method of combining rationalism method and empiricism method is more suitable for the completion of natural language processing, among which many problems are highlighted, among which the innovation of natural language processing research method needs to be improved. In general, natural language processing has become more suitable for machine algorithms and applications, and has played a very important role in practical operation. And scholars in various places also put forward many opinions and suggestions for the development of natural language processing, which enriches the connotation of natural language processing and promotes the further development of natural language processing and application. Natural annotation resources play a very important role in natural language processing. Besides, in some methodological aspects, it can be put forward relevant things.

First, there needs to be a large scale to support the development of natural annotation resources. There are many resources on the Internet, which can provide favorable conditions for the development of natural annotation. Second, in natural language processing, shallow processing can be carried out on a large scale. In the process of computer application, natural language processing cannot be engaged in deeper language analysis, so it can only be processed at a shallow level, which becomes the basic tone of computer language processing. Third, natural language processing cannot be separated from the integration of human wisdom. Relevant technicians should constantly enrich and develop their own professional knowledge, have a deeper understanding of natural annotation, and help natural language processing more scientific and reasonable. When relevant personnel obtain data through “natural Annotation”, they should give full play to their subjective initiative, analyze and judge the acquired data, eliminate the unnecessary interference, minimize the error, and make the natural annotation work more scientific and reasonable as a whole. Fourth, when dealing with natural language, we should analyze and integrate a language resource as much as possible. If the direction of thinking is too single, it will destroy the reasonable development of “natural annotation”, and bring a lot of difficulties to natural language processing. Although people have some research results on natural language processing at present, there are still many problems in the specific application of natural annotation resources, which need relevant personnel to combine their own professional knowledge. According to the characteristics of natural language itself, it needs to be actively thought and innovated, finding more scientific and reasonable ways, and making contributions to the development of natural language processing.

ISSN 2663-8169 Vol. 2, Issue 1: 29-32, DOI: 10.25236/IJNDE.2020.020107

2. Definition and Basic Types of Natural Annotation Resources

2.1 Definition of Natural Annotation Resources

The definition of natural standard resources is basically consistent with the meaning expressed by “user generated data”, which is developed to serve better communication and communication among users on the Internet. It contains a lot of resources, such as microblog, user's log, forum, web page, etc. For ordinary network users, these resources facilitate communication, and also learn a lot of knowledge, record their daily life and so on. However, from the perspective of relevant disciplines, these resources can be regarded as some manual annotations, which are applied to the processing of language information. In fact, when carrying out the above activities on the network, we usually have no concept of natural annotation resources. Of course, for users, these also need not be considered. In a word, when network users communicate on these platforms, they inadvertently help with natural language processing, make various language expressions more accurate, and form a kind of obligation “Annotation”.

2.2 Basic Types of Natural Annotation Resources

The following cases could clarify the basic types of natural annotation resources more clearly. First, when you open the Baidu search page and click the news category, you can find that there are many “spaces”, “punctuation” and “beginning or end of sentence” contents, which are natural annotation contents in the computer language. It can help users understand the content that the page wants to express more clearly and get more information. Users can find a lot of keywords in the page, such as news, webpage, know, stock, people's livelihood, etc., forming a “display natural Annotation”. In addition to this kind of “display natural Annotation”, there is also an “implicit natural Annotation”, which requires professionals with professional knowledge to find out. For example, string is one of them, which can help different concepts to be closely linked and lead out many related contents through one word. If we search “iPad is one kind”, we will get many corresponding sentence patterns, such as “iPad is a hybrid device”, “iPad is a new type of product”, “iPad is a touch-screen computer”, etc. Although only half of the sentence patterns are typed out, many related sentence patterns and contents can be obtained through the natural annotated network resources.

Second, natural annotation resources are not only web pages, but also have many other contents, such as blogs. Users can publish their own articles through the blog. In the process of using, they can add some “tags” that belong to the author's characteristics, so that readers of the blog can be more familiar with the author through the tags, forming a “display natural annotation “. If your article is about health preservation, you can add labels such as “health”, “nutrition” and “health preservation” to make your article more recognizable, and it is also convenient for users with corresponding needs to find articles for reading. In terms of individual needs of individual users, these labels are personalized, but in terms of the overall audience group, many similar groups are classified into the same annotation, which has certain popularity. Compared with the classification by experts, these classifications are closer to the needs of the public, so they are also very easy to be accepted in the application. In addition to blogs, many social networking sites also use tags, such as Facebook, helping users search for relevant content.

Third, Wikipedia is a very important content in the basic types of natural annotation resources. It shows the “show natural Annotation” in a very specific and detailed way, reflecting its scientific and reasonable part. For example, you can try to search for the keyword “ Steve Jobs” in Wikipedia, which can clearly show the introduction of jobs. Among them, there will be some bold and obvious fonts that attract users' attention and form a “display natural Annotation”. These annotations can help computer language improve more descriptive words and express it systematically. Moreover, through the description of relevant words, users can have a very accurate understanding of Jobs and learn about its appearance, background, achievements, etc. In this process, natural annotation resources play a great role in finding the most suitable words to describe “Steve Jobs”. After computer language analysis, it will show and introduce the content of “Steve Jobs” to users. At the same time, the application of Wikipedia also helps users to understand all aspects of basic knowledge. It is a very practical tool, and has a very advanced expression and application in knowledge content and retrieval function.

Fourth, the basic types of natural annotation resources are very rich, and there are many kinds. In short, as long as you are using Internet technology, you will definitely be exposed to the use of natural annotation resources. For some more professional technicians, they can go beyond the superficial search to a deeper “user query log” form. These logs are also a form of natural annotation resources from another perspective, helping relevant personnel to conduct network activities more scientifically and reasonably. And natural annotation is not only limited to the form of text, but also can form the form of link through the intercommunication between web

ISSN 2663-8169 Vol. 2, Issue 1: 29-32, DOI: 10.25236/IJNDE.2020.020107

pages, so as to achieve a closer relationship. For example, the microblog attention function we often use makes the natural annotation more specific, playing an important role in the Internet application, realizeing a richer function and improves the promotion and utilization of various platforms.

3. Case Analysis of Natural Language Processing Based on Internet Natural Annotation Resources

3.1 The Effect of "Showing Natural Annotation".

For those punctuation marks, the natural marking resources in the form of sentence beginning and sentence ending, to a certain extent, have affected the automatic segmentation function of Chinese. At present, there is a huge Chinese Web corpus in the network application. With this corpus as the foundation and support, we can mark the display of Chinese characters according to a certain probability. Under the condition of Internet, we can deal with the natural language of natural annotation resources more scientifically, and help the network users get a better experience when they search for Chinese characters.

3.2 Help the Language Express More Vividly through the Natural Annotation of Quotation Marks.

In some news on the web, we often see the application of quotation marks. Quotation marks can automatically extract some interesting phrases or sentences, and form a quantitative tracking on the network, which can reflect the overall political, economic and cultural state. By adding quotation marks, we can create some catchwords. This kind of "show natural annotation" makes the language more prominent in the huge corpus and more convenient for people's expression.

3.3 The Effect of "Implicit Natural Annotation".

In the text of Internet, there are many "implicit natural annotation". They sort out many adjectives reflecting users' emotions through a lot of combining sentences. On many real-time social platforms on the Internet, there is a lot of information that can be provided to big data in the background, so as to collect and sort out a large number of expressions about emotions, making "implicit natural Annotation" more scientific and reasonable. "Implicit natural Annotation" will make the natural language processing expression based on the Internet natural annotation resources more human-oriented, and promote the further development of natural annotation.

3.4 The Natural Language Processing Based on Internet Natural Annotation Resources Can Do Better in Association Performance.

The algorithm of machine learning is applied to the function of tags mentioned above. It expresses words more specifically through the association among many tags on the network, so that when people input some text without tags. Some related tags will be generated automatically. For example, when you type the text "I should have missed you", some corresponding labels will appear on the platform immediately, such as emotion, mood, breakup, me, love, predecessor, etc. These tags are related to the text that the user typed, and even express the content that the user wants to express from the side. Through the algorithm of machine learning, the "explicit natural Annotation" becomes more vivid, and it is more appropriate to the user's ideas when expressing, which reflects the intelligence of natural annotation resources.

3.5 Application of User Query Log

In the Internet, there is also a kind of "natural annotation" in the background of search. Natural annotation is carried out by users' query logs. For example, we can query the high-frequency words used by Google users for a period of time, finding out which words have corresponding relationship with pneumonia, and give early warning to the public according to the query results, so as to minimize the incidence of pneumonia. This kind of natural language processing based on the natural annotation resources of the Internet, compared with the traditional way of conducting relevant investigations through the Centers for Disease Control and prevention. It does not make much difference in the accuracy of the survey results. Through the method of user query logs, it is very convenient to obtain the data to be investigated, and the time is also short, which improves the work of the investigation efficiency. More importantly, the public can be warned as soon as possible to minimize the harm of the disease.

ISSN 2663-8169 Vol. 2, Issue 1: 29-32, DOI: 10.25236/IJNDE.2020.020107

3.6 Natural Language Processing Based on Internet Natural Annotation Resources Cannot Be Separated from the Support of Professional Knowledge.

The development of natural labeling cannot be separated from the efforts of relevant professionals, which requires a lot of professional knowledge as the basis to complete. For example, there are great differences in the structure of the two syllable-verb complement in Chinese, among which the structure of the words will be more variable. Among them, there are many infixes, forming an implicit natural annotation in computer language. Usually, the closely combined two syllable-verb complement structure will add an object after it or use an object before the verb. For example, the word “broken” can be expressed as “broken cup” or “broken cup”. We also need to have a full understanding of language knowledge and language structure to help the relevant staff get the results of natural annotation more quickly.

4. Conclusions

This paper discusses the related problems of natural language processing based on Internet natural annotation resources, hoping to make positive efforts to the development of natural annotation. First of all, we should clearly understand the definition and basic types of natural annotation resources, and learn and recognize the concept and types of natural annotation from multiple perspectives and levels. Natural annotation has its own characteristics. It needs a strong theoretical knowledge to support it so as to better research and development. It needs relevant staff to study steadily and constantly put forward innovations, which could meet the needs of the development of the times and promote the application and development of natural labeling. Based on the natural language processing of natural annotation resources on the Internet, this paper enumerates many closely related cases to help analyze in order to have a deeper understanding of the natural needs of machines. This paper starts with the case of natural language processing based on Internet natural annotation resources, helps to analyze the development of natural annotation resources, and combines the characteristics of “display natural Annotation” and “implicit natural Annotation”, and enumerates the cases of natural annotation of quotation marks, algorithm of machine learning, application of user query log and so on. Through this process, it would help people understand the nature annotation resources. Moreover, it would have more profound cognition and understanding, and contribute to the further development of computer algorithms.

References

[1] Hou Jiateng, Chang Wei, Lin Guanfeng (2019). Design and implementation of intelligent customer service robot based on natural language understanding technology [J/OL]. Electronic Technology and Software Engineering, no.23, pp.238-240.

[2] Hou Jiateng, Chang Wei, Lin Guanfeng (2019). Design and implementation of intelligent customer service robot based on natural language understanding technology. Electronic Technology and Software Engineering, no. 23, pp.238-240

[3] Zhang Hong (2019). Application Analysis of Artificial Intelligence Technology in the cultivation of English discipline Core Literacy. Software Guide (Educational Technology), vol.18, no.11, pp.86-87

[4] Huang Wenjuan (2019). Application of search engine based on artificial intelligence in digital books. Information Technology, vol.43, no.11, pp.93-96+102.

[5] Liu Yijing, Xu Linjie, Dai Qifeng (2019). NL2SQL technology based on natural language processing and deep learning and its application in BI enhancement analysis. China Informatization, no.11, pp.62-67.