Design and Implementation of Related Literature Recommendation in Nstl Literature Retrieval System

Wang Lina

Baicheng Normal University, Baicheng, Jilin 137000, China

Keywords: Nstl literature retrieval, Recommendation, Digital library

Abstract: China has proposed a subsystem to calculate the similarity of secondary literature and realize the real-time recommendation of relevant articles. They are designed for the National Science and Technology Library (NSTL) literature retrieval system. In this system, recommendations are made theoretically; quantitative evaluation and effect evaluation are carried out to verify the results. This subsystem for literature recommendation can further improve the service quality of the NSTL literature retrieval system [1].

1. Introduction

People's lives have been continuously improved as economy and society develop rapidly. At the same time, there has been growing demand for cultural knowledge. With the dramatic progress of information technology, China has also established a digital library. Providing effective information services through effective digital resources is the key function of a digital library. The National Science and Technology Library is an authoritative literature information service provider in China. In its literature retrieval system, users can retrieve literature by inputting keywords. However, this search method has shortcomings. Maybe hundreds or even thousands of literature share same keywords, and it is troublesome for users to identify the literature they really need. Often, users fail to continue browsing after scanning dozens of irrelevant articles. So they are very likely to miss the literature they need, this inconvenience reduces users' interest in using digital libraries and searching knowledge online. Users that input same keywords probably have different needs, but the system will only present same results under the same keywords, which reduces the practicability of the literature retrieval system to a certain extent [2].

2. Definition of Recommendation of Related Literature

Relevant literature refers to the information recommended for users when they are browsing detailed information like titles and abstracts, and this information often bear a high degree of similarity to the literature that the users are browsing. At present, there are some digital libraries in China, such as CNKI, Stanford University Digital Library, etc., which have realized the recommendation of related literature. The main technologies adopted are divided into the following four types:

(1) The system automatically screen the articles most similar to the searched one in advance, and then collects these articles as a set of related literature, so that when users browse, they can directly see the related articles. This kind of calculation is relatively efficient, but the computation will be massive, thus causing low accuracy. And it is difficult to realize the recommendation of new literature [3].

(2) The system recommends related literature based on keywords, but in this way there will be too many possible results and the efficiency will be lower.

(3) Personalized recommendation. The personal information of the user is usually collected through the user's name and the relevant literature that the user has browsed, and literatures are recommended accordingly. This is an active recommendation, which is also a part of the functions that the digital library has realized in recent years. NSTL will also adopt the personalized recommendation.

  DOI: 10.25236/ictmic.2020.044

(4) Related recommendation of related literature. Traditional association rules need to be applied in the recommendation of relevant literature. Besides, strengthening the association rules and recommending relevant literature need a lot of literature information, and due to some limitations, it cannot be carried out generally.

3. Similarity Calculation of Relevant Literature

According to NSTL literature detection system, this paper proposes a method to calculate the similarity of secondary literature, and realize the real-time recommendation of related literature. In this way, users can get the relevant literature through keywords or related information and be recommended literature with the highest similarity through secondary calculation. This method not only improves the efficiency, but also effectively ensures accuracy, and can better adapt to the changing needs of the whole database. As for the calculation method of literature similarity, it mainly focuses on the calculation method of space model. The method of literature similarity calculation for secondary literature is mainly applied in NSTL system. Actually, a whole article cannot be used for searching, because information will be too large to process. There is no way to put a whole article into searching box, however, through secondary calculation subsystem, users can use title, keywords and abstract to better search for articles with higher similarity. The abstract, keywords and title of an article can be enough to differentiate articles. They can help to find articles with higher similarity, and the results can be more satisfying.

4. Steps for Using Subsystem of Related Literature Recommendation

Fig.1 Is the Subsystem for Related Literature Recommendation, “K” is the Number of Literature Recommended by the System, and the System Preset “K” for Less Than or Equal to “n-I”.

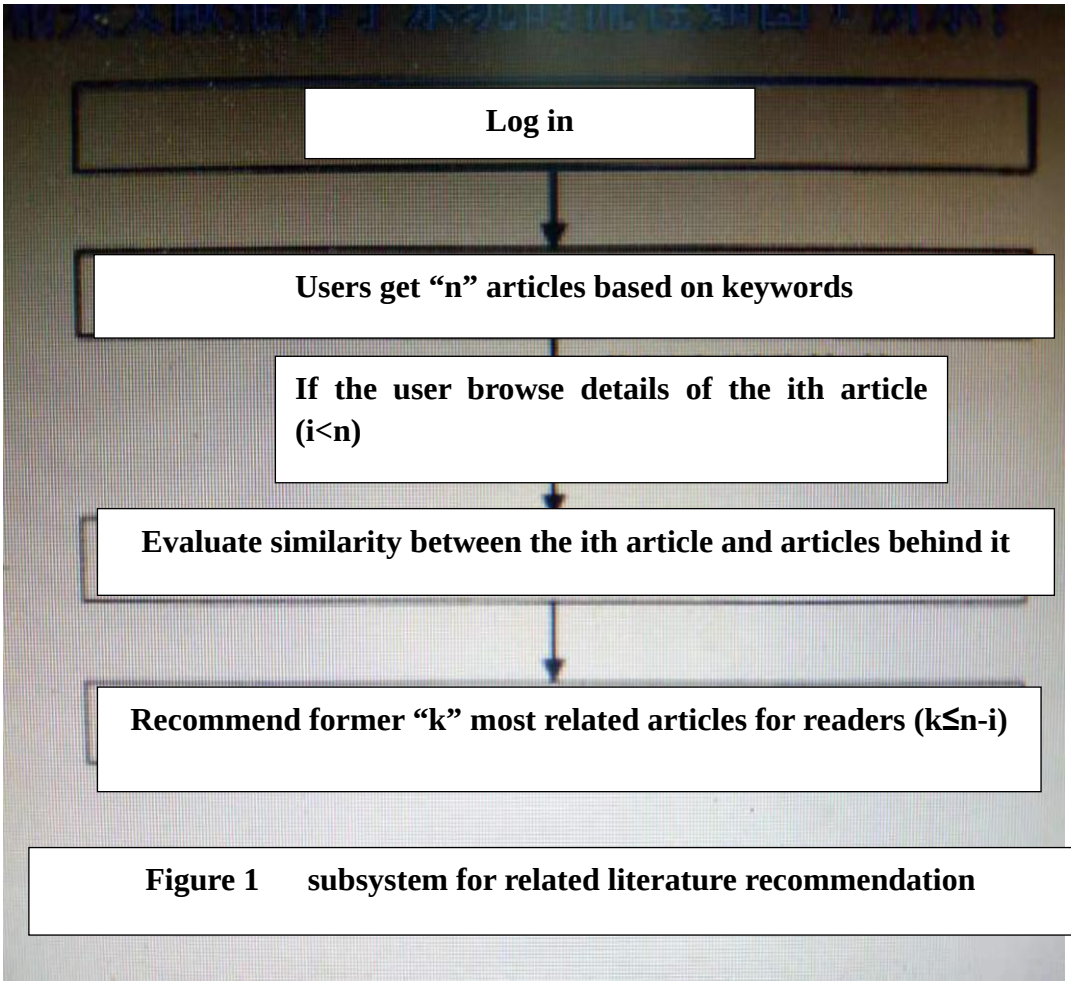

Figure 1 subsystem for related literature recommendation

When using the subsystem for secondary related literature recommendation to calculate, the

NSTL literature retrieval system can automatically return to the list, showing only the part of the title of an article. When a user wants to read this article, he can click the link to get further information like title, keywords, or abstracts; if the user is not interested in this article, he can go back to the former page to read other relevant articles. And keywords can be directly added into the corresponding articles and then messages can be delivered to the system to make a second search for articles with higher similarity. The second search uses keywords, titles and abstracts to better retrieve the available articles. In this way, the recommended article can be more similar to the original one and it is more convenient for users [5].

5. Functions of System

5.1 Login Management

At present, when users first log in this system, they don't need to register for an account. The system mainly serves researchers in this field. Therefore, in order to prevent others from stealing data and to ensure the integrity and security of the data, the system does not support registration. Only if the information uploaded by the administrator, can the user be registered, which not only protects the privacy of researchers belonging to this field, but also strengthens the feasibility of this system [6].

5.2 Text Content Retrieval

The precision of retrieval has increased after being reformed. When users search for the literature they need, they can find the literature they want in a most effective way, and the similarity will be higher. Besides, the system will automatically evaluate the similarity of the detected literature to make sure users get the optimal-related articles. However, domestic technology is immature at the moment. Similarity evaluation of the literature recommendation is currently called NDCG. NDCG mainly scores the literature information from the recommendation list, but all the literature is recommended by the secondary related literature subsystem, so NDCG is added for scoring. This design should also be combined with the literature recommendation subsystem so that the two systems can adapt to each other and recommend articles with higher similarity for users.

6. Conclusion

With the development of China's economy and society, science and technology has also improved a lot. The Internet has generally changed the way people live and work. China's database has also been enlarged due to the considerable development of information technology. However, it also presents a new challenge for China. For sake of the development of NSTL, China needs more automatic detection systems. The literature similarity calculation method of the secondary literature is used to process the articles retrieved by users. And the similarity evaluation remains a challenging task. NSTL literature retrieval system will automatically help users retrieve articles with a high degree of similarity through information like title, keywords, and abstract. But in the process, the evaluation won't be definitely accurate and requires human's participation. There are still more to study about the related literature recommendation in DSTL literature retrieval system.

References

[1] Zhang Zhiping(2010) Li Linna.Design and Implementation of Related Literature Recommendation in NSTL Literature Retrieval System .Modern Library and Information Technology, NO.7,pp.110-113.

[2] Establishment of National Science and Technology Library Lanbai Pilot Area Service Station (2016) .Gansu Science and Technology, Vol.32,No.23,pp. 1.

[3] Zeng Jianxun, Lai Yuangen, Deng Shengli, et al.(, 2011) Users Analysis of National Science and Technology Library Based on Customer Value Theory.Journal of the China Society for Scientific and Technical Information,Vol.30,No.9,pp.973-979.

[4] Meng Liansheng, Zhang Jianyong, Liu Xiaomin.(2010).10-year Overview of NSTL Literature Information Processing .Digital Library Forum, NO.10,pp.18-21.

[5] Zhao Long(2015).Science Tool for Scientific Research Team to Obtain Industry Information--NSTL (National Science and Technology Library Literature Center) CNC Machine Tool Field Information Portal.CADCAM and Informatization of Manufacturing Industry, NO.2,pp.75-76.

[6] Yao Changqing, Qiao Xiaodong, Zhang Lianjun, et al(2009).Spatial autocorrelation analysis based on the number of original NSTL requests.Information Science, Vol.27,No.10,pp.1501-1507.