Developing Visualization Techniques for Semantics-based Information Networks

Richard M. Keller
Computational Sciences Division
NASA Ames Research Center
Moffett Field, CA 94035-1000
rkeller@arc.nasa.gov

David R. Hall
QSS Group, Inc.
NASA Ames Research Center
Moffett Field, CA 94035-1000
dhall@arc.nasa.gov

ABSTRACT

Information systems incorporating complex network-structured information spaces with a semantic underpinning – such as hypermedia networks, semantic networks, topic maps, and concept maps – are being deployed to solve some of NASA's critical information management problems. This paper describes some of the human interaction and navigation problems associated with complex semantic information spaces and describes a set of new visual interface approaches to address these problems. A key strategy is to leverage semantic knowledge represented within these information spaces to construct abstractions and views that will be meaningful to the human user. Human-computer interaction methodologies will guide the development and evaluation of these approaches, which will benefit deployed NASA systems and also apply to information systems based on the emerging Semantic Web.

Categories and Subject Descriptors

H.5.4 Hypertext/Hypermedia – *Navigation*

H.5.2 User Interfaces – *Graphical user interfaces (GUI)*

I.2.4 Knowledge Representation Formalisms and Methods – *semantic networks*

H.3.4 Systems and Software – *Information networks*

Keywords

Semantic network visualization

MOTIVATION AND OBJECTIVES

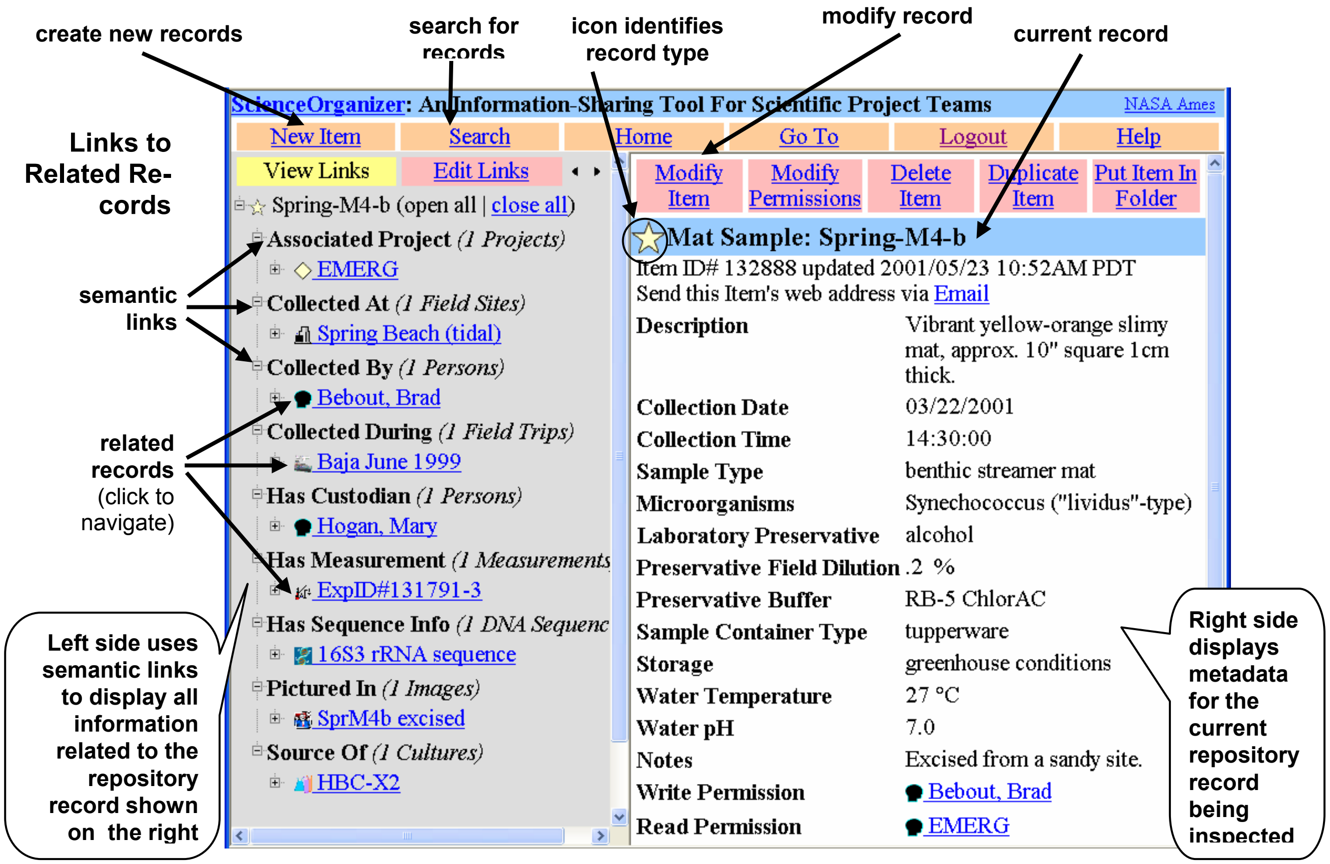
The advantages of network-structured information storage and retrieval systems, such as hypertext and hypermedia systems [11], semantic nets [6], concept maps [12], and topic maps [9] are becoming increasingly important to NASA. NASA has deployed knowledge management and educational tools featuring networked-structured information spaces to support various missions. For example, the Center for Mars Exploration at NASA Ames has applied concept mapping to develop landing site selection criteria for Mars missions, and to describe the science of astrobiology for educational outreach purposes [3]. In addition, we have developed and deployed SemanticOrganizer [7], a semantics-based collaborative knowledge management system designed to support the information management needs of distributed teams of NASA scientists and engineers. This paper presents some observations and research ideas relating to our experience developing and supporting SemanticOrganizer.

A Brief Overview of SemanticOrganizer

SemanticOrganizer consists of a structured semantic hypermedia repository of typed, customizable data records. Each record in the repository characterizes a concrete or conceptual item relevant to a project team (e.g., a specific person, place, document, meeting, event, etc.). A record includes a set of metadata properties and optionally, a file containing an image, dataset, document, or other relevant electronic product. The records are extensively cross-linked via semantically labeled relations to permit easy access to interrelated pieces of information. A master ontology describes different types of data records and defines links that can be used to express relationships between records. The system supports over 300 registered users within 35 different work teams. SemanticOrganizer's network-structured repository currently contains over 25,000 nodes and over 160,000 links.

SemanticOrganizer users enter and interlink data records and files using a specialized Web interface that enables them to navigate through the network of records, view metadata and files, and search for specific records (see Figure 1). A permission management system limits user access to a defined subnet within the information space that contains information relevant to their project team. Even with this limitation, teams with heavy usage have access to large subsets of the network. Unfortunately, SemanticOrganizer's primary user interface presents a highly localized view of the information space; the interface displays the details of a single focal node along with a listing of traversable links to nodes immediately adjacent in the space. While this

Copyright 2003 Association for Computing Machinery. ACM acknowledges that this contribution was authored or co-authored by a contractor or affiliate of the U.S. Government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only..
K-CAP'03, October 23-25, 2003, Sanibel Island, FL, USA.
Copyright 2003 ACM 1-58113-000-0/00/0000…$5.00



Spring-M4-b (open all | close all)
- Associated Project *(1 Projects)*
  - EMERG
- Collected At *(1 Field Sites)*
  - Spring Beach (tidal)
- Collected By *(1 Persons)*
  - Bebout, Brad
- Collected During *(1 Field Trips)*
  - Baja June 1999
- Has Custodian *(1 Persons)*
  - Hogan, Mary
- Has Measurement *(1 Measurements*
  - ExpID#131791-3
- Has Sequence Info *(1 DNA Sequenc*
  - 16S3 rRNA sequence
- Pictured In *(1 Images)*
  - SprM4b excised
- Source Of *(1 Cultures)*
  - HBC-X2

Mat Sample: Spring-M4-b

Item ID# 132888 updated 2001/05/23 10:52AM PDT

Send this Item's web address via Email

Description	Vibrant yellow-orange slimy mat, approx. 10" square 1cm thick.
Collection Date	03/22/2001
Collection Time	14:30:00
Sample Type	benthic streamer mat
Microorganisms	Synechococcus ("lividus"-type)
Laboratory Preservative	alcohol
Preservative Field Dilution	.2 %
Preservative Buffer	RB-5 ChlorAC
Sample Container Type	tupperware
Storage	greenhouse conditions
Water Temperature	27 °C
Water pH	7.0
Notes	Excised from a sandy site.
Write Permission	Bebout, Brad
Read Permission	EMERG

Figure 1: SemanticOrganizer interface displays current record being browsed on the right and semantic links to related records on the left. The interface provides a highly local view of the information space and requires users to use search facilities to move to a non-local node in the space.

interface is appropriate to targeted network viewing and editing, the interface provides no graphical overview for purposes of non-local navigation. Secondary interfaces do provide overview displays for specialized tree-structured and sequenced nodes in the space, but these are suitable for viewing only a small fraction of the nodes in the space. We feel that SemanticOrganizer provides an excellent live testbed for developing and evaluating visualization and interaction techniques to overcome some of the difficulties of interacting with large semantic network topologies.

Difficulties of Network Topologies

Although networks are effective structures for storing and accessing interconnected information, developing effective techniques for interacting with network structures presents distinct human interface design challenges. Three essential tasks must be supported by interfaces to network-structured information systems: *browsing* (exploring within a local network "neighborhood" by traversing links), *searching* (locating a specific node located within the network by issuing a query), and *teleporting* (moving directly from the current neighborhood to a distant neighborhood). Supporting these interface tasks presents few problems when interacting with networks consisting of a handful of nodes. But it becomes very difficult to develop effective interface strategies that apply to larger and more complex network topologies. Consider, for example, the following key difficulties:

- *Lack of distinguishing structural properties*: Network topologies are homogenous, undifferentiated, and lack the kind of distinctive structural characteristics necessary to cognitively orient users within the overall information space. Depicted graphically, any neighborhood in the network appears visually like any other. Further, with highly connected networks, there are so many navigation paths that users quickly get disoriented and become "lost in hyperspace" [4].
- *Lack of suitable scoping facilities*: The ability to define subsets, subgraphs, or partitions of the network that effectively scope user operations is essential for an effective interface. The simple scoping schemes designed for most network information spaces prevent users from carrying out precision scoping. This leads to inefficiencies in user interaction and can result, for

example, in poor information recall when searching the network.

- *Scaling effects*: Simple graphical layout techniques that are computationally tractable for small networks do not scale to hundreds or thousands of nodes and interconnections. Moreover, it is difficult to follow links when they cross in huge interconnected structures. As a result, it is difficult to provide the kind of simple information space overview that is necessary for effective teleporting [14].

Without effective interface strategies for dealing with the difficulties described above, networked information systems become difficult and – at worst – impossible for humans to use. We plan to study browsing, searching, and teleporting in the context of the SemanticOrganizer system and to develop interface approaches that address these difficulties.

APPROACHES TO VISUALIZATION

Our approaches to addressing challenges associated with interaction and visualization of large semantics-based information networks are summarized in the following subsections. We expect these approaches to be applicable to SemanticOragnizer, in particular, but also to systems with similar overall characteristics, including the Semantic Web, concept maps, topic maps, and semantic nets.

Exploit maximal knowledge to inform the visualization process

Two main sources of information have been used by previous researchers to generate effective visualizations for network-structured information spaces: graph-theoretic knowledge (e.g., topological properties of the network) and information-theoretic knowledge (e.g., word-sense and meaning of content stored in nodes of the network). In studying SemanticOrganizer, a third powerful source of information becomes available that we can highly leverage – semantic knowledge in the form of models and ontologies. This sort of knowledge describes the types of information stored in the network, the interrelationships among information, and the user's information tasks and goals. Metaknowledge of this type can be very valuable in developing interaction methods that are appropriate to the use context. Aside from graph-theoretic, information-theoretic and semantic knowledge, we believe it is essential to incorporate direct end-user input and feedback in design and evaluation of visualization techniques.

Develop various methods of abstracting and filtering the information space

Reducing the apparent size of the information space is the key to dealing with scaling problems and can also make the structure of the space more apparent. This reduction can be achieved by developing automated, multi-level filtering and abstraction techniques to hide network nodes and links. These techniques can be informed by model-based, graph-theoretic, and content-based properties of the elements of the network.

With filtering, the approach is to reduce the number of network elements by defining filter criteria and ignoring the items that don't match those criteria. Examples of criteria include manifest node properties such as name, type, creation date, and creator. Computed properties of nodes, such as last reference date, access frequency, etc., can be useful filters for some purposes. Filters can also be defined on links, for example to ignore certain links based on the semantic relevance of those links in a particular task context. Access permissions also can be effective filters, for example to display the nodes that are jointly visible to a set of projects. An adequate user interface must provide facilities for users to define filters on their view of the navigation space and must also provide predefined filters and the ability to dynamically change filter parameters while navigating.

Reducing the information space by constructing abstract nodes using various grouping strategies, including semantic, content-based as well as graph-theoretic, has been an area of interest to several researchers for some time. Botafogo and Shneiderman [2] built aggregate structures based on simple properties of the connected structure of a network; Pirolli [13] exploited metainformation and reference frequency to determine groups; Zhang and Mostafa [15] used properties of the underlying content to group similar items. This body of research can be extended by focusing on abstraction criteria that exploit knowledge of the task context or the information semantics in the underlying space.

Develop flexible scoping methods

All user interactions defined on network structures (e.g., searching, editing, navigating) require a scope of operation. A scope consists of a subset of nodes and links drawn from the overall network. For example, we can use the notion of a scope to identify a connected subgraph constituting a "neighborhood". The user's ability to define a scope that is natural and efficient for the task at hand is essential to effective interaction. Different automated scoping approaches can be developed that utilize model-based, graph-theoretic, and content-specific methods of defining node/link subsets. To support user definition of scope, tools can be developed that allow users to identify subsets by defining generative rules or constraints. Individual users should also be able to define scoping subsets manually for their own purposes. A facility to create, maintain, and use such defined scopes within an interactive graphical navigation tool is an important part of an interface to a network-structured space.

Develop a notion of "semantic distance"

The concept of a "semantic" distance between two nodes is critical to developing user-centered graph layout, navigation, and abstraction techniques. We believe that the user's concept of distance is related to a mental model of the interaction space. We propose to construct distance (i.e., edge weight) measures that reflect mental models based on the semantics of the information space, the information content, and the task being performed. Differing semantic distance measures can be used in conjunction with different abstraction and scoping methods, and can be tuned based on user feedback. In addition, semantic distance can be applied to developing visualization and navigation methods that facilitate user interaction and highlight the focus of a display. For example, distortions applied to the graphical display of networks – including fisheye view [5] and hyperbolic transformations [10] – can show more detail in the area of focus at the expense of items nearer the horizon of interest. These techniques require that geometrical distance in the displayed graph correlate well with the semantic distance.

Exploit substructure

Although the node and link structure of a network appears fairly homogeneous, substructures within the network manifest themselves as patterns that can be identified and visualized distinctively. For example, hierarchical structures, list structures, cross-correlated structures, and time-based structures have well-established visual clichés in the form of trees, single and multi-dimensional tables, and PERT charts. It should be possible to identify substructures using pattern matching techniques, encapsulate substructures using abstraction, and then display them using heterogeneous, substructure-appropriate visualizations.

When substructure is absent, there are various ways to add distinguishing structure to a network. Some of the simplest include providing a facility for expert users to identify, in a persistent manner, nodes and paths that are important for a body of users. Paths could include sequences of direct or macro-links that would allow a user to teleport to other areas of the hyperspace. Graph-theoretic measures can be used to identify distinguished nodes automatically [2,8] or to assist the expert human in this identification task.

Parts of the network that remain without recognized or induced substructure can still be presented in ways that enhance, rather than thwart, a users understanding of the underlying structure. Researchers have long been interested in aesthetic attributes of effective graphical presentations. These attributes include the obvious advice of limiting the number of items displayed and the number of edge crossings. They also extend to manifesting symmetry and other regularities.

FUTURE PLANS

SemanticOrganizer provides an ideal test bed from a human-computer interaction perspective because there is a challenging visualization problem as well as a history of use and an accessible, established user base. We plan to study the appropriateness of the visualization approaches outlined in the previous section for users performing selected interaction tasks in the context of a highly interconnected, semantically linked, collaborative knowledge structure. In conjunction with those studies we will implement new graphical interfaces to improve users' capability to efficiently understand, access, and navigate these networks.

As part these plans, we intend to lay the proper groundwork by adding a graphical navigation workbench to the SemanticOrganizer environment. This workbench will include a visual, interactive browser with graphical layout techniques for handling various graph types. We will include interactive facilities for filtering nodes and links, and for showing and manipulating node and subgraph abstractions. Using this workbench, we will investigate, develop, and evaluate graph-theoretic, information-theoretic, and model-based filtering, abstraction and semantic distance metric strategies. These strategies will undergo human factors evaluations to determine their effectiveness

We expect the resulting tools to have direct immediate impact in improving the effectiveness of teams of scientists and engineers now using SemanticOrganizer. The applicability of these results should extend to other NASA-relevant information systems, including various concept map applications. Finally, these results will have wider applicability to future visual interfaces for interacting with the emerging Semantic Web.

ACKNOWLEDGMENTS

Work on SemanticOrganizer is being funded by the NASA Intelligent Systems Project of the Computing, Information, and Communications Technology Program.

REFERENCES

[1] Berners-Lee, T., Hendler, J. and Lassila, O., The semantic web, Scientific American, May 2001.

[2] Botafogo, R.A. and Schneiderman, B.. Indentifying aggregates in hypertext structures. in Third ACM Conference on Hypertext. (1991, San Antonio, TX). ACM. pp. 63-74.

[3] Briggs, G., CMEX Concept Maps, http://cmex-www.arc.nasa.gov/CMEX.

[4] Edwards, D. M. and Hardman, L., "Lost In Hyperspace": Cognitive Mapping and Navigation in a Hypertext Environment, Chapter 7, Hypertext: Theory Into Practice, Edited by McAleese, R., Ablex Publishing Corporation, 1989.

[5] Fairchild, K., Poltrok, S. and Furnas, G. `Semnet: Three-dimensional graphic representations of large knowledge bases' in Cognitive Science and its Applications for Human-Computer Interaction, R. Guindon, Ed. Lawrence Erlbaum, 1988.

[6] Findler, N.V.(ed.), Associative Networks - Representation and use of knowledge by computers. Academic Press, New York, 1979.

[7] Keller, R.M., SemanticOrganizer Knowledge Management System, http://sciencedesk.arc.nasa.gov.

[8] Kleinberg, J. M., Authoritative Sources in a Hyperlinked Environment. Journal of the ACM Volume 46, Issue 5 (1999).

[9] Lacher, M.S., and Decker, S., "RDF, topic maps, and the semantic Web," Markup Languages: Theory & Practice, vol. 3, pp. 313-31, 2001.

[10] Lamping, J.. and Rao, R.: The Hyperbolic Browser: A Focus+Context Technique for Visualizing Large Hierarchies, Proceedings of CHI'95, ACM Conference on Human Factors in Computing Systems, New York, pp. 401-408, 1995.

[11] Nielsen, J. (1990). Hypertext and Hypermedia. Academic Press, Boston.

[12] Novak, J., and Gowin, D. (1984). Learning How to Learn. New York: Cambridge University Press.

[13] Pirolli, P., Pitkow, J., Rao, R., "Silk from sow's ear: Extracting usable structures from the web," in Proc. ACM Conf. Human Factors in Computing Systems, CHI. 1996, ACM Press.

[14] Schneiderman, B.. Direct manipulation: a step beyond programming languages. IEEE Computer. Aug. 1983.

[15] Zhang, J. and Mostafa, M.. Information retrieval by semantic analysis and visualization of the concept space of D-Lib Magazine. D-Lib Magazine, 8:11, (2002)