Combining Context Ontology and Landmarks for Personal Information Management

Khalid Latif, A Min Tjoa
Institute of Software Technology and Interactive Systems
Vienna University of Technology,
Favoritenstrasse 9-11/188,
A-1040 Vienna, Austria
{klatif, amin}@ifs.tuwien.ac.at

***Abstract*-In this paper we present our work on personal information management by exploiting both context ontology and landmarks. The context model introduced is based on location, agent, time, and category of information corresponding to the 4Ws: where, who, when, and what. Each of them represents one dimension of an information item. Semantics of each dimension are explicitly described in the ontology for personal information items along with their inter-relation. Therefore a necessary modification of the LATCH model has been investigated in-depth for its usage for personal information management and further retrieval. Navigation in the information space is supported by the use of landmarks.**

I. INTRODUCTION

Ever increasing capacity of contemporary storage devices inspires the vision to accumulate information without the need to delete old data. An information item is useful only when it is stored and later on being possible to look at it. Now the technology is at such a point that the enormous amount of information can be stored, but is not being exploited effectively and efficiently due to lacking semantics. For personal information the semantics emerge from multiple dimensions such as by analyzing the contents and discovering ontology instances. Consequently personal information can be structured in a suitable way by exploiting the fact how most people organize their life bits [18] to facilitate efficient retrieval.

Modern day desktop applications allow humans to benefit from different approaches for organizing information items. Such applications rarely exploit their semantics and also do not use a common conceptual scheme for information management. Thus information items organized by an application following a specific metaphor can not be automatically linked to the items managed by other applications. Nevertheless, such an inter-relation and inter-linking is important. Otherwise users have to redundantly *re-enter on a new path* to find the required information with the associated counterparts managed by different applications – users are captured in a “prison of metaphors”.

This work as part of the SemanticLIFE project is generously supported by ASEA-UNINET (ASEAN-EU Academic University Network) and HEC (Higher Education Commission of Pakistan). This research is also supported by the DynamOnt project which is funded by the Austrian Government’s FIT-IT Research Initiative on Semantic Web under the contract 809256/5512.

1-4224-0220-4/06/$20.00 © 2006 IEEE.

The problem could be demonstrated by an example: *Alex* is searching an article on *association of thoughts* knowing the fact that it was saved in a (file system) folder after following the web link forwarded by *John* in his email from last summer (see Fig 1). To add to the complexity say Alex has read a lot of articles from the web on the same topic and have saved them in his workspace. Now with traditional information retrieval techniques he can try to search for the said article based on the keywords “John” and “association of thoughts”. Such keyword based search will not retrieve the desired article effectively (may rank it too low) because of the fact that no document actually contains both search phrases.

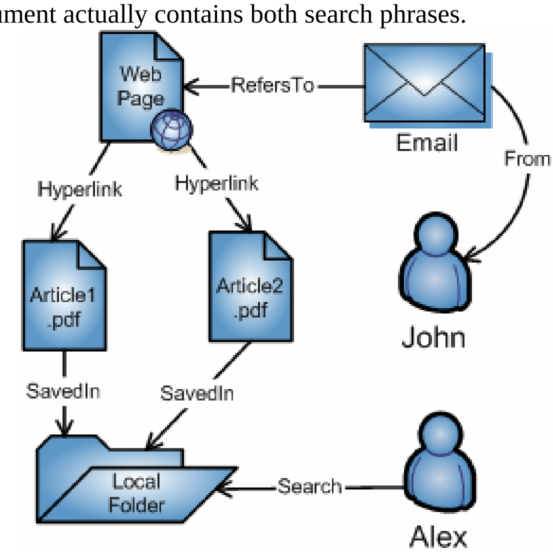

Fig 1. Interrelation of different information items.

While this trivial example shows the usefulness of portraying *involved agent* from the context, exploiting more associations emerging from other dimensions like location and time can realize even very complex scenarios.

This kind of interrelation was suggested half a century ago by Vannevar Bush. He suggested to draw on the principle of **trails** for personal information organization due to the congenital nature of human mind to follow the **association of thoughts** [6]. It is worthwhile to quote his statement about how human mind uses this principle to follow an information item:

When data of any sort are placed in storage, they are filed alphabetically or numerically, and information is found by tracing it down from subclass to subclass. It can be in only one place, unless duplicates are used; one has to have rules *as to which path will locate it, and the rules are cumbersome. Having found one item, moreover, one has to emerge from the system and re-enter on a new path.*

The human mind does not work that way. It operates by associations. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails …

A. Problem Statement

The primary goal of this research is to build a personal information management system by exploiting the semantics of the information and thus realizing the vision of trails and association of thoughts. In our SemanticLIFE project which have been started in 2003 we pursuit this goal in a fundamental way. The architecture of SemanticLIFE system is presented elsewhere [2]. The range of data sources starts from communication data (emails, phone calls, chat sessions) to personal documents, pictures, web-browsing sessions and calendar data, and may include a whole range of additional sources up to sensory data (e.g. temperature, geographic location, blood pressure). In this paper we specifically focus on the issue of semantic associations between information items by exploiting the context. We discuss the main issues that arise when realizing the vision of association of thoughts for personal information.

Continuous Extension: In the words of Vannevar Bush *a record if it is to be useful to science, must be continuously extended* [6]. Storing information object for later retrieval is one type of extension, but the desired augmentation is gradual semantics enhancement – thus transforming information objects to knowledge objects.

Information Context: Personal information items have varying content types. A uniform conceptual schema is required for modeling the semantics of information context for most of the items.

Information Network: For realizing trails it is necessary to structure the personal information space where each information item is organized by its association with some other information item(s).

The Relative View: In some cases, characteristics of an information item are identified relative to a significant event, or state of the same or other item. Such as for a picture declaring that it was taken after few days of a momentous event entails reconciliation of subjective and objective views.

B. Related Work

The vision of log-term archival and organization of personal information was coined by Vannevar Bush half a century ago [6]. Many research projects are targeting to realize this vision, [1, 11, 13]. A comprehensive analysis of such systems is out of scope for this paper. Here we briefly discuss existing efforts in establishing associations between information items and landmarks.

Information Semantics: One widely used metaphor for organizing personal information is time such as for SIS project [9] time line was used to represent information items [19]. Lifestreams project used time-ordered streams as storage model for organizing user's personal workspace [11]. Jim Gemmell and colleagues explored location and time in organizing personal photo archives [4] for the MyLifeBits [13] platform. Different studies have revealed that although important but, time should not be the only principle to organize personal information [9, 23]. A little effort has been put to identify other generic dimensions.

Association of Thoughts: MyLifeBits system uses typed-links and *transclusion* for associating one item to other [13], and trails are realized by manually constructing collections. Amit Sheth and colleagues are researching new ways for modeling, discovery, and ranking of semantic associations [3] as part of SemDis[1] project. In their hypothesis two items are semantically associated if a property sequence of a certain form is present in the graph representation of items. Their technique doesn't focus on building the trails.

Landmarks: [20] identified three categories of landmarks (1) *visual*, (2) *semantic*, and (3) *structural* for information management intended to be used in the context of web and hypertext. Although the work done by the author is remarkable but may not be re-used for personal information space as most of the personal information items such as contact detail, meeting reminders, phone calls, and activity logs are inherently different from hypertext. The notion of landmarks is also used in graph drawing of co-citation networks [7] for representing the importance of a node such as a highly cited article.

C. Contribution

To grasp the enormity and flood of information in case of life time capture of personal experiences [5, 13]; we propose to make use of context ontology coupled with landmarks. We use the term "Context" in the following sense: The context of a life-item is the semantic insights of its contents and relation with other items. While context ontology will provide semantics to life-items, landmarks guarantee efficient retrieval in large information space by exploiting associations.

First we will briefly discuss existing approaches for personal information organization. In the subsequent section we will introduce context ontology for capturing and enhancing semantics of life-items. The context ontology provides a binding for information items and it further benefits from the landmarks which are presented in section 4. Finally we summarize the status of our work and an outlook of open issues.

II. Personal Information Organization

Driven by the vision of trails and association of thoughts we analyzed different desktop applications to study their information organizing principles and how trails could be formed by adding information context to conceptual schemata of life-items. The choice of how to organize information is not

[1] http://lsdis.cs.uga.edu/projects/semdis/

always obvious, since more than one scheme can apply. For that reason applications usually support multiple presentations of their confined contents. For example emails could be listed based on their sent/receive date, persons involved in the email, the thread, or even based on their contents (cf. Fig 2 for later case).

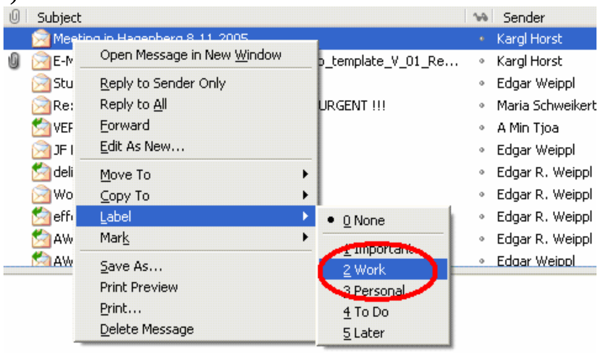

Fig 2. Tagging emails based on their contents

Context principles such as time and categories although used somehow in different applications mostly implicate that their inter-relation is missing in two ways. *Firstly* semantics of the information items are not modeled explicitly and there is no binding of properties possessed by one information object with the others. Providing a unified view of the information space consisting of such objects becomes substantially difficult and hence implies the absence of morphism from one contextual organization to any other context in personal information space.

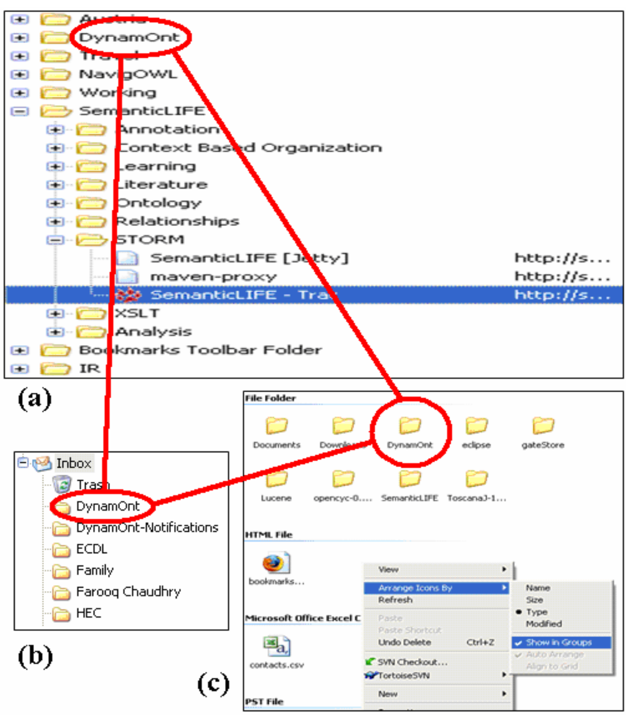

Fig 3. Categories in (a) Bookmarks, (b) IMAP Folders, and (c) File-system directories – similar category is highlighted.

Secondly categories, hierarchies, or other principles used in an application have no inter-relation with those used in other applications. For example the bookmarks items, emails, working documents, and presentations for a specific project live in separate homes (see Fig 3). Thus an information item present in a category in one application has no explicit associations with its counter parts in other applications in the similar category, which is necessary for building trails and allowing humans to follow the association of thoughts to locate an information object. One possible solution to this problem is (1) using a shared conceptualization of information items, (2) exploiting the semantics using a context ontology, and finally (3) annotating and linking the information items based on that ontology.

III. Semantics Enhancement with Context

Several definitions of context are present in the literature and variety of context models are in use by researchers [21]. Context could be used for personal knowledge management in (1) the information capture and organization stage and (2) during the retrieval time. The importance of capturing the context during the first stage for building associations is obvious as demonstrated earlier. The diversity of context models raises the question of what aspects should be captured for personal information as part of the context. We propose to model aspects which may well be used to organize information objects. Thus the information context will symbolize the personal information space in which each aspect represents one dimension or view of the information.

A. LATCH as Context Model

Richard Wurman identified that organization of information is finite and there are only five principles ***L****ocation*, ***A****lphabet*, ***T****ime*, ***C****ategory*, and ***H****ierarchy*, known as LATCH [24]. Our study of existing desktop and personal information management systems has revealed that most of them use one or more LATCH principles for information organization. But, all of its principle may not be taken as input to context model due to their non-contextual nature. Alphabet, for example, is more an ordering principle and could be applied to any other context metaphor such as lexical ordering of locations.

On the other hand hierarchy could be taken as a mean to organize categories and locations. This is because the hierarchy typically implies arrangement of items in a tree structure. But the principle of hierarchy in LATCH refers to a continuum organization where each information item is organized in relation to the other items based on some property; e.g. picture *p1* was taken after picture *p2* that was taken after picture *p3*, etc. A story from the selected pictures could easily be created provided the information space is organized in such a way. For our context ontology we have captured both the meanings of hierarchy, tree structure and continuum. More detail on the later aspect is present in the next section.

Another aspect missing in the LATCH, if considered as context model, is the **agent**. Most of the activities such as personal communication (e.g. emails, instant messages, and phone calls) and collaboration (as in research projects or in office work) embody other persons. Photos also encircle agents – mostly human agents. Yet there are information items which depict non-human agents such as correspondence with some research funding agency or university. Thus replacing alphabets with agents in the LATCH make it suitable for modeling as context metaphor in personal information management.

Interestingly location, agent, time and category match with *where*, *who*, *when* and *what*[2] respectively. Additionally the hierarchy represents the relationship of information object with other objects of similar kind. Modeling these dimensions in personal information as ontology not only explicitly amplifies its semantics but also provide a foundation for binding between the information objects. Having described the drive behind the context ontology we proceed toward its development.

B. *The Context Ontology*

Ontologies could be developed and arranged by following different approaches such as taxonomic and faceted based [8]. For this research we have restricted ourselves to hierarchical approach, and hope to benefit from the other approaches in future.

Instead of developing the ontology from scratch we have adopted concepts mainly from SUMO [17] and OpenCyc (open source version of Cyc [15]). This satisfies a fundamental feature of conceptual ontology design patterns [12] and DynamOnt methodology[3]. The extracted fragments of SUMO and OpenCyc provide taxonomic and axiomatic context of the ontology. For the final ontology we mapped each concept to appropriate WordNet synset [10]. The location, agent, time, and category part of the context ontology are briefly discussed below:

Location describes a point or extent of a life-item in space. OpenCyc describes a comprehensive vocabulary of geographic concepts including *GeographicalRegion*, *Continent*, and *Country* etc. Some reference ontologies do not define such vocabulary and instead use the notion of *longitude* and *latitude* as properties of the concept *Location*. We agree that a geographic region could be described as longitude and latitude values but for the first version of our LATCH Ontology we have modeled the Cyc approach. Interestingly OpenCyc includes vocabulary on proximity containing the concepts such as *near*, *adjacentTo*, and *onPath* which we found very useful for supporting continuum and relative values.

Agent is a generic notion, an individual or an organization, which can take a role for carrying out operations. The agents are further categorized into (1) *Individual*, (2) *Group*, and (3) *Organization*. The concept of group, although present, is not used in the context of agent in Cyc. Fig 4 describes the agent model of our ontology.

Time is modeled as a *Time-Quantity* and as *TemporalThing* having temporal extent by Cyc and we have adopted the same. *TimePoint* is a specialization of *TemporalThing* which represent one point on the timeline axis. Many relationship predicates were also modeled. For example the predicate *endsDuring* in the statement *(S endsDuring O)* means for the *endPoint* (ep_s) of *TemporalThing S* the following holds *(O temporalBoundsContain* ep_s*)*. Recently OWL-Time [14] was suggested by Semantic Web Best Practices Working Group (SWBPD) for modeling the time in RDF. In future, we will examine how the time model we adopted from Cyc could be aligned with OWL-Time.

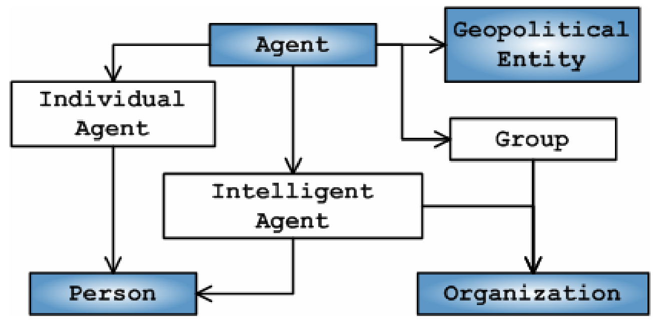

Fig 4. Abridged agent model, fragments are extracted from the reference ontologies.

Categories: We use categories to illustrate what the information contents are about. One can argue that location, time, and agent could also be realized as categories. We elucidate it with one example: Both, an article on sight seeing tours in Vienna and an email for registration in European Conference on Digital Libraries (ECDL) held in Vienna, can be positioned close to each other on the location axis. But their topic is different from each other, characterized by placing them in distinct categories. Still both artifacts can belong to one collection (say *ECDL Participation*). Thus by placing a life-item in a category we describe *what* the contents are about instead of *when*, *where*, or *who* is involved.

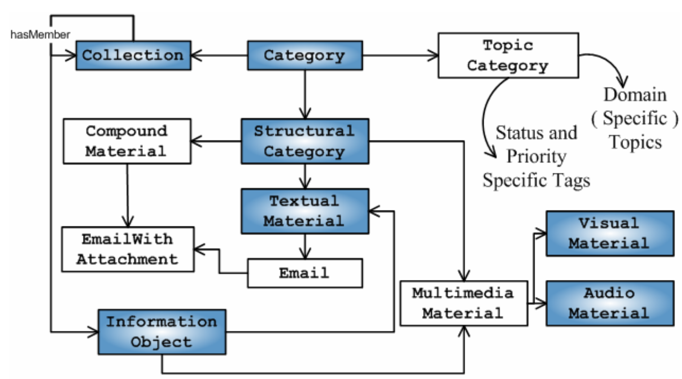

Fig 5. Category hierarchy, arrows represent specialization and highlighted concepts are mapped to the reference ontologies.

We support three kinds of categories based on (1) *content structure* cf. Fig 5, (2) *content topic* after following the

[2] Because the categorization is based on the contents describing what the information object is about.

[3] The development process started with use case description of the scenario and then gradually building the ontology model by aligning the concepts with reference ontology. See http://dynamont.factlink.net

domain ontology (we have selected Academics and Research as the domain and used SWRC [22] and AKT Reference Ontology [16] for modeling the topic categories), and (3) *collections* built by the users. Collections are distinct from other categories because they can contain variety of life-items and also one collection could be placed in more than one category. As with above example, the collection *ECDL Participation* could be placed in the category of *Attending-A-Conference* (taken from AKT Reference Ontology). Collections are useful for the users to organize life-items while doing a specific task and as demonstrated in [13] could be valuable for building trails.

IV. LANDMARKS

Humans make use of variety of practices to build associations and trails. *Method of loci* (also known as *mnemonics*) is one example of such practices originated with the ancient Greeks. The idea is to relate parts of the information to well-known landmarks. Recent example of its use is in rescue operations after earthquake in northern areas of south Asia where American pilots where having difficulty to remember south Asian city names. For efficient communication they virtually named the effected cities (*Balakot*, *Bagh* etc.) after city names in USA. Another use of landmarks is as reference point such as in hypertext systems. The opening web page (*home page*) is considered a landmark and every other web page in that particular web application is linked with it. While the first example shows random association the later is more logical. For us it is not important if a landmark is used as mnemonic or as a reference point more important is building the trails by linking together different items. Though we focus on modeling the context and landmarks for automatically building logical associations, users are not impeded in manually constructing random ones.

For the case of personal information any object or its significant state can be associated with other objects and thus creating a cognitive map of the life-items. In contrast to public and personal landmarks on time axis as proposed in [19] we argue that (1) landmarks could be located in various axis not only time and (2) the significance of landmarks is better remembrance of items so they should be all personal. By saying this we do not negate that a momentous news story can be a landmark. The point is that a landmark should be induced by the user and not by the system so it is personal in that sense.

For declaring an information item as landmark user simply selects it and assign a non negative weight value $w>0$. The weight w provides the attraction force and determines the strength of the landmark in terms of semantic depth. Based on the value of weight w we decide if an item is linked and could be followed through the landmark even if it is not directly linked with it. An item I is said to be semantically associated with a landmark L_K having weight w_K if the following holds:

$$SemanticDistance(I, L_K) \leq w_K \quad (1)$$

The semantic distance is computed in several ways such as the manual associations, property-entity associations [3], and the topic association. The later is measured from the hierarchical distance of the topics of the items from the RDF graph. For manually associating one or more information items with a landmark user drags and drop those in an item list widget (see Fig 6). The desired landmark is later on selected and finally the user commits to establish the association. Comments, both (1) free text or (2) using a category hierarchy, could be attached with the association. In principle this process of associating items with landmark could be applied to connect an information item with any other information item.

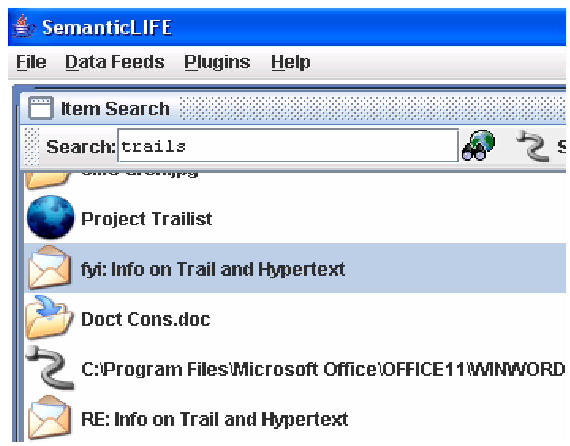

Fig 6. Item list widget

V. KNOWLEDGE SPACE NAVIGATION

Driven by the vision about trails, SemanticLIFE allows the user to select any path. For example user can select agent axis as starting point, and *Pakistan* as the agent. This will get him/her information items related with *Pakistan* such as the news story *Austria helped Pakistan with water processing plant in earthquake rescue operations* as the most recent landmark item associated with the concept *Pakistan* in the context of agent. Noticeably *Pakistan* and *Austria* both are instance of *GeopoliticalEntity* which in this scenario means an *Intelligent Social Agent*. Selecting the news story will present the user with all information items associated with the landmark such as the news stories of the earthquake, photos of the scenes and the fact that a fund raising lunch was arranged in the United Nations headquarter in Vienna.

Now the user can look for the life-items on the location axis by zooming in to *Vienna*. Items with fine-grained locations of Vienna will also be presented to the user such as the collection "Talks at institute IFS building". The only limit remains imagination as the user can choose the time axis to view items close to a specific talk on timeline axis.

Citing Vannevar Bush once again – *Man cannot hope fully to duplicate this mental process [of association of thoughts] artificially… [But] it should be possible to beat the mind decisively in regard to the permanence and clarity of the items* *resurrected from storage*.

VI. Conclusion and Open Issues

We have presented our research on enhancing semantics of personal information for building trails by using context ontology and landmarks. We have customized the LATCH principle, by replacing lexical ordering with Agent, for using it as context model. Although we have used a hierarchal approach to ontology building, work on facet approach is planned for future. Additionally, the ontology contents need more efforts to make them comprehensive e.g. longitude and latitude values for geographical entities are not dealt with so far.

Automating the ontology building task so that the user can manage his personal view of the domain is not an easy task. With in the DynamOnt project we are investigating a methodology for dynamically building ontologies. We hope to benefit from its question answering approach in future to allow the user for creating new content topic categories and placing them in reference ontology.

References

[1] E. Adar, D. Karger, and L.A. Stein, "Haystack: Per-User Information Environments", In *Proceedings of 8th International Conference on Information and Knowledge Management*, Kansas City, Missouri, 1999.

[2] M. Ahmad et al., "SemanticLIFE - A Framework for Managing Information of a Human Lifetime", In *Proceedings of 6th International Conference on Information Integration and Web-based Applications and Services*, Jakarta Indonesia, 2004.

[3] B. Aleman-Meza, C. Halaschek-Wiener, B. Arpinar, C. Ramakrishnan, and A. Sheth, "Ranking Complex Relationships on the Semantic Web", *IEEE Internet Computing*, vol. 9(3), May/June 2005, pp. 37-44.

[4] A. Aris, J. Gemmell, and R. Lueder, "Exploiting Location and Time for Photo Search and Storytelling in MyLifeBits", *Microsoft Research Technical Report*, MSR-TR-2004-102, October 2004.

[5] R. Boardman, R. Spence, and M.A. Sasse, "Too many hierarchies? The daily struggle for control of the workspace", In *Proceedings of 10th International Conference on Human-Computer Interaction*, 2003, pp. 616-620.

[6] V. Bush, "As We May Think", *The Atlantic Monthly*, July 1945.

[7] C. Chen, *Information Visualization: Beyond the Horizon*, 2nd ed., Springer-Verlag London, 2004, pp. 285-287.

[8] R. P. Díaz, "A Faceted Approach to Building Ontologies", In *Proceedings of IEEE International Conference on Information Reuse and Integration*, October 2003.

[9] S. Dumais, E. Cutrell, J.J. Cadiz, G. Jancke, R. Sarin, and D.C. Robbins, "Stuff I've Seen: A System for Personal Information Retrieval and Re-Use", *SIGIR*, Toronto, Canada, 2003.

[10] C. Fellbaum, *WordNet: An Electronic Lexical Database,* The MIT Press, 1998.

[11] E. Freeman, and D. Gelernter, "Lifestreams: A Storage Model for Personal Data", *ACM SIGMOD Bulletin*, vol. 25(1), March 1996, pp. 80-86.

[12] A. Gangami, "Ontology Design Patterns for Semantic Web Content", In *Proceedings of International Semantic Web Conference* (ISWC), LNCS 3729, 2005, pp. 262-276.

[13] J. Gemmell, R. Lueder, and G. Bell, "Living with a Lifetime Store", ATR Workshop on Ubiquitous Experience Media, Keihanna Science City, Kyoto, Japan, September 2003.

[14] J.R. Hobbs, and F. Pan, "An Ontology of Time for the Semantic Web", *ACM Transactions on Asian Language Processing: Special issue on Temporal Information Processing*, vol. 3(1), March 2004, pp. 66-85.

[15] D. Lenat, "Cyc: A Large-Scale Investment in Knowledge Infrastructure", *Communications of the ACM*, vol. 38(11), 1995, pp.33-38.

[16] E. Motta, and N. Gibbins, "AKT Reference Ontology", v2.0 2003, http://www.aktors.org/publications/ontology/

[17] I. Niles, and A. Pease, "Towards a Standard Upper Ontology" In Proceedings of 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), October 2001.

[18] P. Ravasio, S.G. Schär, and H. Krueger, "In Pursuit of Desktop Evolution: User Problems and Practices with Modern Desktop Systems", *ACM Transactions on Computer-Human Interaction*, vol. 11(2), June 2004, pp. 156-180.

[19] M. Ringel, E. Cutrell, S. Dumais, and E. Horvitz, "Milestones in Time: The Value of Landmarks in Retrieving Information from Personal Stores", In *Proceedings of 9th International Conference on Human-Computer Interaction*, INTERACT'2003.

[20] M.E. Sorrows, "Recall of Landmarks in Information Space", *PhD Dissertation*, School of Information Sciences, University of Pittsburgh, 2004.

[21] T. Strang, and C. Linnhoff-Popien, "A Context Modeling Survey", In *Proceedings of the Workshop on Advanced Context Modelling, Reasoning and Management*, 2004.

[22] Y. Sure, S. Bloehdorn, P. Haase, J. Hartmann, and D. Oberle, "The SWRC Ontology – Semantic Web for Research Communities", *Lecture Notes in Artificial Intelligence*, vol. 3808, Springer-Verlag, 2005, pp. 218–231.

[23] J. Teevan, "How People Re-find Information When the Web Changes", MIT-CSAIL AI Memo 2004-012, June 2004.

[24] R. Wurman, D. Sume, and L. Leifer, *Information Anxiety 2*, Que, 2000.