Xml and Its Application in Library and Information Retrieval

Cai Lianghui
School of Media Arts, Wuhan Qinchuan University, Wuhan, 430204, China

Abstract. the Extensible Mark-Up Language Xml is a Simple Data Storage Language, Using a Series of Simple Tags to Describe Data, and These Markers Can Be Used to Establish a Convenient Way, Although the Xml Space to Occupy More Space Than the Binary Data, But Xml is Extremely Simple and Easy to Master and Use. Xml Organized Management of Web Information Resources Distributed in Logic, Web Information Retrieval System Structure to Design a Professional Content, and Provides a Way for Users to Obtain the Special Contents Oriented Information Resources. a Good Data Storage Format, Scalable, Highly Structured and Easy for Network Transmission Characteristics So, and Xml is Determined by the Excellent Performance. the Paper Presents Xml and Its Application in Library and Information Retrieval.

1. Introduction
The Library as a Cultural and Educational Institution and Social Information Center in Advocacy and Development Has a Irreplaceable Role in the Process of the National Reading. Reading is Related to a Nation's Core Value System and the Ideological Foundation, is to Strengthen the Cultural Identity, Cohesion of Country People, Inspire the National Spirit, and Improve the Quality of Citizens. Chunhua Society, an Important Way to Construct Core Values. in the Past Ten Years, the National Reading Has Gradually Risen to Improve National Quality, Improve the Comprehensive Competitiveness of the Country, to Promote an Important National Strategy to a Harmonious Society Civilization [1].

The Study of Literature Information Retrieval, Can Be Very Good to Help Us to Make Use of the Resources of the Library, Whether Paper or Electronic Resources. Before, I to Our School Library is Not Very Familiar with, Each Search Books Are Always Like Look for a Needle in the Ocean, Wasting a Lot of Time, and Don't Know Our School the Room Can Also Find a Lot of Online Resources, Can Do, Reading and So on. the Learning of This Course, I Can Quickly Find What I Need the Book, Not to Waste a Lot of Time to Go. Moreover, I Will Also Use Electronic Room Database to Help You Learn. I Personally Think of Students This Course is Very Necessary, Let Students Faster, Better Use of the Resources of the Library, This Library is to Maximize, for Our Students' Learning Will Play a Great Promoting Come into Action.

The Library Collection of Various Types of Human for Thousands of Years Accumulated, Various Kinds of Vectors of Document Information in Germany [2]. the Berlin Library Door is Emblazoned with the Phrase: “Here is the Treasure House of Human Knowledge, If You Can Master It Key Words, Then All the Knowledge You Are Here Said.” the Key is That the Ability of Using the Library.
Learn to use the literature information retrieval is the basic ability of students must have the authority of the board of education in twenty-first Century 1996 entitled “Learning: the treasure within” report put forward four kinds of basic ability, “learning” is one of the four basic the ability to “learn knowledge.” in a sense is to learn to read a lot of knowledge, but the knowledge of the people of the most simple and direct way or learn from books. So, students will to grasp the method of literature retrieval, can be learned in books. Through reading, can accept predecessors to explore nature observe social achievements, but also through newspapers, books and network to collect the latest information needed for innovation, and on this basis. Students only learn to use the library to open the treasure house of knowledge gate go, get master and use, the human for thousands of years of accumulated all spiritual wealth. Learn how to use library students need to improve their own quality.

XML extensible markup language (eXtensible Markup Language). Mark refers to the information symbols computer can understand, by this marker, computer can deal with various kinds of information between the articles. How to define these markers, can choose the international general markup language, such as HTML, can also be used by relevant parties such as XML this decision markup language is the language of the scalability of. XML is simplified from the SGML. It is mainly used in XML, XSL and XPath.

This section is a basic definition of XML, a widely accepted explanation. Simply, XML is a data description language, although it is the language, but usually, it does not have the common language of the basic functions of computer recognition and running. Only rely on another language to explain it, make it to achieve the effect you want or computer to accept.

360 browser search tools, it is a retrieval tool of our daily with the operation of its understanding, so you can save some time [3]. The database selection of Tsinghua Tongfang CNKI database, Tsinghua Tongfang journals into 8712 kinds, including 1750 kinds of core journals, Tsinghua Tongfang covers arts and medicine all subjects provided the search field most, Tsinghua Tongfang in addition to providing some necessary search fields, also provides the full text, such as references, intelligent retrieval, fund classification, ISSN and other fields can be more convenient and more accurate retrieval of desired articles. Retrieved from the navigation function, navigation Tsinghua Tongfang way up, there are a variety of navigation system, the search query is more convenient freely. On the citation query, Tsinghua Tongfang has specialized library, from the language interface, Tsinghua Tongfang, traditional Chinese, The three English language, which is used by different languages, is more widely retrieved.

2. On the Retrieval Report of Chinese Database Information Retrieval

No matter is the primary school, middle school, or high school textbooks, everything on the tip of the iceberg, we must constantly look for some of the knowledge they need and, before we are more asked the teacher, classmates, because there is no library so find the book is not possible, not familiar with the network technology is also very [4]. The lack of methods and so they can find information before high school really. Especially after entering the university to learn more, how to improve their ability to acquire and utilize information; how to update their knowledge structure after entering the workplace, so they don't fall behind has become very important. So, the ability to acquire knowledge and information has become one of the important ability of college students must have.

\[
\tilde{W}_{j,k} = \sum_{l=0}^{L-1} \tilde{h}_{j,l} x_{k-l \text{ mod } N}
\]

(1)

The majority of researchers to attract eyeball semantic retrieval based on tradition. To realize the information retrieval from keywords to improve the level of knowledge level. Based on the traditional information retrieval based on keywords, has achieved great success, but it can not express the user's query request fundamentally [5]. Semantic retrieval is the information retrieval and artificial intelligence
technology, Natural Language Processing search technology that combines the technologies of it from the point of view of semantic understanding of information retrieval and analysis object the search request, and improves the retrieval performance. But due to the field of natural language understanding and artificial intelligence and other limitations, the semantic retrieval technology will have a long-term in-depth research Process.

The development of network technology. To provide services through the Internet database types increasingly rich [6].

If you are a novice new to XML, so may not by definition is to know what is XML. Maybe you can change an angle to understand what is XML; from the application to XML from XML, what can be done to recognize it, it should be defined more than the empty to you more help.

$$\frac{\partial W}{\partial P} = \begin{pmatrix} \frac{\partial W_1}{\partial a_1} & \frac{\partial W_1}{\partial a_2} & \frac{\partial W_1}{\partial d_1} & \frac{\partial W_1}{\partial a_3} & \frac{\partial W_1}{\partial d_2} \\ \frac{\partial W_2}{\partial a_1} & \frac{\partial W_2}{\partial a_2} & \frac{\partial W_2}{\partial d_1} & \frac{\partial W_2}{\partial a_3} & \frac{\partial W_1}{\partial d_2} \\ \frac{\partial W_3}{\partial a_1} & \frac{\partial W_3}{\partial a_2} & \frac{\partial W_3}{\partial d_1} & \frac{\partial W_3}{\partial a_3} & \frac{\partial W_3}{\partial d_2} \\ \frac{\partial W_4}{\partial a_1} & \frac{\partial W_4}{\partial a_2} & \frac{\partial W_4}{\partial d_1} & \frac{\partial W_4}{\partial a_3} & \frac{\partial W_4}{\partial d_2} \\ \frac{\partial W_5}{\partial a_1} & \frac{\partial W_5}{\partial a_2} & \frac{\partial W_5}{\partial d_1} & \frac{\partial W_5}{\partial a_3} & \frac{\partial W_5}{\partial d_2} \end{pmatrix}$$

(2)

XML applications are mainly divided into two types, document type and data type. Here are several common XML applications:

1. custom XML+XSLT=>HTML, one of the most common document applications, XML stores the XML data of the entire document. Then XSLT transforms XML and parse, combines HTML tags in XSLT, and finally becomes HTML, which is displayed on browsers. A typical example is the post on CSDN.

2. XML as a database, the data type is one of the most common applications. We use the XML API (MSXML DOM, JAVA DOM) to access and query of XML. To achieve the message board, you can often see with XML as the database [7].

The connection and difference between XML and HTML:

- The HTML document mainly contains the display mark-up, mainly for browsing.
- HTML contains a predefined set of tags.
- XML can define its own tag set, which is extensible.
- The XML tag represents the logical structure of the data, and carries out strict syntax and semantic checks.
- The XML document separates the data from the display format.

As we are learning in school students, is to learn and master the knowledge, but it is more important to learn to acquire knowledge [8].

3. Information Retrieval on the Service of University Library Based on Xml

Before going to the library to borrow books, often find the classification of books, but do not know in the first few floors. So sometimes to a layer to find in the library is very troublesome. Now, the study of information retrieval, know the classification of documents, find books on the side it is more. Such as “H language”, “I representative literature”, “K represents the history, geography class, they all belong to the nature of the arts, so I knew what to go to the library to find a few buildings. Learning information retrieval harvest is far more than that, such as the flexible use of various retrieval methods and in order to achieve satisfactory retrieval results, as I know that several of the world's Encyclopaedia of knowledge and so on, these make me feel benefit.
A UN official said: “the future of illiteracy is not illiterate, but did not learn learning methods.” Literature retrieval course is a basic course of practice oriented and practical, with the rapid development of information technology and extensive application of information network, information retrieval course it is very necessary. As a kind of information retrieval methods and skills, help people adapt to technology and production, ready to add, update their knowledge, improve the knowledge structure is reasonable, improve the ability of solving problems of science and technology [9]. The use of methods and skills to master a variety of retrieval tools, is to strengthen the information consciousness, improve the ability of self-study and also provides an effective way for the realization of lifelong learning.

XML applications and related technologies:
- DTD, XML Schema, CSS, Data Island, XSLT, DOM, SAX, XPath, XLink, XPointer

Application:
Electronic commerce, instead of database storage, data exchange, software configuration document record, intermediate transformation, is more convenient to use XSL parsing, and it is also easy to configure, Web services, content management, and Web integration [10]. The most typical is the J2EE configuration WEB server with the web.XML. Application of this estimation is very easy to understand. We just need the data into XML, and then load the application runs on us, according to the different data, do the corresponding operation. It has 2 and application. Similar points, different is, change the data in the database is normal, and the configuration information is often relatively static, the lack of change.

Wanfang Data Resource System is a large scale science and technology created on the Internet, business information platform, content involves natural science and social science disciplines. Including: academic journals, dissertations, conference papers, patents, foreign standards, scientific and technological achievements, policies and regulations, new records, institutions, technology experts and so on.

The source database created in 1989 Chinese science and technology magazine database. “The text and bibliography Abstract edition correspondence. The database contains since 1989 in natural science, engineering, agriculture, medicine and health, economy, education and Library and information science journals. More than 8000 kinds of database are classified according to < Chinese library classification > and all the literature is divided into 8 Albums: Social Science, economics and management science, education, library and information science, natural science, agricultural science, engineering and technology.

DTD basic concepts; declarations of DTD to elements; declarations of DTD to attributes; declarations and references of entities; reference to DTD by XML; validation of XML document validity;

DTD is the abbreviation of the document type definition (Document Type Definition).
Declare internal DTD:
<! DOCTYPE root element name [
<! ELEMENT... >]>

Declare external DTD:
<! DOCTYPE root element name SYSTEM “DTD-URL” >

External DTD:<! ELEMENT... (do not need to write the DOCTYPE.)

Element declaration: <! ELEMENT element name EMPTY|ANY|#PCDATA|Mixed|Children>
Children: all subelements must appear once and have to appear in turn:
<! ELEMENT element A (name, name2|name3) >
No other subelement <! ELEMENT name #PCDATA>
<! ELEMENT Name2 #PCDATA>
Element content type:?: 0 or 1 times
Selection of retrieval methods, commonly used method: divided method, investigation method, sampling method, tracing method, comprehensive method

The literature search clue document types according to various features of literature

It can adjust the search strategy to obtain the required information to adjust search strategy in the retrieval process; the retrieval strategy may be adjusted several times. Such as literature retrieval too much, to reduce the search range and it is necessary to expand the search scope. Based retrieval approaches and methods selected, through various index retrieval tools check the sequence number, and then check the documents or titles, obtained from the original literature source, at last.

4. Xml and Its Application in Library and Information Retrieval

Web information resources can be roughly divided into two categories: text and multimedia information (image, video) information, in order to effectively extract the needed information, Web information retrieval system of the new generation should toward faster search speed, higher accuracy, and can meet the personalized needs of users in the direction of development. This needs us in Natural Language Processing, a breakthrough in data mining and machine learning, and these technologies will be fully applied to information retrieval, so as to make people feel better Web information convenience.

Google's mission is to provide the best online inquiry service for you, to promote global exchange of information. Google developed the world's largest search engine, provides the most convenient method of online information search. The about 2000000000 page arrangement, Google may need to provide appropriate search results for users around the world, and the search time usually less than half a second. Now, every 150 million Google need to provide query service. Google innovative search technology and elegant user interface design of the Google from the first generation of today's search engine. Google is not only the use of keywords or talent shows itself the search technology, it found itself in advanced PageRank (TM) (PageRank) on the basis of technology. This patent pending technology can ensure that always will be the most important first search results presented to the user.

$$d_m^*(f) = \begin{cases} \frac{d_m^*(f_1)}{w_m(x,y)\times d_m^*(f_1)} + (1 - w_m(x,y))\times d_m^*(f_2) & x < N_m - w_m \\ \frac{d_m^*(f_2)}{N_m - w_m} & N_m - w_m \leq x \leq N_m + w_m \\ \frac{d_m^*(f_2)}{x > N_m + w_m} & \end{cases}$$

5. Conclusion

The collection of University Library Literature Resources expanding, continuously push forward the process of informationization. At present, the university library has a collection of large amount of electronic information resources and other advantages. As an important source of public library knowledge acquisition, information retrieval technology is the main tool of library service to the public, only the development and improvement of the information retrieval technology, in order to the library resources effectively. To convert the power of social development. As the higher school library, the task is arduous and challenging, only in the service based on the campus, on the public, to maintain the leading domestic technology, follow the international forefront, and continuously forge ahead, on the development of the times, to meet the needs of society, the development of services to colleges and universities the promotion of social progress, its complete preservation of the cultural heritage of civilization, the mission.

XML is a fast growing standard to exchange data between computer systems. Microsoft uses this copyright language (or standard) to describe the XML data of many Microsoft applications. Just like
HTML, extensible markup language XML (eXtensible Markup Language) is a markup language. It also depends on the description of certain rules can read the label and label application processing tools to play its strong function. This, from the XML name can also get a glimpse.

References
[1] Lu Juan, analysis of the current network information retrieval and filtering technology, scientific and technological information development and economy, 2016, 16, 23-29.
[2] Yang Xiaohong. The development and service innovation of literature and information resources in University Library. Science and technology information development and economic, 2015 (21), 12-19.
[3] Sun Guangwei, research on multimedia information retrieval technology, Journal of Jilin Architecture and Civil Engineering Institute, 2012, 6, 79-81.
[4] Zhang Yufeng. On knowledge retrieval and information retrieval. Chinese Library Journal, 2013, 5, 29-33.
[5] Zhao Jing, Zhang Hongye, Web information retrieval technology, scientific and technological intelligence development and economy, 2014, 20, 105-109.
[6] Ren Yan. Discussion on Library Information Service under the network environment. Information science, 2013, 1, 69-76.
[7] Party Yang Yang, XML information retrieval technology overview, Gansu science and technology, 2015, 10, 41-16.
[8] Cao Peng, the development of modern information retrieval technology, higher correspondence Journal (NATURAL SCIENCE EDITION), 2014, 8, 73-75.
[9] Wang Ran, the application of information retrieval technology in the University Library, the development and economy of scientific and technological information, 2015, 18, 30-32.
[10] Huang Zhenjiang, Li Yong, the current situation and development trend of network information retrieval, Journal of Heze Medical College, 2014, 2, 95 - 96.