Exploring the Semantic Types of Relationships for Visual Query Development

Ushakov A.S., Agapova O, Smith T, Gerber M, Ushakov A.A
University of California Santa Barbara, Stanford University.
aushako@alexandria.ucsb.edu

The paper describes a cognitively and semantically appropriate, simple-to-use, graphical query interface KnowVis for NSDL, viewed as a vitally important tool for educators and students. The Paradigm and Principal Hypothesis are formulated. The four basic research questions are put forward. Then, fundamental vision of Query Interface is formulated. A vignette of practical in-class implementation is provided. The Evaluation methods are formulated and Research plan is drafted. Finally, the relevant research is sited. The work is based on developments and findings of Alexandria Digital Earth Prototype project at UC Santa Barbara.

1. Statement of Need.

The importance of the Internet and Digital Libraries (DL) in the educational community is indisputable. They have penetrated every aspect of society—school, homes, and workplace. For example, a recent study shows that approximately 85% of middle- and high-school students use the Internet in school. Approximately one-half of public school teachers who have computers or Internet access report that they use them for classroom instruction.

However, to find appropriate instructional materials within the National Science Digital Library (NSDL) collections, users must formulate their queries, locate sites, analyze search results, adapt the materials, and organize them for the needs of their particular classroom learning activities. Despite advances in NSDL services over recent years (Zhang '02, Veerasamy '96), a significant gap exists between the actual needs of teachers and students and how the present NSDL system processes queries and displays NSDL discovery results (Cook '03). These services do not differ from general Web-based query and display operations and they ignore users' cognitive individuality.

The Institute of Museum and Library Services calls for research showing how knowledge organization tools can be used to improve the user's experience with digital collections, and for more projects beginning with evaluation of user and organizational needs that can be incorporated into DL design .

1.1 An advanced educational query interface needs to be designed and created. The interface must employ an appropriate cognitive paradigm and a semantic-based graphical format to address the user's information needs. Current DL and most Web search systems, including those utilizing powerful ranking algorithms, do not meet user needs completely (Tomita '00). This is because existing search systems expect users to submit well-specified queries, a difficult task for non-specialist users. Teachers and students with weak technological skills and students unfamiliar with scientific knowledge domain often search the DL with a poor understanding of how to specify their information needs.

1.2 It is vitally important to develop an interface that assists educators and students in sifting through multiple pages of metadata identified by their query. The number of NSDL and Web-based learning resources are growing rapidly. This complicates the processes of analyzing unwieldy and poorly organized information retrieval (IR) results. Users experience cognitive and semantic difficulties in sifting through DL search results, usually represented in the form of metadata-derived lists of links.

1.3 Goal. The answer to this challenge is a cognitively and semantically appropriate, simple-to-use, graphical query interface called KnowVis, which would be validated and used by science teachers and students, and which will improve existing NSDL query formulation.

We expect that KnowVis could trigger a breakthrough in NSDL users' productivity. It will assist with the following:

- ***Visualizing the NSDL scientific concept space*** (controlled vocabulary, thesauri) and multiple relationships among the concepts, to make the structure of science understandable and visible, and to help individuals navigate through scientific knowledge;

- ***Transforming a traditional Digital Library query*** from menu selection or form-filling into a graphical, concept-map-like, dynamic environment, which semantically and cognitively supports clients in meaningful query formulation;

Implicitly modifying the traditional Digital Library discovery results display interface from plain metadata-derived lists to an ergonomically appropriate, visual display that assists with self-organization of the discovery results.

2. Innovative Paradigm and Technological Advances

2.1 Paradigm. In order to bring improved visual and semantic quality into NSDL interfaces, we will shift the focus of graphical representation of query and discovery displays from visualizing the technological features of information retrieval to visualizing and supporting the human aspects of knowledge acquisition. KnowVis will provide a new mechanism to help content developers combine resources from different collections semantically. This novel approach takes advantage of the synergy between two highly correlated, known models:

1. *Thinking Maps* (Hyerle '00), which are well-studied, visual models of cognitive processes that support meaningful reading, writing, and problem solving , and

2. *Ontological Model of scientific concepts* from Alexandria Digital Earth Prototype (ADEPT OM). ADEPT OM describes the knowledge space using seven basic Semantic Types of Relationships (STR) among concepts (Smith). The model reflects perspectives of both the natural and informational sciences communities, rather then the vision of a few individuals .

Integrating these models will create a graphical user interface for effective NSDL query and search results representation, especially useful for an educational audience.

2.2 Technological advances and Interoperability. This method will be interoperable with the NSDL core integration interfaces and compatible with the Alexandria Digital Library and the Digital Library for Earth System Education. KnowVis utilizes an interactive, browser- and XML-based approach.

3. The Hypothesis

3.1 The Hypothesis. KnowVis development is supported by the central research hypothesis consisting of several assumptions:

1. There are similar cognitive processes present in both regular human reading comprehension and mental processing of an IR results list generated by a search engine.

2. All scientific concepts in any knowledge space are nominally bonded between each other through various semantic types of relationships (Smith '03).

3. These cognitive processes of reading comprehension are based on realizing the relationships through which the concepts could be linked, both for text and IR list.

Therefore, a system designed on these cognitive principles, coherent with the semantic types of relationships among the concepts, can guide and support users in formulating queries and exploring IR results.

The testing of this hypothesis will be based on comparative analysis and by combining two independent models, such as Thinking Maps and a ADEPT OM, into the KnowVis system. In our preliminary research, we discovered a certain degree of correlation that exists between these two models (see table 1). The correlation exists between (a) types of human cognitive processes of comprehension reading, and (b) the semantic types of relationships among scientific concepts, which are used for information space description (Agapova '03, Smith '03).

3.2 Thinking Maps were developed as a language for learning in 1988 by Dr. David Hyerle. They are theory-embedded (Hyerle '00) and each map brings together a form and function (Novak '00).

The *function* is defined by a respective cognitive structure for thinking. Cognitive patterning includes: (1) defining a perspective or a point of view, (2) sequencing or determining how elements fit within a structure, (3) determining parts, (4) cause-effect reasoning, (5a-b) comparing and contrasting, (6) characterizing or attributing, and (7) applying: executing (applying a procedure to a familiar task) or implementing (applying a procedure to an unfamiliar task) (Anderson '01) They support mental processes that help a reader to (a) highlight the important concepts, (b) define relationships among concepts, and (c) visualize them. (Table 1).

The *form* is represent by a visual pattern of each cognitive process: (1) Circle map, (2) Flow map, (3) Tree-map, (4) Multi-flow map, (5a-b) Bridge map and Double-bubble map, (6,7) Bubble map (see Table 1).

3.3 ADEPT Ontological Model (OM) of Scientific Concept differs from other ontological models by introducing semantic types of relationships among scientific concepts. The model was developed and elaborated in the framework of the ADEPT project (Smith '03). It was implemented in XML SPY; alpha and beta tested; and was populated for the ADEPT Visual Learning Environment knowledge base. Test-bed repositories were devised for both Tamino (XML) and MySQL (relational) databases. Later this model was modified for needs of K-12 science education (Agapova ’03).

3.3.1 The model structure. Two major blocks can be highlighted in the modified ontological structure: (a) the identity of a concept and (b) the Semantic Types of the Relationships (STR) or, in other words, the semantic code of the concept. The identity of a concept holds its unique attributes: Facet, Term, Explanation, and Representations. The STR part of the model describes possible ways (relationships) in which the concept could be incorporated into the scientific knowledge space. The Semantic types of relationships are: (1) disciplinal contextual area, (2a-b) changes in time and location, (3a-b) hierarchies, (4) cause-and-effect, (5a-b) association by analogy or contrast, (6) properties, and (7a-b) implementation of knowledge and operations (see Table 1). Several sub-types can extend each type of relationships. The model was developed by analyzing and integrating numerous classification and integration approaches for scientific knowledge (Ranganathan ’78, Sowa ’00, Barfourosh ’02).

3.3.2 Vision of NSDL Scientific Concept Space. Applying the inductive approach and multi-relational descriptions, our ontological model allows for an innovative description of the scientific concept space (NSDL-controlled vocabulary).

Deductive method. Usually, librarians use deductive methods of categorizing the knowledge space by clusters: topics, categories, or taxonomies (Hill ’02). This is a cluster-centered system. The *concept* here is a component of the cluster, and one concept can be found in many clusters.

To find information about a concept, the user has to search many clusters. Very often, students do not know to which area a particular concept can belong.

Inductive approach. In contrast to the deductive method, the ADEPT ontological model takes advantage of an inductive approach (thesauri-like). This is a concept-centered system. The model allows deriving the clusters from multiple interconnections of individual scientific concepts. This means that:

(a) NSDL concept space structure is built of nodes (concepts) and links (relationships);

(b) Both links and nodes are independent and variable components;

(c) A link has multiple semantic meanings expressed in STR. In contrast, the link in semantic network structures has only one meaning—semantic similarity .

(d) The concepts connected by the same STR belong to one cluster. *To find information about the concept of interest,* the user needs to know only the term and to choose the appropriate relationship out of the seven proposed.

3.3.3 Priorities in the approaches. We favor the inductive approach, since: (a) there is no standard for scientific knowledge categorization in education and DL; (b) individuals rarely agree or accept uniform categorization approaches ; (c) our brain builds unique knowledge structures based on previous knowledge of individuals and personal characteristics.

3.4 Similarity between the Models. Although ADEPT OM and Thinking Map models were developed independently, they both demonstrate a high level of correlation (Table 1). This correlation between cognitive processes of reading comprehension and semantic relationship types suggests that the KnowVis interface could be developed by merging the Thinking Map and ADEPT OM models.

4. Research Questions

4.1 How do users naturally analyze the IR results list without technological and cognitive support? Are the cognitive processes of reading comprehension and mental processing of the IR result lists similar?

4.2 What semantic type of relationships among scientific concepts do participants use to connect the concepts meaningfully? Are those semantic types different from relationships that we plan to implement in the KnowViz system? What keywords, if any, can indicate each semantic type of relationship?

4.3 How can each semantic relationship type be visually represented to best satisfy a majority of users? What are the mental models that participants build when they connect scientific concepts with different types of relationships? How much data, or in other words, how many nodes and links should be presented to the user? Does the rule of human perception “7 plus minus 2” apply in this case?

4.4 How do individual differences such as (a) relevant IT experience, (b) information management behavior, (c) spatial and verbal ability, and (d) structuring of the scientific domain influence users’ behavior and outcomes, as measured with and without KnowVis support?

5. Vision *of the KnowVis* Query Interface

5.1 At a Glance. This approach allows, on one hand, designing a well organizing query instead of typing plain keywords, and on the other hand, receiving a self-determined, spatial layout for search results instead long list of “hits.” A vignette and a hypothetical model of KnowVis are shown below.

Table 1. Correlation between the ADEPT Ontological Model and Thinking Maps

<table>
<tr><th colspan="2">ADEPT OM</th><th colspan="2">Thinking Maps</th><th colspan="2">Examples</th><th>Visualization</th></tr>
<tr><td colspan="2">1. Contextual area</td><td colspan="2">Defining a point of view</td><td colspan="2">Water, Chemistry</td><td>Circle map</td></tr>
<tr><td rowspan="2">2. Changing</td><td>Time</td><td rowspan="2">Sequencing</td><td>Chronological</td><td colspan="2">History</td><td>Flow map</td></tr>
<tr><td>Location</td><td>Special</td><td colspan="2">Earth Distribution</td><td>Globe, geo-map</td></tr>
<tr><td rowspan="2">3. Hierarchies</td><td>Whole/par</td><td rowspan="2" colspan="2">Determining parts</td><td colspan="2">Structure of Water</td><td rowspan="2">Tree-map</td></tr>
<tr><td>Is a</td><td colspan="2">Water Resources</td></tr>
<tr><td colspan="2">4. Cause-and-effect</td><td colspan="2">Cause-effect reasoning</td><td colspan="2">Chemical Reactions</td><td>Multi-flow map</td></tr>
<tr><td rowspan="2">5. Association</td><td>Analogy</td><td colspan="2">Comparing</td><td rowspan="2">Theories</td><td>Similarity</td><td>Bridge map</td></tr>
<tr><td>Contrast</td><td colspan="2">Contrasting</td><td>Differences</td><td>Double-bubble map</td></tr>
<tr><td colspan="2">6. Properties</td><td colspan="2">Characterization</td><td colspan="2">Properties of Water</td><td>Circle-bubble map</td></tr>
<tr><td rowspan="2">7. Implementation</td><td>Knowledge</td><td rowspan="2">Applying</td><td>Executing,</td><td rowspan="2" colspan="2">In Ecology, Pollution</td><td rowspan="2">Square-bubble map</td></tr>
<tr><td>Operations</td><td>Implementing</td></tr>
</table>

Vignette. Ms. Jones intends to prepare her next lesson. She opens the browser interface and enters the name of her desired topic, "Importance of Water to the Community," to find supplemental instructional materials. The search engine displays a list of references 88 web pages long. Ms. Jones sighs, but on exhalation, she remembers the *KnowVis* button and clicks it. Instead of a long list of references, an appealing visual structure for her query reformulation is displayed. She heard from a colleague that this tool could help her specify the query, sort the references, and represent them in a graphical way, which anyone could easily perceive.

Ms. Jones clicks on three relationships arrows in a sub-query map and KnowVis builds the corresponding consolidation maps.

First– Hierarchies or Types of water resources – looks like a tree graph.

Second – Location map shows a state map with major water resources.

Third – Implementation – displays the references about agricultural, industrial, and medical applications in a daisy-like form.

Then she selects several concepts in each map and clicks on "Search." Soon, icons representing instructional materials appear, clustered around each node of her consolidation map. The icons represent particular parts of documents and could be opened in a separate window. Additionally, *KnowVis* sorts instructional material in terms of their use with students having low/medium/high motivation and for students with low/medium/advanced achievement. Ms. Jones checks the DL resources, then drag and drops the appropriate materials into her lesson plan. She finalizes her work and saves the plan to a school server. Done! Ms. Jones glances at her watch. It was three times faster than completing the same task without the assistance of the KnowVis.

To achieve the functionalities described in Vignette, the KnowVis architecture has to have two main components: a human-designed Knowledge Base and a software client with server-side components, both supported by corresponding services.

5.2 Knowledge Base (KB). The KB is defined by our ontological model of a scientific concept and stores background knowledge as records of scientific concepts and relationships among them. The array of concept records in XML format builds the content of the knowledge base. The content is coordinated with National Science Educational Standards (NSES): Physical and Earth science K-12 content standard.

5.2.1 Functionalities. Two principal functionalities can be highlighted in the knowledge base:

First, KB functionalities imitate the activity of human working memory (Wolfe '01) and compare current "perceptual" keywords with ones previously stored in the KB (matching the key words to the concepts that are stored in KB).

Second, KB functionality simulates the operation of human semantic memory. The KB search engine is able to find additional concepts meaningfully related to the initial keywords through the seven basic semantic types of relationships.

5.3 Software Client

The KnowVis KB functionality requires a User Interface (UI). The UI is located on the client side and is organized into two windows to sustain each process in series: Sub-Query and Consolidation.

5.3.1 Sub-Query map is a semantic pattern for revising the query via direct graphical manipulation of the background knowledge.

Now let us follow the previous **Vignette** step by step with the respective comments about the KnowVis actions.

Step One. User enters initial keywords.

The map represents searching keywords visually in the ADEPT concept model format, which includes the identity of a concept and arrows with STR names.

Step Two. To revise the query, the user simply chooses the discipline area.

After that, the system replaces general STR names with names that are used within particular disciplines and are familiar to students and teachers. For example, Hierarchy would be changed into Structure of Matter. The system also changes the order of arrows appearing around the keyword in a way that is traditional for the particular discipline.

Step Three. In order to complete the query revision, the user must define the STR by clicking on the corresponding arrow and examine the consolidation maps.

KnowVis sends keywords and the STR to the local Knowledge Base, searches for relevant concepts, which are bonded via STR to the initial keywords, and builds consolidation maps.

5.3.2 Consolidation Map is a mechanism for transforming the list-based retrieval results into a graphical and cognitively appropriate format. The map is a thinking map-like, graphical representation of the relevant concepts. It consists of nodes of key words in the center, which are bonded with nodes of concepts.

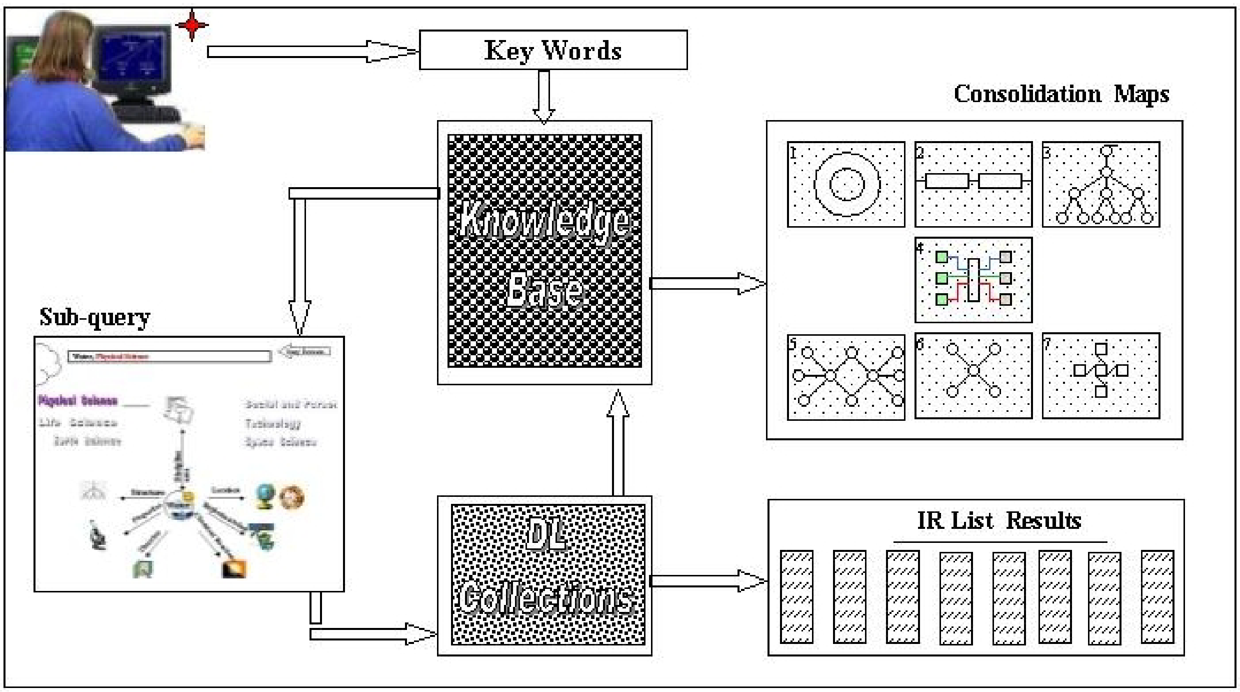

Diagram 1. KnowVis **Hypothetical Model**

Each STR is represented by a specific graphical pattern

Step Four. Users can assign the importance of the relevant concepts suggested by the system by clicking on corresponding nodes. He or she can also rearrange the system-generated spatial layout of search results using a simple drag-and-drop interface.

The UI service then (a) removes the concepts and relationships that the user did not mark up in the initial maps; (b) displays the new, user-defined map graphically; and (c) sends the new query to the DL search engine.

To display the search results we plan to use TileBars – a graphical software tool from Xerox PARC. Using this software our KnowVis UI depicts discovered documents as icons. Icons are grouped in segments around the nodes in a graph. Each icon shows the relationship between the terms in a query and the documents retrieved in response to that query.

Since 2003, NSDL employs the Lucene search engine and continues using several protocols such as AQL, SDLIP+Z39.50 T102. The KnowViz protocol will be built on top of the NSDL implementations to provide other developers with full access to the visual query and IR display.

Step Five. At this stage, the user can examine the organized discovery results and save the consolidation map to a server.

6. Evaluation Plan and Research Method

The evaluation process will use in-depth qualitative and quantitative methods, collecting the systematic data in support of KnowVis development (formative evaluation purposes). Analysis of observations, interviews, documentation and other evaluation materials will be applied to monitor project activities, and also to serve the dataset for studies that focus on project outcomes (summative evaluation purposes).

6.1 Participants

To attain scalable, cumulative data that addresses the research questions, the project will identify and recruit approximately 24 teachers (in equal proportions for each discipline, including chemistry and earth science) and about 120 of their middle and high school students (5 students per teacher). A diverse social population will be represented, including rural, urban, minority, and private schools, both with and without technological expertise.

6.2 Developing the KnowViz Prototype

The team will develop a scalable prototype of KnowViz in time to start multidisciplinary research. The prototype will provide necessary and sufficient functionality for conducting all experiments described below. Design of the KB and UI will be separated from the retrieval engine specifics.

6.2.1 Discipline Topic and Sub-topics. The KnowVis prototype will be based on an interdisciplinary topic such as "Water." Subtopics about water, which are based on basic STR, are: (1) Relevant disciplines – Chemistry, Physics, and Earth science; (2) 2.a History of water (origins on Earth); 2.b Distribution of water sources on the Earth; (3) Structure of water; (4) Chemical reaction with water; (5) Theories of substance (similarity and distinguishing; (6) Properties of water; (7) Using water. Selection of the knowledge domain "Water" is justified by the fact that both physical and earth science teachers and students can interact with the domain (Aivazian '03).

6.2.2 The knowledge base will be populated by XML records of scientific concepts that belong to the topic and subtopics mentioned above. Information about "Water" will be represented through STRs for more than 140 related concepts. The relationship types are: 1) Discipline area, 2) Time and Location, 3) Hierarchies, 4) Comparison and Contracting 5) Cause-Effect, 6) Properties, and 7) Implementation. Each concept record will contain all completed elements that were described in the KnowViz ontological model (part 5.2.1).

🛠 Using Prior Results. In the partnership framework, the ADEPT project will offer to KnowVis the XML Model of the scientific concept for KB creation (developed by T. Smith and the ADEPT team and modified by O. Agapova). Also, our team will have the opportunity to choose among 1000 short records of geography concepts and 30 completed records of scientific concepts to begin the KnowVis KB population.

6.2.3 Collection of DL documents for the experiments will consist of 1400 documents residing on the project's server.

🛠 Using Prior Results**.** The DL documents will be selected from repositories of ADEPT, NSDL, and DLESE (provided by projects liaisons D. Fulker, M. Marlino, T. Smith). In case there's a lack of documents to introduce chosen concepts, documents from the ChemDiscovery learning environment will be used (provided by ChemDiscovery developers O. Agapova and A. Ushakov).

6.2.4 Experimental online software client interfaces will be developed. They include several visualization options for sub-query and consolidation map interfaces. The UI will support each experimental condition (part 8.5). Our attention will focus on a reasonable combination of ADEPT OM and Thinking map representations and their modification.

🛠 Using Prior Results**.** The ADEPT project will offer several online software packages to KnowVis—the Lecture Composer and graphical software called Grapher. The Grapher implements the four-spring algorithm and models a concept space in 2D constructions of nodes (squares with a concept name) and dynamic links among them. However, Grapher makes no distinction between relationship types and depicts all relationships as a plain link.

6.3 Equipment. The computer hardware and equipment employed for conducting the evaluation workshops will closely resemble the environment that can be found in today's school computer labs.

6.4 Pre-test. All participants will be required to complete a demographic survey, answer a series of questions, and take the pre-tests to produce the data about individual user differences for comparative research. Qualitative data will include: The Level of web and DL expertise (WE), which are defined via self-assessment and personal specifics such as: *(a) Information Management Behavior (IMB)* – Czerwinski and al. have indicated ('99) that most users adopt a knowledge organization based on semantic categories. However, some users in their study augmented this organization with temporal or alphabetical cues. To determine the IMB type, our participants will be given a test similar to the one described by Czerwinski.

(b) Knowledge Structure Abilities (KSA) – all participants will be given the topic "Water" and will be asked to draw a concept map that shows related

concepts (within 30 min). Those maps will be compared among the participants and also with the knowledge structures that were developed by the same participants during Condition III.

<u>(c) Spatial and (d) Verbal Abilities (SA, VA)</u> – participants will be given Woodcock-Johnson tests of (i) cognitive abilities, (ii) spatial relations and (iii) verbal comprehension and also paper-folding tests (VZ-2).

After the pre-test data is analyzed by the evaluation team and experts, the participants will be categorized for better interpretation of the experimental progressive variables. We expect such categories of participants as: (a) WE: high, medium, low; (b) IMB: semantic clustering, networking, alphabetical, and new type; (b) SA: high, medium, low and (c) VA: high, medium, low. Expected also is that some correlation exists between the individual categorizations mentioned above and the ability to work with visual IR representations. All individual characteristics will be taken into account in the final metrics to make the project evaluation more objective .

Table 2. Experimental Conditions

Presentations	I. Controlled Condition	II Condition	III Condition	IV Condition	V Condition:
Query Format	Fill-in-form format	Thinking map-like	Thinking map-like	Alphabetical menu	Users self-determine both Presentations
Search Results	Plain list of references	List grouped by clusters	Thinking map-like	List grouped by clusters	

6.5 Experimental Conditions. We will employ five experimental conditions (Table 2). All participants will use the same conditions, instruments, topics, and tasks to provide some generality of research and evaluation results. Individuals will be randomly assigned to one of the 24 different orders of conditions.

6.6 Evaluation Experiments and Measurements

We plan to conduct a careful examination of individuals' vision of IR, their interaction with the KnowViz prototype, and participant outcomes. User feedback will be incorporated into the KnowVis design practices.

6.6.1 Experiment 1: IR from the Users' Perspective

<u>a) Defining the semantic types of relationships (STR) by users.</u> Participants will conduct a series of tests in which they are required to define an STR among the given concepts and provide a name for each type.

Instruments: A set of eight exercises for defining the STR between concepts within the topic "Water" (one for each STR and one combined exercise).

Measures: i) STR suggested by the participant; ii) Similarity and differences between the STR produced by the participants and KnowVis.

Variables: Quantities and types of new STR.

Predictions: We expect new STRs, not previously listed in the KnowViz hypothetical model

<u>b) Comparing Visual Patterns of the STR.</u> Participants will be asked to connect the concepts listed in experiment 1.a. This test will indicate the correlation between visual KnowVis STR presentations and how individuals see those visual patterns. Additionally, a special memory test will take place to find the optimal number of nodes and relationships that our users can perceive in one glance .

Instruments. A set of eight exercises for drawing the relationships among the concepts within the topic "Water."

Measures: (i) types of the graphical patterns; (ii) similarity and differences in relation to the Thinking Maps representation.

Variables: Number and type of new patterns.

Predictions: We expect to find differences in the visual perception of participants in comparison with Thinking Maps.

<u>c) Human Operations during the exploration of IR lists.</u> Participants will be asked to explain the mental and physical operations that they use for IR list analysis and select the appropriate documents during the Controlled Condition I.

Instruments: Sets of 5, 20, and 50 Web pages with lists of titles and abstracts will be collected for the topic "Water."

Variables: Number of pages that were explored by the user in each set, and the cognitive patterns that the user applied in order to analyze the sets.

Measures: (i) types of user operations; (ii) similarity and differences among the cognitive patterns that users are used and thinking maps are proposed (Table 1).

Predictions: We expect similarity with cognitive operations during meaningful reading.

6.6.2 Experiment 2: KnowVis Prototype Evaluation

Evaluation of the KnowVis interface will be based on a collection of specific Qualitative and Quantitative systematic data for following evaluation categories:

Effectiveness – accuracy and completeness with which users achieve task goals;

Efficiency – the amount of time spent to achieve task goals;

Subjective satisfaction – positive attitudes toward the use of KnowVis. Here are some examples of questions: Did it go as expected after we explained the purpose of KnowVis? If you were to design KnowVis, would you plan anything differently? Users also will be asked to explain what they were thinking during the sessions, and what they liked and disliked.

Tasks A. A variety of tasks will be applied in order to provide some generality of evaluation results. They include location, skimming, and comparing discovered documents within the topic “Water” and their sub-topics. The research and statistical methods, which were described in (Reiterer ‘00), will be used.

Usability test will apply a form of self-assessment (Newby ’02) with the focus on a group of three questions: (a) To what extent do you think you understand how to use KnowViz? (b) To what extent do you think the KnowVis interface may be useful? (c) What are some of the things you think KnowVis might be useful for?

Used Conditions. The first three evaluation categories will be investigated during the implementation of conditions 1-V; Condition I is controlled (8.7.2) in those series. The usability test will apply only during condition V.

Interoperability test will take place to control the compatibility with NSDL, ADL, and DLESE portals.

6.6.3 Experiment 3: Evaluation of Participants' Outcomes. Some aspects are usually not measured during IR system evaluation. For example, “Can information IR visualization provide a better overview of a discipline domain than a text-based system?” (Newby ‘02) To begin to investigate that area and to measure participants’ outcomes, we will design additional tasks for this part of the evaluation.

Tasks B. The tasks will simulate basic activities that are often included in educational practice: (a) *for students* – homework on literature review for scientific projects; and (b) *for teachers* – lesson preparation. However, teachers will be asked to create both literature reports and lesson plans. Teachers’ reports will be used as control documents, which will be compared with students’ outcomes.

7. Relevant Research

7.1 Research on Visualizing Implicit Queries for Information Management and Retrieval.

A 1998 survey (Kehoe ’98) of over 10,000 web users and Abrams’s 322 web users survey (Abrams ‘98) revealed that one of the most common problem users have with the web is organizing the information they gather. Czerwinski (’99) presents a successful visual tool, called Data Mountain, which allows users to manually create a spatial layout of thumbnails of their documents in a 3D environment.

7.2 Research on Memory and Cognitive Factors of the IR and Visualization.

Without landmarks, paths or cues, users are inclined to become disoriented in virtual IR worlds (Horvitz ’98). However, little has been done to make a link between human cognitive processes of knowledge understanding and the processes of navigation through the Web-knowledge. There have been a number of works exploring the role of cognitive processes and memory lines in navigation through large information space (Chennawasin ’99, Chen ’97, Chun ’98, Leplow ’98, Mallot ’98). Most of the research was focused on associative memory and special abilities. For example, Chennawasin (’99) studied the relationships between associative memory and the use of a spatial user interface. Findings indicate that a key design element of an effective and usable 3D interface relies on human memory factors being taken into account. Swan and Allan (’98) tested users’ verbal fluency and spatial ability. Darken and Sibert (’93, ’96) discovered that people tend to take advantage of environment cues, which make it easier to locate search targets.

Table 3. Visualization Techniques

Arrangements	2D	3D	Hyperbolic
Graph	TileBars, Tking	Microsoft Office, GT-VGIS, StarWalker,	
Tree	Cluster tree, Mondesa’s Navigator, Techquila: TM4J, Mind Map from Axon Idea, Kartoo, PDQ Tree-Browser	Cone, UNIVIT, Narcissus, CoBrow	Sphere, Fish-eye H3/H3Niewer libraries, Site Manager, Library of Congress,

Map	Tree-Map, MDS-Map, BEAD Themescape Map, WEBSOM Self-Organizing Map, ET-Map		
Metaphors		Active Worlds, Virtual-City, Landscapes, Perspective Walls, Rooms, VIBE, WebBook	
Combined	LifeLines, Storyspace, Squirrel	CAT-A-Cone, Butterfly	

7.3 Visualization Techniques

There are significant amounts of visualization tools for representing different data resources including IR results (Spence '01, Kobayashi '00, Newby '02, Reiterer '00, Sebrechts '99, Cugini '00). The most important tools, including examples of software, are summarized in Table 4. However, there was no single recommendation of the use of hyperbolic over 3D over 2D over text, but instead a series of situations in which each type of the system might be the best (Newby '02). Other research has shown (Le Grand '02, '03) that among graphs, trees, and maps representations, trees are the ones that are most understandable by users. Moreover, trees traditionally are used for educational concept map representation. They are habitual and recognizable.

7.4 Related Ontological approaches. The idea of the KnowViz ontological model of the scientific concept originated from the Chemical Abstracts Service (CAS '01) models, the MatML Working Group of the National Institute of Standards and Technology (NIST '01), and Strongly Structured Model (SSM) for scientific concepts that was developed at the ADEPT project (Smith a-b). All three models have been successfully used.

7.4.1 American Chemical Society developed the CAS model for the largest and most comprehensive databases of chemical information, including Chemical abstracts (CA) and Registry. CAS databases include more than 22 million abstracts of chemistry-related literature and patents. CAS indexes and abstracts patents, articles from approximately 9,000 scientific journals, conference proceedings, and other documents. The CAS ontological model describes the substance and are expressed by: *Name* – systematic, generic, proprietary or trade, trivial; *Description* – CA abstract text; and *Representations* – molecular formulas, illustrative structural diagrams; index of Ring Systems.

7.4.2 NIST ontological model looks wider, represents the materials, and embraces Name, Class, Specification - Source, Form, Processing, Geometry, Characterization, Property, and Associations.

7.4.4 Differences. Despite the visual resemblance, the ADEPT ontological model of scientific concept differs from the NIST and CAS models by diverse integration. Our model is not dependent upon any single knowledge classification approach. It represents each individual concept (can be scientific concept, principal, substance, material, and et. al) from the perspectives of (a) belonging to, (b) position in, and (c) describing via several knowledge organizational approaches (Sowa '00; Ranganathan '78), such as glossaries, dictionary, subject headings, scientific categorization, classification, taxonomies), and thesauri. The ADEPT concept model differs from the CAS and NIST models by showing similarity with the organization of human semantic .

8. References

Hill, L. L., Buchel, O., Janee, G., & Zeng, M. L. (2002, November 17). Integration of Knowledge Organization Systems into Digital Library Architectures: Position Paper. Paper presented at the Thirteenth ASIS&T SIG/CR Workshop on "Reconceptutalizing Classification Research", Philadelphia, PA. http://alexandria.sdc.ucsb.edu/~lhill/paper_drafts/KOSpaper7-2-final.doc.

T. Smith, D. Ancona, O. Buchel, M. Freeston, W. Heller, R. Nottrott, T. Tierney and A. Ushakov. The ADEPT Concept-based Digital Learning Environment. *Proceedings of the Seventh European Conference on Research and Advanced Technology for Digital Libraries (ECDL)* (Trondheim, Norway; August 17-22, 2003).

Agapova O., Mayer R., Smith T., Ushakov A. S., Decker Stefan, Ushakov A.A., (2003), "Using Digital Library Visual Services for Building *a* Lesson-Design Environment Prototype", RCDL, Saint-Petersburg, October 29-31, 2003.

Ranganathan S. (1978). Prolegomena to library classification. Bombey: Asia Publ. House. 640 p.

Sowa J.F. (2000). Knowledge Representation. Brooks/Cole. Pacific Grove. Pp 417-427, 476-489.

Barfourosh A.,A., Motahary Nezhad H.R., Anderson M. L., Perlis D., (2002), Information Retrieval on the World Wide Web and Active Logic: A Survey and Problem Definition, http://citeseer.ist.psu.edu/correct/535484

Steyvers, M., & Tenenbaum, J. (submitted). Graph Theoretic Analyses of Semantic Networks: Small Worlds in Semantic Networks. *Cognitive Science*

Novak J. (1998). Learning, Creating, and Using Knowledge. Lawrence Erlbaum Associates, Publisher. New Jersey. 19-35

Anderson, L. et al. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* New York: Longman

Hyerle D. and Thommasina D. P., (2000). Thinking Maps: The Cognitive Bridge to Literacy: A Visual Language for Bridging Reading Text Structures to Writing Prompts, Designs for thinking. Alexandria, VA: ASCD

Zhang J., (2002), Information Tryrieval by Semantic Analisis and Visualization of the Concept Space of D-Lib Magazine, Ovtober, V.8, N. 10

Veerasamy A., Navathe S., (1995), Querying, Navigating and Visualizing a Digital Library Catalog, Proceedings of the Second Annual Conference on the Theory and Practice of Digital Libraries.

Cook C., Heath Fred and all (2003) Developing a National Science Digital Library (NSDL) LibQUAL+TM Protocol: An E-service for Assessing the Library of the 21 Century, NSDL Evaluation workshop.

Tomita J., Kikui G. Interactive, (2000), Web search by graphical query refinement, IEEE Symposium on Visual Languages, Seattle, Washington, September 10-14,http://web.engr.oregonstate.edu/~burnett/vl2000/schedule/session2.html

The *KnowVis Mock Up:*
http://www.alexandria.ucsb.edu/~ushakov/knowvis/