Personal Knowledge Models for More Productive Knowledge Workers

Max Völkel

Forschungszentrum Informatik, Karlsruhe
voelkel@fzi.de

Abstract: In this paper we motivate how *personal knowledge models* can make knowledge workers more productive. The external representation can help him to tackle typical cognitive limits. They do so by allowing a knowledge worker to efficiently and effectively model her or his personal knowledge in a mix of informal and formal knowledge. As a result, more knowledge can be re-used cheaper and hopefully more complex problems can be tackled.

1 Motivation

Our world is constantly changing and the rate of change has constantly increased. Today, changes are in large part caused by the humans themselves, due to the growth of their global population and the ability to use technology to change matter, i. e. in agriculture, energy production, goods production, and transportation. Part of this ability to steer the forces of nature and humans in controlled ways can be contributed to the invention of management: "The most important contribution of management in the 20th century was to increase manual worker productivity fifty-fold (Dru99)." The fast rate of change in the environment and in human societies causes great problems. Humans must understand these problems and develop suitable solutions. Drucker goes on and foresees: "The most important contribution of management in the 21st century will be to increase knowledge worker productivity – hopefully by the same percentage. [. . .] The methods, however, are totally different from those that increased the productivity of manual workers." (Dru99) What could be methods to increase the productivity of knowledge workers?

Acknowledgements Part of this work has been funded by the European Commission in the context of the IST NEPOMUK IP - The Social Semantic Desktop, FP6-027705.

2 Approach

In this section we define our approach and motivate that *personal knowledge models* can amplify the human ability of understanding and solving problems.

Focus on Individual Knowledge Workers The field of knowledge management investigates since about 1995 (Sta05) how people and knowledge work together. North (Nor07) defines *knowledge work* as "work based on knowledge with an immaterial result; value creation is based on processing, generating and communicating knowledge." Knowledge management approaches have mostly focused on sharing of knowledge within organisations and teams. Polanyi (Pol66) makes a distinction between *explicit knowledge* encoded in artefacts such as speech, books or web pages, and *tacit knowledge* which resides in the individual. The SECI-model of Nonaka (Non94)[p. 57 ff] describes knowledge conversions between tacit and explicit knowledge.

Initial approaches tried to improve knowledge sharing by installing central repositories for codified knowledge. However, the high expectations towards them have often not been met (BM02). Later approaches concentrate on tools such as expert finders and corporate white pages, hence less on managing codified knowledge and more on direct communication among people.

Schütt (Sch03) defines a *knowledge worker* based on the works of Drucker and Taylor (Tay11): Simplified, workers (doing) are instructed by managers (thinking). These managers have to manage themselves. This self-managing is considered an important characteristic of a knowledge worker. Knowledge workers have to manage themselves, because their tasks are constantly changing. Increasing the knowledge worker productivity has to be a companies main goal, because as Davenport (Dav05) notes: "The knowledge-based organisation is no more effective than the sum of its knowledge workers". One should focus on the individual and give individual users incentive and benefit before focusing on the social network (Ore06) In this paper we focus on the *individual* knowledge worker and ways to make her or him more productive.

Expanding Cognitive Limits with External Artefacts Seminal articles by Bush (Bus45) and Engelbart (Eng63) describe *tools* that allow an individual to work more efficiently and more effective with *generic external representations* of knowledge. Both projects let a user create and connect knowledge cues in flexible and sophisticated ways. As a result, a user is no longer restricted to editing linear documents or two-dimensional drawings but can instead create a complex, abstract representation of a mental model.

In knowledge work, people are frequently confronted with two limitations of the human mind: long-term memory recall and short-term memory capacity. Limits of the long-term memory can be overcome partially with tools to help remembering or reconstructing knowledge. Human short-term memory can hold only around seven objects at a time (Mil56). For user interfaces, Shneiderman (Shn98) advises to "Do everything possible to free the user's memory burden". Interestingly, also the limitation of the short-term memory can be partly relieved by using external knowledge representations, e. g. by taking short notes, or drawing a diagram or mind-map that helps keep an overview over a somewhat larger set of items and quickly bring each single one into full conscience on demand. We conclude that both of these cognitive limits can be addressed by providing an adequate external knowledge representation tool.

The research field of *Personal Information Management* (PIM) has first been established by software tools like contact managers, calendars, to-do-lists and notes. The first publications explicitly mentioning the field come from 2001 (JBD01), since 2004 international workshops discuss the topic. The report of the first Personal Information Management (PIM) workshop (JB05) defines the *personal space of information* (PSI) as a set that "includes all the information items that are, at least nominally, under that person's control (but not necessarily exclusively so)" (JB05).

In this paper, we focus on the role of external artefacts to overcome cognitive limits of the knowledge worker. We further limit the research to knowledge artefacts that are created with the intention to be used by its creator, i. e. "notes to self".

Knowledge Cues Building on the distinction between explicit and tacit knowledge made by Polanyi (Pol66), later works (DC00, NK98) conclude that external and tacit knowledge are actually two extremes on a spectrum. Maurer (Mau99) states that knowledge resides in the heads of people and the computer can only store "computerized knowledge" which is to be understood as "shadow knowledge", a "weakish image" of the real knowledge.

In PKM, we often deal with knowledge that is somewhere in the middle of these extremes. Note-taking e. g. is a core activity of PKM: An individual creates an external representation for internal concepts. Later, the external representation is internalised again to re-activate the knowledge in the individuals mind. If somebody writes a short informal note to himself it is often completely meaningless to others. The knowledge is thus not fully externalised – Yet this note is an external reminder about some knowledge that the author would otherwise forget.

North (Nor02) defines the terms signal, data, information, knowledge, and wisdom in a layered fashion, one building upon another. By this definition, knowledge itself cannot be stored in information systems, only information can. Kidd (Kid94)[p190] concludes a tool should concentrate on capturing and reproducing the appearance of marks made by knowledge workers rather than interpreting them. However, only processing of knowledge that is sufficiently structured and formalised can automated.

Definition: In this paper we use the term *knowledge cue* to denote an external artefact created with the intent to be internalised later again by its author, then hopefully re-evoking the knowledge the author wanted to remember.

In the digital world, a *knowledge cue* can be represented in a text file, image, filename, or even in a part of a filename. It can also be the fact that a certain file has been placed into a certain folder, the assignment of a tag to a document or a link connecting two documents. In general, knowledge cues can be represented both *in content entities such as text or images and as connections among content entities*.

Storing text and hypertext in a computer allows retrieval by full-text search and browsing links. Structured databases and semi-structured document formats (e. g. XML) allow to answer queries with aggregated answers. E. g. an address data base can answer a query – if sufficiently formalised – like “number of people living in Germany”. By representing knowledge in a formal knowledge representation language, e. g. OWL (SD04) or RDFS (DB04), the computer can deduct new knowledge and answer queries about concepts. Obviously, the more structured knowledge is and the more explicit semantics of types and structures have been mode, the more powerful become ways to work with this knowledge in a computer. Oren (Ore06) requires to leave users their freedom and do not constrain them into rigid schemas . This paper concentrates on the representation, manipulation and use of *generic*, i. e. not domain specific, *knowledge cues* on different levels of granularity and formality.

User Interaction In 2008, there were 400 million personal computers (PCs) in use[1]. In this work we restrict ourselves to manipulation of digital knowledge artefacts by means of graphical user interfaces operated by keyboard and mouse. On a standard PC, text is much faster to create than diagrams. I. e. by means of information theory pressing a single key on a standard computer keyboard with 102 keys sends 6,6 bits of information, whereas moving the mouse to one of 102 defined locations on a screen and clicking there requires more time and haptical effort[2].

Visual representations play an important role in architecture, engineering, and research as “artefacts of knowing” (EW07). The commonest types of images are not purely pictures, notation or writing but are a *hybrid* between pictorial, geometric and scripted elements (Elk). As such they can been seen as semi-formal models of a domain. A study on 28 people in meetings using pen-computers (Kha94) found out that 50% copied diagrams from white-boards only, 17% used diagrams very occasionally, and 33% did not drew diagrams at all. In our model, we treat pictures and images drawn by the user as atomar symbols. We restrict our approach to authoring of text, structured text, and links and formal statements between textual items. We exclude free-form drawing for sake of feasibility. Note that this does not exclude interactive visualisations that show e. g. connections between items.

[1] According to the German magazine COMPUTERWOCHE
[2] Source: own calculations

Cost and Benefit The basic processes in PIM have been identified as (JB05): *Keeping*: input of information into a PSI; *Finding or re-finding*: output of information from a PSI, and *Meta-activities*: e. g. mapping between information and need, maintenance and organisation.

We can describe the *knowledge cue* life-cycle consisting of six phases: (1) *Creation* of knowledge cues, e. g. by authoring or by import from other sources. This maps to "keeping". The next steps are subsumed under meta-activities: (2) *Structuring* within the knowledge cues, e. g. by using text formatting and (3) *Organisation* among knowledge cues, e. g. by linking, tagging, and classifying. To obtain any benefit for the effort put into steps 1–3 on needs step (4) *Retrieval* of knowledge cues, e. g. by searching, browsing and following links. Knowledge workers often have to create information artefacts such as presentations, reports, speeches or books. Instead of creating them from scratch, we can sometimes re-use the existing content of and structures among knowledge cues. Hence we have the optional step (5) *Transformation* of knowledge cues to other formats or visualisations. Finally we have step (6) *Using* knowledge cues. This step happens outside of the system is just part of our model for economic reasons: This is the only step that gives value to the user; steps 1–5 are investments for future cost savings. Side remark: Some people do take notes solely because the act of writing helps them to remember the content better (Kid94).

If our task was to re-inform ourself about some previous knowledge, the value of the knowledge cue will be reflected as the value of decisions taken on the basis of the knowledge re-evoked by the knowledge cue. If our task was the creation of an information artefact, the price paid by others can be an indicator of its value.

Existing PIM tools either focus on specific structured data such as appointments, to-dos or contact data – or tackle only free-form note taking. Management of CD collections, cocktails recipes, text fragments, ideas, the personal social network, structured argumentation, bibliographic data, or web site logins requires domain-specific tools.

As explained in Sec. 2, more structured, more formalised data is easier to retrieve and transform, i. e. has lower costs. On the other hand, more structured data implies higher costs of creation, structuring and organisation. The optimal allocation of costs therefore depends on the number of times the externalised knowledge is re-used. A more detailed analysis of costs and benefit in Personal Knowledge Management (PKM) can be found in (VA08).

Personal Knowledge Models Computers allow us to partially automate operations on a knowledge model. Ontologies and formal reasoning can help to reduce retrieval costs when from a set of explicitly stated formal relations further formal statements can be inferred *automatically*. This deduced knowledge has not to be constructed by hand but is already available for browsing and queries.

Additionally, ontologies are a technology for data integration between heterogenous data stores using different schemas. The same situation arises if e. g. a number of notes have been taken about two topics, and later the user decides that one of the topics can be considered a sub-topic of the other one. After adding a formal super-topic statement between these two topics, a reasoning engine can find all relevant notes when querying for the super-topic. However, the more formal the statements are, the more thinking and modelling effort has to go into their creation.

The user must have the freedom to decide how much effort to use for knowledge modelling. Therefore, we let her or him use plain text, structured text, and formal statements together.

Definition: A *Personal Knowledge Models* is a digital artefact with the purpose to represent a set of knowledge cues in a unified way. The knowledge cues can vary in size, structured-ness and degree of formality.

3 Requirements for Personal Knowledge Models

In this section we analyze requirements for personal knowledge models from literature.

One obvious way to make knowledge workers more productive is by *re-using as much knowledge they have created as possible*. Knowledge workers not only create and transform knowledge, they also have to communicate knowledge to other parties. For this, they have e. g. to write documents and reports (Roc02). The cost of transforming explicit knowledge encoded in one kind of formalism into another one, e. g. from a spreadsheet to a presentation, should be taken into account. The overall idea of this section is therefore to make re-use of knowledge and content as cheap as possible.

To re-use content residing in one kind of representation in another tool, it needs to be transformed. Transformations between datamodels come not for free. A naive approach to convert between n formalisms would require writing n^2 transformations. However, if a common intermediate formalism can be used, the costs come down to $2n$. To save costs in content transformation, the conceptual knowledge model should therefore be a *super-set of the conceptual models of all other relevant PKM tools*. Due to lack of space we list only those requirements that are most novel.

Step-wise Formalisation Users needs a simple way to express content in an informal way, e. g. as plain text, formatted text or box-and-arrow diagrams (Ore06, AvE04). People need to be able to work at any level of formality (or informality), and to freely mix such levels (Let91). Such a mix of data could be called "semi-semantic", analogous to "semi-structured" data. Then the user should be able to migrate the knowledge into more formal structures, if desired (c. f. (VH06)). The need of incremental formalisation has already been recognized and described in (FMSM99): Incremental formalization requires a system architecture that (1) integrates formal and informal representations and (2) supports progressive formalization of information.

Knowledge Model Refactoring In a similar way as stepwise formalisation, knowledge cues in general have also to be changed. Borrowing a term from software engineering this process can be called *knowledge refactoring*. Efficient knowledge refactoring should let a user perform filing, categorisation and annotation operations on single or multiple knowledge cues efficiently. All kinds of structures and formal statements within and among knowledge cues should be easy to change. Schreiber (SH04) emphasises flexibility of knowledge models and the need for restructuring:

> A central question is how to understand a structure of the collected information. [...] But, as known, the problem is that the process of learning will continue and all the time the individual will be inspired to look after new subjects or new elements of the subject. After a while, the individual will experience that the structure of the information is not the right one anymore. It is necessary to choose a new way to structure the information. [...] Thus, a flexibility concerning how to structure the information is necessary.

Thus flexible ways to *restructure* a knowledge model are needed.

Knowledge Model Maintenance Externalised personal knowledge artefacts are usually organised in a systematic manner, e. g. files are sorted in folders and subfolders. Unfortunately, a good structure today is not a good structure tomorrow, therefore personal organisation schemes change. Weeks, months or years later, an existing folder structure sometimes "does not make sense" and cannot help well in locating files. Even if good refactoring support is available, it would be too costly to re-organise all personal artefacts frequently. Additional metadata about the usage of the knowledge cues by the system is required: How often did the knowledge cue appear in search results? How often has it been changed? When has the most recent statement been made about this knowledge cue? When was the last time this knowledge cue was used for inferencing? Such metadata can be used by the system to ask a user specifically and actively about the status of certain knowledge cues.

4 Conclusions

In this paper we have motivated the need for personal knowledge models and described some requirements for them. Knowledge models need to be able to express knowledge cues as atomar units of content – represented as text or images – as well as connections among them. In order to partially automate processes on the knowledge models – such as deriving new connections or aggregate views – the user must be able to explicate the semantics of connections. As formalisation is a hard task, the user must be able to perform it in a step-by-step fashion. It must be easy for the user to restructure his knowledge model with low effort. Otherwise an existing knowledge model becomes more an obstacle than a help in modelling a mental model. Only if a users mental model can be modelled precisely enough, a user can perform reasoning-like tasks on the symbolic representation. Active support for maintenance of knowledge models is required, to keep a knowledge model up-to date with changing mental models.

We believe that PKM and knowledge work in general will move away from the tight coupling of "one formalism to one tool" and instead move to generic formalisms that can be edited in a number of tools. Such generic formalisms will allow new scientific or personal insights as well as provide cheaper ways for storing, retrieving and transforming knowledge. These requirements should be taken into account when designing personal knowledge management tools.

References

[AvE04] Andreas Abecker and Ludger van Elst. Ontologies for Knowledge Management. In *Handbook on Ontologies*, International Handbooks on Information Systems, pages 435–454. Springer, January 2004.

[BM02] A. Braganza and G. J. Mollenkramer. Anatomy of a failed knowledge management initiative: Lessons from pharmacorp's experiences. *Knowledge and Process Management*, 9(1):23–33, 2002.

[Bus45] Vannevar Bush. As we may think. *Atlantic Monthly*, 176:101–108, 1945.

[Dav05] Thomas H. Davenport. *Thinking for a Living: How to Get Better Performances And Results from Knowledge Workers.* Harvard Business School Press, Boston, MA, USA, September 2005.

[DB04] R.V. Guha Dan Brickley. RDF Vocabulary Description Language 1.0: RDF Schema. `http://www.w3.org/TR/rdf-schema/`, February 2004.

[DC00] Charles Despres and Daniele Chauvel. *Knowledge Horizons: the present and promise of Knowledge Management.* Butterworth-Heinemann, October 2000.

[Dru99] Peter F. Drucker. *Management Challenges for the 21st Century.* HarperBusiness, 1999.

[Elk] James Elkins. *The Domain of Images.* Cornell University Press.

[Eng63] D.C. Engelbart. *A Conceptual Framework for the Augmentation of Man's Intellect*, pages 1–29. Spartan Books:

Washington, DC: London, 1963. Technical Report version: http://sloan.stanford.edu/mousesite/EngelbartPapers/B5_F18_ConceptFrameworkInd.html.

[EW07] B. Ewenstein and J. K. Whyte. Picture this: visual representations as 'artifacts of knowing'. *Building Research and Information*, 35(1):81–89, 2007.

[FMSM99] III Frank M. Shipman and Raymond J. McCall. Incremental formalization with the hyper-object substrate. *ACM Trans. Inf. Syst.*, 17(2):199–227, 1999.

[JB05] William Jones and Harry Bruce. A Report on the NSF-Sponsored Workshop on Personal Information Management. report, January 2005.

[JBD01] William Jones, Harry Bruce, and Susan Dumais. Keeping found things found on the web. In *CIKM '01: Proceedings of the tenth international conference on Information and knowledge management*, pages 119–126, New York, NY, USA, 2001. ACM Press.

[Kha94] Fawzia Khan. A Survey of Note-Taking Practices. Technical Report HPL-93-107, Hewlett-Packard, Feb 1994.

[Kid94] Alison Kidd. The Marks are on the Knowledge Worker. In *CHI '94: Proc. of the SIGCHI conf. on Human factors in computing systems*, pages 186–191. ACM Press, 1994.

[Let91] Timothy C. Lethbridge. A Model for Informality in Knowledge Representation and Acquisition (an extended abstract). presented at Workshop on Informal Computing, May 29-31 1991, Santa Cruz CA, 1991.

[Mau99] H Maurer. The Heart of the Problem: Knowledge Management and Knowledge Transfer. In *Proc. ENABLE'99*, pages 8–17. Espoo-Vantaa Institute of Technology, 1999.

[Mil56] George Miller. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *PSYCHOLOGICAL REVIEW*, 63:81–97, 1956.

[NK98] I. Nonaka and N. Konno. The concept of "ba": Building a foundation for knowledge creation. *California Management Review*, 40(3):40–54, 1998.

[Non94] Ikujiro Nonaka. A Dynamic Theory of Organizational Knowledge Creation. *Organization Science*, 5(1):14–37, Februar 1994.

[Nor02] Klaus North. *Wissensorientierte Unternehmensführung*. Dr. Th. Gabler Verlag, 3rd edition, 2002.

[Nor07] Klaus North. Produktive Wissensarbeit. In *5. Karlsruher Symposium für Wissensmanagement in Theorie und Praxis*, October 2007. CD-ROM.

[Ore06] Eyal Oren. An Overview of Information Management and Knowledge Work Studies: Lessons for the Semantic Desktop. In Stefan Decker, Jack Park, Leo Sauermann, Söen Auer, and Siegfried Handschuh, editors, *Proc. of the Semantic Desktop and Social Semantic Collaboration Workshop*, volume 202 of *CEUR.ws*, 2006.

[Pol66] Michael Polanyi. *Tacit Dimension*. Routledge & Kegan Paul Ltd, London, 1966.

[Roc02] Ann Rockley. *Managing Enterprise Content: A Unified Content Strategy*. New Riders Press, October 2002.

[Sch03] P. Schütt. The post-Nonaka Knowledge Management. *Journal of Universal Computer Science*, 9(6):451–462, 2003. http://www.jucs.org/jucs_9_6/the_post_nonaka_knowledge.

[SD04] G. Schreiber and M. Dean. OWL Web Ontology Language Reference. http://www.w3.org/TR/2004/REC-owl-ref-20040210/, February 2004.

[SH04] Trine Schreiber and Karen Harbo. Information literacy and personal knowledge management. In *12th Nordic Conference on Information and Documentation*, pages 106–114, 2004.

[Shn98] Ben Shneiderman. *Designing the User Interface*. Addison Wesley, May 1998.

[Sta05] Michael Stankosky. *Creating the Discipline of Knowledge Management: The Latest in University Research*. Butterworth-Heinemann, February 2005.

[Tay11] Frederick Winslow Taylor. *The Principles of Scientific Management*. Harper & Brothers, New York and London, 1911.

[VA08] Max Völkel and Andreas Abecker. Cost-Benefit Analysis for the Design of Personal Knowledge Management Systems. In José Cordeiro and Joaquim Filipe, editors, *ICEIS 2008 - Proceedings of the Tenth International Conference on Enterprise Information Systems, Volume AIDSS, Barcelona, Spain, June 12-16, 2008*, pages 95–105, JUN 2008.

[VH06] Max Völkel and Heiko Haller. Conceptual Data Structures (CDS) – Towards an Ontology for Semi-Formal Articulation of Personal Knowledge. In *Proc. of the 14th International Conference on Conceptual Structures 2006*, Aalborg University - Denmark, JUL 2006.