

# The Study of M- Learning Resource's Intelligent Search in the "Internet+" Environment

Shu Jing Zhao<sup>1,a</sup> Zheng Hong Wang<sup>2,b</sup>

<sup>1</sup> Continuing Education Department, Beijing Institute of Technology, Beijing, Haidian Zone, China

<sup>2</sup> Continuing Education Department, Beijing Institute of Technology, Beijing, Haidian Zone, China

<sup>a</sup> Email:1414415057@qq.com, <sup>b</sup> Email:13391729985@163.com

**Keywords:** M-learning Resource; search engine; automatic text classification; heuristic intelligent search technology.

**Abstract.** In the development of today's human science and technology, course content and curriculum system are constantly changing; new concepts put forward frequently; subject border gradually becoming blurred, Course cross and fusion becoming more and more intense in the "Internet+" Environment. Thus, it is urgent to set the rules for researching the historical and current learning resources by Category. This paper based on the existing technology of search engine and automatic text classification makes an attempt to put forward a new heuristic intelligent search technology.

## Introduction

With the development of mobile computing technology and wireless communication technology, mobile learning, as a branch of digital learning, attracted much attention. At the same time, in many fields it has put into practice and made many achievements. But so far, the development of mobile learning is still in its primary stage and there are still many problems need to be solved. The most important one is how to research and collect learning resources on the internet. What's more, Resources bank has turned up in the resources construction agenda of the Ministry of Education which means we have to make a breakthrough on the intelligent search technology.

## Search Engine

Search Engine is a system which can provide users with the retrieval service by collecting information automatically from the Internet, filtering and analyzing. Specifically, it is an information retrieval tool. Firstly, search engine makes an uninterrupted web scraping for a given URLs, then extract information and establish index. At the same time store it to the index database. When the user submits the search conditions, it can quickly find out information in the database for users. Different search engine has different types in the implementation, but the basic function structures are the same. General search engine is based on the internet which is including five parts: page collector, text analyzer, indexer, Information searching device and UI. As shown in figure 1.

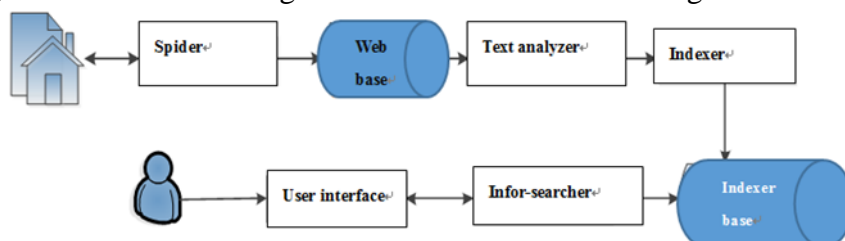


Figure1 Flow Chart of Searching

**Page collector.** Page Collector is exactly a kind of spider. The main function is collecting large numbers of page information on the Internet according to the demands of customer. And at the same time, it stores the new page information in the database in the form of document in case of the

following indexing and retrieval. In order to avoid the invalid link, it gives a regular access to all the web pages.<sup>[1]</sup>

**Text analyzer.** Text analyzer will deal with the original web pages according to the form of text in advance. Filtering out noises of the text, such as advertising, navigation and other useless information, extracting parts are the useful words. Then cut the documents into segmentation, delete meaningless operations. Different search engines have different index models and so as text analysis technology.

**Indexer.** Firstly, the indexer analyzes the existing text, then takes out key words as term which including web page, URL address, web page code, page keywords, positions of the key words, page size, generated time and the relationship with other web chain. Building positive index and inverted index, so as to inversion table and index glossary. Mainstream search engine generally also supports a variety of formats such as doc, TXT file, PDF, etc.<sup>[2]</sup>

**Information searcher.** Information searcher looks up learning resources in the index database according to users' demands. Ranking the qualified information according to specific algorithm and withdrawing the documents which are greater than set threshold. Finally, returning search results to users according to the ranking from high to low.

**User interface.** The user interface provides users with a visual input/output place. It will accept users' query requests and return results to the users in the form of a list displayed on the page, at the same time, users can choose what results to see.

The above five parts implement all the functions that a relatively complete search engine system should have. The biggest advantage of search engine is that it can help users find what they truly want efficiently in the voluminous information flow. However, nothing is perfect, the shortcoming of search engine is that the searching results usually interfered by advertisements which will mislead the users.

### **Automatic text categorization**

A completed learning activity demands a lot of corpora. However, the text information climbed down from the Internet is messy and redundancy, so it is necessary to deal with the climbed information and filter out the useless information. Firstly, to get more accurately target text retrieval, it is necessary to process contents of the text information by Chinese word segmentation. Secondly, refine entry content, stop entry the unspecified segmentation like "I", "the" and "they". Thirdly, mind the entry selection. To reduce the workload and unnecessary interference terms, it is necessary to choose entry what the users truly want. Fourthly, mind the digitizing text. In order to have an efficient utilization of selected entry, the text needs to be in digital form which computer can analyze and process. Fifthly, get text categorization model. Finally, after the same processing, user should give the text which has been classified to classifier who is responsible for distributing the text into specific category automatically.<sup>[3]</sup>

### **Heuristic intelligent search platform**

Based on the single function and user's inefficient experience, the author combines automatic text classification technology in the field of data mining with the structure characteristics of current network education resource. Then, a heuristic intelligent search platform for online education resource is put forward which is on the condition of keeping the easy using of search engine. The main goal is to improve user's experience and integrating existing education resources effectively. Recombining, reutilizing and subliming knowledge are main features.<sup>[4]</sup> The specific system framework as shown in figure 2.

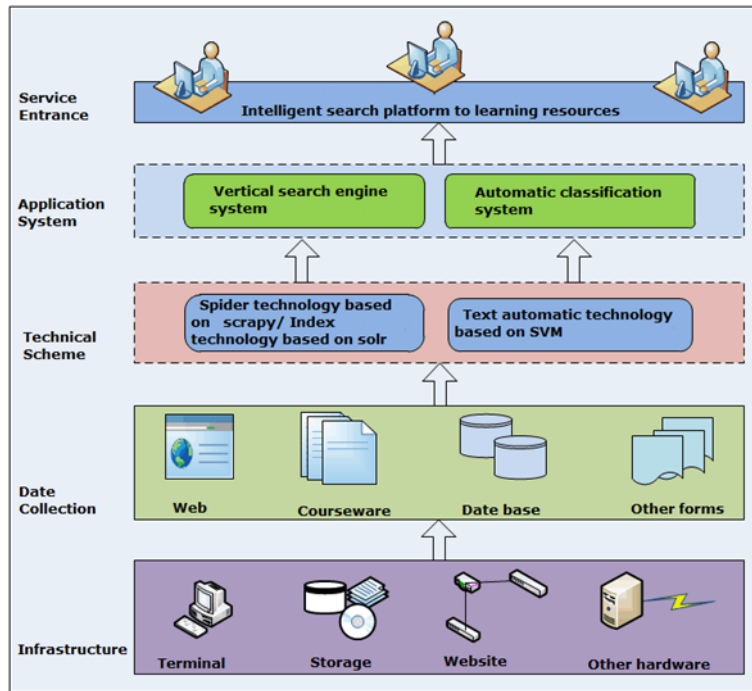


Figure 2 Framework of intelligent search platform

Analyzing the technical scheme of this platform we can know it well, as shown in figure 3. After acquainting resources from online in data module, it is time to preprocess the data including redundancy removal, word segmentation and demonizing. At the same time, data store should be complimented. The following task is classifying the collected resources including sample feature selection, feature representation and classifier training. Of course, the new resources also need to be classified by the classifier training. Finally, resources index should be established on the database.<sup>[5]</sup> Only in this way, can it be meeting the user's individual needs better.

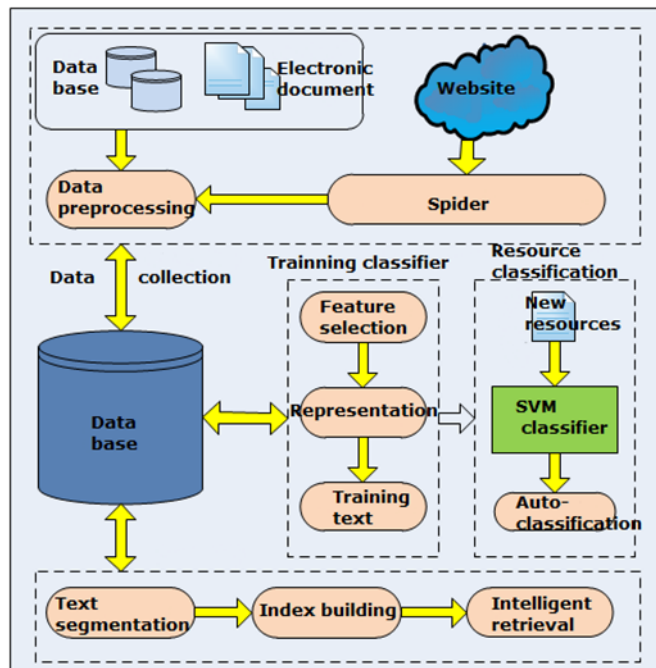


Figure3 Work Flow Chart of intelligent search platform

## Conclusion

In the conclusion, the general framework of M-learning resources' smart searching system includes three parts: resources platform, management platform, user's website. Searching system collects target resources by spider and stores them into resources platform. Intelligent sorting, marking keywords and classifying taking place on the resources platform. At the same time, a structured storage can be realized on this platform. Management platform also can achieve routine management of resources which includes adding, removing, fixing and inquiring. In addition, management platform also can combine different resources into new learning course according to resources classification and import them into the trading website to meet user's needs. Furthermore users can have trading activity on the resources platform, such as buying what they really need. All in all, the smart searching system gives users a very efficient way to get M-learning resources. However the research in this paper is not deep enough and lack of practice testing. The future researchers in this field should make this intelligent system deep exploited and more maturity. What's more, the reprocessing of M-learning resources is also a main field which should attract more attention in the future.

## References

- [1] Jialing Zhou. Focused crawler principle and key technology research [J]. Science & Technology Information. 2008(22)
- [2] Srinivasan P, Pant G, Menczer F. Target seeking Spiders and their topical performance. Proc of SIGIR Conference on Research and Development in Information Retrieval . 2002
- [3] Aggarwal C, Al-Garawi F, Yu P. Intelligent crawling on the World Wide Web with arbitrary predicates. Proceedings of the Tenth International World Wide Web Conference . 2001
- [4] Christopher J.C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition[J]. Data Mining and Knowledge Discovery . 1998 (2)
- [5] Jianbao Wu, Zhenguo Ding. Distributed search engine based on Map/Reduce research [J]. Modern Technology of Library and Information. 2007(08)