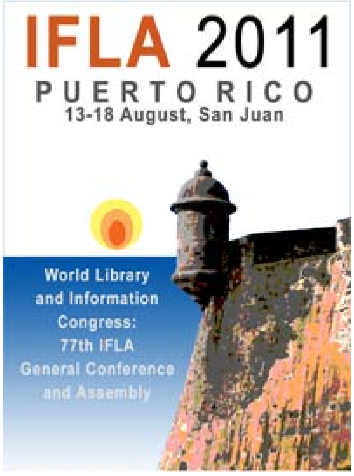

A system model for art information retrieval

Zhengwu Guo[1]
Ninde Normal University library
Ningde, Fujian, China

Meeting: **79 — *Tackling the challenges of multilingualism in the arts: catalogues, databases, digital collections and other material in the global context* — Art Libraries Section**

Abstract:

How to transcend the barriers between natural language and retrieval language in art information retrieval? By up to 6 and a half years of project development and studies, the author found some effective Countermeasures. The core contents are as follows:

1. Gather on line a tremendous amount of information retrieval resources of different countries; build up federated search and unify retrieval searching entry; and keep application functions and habits of different databases. "One page" integration of multi-language and different kinds of multiple cluster resources is more convenient for users than that "one-stop".

2. Construct a new intelligent navigation system by semantic retrieval and visualization methods to transcend the retrieval languages barriers, solve the accessibility conundrum of information retrieval and popularize high level information retrieval at speed.

3. Open information access and let users build their own cluster resources to meet further multivariate needs.

Key Words: art information; integrated retrieval; intelligent navigation

[1] Zhengwu Guo
Address:No.98-1 Jiaocheng South Road,Jiaocheng District,Ningde City,Fujian Province, 352100 China
Email:dsdhdh@gmail.com Phone:86 593 2912573 86 13559900393
Associate Professor of Ninde Normal University library, Deputy Curator. Having been engaged in information management research for 23 years, study results were awarded the provincial government prize and First Prize of China Library Association.

Art is very difficult to define. Many people believe that it is the practice of activities which create a symbol of human affection and entertainment .In the social dissemination, it has the role to convey values and emotional images. In fact, Art has become an important factor which affects the development process of social life. It does not just belong to the art of actors, but the whole human society.

Art cannot do without information. With the development of art practical activities, countries and regions have had a lot of art information. Much of the art information has been digitized into searchable information resources. How can we retrieve art information from a wealth of information resources? How the art information retrieval can stride across the barrier of many kinds of natural languages and retrieval languages. These are issues of great concern about the use of art information resource globalization. With this in mind, after six and a half years of project research , the author found some effective coping methods and established an integrated model of information retrieval which can apply to information retrieval for arts.

1 "One page" management services on art information resources

When we talk about "one page" management services on information resources, it first bases on the idea of federated search. The essence of federated search is to achieve union search for heterogeneous data sources. That is regarding multiple distributed heterogeneous data sources as objects to provide users with a unified search interface and transforming retrieval demand of users into a search expression of different data sources. At the same time, multiple distributed heterogeneous data sources through local network and wide area network can be retrieved and search results can be integrated. After deduplication and sorting, we will present the results in a uniform format to users.

Being different from a number of federated search, regardless of quantity of information resource database, "one page" management services try their best to visualize all their resources in a single page; regardless of the many differences between the retrieval results of information resources which have been distributed on the network, all of them can be presented in a single page; In the question-type translation process, it is practicable to try to overcome language barriers and translate search terms directly. Then match them to the relevant resources, so that they can be retrieved automatically; in the presentation of search results, it is practicable to try to overcome language barriers to provide an automatic machine to translate them. In short, it is wanted that all the functions can be achieved in a page. That will be more convenient than that "one stop". Figure1 indicate the basic logic.

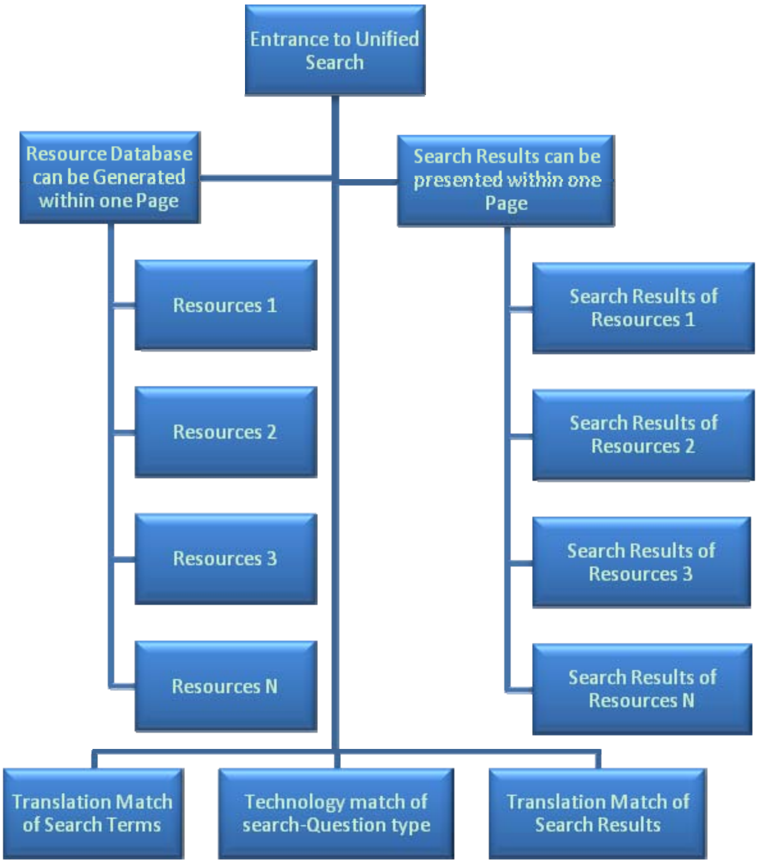

Figure 1: Basic Logic Structure of “One Page” Management Services

1.1 Comprehensive selection of art information resources

Art information is divided into two categories: one is the art itself, the information belongs to the content and forms of artwork, such as the social life which the artwork reproduces, emotional which the writer performs in the arts and the elements and constitution which perform in the arts. In addition, it also includes the information which is the appreciation or criticism to artwork and the comment or review etc. which have described above. The other one is the information which relates to art behaviors and art exchanges. It belongs to art culture and art exchanges, such as artistic creation information (author, time, background etc.), disseminating information of the artwork (art auction, art prices, film Festival, etc.), art theory and research papers and so on.[1] Therefore, the scope of art information is very wide and the types of art information resources are very complex. Especially, with the further development of information technology, art information not only has structured repositories but also the unstructured resources. At the same time, art information resources also present the trend of distributed and rapid development. Therefore, the use of integrated retrieval for art information is effective. All of the following regards federated search or integrated search as a precondition.

In 1998, WebFeat decided to “change the way people search for”. His idea is simple-------“to allow libraries to query any part or all of their own database simultaneously "[2]. To use the basic concept of federated search and to search art information better, selecting the art of information resources is as the first.

Many federal retrieval systems have deficiencies in the realm of selecting and accommodating/containing art information resources. The first one is inadequate internet information resources. Early federal retrieval focused on building a unified retrieval portal to multiple local resources. And the searching of information resource database on the internet cannot be integrated or less integrated, which does not match the rich art information retrieval resources. Secondly, federal retrieval system can automatically identify the data source and judge whether it is useful. However, this recognition is not transparent to user. If the whole base of a data source is undetected, users really don't know. A given question term is used to judge which search engine or database retrieval technique is the best option for a given search term, and which database defeats its own retrieval purpose. This technology is much more difficult to realize than content selection. Therefore, there is no need to judge automatically and entirely by the system. Users also need some opportunities and navigations to judge independently. Thirdly, for the tolerance of retrieval time, retrieval systems often do not extract all search results of various resources, possibly resulting in omission of valuable information. Fourth, some users have acquired the habit of using the specific database. Actually they need to keep the usage of different databases. To take into account these realities, integrated retrieval should maintain single retrieval of varieties of art information resource database or channel.

If users are given more opportunities for self-search, to maintain and list significant resource database and channel. Usually, the search interface is not simple and users' choice to use is very complicated, thus it cannot achieve the purpose of federal search which is simple and practical. To overcome this contradiction, we have designed a variety of solutions.

First, navigate the new classification of resources, so that all resources can be focused on one page simple for fast navigation. We cannot use the traditional classification navigation of static tree structure. If static tree structure is more than class, it will be very inconvenient for use. So the author developed a dynamic tree structure. It is very suitable for layers of the column class and fast guidelines. No matter how many art information resource database or channel can be fully accommodated. (Figure 2 is a case which it can accommodate more than 100 search resources.) Secondly, develop and design the function which bases on search terms to recommend data source so that the function which federal retrieval system can automatically identify and judge the data source is clear to users. Thirdly, consider giving users the chance to add or delete and so on to customize personalization options of personal data sources.

Through this approach, art information retrieval users can easily retrieve a very large number of artistic resources.

Figure 2

1.2 "Resource that is retrieval" is a unified entrance method to retrieve

From the perspective of users' retrieval demand, they not only have comprehensive and unified search needs but only have separate needs. What users enter into the database separately to search resources is the most basic retrieval needs and long-term usage habits of users. Basing on the habits to improve distributed cross-database search way, we studied and developed a unified search way which is "Resource that is retrieval". When users fill out search terms, no matter what search resources they click on, search results can be obtained directly. If user wants to retrieve the relative information such as the Louvre, Buckingham Palace, Sydney Contemporary Art Museum in Australia, the Bonn Federal Art Gallery in Germany and so on, they can fill in the search terms and switch directly to retrieve. For another example, if users enter BaishiQI in the search box and click on "Interactive Encyclopedia" channel, they will obtain the search results immediately; if users click on "Youdao Dictionary", they will get search results which are translated by "Youdao Dictionary"; if users click on "Art Network", they will straightway get relative search results.

This approach not only overcomes tedious work to land the multiple sites one by one and to fill in the search terms one by one. It solves diversity problem of the search entrance of digital resources. What's more, it maintains best the different use function, search speed and use habit to different databases. It integrates and unifies navigation and search function organically and enhances and meets the basic needs and habits of users. This approach with the federal search solves the problem of unified search of various resources as well as the separate one in a high speed.

By the way of unified entrance method which is "Resource that is retrieval", art information search will become very convenient.

1.3 The translation of question-type search

Different information management systems have different methods of information technology processing. When federal search engine submits search terms to independent search engines, it needs to process them grammatically and semantically, and then translate them into those that fit for different search techniques and processing formats of different search engines, and finally push the processed search terms to the appropriate search database. The key of question-type search translation is to minimize the loss of semantic information of users' search terms. The translation should ensure that users can get an optimal search result in various databases and achieve the coordination and optimization of the recall ratio and the precision ratio. There are usually two methods in question-type search translation: the first one is background process of federal retrieval system. Through retrieving federally metadata in index database, the data mapping will be implemented between the source databases. The machine will automatically translate users' search terms into those that are suitable for different search interfaces and search techniques of databases. The other one is the design of interactive search interface. When users are searching, relevant interactive terms and instructions will appear in search box, providing users with options or retrieval guide and help. Many former federal systems have made a lot of outstanding work.

In question-type search translation, the author mainly added a new development. By system automatically judging, the system translates the search term into those that adapt to all sorts of database language so that users can overcome the language barrier of search terms. This development quickly resolves the retrieval problem among databases of different languages, and offers the convenience of cross-linguistic search. For art information retrieval, when users attempt to search information in multinational institutions or databases, the efficiency will be greatly enhanced.

1.4 "One- page" show of search results

The feedback of federal retrieval system in the search results has made many important achievements. For example, Federal retrieval system owns basic metadata search capabilities. It also has the capacities of deduplication and sorting， personalized customization function and flexible clustering methods.[3]

The aim of my development research is to make the research results show in "one- page" style. Because of different databases, the technology and pattern are very different, we strive to achieve the aim that research results of all resources can be basicly maintained original sample in one page. It has practical significance that users can view search results while they do not have to jump to a new page. At the same time, the author let "one page" research results translated by multilingual machine and try his best to minimize the language barrier of users as well as he can.

One-page search is based on the development of one-stop search. One-stop search is more convenient than that each information resource logins and search one by one. But when many one-stop searches choose different databases, it causes the problems such as switching search problem and too many result pages. Users are easy to get lost. Therefore, what users fill out the search terms from the beginning and implement multiple search, then search results are presented in one page is more convenient and intuitive than that one stop. Because what users see is what they obtain. Search will become more concise.

2. Construction of new intelligent knowledge-information navigation

Knowledge and information institutions use body tissue such as classification, thematic approach and document searches. But users cannot express document needs and access to document. The contradiction exists for a long time from manual search time to network retrieval times. Though knowledge and information institutions adopt many measures such as users' education, organizing classification, subject index, making classification, topic navigation and so on to help users solve the contradiction, the problem is still serious. To take online bibliographic retrieval for an example, users rarely retrieve by means of classification and subject. So Caiying Zhang etc. think the retrieval function of classification is on the decline, it cannot mention in the same breath with its document organization because it is irreplaceable.[4] In a network environment, it is popular to search by keywords currently. The search form basing on keyword matching cannot satisfy users' needs which they need the information content instead of information form. So it is upset to information retrieval and knowledge navigation.

The advantages of classification search and subject concept search cannot be ignored, so it is still a major issue of library science, information science and information retrieval to solve the difficult problem to make classification and subject retrieval use easily. For this reason, the author tries his best to do some meaningful thing. Users would not master classification, thematic approach while they can carry out question-type search directly to make knowledge navigation service more easily achieved by public. At present, the try has had substantial progress. It can serve as a kind of knowledge map software to use. The paper will take classification of usability for an example to talk about the retrieval of art Information

2.1 Embedded Intelligent Navigation, a Clean Interface

Here take an OPAC page of China's National Library (CLC) as a prototypal example, shown as figure 3.

Figure 3

2.2 Real-time navigation, without requirement for users' grasp of classification

In the information retrieval system, class numbers are usually a searchable field in document information databases. Users must input class numbers during question-type classification search. As the class number is a kind of mark symbol composed of letters and figures, users can not use it as accurate as using natural language to express their document requirements. And users can hardly express some CLC class numbers, especially for those that are very systematic, structured with deep category and complicated symbolizing system, which results in performing practically no function and being used by few people.

We cannot expect users to master the abstruse classification. Since users do not understand classification, how can the system easily achieve classification retrieval questioning? To overcome the problem, the author designed a series of implementation methods.

2.2.1As soon as mouse clicks classification search box, 22 basic categories of CLC are immediately presented.

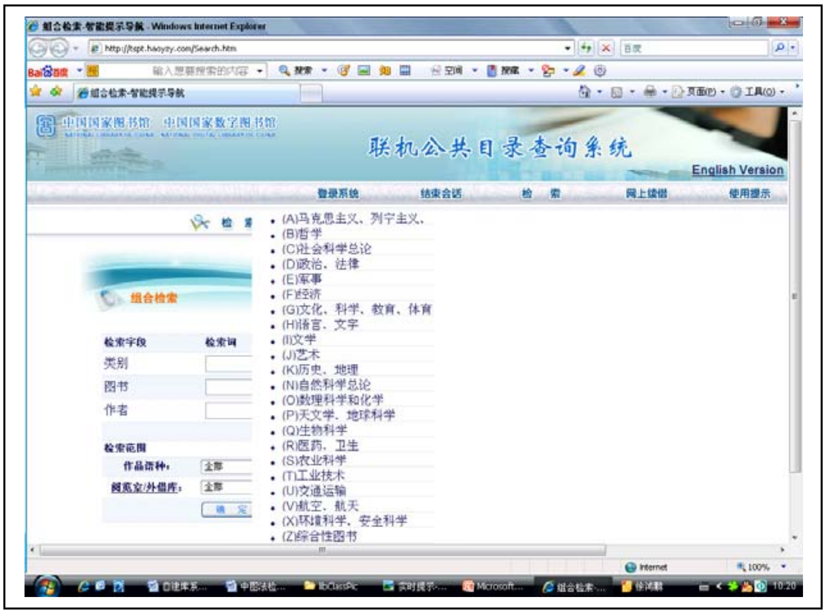

Figure 4: Real-time tips when mouse clicking the classification search box.

Description: the implication of a single class number is abstract. Whether users know a little about the class number or not, they are generally difficult to accept under the condition of real-time displaying only class number. Thus it is necessary to display class numbers and Chinese names of category simultaneously and to make users understood fully at one glance.

2.2.2 Classification navigation's real-time dynamic. As long as users move the mouse, guidance at all levels is implemented for users to browse by category until the bottom, as shown in Figure 5.

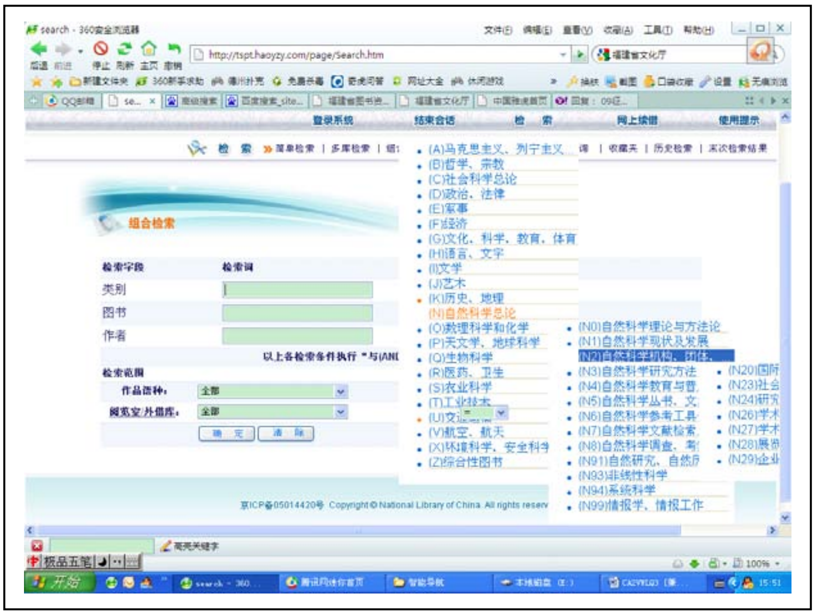

Figure 5

When users need to search the art information, they can move the mouse to point the "J Art". And it will be able to achieve the art guide at all levels and browse by category.

2.2.3 When the mouse clicks and selects the category, class numbers will directly fill in the search box, as shown in Figure 6.

Figure 6

Description: class numbers and class names need to show simultaneously. When users select a category, search box can only express the chosen class numbers but no Chinese names of category for the convenience of search. Certainly the tips of category names can also take into account all kinds of annotations reflecting the explanatory comment and the category reference, etc. of the upper category names for ease of reference. However, if the annotation is specified to much in detail, readers usually don't like. Therefore, it calls for much deliberation.

2.2.4 Multiple navigation functions: when the mouse clicks the chosen class numbers, the system can do secondary navigation, as shown in Figure 7.

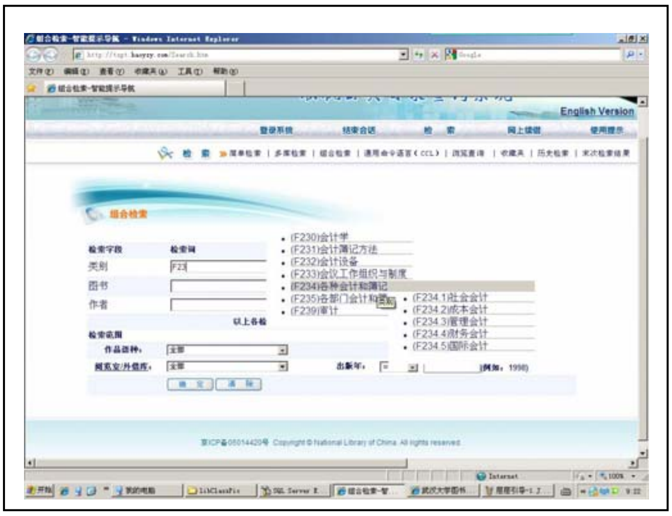

Figure 7

2.2.5 Readers don't need to know class numbers and they can do class numbers retrieval through the semanteme of key words (subject terms), as shown in Figure 8.

Figure 8

2.2.6 There is no need to return to the basic categories. Navigation can begin with any node, as shown in Figure 9.

Figure 9

2.2.7 To overcome multilingual obstacles, navigation retrieval language will be automatically translated.

By automatically translating retrieval language, people of different languages can use multilingual ontology navigation. For example, English speakers can also use our CLC.

2.3 Making classification symbols full compatible with practical class numbers and citing practical class numbers in smart navigation

Chinese Library Classification (CLC) is composed of a huge Classification System and lots of Classes. It has 22 classes and each class covers many multilayer subclasses. According to the common auxiliary principle, there are more than 50000 basic classes and the total class approaches 400000 in the 5th Printing Edition of CLC.[5] Can you accept the response speed which takes such many categories as intelligent guide? Combining with multilayer navigation experiments of CLC and the monolayer experiments which studied on about 50000 normal drug intelligence guiding words, the author reached a conclusion that at present the response speed of navigation algorithms works well and if we standardize more guiding words is feasible. Therefore usability design of category navigation is theoretically suitable for almost class number of each classification. Take CLC for an example, after a basic category has been subdivided into six or seven layers, using the alphabetical guide instead of the mouse guide, the user will find the efficiency of navigation is equivalent to entering a seven-letter English word or seven Chinese phonetic alphabets(about two Chinese characters spelling).

When users enter the class number of the bottom layer, which most of basic categories have been subdivided, the response speed of navigation is quite easy to accept.

But the application of classifications encountered in theoretical and realistic aspects. Most of literature class numbers are set according to the classification rules in the practice. They are more detailed than basic categories. For the ordinary users' need, it is better to cite practical class numbers in smart navigation. Except some users (mainly including librarians), who indicate that they prefer classification symbols, the navigation is going to cite classification symbols.

There is a great disparity in quantity among different subjects because of the scales of document resources in different systems and different subjects. Such as, some classifications have abundant resources, others have a few resources, and still others have none. In hence, the quantity of intelligent guide words has to mesh with the words' actual quantity and controls their category layers. In practice, the navigation works efficiently. Basing on the 22 classes and eight division method, the author calculated that the quantity of intelligent guide words should be limited within 720000.The author also thinks that almost category layers usually come to bottom within seven subdivisions, and irrelevantly controls their layers.

2.4 Apparent advantages in the integration of a variety of data analysis and the intelligent adjustment to guide the content

Intelligent navigation can conduct real-time intelligent learning and adjusting guidance according to Category-hierarchy relation, the state of the number of document and frequency of use, etc. It greatly facilitates the retrieval needs of users and it is much better than classification navigation of "browsing function". Basic comparison is as follows:

Function / Navigate mode	Browsing navigation of classification	Question-type navigation of classification
Realizing state	Having been realized for many years	Realizing the breakthrough first
Hierarchy guidance, category browse	Yes, less category hierarchy in general	Yes, covering any bottom of category
Matching condition of category and practical document	Complete matching	Incomplete matching
Guiding any node at all levels	Difficult	Easy
Cross-level navigation function	Difficult	Yes
Real-time guiding function	Static	Dynamic
Classnames and classnumbers simultaneously display the function	Part retention	Full retention
Classnumber input	No	Yes
Transformation classnumbers after inputting key words	No	Yes
Translate automatically	No	Yes
Function of retrieval memory	General non-retention	Yes
Retrieval speed and efficiency	Slow, low	Fast, high
Friendly users	General	Good

2.5 Hoping a variety of co-operation for huge project

To take achieving intelligent navigation for an example, the main difficulties are as followings:

2.5.1 The basic Category of CLC 5ed. is more than 50,000. It will be more than 400,000 after the common auxiliary. To organize the category, the project is very huge and possesses high specialization.

2.5.2 To convert keywords into classification, the current system has already compiled vocabulary more than 600,000. The amount of data is enormous.

2.5.3 TO navigate the actual books in various libraries, we need to deal with large amounts of data; To convert keywords into class numbers and to convert practical book classification into taxonomic name classification, the workload is even more enormous.

2.5.4 To achieve real-time intelligent navigation of classification, the amount of calculation is large and function algorithm must be very advanced.

2.5.5 To embed in the different document information management system, it is needed to have the different document information management system designed different procedures.

Therefore, in the international, to achieve the knowledge intelligent navigation which is based on various ontology and languages, we need to seek a great deal of international cooperation. Only in this way can we absorb the different countries the latest achievements such as classification, thematic approach, keyword search technology and search navigation technology. What’s more, we can update and apply it according to variation of hypostatic document in these information institutions such as libraries, museums and so on. Only so can each information institution save large amount of time and energy.

My development in the intelligent navigation is mainly as follows: 1. Facilitating semantic retrieval navigation; 2. The infinite visual navigation retrieval. 3. The mouse guiding the navigation by layers. 4. Any retrieval approach navigation. 5. Specialized precision navigation. 6. Real-time dynamic navigation. Through the author has made some breakthrough in the intelligent navigation, the amount of work that the author do is limited. The author hopes more co-operation very much, including scientific research project or marketization co-operation.

3. Conclusion

In addition to the above development, the author also did some work in integrated retrieval of information, including opening information access, and trying to meet the diverse needs, allowing users to easily build their own cluster resources, intelligent navigation and one-page retrieval. These developments are now under consideration. In the view of the developed platform, they can bring lots of convenience to art information retrieval or other integrated retrieval. Characteristics have emerged in the respect of transcending various natural language barriers and retrieval language barriers, knowledge intelligent navigation and “one page” management services on information resources. The author hopes to get your favor, help and deepened cooperation.

References:

1.金胜勇.艺术信息资源共享研究[D].河北：河北大学，2004：5-6

2.马骅.国外主要联邦检索系统的兴起、现状及发展趋势[J].图书馆建设,2009(3) :1

3.陈家翠.联邦检索机制及其存在的问题[J].图书情报工作, 2006(6):89

4.张英彩,董素音,蔡丽静.书目分类导航系统的快速实现[J].图书情报工作,2006(2):84-86

5.Ideapub.net 2005 智能信息集成系统，http://www.soft6.com/business/3/33350.html