

THE DESIGN AND DEVELOPMENT OF PROJECT-ORIENTED INFORMATION SYSTEM

Yuehong Zhao^{*}, Chao Liu, Hao Wen, Hezhen Zhang and Zhihong Xu

Institute of Process Engineering, Chinese Academy of Sciences

*Email: yhzhao@lcc.icm.ac.cn

ABSTRACT

In this paper, the idea of building a project-oriented information system based upon a specialized information database was discussed. It attempts to provide tools for helping researchers use Internet resources effectively in the course of their research. Based on this idea, a web-based project-oriented information system was constructed. The paper systematically expounds the design and development process of the project-oriented information system. Furthermore, examples of utilizing the project-oriented information system to obtain useful information and suggestions for specific projects were described. According to our discussion and utilization of the system, we believe that building a project-oriented information system can help researchers with their research projects.

Keywords: Project-oriented, Information system, Internet, Database

1 INTRODUCTION

It is well known that the quantity of information on Internet is immense, and increasing by over one million pages a day. The rapid growth of the global network offers a new information service for the scientific research, but the process of searching for information is difficult and time-consuming using general search engines. These retrieve much irrelevant information, especially when searching for information on a specific field. Under these conditions, developing a specialized information system is a good way to solve this problem (Xu, 1998; Peng, & Kong, 2000). Much effort has been devoted to this aspect, for example, in the field of chemistry, ChIN (Li, Yang, & Xu, 1999; The Chemical Information Network, 2002), ChemCenter (The American Chemical Society, 2002), and Chemweb.com (The World Wide Club for the Chemistry Community, 2002) are successful chemistry-oriented information systems. Most of these information systems focus on providing an interface for researchers to navigate and search related information. But the information requirements for scientific research projects focus on topics and ideas that need further analysis and study based on information stored in a specialized information system. In order to utilize the information in a specialized information system effectively in a detailed research project, building a project-oriented information system that can implement project-oriented information mining is an effective way to obtain and exchange information.

The objective of our work is to discuss a method for the overall design of a project-oriented information system, the collection of network resources, the organization of information, as well as the design and development of a related database system which gives researchers an interface for exploring and tracking cutting-edge and developing trends in a specific field.

2 THE FRAMEWORK OF A PROJECT-ORIENTED INFORMATION SYSTEM

The intention behind building a project-oriented information system was to set up an information service network and communication system based on database and Internet technology, involving many functional parts. Among them, the design and development of the computer platform and related database system are core parts of building such a specialized information system (Xu, Taylor, Geisler, Li, Lan, & Stevens et al., 2002; Liu, Li, Guo, Huang, & Yang, 2000; Wexler, 2001). There are many

development toolkits and software platforms available that support the construction of such an information system. But for the construction of a web-based project-oriented information system, there are other tasks that need to be considered seriously:

- The realization of mining web information for a specific field.
- Building a specialized information system with the proper organization of information according to a specific knowledge system.
- The construction of a project-oriented repository based on the specialized information system.
- The development of query tools for the information system.
- The design and development of a functional module that allows users to publish information and to communicate with each other.

Based upon the considerations above, the framework of a project-oriented information system can be designed. It consists of three major parts: a database, a database management system and a system for collecting network resources as illustrated in Figure 1. The detailed description of important parts of this system will be provided in following sections.

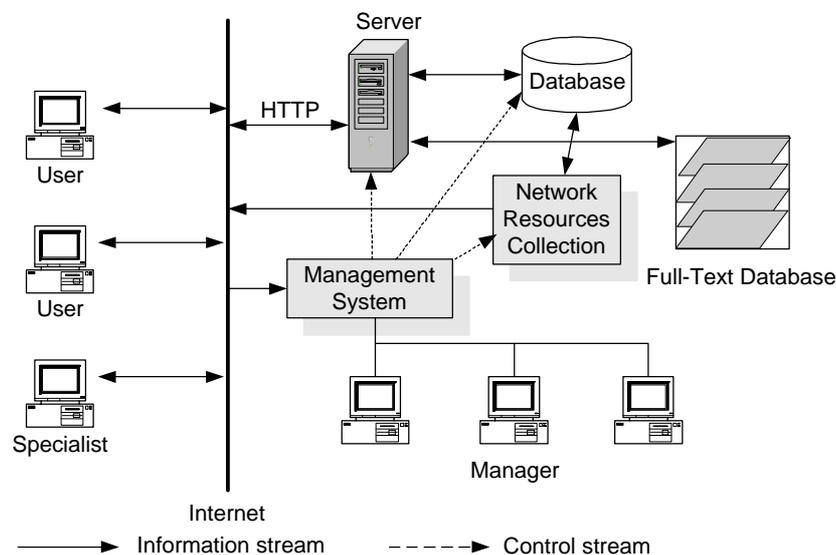


Figure 1. The framework of the project-oriented information system

3 THE DEVELOPMENT OF THE PROJECT-ORIENTED INFORMATION SYSTEM

A database system that records and manages of information to help researchers navigate and search for information for a specific project is the core of a project-oriented information system. It is needed because there are a wide variety of information resources on the Internet, such as technical reports, government documents, the journal literature and other types of information. Proper classification and organization of the information collected from Internet is the basis for design of the database system, which affects the information's validity to some extent. Giving researchers a user-friendly search interface that does not require prior knowledge of search technologies is another key part of this information system. In this section, key steps in developing the project-oriented information system, such as the classification and organization of information, the design of the database for organizing the information, the information search strategy, and the collection of network resources are discussed.

3.1 Classification and organization of information

There is no commonly accepted standard method for classifying information resources on the Internet. Only a few methods are usually used. For example, some suggest that information on the Internet can be classified using similar methods to those used in general information classification or using methods

based on how the information is disseminated and its interactivity (Si, Peng, & He, 2001). In specialized information classification, Li, X. X., Yang, Z. Y., & Xu, Z. H. (1999) systematically investigated related chemistry information resources on the Internet, and provided a suite of classification methods (Li, Yang, & Xu, 1999; The Chemical Information Network, 2002). In their proposed method of classifying information, information is divided into different levels and organized in a hierarchical manner, which helps users navigate and find related information.

In our work, the information in the project-oriented information system was organized into two administrative levels. One was for the overall information on the specific field. The other was for project-oriented information for detailed subjects or cutting edge projects. The information classification method proposed by Li X. X., Yang Z. Y., & Xu Z. H. (1999) was adopted for overall information organization and for each specific project or subject. It is assumed that information resources on Internet have the following first level categories (Liu, 2002):

- Publications & Libraries.
- e-business & Information Providers.
- Research Center & Institutes & College.
- Relative Organizations.
- Other Websites and Internet Services.
- Miscellaneous on-line Knowledge.
- Patents.
- Engineering databases.
- Newspapers.
- New Technology.
- Companies and Enterprises.

Furthermore, first level category can be divided into sub-categories. For example, in our work the category of publications and libraries was divided into CD, books, libraries, book subscriptions on the Internet, publishing companies, newspaper indices and periodicals.

3.2 Database design for information organization

In our work, a tree-structure was used to manage information in a hierarchical manner, consistent with multi-level information classification. The tree-structure consists of many nodes, each representing one category level. Relationships between nodes include brother relationships between nodes on the same level and parent-child relationship between nodes of the nearest two levels. In order to express the relationships between nodes, a relational database was used. The Entity-Relationship diagram for the database is given in Figure 2. The entities in Figure 2 are organized into three tables: category, url_class and url_content. Descriptions of these 3 tables are given below:

- The category table stores information about the information categories coupled to the nodes in the tree-structure, ClassID is the primary key, MeaningE, MeaningC, and ParentClassID for the identifier of Parent category are defined to describe one category. Nodes having same ParentClassId are brother nodes. The node without a parent is the root. Then the tree-structure is realized.
- The Url_content table maintains detailed information data downloaded from Internet. It gives a detailed description of the information collected. In order to keep track of the most of up-to-date information on the Internet, a field called updatetime was defined. As a piece of information can belong to several categories, the relationship between tables Category and url_content is therefore many-to-many. In order to resolve this problem, a relational table named url_class representing the many-to-many ship relation between the tables Category and url_content was designed. There are only two fields in this table, the field UID that represents the information identifier and the field ClassID that represents the category identifier.

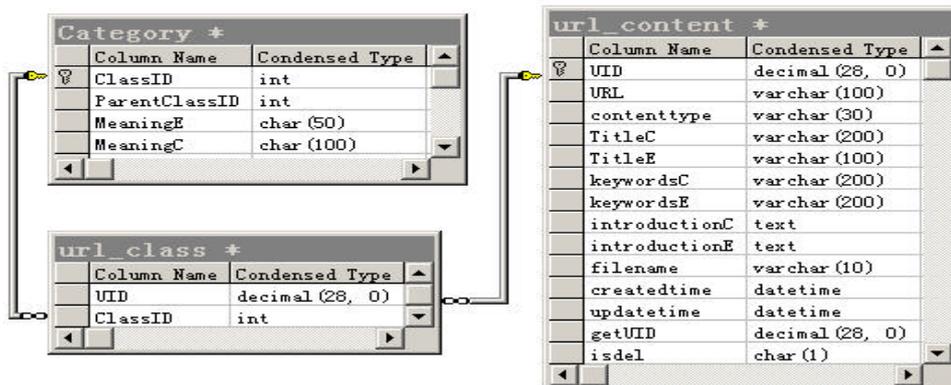


Figure 2. The structure of the category, url_class tables and their relation table

3.3 Network resources collection

General search engines are commonly used tools for searching for information on the Internet. However, the search results usually contain plenty of irrelevant information, which makes it difficult and time-consuming for the user to find the information they want (Liu, 2002). The poor precision of such search results usually cannot satisfy the information requirements of researchers in specialized fields. Because information on a specialized field always appears on a fixed number of web sites dedicated to such a field, and the keywords used in such fields are constant, modified search methods were developed based upon these two characteristics. Searching the selected web sites in a set of collected URLs or searching the Internet using general search engine with keywords from set of keywords compiled for a specific field is the method usually adopted. The compilation of the set of URLs or keywords for a specific field is the most important and creative step, and depends on experts in the specific field. In our work, the former method was adopted to help collect network resources. The URL set was obtained from experts that collected and evaluated URLs in a specific field. In order to improve precision and to reduce the amount of irrelevant information, the URL set is continually updated. The management and updating of the URL set was realized by using a database in our work. Figure 3 shows the table definition for this database.

Column Name	Condensed Type	Nullable
UID	decimal (15, 0)	NOT NULL
url	varchar (255)	NOT NULL
modifiedtime	datetime	NULL
gettime	datetime	NULL
status	char (1)	NULL
getflag	char (30)	NULL
referenceurl	varchar (100)	NULL

Figure 3. Table definition of table get_urls

Based on the collected URL set, a program was developed to fetch and analyze the web information automatically. Then the keywords, title, description and related links in web page were separated and automatically inserted into database by the program. In order to ensure the scientific validity of the information, each data statement underwent examination by the manager. At the same time, a brief description was added to make it more convenient for user.

3.4 The data search system

The data search is another important part of the system. It is the interface that the user uses to find the

information stored in database. In our work, two search methods were realized: a keywords search based on SQL for the overall information in the database, and a full-text search for project-oriented information.

Keywords search This search method is the basic method used in every part of the information system, and provides Boolean searching (AND, OR, NOT) to users. To allow the user access to information on the Internet, web-based technology was used to deal with user requests and to manipulate the database. In our work, PHP was used to implement this function.

Full-text search By using the full-text search, a search result contains more information than by using the keywords search. However, slower search speed and high index space consumption are disadvantages of full-text search. So we only use the full-text search for the project-oriented information library because of its smaller size. In our work, CGlimpse (Chen, 1998) was adopted to implement the full-text search, and it can process both English and Chinese characters. In CGlimpse, a special automatic indexing technique was adopted to optimize the construction of the indices, which results in a compromise between index space and searching speed. The search can therefore be carried out at higher speeds using a reasonable amount of index space. In order to support search requests from the Internet, CGI script is used.

4 APPLICATION IN PROCESS ENGINEERING

In our work, a process engineering project-oriented information system was built on Linux operating system. A web site (<http://159.226.63.156/metalnano1/>) based on the database system was constructed. The home page of the project-oriented information system is displayed in Figure 4. This web site collects and classifies related subjects or projects in the field of process engineering, such as chemical metallurgy, cleaner production and nanomaterials etc. To date, the number of information items is nearly 600. An information query function and a method of communicating with users were provided in this web site. Users can communicate with each other by clicking the 'Forum' hyperlink, which enables users to benefit from each other's knowledge. If a user has valuable information, he or she can submit the information by clicking the Your Words or Add URL hyperlinks. The keywords search input box was placed at the top of the home page allow the user to implement Boolean searching. Selectable lists were set up to allow the user to choose specific projects or subjects to search. Clicking on the 'Full-text search on scientific subjects' hyperlink will lead users to the full-text search interface and allow the user to define search options. This page also allows users to navigate information on specific projects. The interface for the full-text search on scientific subjects is displayed in Figure 5. For both search methods, the information retrieved was displayed in pages containing records with a brief introduction and a link to the original source.

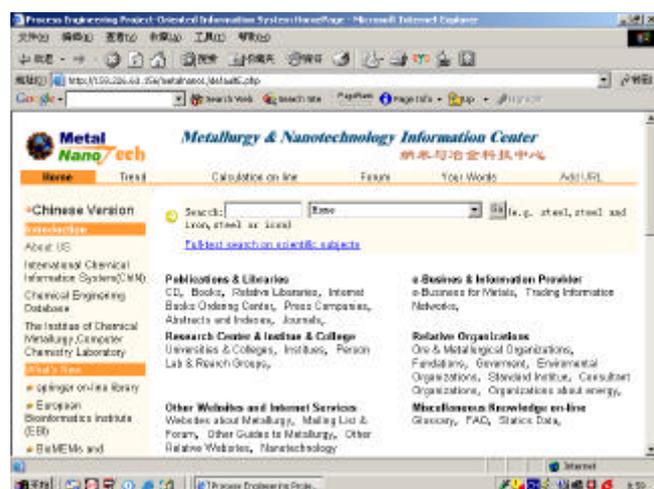


Figure 4. The Process Engineering Project-Oriented Information System Home Page

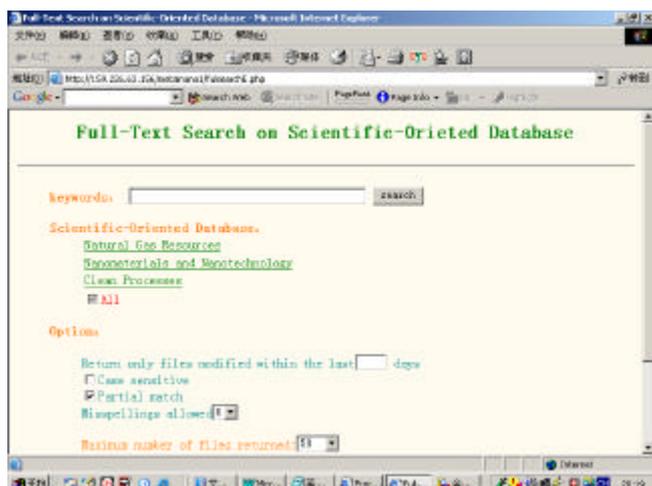


Figure 5. The interface of the Full-text search on scientific subjects

We used the process engineering project-oriented information system that we built as an information service on other related research projects and in the compilation of two books, for example:

- An information category on the utilization of clean energy was designed to help the chemical engineering, metallurgical and petrochemical industries promote the wise use of natural gas. Information from all over the world about the status of energy and the environment, the discovery of a new gas energy source (methane hydrate) and the advanced process technology of natural gas was collected and reviewed. The needs and problems encountered in developing the natural gas industries in China were also discussed. Moreover, based on an analysis of the relevant information, some integrated processes for liquid fuels and chemical products in China have been proposed. This provided the information for our research project (coal & gas to liquid technology research) and the compilation of a book entitled *Development of Green Process Engineering in 21st Century* (Xu, Wen, Guo, Li, Liu, & Zhao, 2002).
- With more and more researchers around the world focusing their attention on the application and development of nanomaterials, an information subject named *Nanomaterials & Nanotechnology* was reviewed and discussed. This provided the information service for the collaborative effort that produced the book entitled *The Application of Nanomaterials in Foreign Countries* (Zhu, 2002). This book also gives some advice on the national strategy for the study of, and research directions in, nanotechnology.

5 CONCLUSION

This paper presents a method of designing and developing a project-oriented information system that records information collected by project-oriented web information mining. The problems encountered in building the web-based project-oriented information system were also discussed. A process engineering project-oriented information system was set up as an example of a project-oriented information system based on advanced database and web technology. Users from anywhere in the world could access the information collected. Some application examples of this system were illustrated in the paper. These applications showed that building a project-oriented information system based on a specialized information system was a good way to explore and track the frontiers and research direction for the specific project in the conditions of rapid global network growth. It may be of help in the construction of project-oriented information systems in other fields.

In our work, the project-oriented information system only provided an information service for the research projects or issues we care about, it cannot flexibly provide valuable information according to other user's interests. More work remains to be done to provide an information service that can be individually tailored to a user's interests using the technologies of project-oriented search engines and artificial intelligence.

6 ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support from National Natural Science Foundation of China, Contract No. 59904004.

7 References

- Chen, F. (1998) *Study on Application of Internet Resource Discovery in Petro-Chem Information System*, PhD thesis, Institute of Process Engineering, Chinese Academy of Sciences, Beijing, China.
- Li, X. X., Yang, Z. Y., & Xu, Z. H. (1999) Access to Chemical Information on Internet ChIN's Web Page. *Chemistry Online 1*, 43-47
- Liu, C. (2002) *The Design of Project-Oriented Information System Based on Internet and its Application in Process Engineering*, PhD thesis, Institute of Process Engineering, Chinese Academy of Sciences, Beijing, China.
- Liu, Z. H., Li, X. X., Guo, L., Huang, S. H., & Yang, Z. Y. (2000) The Web Interface and Database Based Management for an Internet Chemical Resources Navigator. *Computers and Applied Chemistry 17* (5), 421-426
- Peng, X. R., & Kong, T. (2000) Development Study of a Specialized Information Service Network. *Information Studies: Theory & Application 23*(1), 29-33, 57
- Si, L., Peng, F. Z., & He, J. F. (2001) The Organization of Network Information Resources and Innovation and Development of Bibliographic Science. *Library and Information Service 9*, 21-24
- The American Chemical Society* (n.d.) Homepage of ChemCenter. Retrieved October 17, 2002 from World Wide Web: <http://www.chemcenter.org/>
- The Chemical Information Network* (n.d.) Homepage of ChIN. Retrieved October 17, 2002 from World Wide Web: <http://www.chinweb.com>
- The World Wide Club for the Chemistry Community* (n.d.) Homepage of *Chemweb*. Retrieved October 17, 2002 from World Wide Web: <http://www.chemweb.com>
- Wexlex, P. (2001) TOXNET: An evolving web resource for toxicology and environmental health information. *Toxicology 157*, 3-10
- Xu, X. F., Taylor, V. E., Geisler, J., Li, X., Lan, Z. L., & Stevens, R. et al. (n.d.) *Design and Development of the Prophecy Performance Database for Distributed Scientific Applications*. Retrieved September 6, 2002 from World Wide Web: <http://prophecy.mcs.anl.gov/publications/siam01.pdf>
- Xu, Z. H. (1998) A Study of Problem-Oriented Scientific Database System. *Scientific Database and Information technology paper collection (4th Volume)*, 19-21, Beijing, China: Science Press
- Xu, Z. H., Wen, H., Guo, Z. C., Li, J. L., Liu, C., & Zhao, Y. H. (2002) *Development of Green Process Engineering in 21st Century*, Beijing, China: China Petrochemical Press
- Zhu, T. (2002) *The Application of Nanomaterials in Foreign Countries*, Beijing, China: Chemical Industry Press