Journal of Zhejiang University SCIENCE
ISSN 1009-3095
http://www.zju.edu.cn/jzus
E-mail: jzus@zju.edu.cn

Resource discovery in distributed digital libraries through visual knowledge navigation

GU Qian-yi†, AHMAD Faisal, SUMNER Tamara
(*Department of Computer Science, University of Colorado, Colorado 80309-0430, USA*)
†E-mail: Qianyi.gu@colorado.edu

Received Aug. 5, 2005; revision accepted Sept. 10, 2005

Abstract: In order to support users to search and browse for various resources, digital libraries are composed of discovery systems which provide user interface and information retrieve system. Recent researches in Information Retrieval have investigated different techniques through improving precision and recall to enhance the effectiveness of discovery system in digital libraries. In this paper, we present our work to enhance discovery system effectiveness with a different approach, through resource discovery based on visual knowledge navigation. In our Strand Map Services project under National Science Digital Library, we introduce the visual resource discovery system called conceptual browsing interfaces, to help educators and learners to locate, comprehend and use educational resources in digital libraries. The paper begins with a short introduction of the Strand Map Services. Then we illustrate the service architecture, the design and implementation of its major components. We will focus our discussion of how the visualization system of the Strand Map Services supports the visual knowledge navigation for distributed digital libraries. This includes the knowledge acquisition of the conceptual browsing interfaces, different knowledge representations in the system perspective and user interface perspective, visualization system modules, algorithm and Web services integration to use visual knowledge navigation to enhance resource discovery in digital libraries.

Key words: Discovery system, Digital libraries, Visualization, Concept space, Web services
doi:10.1631/jzus.2005.A1306 **Document code:** A **CLC number:** TP391

INTRODUCTION

In this paper, we present our work in the Strand Map Service (SMS), which provides concept map based browsing interfaces for resource discovery in a library independent manner within American National Science Digital Library (NSDL).

As critical components of digital libraries, discovery systems provide interfaces that support user to search and browse for resources in digital libraries (Hall *et al*., 1999). The effectiveness of discovery systems has big impact on the usability of digital libraries. Various techniques have been used to enhance discovery system effectiveness by improving precision and recall. SMS looks at another way of supporting resource discovery based on visual knowledge navigation (O'Donnell *et al*., 2002). In SMS, we introduce the visual resource discovery system called conceptual browsing interfaces, to help educators and learners to locate, comprehend and use educational resources in digital libraries.

The concept space of conceptual browsing interfaces is based on science education concept maps created by the American Association of Advancement of Science (Project 2061) to depict the concepts of learning objectives and their relationships for K-12 students in the U.S. (Project 2061, 1993). These interfaces are comprised of interacting visual components containing different views onto this concept space. By providing navigational and orientation cues that are typically lacking from traditional keyword or fielded search interfaces, the conceptual browsing interfaces try to help users understand the learning objectives and their relationships by providing visualizations and to facilitate the formulation of search requests for educational resources related to learning goals and the retrieval of these resources from distributed digital libraries.

For forming search requests and retrieving resources, the conceptual browsing interfaces serve as a single, unified gateway for users to explore distributed resources. By interacting exclusively with our conceptual browsing interfaces, the user can search directly or indirectly for distributed Web resources residing in different digital libraries such as the National Science Digital Library (NSDL, http://nsdl.org/community/documents.php), the Digital Library for Earth Science Education (DLESE, http://www.dlese.org), the Harvard-Smithsonian Digital Video Library (HSDVL) and in a larger domain such as Google.

In this paper, we focus one major component of SMS: Visualization System and how it is used to support the visual knowledge navigation for distributed digital libraries. This includes the knowledge acquisition of the conceptual browsing interfaces, different knowledge representations in the system and user interfaces, visualization system components, algorithm and Web services integration to use visual knowledge navigation to enhance resource discovery in digital libraries.

STRAND MAP SERVICES

Benchmarks and Strand Maps

Benchmarks are learning goals which describe what learners should know, or be able to do, at key stages in their primary and secondary education. Strand maps are constructed based on the benchmarks stated in Benchmarks for Science Literacy (Project 2061, 1993). The Atlas of Science Literacy (Project 2061, 2001), published by the American Association for the Advancement of Science (AAAS) and the National Science Teachers Association, features strand maps on topics that are important to science literacy. Each map consists of node-link representations illustrating a set of relationships between benchmarks organized around a topic. High-level descriptions of the benchmarks are provided in the nodes, while the links depict the interrelationships between benchmarks. Each map contains vertical strands reflecting key ideas in that topic. Each strand is cross-referenced by grade levels (K-2, 3-5, 6-8, 9-12) to illustrate how student understanding develops over time. They provide a visual representation that emphasizes the coherence intended in the benchmarks and encourage both teachers and learners to make connections between ideas.

Strand Map Services Overview

We build up Strand Map Service on the knowledge base of strand maps and extend the services support the needs of K-12 (primary and secondary school) educators and learners, and digital library developers through the provision of graphical conceptual browsing interfaces and programmatic Web service interfaces. The graphical browsing interfaces help learners and educators to locate and use learning resources in educational digital libraries (Sumner *et al*., 2003). Their graphical representations help learners and educators to understand the learning objects and their internal relationships. To achieve these learning goals, the user needs to retrieve educational resources aligned to those objects and use the resources effectively.

The Web service interfaces enable digital library developers to easily construct conceptual browsing interfaces appropriate to the needs of their specific library audiences using dynamically generated visual components provided by the Service.

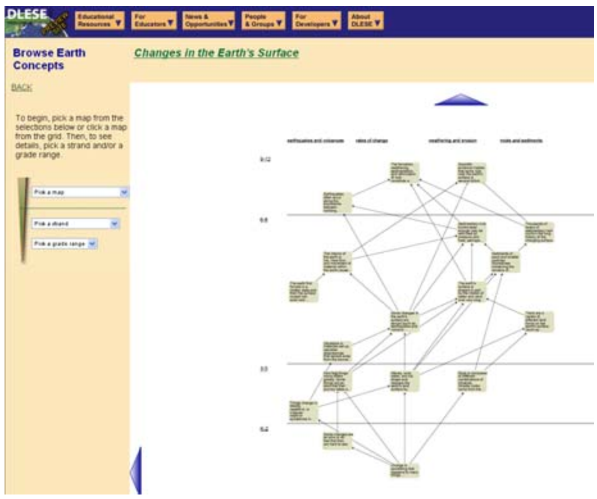

Fig.1 Example of Strand Map Services embedded with DLESE (http://preview.dlese.org/sms/start?conceptMaps=SMS-MAP-0048&browseQuery=GO)

Strand Map Services components

To enable the construction of conceptual browsing interfaces, the Strand Map Service (SMS) is composed of three major components:

(1) The Concept Space Interchange Protocol (CSIP). CSIP is implemented as REpresentational State Transfer style Web service. The protocol is used to construct the queries to retrieve strand maps and their fragments from strand maps information space, and is also used to specify the representation of the dynamically generated strand maps from underlying data model. By supporting a rich query language, the CSIP supports comprehensive capabilities for strand map management, extraction, and generation (Ahmad, 2004).

(2) The SMS visualization system. It is composed of the visualization algorithm to dynamically generate strand maps and strand map fragments from the underlying data model. It also contains the component which rendering the visualization into different formats such as SVG and XML for visual browsing interfaces.

(3) SMS Data Model. It captures rich data from strand maps that enable querying the information space and dynamic rendering of strand maps. The data model captures the concepts, the relationships between concepts, and the educationally relevant metadata.

VISUAL KNOWLEDGE NAVIGATION

As we discuss earlier, SMS provides conceptual browsing interfaces for learners to navigate in the information space to build up better scientific understanding. Further more, the visual knowledge navigation is used to enhance resource discovery. In this session, we discuss in details on the acquisition of the knowledge base to create the conceptual browsing interfaces; the knowledge representation in system perspective and user perspective; the implementation of the visualization; and finally, how the visual navigation is used to enhance resource discovery for digital libraries.

Knowledge acquisition for visual interfaces

The knowledge base of the visual interfaces is composed of strand maps and their semantics underlying. We design and develop a cataloging process to acquire such knowledge: the semantics of these containers were defined and the best practices for cataloging were articulated to achieve consistency. The cataloging process involves two steps:

(1) Text Acquisition: The cataloger is given an Excel sheet to input the text data associated with each object. The fields of each kind of objects are defined in CSMF.

(2) Graphical Relation Acquisition: To capture the graphical relationships between objects in strand maps, the process provides the cataloger with an E-learning tool called GetSmart (Marshall *et al*., 2003). The GetSmart system was built based on a model of how communities and individuals create and share knowledge. The system is composed of concept map building applet as shown in Fig.2. The cataloger uses this applet to input relationship data into our knowledge base by generating XML data through exporting concept maps from GetSmart. The integration of GetSmart with SMS data model is through style sheet which transfers GetSmart XML data to data validated through SMS data model.

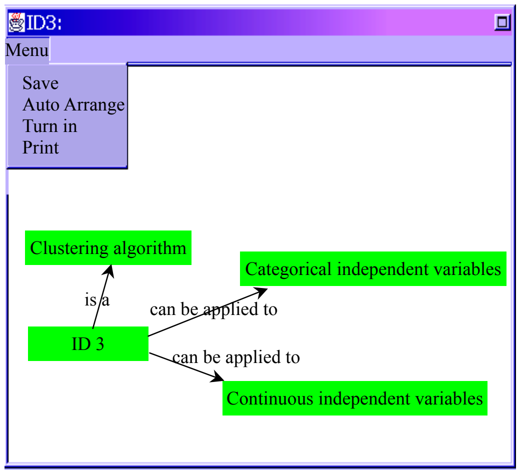

Fig.2 Concept map building applet from GetSmart

To reach the goal of generating graphical representation of concept map in a dynamic manner, our approach is to build an intelligent system which simulates the human map drawing. To understand the cognitive process of human's reasoning on map drawing is important. Also, the understanding of these processes needs to be formalized in a programmatic way so that these understanding are able to help the system be built up.

Our methodology intertwined expert knowledge acquisition activities, to inform design and evaluation of algorithm for generating visual browsing interfaces. Our knowledge acquisition activities involved analyzing the published strand maps and interviewing professional strand map developers. These activities enabled us to articulate the semantic constraints that needed to be preserved and the desirable aesthetic heuristics used by human experts who created the published maps.

The readability of visual interfaces can be evaluated by graph drawing conventions, aesthetics, rules and efficiency (Tollis, 1999; Sumner *et al*., 2004; Tufte, 1990). The goal of the visualization system is to generate visual interfaces which have the readability closed to strand maps with human heuristic effort. Meanwhile, another important requirement is that it preserves the semantics of the strand map in the graphical presentation. We have defined the following heuristic matrix to guide the design of the algorithm of visual interface:

Required Element:
- Preserve sub domain relation
- Preserve the internal relations between nodes
- Nodes overlapping should be avoid

Optional Element:
- Minimize the white space in the graph
- Minimize (optimal to avoid) edge cross node
- Minimize edge crossing
- Minimize length of any given edge
- Minimize total length of edges
- Avoid bend of the edge

Visualization system architecture

In the following we describe the different modules that are involved in generating the visual interfaces.

(1) User Interaction Module: Includes the user interface and query generator. User interface interacts with users so that their specific learning inquiries and search contexts are captured by the system. Query Generator: User input triggers the sending of a CSIP query to the concept space (Sumner *et al*., 2004). The query generator is bound to the user interface so that the CSIP query will be able to retrieve the data in response to the user's learning inquiry. The query service is used as the service middleware for generating the conceptual browsing interface. This service request is used to retrieve the strand map information that can be used by client digital libraries for resource discovery and navigation.

(2) Data Interaction Module: Retrieves the data from knowledge base represented as CSMF data in the back-end database. It is represented according to the object model of the novel visualization module. This module: Parses the CSIP query and generates the corresponding SQL query; Executes the SQL query and retrieves the result set from the database.

(3) Visual Interfaces Initialization Module: This module initialize the generation of visual interfaces. It serves to implement the visualization algorithm in Java; use the Rules and Aesthetics component to create the requested conceptual browsing interface; and encode the result in SVG format.

(4) Visual Interfaces Composition Module: As discussed earlier, the visualization algorithm can dynamically generate a graphical representation of concepts and their relationships (Gu *et al*., 2004). Novel visualization is based on the successful retrieval of the requested visual components and their organization or assembly into a coherent interface using our visual algorithm. The composition module includes several units: Visual Component Identifier is to identify the type of visual components it is going to generate. Visual Component Retriever works with the data collector to retrieve the requested data from the concept space. Visual Components Synthesizer synthesizes the data from different visual components into a single, new model object. The object is then used by the visual algorithm to dynamically generate a novel visualization.

Knowledge representation of conceptual browsing interfaces

1. Representation in system perspective

In order to guide the acquisition of the knowledge base of conceptual browsing interfaces and information discovery process, we design the metadata framework of SMS called the Concept Space Metadata Framework (CSMF). With the definition of the CSMF, we are able to acquire of strand maps and its constituents along different semantic dimensions and are able to retrieve according to different criteria. The principles that are used to guide the development of CSMF are:

(1) Accurate representation of AAAS information space for future innovative use;

(2) Making it useful in current practices of edu-

cational digital library domain (DCMI, 2003) (Specifically NSDL).

2. Representation in user perspective

With the acquisition of the knowledge base composed of numerous strand maps, the goal of our service is to represent the information space to end users as conceptual browsing interfaces. The conceptual browsing interfaces will enable educators and learners to discover educational resources that support the learning goals, or benchmarks, articulated in the Strand Maps; to browse the interconnected learning goals in the Strand Maps; and to enhance their own content knowledge by using the service to explore important background information on the learning goals, such as prior research on student misconceptions.

Here we report on the algorithm that we have created to dynamically generate visual interfaces. The goals of the visualization algorithm are three-fold. First, the algorithm must be able to generate all the promising components identified in the earlier studies. Second, the algorithm must preserve both the semantics underpinning the strand maps and the aesthetic standards of the AAAS human experts who made the original paper based maps. Third, the algorithm must enable the NSDL Strand Map Service to extend the maps modeled in the Atlas, by representing interdisciplinary relationships between benchmarks and strands that cross map boundaries imposed in the two-dimensional paper-based publication and by enabling users to navigate through this concept space using new graphical representations.

Enhance resource discovery through Web services integration

As we have seen, conceptual browsing interfaces provide learners with novel visualization in particular learning domain. Through interacting with these visual interfaces, learners are able to explore the information space with visual knowledge navigation. These kinds of activities provide learners with opportunities of building up better understanding of scientific knowledge. In this session, we will show our approach to enhance resource discovery of digital libraries by visual knowledge navigation through conceptual browsing interfaces.

Recent studies (Sonal, 2004; Butcher *et al*., 2005) showed that educational value of digital libraries can be enhanced by using strand maps for navigation and resource exploration. Students using the strand map interface engage more with science content compared to those using a keyword-based searching interface. Another finding is that the educational benefit of strand maps diminishes when students move from strand maps to the resource list resulting from keyword searching or clicking at a concept in the strand map. This can be attributed to the disappearance of the context that was provided by the strand maps. To overcome this problem we have embedded the digital library resources in the strand maps so that the user (e.g., a student) presented with a strand map can see the digital library resources within the context of the strand map. Fig.3 shows a portion of a strand map along with the retrieved resources.

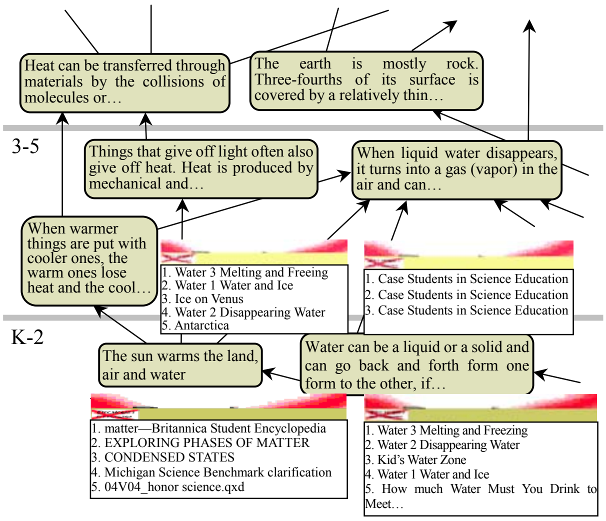

Fig.3 Resource discovery from Google, DLESE, NSDL and HSDL through SMS visual interfaces navigation

In order to enable flexible integration of digital library resources into the strand maps we have adopted a Web-service composition approach. Many resource discovery systems of digital libraries such as Google, NSDL and DLESE support Web services. Such systems provide Web search API through HTTP protocol. The APIs are integrated into conceptual browsing interfaces by using SVG post-processing. Through such processing, the search APIs for distributed digital libraries discovery system are parameterized and encoded into visual interfaces represented in SVG. The users are able to dynamically retrieve resources aligned to objects in the visual interfaces from distributed digital libraries.

The process of the resource discovery in digital libraries with visual knowledge navigation embedded is illustrated as following:

First of all, a digital library user enters a digital library Web site. Then the user clicks on a link to open a strand map. This action will trig the digital library use the SMS client library to send a request to retrieve the target map from the SMS server. The retrieval request also defines which query string should be attached to each concept in the strand map. SMS server constructs the map and then attaches the requested query string, as a hyperlink, to each concept within the SVG. Once the SMS client library receives the map it parses the SVG and uses the embedded hyperlinks to retrieve search results e.g. Google, NSDL, DLESE, etc. The SMS client library then parses the search result for each concept and adds those results in the strand map SVG. Finally, the search result embedded strand map is returned to the user/client.

CONCLUSION

Our research presents a novel approach to enhance the digital library discovery system by contextualizing the educational resources through conceptual browsing interfaces. The conceptual browsing interfaces visualize the key concepts and their relations in particular scientific domain (strand maps) and resources in digital libraries. Through navigation and exploration in the interfaces, the user is provided services for more thorough understanding of the domain knowledge and related resources. This approach has proved to be useful because users engage more with content compared instead of search strategy compared to those using a keyword-based searching interface (Butcher *et al*., 2005). Future work will examine the generalizability of this approach to other visualization systems for digital libraries discovery system.

Reference

Ahmad, F., 2004. A Strand Map Service for Educational Digital Libraries. Unpublished Master's Thesis, University of Colorado, Boulder.

Batttista, G., 1998. Graph Drawing: Algorithms for Visualization of Graphs. Princeton Hall.

Butcher, K.R., Bhushan, S., Sumner, T., 2005. Multimedia displays for conceptual search processes: information seeking with Strand Maps. Submitted to ACM Multimedia Journal.

DCMI (Dublin Core Metadata Initiative), 2003. Dublin Core Metadata. http://dublincore.org/documents/dces/.

Gu, Q.Y., Ahmad, F., Molina, F., Sumner, T., 2004. Dynamically Generating Conceptual Browsing Interfaces for Digital Libraries Using SVG. 3rd Annual Conference on Scalable Vector Graphics, Tokyo, Japan.

Hall, R.H., Hall, M.A., Saling, C.B., 1999. The effects of graphical postorganization strategies on learning from knowledge maps. *Journal of Experimental Education*, **67**(2):101-112.

Marshall, B., Zhang, Y.W., Shen, R., Fox, E., Cassel, L.N., 2003. Convergence of Knowledge Management and E-Learning: the GetSmart Experience. Proceedings Joint Conference on Digital Libraries, Houston, TX.

O'Donnell, A.M., Donald, F.D., Richard, H.H., 2002. Knowledge maps as scaffolds for cognitive processing. *Educational Psychology Review*, **14**:71-89.

Project 2061, AAAS (American Association for the Advancement of Science), 1993. AAAS Benchmarks for Science Literacy. Oxford University Press, New York.

Project 2061, AAAS & NSTA (American Association for the Advancement of Science, and the National Science Teachers Association), 2001. AAAS Atlas of Science Literacy. NSTA Press, Washington, DC.

Sonal, B., 2004. Designing Visual Components for Conceptual Browsing Interfaces. Master's Thesis, University of Colorado at Boulder.

Sumner, T., Bhushan, S., Ahmad, F., Gu, Q.Y., 2003. Designing a Language for Creating Conceptual Browsing Interfaces for Digital Libraries. Proceedings of the third ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL 2003, Houston.

Sumner, T., Ahmad, F., Bhushan, S., Gu, Q.Y., Molina, F., Willard, S., Wright, M., Davis, L., Janee, G., 2004. A Web service interface for creating concept browsing interfaces. *D-Lib Magazine*, **10**(11).

Tollis, I., 1999. Graph Drawing and Information Visualization. ACM Computer Surveys.

Tufte, E.R., 1990. Envisioning Information. Graphics Press.