# Comparative Study on Subject Information Gateway of the World

Zhang Xiya，Xiao Xiaobo，Zhang Huijun (Xian Jiaotong University Library，Xian，China，710049)

***Abstract:*** *This paper compare several differences with other subject information gateway systems, such as embody scope, resource types, organization system, resource description, retrieval function , appreciation service, renewal and maintenance, cooperation model ,amount of data etc. For the more, the paper put forward some good ideas of developing Subject Gateway system.*

***Key word:*** *Network Resource；Subject Navigation；CALIS；Comparative Study；Subject Information Gateway*

It has become an important way for researchers in universities to obtain resources from Internet. Search engines such as Google are often the choice to obtain particular network information. However, due to lack of evaluation on qualities and standards of descriptions, sometimes it is very difficult to get valuable information of a search engine. Therefore, researchers urgently need a new model to organize and explore online academic information. Under this particular situation, starting from the mid 90's of 20th century, Hundreds of subject information gateway with valuable online academic resources to users were built up worldwide, providing valuable online academic resource to users. It helps the users to obtain valuable information more quickly and accurately. Currently, the very important issue we face is to objectively evaluate and compare the subject information gateway in order to improve system functions and serve users in a better way.

## 1. Current practice on subject information gateway in the world
## 1.1 Current practice in China

One subject information gateway built in China was one of CALIS' projects starting in 2000. Later on, Chinese Academy of Sciences, National Science Digital Library and Beijing Leisun Technology Inc. started to build up CSDL gateway and other subject information gateway.

**1）CALIS Subject Navigation Database**(http://navigation.calis.edu.cn) [1]

CALIS subject navigation database was the first program of collecting network resources in China. During the 10th Five-year plan period, based on the policy of "unified platform, unified standards, cooperation-building, separately maintenance, centralized services, resource nationwide shared ", according to rules of "pursuit of quality, emphasis on the application, taking into account the quantity", we realized the goal of " building up a international network resource navigation database to provide important academic website navigator and free academic resource navigator, based on the subject classification system which issued by Ministry of Education of PRC as a classification basis of navigation database".

The participants of CALIS subject navigation database are consisted of 53 University Libraries. So far, it contains one hundred and sixty thousand free academic resources, relating to 11 subjects, 79 first-class subjects. Resource types include references, full text,

multimedia, interacting resources, yellow pages, events, etc. This database was opened to all participant libraries starting from June 2006. It has been clicked seven hundred and sixty thousand times, providing students, instructors and researchers in universities free access of academic network information. It has the following main characteristics:

（1）Including 79 first-class subjects, one hundred sixty thousand records with comprehensive coverage and huge quantity.

（2） Several criteria were established and carried out, such as a unified resource evaluation and selection criteria , metadata and cataloguing rules. The resources are selected and catalogued by librarians with relevant academic background so the accuracy and authority of the database is guaranteed.

（3） A model with "cooperation-building and separately maintenance" is established , so does a model of "centralized services and resource shared nationwide" .

（4）Organized by subject classification system issued by Ministry of Education of PRC.

**2）CSDL Subject Information Gateway** （http://159.226.100.135/ejournal/CSDL－SubjInfo.php）[2]

Chinese Academy of Sciences, National Science Digital Library（CSDL）officially started the built-up work on subject information gateway in December 2001. Until July 2008, 5 subject information gateways had been built, they are: physics and mathematics, life science, book information, chemistry, resource and environment science; 9 unique gateways had been built, they are: microbiology，Qinghai-Tibet Plateau、Yangtze River drainage area ecosystem and environment, natural products and pharmacy, technology policy, new infectious diseases topics, marine science、seed plants in China, patent information；4 NSTL portal had been built, they are: Nano science and technology, cognitive science、food and nutrition、AIDS prevention and control.

CSDL subject information gateways refer related international standards on categorization system based on the uniqueness of each subject. For example, in the science information gateway, mathematics applies the method of classification used in American "Mathematics Evaluation" and German "Mathematics Abstracts". In the resource and environment information gateway, classification system of RESC and DDC are applied.

CSDL subject information gateway and software platform use abroad free software. Every single subject builds up its own relatively independent software, and sets up its own standards of resource selection and description. It mainly selects all kinds of free resource as well as important commercial resource. Its resource types are different, for example, life science resource has different types: newest resource, database, software, updating relevance, organization information, book resource, hotspot research topics, subject topics etc. Resource and environment science has the following types: data sets, software, text resources, interacting resources, organization website, multimedia etc.

So far, five CSDL subject information gateways have forty thousand data.

**3）Firstlight Knowledge Management System** （http://www.firstlight.cn）[3]

Firstlight Knowledge Management System is a subject information gateways released by Beijing leisun Technology Inc. in May 2005. It is a Wikipedia website provided database service, advanced research and consultant, all of which combined academy, research,

education, examination, management etc. It includes philosophy, social and human science, nature science, engineering technology, medical science, architecture science etc.

Firstlight is an academic website group constituted by 28 subject websites. Every subject website has knowledge news, international news, features, subject navigator, graduate admissions information, conference center, academic navigate, database, Firstlight stars, Firstlight focus, Firstlight discussion, Firstlight blog etc.

In 2007, Firstlight achieved further improvement in resource selection, resources Processing、Full-text index and Site Archive, particularly, The built-up of original ecological processing systems efficiently solved the problem of instability of navigator resources.

So far, Firstlight has more than four hundred thousand data.


## 1.2 Current Practice Abroad

Subject information gateways abroad started earlier. Early subject information gateways were usually built up and supported by big research projects. Later on, university libraries and research organizations joined in. Main ones are: British INTUTE(used to be RDN)、European DESIRE、German SSG—FI、American INFOMINE、LII etc. In this paper will mainly introduce and compare INFOMINE and INTUTE.

**1）INFOMINE**（http://infomine.ucr.edu）[4]

INFOMINE was started in January 1994 by University of California Riverside as a library project. Later on, other six libraries or colleges such as Wake Forest University joined in the project. It was the first tool of web resource organization provided by libraries. It was also the first website based on the virtual academic resource on web.

INFOMINE information resource has biology, architecture and medical science, commercial and electronics, culture and religion, social and human science, physics, mathematics and engineering etc. It has content development group、programming and system management、and iVia software group. Its main research focuses on web crawling and metadata assignment. Web crawling mainly uses some crawling tools to find new resource, and the latter it can allocated different field content with the application of iVia automation metadata assignment model.

INFOMINE project is founded by FIPSE, IMLS, and UC Riverside library. It has more than one hundred thousand data.

**2）INTUTE** (http://www.intute.ac.uk) [5]

INTUTE is the largest subject information gateway website in British as one of eLib projects started in 1994. In October 1999, eLib projects were substituted by RDN (Resource Discovery Network) projects. RDN was initiated by MIMAS of Manchester University, cooperated with many cooperators and funders. The core of the whole organization is an association including seven universities. RDN combined eight very famous subject information gateways, such as British Social Science Information Gateway (SOGIG), Biology Resource Navigate (BIOME), Physics Information Gateway, Engineering and Computer Science Gateway, Geography and Environment Information Gateway, Human Science Gateway, Art Science Information Gateway and Sociology Gateway (ALTIS) etc. There are four areas: Science and technology, Arts and Human

Science, Social Sciences, Health and Life Science etc. Free high qualified network resource services are provided to instructors, researchers and students. In July 2006, RDN was renamed as INTUTE.

INTUTE organization includes an executive committee, which responds for INTUTE policies, direction of development and management. There are four topical groups (Science and technology, Human science and Art, Social Sciences, Health and Life Science). Different universities are responsible for database setup and maintenance of each topic. Moreover, Bristol University is responsible for training service and virtual training lectures of INTUTE itself.

The funding of INTUTE is provided by JISC and AHRC. There are 123,765 data in the database.

## 2. Comparison and analysis of international building of subject information gateway

So far, research on evaluation of network resource subject navigator indicates that there are following standards to evaluate the building of subject information gateway: coverage, resource types, organization system, resource description, retrieval function, value added services, renewal and maintenance, cooperation model and amount of data. [6]Based on these nine standards, we sort out the evaluation standards of some representative network resource subject information gateway (see diagram 2.1). We compare CALIS navigator database, INFOMINE, INTUTE, CSDL subject gateway, Firstlight knowledge management system based on the nine aspects above.

diagram 2.1：Evaluation standards of representative subject information gateway

|  | INFOMINE | INTUTE | CALIS subject navigator database | CSDL subject information gateway | Firstlight knowledge operation system |
|---|---|---|---|---|---|
| Coverage | Subject: biology architecture medical science, commercial and electronics, culture and religion, social and human science, physics math and engineering etc. 9 areas Language：English | Subject：Art and Letter Science, Health and Life Science, Science and Technology, Social Science etc. 4 Areas Language：English | Subject: 11 subjects except for military science, 79 first-class subject; Language: Chinese, English | Subject: Physics and Mathematics, Life Science, Book Information, Chemistry, Resource and Environment etc. 5 areas. Language:: Chinese, English | Subject: Classic Theories, Philosophy Religion, Social Science, Politics, Law, Economics, Culture and Education, Language and Literature, Art, History etc. 29 areas Language: Chinese, English |
| Resource types | Data base, e-journal, e-book 、 bulletin, mail list, online catalog, articles, research catalog, subject introduction, textbook etc | Journal article, abstract, conference dissertation, e-book, patent, software, policy, FAQ, government publication | reference resources, full text resources, Industry resources, Multimedia resources, Interacting resources, Yellow | Newest resources, database, software, updating relevance, organization information, hot topic resources, books information, research | Knowledge news, international developments, figures, subject navigator, graduate admission resources, conference center, academy |

| | | | | | |
|---|---|---|---|---|---|
| | (including paid resource) | (including paid resource) | pages, events etc. 7 first-class categories; 29 second-class categories. | topics, subject topics(including paid resource) | instruction, knowledge database etc. |
| Organization system | Use LCC classification system and LCSH vocabularies. | Use UDC classification system and HASSET vocabularies | Organization based on the' confer doctor degree and master degree and the subject catalogue of training post graduate' issued by Ministry of Education in 1997 no vocabulary list | Each subject develops its own unique classification system based on international standards. For example, subject of math and physics, physics applies "International Categorization Method of Physics " no vocabulary list | The bottom classification based on the National Standard 'subject classification and code rule'; user interface based on the' confer doctor degree and master degree and the subject catalogue of training post graduate' issued by Ministry of Education of PRC in 1997 no vocabulary list |
| Resource description | Use self defined metadata, including 10 data. Some are Indexed by professional librarians. Some are automatically Indexed by machines. | Use self defined metadata, including 10 data. Indexed by professional librarians manually. | Use self defined CALIS navigator database "Norms of network resource metadata"18 descriptive elements. Indexed by librarians | Each subject define amount of elements based on its own uniqueness. For example subject of resource has 28 elements. Chemistry has 11 elements. Indexed by librarians | Use self defined metadata based on DC metadata. Includes 16 elements. Indexed by company full-time staff |
| Retrieval function | Provide basic search (including full article search), advanced search, and multi-way review function | Provide basic search, advanced search and topic review function. Cross topic search available | Provide basic search (full text search available), advanced search, category search and subject catalog tree review. Could save search method and history. | Provide key word search, advanced search, and review based on subject system and resource types. | Support full text search and subject review under all subjects. |
| Value added services | Contain recommended resource, INFOMINE blog, recommendation, IINFOMINE quick review, links of other search tools | Contain INTUTE blog, my INTUT, conference and event, hot topics, recommended website, VTS、SPP etc. | Contain resource recommendation, use help, FAQ 、 forum, online questionnaires, personal service (my space, search method and history. | Contain newest resource, information subscription, resource recommendation, subject forum, information recommendation, academy communication, etc. | Contain knowledge news, international development, Firstlight star, Firstlight focus, Firstlight forum, Firstlight blog, original ecological archives, etc. |
| renewal | Contain automatic | Data date renewal, | Weekly data date | Subject Weekly and | Daily data renewal. |

| and maintenance | renewal and expert renewal, dead link check software | dead link check software | renewal, dead link check software | Monthly renewal | Resolve dead links by original ecology. |
|---|---|---|---|---|---|
| Cooperation model | University libraries share models and cooperate. Funded by foundation and universities. | Universities and organizations share models and cooperate. Funded by foundation | University libraries share models and cooperate, funded by government. | Each subject build-up independently, CSDL cooperate and instruct, funded by government. | Company's business operations and concentrate production |
| Amount of data | One hundred thousand data | One hundred and twenty thousand data | One hundred and sixty thousand data | Forty thousand data | Four hundred thousand data |

## 1）Coverage

Coverage is the main problem of subject information gateway. It relates to subject and languages scope. As for subject scope, except for CSDL subject gateway which contains to 5 subjects, others contain comprehensive subjects which almost including every subject. It is very convenient for users to search resources. As for languages, INFOMINE and INTUTE aboard only provide English. Subject navigator in China has both Chinese and English resources. For example, CALIS subject navigator requires English resources has to be more than 40%. In fact, the average English reaches 34%, some subject reaches 90%. [7] The policies on Firstlight resources building require abroad resources should be between 24%-30%. [8] So, domestic subject information gateways are more suitable for Chinese users to search for resources including Chinese resource. In another hand, the high percentage of English resource indicates that domestic subject information gateway including CALIS indeed contain updated international information resources.

## 2）Resource Types

INFOMINE、INTUTE、CSDL and CALIS navigator contain college education and research leveled resources. Firstlight system is knowledge relevant. It serves more people. Detailed resource types relates to authority, accuracy, practicability and popularity. Therefore, building up resource types should balance main themes and details. It can't be too detailed to be practicable. For example, CALIS Navigator database originally set up 7 first class classifications, 29 second class indexing classifications. Now, it is clear that those classifications are not easy to be practiced. Librarians' labeling and users' searching system look complex and hard to be learned. Moreover, INFOMINE、INTUTE、CSDL gateways contain both free and paid resource, CALIS navigate and Firstlight system contain only free resources.

## 3）Organization System

Subject information gateway relates to the resource organization system which includes classification method and theme list. Resource organization system could be used for different types of resource review, topic search, indexing classification, and other resources system search and review. The advantages of subject information gateway is not only its high qualified resource but also the convenience of classification system and topic access. [9] This is also the main difference between subject information gateway and

common search engine. Domestic subject information gateway pay attention to the scientific, authority, and popularity. At this point, it is better than abroad systems. However, it lacks of vocabulary list and is not good for resource index and relevance searching.

**4）Resource Description**

Metadata description is the basis of subject information gateway search and review. The amount of metadata directly relates to the resource characteristics and details of that particular subject. In contrast, the metadata of two subject information gateway abroad are less than 10, and its resource description is relatively simple. Meanwhile other three domestic subject information gateway are more detailed and deeper.

Except for Firstlight system which is labeled by fulltime staff, all labeling work on subject information gateway data is done by librarians. Some INFOMINE data is indexed by professional librarians manually, and some is indexed automatically by machines. CALIS Navigator data is also mainly indexed and classified by professional librarian. Its local system could by interrupted by staff in order to obtain automatic classification on some resources.

Webpage print screening or webpage saving is a special metadata. This could be used to review websites that contain particular resources and to save search time. It could also provide information while websites are not available. Currently, only Firstlight provides this service.

**5）Retrieval Function**

Retrieval function is an important standard of subject information gateways . Its easiness directly effects the frequency it is used. Subject information gateways currently contain huge resources. Therefore, many subjects provided basic search, advanced search, as well as topic review function. As for this, domestic and abroad subject information gateway functions do not differ too much.

The show of CALIS and other Navigator search results is flexible. Users could choose the way of showing based on their own needs.

**6）Value Added Service**

Subject information gateway does not only provide professionals particular subject resources, it also provides value added service to support users to obtain resources. The topical value added service includes: web mirror image, forum, training materials、new resource introduction, recommend resource and message delivery service, personalized webpage customization, personal resource management etc. Observed from the comparison, both domestic and aboard subject information gateway provide many types of value added service, but instead of practice value added service, domestic subject navigator still remains academic and conceptive. For example, the forum set in CALIS navigator is blank, and it's online questionnaire and FAQ is not for users to learn how to use subject navigator. As for service provided to signed users, it is either not practicable or couldn't be delivery. [10] It needs to be improved. Moreover, INTUTE subject navigator provides a virtual training system （VTS） which is a unique and successful service project. Its online instruction is very popular to improve the user's information quality and IT technique which based on subjects/topics. Firstlight system built up the original

ecological archives, which fundamentally resolves the most serious problem of unstable resource coverage.

**7）Renewal and Maintenance**

Due to variance of network environment and instability of network resources, when a subject information gateway is built up, it is become important to renew and to maintain. The renewal and maintenance of navigator resources include three contents, the first is to record. Make sure records are accurate. The second is to check links regularly. Delete wrong and dead links. The third is to renew resource consistently. Renewal and maintenance is very important work for the abroad subject information gateway. Wrong and dead records are very few, and new resources have been being updated. [11] While CALIS was originally built up, due to lack of renewal and maintenance, there were many dead links. For some subjects, dead links even reached 20-30%. It dramatically decreased the degree of users' usage and recognition. Firstlight system built up the original ecological archive, which is an important method to resolve the problem of dead links.

**8）Cooperation Model**

The cooperation model of the abroad subject information gateway is cooperation between universities and libraries. They build up shared models and funded by foundation organizations. Due to shared needs, members are motivated to participate. While enjoying the service, they are responsible to provide other members information and service. This cooperation model is good for long-term stable development. Domestic CALIS navigator data and CSDL subject gateways are shared models funded by government. Although members have needs, subject information gateway can't stably develop if government funding couldn't consist. Building up shared models does not only have the advantage to avoid repeat work, which causes waste of resource, but it also combines advantages of every member to develop more valuable network resource navigator. [12] The disadvantage is the hardness and inefficient of organization due to huge members and while librarians also have to work for data. Firstlight system is a business system. Its advantage is to produce efficiently, but the premise of its long-term development is to be appreciated by users.

**9）Amount of Data**

Any subject information gateway has to hold certain amount of data. Except for CSDL gateway, other four comprehensive subject navigator's data are more than one hundred thousand. They all have enough data to serve users, but it does not mean the more the better. INFOMINE subject navigator has been containing more than one hundred thousand data. From July 2006 to July 2008, INTUTE subject navigator only increased ten thousand data. Currently, CALIS navigator has reached one hundred and sixty thousand data, but it obviously contains some low unqualified resources which negatively effect users' usage. Firstlight system plans to keep increasing ninety thousand data every year from 2008. So far it has reached four hundred thousand data. Based on the initial intention of subject navigator, coverage resources should be seriously selected and labeled by librarians. Therefore, we should insure the selection and coverage of core resource in order to indicate the uniqueness of subject navigator instead of only paying attention to the amount of data. To avoid shortness of data, INFOMINE subject navigator

provides user links of common search engine on its website alternatively.

## 3. Conclusions

According to above analysis, we recommend the following works in order to improve the building of various subject information gateways:

**1）Strengthen management and organization and improve project execution.**

Decrease indirect management and divide the whole subject navigator into small categorizations based on those main subjects, such as: literature, mathematic science, technology, medical science, architecture, management etc. Each main subject should be served by a center library for every day management, coordination, and regular inspection. Under each center library, there are cooperated libraries to improve the management efficiency.

**2）Improve horizontal cooperation and expand shared resources**

Cooperate with international organizations to build up a model of English resource navigator and data resource sharing. Cooperate with domestic organizations to directly purchase or utilize partial metadata resource in order to resolve the problems of overwhelming work and slow update by librarians.

**3）Combine resource types, fully save core academic resource and provide relevant service.**

Recombine currently resource types in order to facilitate librarians' works and users' review. Increase the amount of academic guidance categories for news, newest subject websites, and academic trend etc. And pay more attention to open access and the access for academic resources such as blog etc. And fully save core academic resource and provide relevant service.

**4）Update system platforms and add long-distance catalog function.**

System software should update systems or functions based on new needs and experiments during the Tenth Five- Year Plan period. Improve online resource automatic tracking, automatic classification, indexing techniques and more human-machine interacting methods in order to improve librarians' efficiency. Increase long-distance catalog function based on web in order to facilitate cooperated libraries to catalog online. No longer configure local systems.

**5）Increase descriptor list to improve efficiency of system search**

Providing descriptor list or standardizing theme word is good for indexing resource, it is benefit users diminishing search range, quick searching for related resource. We need to take the advantage of subject navigator to improve search efficiency.

**6）Tracking users' needs, provide value added service mainly characterized by personalized services.**

According to users' personal needs of information, accurately insure their needs by tracking registered users' information. Provide personalized service such as my space, search strategies, search history, customization of information etc.

**7）Strengthen propaganda，promotion and training to increase website Click-through rate.**

Propaganda, promotion and training are important ways to improve the utilization rate of subject information gateway. We need to adopt various ways to solve the problem that users do not know about the gateways. And insure more universities and research organizations link navigator websites in any possible ways. Pay attention to training in order to solve users' problems while they are using those websites.

**Reference**:

[1] CALIS Key Subject Network Resource Navigate Gateways. [2009-6-5]. http://navigation.calis.edu.cn

[2] Chinese Academy of Sciences, National Science Digital Library Subject Resource Gateways. [2009-6-5].http://159.226.100.135/ejournal/CSDL－SubjInfo.php

[3] Firstlight Knowledge Management System. [2009-6-5]. http://www.firstlight.cn

[4] INFOMINE. [2009-6-5]. http://infomine.ucr.edu

[5] INTUTE. [2009-6-5]. http://www.intute.ac.uk

[6] Li Lili．Comparative Study of International Architecture Information Gateway [Master dissertation]．Beijing：Chinese Academy of Agricultural Science，2007

[7] CALIS Navigate Database Project Management Group．Mid-term Inspection Circular about CALIS "Key Subject Network Resource Navigate Database" [2009-6-5]. http://navigation.calis.edu.cn/cm/allnotice.jsp

[8] Hu Zhiqiang. Landmark product of Firstlight－Subject Navigate 3.0. Subject Navigate 3.0，Beijing：Beijing leisun Technology Inc，2006：6-12

[9] Zhang Rui. Research on Built-Up of Subject Information Gateway [Master dissertation].Chengdu: Sichuan University，2006

[10] Liu Xiuhua. CALIS Research and Analysis on Built-Up and Utilization of Key Subject Information Gateway. *Library and Information Works Online*，2008(3)：1-6

[11] Wei Ruimin. Evaluation and Analysis on CSDL Subject Information Gateway in China. *Information Exploration*，2007(1)：116-118

[12] Zhang Chunhong etc. Current Situation and Development on Network Academic Resource Navigate. *Library Development*，2003(04)：17-19

Contact Information：
Name：Zhang Xiya
Email：xyz@mail.xjtu.edu.cn