Personal Knowledge Management with Semantic Technologies

Max Völkel, FZI Karlsruhe, Haid-und-Neu-Str. 10-14, Karlsruhe, Deutschland, voelkel@fzi.de

Sebastian Schaffert, Salzburg Research Forschungsgesellschaft mbH, Jakob Haringer Straße 5/III, A-5020 Salzburg, Austria, sebastian.schaffert@salzburgresearch.at

Eyal Oren, Digital Enterprise Research Institute (DERI), National University of Ireland, Galway, Ireland, eyal.oren@deri.org

Keywords: *knowledge, cost-benefit, PKM requirements, gradual formalisation, PIM, note taking, personal wiki, semantic wiki*

1. Introduction

Managing and enabling knowledge is a key to success in our economy and society (Wenger, 2002, p. 6). The problem of knowledge management can generally be tackled from two sides: top-down and bottom-up. Many approaches have been taken from the top down, in which the organisation aimed to better manage their internal knowledge by installing central knowledge repositories. Many of these systems were less accepted than expected (Braganza, 2002). Along with the Web 2.0 notions of user-provided content and collective intelligence, more bottom-up approaches to knowledge management were developed.

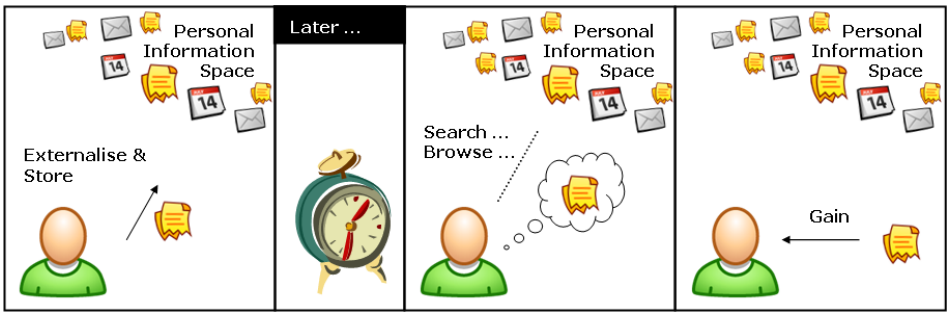

Figure 1: The Personal Knowledge Management dilemma.

In this chapter we describe an individual-centric, bottom-up approach to *personal knowledge management* (PKM). PKM is the individual management of knowledge from a subjective perspective.

Codifying knowledge. Although knowledge cannot be externalised as such (knowledge exists in the interpretation and experience of individuals), people can still *codify* knowledge as information: symbols with meaning (semantics). In this context, symbols are mostly understood as words although images could be used as well. Personal knowledge management involves the management of such information symbols, and by taking into account the semantics of these symbols, tools can support individuals to a larger extent.

Personal insights and associations are very hard to capture, yet worth to keep. It is unclear to the person writing down the notes when and in which context a particular note will be relevant again. Even worse: when we make personal notes, we often do not know if we will ever retrieve those notes again, nor do we know which gain we will get from these notes. Therefore we estimate intuitively the expected cost of externalisation and storage (C_E), cost of search (C_S) and the expected gain (G). Our goal is to have a positive balance, hence $G > C_E + C_S$. Figure 1 shows the general personal knowledge management dilemma: We never know if our costs of authoring and retrieval are worth the effort.

Cost of authoring. A rather simple form of authoring is writing plain text. In a next step, a user can add formatting to the text, which gives it a structure, e. g. different levels of headlines imply a tree structure. Upon retrieval, one can than more easily navigate to the relevant part of a document. Furthermore, a user can add cross-references to other parts of the same text or to other documents. In academic works such cross-links enhance the chance for successful retrieval; e. g. a citation to another work makes the other work more likely to be found. Each step of externalisation described so far lowers the cost for retrieval (C_S) but comes at the expense of higher externalisation costs (C_E). Articulating semantic statements about concepts or semantic links between concepts has high costs in many tools (e.g. ontology editors such as Protégé), since the user must change his familiar user interface paradigm (document and text editing) for a different paradigm (ontology editing) and must fully formalise all content or loose the relation between the concepts in the document to the concepts in the ontology.

The cost gap. To sum up the situation, current tools offer two primary choices: On the one hand, one can make very "cheap" notes, which have little structure and thus have a low chance to be found other than through a keyword match on the full-text. This low-cost strategy (comparable to e.g. post-it stickers) does not scale, if many documents about the same topic exist. More formally, the cost of externalisation (C_E) is low, but the costs of search (C_S) are high, as one needs many searches before finding a note again – if it is found at all.
On the other hand, one can use a custom database application or ontology editor with high costs of externalisation to lower the search costs. The high costs, due to the paradigm switch and the strict constraints of the editing environment, are often too high for everyday information management as one cannot sit down and design a database or ontology when trying to externalise some fact obtained e. g. from an email.
This PKM cost gap can be bridged in two ways: One can lower the cost of search on collections of loosely structured notes. Desktop search engines do this by offering fast full-text search. Other approaches like ontology learning try to extract structures from unstructured data. On the other hand, one can lower the cost of authoring information with explicit structure and/or semantics. This will also lead to lower search costs, as more structured data leads to higher precision (Kando, 1997). The second approach, lowering the cost of authoring structured knowledge, is the one we take in this chapter.

Outline. This chapter first presents general requirements for PKM obtained from a literature study (Sec. 2). Then existing "classic" (non-semantic) PKM tools are reviewed (Sec. 3). In Sec. 4 the notion of a *semantically-enhanced PKM tool* (SPKM, c. f. Fig. 3) is introduced as a novel approach to the PKM problem. Sec. 5 reviews existing prototypes of SPKM tools and lists their drawbacks. Sec. 6 presents personal semantic wikis as a best-of-breed implementation of the SPKM vision. In this approach we do not use wiki technology as a community platform but as a personal authoring environment. Sec. 7 discusses future trends and concludes the chapter.

2. Requirements for Personal Knowledge Management

We live in a knowledge society in which people have access to tremendous amounts of information and communicate with others all over the world. Although most of this information and communication is digital, people still predominantly use traditional analogue techniques: several studies (Jones, 1997; Blandford, 2001; Campbell, 2003) show that simple to-do-lists, paper calendars, address books, and diaries are the most commonly used tools, all of which are analogue paper-based tools. Although email is increasingly used for some personal information tasks (Bellotti, 2004; Whittaker, 2006) it is mostly unstructured (Dai, 2005).

Requirement 1: Acceptable Costs. As argued in Sec. 1, a working PKM solution must have acceptable costs both for externalisation and for search. Based on an exhaustive literature analysis of information management and knowledge work (Oren, 2006a),[1] we derive six additional requirements for personal knowledge work:

Requirement 2: Focus on the individual. Many works on information and knowledge management indicate that "knowledge is ultimately created by individuals" (Polanyi, 1958). This motivates our conceptual shift from organisational knowledge management to personal knowledge management. Individuals are personally committed to knowledge creation (Nonaka, 1994) but this commitment relies on their intentions (understanding and actions of individuals) and on their autonomy (self-motivation and freedom to pursue their intentions). Personal autonomy is thus crucial and individuality and freedom should be supported by tools.

Traditionally, three types of office workers can be distinguished (Kidd, 1994): the knowledge worker (who creates new knowledge), the communication worker (who amplifies information and connects people), and the clerical worker (who manages documents). All three of them perform administrative tasks, which are repetitive, structured and document-driven, and research tasks, which are flexible, unstructured and information-driven, but do so in different proportions: clerical workers perform mostly administrative tasks, while knowledge workers mostly do

4

1 The literature survey started in Nov 2006 from recent issues of relevant conferences (ESWC, ISWC, CHI, and SIGIR conferences and the CACM journal) and followed both backward and forward citations.

research tasks. These research tasks usually have no structured procedures and workers have no clear idea of their next steps. The unstructured nature of their work makes it difficult to support knowledge workers in general, and highlights the need for their personal autonomy and freedom.

Requirement 3: Forget rigid classification. Strict hierarchical classification does not work for many individuals (Hearst, 2006); therefore we see a rise of more flexible structures: tagging and ontologies. Many people categorise information mainly by context (Lansdale, 1988) and retrieve information based on contextual clues (associations). Contrasting, most classifications in the computer are hierarchical (e.g. file folders) which leads to usability problems (Randall, 2001). Although research has shown that people prefer category-based search over keyword-based search (Yee, 2003), using *strict* hierarchies of categorisation worsens the results since they do not work well for most people. The activities of knowledge workers are highly unstructured; the documents intended use (relevant for later retrieval) is often unclear at the time of filing (Bondarenko, 2005; Kwaśnik 1983).

Requirement 4: Follow the links. Although desktop search greatly helps individuals in retrieving their own files, users actually prefer to find information not by searching and jumping to it but by orienteering and browsing between related items (Teevan, 2004). Such exploration techniques are not only preferred in an unfamiliar domain (or when browsing somebody else's data) but also when investigating own personal information spaces (Schraefel, 2006; Cutrell, 2006). This preference stems from the fact that people often have difficulties to unambiguously specify what they are looking for and thus prefer browsing over searching (Barreau et al, 1995). Furthermore, human memory can greatly benefit from context information which is enabled by exploratory browsing between related items. Users clearly benefit from being able to follow the links in their data and browse their information space. The ability to automatically identify and present links between relevant data requires the data to be interlinked and semantic relations between data elements to be identified.

Requirement 5: Remember the context. Contextual information is crucial in human information retrieval. People categorise information based on its context and recalling therefore also benefits greatly from contextual information (Lansdale, 1988). In a personal classification system, the classification of a document is strongly affected by its intended use and context (Kwaśnik, 1983); that means that we should capture such document context to facilitate retrieval. Dumais (2003) reports that rich contextual information such as people names and dates, i.e. the personal and temporal dimensions, provide beneficial cues for retrieval in a personal information system and seem even more important than standard ranking functions. Which context information to use in knowledge management is still an open question, often used dimensions are the five "who, what, when, where, how" or the "who, what, when, where, why" dimensions (Ranganatan, 1962).

Requirement 6: Value the power of paper. Despite all the progress of our PC desktops, physical paper remains a crucial tool for most people. Even the advent of the internet has only increased the amount of printed paper (Sellen, 2003, p. 7). This seeming paradox (a continuous rise of paper usage together with an increased use of digital media) is understandable since (i) the Internet gives people access to more and more information which often ends up printed, since (ii) printing technology is evolving, making printing easier, faster, and better, and since (iii) paper is extremely well-suited for authoring, reading, and reviewing documents (Sellen, 2003, p. 13 - 15). Research has shown that reading stimulates writing (Adler, 1998) and that paper is very well-suited for combined reading and writing because of its possibilities for flexible, spontaneous annotation (draw or write with a pen), quick navigation between documents (simply flip a page) and spatial layout to form a mental model of documents (O'Hara, 1997). A recent study showed that 85% of reading activities were paper-based although almost all information was available to people digitally, on their computer. It seems that digital technologies, rather than replacing paper, shift the point where paper is used: from publishers to the individuals which download, print, and read documents. Also, paper fits the requirements for individual autonomy and freedom very well: compared to digital media, paper puts almost no restrictions on the content that can be written on it (sketches, text, or both).

Requirement 7: Keep it simple. Perhaps the most important requirement shown in the literature is the simplicity of tools. The adoption rate of digital technology for personal information (even in todays age of ubiquitous mobile phones and PDAs) is quite low (Jones, 1997). This relatively low adoption may be explained by the unawareness of most users with the features of their existing tools; most people for example are even unaware of long-existent features such as automatic email filtering (Bondarenko, 2005). Research has also shown that people use many tools quite differently than expected; for example, people use folders on their computer not only to organise their files but also to remind them of tasks, to decompose problems, and to plan their work (Jones, 2005; Barreau, 1995). Simple tools do not constrain users workflows but support their autonomy and are easy to understand and use.

Requirement 8: Keep the flow. As an additional requirement from the authors own observations, a support for personal work processes is crucial: One must allow the information and structures to travel from one application in a workflow to the next. We illustrate this requirement with an article writing scenario: Writing is a central activity for all kinds of knowledge workers. They have to compile reports or write articles for scientific journals. Writing an article includes reading related work – and often also taking personal notes on it. Then an outline or other kind of high-level structure is usually created (Esselborn-Krumbiegel, 2002). This structure is then refined to the argument-level and fleshed out into a linear text. Next, the text is partially rewritten and references to figures, other sections and literature are added. Eventually the text is typeset and thus the final layout is created.

3. State of the art in "classic" PKM approaches and their problems

Several different approaches for personal knowledge management exist. In the following, we concentrate on popular tools used for this task: weblogs, wikis, mind-mapping, and personal information management (PIM) tools. We also discuss common problems for personal knowledge management in these tools.

Weblogs. A weblog is a specific kind of personal website where content is published in the form of small articles and displayed in reverse-chronological order (latest article first), similar to a personal but

publically readable diary. Readers can leave comments to articles. A particularly interesting feature of weblogs is the so-called *Trackback*-function which allows a weblog to become aware of other weblogs that refer to its articles, and the so-called *Blogroll* which lists or even syndicates other weblogs that are interesting to the owner. These two features together allow forming the so-called *Blogosphere*, which describes the network formed between personal weblogs by hyperlinks.

Weblogs are primarily interesting not because of their technology but rather because of the social interactions they enable by commenting and referring to each others weblog articles. A weblog unfolds its true value only when connected with others.

In personal knowledge management, weblogs are mainly used to record ideas and other *knowledge bits* quickly. Since articles are usually publically readable, external users can comment on them and thus help improving the quality. On the other hand, private, confidential, or not yet very elaborate ideas should often not be publically available; some weblogs support this by restricting the group of people that can read articles. In weblogs, access to knowledge is primarily oriented on the time-axis, i.e. browsing through articles primarily by publication time. Some weblogs allow to associate articles with additional categories that support a more selective browsing through the articles.

Wikis. A wiki (Leuf, 2001) is essentially a collection of web pages connected via hyperlinks. Many different wiki systems exist, but they commonly have a straightforward interface for editing content with a simplified syntax that makes it very easy to set hyperlinks to other pages within the wiki. Therefore, content in a wiki is usually strongly connected via hyperlinks. Furthermore, editing of content in wiki systems is web-based and access is often unrestricted or at least hardly restricted. Most wiki systems also provide a rollback mechanism for reverting back to previous versions in case of accidental or undesired changes. Wikis are used in many areas, like encyclopaedia systems (e.g. Wikipedia), as personal or group knowledge management tools, as collaboration tools, or in collaborative learning environments.

Concerning personal knowledge management, wikis can be used as a very flexible and powerful tool (Wagner, 2005). The simple wiki syntax allows to easily creating content with a low technical barrier, and since it allows entering free text, the user is free to choose his own process and workflow. Easy hyper-linking allows connecting knowledge for

navigation purposes and improves the retrieval of relevant knowledge. Cleverly used links and back-links allow also for hierarchies, and categorisation. The plug-ins available for many wiki systems provide additional means of entering and visualising content, e.g. as a calendar, as a task list, etc. Content can easily be made publically available, and most wikis allow restricting access to content to specific user groups. It is also possible and useful to install a wiki on the user's personal computer. There even exist projects that aim to install a personal wiki on a USB stick.

The main disadvantage of wikis with regard to knowledge management is that whereas creating content is easy, retrieval usually is not. Users are usually limited to full text search and hyperlinks, which often does not allow to find the relevant content, e.g. when synonyms are used or when the hyperlinks are not set properly. Also, wikis that are not maintained very well tend to be rather chaotic, which further aggravates the problem. For note-taking, wikis typically do not allow users to indicate the context of notes (notes are usually related to projects, tasks, responsibilities, email communication, etc.). Therefore users cannot query their personal notes e. g. to find all notes about a certain email discussion or navigate between two notes related to the same project. All navigation links have to be added explicitly because current applications do not support the personal meaning (semantics) of the notes.

Mind-Mapping. Mind-mapping and concept mapping are methods for structuring ideas, words, topics, or other items using graph representations. Centred around a key concept, further items are arranged around the centre concept. Mind maps are used in many different areas, e.g. brain storming and collection of ideas, note taking, learning, summarising, structuring, etc (Haller, 2002). Mind-mapping is not bound to computers; the method is much older and can equally or even superiorly be used on paper.

Mind-mapping is a traditional tool for personal knowledge management. Its main advantages lie in the structuring of content and the quick overview one gets when looking at a mind-map. The disadvantages are the limited scalability (a sheet of paper / the computer screen only suffices for the visualisation of relatively few concepts at once) and the fact that a mind-map is always centred on a single concept.

Personal Information Management (PIM) tools. Personal Information Management tools usually offer ways to manage a calendar, appointments, tasks, and addresses. Often, email and simple note taking are also included in the definition. Popular PIM applications are MS Outlook, KDE Kontact, and Gnome Evolution. Recently gaining importance is the combination of the Google services Google Mail, Google Calendar, and Google Docs.
The main advantage of PIM systems is that they integrate the different knowledge sources of office workers into a single interface. Unfortunately, most PIM tools are simply a collection of different systems and the actual information integration is not very deep. Neither is it possible to query the information nor are the relations between information items explicitly represented.

3.1 Summary of existing approaches

Weblogs are meant as a tool for the individual (Req. 2) for publishing and do not work well for restructuring knowledge. Strengths of weblogs and wikis are handling of links (Req. 4) and simplicity (Req. 7), but weblogs lack acceptable search costs (Req. 1), context of notes (Req. 5) and import/export support (Req. 8). A comparison table can be found in Fig. 2.
A *wiki* is great to keep a big repository of linked snippets, but it is hard to get an overview of the stored content. Wikis focus mainly on collaborative usage (Req. 2); their interlinked nature allows finding content by associative browsing (Req. 1).

		Weblog	Wiki	MindMap	PIM	iMemex	Art. Mem.	Haystack	NEPOMUK	PSW
Acceptable Costs	Req. 1		~						~	+
Focus on the individual	Req. 2	~		+	+	+	+	+	+	+
Forget rigid classification	Req. 3	~	~				+	+	+	+
Follow the links	Req. 4	+	+				+	+	+	+
Remember the context	Req. 5							+	+	
Value the power of paper	Req. 6			~						
Keep it simple	Req. 7	+	+	+	~	~				+
Keep the flow	Req. 8			~	~	+			+	~

Figure 2: An overview of existing PKM and SPKM tools w.r.t. requirements.
Legend: + = feature present, ~ = feature partially present

Mind-map tools are probably the only tools which have a paper-like work style (Req. 6). They are simple to use (Req. 7) and often have export functions from mind-map to textual outline (Req. 8), however usually additional links and icons are lost and re-import is impossible.

PIM tools are good for the narrow domains they are designed for, but fall short when a set of structured notes has to be managed (Req. 1). At least they often can export structured content in usable formats (appointments, contact data, etc.).

Overall, acceptable search costs and integration into workflows are the weakest points of existing PKM tools.

4. Introducing Semantic Personal Knowledge Management (SPKM)

We define an SPKM tool as a PKM tool that supports personal knowledge management by managing additional *explicit semantic representations*. An SPKM tool allows a user to maintain a personal explicit representation of his world view. Instead of large document collections a knowledge worker can centre his work on a single personal knowledge model (Völkel, 2007).

For an SPKM tool to be useful, it should offer a good ratio between the effort it takes to externalise and (partially) formalise knowledge and the benefit gained by enhanced query answering and browsing abilities.

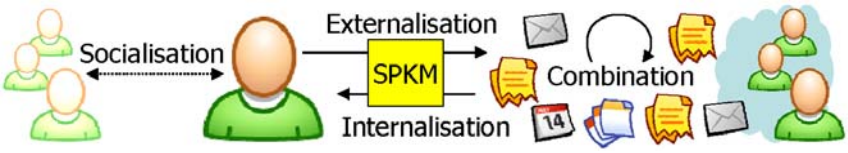

Figure 3: Semantic Personal Knowledge Management supporting externalisation (authoring) and internalisation (learning) of personal knowledge.

Fig. 3 depicts the knowledge processes described in (Nonaka, 1995) as they happen in PKM. Socialisation (direct face-to-face) becomes rare in an always online world. As most communication with oneself (PKM) and others (collaboration) happens through digital channels, efficient externalisation and internalisation is becoming the PKM bottleneck.

By mapping structural features of a document to semantic statements of the underlying knowledge model, the computer could assist the user much better. Tables are a particular example of structures with implicit semantics, which are quite difficult to formalise (Pivk, 2005).

Existing tools do not handle semantics of structures: For example, John might write a document, and mark some parts in red; these markings mean "statements do be discussed with Alice". If he exchanges the document he needs to communicate these ad-hoc semantics (discuss with Alice) implied by structures (red) as well. Currently, he would do so separately, by email or phone. Making the semantics of structures explicit allows for better search ("what has to be discussed?") and more efficient knowledge exchange.

A truly usable SPKM system should allow the user to use the different editing paradigms (wiki, document, mind-map) at wish – *on the same data*. Existing PIM objects (tasks, addresses, appointments) and desktop resources (files, folders) have to become first-class citizens, so that they can be annotated and interlinked as well. The importance of flexible authoring methods has been recognised in knowledge engineering (Fensel, 1994). An SPKM tool supports the user in these processes.

Each individual uses his own SPKM tool as a personal knowledge repository; he benefits personally from this system by having better retrieval of his knowledge. His personal SPKM tool is connected to other applications and in a loose peer-to-peer fashion to other SPKM tools. This network allows individuals to combine their knowledge through sharing and exchanging. If designed correctly, SPKM tools are easy to use and cognitively adequate for modelling and refactoring knowledge; enable information sharing and reuse within the personal knowledge space, with other knowledge workers, and with existing information systems. And they enable structured access to personal and collaborative knowledge through queries, categorisation, and associative browsing.

5. Existing SPKM tools and their drawbacks

Currently, not many SPKM tools exist. A brief overview is presented in this section.

iMemex (Dittrich, 2006a) calls itself a "A Platform for Personal Dataspace Management" and is centred around a data model (IDM (Dittrich, 2006b) for the semantic desktop. IDMs core goal is to cross the boundaries between files and resources in files. All items, be it folders, files, sections or paragraphs are mapped into a single model. IDM emphasises performance, expressivity, lazy data sets and data streams. There is no publically available end-user tool, but first prototypes focusing on search will be available soon. However, authoring of structured data seems not to be the scope of iMemex.

Another existing SPKM system is *Artificial Memory* (Ludwig, 2004), which uses a tree-like structure in a web interface to interact with the user. It represents semantics with icons and allows the user to add new statements. However, unlike a semantic wiki, the user does not have the freedom to *not* structure his writings.

The European project NEPOMUK aims to bring forward the state-of-the-art by developing a *social semantic desktop*. Within NEPOMUK a data model for structuring knowledge in explicit yet vague relations is developed as *Conceptual Data Structures* (CDS) (Völkel, 2006a). NEPOMUK builds on *Gnowsis* (Sauermann, 2005) which included a semantic wiki from the start to allow a user to make arbitrary statements. In Gnowsis, a central ontology could be browsed and edited in parallel, using the same concepts as those of the semantic wiki.

Haystack (Quan, 2003) is another semantic desktop, which emphasizes browsing through and interacting with a heterogeneous collection of RDF resources. However, Haystack does not foresee to create completely unstructured items first, and thus places the burden on the user to formalise her knowledge in one big step.

Summary: Although all these tools focus well on the individual (Req. 2, c.f. Fig. 2) and offer plenty structuring abilities (Req. 3, 4, and 5), they are neither simple to use (Req. 7) nor have they acceptable costs for externalisation (Req. 1).

6. Personal Semantic Wikis (PSW)

Semantic wikis enhance classical wikis with the ability to author and use formal semantics: in the same wiki style of free text editing, semantic statements can be added. They can describe the page or parts thereof semantically. A semantic wiki offers a uniform way to work with all levels for formality (text, structured text and formal statements). As shown in Fig. 4, they allow users to structure and annotate their data but they do not force them to do so. Using enhanced wiki syntax (plain-text with few mark-up commands and few semantic annotation commands) has several benefits: (a) most users are used to text typing and avoid familiarising with yet another user interface; (b) existing skills for text manipulation (e. g. copy and paste of text blocks) are leveraged to edit a document structure; (c) users refine interactively the input until the result matches the intended structure; (d) wikis allow soft transitions between knowledge layers including free-text, therefore no knowledge of any syntax is required to start authoring; (e) wiki syntax has little layout

options, forcing the user to focus on the structure and the content; (f) text is in general a faster method of entering semi-structured information than graphical approaches.

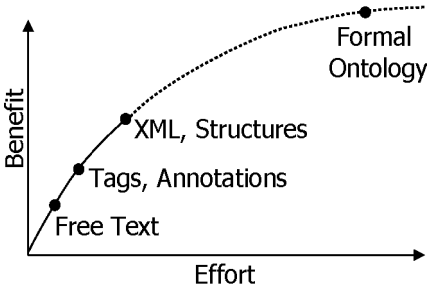

Figure 4: Semantic wikis handle a continuous spectrum of knowledge types.

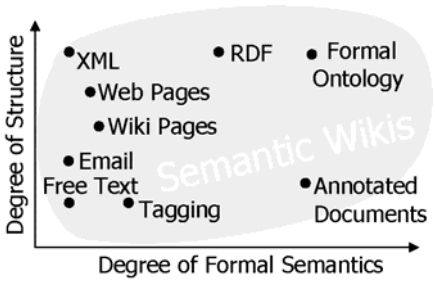

Figure 5: Semantic Wikis can author in different paradigms simultaneously.

Several semantic wikis exist today. A full-day workshop on semantic wikis was held in 2006 (Völkel, 2006b). By using the authoring capabilities of semantic wikis for personal knowledge management – ignoring wiki community aspects – we get a *personal semantic wiki*. One of the systems of particular interest is the system SemperWiki (Oren, 2006b), which has been designed for personal use and allows the user to integrate simple wiki texts with formal semantic annotations. SemperWiki can be used as a traditional personal wiki for note-taking, but can also be augmented with simple semantic annotations. These annotations can be shared with others (using RDF, the standard semantic web language) but more importantly, enable improved navigation to related items and intelligent queries. Arbitrary annotations can be made and the tool can be customised for usage in a certain domain that use specific terms and ontologies. Such domain ontologies can contain background knowledge (e.g. in the biology domain possibly a taxonomy of biological species and their relations) that can be used in conjunction with the manually entered wiki annotations. SemperWiki integrates with the Gnome (Linux) desktop: it is available across the whole desktop with a single keypress, it automatically saves all changes and updates all navigation links instantly.

PSWs allow users to explicitly indicate the semantics of information items and their relations to other information objects. Linking to desktop objects is often still limited or difficult to use. Semantic relations

between information objects allow for contextual navigation even between pages that have not explicitly been related to each other but that can be related based on their meaning, such as two notes about the same project. The flexibility and freedom of wikis is maintained: users can annotate notes with their own terms and can later add background knowledge that relates their terms to those used by others in their working environments, enabling those others to reuse their annotations without requiring a-priori agreement on terms or classifications.

However, even semantic wikis do not fully exploit the opportunities of supporting the user. They have currently a severe lack of refactoring tools, a feature found in most advanced IDEs today, which allows renaming or moving items while maintaining all references. Second, semantic wikis do not exploit the structural parts of a page and hence cannot return e.g. short sections relevant to a query but instead return only full documents.

6.1 Scenario: Writing an Article using a PSW

Writing is a central activity for all kinds of knowledge workers. They have to compile reports or write articles for scientific journals. Writing an article includes reading related work – and often also taking personal notes on it. Then an outline or other kind of high-level structure is usually created. This structure is then refined to the argument-level and fleshed out into a linear text. Next, the text is partially rewritten and references to figures, other sections and literature are added. Eventually the text is typeset and thus the final layout is created.

First, the related work can be managed in a PSW. Different from standard reference management tools, a PSW allows annotating and relating references. Second, personal notes on the topic at hand can be managed, structured, refined and related in the same environment. The benefit of using a PSW is the unified management of heterogeneous artefacts in one environment with one language. Here, language refers to the names used for relation and entity types. A consistently used vocabulary with clear semantics – even if only known to the user in the first place – is of great value in information processing, e.g. for posting queries. Such queries can then return a heterogeneous item collection. A second benefit is, that a PSW allows users to externalise the semantics of relation and entity types. Once a relation has e. g. been defined as transitive or symmetric, further relations can be computed by inference, leading to better search results. The usual top-down refinement process in

article writing is well supported by PSWs: As entities are supposed to be fine-grained and referenceable, items can be re-used. An outline evolving into a draft document can thus be represented by a hierarchy of interlinked items. Smaller items allow for better re-use in other contexts, and semantic annotations help to find the items needed. In collaborative work, multiple users are able to work together explicitly on argument and text structures, something rarely possible with classic tools, i. e. only a few dedicated argumentation management tools exist. Another possibility for closely collaborating parties is the direct exchange of model parts from one PSW to another, without the need to create a document at all. The other user can than import and link the other's knowledge model. A more detailed description how and why a cultural shift from a document-centred to a model-centred society could happen is explained in (Völkel, 2007).

6.2. Evaluating Personal Semantic Wikis

The main advantage of PSW compared to other PKM and SPKM approaches is the use of stepwise, gradual formalisation, leading to acceptable costs (Req. 1) both for externalisation and search. PSW offer the same support for classification and links (Req. 3 and 4). Until PSW are integrated with semantic desktops, their ability to track context is still limited (Req. 5). PSW are simple to use (Req. 7), but have no clear way of working together with paper (Req. 6). As the resulting knowledge base is partially structured, an export to and import from existing formats is partially possible (Req. 8).

7. Conclusion

In this chapter we have introduced the concept of personal knowledge management enhanced with semantic technologies (SPKM). We presented results from an exhaustive literature study on requirements for information management. Then we analysed existing approaches for PKM and evaluated them with respect to the requirements. We found existing tools lead to a high degree of information fragmentation: structures made explicit in one tool become implicit again in the next tool. There are currently no tools allowing the user to use the different editing paradigms (wiki, document, mind map) at wish – on the same data. Existing PIM objects (task, address, appointment) and desktop resources (files, folders) can currently not be annotated and linked.

7.1 Outlook

A semantic wiki by itself does not address the requirements: detection of user context and importance of ordinary paper, which are open research issues that are actively worked on in for example ubiquitous computing and in electronic paper. Both are likely to minimise the gap from real to digital world. Advances in scanner technology, optical character recognition (OCR) and natural language processing (NLP) are also likely to integrate paper tighter into the digital world. For privacy and cultural adoption it seems the only way is to educate people in the same manner as we teach them reading, writing and calculating. On future agendas we will probably also find the subjects semantic modelling and data privacy.

7.2 Future Research Trends

An SPKM tool should analyse and cluster the information to find related items. All information should be stored in a content repository that can apply rules with background knowledge, and one should be supported in reusing existing terminology. SPKM systems should be connected to each other forming a network of knowledge; this network allows sharing knowledge with others. The interface of the SPKM system should be able to be personalised to each user and adapt to his preferences.

Import and export with existing and future non-semantic tools is an issue. Also the integration between different semantic applications – although easier – is still a research issue. Another concern is privacy. If people continue to publish a lot of their semantic data, they will be surprised to learn how easy such information can be aggregated later on. Non-technical issues such as semantic data handling as a cultural technique, the gap from paper to the digital world and simplicity in use are still open.

A critical success factor of SPKM solutions will be the soft introduction of formalization: A user must e.g. always have the option not to formalize at all. A gradual, step-wise formalization of the existing knowledge structures should let the user decide how much effort (K_E) he is willing to spend. At any point in time, more formalization effort has to be rewarded with better querying and browsing capabilities (lower K_S). The Achilles heel of semantic tools will be the authoring part. Only if the end-user has an easy-to-use environment to create semantic structures on the fly and explicitly, semantic technologies can improve searching and browsing. Of course, this approach should be accompanied with automatic generation of semantic data from existing content.

References

Adler, A., Gujar, A., Harrison, B. L, O'Hara, K., & Sellen, A. (1998). A diary study of work-related reading: Design implications for digital reading devices. In *CHI*, (pp 241-248).

Barreau, D., & Nardi, B. A. (1995). Finding and reminding: File organization from the desktop. *SIGCHI Bulletin, 27(3)* (pp. 39-43).

Bellotti, V., Dalal, B.,Good, N., Flynn, P., Bobrow, D. G., & Ducheneaut, N. (2004). What a to-do: Studies of task management towards the design of a personal task list manager. In *CHI*, (pp. 735-742).

Blandford, A., &Green, T. R. G. (2001). Group and individual time management tools: What you get is not what you need. *Personal and Ubiquitous Computing, 5(4)* (213-230).

Bondarenko, O., & Janssen, R. (2005). Documents at hand: Learning from paper to improve digital technologies. In *CHI*, (pp. 121-130).

Braganza, A., & Mollenkramer, G. J. (2002). Anatomy of a failed knowledge management initiative: Lessons from PharmaCorp's experiences. *Knowledge and Process Management, 9(1)*, (pp. 23-33).

Campbell, C., & Maglio, P. (2003). Supporting notable information in office work. In *CHI*, (pp. 902-903).

Cutrell, E., & Dumais, S. T. (2006). Exploring personal information management. *Commun. ACM, 49(1)*, 50-51.

Dai, L., Lutters, W. G., & Bower, C. (2005). Why use memo for all? restructuring mobile applications to support informal note taking. In *CHI*, (pp. 1320-1323).

Dittrich, J.-P. (2006). iMeMex: A Platform for Personal Dataspace Management. *SIGIR PIM*.

Dittrich, J.-P., Marcos Antonio Vaz Salles (2006). iDM: A Unified and Versatile Data Model for Personal Dataspace Management. *VLDB*: (pp. 367-378), Seoul, Korea.

Dumais, S., Cutrell, E., Cadiz, J. J., Jancke, G., Sarin, R., & Robbins, D. C. (2003). Stuff I've seen: A system for personal information retrieval and re-use. In *SIGIR* (pp. 72-79).

Esselborn-Krumbiegel, H. (2002), *Von der Idee zum Text. Eine Anleitung zum wissenschaftlichen Schreiben*., Utb.

Fensel, D., et al (1994). Integrating semiformal and formal methods in knowledge-based systems development. In *Proc. of the Japanese Knowledge Acquisition Workshop* (JKAW-94) (pp. 73–89), Hitachi, Japan.

Haller, H. (2002). *Mappingverfahren zur Wissensorganisation*. diploma thesis at Freie Universität Berlin.

Hearst, M. A. (2006). Clustering versus faceted categories for information exploration. *Communications of the ACM, 46(4)*.

Jones, S. R., & Thomas, P. J. (1997). Empirical assessment of individuals' "personal information management systems". *Beh. & Inf. Techn., 16(3)* (pp. 158-160).

Kando, N. (1997). *Text-level structure of research papers: Implications for text-based information processing systems*, in `BCS-IRSG Annual Colloquium on IR Research', Workshops in Computing, BCS.

Kidd, A. (1994). The marks are on the knowledge worker. In *CHI* (pp. 186-191).

Kwaśnik, B. H. (1983). How a personal document's intended use or purpose affects its classification in an office. In *SIGIR* (pp. 207-210).

Lansdale, M. (1988). The psychology of personal information management. *Applied Ergonomics, 19(1)* (pp. 55-66).

Leuf, B., & Cunningham, W. (2001). *The Wiki Way: Collaboration and Sharing on the Internet*. Addison-Wesley.

Ludwig, L., O'Sullivan, D., & Zhou, X. (2004). *Artificial Memory Prototype for Personal Semantic Subdocument Knowledge Management (PS-KM)*. ISWC demo.

Nonaka, I. (1994, February). A dynamic theory of organizational knowledge creation. *Organization Science, 5(1)* (pp. 14-37).

Nonaka, I., & Takeuchi, H. (1995). *The Knowledge-Creating Company*. Oxford University Press, New York.

O'Hara, K., & Sellen, A. (1997). A comparison of reading paper and on-line documents. In *CHI*, (pp. 335-342).

Oren, E. (2005, November). SemperWiki: a Semantic Personal Wiki. In *Proceedings of the ISWC Workshop on the Semantic Desktop*.

Oren, E. (2006, November). An overview of information management and knowledge work studies: Lessons for the semantic desktop. In *Proceedings of the ISWC Workshop on the Semantic Desktop*.

Oren, E., Völkel, M., Breslin, J. G., & Decker, S. (2006, September). Semantic wikis for personal knowledge management. In *Proceedings of the International Conference on Database and Expert Systems Applications (DEXA)*.

Polanyi. M. (1958). *Personal Knowledge: Towards a Post-Critical Philosophy*. Routledge & Kegan Paul Ltd, London.

Pivk, A., Cimiano, P., & Sure, Y. (2005, October). From tables to frames. *Journal of Web Semantics, 3(2)* (pp. 132-146).

Quan, D., Huynh, D., & Karger, D. R. (2003). Haystack: A platform for authoring end user semantic web applications. In *Int. Semantic Web Conf.* (pp. 738-753).

Randall, D. W., Bowker, G., & Leigh Star, S. (2001). Sorting things out: Classification and its consequences - Review. *Computer Supported Cooperative Work, 10(1)* (pp. 147-153).

Ranganathan, S. R. (1962). *Elements of library classification*. Bombay: Asia Publishing House.

Sauermann, L. (2005). The gnowsis semantic desktop for information integration. In *Wissensmanagement* (pp. 39-42).

Schraefel, M. C., Wilson, M., Russell, A., & Smith, D. A. (2006). mSpace: Improving information access to multimedia domains with multimodal exploratory search. *Communications of the ACM, 49(4)* (pp. 47-49).

Sellen, A. J., & Harper, R. H. R. (2003). *The Myth of the Paperless Office*. MIT Press, Cambridge, MA, USA.

Teevan (2004) J. Teevan, C. Alvarado, M.S. Ackerman, and D.R. Karger. The perfect search engine is not enough: A study of orienteering behavior in directed search. In CHI, pages 415--422, 2004.

Völkel, M., & Haller, H. (2006a). Conceptual Data Structures (CDS) - Towards an Ontology for Semi-Formal Articulation of Personal Knowledge. In *Proc. of the 14th International Conference on Conceptual Structure*. Aalborg University – Denmark.

Völkel, M., & Schaffert, S. (eds.) (2006b). *Proceedings of the First Workshop on Semantic Wikis - From Wiki to Semantics*, ESWC.

Völkel, M. (2007). From documents to knowledge models. In *Proc. of ProKW workshop, Konferenz Professionelles Wissensmanagement*, Potsdam, Germany.

Wagner, C., & Bolloju, N. (2005, April - June), "Knowledge Management with Conversational Technologies: Discussion Forums, Weblogs, and Wikis", *Journal of Database Management, Vol. 16, Issue 2*, i-viii.

Wenger, E., McDermott, R., & Snyder, W. M. (2002, March). *Cultivating Communities of Practice*. Harvard Business School Press.

Whittaker, S., Bellotti, V., & Gwizdka, J. (2006). Email in personal information management. *Commun. ACM, 49(1)* (pp. 68-73).

Yee, K.-P., Swearingen, K., Li, K., & Hearst, M. (2003). Faceted metadata for image search and browsing. In *CHI*.

Additional Reading

Bardini, T. (2000), *Bootstrapping: Douglas Engelbart, Coevolution, and the Origins of Personal Computing* (Writing Science), Stanford University Press.

Davenport, T. H. (2005), *Thinking for a Living: How to Get Better Performances And Results from Knowledge Workers*, Harvard Business School Press.

Engelbart, D. (1963), *A Conceptual Framework for the Augmentation of Mans Intellect*, Spartan Books: Washington, DC: London, pp. 1-29. Technical Report

Friedewald, M. (2000), *Der Computer als Werkzeug und Medium. Die geistigen und technischen Wurzeln des Personalcomputers* (Taschenbuch).

Goodman, D. (1988), *The Complete HyperCard Handbook*, Random House Information Group.

Lethbridge, T. C. (1994*), Practical Techniques for Organizing and Measuring Knowledge*, PhD thesis, University of Ottawa.

Rheingold, H. (2000), *Tools for Thought: The History and Future of Mind-Expanding Technology*, The MIT Press.

Sowa, J. F. (1999), *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Course Technology.

Acknowledgements

Parts of this work have been funded by the European Commission 6th Framework Programme in the context of the Knowledge Web Network of Excellence project, FP6-507482 and the EU IST NEPOMUK, the Social Semantic Desktop, FP6-027705. Parts of this material are based upon works supported by the Science Foundation Ireland under Grants No. SFI/02/CE1/I131 and SFI/04/BR/CS0694 and parts of this work have been funded by Salzburg Research, Austria.