

Design and Implementation of Science and Technology Information Collection System Based on Web Block

Shuming Wei

Tianjin Nankai High School , Tianjin ,P.R.China 454000

Key words: Information collection; Web block; Design

Abstract: This article simply introduces the existing information acquisition system, this paper expounds the system of science and technology information collection system based on web block frame, key technology of the system carried on the thorough research, and proposed the corresponding solution measures, hope to be able to realize the design of the science and technology information collection system based on web block and the application.

1.Introduction

With the rapid development of Internet, people on the ways and means to obtain information has made great progress, has changed the way people's production and living. But at the same time of all kinds of information bring convenience, also faced with the situation of the information explosion, facing the vast amounts of information resources, only to rely on artificial to complete the collection and processing of information, not only waste a lot of time and energy, but also difficult to obtain the required information. In this case, the science and technology information collection system based on web block can effectively avoid this problem, improve the efficiency and accuracy of people to gather information.

2.Existing information acquisition system

Because modern science and technology information update quickly, at the same time has a strong timeliness, information is mostly concentrated in the BBS, announcement or news. Scientific and technical workers in the scientific research information query, usually into several specialized technology site to collect the required information, the information collection channels are fixed. Based on the technology of web blocking information collection technology, a fixed source of science and technology information can be collected, automatic data collection, and for scientific and technical workers to provide the latest information you need, reduce the burden of the scientific and technological personnel to consult a large number of web sites, and also reduces the omission phenomenon in the process of collecting information.

At present, the popular information acquisition software is more in the market, such as a madman, companies search through, octopus, shadow, vehicle, etc., they are used to collect the template configuration way to customize user needed to realize the collection of the function. Although this method is relatively cumbersome, but the rules of the acquisition is relatively perfect. In the process of use, users need to have some knowledge of web page code, so it is difficult to get a wide range of applications. Science and technology information collection system based on web block design, the web technology is applied to the design process, web content can be modeled, and then presented to the users, the user again carries on the simple operation to complete the information collection. Operation is not only very simple, but also relatively fast, can be more widely used.

3.Science and technology information collection system based on web block system framework

Science and technology information collection system based on web block can be based on a fixed research website or section to track information content, while understanding the release of information in the plate and update the situation, can help technical personnel more timely to master of science and technology development, so as to adjust the research strategy.

Specific design ideas: first of all, according to the fixed sites or plates capture the target page source code, according to the structure characteristics of the web page divides into blocks, learn the user's specific needs, design positioning of the plates to do well. To extract the content, using keywords to realize information filtering, finally, the information collected will be collated, and then presented to the user to browse.

The system design is mainly divided into three modules, respectively information collection module, task management module and database management module. The core of the whole system is information collection module, according to the requirements of science and technology personnel from the fixed site or plate to collect the required information, and then process the acquired information. Task management module can be achieved on the existing user configuration tasks in the system to run and maintain, the user to search information for the first time, can record the user customized information, and then store this information to the task file. When the users browse the site again, direct access to the task before you can meet the needs of users, can also automatically update their mission. Data management module's main function is to do a good job of the control of database data, such as data storage, data deleting, data to be obtained. When making the development of the system, using B/S framework, to choose the Java as a development language.

4.Key technology of the system

At present, most of the acquisition software requires the user through the html code for web page to do a good job of configuration template, the users need to have knowledge of the relevant skills, due to a lot of science and technology personnel lack of relevant knowledge, so it is difficult to use software for information collection. Through the web block technology, no longer need to master relevant knowledge of the code, can be directly to locate the information they need, and then complete the collection work, has a strong suitability. At the same time, the system can be finished according to the user's specific setting module to the corresponding website information collection, such as once a day. But usually technology website information release frequency is not high, the collected information tends to have a repeat, if the acquired information is not filtered, all saved into the database, not only takes up a lot of space, but also lead to data redundancy. When carries on the system design again, also need to think carefully in this respect.

4.1Web block and positioning technology

Because the html file in the browser is usually in the form of a number of partitions presented, each partition belong to a piece of content. When block of html web pages, can use the `< div >`, `< table >` and other tags to complete. Because when mark and mark on the display can be tied for display, can also be nested, can effectively combine mark content, the nested structure of an html page into more than one piece of content without contact.

In some html language writing a relatively standard web pages, tag appears in the form of are in pairs, each start tag must be equipped with an end tag. When to read files, only need to find the start tag, you can to determine the beginning of the regional boundaries, when have an end tag appears, can be made into a partitioned according to the content of the beginning and end of web pages.

4.2Data elimination heavy

System in the process of data acquisition, if the web site update frequency is lower than the sampling frequency, there will be a lot of duplication of collection, long repeat acquisition makes database data redundancy and reduce the use of performance of the whole acquisition system. When this system to carry on the design, in order to avoid the occurrence of this phenomenon, select the data inserted into the database.

Generally in the completion of information collection, the information is stored in the database in the order in which it was collected, the information mostly arranged in time order. The first information collected is often the latest information, it will also be the first to the database insert, the order in which the system inserts information is sorted by time. In the second data collection, in order to effectively reduce the repetition data, before the first data compared with the first piece of

data, if the information is different, the choice before the second data compared with the first data, repeated operation, until the encounter with the acquisition of the first data of the same data so far, then the data is the latest data for the web site, and then insert the latest data into the database. In every time to the acquisition of new information, you need to compare the new information with the previous information, check the repeatability of information, although this method can largely reduce the repetition rate of information, but need every piece of data comparison, greatly reduce the efficiency of the system.

When this system to carry on the design, using reverse way to insert information, thus avoiding the step of comparing data, improve the operation efficiency of the system. This system in the information acquisition, will be the last message as the first insert the data inserted into a database, so the database last recorded by the data will be presented by the latest data. During the data collection system for the second time, just need to take out the last data in the database compared with the first piece of data, if the last data is the same as the fifth data, article 5 before data does not exist. The article 5 before data record, new data can be collected, and can effectively avoid data to a recurring problem, greatly reduces the complexity of the system to collect information, guarantee system high-efficient operation.

5. Conclusion

Science and technology information collection system based on web block mainly has two major difficulties when the design. The first difficulty is the web block and positioning technology, can use the start and end tags to locate. The second difficulty is data elimination heavy, the use of information inserted in reverse chronological order, has carried on the consummation solution, hoping to really design a scientific and technological information collection system, real realize the goal of its design and performance.

References

- [1] Zhou Yuzhi, Liu Zhanming, Wang Bo, Han Yuchen, Wu Jinying. Network information on the automatic collection of technical difficulties and solutions [J]. Science and Technology Communications, 2013,06: 204-205.
- [2] Wu Zhenxin, Xie Jing, Zhang Zhixiong, Hu Jiying, Li Wenyan. Construction of network information collecting and archiving management platform of international important scientific research institutions [J]. Notes, 2014,22: 100-104.
- [3] Chen Jianguo. Web-based Web site news collection system design and implementation [J]. Journal of Jinggangshan University (Natural Science), 2012,02: 54-57