Ontology Models for Supporting Exploratory Information Needs

Maria Lee* and Riichiro Mizoguchi**

* Research Data Networks Cooperative Research Centre
CSIRO Mathematical and Information Sciences
Locked Bag 17, North Ryde, Australia
Fax: +61-2-9325-3101, E-mail: maria.lee@cmis.csiro.au

** The Institute of Scientific and Industrial Research, Osaka University,
8-1, Mihogaoka, Ibaraki, 567, Japan. E-mail: miz@ei.sanken.osaka-u.ac.jp

Abstract

Exploratory browsing is potentially an important technique for retrieving text information from large knowledge bases. However, the task of information seeking is complex and it is easy to get lost in a composite network of nodes representing concepts. In this paper, we show how an exploring capability can be supported by ontology-based knowledge. The disadvantages mentioned above are avoided by providing facilities for guiding users' exploratory tasks particularly if they are not experts in the domain. We offer a way for information seekers or explorers to pursue information objects relevant to their tasks and apply task action steps to achieve their goals and intentions. Three types of ontology-based models were proposed that structure and organise information to support information access. The retrieved information is designed for navigating users' information seeking, but not restricting users' options. In addition, we discuss the maintenance of users' task actions to support information re-use purposes.

1. INTRODUCTION

The World Wide Web can be argued as being designed to support exploratory (browsing) behaviour. However, exploring can be a time-consuming task as soon as the number of potentially relevant pages become large. When exploring, users may have difficulty in addressing their needs if they are not domain experts, e.g. when browsing government information. Under such circumstances, a user might wish to organise concepts into a cluster and provide cross-linked references to complete a particular task to explore related information (Smith, Newman, and Parks, 1997). Therefore, it requires the user to find related items of information within an information structure.

However, it is difficult for users to cluster concepts if they are not the experts in the area of interest. One of the most frequently encountered problems by a user is the difficulty in querying information distributed across multiple sources using familiar terminology. Research has shown that the probability of matching between the terms in the domain and the vocabulary used by the user is <0.2 (Furnas *et al*, 1987).

In order to address these problems in part, an attempt has been made to simulate exploring behaviour by matching users' exploratory tasks with the necessary information required to carry out these tasks. The retrieved information therefore can be designed to guide users to avoid getting lost in a complex network of nodes representing concepts. For example, if a user is exploring information about an "airport", the necessary input required for such information can be supported by showing the transportation route, destination, frequency of departures, fare and time required. If a user is exploring the information for entering "educational organisations", then pre-requisite and application procedures can provide users with the necessary information for entering schools and therefore guiding the exploratory task.

As the Web expands exponentially in size, the lack of organisation makes it very difficult to efficiently glean knowledge from the Web. In order to provide navigation aids to assist users' information exploratory tasks on the Web, it is important to provide semantic structure of the information environments. Ontology has been proposed to be used in various areas: knowledge-based systems, information retrieval, standardisation, etc (Lenat and Guha, 1990, Neches *et al*, 1991, Gaines, 1997a). Ontology has been viewed as general organisational devices for the specification of conceptual knowledge. Ontology provides the foundation to explicitly represent concepts with their relationships (Gruber, 1992, Mizoguchi, 1993). Top-level categories can be a helpful guide in decomposing a complex information problem and fashioning a comprehensible and effective information access method (Sowa 1995 and 1997). As the need for using more than one type of ontology has been necessitated by the increasing distributed nature of information (Amba *et al*, 1996), we have proposed three types of ontology-based models to suit classification and retrieval purposes.

2. ONTOLOGY-BASED MODELS

Our efforts have focused on the use of ontology to support information needs in exploring government information. Scenarios that arise in the applications motivate the development of ontology. Generally speaking, most information explorers are interested in three types of information when exploring government contents: organisational structure, business functions, and human activities. In this paper, we concentrate on defining organisational structure and business activities of most government organisations.

Two types of users are involved in the development of a computer system, an Intranet: Ontology Engineers/Record Managers and end-users. The prime responsibility of the Ontology Engineers/Record Managers is to provide control and consistency over the vocabulary used for classification and indexing records. Classification requires an understanding of the context and content of the record and a knowledge of existing records in the system and indexing includes attaching one or more searchable terms to each recorded item, such as a file of a paper or electronic document. This facilitates the end-users' ease of searching and retrieval of records. Based on the information needs, three types of ontology-based models are proposed: organisational ontology, domain activity ontology and user activity ontology. The organisational and domain activity methods are for indexing and titling purposes whereas user activity ontology facilitates searching and retrieval purposes.

2.1 Organisational Ontology

The organisational ontology provides semantic information of government structure. Complex structure is often resolved by decomposition the structure into hierarchical units in a computer system. Figure 1 shows a hierarchical tree for government organisational ontology. The tree represents the concepts of entities. For example, the government department consists of state departments and federal departments. The state departments include ACT, New South Wales, Northern Territory, etc. The federal departments include Attorney-General Department, Department of Defence, Department of Health and Family Services, etc. Each entity links to correspondent relationships and attributes. For example, the attribute for employee is employee name and the employment relation defines a term as work-for.

The organisational ontology is able to answer sophisticated knowledge-based queries, such as finding "Wood's supervisor's telephone number". Searching existing web indices for "Wood" may yield to thousands of pages about wood in forest. In this example, the ontology would include attribute like "surname", classification like "person", and relationships link "employee".

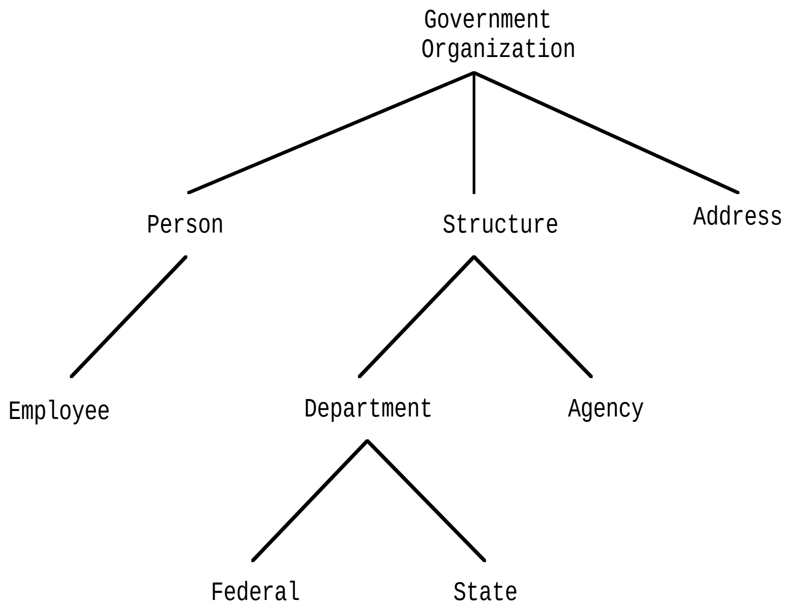

Figure 1. A Hierarchy Tree for Organisational Ontology.

2.2 Domain Activity Ontology

Domain activity ontology is designed to organise and structure business functions and activities taking place in a domain. Unlike the traditional library classification, where books are classified according to subjects, the domain activity ontology provides a hierarchical structure to classify records according to the business functions and activities documented by the records for classifying and indexing purposes. Thus a sequence of business activities may be followed using a group of records which have been classified together.

The development of domain activity ontology is based on CSIRO Keyword AAA (CSIRO 1996), which is an alphabetical list of terms developed for use with the document management for Australia commonwealth government. Keyword AAA covers general terminology, which describes terms common to the business functions and activities of most organisations. It is used in conjunction with functional terminology, which describes unique terms to the individual organisation's business functions and activities. Bringing the general and functional terminology in one domain activity ontology provides comprehensive coverage.

The components of the domain activity ontology include keywords (KW), activity descriptor (AD) and subject descriptor (SD). Keywords are usually presented by nouns or noun phrases, e.g. accident, asset management, community relations, etc. A keyword is followed generally by one activity descriptor, which reflects the activities performed within the function. Narrower terms of activity descriptors further define the subject content of the activities represented by the subject descriptors. For example,

1. (KW) (AD) (SD)
COMMUNITY RELATIONS -COMMITTEES - Olympic Games Celebrations

Example 1 shows that a number of records may be created for the Olympic Games Celebrations so the COMITTEES activity descriptor is used to highlight the activities undertaken by that particular committee.

2. (KW) (AD) (SD)
COMMUNITY RELATIONS - MEETINGS - Olympic Games Celebrations

Example 2 shows that the prime activity of this record is MEETINGS from the community relations functions of the organisation.

The role of domain activity ontology is to define records by their relationship to the activity, which they document. Records classified and indexed using the domain activity ontology therefore will have meaning and will be placed within the context in which they were created. This also means that all files on one function or activity are grouped together in the index.

2.3 User Activity Ontology

User activity ontology is related to information seeking activities carried out by the end-users. The task of information seeking is complex. From a knowledge based analysis point of view, we need to look at what users need to know about the information objects and actions involved in the information seeking task, and how that knowledge is organised. User's activities normally can be described by performing action tasks and information objects related to the information needs. For example, a user is looking for information in a music library. S/he may perform activities such as entering a collection, searching the index to a shelf and listening to a song.

The development of user activity ontology begins by building taxonomic classification of task knowledge and information objects. The task knowledge includes vocabulary for representing execution process of actions, e.g. entering, searching, listening, etc. The information object can be a CD, a tape, etc. The combination of task actions and information objects present a dynamic reasoning knowledge. For example, we can search for a CD or a tape, or we can listen to or sing a song.

Figure 2 shows a conceptual model for user activity ontology. *Task* represents users' information seeking actions. Generic verb or verb phrases are used to represent task-action vocabularies, which are domain-independent and can be handcrafted in a generic way.

A collection of generic verbs are used to describe user's information seeking tasks, such as:

- Access to: to reach, use or approach something.
- Apply: to formally ask for something.
- Immigrate: to come as a settler, not as a tourist or visitor

- Lodge: to place (a statement, etc) with the proper authorities.
- Seek: to look for; try to find something.
- Travel: to make a journey from place to place.

Information object represents a set of objects related to the information. From the end-user point of view, they may not be familiar with organisational structure and terminologies used for business functions and activities with the domain. Therefore, the information objects include a higher-level description of organisational and domain activity ontology, e.g. only include keywards and activity descirptors in domain activity ontology. A collection of nouns or nouns phrases are used to describe the information objects, such as:

- Health Services (e.g. aged, people with disability, family with children, etc)
- Public transportation (e.g. train, bus, air, or ship)
- Tax (e.g stamp duty, land tax, debits, tax, etc)
- Welfare (e.g. pension, special benefit, sole parent benefits, etc)

A combination of a generic verb and a noun such as "verb + noun" can be referred to as generic problem solving process (Mizoguchi and Vanwelkenhuysen, 1995, Mizoguchi *et al*, 1995, Ikeda *et al*, 1996). Such a combination facilitates the information seeking process.

The combination of task and information objects refer to user's information seeking activities, such as:

- Access to <u>public transportation (</u>e.g. train, bus, airports, etc)
- Apply <u>welfare</u> (e.g. pension, special benefits, sole parent benefits, etc)
- Lodge <u>tax</u> (e.g stamp duty, land tax, debits tax, etc)
- Travel <u>location (</u>e.g. Australia, Sydney, etc)
- Seek <u>health services (e.g </u>aged, people with disability, family with children, etc)

The above underlines indicate generic nouns for information object.

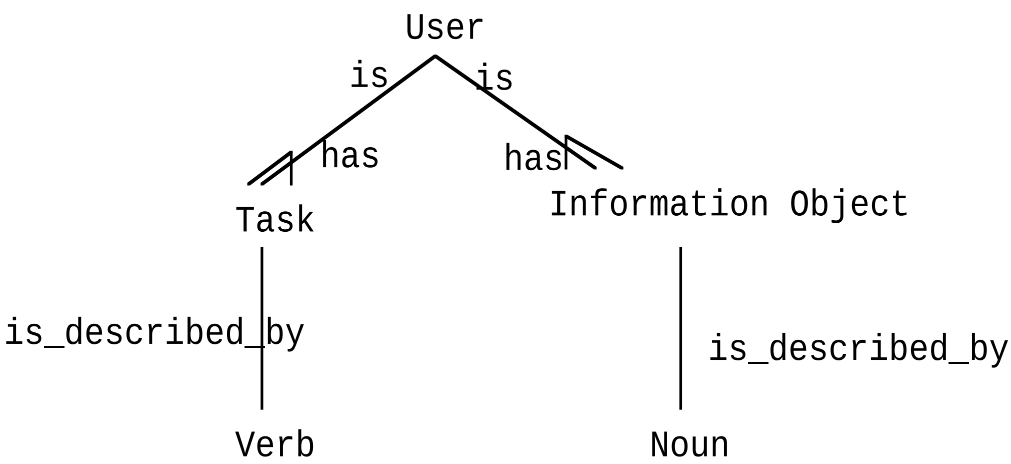

Figure 2. A conceptual model for a user activity ontology.

3. MATCHING TASKS AND NAVIGATION AIDS

Information explorers pursue information objects relevant to their information needs and apply task action steps to achieve their goals or intentions. It is desirable to design a computer system to match users' information needs and the type of tasks performed. In order to accomplish such an information seeking task, organisational and domain activity ontology are used for indexing and titling records where the end-users base on user activity ontology for retrieval purposes.

When a user starts an exploratory session, s/he selects the required task and information object within the user activity ontology for information seeking. The necessary information for carrying out the exploratory tasks is shown in a cluster. The retrieved information is used for navigating information seeking, but not restricting user options. The user can always ignore the information and go off in a direction of their own choice.

The type of task being performed by the user is domain independent whereas the necessary information required for these tasks is domain dependent. It is necessary to structure domain-dependent information for supporting the domain-independent task actions. The retrieved information therefore assists users for exploring information if they are not familiar with the domain. Table 1 shows sample examples to match users' exploratory tasks with the necessary information required to carry out the tasks. The task action column represents vocabularies for information seeking tasks. The information object column describes the domain activities. The information-guide column shows the relevant information required to carry out the search task. With a few mouse clicks, users can efficiently find the required information. It should be noted that the combination of task actions and information objects is not a one-to-one mapping. For example, different task actions such as "audit" and "budget" relating to the base term "asset management" or same action such as "apply" relating to different terms "welfare" or "permanent resident". The categorising using verbs and nouns provide intelligent associations to tailor users' information seeking needs. The guided information depends on task context.

The proposed system suggests related concepts for the user to explore. It is up to the user to determine the topic of interests. The advantage of this is that retrieval is simplified and made more accurate by limiting the range of permissible terms through the use of controlled vocabulary. The result, then, is increased efficiency: less time is spent searching for records and the retrieval of the right records.

Task Action	Information Object	Information Guide
Access to	Public Transportation (e.g. train bus, airports, etc)	- Transportation route - Fare - Destination - Time required - Frequency of departures
Apply	Welfare (e.g. pension, special benefits, sole parent benefits, etc)	- Welfare legislation - Eligibility - Application procedures
Apply	Permanent Resident	- Resident in Australia - Sponsor for permanent resident - Assurance of supports - The point system
Audit	Asset Management	- Purchase orders - Sales - Tax

Budget	Asset Management	- Business plans - Cost-benefit analysis - Fraud - Guidelines/regulations - Legal - Purchase orders - Sales - Tax
Lodge	Tax (e.g. stamp duty, pay-roll tax, land tax, etc)	- Tax rates - Tax legislation - Tax guide - Electronic lodge procedures

Table 1 Sample examples of guiding user information seeking tasks.

4. INFORMAITON RE-USING

One of the advantages of using task actions for information seeking is that it aids query formulation. However, there may have some negative aspects of applying task actions. For example, it may restrict user's search freedom a bit. An important consideration is to maintain the user's task actions for re-using purposes. In order to increase the user’s successful experience and positive attitudes. It is desirable to capture specific fact intentions (Gaines 1997b) required by the user. Therefore, the ability to support information grows and evolves and becomes extremely important.

Three considerations are involved to improve the consistency and effectiveness of users' information searching tasks:

1. If a new task action is entered by the end-user.
2. If a new information object is entered by the end-user, and
3. If a new link between task and information action is created by the end-user.

For the first consideration, if a query includes a new task action, then the system will automatically insert the task into a task database. In order to consistently record concepts with their relationships, the relevance of the new object and the new relationship must be checked. For example, if an end-user is looking for "migrate to Japan" information, the system will check:

1. If this is a new task, "migrate", and
2. If this is a new information object, "Japan", and
3. If this is a new relationship between the task and object.

Since the built system - the Intranet, is related to Australia government information, the "migrate to Japan" query is not relevant to the topic and therefore won't be inserted. However, the Ontology Engineer/Record Manager can base on the number of these kinds of queries to include relevant links in the Intranet later. With the re-use of the query formulation capabilities, the ontology models would therefore evolve over a period of time and improve with increased usage.

5. CONCLUSION

We have provided three types of ontology models that structure information in a way to support navigation aids for information access. We have shown that exploratory browsing can be coupled with ontology-based knowledge to avoid getting lost in cyberspace particularly if users are not experts in the domain. In addition, we have discussed the maintenance of users' task actions to support information re-use purposes.

Research in ontology has led to the idea of the Ontolingua Server (Farquhar et al, 1996) for knowledge interchange. The Ontolingua Server provides a set of tools for browsing, creating and editing ontology on the Web. A major difference between the Ontolingua-based ontology and our proposed ontology methods concerns the task dependency of ontology. Ontolingua-based ontology is task-independent, whereas ours includes both task-dependent and task-independent ontology. We believe that the Ontolingua Server would be useful for providing information services on the Internet, whereas our ontology models would allow users to explore information from Intranets. SHOE (Simple HTML Ontology Extensions) allows the Web authors to annotate their pages with ontology-based knowledge (Luke, Spector and Rager, 1996). The ontology created in SHOE can be classified as organisational ontology, but SHOE lacks the matching task between domain activity ontology and user activity ontology.

In terms of our future work, an important consideration for information navigation is the challenge posed by the situated cognition movement (Menzies, 1996, Chandrasekaran and Josephson, 1997). Situated cognition is based on the idea that the way in which knowledge is used is highly dependent on situation. We feel that the navigation information issues should be investigated by varying the situations, in such a way that we are able to track the necessary changes of ontology models. This also implies that we need systems that support the changes needed for ontology medels as situations are changed. To tailor the ontology models to the situated cognition paradigm is needed.

ACKNOWLEDGMENTS

The work reported in this paper has been funded in part by the Research Data Networks (RDN) Co-operative Research Centre (CRC) program, Australia.

REFERENCES

Amba, S. Narasimhamurithi, N., O'kane, K., and Turner, P. (1996) Automatic Linking of Thesauri, Proceedings of the 19th International ACM SIGIR conference, 181-186.

Chandrasekaran, B. and Josephson, J. (1997) The Ontology of Tasks and Methods, the Symposium on Ontological Engineering, AAAI Spring Symposium Series.

CSIRO Keyword AAA (1996), CSIRO Resource Management Office..

Farquhar, A. Finke, R. and Rice, J. (1996) The Ontolingua Server: a Tool for Collaborative Ontology Construction, Proceedings of the Banff Knowledge Acquisition Workshop (KAW96).

Furnas, G., Landauer, I., Gomez, L., and Dumais, S. (1987) The Vocabulary Problem in Human-System Communication, Communication of the ACM, 30 (11):964-971.

Gaines, B (ed) (1997a) Using Explicit Ontologies in Knowledge-based System Development: a special issue, International Journal of Human-Computer Studies, vol 46, no 2/3.

Gaines, B (ed) (1997b) World Wide Web Usability: a special issue, International Journal of Human-computer Studies, vol47, no 1.

Gruber, T. (1992) A Translation Approach to Portable Ontology Specification, Proceedings of Japanese Knowledge Acquisition Workshop (JKAW'92), 89-108.

Ikeda, M., Seta, K., Kakusho, O., and Mizoguchi, R. (1996) An Environment for Building Conceptual Model of Problem Solving, Proceedings of the Pacific Knowledge Acqusition Workshop (PKAW'96), 210-225.

Lenat, D and Guha, R. (1990) Building Large Knowledge Bases, Addison-Wesley.

Luke, S., Spector, L., and Rager, D. (1996) Ontology-based Knowledge Discovery on the World-Wide Web, Proceedings of the Workshop on Internet-based Information Systems, AAAI-96.

Menzies, T. (1996) Assessing Responses to Situated Cognition, Proceedings of the 10th Banff Knowledge Acquisition for Knowledge-based Systems Workshop.

Mizoguchi, R. (1993) Knowledge Acquisition and Ontology, Proceedings of KB&KS'93, 121-128.

Mizoguchi, R. and Vanwelkenhuysen, J. (1995) Task Ontology for Reuse of Problem Solving Knowledge, Proceedings of KB&KS'95, 46-59.

Mizoguchi, R., Ikeda, M., Seta, K., and Vanelkenhuysen, J. (1995) Ontology for Modeling the World from Problem Solving Perspective, Proceedings of IJCAI Workshop on Basic Ontological Issues in Knowledge Sharing.

Neches, R., Fikes, R., Finn, T. Gruber, T. Senator, T. and Swartout, W. (1990) Enabling Technology for Knowledge Sharing, AI Magazine, 12, 36-56.

Sowa, J. (1995) Distinctions, Combinations, and Constraints, Proceedings of IJCAI Workshop on Basic Ontological Issues in Knowledge Sharing.

Sowa, J. (1997) Knowledge Representation: Logical, Philosophical, and Computational Foundations, PWS Publishing Company.

Smith, P., Newman, I., and Parks, L. (1997) Virtual Hierarchies and Virtual Networks: Some Lessons from Hypermedia Usability Research Applied to the World Wide Web, International Journal of Human-Computer Studies. (1997) 47, 67-95.

ter Stal, W. van der Vet, P., and Mars, N. (1995) The Role of Ontologies, Proceedings of IJCAI Workshop on Basic Ontological Issues in Knowledge Sharing.