Towards a Conceptual Framework and Metamodel for Context-aware Personal Cross-Media Information Management Systems

Sandra Trullemans and Beat Signer

Web & Information Systems Engineering Lab
Vrije Universiteit Brussel
Pleinlaan 2, 1050 Brussels, Belgium
{strullem,bsigner}@vub.ac.be

Abstract. Information fragmentation is a well-known issue in personal information management (PIM). In order to overcome this problem, various PIM solutions have focussed on linking documents via semantic relationships. More recently, task-centered information management (TIM) has been introduced as an alternative PIM paradigm. While these two paradigms have their strengths and weaknesses, we aim for a new PIM system design approach to achieve better synergies with human memory. We further envision a cross-media solution where physical information is integrated with a user's digital personal information space. We present the Object-Concept-Context (OC2) conceptual framework for context-aware personal cross-media information management combining the best of the two existing PIM paradigms and integrating the most relevant features of the human memory. Further, we outline how the OC2 framework has been implemented based on a domain-specific application of the Resource-Selector-Link (RSL) hypermedia metamodel.

1 Introduction

The research field of Personal Information Management (PIM) investigates complex interactions with personal information. An emphasis is on the organisation and maintenance of personal information in file hierarchies, emails, on desks or in bookshelves. This personal information is spread over devices, tools and Web Services or managed in physical archives and piles. Several researchers investigated the effects that information fragmentation has on the organisation and re-finding of personal information. Bergman [1] observed that users experience a cognitive overload when re-finding information to fulfil their tasks since the information is stored in different tools (e.g. email client) as well as at different places in the hierarchical file system.

In order to overcome the information fragmentation problem, two main PIM paradigms have been introduced. Bush [2] described the Memex as a potential future PIM system and criticised the bad practice of information organisation in hierarchical storage systems. He proposed a new *unified view* paradigm where

information items are linked with each other to form an interlinked personal information space. With the rise of semantic web technologies, a number of promising unified view-based PIM systems have been realised. The more recent *task-centered information management (TIM)* [3] paradigm focusses on providing task-related personal information.

We introduce a new PIM approach addressing the problem of information fragmentation. While the unified view and TIM paradigms are based on some human memory features, we aim for better synergies between PIM systems and the human memory where the PIM system acts as a memory prosthesis and supports intuitive and natural interaction [4]. We introduce the Object-Concept-Context (OC2) framework which combines the advantages of both, the unified view and TIM paradigms, and conforms to human memory in organising and re-finding activities. We further discuss an implementation of the OC2 framework based on a domain-specific application of the Resource-Selector-Link (RSL) hypermedia metamodel [5].

2 Background

The long-term information storage in human memory involves two memory systems [6]. The *episodic memory* includes the process of receiving and storing spatio-temporal information about events and their autobiographical references to previous events as well as perceptible properties (i.e. contextual factors). In contrast to the episodic memory, the *semantic memory* consists of semantic relationships between concepts. Each of these relationships has a weight which indicates how relevant the relationship is in a given context and might differ for the two directions of a bidirectional relationship between two concepts [7]. Note that the semantic memory does not store perceptible properties but rather references to the episodic memory [6]. Concepts instantiated in the semantic memory are general ideas formed in our mind to abstract the complexity of the real world. On the other hand, objects are observable artefacts which are categorised in concepts in order to give them semantics [8]. Barsalou [9] states that this classification depends on the goal or purpose people have in mind. Therefore, an information item which is present in the real world can be classified in more than one concept (cross-classification) and with a certain context relevance [10]. In order to retrieve information from semantic memory, the spreading activation theory where external stimuli activate various concepts based on their perceptible properties in the episodic memory is applied [11]. We can identify four main features of the human memory including the storage of contextual factors for each concept, context-dependent relationships between concepts, the context-dependent categorisation of objects and the use of spreading activation to retrieve the right information for a given context.

In order to provide a unified view of a user's personal information space, two approaches have been introduced. First, systems such as MyLifeBits [12] follow the idea of episodic memory by organising personal information on a timeline, but these systems lack the functionality to create associative relationships. A

second category of PIM systems focusses on semantic memory based on Bush's vision of associative trails between documents [2]. HayStack [13] enables the linking of documents and supports the categorisation of objects into concepts. In order to provide machine-readable semantic relationships, semantic desktops such as Gnowsis [14] specify these relations through a personal ontology. TIM-based solutions step away from focussing on information organisation [3]. Similar to the unified view paradigm, they define a personal ontology for the modelling of semantic relationships [15]. In OntoPIM, users can allocate objects to ontology classes and extend the personal ontology through inheritance. In order to provide task-centered information, Katifori et al. [16] implemented a spreading activation algorithm for OntoPIM. By allocating a weight for a given task to the personal ontology's entities, the most relevant concepts can be activated.

3 The OC2 Conceptual Framework for PIM Systems

In order to achieve a better synergy between PIM systems and the human memory, we extended the unified view and TIM paradigms to address the previously mentioned four main features of the human memory as well as to support more natural interactions. Our *Object-Concept-Context (OC2)* conceptual framework for context-aware personal cross-media information management systems consists of the three *object*, *concept* and *context* layers outlined in Fig. 1.

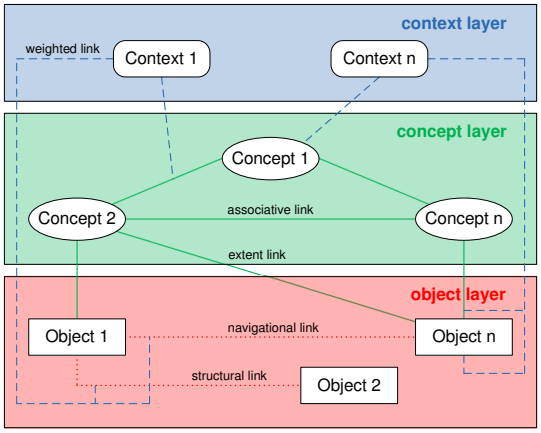

Fig. 1. Object-Concept-Context (OC2) conceptual framework

The object layer contains objects representing real-world elements which can be observed in physical or digital space. An object might, for example, stand for a physical paper document or a post-it note but also an email or a website. Since objects are proxies for real-world elements, they need to be uniquely identifiable via a uniform resource identifier. Note that objects are not restricted to single documents but can represent complex structures composed of individual objects. For instance, an object representing a document can also be treated as

a composition of objects representing the paragraphs, allowing us to organise and retrieve information at a finer granularity level than the commonly used information units such as documents.

Elements of the OC2 object layer may have relationships with other objects via *navigational* or *structural links* as illustrated by the dotted red lines in Fig. 1. While currently most users navigate between files via file hierarchies, we would rather like to have the possibility to associate different objects across tools and devices in order to retrieve information in a more natural way. Since an object can be composed out of other objects, we also need structural links to represent these compositional relationships. Furthermore, structural links can be used to express organisational structures within a personal information space. The unified view paradigm is criticised since users often prefer their own organisational behaviour [17]. We should therefore offer the possibility to integrate a user's current organisational structures across the physical and digital space. Our conceptual model goes beyond mimicking the human memory and enables the description of a user's organisational behaviour.

The concept layer consists of conceptual elements (e.g. words or sentences) representing a user's conceptualisation and abstraction of observed objects similar to the definition of concepts in psychology [8]. On the concept layer, concepts may have *associative links* to other concepts as well as *extent links* to objects of the underlying object layer as indicated by the solid green lines in Fig. 1. As mentioned earlier, the human memory builds semantic associations between its elements. In our OC2 framework, we represent these semantic associations via bidirectional associative links. Existing PIM solutions are often based on semantic web technologies and ontologies are used to describe semantic associations, whereas in OC2 we step away from technical restrictions. The concept level further introduces extent links representing any categorical relationship between a concept and various objects whereby an object might participate in more than one extent link. Thus, our conceptual framework addresses the psychological theory that concepts are internal while objects are external to our mind [8]. In HayStack [13], automated extraction modules allocate objects to their semantic concept, whereas OntoPim [15] supports the manual categorisation of objects in a personal ontology. We argue that both, the manual as well as the automatic classification of objects should to be supported.

As elaborated earlier, the human memory does not only consist of the semantic memory but also includes an episodic memory which is responsible to store contextual factors about events. By preserving these contextual factors, users may search in their semantic memory based on contextual cues. Previous descriptive PIM research has observed the importance of contextual cues in refinding activities. For example, during retrieval users often use the contextual information squeezed into folder labels [18]. While the TIM paradigm mentions the exclusive use of personal information in executive tasks [3], we do not agree that personal information is only used in specific tasks such as managing a photo album. The OC2 contextual layer contains *context* elements for managing contextual factors about the semantic graph induced by elements of the concept

and object layers. A context element describes a composition of contextual factors and a contextual factor can be any condition or observation. While most context-aware systems apply technology-driven contextual factors such as time and place, our context elements can be freely defined by a user [19].

Due to the fact that the human memory retrieves information by spreading activation [11], we should offer the basic building blocks to enable the application of a spreading activation algorithm. In the spreading activation theory, concepts are activated by external stimuli. We interpret these external stimuli as an activation of a context element representing a user's current context similar to the approach by Katifori et al. [16]. While Katifori et al. use a task as stimuli, we present a broader solution by using context as stimuli. In order to conform to human memory, a concept and its semantic relationships have to be relevant in a given context. Therefore, the OC2 framework specifies that concepts and their associative links can have a relevance (weight) for a given context element which is illustrated by the blue dashed lines in Fig. 1. Furthermore, as mentioned previously, objects are assigned to their semantic concepts in a context-relevant manner in human memory [10] and therefore our extent links can also have a weighted link to a context element. In contrast to existing work, we provide a context-sensitive categorisation of objects in their semantic concept and further support context-dependent structural links. Finally, navigational links have a relevance factor for given context elements in order to provide users with shortcuts to the most relevant digital or physical artefacts for a user's current context.

4 A Generic Context-aware PIM Metamodel

The implementation of the OC2 framework is based on a domain-specific application and extension of the Resource-Selector-Link (RSL) metamodel for hypermedia systems by Signer and Norrie [5]. The main components of the RSL metamodel are illustrated in light grey in Fig. 2 whereas the necessary extensions are highlighted in blue. Note that the metamodel has been defined by using the OM data model [20] offering the concept of collections of objects as well as associations between objects. A collection is represented as a rectangle with the type of the included objects indicated in the shaded part. Associations and their cardinality constraints are visualised via oval shapes between collections.

The core elements of the RSL model are the `Resources`, `Selectors`, and `Links` collections which are all subcollections of the `Entities` collection. For each information item to be managed, a resource is instantiated and added to the `Resources` collection. We can not only link entire resources but also specific parts of a resource by using `Selectors`. Furthermore, RSL offers a plug-in mechanism to subtype the resource and selector concept for a particular type of media such as web resources, movies or even physical paper. Links in RSL are directed bidirectional many-to-many links between entities. Since both the `HasSource` and `HasTarget` associations have an entity as domain, we can define links between any of the three `Entities` subcollections. Note that the RSL metamodel also provides some user management on instantiated entities. A complete description of the RSL model can be found in [5].

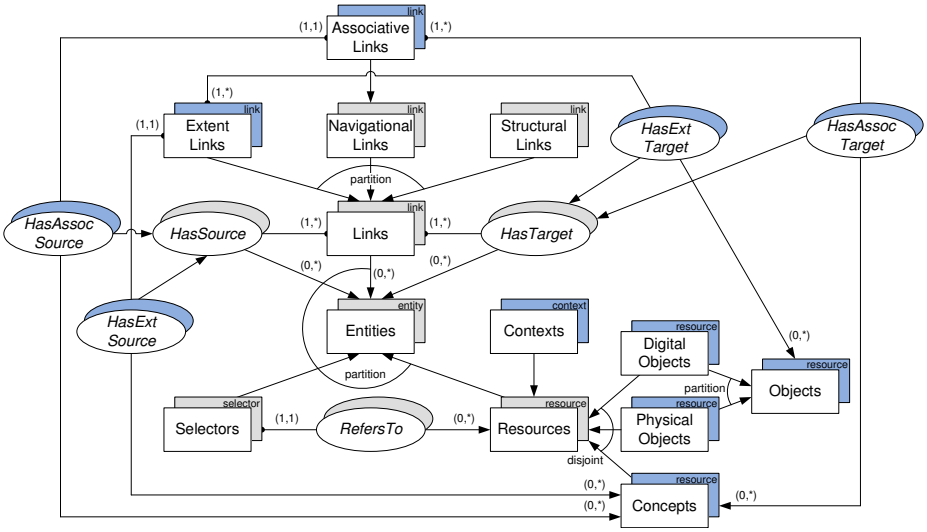

Fig. 2. OC2 framework as domain-specific RSL application

The domain-specific application of the RSL metamodel consists of three main extensions. Our OC2 conceptual framework distinguishes between concepts which are internal to the memory and objects external to the memory. Additionally, objects can be physical or digital objects. Due to the fact that our memory uses the extra information whether an object is physical or digital as a re-finding cue, this has to be taken into account in the modelling process. We therefore extended the `Resources` collection with the `Concepts`, `Physical Objects` and `Digital Objects` subcollections as shown in Fig. 2. A disjoint constraint between these three subcollections indicates that an information item (resource) can only be in one of these subcollections. Since `Physical Objects` and `Digital Objects` are objects external to the memory, we have introduced a more generic `Objects` collection. By modelling `Context` as a subcollection of `Resources`, a context can be linked to other entities and in particular to other contexts. Furthermore, the OC2 framework imposes a distinction between extent and associative links. `Extent Links` are a subcollection of `Links` since they link `Objects` to their semantic concept. The `HasExtSource` and `HasExtTarget` define the objects and the concept of a given OC2 extent link. `Associative Links` are a specific type of navigational links since they represent the navigation functionality between `Concepts`. Similarly to the extent links, an associative link has to participate in the `HasAssocSource` and `HasAssocTarget` associations since an associative link can only have one concept element as source and one or multiple concept elements as target.

In line with our OC2 conceptual framework, we have to define a weight for a specific context on concepts and objects as well as on the four different types of links. Since these elements are in fact subcollections of the `Entities` collection, we model the context relevancy via the `InContext` association as shown in Fig. 3.

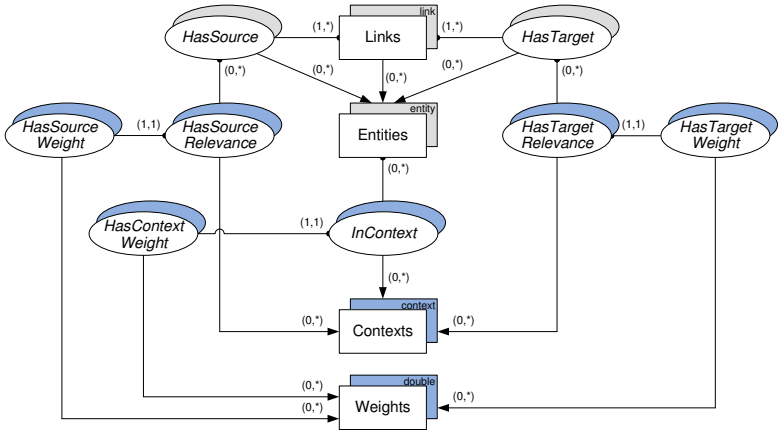

Fig. 3. Integration of context elements in the RSL metamodel

Furthermore, the weight is given through the participation of the `InContext` association and the `Weights` collection in the `HasContextWeight` association. Since the RSL metamodel defines links as many-to-many links, this implies that only the link itself can have a contextual relevancy. We aim for a finer granularity by supporting contextual relevancy at the level of the `HasSource` and `HasTarget` associations. The `HasSourceRelevance` and `HasTargetRelevance` associations together with the corresponding `HasSourceWeight` and `HasTargetWeight` associations are used to the define this relevancy. In this way, each source and target of a link may have their own relevance for a given context.

5 Discussion and Future Work

We have introduced the OC2 conceptual framework and its implementation integrating the strengths of both the unified view and TIM paradigms. While these two paradigms are based on principles of the human memory, they do not provide navigational and structural relationships on the object level. Additionally, they do not support the context-sensitive categorising of objects in their semantic concepts and no context relevancy can be defined on object elements.

Our current implementation focusses on the user-centric aspect to create a conceptual model. Nevertheless, machine-readable conceptual models do have advantages in PIM settings and we investigate the integration of conceptual models created by individual users with a more generic PIM ontology. While Katifori et al. [16] provide a spreading activation algorithm for TIM applications, we plan to extend this algorithm to include the full functionality offered by the OC2 framework and use context as activation element.

We have presented a new approach to PIM system design addressing the problem of information fragmentation and combining the strengths of the unified view and TIM paradigms with the most relevant aspects of the human memory. The OC2 framework has been implemented as a domain-specific application of

the RSL hypermedia metamodel. Furthermore, he have developed a proof of concept semantic desktop application based on the OC2 framework. Last but not least, the presented OC2 framework for context-aware personal cross-media information management systems represents a research platform for innovative ideas in managing information across digital and physical information spaces.

References

1. Bergman, O., Beyth-Marom, R., Nachmias, R.: The Project Fragmentation Problem in Personal Information Management. In: Proc. of CHI 2006, Montreal, Canada
2. Bush, V.: As We May Think. Atlantic Monthly **176**(1) (1945)
3. Catarci, T., Dix, A., Katifori, A., Lepouras, G., Poggi, A.: Task-Centered Information Management. In: Proc. of the 1st DELOS Conference, Pisa, Italy (2007)
4. Whittaker, S., Kalnikaite, V., Petrelli, D., Sellen, A., Villar, N., Bergman, O., Clough, P., Brockmeier, J.: Socio-Technical Lifelogging: Deriving Design Principles for a Future Proof Digital Past. Human-Computer Interaction **27** (2012)
5. Signer, B., Norrie, M.C.: As We May Link: A General Metamodel for Hypermedia Systems. In: Proc. of ER 2007, Auckland, New Zealand (November 2007)
6. Tulving, E.T., Donaldson, W.: Organization of Memory. Academic Press (1972)
7. Collins, A., Loftus, E.: A Spreading Activation Theory of Semantic Processing. Psychological Review **82**(6) (1975)
8. Braisby, N., Gellatly, A.: Cognitive Psychology. Oxford University Press (2012)
9. Barsalou, L.: Ad-hoc Categories. Memory and Cognition **11**(3) (1983)
10. Roth, E., Shoben, E.: The Effect of Context on the Structure of Categories. Cognitive Psychology **15**(1) (1983)
11. Anderson, J.: A Spreading Activation Theory of Memory. Journal of Verbal Learning and Verbal Behavior **22**(3) (1983)
12. Gemmell, J., Bell, G., Lueder, R., Drucker, S., Wong, C.: MyLifeBits: Fulfilling the Memex Vision. In: Proc. of Multimedia 2002, Juan-les-Pins, France (2002)
13. Karger, D., Bakshi, K., Huynh, D., Quan, D., Sinha, V.: Haystack: A Customizable General-Purpose Information Management Tool for End Users of Semistructured Data. In: Proc. of CIDR 2003, Asilomar, USA (January 2003)
14. Sauermann, L.: The Gnowsis Semantic Desktop for Information Integration. In: Proc. of IOA 2005, Tuscon, USA (February 2005)
15. Katifori, V., Poggi, A., Scannapieco, M., Catarci, T., Ioannidis, Y.: OntoPIM: How to Rely on a Personal Ontology for Personal Information Management. In: Proc. of ISWC 2005, Galway, Ireland (November 2005)
16. Katifori, A., Vassilakis, C., Dix, A.: Ontologies and Brain : Using Spreading Activation Through Ontologies to Support Personal Interaction. Cognitive Systems Research **11** (2010)
17. Bergman, O., Beyth-Marom, R., Nachmias, R., Gradovitch, N., Whittaker, S.: Improved Search Engines and Navigation Preference in Personal Information Management. Transactions on Information Systems **26**(4) (2008)
18. Jones, W., Phuwanartnurak, A., Gill, R., Bruce, H.: Don't Take My Folders Away!: Organizing Personal Information to Get Things Done. In: Proc. of CHI 2005, Portland, USA (April 2005)
19. Bellotti, V., Edwards, K.W.: Intelligibility and Accountability: Human Considerations in Context-Aware Systems. Human-Computer Interaction **16** (2001)
20. Norrie, M.C.: An Extended Entity-Relationship Approach to Data Management in Object-Oriented Systems. In: Proc. of ER 1993, Arlington, USA (December 1993)