

Organizing and Integrating Knowledge on the Desktop with Repositories

Gontlafetse Mosweunyane and Leslie Carr

School of Electronics and Computer Science, University of Southampton

{gm05r}{lac}@ecs.soton.ac.uk

ABSTRACT

The institutional repository, as a collection of research knowledge, is being encouraged to integrate with the researcher's desktop environment. However, the desktop metaphor used on personal computer systems provides only limited abilities to organize an individual's information; the same is true for the digital library metaphor adopted by repositories. Neither environment provides many tools for relating its contents to external information sources.

The need for automatically organizing information on the desktop to meet a user's various needs to accomplish knowledge tasks has been identified by researchers but there is still no widely-adopted solution to tackle it. Semantic web concepts are being used to organize web data into structures that enable machines to interpret and combine it easily with other information to build knowledge. In this paper we discuss the need for an organization mechanism for the desktop that will help integrate desktop data with other data from repositories and from the web, or vice versa.

1. INTRODUCTION

The institutional repository, as a collection of research knowledge, is being encouraged to integrate with the researcher's desktop environment, as it embodies their workflows and working practises [8]. To do this we need to understand how information is (or can be) organised on the desktop.

Systems based on the desktop metaphor allow for information to be stored in documents in files that can be named and placed in folders that can be nested to form hierarchical structures. This storage model is used on the most pervasive computer systems such as Microsoft Windows and Mac OS and is the only work environment

known to many users and designers [1]. This method of organization is static, and presents problems in finding items later and reminding users of what items they have [2], among other problems.

Current facilities on the desktop also make it difficult, if not impossible, to associate information within the desktop and with other separate information sources. Semantic data on the web is published by the use of linked data browsers to ensure that people can interlink different sources. In a similar way, we need structured data derived from data on our desktops. We also need to create applications and browsers that make it easy to interlink information on the desktop and make it easy to find and link to other information outside the desktop that may be related to what they have.

2. PROBLEMS WITH FINDING AND RELATING ON THE DESKTOP

Researchers have also found that information does not fall into neat categories [3], but rather falls into overlapping and fuzzy ones which cannot remain unambiguous over time [4]. Malone [5] and Rekimoto [6] also identified the problem users have with classifying documents into specific folders because they may belong to more than one category. The categorical structure for files also changes over time and is dependent on the task one might be doing at that time [7].

The separation of information in files and folders and unavailability of mechanisms allowing the users to dynamically organize it to allow for easier finding and reminding hinders the efficiency and effectiveness of carrying out work, and therefore creating of knowledge by users. Users are also been encouraged to define and add metadata in the form of attributes and keywords to accompany their documents, but

there is little reward for putting an effort into this in current systems. Classifying documents into folders itself involves a major cognitive task [5]. Users can be assisted in locating and integrating information if these attributes can be used to dynamically restructure and view information on the desktop.

Organizing on the desktop will make it easier to interlink with similar data outside the desktop. A repository is an institutional scale collection which feeds off individuals and is organized in a particular way. The desktop, on the other hand, represents an individual's "collection" and "organization". Users will benefit from an organization mechanism for the desktop that will help integrate desktop data with data from the repository.

3. DYNAMIC ORGANIZATION USING KEYWORDS

The semantic web emphasizes the use of metadata to express the meaning of data and encoding relationships between these such that it can be shared and easily integrated by applications and intelligent agents. Most applications developed for the semantic web so far simply utilize the inherent semantics provided by the structure of the data encoded in Resource Description Framework (RDF) to visualize relations between data sources. Analysing the structured data itself may help in deriving more knowledge from it such that it can be related and therefore connecting to similar data.

Keywords have always served an important part in locating and relating information, especially in search systems where their use and frequency is used to locate more relevant items. The experiment undertaken involved the extraction of keywords from metadata derived from the inbuilt properties of Microsoft Office documents and using the alphabetic-ordered keywords to dynamically categorize them.

The future part of the project involves the definition and implementation of methods that help connect these categories to relevant semantic information/resources on the web (see figure 1). We are also investigating how to

compare and integrate similar data sources based on their metadata.

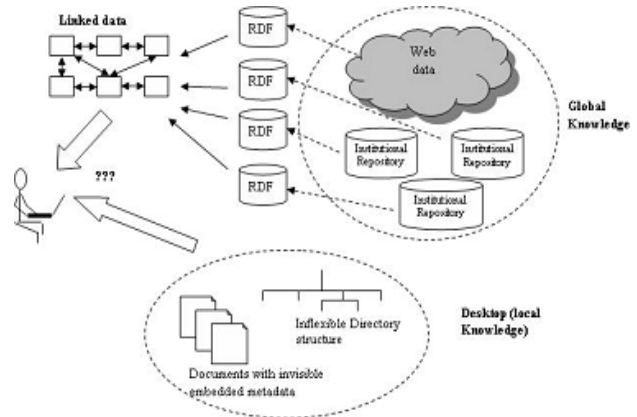


Figure 1: Problem Integrating Different Knowledge Sources on the Desktop

4. CONCLUSION

The problem of integrating repositories with users' desktops is the lack of organising principles which are available to the desktop user. There are few mechanisms provided by the system apart from the file system structure and invisible, application-defined metadata. A proposition for dynamic and flexible organization on the desktop for finding information in the repository and the desktop is presented as a solution. Dynamic organization gives the user the freedom to use attributes to record ideas over time and be able to locate them later with similar information. It also helps visualize implicit relations between the data. Integration with other data sources is also important for building knowledge.

The envisioned framework will help in locating and associating information users have on their desktops and that available from institutional repositories and other web sources.

5. REFERENCES

- [1] Victor Kaptelinin and Mary Czerwinski, *Introduction: The Desktop Metaphor and New Uses of Technology*, Beyond the Desktop Metaphor: Designing Integrated Digital Work Environments, MIT Press, 2007, 1-12.
- [2] Eric Freeman and David Gelernter, *Beyond Lifestreams: The Inevitable Demise of the Desktop Metaphor*, Beyond the Desktop Metaphor: Designing Integrated Digital Work Environments, MIT Press, 2007, 19-48.
- [3] Lansdale M., *The Psychology of Personal Information Management*, Applied Ergonomics, 19, 55-66, 1988.

[4] Dumaïs, S.T., and Landauer, T.K., *Using Examples to Describe Categories*, Proceedings of the 1983 ACM SIGCHI Conference, 112-115, December 1983

[5] Thomas W. Malone, *How do people organize their desks? Implications for the Design of Office Information Systems*, ACM Transactions on Office Systems, 1(1), 99-112, 1983.

[6] Jun Rekimoto, *Timescape: A Time Machine for the Desktop Environment*, CHI 99 Extended Abstracts on Human Factors in Computing Systems, pages 180-181, 15-20, May 1999.

[7] P. Dourish, W. K Edwards, A. LaMarca, J. Lamping, K. Petersen, M. Salisbury, D.G.Terry, and J. Thornton, *Extending Document Management Systems with User-Specific Active Properties*, ACM Transactions on Information Systems, 18(2):140-170, April 2000

[8] -, 2007, "Bring the Repository to the Workflow". CRIG Unconference, Birkbeck College, 6th-7th Dec 2007. http://www.ukoln.ac.uk/repositories/digirep/index/CRIG_Unconference