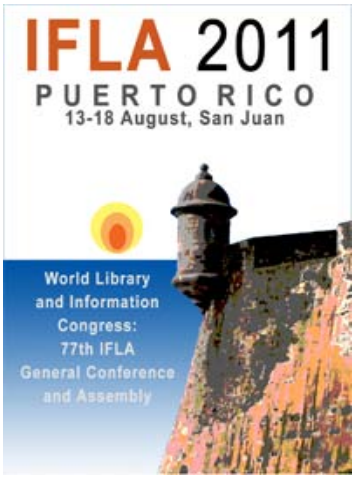

一种集成的艺术信息检索系统模型

(A system model for art information retrieval)

郭正武[1]

(Zhengwu Guo[2])

宁德师范学院图书馆 中国 福建 宁德 352100

(Ninde Normal University library, Ningde Fujian 352100 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**

摘 要：

艺术信息检索如何跨越多种自然语言和检索语言的障碍呢？作者通过长达 6 年半的项目开发研究，找到了一些有效的应对办法，这一模型设计的核心内容：1. 通过网络大量集聚不同国家的信息检索资源，对集群的资源进行联邦检索，统一检索入口，并且保持不同数据库的使用功能和使用习惯，对不同文种、不同类型的多种集群资源进行“一页式”管理服务，比“一站式”更加方便用户检索利用。2. 通过语义检索、可视化等方法，构建新型智能导航系统，跨越多种检索语言障碍，破解信息检索的易用性难题，快速普及高水平的信息检索。3.开放信息存取，并让用户方便构建自己的集群资源，进一步满足多元需求。

关键词 艺术信息 集成检索 智能导航

[1]郭正武，宁德师范学院图书馆副研究馆员，副馆长。从事信息管理研究 23 年，研究成果曾获省政府奖和中国图书馆学会论文一等奖。

通讯地址：中国福建省宁德市蕉城区蕉城南路 98-1 号，邮编：352100

Email:dsdhdh@gmail.com 电话：86 593 2912573， 手机：86 13559900393

[2] 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.

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, I found some effective Countermeasures. The core contents :

1.Gather on line a tremendous mount 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. "a page" integration of multi-language and different kinds of multiple cluster resources is more convenient for users than that "one-stop " .

2.Constract 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.

艺术非常难以定义。许多人认为它是人类创造情意符号和娱乐的实践活动，在社会传播中，它有传达价值和感情意象的作用。事实上艺术已经成为影响社会生活发展进程的一种重要因素，它不单单属于艺术的行为主体，更属于整个人类社会。

艺术离不开信息。随着艺术实践活动的开展，每个国家和地域都产生了大量的艺术信息。这些信息有许多已经被数字化，变成可检索的信息资源。如何从大量的信息资源中检索艺术信息？艺术信息检索如何跨越多种自然语言和检索语言的障碍呢？这是一个艺术信息资源全球化利用非常关注的一个问题,也是任何信息资源全球化利用的重要问题。为此，本人通过长达 6 年半的项目开发研究，找到了一些有效的应对办法，建立了一种集成的信息检索模型，适用于艺术信息检索。

1 艺术信息资源的“一页式”管理服务

信息资源的“一页式”管理服务，首先基于联邦检索（Federated Search）的理念。联邦检索的本质是对异构数据源实现统一检索。即以多个分布式异构数据源为对象，向用户提供统一的检索接口，将用户的检索要求转化为不同数据源的检索表达式，并发地检索本地的和广域网上的多个分布式异构数据源，并对检索结果加以整合，在经过去重和排序等操作后，以统一的格式将结果呈现给用户。

与许多联邦检索不同的是，不论信息资源库数量多少，“一页式”管理服务试图把所有的资源都可视化集成在一个页面内；不论联网分布的每一个信息资源的检索结果有多大的不同，都试图把它集成在一个页面内呈现；在提问式转译过程中，试图克服语言障碍，直接翻译检索词，匹配到有关资源中自动检索；在检索结果的呈现中，也试图克服语言障碍，提供一个机器自动翻译。总之希望所有功能在一个页面中实现，这将比“一站式”更方便。其基本逻辑示意如图 1

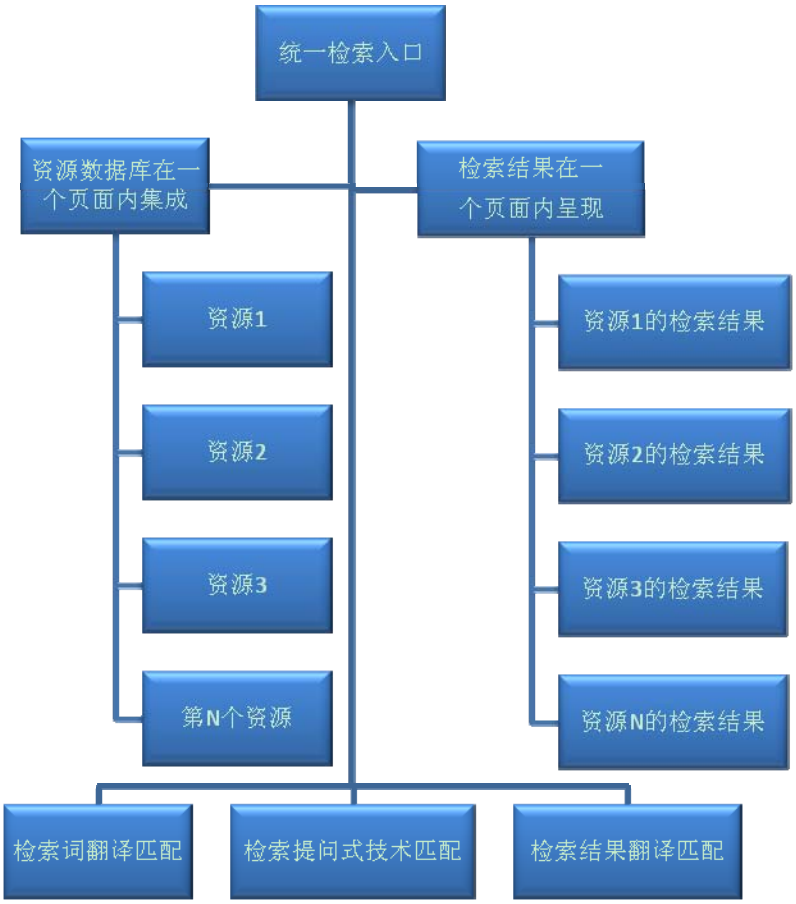

图 1 “一页式”管理服务的基本逻辑结构

1.1 艺术信息资源的全面选择

艺术信息主要分为两大类：一是艺术品本身的信息，属于艺术品的内容和形式方面的信息，如艺术品所再现的社会生活、作家寄于艺术品所表现的情感、表现形式的元素与构成、对艺术品进行欣赏或批评的信息以及关于上述各种信息的评论、综述等；再一类是与艺术行为或艺术交流相关的信息，属于艺术的文化和交流方面的信息，如艺术的创作信息(作者、时间、背景等)、艺术品的传播信息(艺术拍卖会、艺术品的价格、电影艺术节等)、艺术理论研究论文等。[i]因此艺术信息的范畴非常广，艺术信息资源的类型非常复杂。特别是随着信息化的深入发展，艺术信息既有结构化的资源库，也有无结构的资源，同时艺术信息资源还呈现分布式的快速发展趋势。因此对艺术信息使用集成检索是有效的，以下的所有探讨都基于联邦检索或集成检索为前提。

1998 年，WebFeat 决定“更改人们搜索的方式”，其思想非常简单——“允许图书馆同时查询其所拥有的任意部分或者全部数据库”[ii],运用联邦搜索的这一基础理念，为了更好地搜索艺术信息，首先要选择艺术信息资源。

在选择和容纳艺术信息资源方面，有许多联邦检索系统存在不足。（1）互联网信息源不足。早期的联邦检索主要针对多个本地资源构建统一的检索入口，对互联网上的信息资源库不能集成检索或较少集成检索，这与互联网上丰富的艺术信息检索资源不相匹配。（2）联邦检索系统自动识别判断数据源是有用或无用，这种识别对用户并不透明，如果出现某个数据

源全库漏检，用户并不知道。对一个给定的提问词，判断哪些搜索引擎或数据库检索技术对给定检索词是最佳选择，哪些数据库达不到应有的检索效果，这是比内容选择更难实现的技术。因此不必完全由系统自动判断，需要给用户一些自主判断的引机会和导航。（3）为了检索时间的耐受性，检索系统往往没有提取各种资源的所有检索结果，有可能把稀缺珍贵的信息遗漏。（4）有些用户对特定的数据库已经形成使用习惯，需要尽量保持不同数据库的使用习惯。为了兼顾这些现实，集成检索需要保持各种艺术信息资源数据库或频道的单个检索。

如果给用户更多自主检索的机会，需要保持和罗列大量资源数据库和频道，通常检索界面不简洁，而且用户选择使用非常繁琐，不能达到联邦检索简洁实用的目的。为了克服这一矛盾，我们设计了多种解决方法。

一、对资源进行新型分类导航，让所有资源在一个页面简洁集中、快捷导航。我们不能使用传统的静态的树形结构的分类导航方法，静态树形结构如果超过 3 级，使用就非常不方便。因此这里开发了一个动态的树形结构，非常适合层层列类、快捷指引。不论有多少艺术信息资源数据库或频道都能全面容纳。（图 2 就是容纳 100 多个检索资源的一个案例。）二、开发设计根据检索词推荐数据源的功能，让联邦检索系统自动识别判断数据源的功能对用户透明。三、考虑给予用户增加或删减等定制个人数据源的个性化选择。

通过这一方法，艺术信息检索用户能轻松拥有极其大量的艺术检索资源。

图 2

1.2“资源即检索”方式的统一检索入口方法

从用户检索需求的角度看，除了有全面统一检索的需求外，也有分别数据库的检索需要。分别进入数据库查找资料是用户最基础的检索需求和长期的使用习惯，基于这种习惯来改进分布式跨库检索方式，我们研究开拓了“资源即检索”的统一检索入口方式。当用户填写好检索词后，无论点击哪个检索资源，均可直接得到检索结果。假如用户想检索卢浮宫、白金汉宫、澳大利亚悉尼当代艺术馆、德国波恩联邦艺术展览馆的信息，填写检索词后即可直接切换检索。又如，在检索框输入齐白石，点击“互动百科”频道，立即得到检索结果；点击“有道词典”，立即得到有道词典翻译的检索结果；点击“艺术网”，立即得到相应的检索结果。

这一方式既克服多个站点逐一登录、逐一填写检索词的繁琐，解决数字资源检索入口的多样性问题，又最好地保持了不同数据库的使用功能、检索速度和使用习惯，有机地结合和统一了导航与检索的功能，提升满足的是用户的基础需求和基本习惯。使用这种方式加上联邦检索，既解决了多种资源统一检索的问题，也解决个别资源的快速统一检索问题。

通过统一检索入口的“资源即检索”方式，艺术信息检索变得非常便捷。

1.3 检索提问式转译

不同信息管理系统有不同的信息技术处理方式。联邦搜索引擎在将检索词提交给各独立搜索引擎时，需要对检索提问词进行语法和语义处理，将检索词转译成适合各种搜索引擎检索技术和处理格式的检索词，并将处理后的检索词推送到相应的检索数据库中。检索提问式转译的关键是尽可能减少用户检索词语义信息的丢失，保证用户在各数据库中的检索结果最优，实现查全率和查准率的协调优化。检索提问式的转译方式通常有两种:一种是联邦检索系统的后台处理，通过联邦检索索引数据库中的元数据，实现与源数据库之间的数据映射，由机器自动将用户检索词转译为适合各数据库检索界面和检索技术的检索词;另一种转译方式则是设计交互式的检索界面，在用户进行检索时，在检索框中出现相应的交互式的术语和使用说明的提示，供用户选择或为用户提供检索指导和帮助。过去的许多联邦系统已经做出了许多杰出的工作。

在检索提问式的转译方面，我主要是增加了一个新的开发。通过系统自动判断，将用户检索词翻译为适合各数据库语言的检索词进行检索，因此用户可以克服检索词的语言障碍。这一开发，快速解决了跨越不同语种数据库的检索，实现跨语言的检索便利。这对艺术信息检索而言，试图在多个国家的机构或者数据库中查询信息时，效率将大大提高。

1.4 检索结果的“一页式”呈现

联邦检索系统在检索结果的反馈方面已经取得许多重要成果。如联邦检索系统除了具备基本的元搜索能力外，对检索结果还具有强大的去重和排序能力、个性化定制功能和灵活的聚类方法。[iii]

我的开发研究主要在于让检索结果能实现“一页式”呈现。尽管不同的数据库，其技术和样式有多大的不同，我们努力实现了针对每个资源的检索结果都能基本保持原样地在一个页面中呈现。用户可以不必跳到新的页面查看检索结果，这有现实意义。同时，我还针对这种一页式的检索结果进行多种语言的机器翻译，尽量减少用户阅读的语言障碍。

一页式检索是在一站式检索的基础上开发的。一站式检索比起单个信息资源的逐一登录检索有其方便性，但是许多一站式检索在选择多种不同数据库进行检索时，要么检索切换麻烦要么结果页面太多，使用者容易迷航。因此用户从开始填写检索词，再到实施多次检索，再到检索结果都在一页呈现，它比一站式来得更方便直观，可视化效果更好。因为所见即所得,所得还都在一个页面呈现，检索变得更简洁明了。

2.构建新型智能知识信息导航

知识信息机构利用分类法、主题法等本体组织与检索文献，但是用户却难以根据分类或主题准确表达文献需求并获取文献。这一矛盾从手工检索时代到网络检索时代长期存在。尽管知识信息机构通过用户教育、组织分类与主题索引、制作分类与主题导航等方式来帮助用户解决这一矛盾，但是问题依然严峻。以联机书目检索为例, 用户现在很少利用分类与主题检索，以致于张英彩等人认为：分类法的检索功能日趋弱化, 已不能与其在文献组织方面独特的不可替代性相提并论[iv]。在网络环境下, 目前备受用户青睐的是使用关键词检索。基于关键词匹配的检索形式，因为缺乏科学准确的专题分类和概念组织，不能很好地满足人们对信

息的内容而非形式的需求。这是信息检索和知识导航的无奈之举。

由于分类族性检索的优势和主题概念检索的优势都不容忽视,因此进一步开发好知识导航，破解分类和主题检索功能易用性的难题，仍然是图书馆学、情报学、信息检索的大课题。在这方面，我做了一些有意义的尝试，希望用户不需要掌握分类法、主题法就能直接进行分类和主题准确提问检索，让知识导航服务更容易实现大众化。目前已经取得实质进展，可以作为知识地图的一种软件来使用。本文以分类的易用性为例来谈艺术信息检索。

2.1 内嵌式智能导航，界面整洁

这里以中国国家图书馆的其中一个OPAC页面作为示例雏形，如图3。

图3

2.2 实时导航，不需要用户掌握分类法

在信息检索系统中，分类号通常是文献信息数据库中一个可检索的字段，用户进行“提问式”分类检索时必须输入分类号。由于分类号是由字母和数字组成的一种标记符号,用户不可能像使用自然语言那样用分类号准确表达自己的文献需求,特别是对于系统性强、结构完善、类目专深、标记制度复杂的中图法的分类号，用户更是难以表达，致使“提问式”分类检索途径长期形同虚设,用户寥寥无几。

我们不能强求用户掌握深奥的分类法。既然用户不懂分类法，又怎样才能轻松实现分类检索提问呢？为了突破这一难题，笔者研究设计了一系列实现方法。

2.2.1 鼠标点击分类检索框，立即呈现中图法的22个基本大类

图4：鼠标点击分类检索框的实时提示

说明：单纯的分类号含义抽象，不管用户是略懂分类号还是根本不懂分类号，仅仅实时显示分类号，用户一般难以接受。因此，类目的分类号和类目中文名称同时显示，让用户一目了然是必要的。

2.2.2 分类导航实时动态，只要移动鼠标即可快速实现层层引导，逐类浏览，直至类目最底层。如图5.

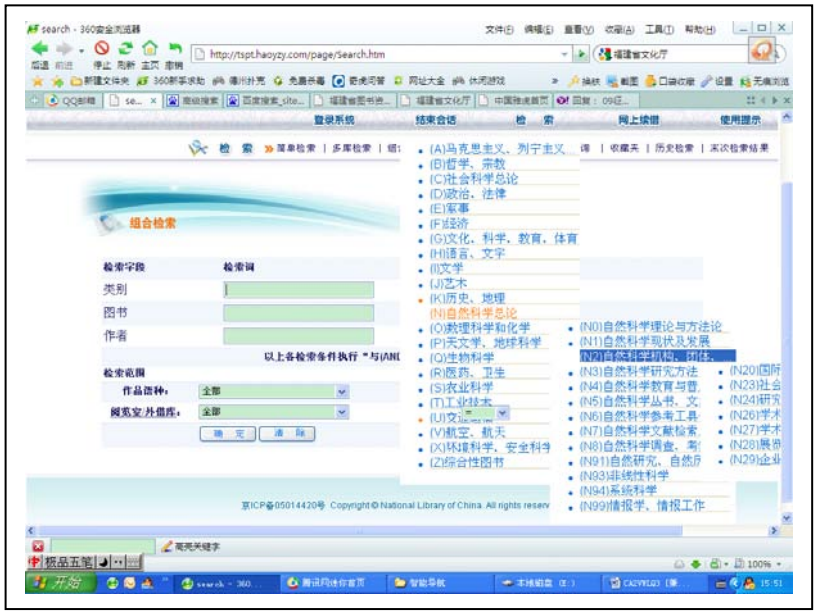

图5

当我们要查艺术信息时，可以用鼠标指向“J 艺术”，便能实现艺术类的层层引导，逐类浏览。

2.2.3 鼠标点选类目，类号直接填入检索框。如图6.

图6

说明：类号与类名需要同时显示，但用户选中某一类目时，却要求检索框只能表达选中的分类号，而不可表达类目的中文名称，以便检索。当然类目名称提示还可考虑体现上位类名称的说明注解、类目参见等各种注释，以便用户查考，但如果详细指明注释，一般读者不喜欢，有待斟酌。

2.2.4 多次导航功能：当鼠标点击所选类号，还能进行二次导航。如图7.

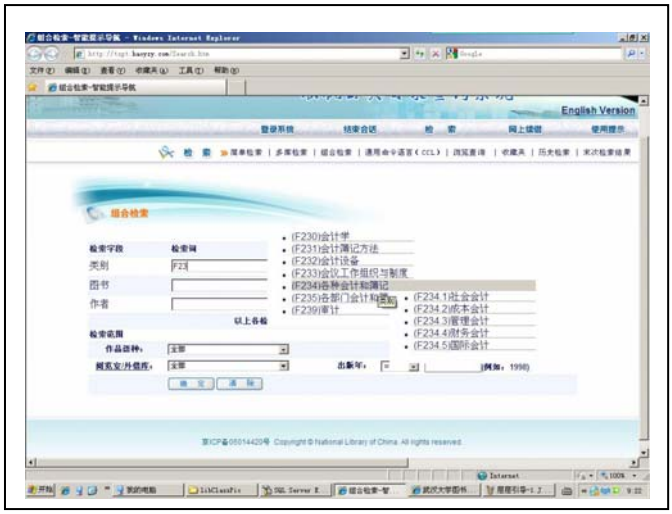

图7

2.2.5 读者不需懂得类号，通过关键词（主题词）的语义可实现分类检索。如图8.

图8

2.2.6 不需返回基本大类，可从任一节点进行导航。如图9.

2.2.7 克服多种语言障碍，自动翻译导航检索语言

通过检索语言的自动翻译，让不同语种人士可以使用多语种本体导航。例如，使用英语人士也能使用我国的中图法。

2.3 全面兼容分类法类号与实际文献类号，可按实际文献类号进行智能导航

中图法分类体系庞大、类目等级多,22个基本大类下包含多层子类目,其中第五版印刷版基础类目总数达5万多条,若按复分原则进一步细分，类目总数可达到近40万条[v]。如此众多的类目作为智能引导词，响应速度可以接受吗？基于中图法的多层导航实验和笔者曾经研究规范的药品智能引导词近5万个的单层实验看，目前的算法响应速度效果极好，如果规范更多的引导词也是可行的。因此分类导航的易用性设计，理论上可以针对目前世界上各种分类法的

所有分类号。特别是象中图法，在基本大类经过六、七层子类目后，一般即达底层，即使不使用鼠标引导而采用字母输入引导，其使用效率仅相当于输入一个7个字母的英语单词或者7个汉字拼音字母（约两个汉字的全拼），因此即使是达到多数基础类目底层的分类号精准请求，也很容易被用户接受。

分类法类号和实际文献类号有一定区别，许多文献的实际类号是根据一定的分类细则编制而成，比分类法基础类目号更详细。由于用户的检索目的是为了查找文献，因此智能引导应以实际文献类号为主。只有用户特别指明需要分类法全面类号时（主要是参考馆员用户），才考虑智能引导以分类法的分类号为主体。

不同系统的文献资源规模不同，各学科文献分布不均衡，各类目对应的文献量相差很大。有的类目聚集了大量文献，有的类目文献很少，还有大量的类目下未收录文献。因此，智能引导词应该根据文献实际数量量体裁衣，合理控制智能引导提示词数量和类目级别。由于智能引导的效率极高，以中图法22个基本大类，完全按照八分法扩展类目进行理想推算，笔者认为分类智能引导词在72万个以内，类目层级一般在7级以内到达底层，完全可以不必控制类目级别。

2.4 融合多种数据分析，智能调整引导内容，优势明显

智能引导导航可以根据类目层级关系、文献数量状况和使用频率等情况进行实时智能学习和调整引导，极大地方便了用户检索的需求，明显优于“浏览式”的分类法导航。基本比较如下：

导航方式 / 功能	分类法浏览式导航	分类法提问式导航
实现状况	已实现多年	首次实现突破
层层引导、逐类浏览	有，一般类目层级较少	有，可达任何类目底层
类目与实际文献匹配状况	不完全匹配	完全匹配
任一节点层层引导	难度大	容易
跨级导航功能	难度大	有
实时引导功能	静态	动态
类名类号同时显示功能	部分具备	完全具备
类号输入	无	有
关键词输入转换类号	无	有
自动翻译	无	有
检索记忆功能	一般不具备	有
检索速度及效率	慢、低	快、高
用户友好	一般	好

2.5 工程浩大，希望多种合作

以实现中图法智能导航为例，主要难度体现在如下几个方面：

2.5.1 中图法（第五版）基本类目五万多条，经过复分后的类目可达四十多万条，组织这些

类目工程浩大，而且专业性强；

2.5.2 要实现关键词（主题词）转化为类号，目前系统已经编辑词汇量达到六十多万个，数据量庞大；

2.5.3 根据各个图书馆收藏的实际图书进行导航，需要处理大量数据；关键词（主题词）转化为实际图书类号，实际图书类号转化为类名类号同时显示等，转化工作量更是巨大；

2.5.4 实现分类法智能实时导航，计算量巨大，函数算法要十分先进；

2.5.5 嵌入不同文献信息管理系统，需要对不同的信息管理系统做出不同程序设计；

因此，要实现国际上基于多种本体、多种语言的知识智能导航，需要寻求大量的国际合作。这样才能不断吸纳各国分类法、主题法、关键词技术以及检索导航技术的最新成果，并根据图书馆、博物馆等信息机构的实体文献变化，及时更新、形成应用，这样才能为每一个信息机构省却大量的时间和精力。也为各国用户省却大量的时间和精力。

在智能导航方面我的开拓主要在于：1.方便语义检索导航；2.无穷级可视化导航检索；3.鼠标层层引导导航；4.任意检索途径导航；5.专业化精准导航；6.实时动态导航。尽管在智能导航上实现了一些突破，但我所做的工作是有限的，非常欢迎有更多的合作，包括科研项目或市场化合作。

3 结束语

在信息的集成检索上，除了上述开发外，我也做了一些开放信息存取方面的工作，并试图进一步满足多元需求，让用户方便构建自己的集群资源、智能导航和一页式检索。这些开发目前正在考虑之中。从已经开发的平台看，可以为艺术信息检索或其它集成检索带来许多便利,在跨越多种自然语言和检索语言的障碍方面、在知识智能导航方面和信息资源管理“一页式”服务方面，已经显现了特色，希望能得到您的喜欢、帮助和深化合作。

参考文献

[i] 金胜勇.艺术信息资源共享研究[D].河北：河北大学，2004：5-6

ii 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

iii 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

iv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

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