A Lightweight Approach for Proactive, Task-Specific Information Delivery

Harald Holz, Heiko Maus, Ansgar Bernardi, Oleg Rostanin
(DFKI GmbH – German Research Center for AI, Germany
<firstname>.<lastname>@dfki.de)

Abstract: Knowledge management approaches for weakly-structured, ad-hoc knowledge work processes need to be lightweight, i.e., they cannot rely on high upfront modeling effort. This paper presents a novel prototype for supporting weakly-structured processes by integrating a standard to-do list application with a state-of-the-art document classification system. The resulting system allows for a task-oriented view on an office worker's personal knowledge space in order to realize a proactive and context-sensitive information support during her daily, knowledge-intensive tasks.

Key Words: Weakly-structured workflows, agile workflows, proactive information delivery
Category: H.3.3, I.2.0

1 Motivation

The recent emergence and popularity of several new desktop search engines such as Google Desktop Search[1], x-friend[2], MSN Desktop Search[3], etc. has clearly shown the need for tools that help users in managing their personal knowledge space (PKS). Typically, the documents needed by a knowledge worker for the task at hand are spread over various places such as e-mail folders, file system folders, or paper stacks on the desk. While the concept of a desktop-wide search certainly relieves the user from the burden of querying several different information sources (e-mail, local and network drives, etc.), current desktop search engines still follow the standard, passive query/retrieve model: the user has to explicitly 'pull' for information that might be relevant for a task he is currently trying to accomplish. Besides being inefficient, empirical studies have shown that such pull approaches typically lead to suboptimal reuse rates of available documents [Mahe and Rieu, 1997].

In order to address this issue, several business process-oriented knowledge management approaches have been developed for proactively providing process participants with information that is relevant with regard to their current tasks [Abecker et al., 2002]. However, as most of these approaches rely on

[1] http://desktop.google.com/
[2] http://www.x-friend.de/
[3] http://toolbar.msn.com/

static workflow/process specifications, they typically are inadequate for weakly-structured workflows such as knowledge-intensive office work processes. Currently, state-of-the-art workflow and document management systems offer valuable support only for routine activities in office work. In spite of such support, it has been claimed that knowledge-intensive office work has not reached satisfying increases in productivity in recent years (cf. [Schütt, 2003]). The reason for this perceived lack of productivity increase in such office work is seen in the insufficient understanding of the nature of knowledge-intensive work and the lack of adequate integration of information support and work activities.

In the following, we present a novel prototype for lightweight information support within knowledge-intensive processes and work environments by realizing proactive knowledge delivery in agile knowledge workflows (cf. [Elst et al., 2003]). The prototype results from an integration of the flexible FRODO TaskMan[4] workflow system and BrainFiler, a state-of-the-art document classification system. [Section 2] describes how BrainFiler is used to construct task-oriented, personal information structures. In [Section 3], we outline how the integration with TaskMan facilitates the proactive recommendation of documents in the context of a given task based on BrainFiler's classification functionality. Related work is reviewed in [Section 4], followed by a conclusion in [Section 5].

2 Task-Oriented, Personal Information Structures

The importance of integrating knowledge management activities into business process modeling and enactment is being increasingly accepted, and several different approaches have already been proposed and successfully realized [Abecker et al., 2002]. One of the primary goals of these business process-oriented knowledge management initiatives is to establish, run and maintain an organizational environment that provides process participants with the information needed to successfully perform their tasks/activities as defined in process models. Consequently, most of the approaches rely on the existence of generic process models or workflow specifications, around which the knowledge capturing and provision strategies are organized. However, a considerable amount of knowledge work processes that occur daily in the context of office work are highly dynamic, ad-hoc, and weakly-structured by their nature, and cannot be modeled in advance at a sufficient level of detail.

In order to address this issue, we aim at a lightweight, bottom-up approach to proactively provide knowledge workers with information items that are relevant in the context of their current tasks. Our approach starts from two assumptions:

– Information that is relevant to knowledge workers during their daily tasks is available from their desktops in the form of electronic documents.

[4] `http://www.dfki.de/frodo/taskman`

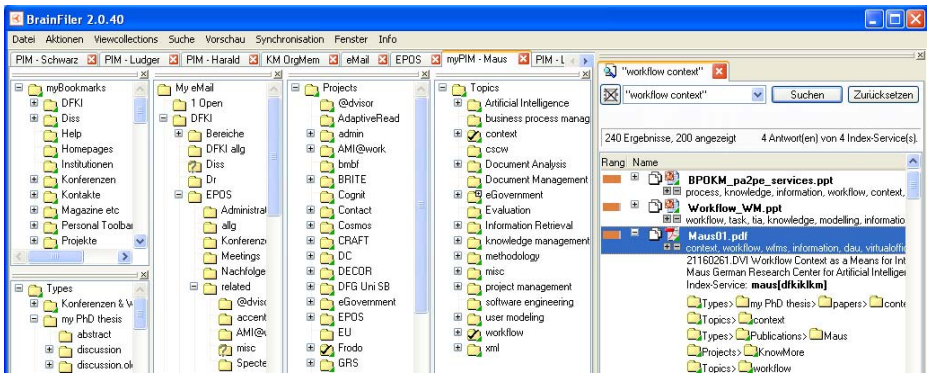

Figure 1: Multi-criterial indexing of documents with BrainFiler

– Knowledge workers are willing to maintain lists of their current tasks using a to-do list (or: task list) application.

While most of the information that is relevant to the knowledge workers during their daily tasks is available from their desktop, the current popularity of desktop search engines indicates that a considerable amount of time (and hence: money) is spent searching for that information [Delphi Group, 2002]. Two main reasons seem to be responsible for this:

1. Documents are stored in several different systems (e.g., e-mail, various local and network drives, etc.)

2. Most folder hierarchies only allow to place a document in at most one folder, i.e., a document cannot be indexed under more than one topic/concept (without creating redundant copies).

While current desktop search engines address the first issue, they do not provide a solution to the second issue. In the ideal case, what would be required in order to reduce the necessity of search is a *task-specific organization* of documents, i.e., a (logical) folder or view on the document space containing all available documents relevant to a current task. As a first step in this direction, our approach makes use of the tool BrainFiler™, a system being developed together with brainbot technologies[5] as part of the EPOS project[6].

BrainFiler realizes a personalized document management environment allowing multi-criterial classification of documents, search functionality such as

[5] `http://www.brainbot.com`
[6] `http://www.dfki.de/epos`

boolean search and document similarity evaluation, as well as incorporation of remote (peer-to-peer) BrainFiler instances. BrainFiler enables a user to build a PKS by allowing to import native structures such as e-mail folders, bookmarks, and file directories together with contained documents [see Fig. 1]. The imported structures are shown as trees (usually interpreted as *is-a* hierarchies). The nodes (interpreted as concepts) get their meaning by a document term-similarity vector determined statistically from the assigned documents.

A user is now able to elaborate the PKS by creating new or rearranging existing structures, making relations between concepts (a concept can have multiple parents), and assigning documents to several concepts. Furthermore, a user is able incorporate published structures from peers. These structures then can be used for a conceptual search (all documents having the concepts X and Y) as well as a combination with the keyword-based search.

While BrainFiler thus allows to group a document under all thematically relevant topics/concepts, a task-oriented view on the PKS is still lacking. For this, we make use of the knowledge worker's tasks maintained using a to-do list application. In general, the to-do list application allows users to manage their current tasks, e.g., such as in MS Outlook, Mozilla Calendar, or standard workflow systems. Typically, the representation of a task covers a short task name and a due date, together with an (optional) longer task description that describes the task's goal and objective in more detail, or – depending on the application – is used as scratchpad to jot down things to remember with regard to the task. For our prototype implementation, we made use of the to-do list application provided by the flexible FRODO TaskMan[7] workflow system ([see Fig. 2], left-hand pane), and coupled it with BrainFiler: for every task added to the to-do list, a corresponding folder node is automatically created within the PKS.

While a knowledge worker is working on one of his tasks, usually he needs access to certain documents (e-mails, PDF documents, etc.) in order to successfully perform the task. Typically, these documents are distributed over several different e-mail or file folders, depending on individual preferences with regard to file organization. For knowledge workers who experience frequent task context switches during their work, or for tasks that take longer than one day, this means that the knowledge worker has to repeatedly either browse manually through his file structures, or repeatedly perform a desktop search in order to find the required documents/folders.

Therefore, we extended the FRODO TaskMan to-do list application by allowing knowledge workers to associate bookmarks (i.e.: links) to relevant file folders and documents with their tasks. Technically, for every task-specific bookmark, a corresponding subnode is automatically created under the task node within

[7] `http://www.dfki.de/frodo/taskman`

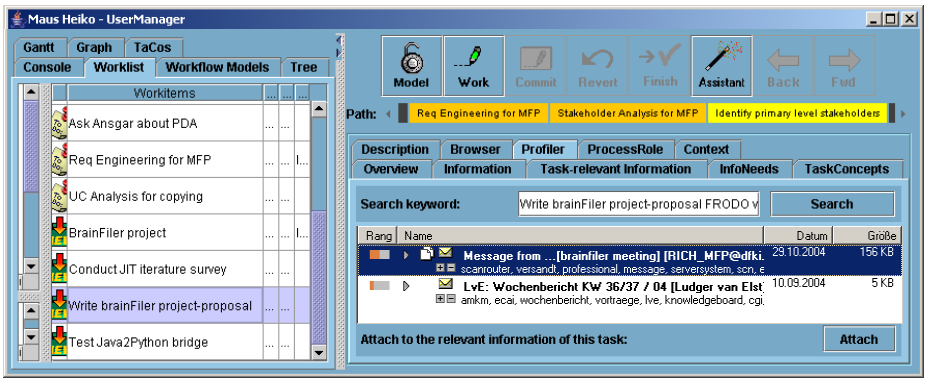

Figure 2: Task-specific, proactive document provision from the personal knowledge space within FRODO TaskMan

BrainFiler. This extends the knowledge worker's PKS with a task-oriented view and yields the benefit of providing him with immediate access to the heterogeneous set of relevant documents in the context of a given task. The resulting task-specific organization of documents also provides the basis for a proactive delivery of other documents (cf. below).

3 Task-Specific Document Delivery

So far, we have assumed that knowledge workers manually associate relevant folders and documents from their PKS with their tasks. Desirable here would be the concept of an automated assistant that "looks over the task enactor's shoulder" and (pro)actively provides him with available documents that are relevant for the task currently being enacted. In order to achieve this, we make use of BrainFiler's document classification functionality: for a given document, BrainFiler can suggest those of the user's concepts which fit the best, based on the statistically induced term relevance information.

In order to proactively provide knowledge workers with access to documents stored within their PKS that might be relevant in the context of their current task, we extended the TaskMan to-do list application by a component that displays the results of BrainFiler's classifications of documents with regard to the current task. [Fig. 2] shows a screenshot from the FRODO TaskMan to-do list application: the left-hand pane shows the user's to-do list, with the task "Write brainFiler project-proposal" currently being selected. In the right-hand pane, two emails being provided to the user by the component in the context of the currently selected task; a double-click on one of these emails will open the email

with the user's default email application. The two emails have been automatically retrieved by using the relevant terms displayed in the text field labeled "Search keywords", that have been extracted from the task name and already associated documents. That way, relevant e-mails are no longer easily overlooked, e.g., because important e-mails with regard to a given task can now be automatically identified among the unorganized flood of continuously incoming e-mails, and displayed to the knowledge worker in their proper workflow resp. task context.

Technically, this functionality has been realized by automatically creating a file with the task name and description, that is being placed within the task's BrainFiler node folder, in addition to the folders and documents (including e-mails) that the knowledge worker manually associated with the task.

All other documents within the user's personal document space, as well as any newly "incoming" documents, are automatically analyzed by the component and tentatively associated with those of the worker's current tasks that the documents seems to be related to, by making use of the BrainFiler's classification suggestions with regard to a task's folder node.

Currently, our prototype can cope with three different ways in which a document can be "incoming": the document can be sent by e-mail, scanned and delivered via a multi-functional product (MFP; creating an intelligent office appliance, see [Maus et al., 2005]), or saved into a file directory that is being synchronized with BrainFiler's concept hierarchies.

4 Related Work

The issues addressed by the approach presented here stem mainly from the areas of process-oriented knowledge management and desktop search engines. In the following, we briefly compare existing work with the approach described in this paper. Most work on integrating knowledge management and process support has been done in the field of business processes (see [Abecker et al., 2002] for a recent overview of Business Process-Oriented Knowledge Management). Prominent approaches such as EULE [Reimer et al., 2000], OntoBroker [Schnurr et al., 1999], WorkBrain [Wargitsch et al., 1998], PreBIS [Delp et al., 2004], or DECOR [Abecker et al., 2001] focus mainly on fairly static (in contrast to weakly-structured) processes with regard to proactive information delivery; hence, they rely on structured task representation and ontologies. Caramba [Dustdar, 2004] realizes an activity-based knowledge management approach for ad-hoc processes by enabling knowledge workers to link knowledge artifacts to tasks. However, only artifacts that have already been linked to a task are made accessible for the task's enactors; a proactive distribution of potentially relevant artifacts based on the content of artifacts already linked to the task is not provided.

The CALVIN project [Leake et al., 2000] investigates lessons learned systems supporting the process of finding information relevant to a particular research task. CALVIN learns about information sources by automatically recording cases that represent the consulted information sources. As the user browses for information, the system maintains the user's current research context (e.g., a set of keywords describing the main topics) and compares it with former contexts. If the similarity between the current and a former context exceeds a certain threshold, the resources associated with the former context are presented to the user as relevant in his current context. Other approaches to provide light-weight, proactive information delivery are based on collaborative filtering (CF) technology, e.g., GroupLens [Resnick et al., 1994] or Entree [Burke, 1999]. In [Holz, 2003a], we described a system that makes use of both a heuristic approach based on CF/CBR functionality and formal specifications of when to recommend certain information resources. Current desktop search engines (e.g., Google Desktop Search, x-friend, MSN Desktop Search) do not yet have a notion of a user's task or some other retrieval context. An exception is blinkx[8], that provides on-the-fly recommendation links to available documents that are relevant to the user's active window (e.g., an open document or e-mail editor).

5 Conclusion

We presented a prototype that realizes a light-weight approach to task-specific, proactive document delivery. The term vector similarity-based approach used here is intended to complement our earlier work on more heavy-weight approaches based on process models and ontologies [Elst et al., 2003, Holz, 2003b], which require considerably more modeling effort on behalf of the user. Although we used a workflow system as a basis for the prototype, the presented approach is also applicable to to-do list applications as found in the personal information management tools (e.g., PDAs) of today's office workers.

The prototype presented here is currently under development and will be evaluated as part of a distributed software development case study that is scheduled for this year. Based on the positive evaluation results for our process-embedded information support [Elst et al., 2003], we believe that an efficiency gain can also be achieved in an everyday office setting with the approach presented here, by making documents more easily available during the office worker's tasks, and helping to prevent that relevant documents might be overlooked.

References

[Abecker et al., 2001] Abecker, A., Bernardi, A., Ntioudis, S., Mentzas, G., et al. (2001). The DECOR Toolbox for Workflow-Embedded Organizational Memory Ac-

[8] http://www.blinkx.com

cess. In *Proc. 3rd Int. Conf. on Enterprise Information Systems (ICEIS 01).*
[Abecker et al., 2002] Abecker, A., Hinkelmann, K., Maus, H., and Müller, H.-J., editors (2002). *Geschäftsprozessorientiertes Wissensmanagement.* xpert.press Springer.
[Burke, 1999] Burke, R. (1999). Integrating knowledge-based and collaborative-filtering recommender systems. In *Proc. Workshop on AI and Electronic Commerce.*
[Delp et al., 2004] Delp, M., Böhm, K., and Engelbach, W. (2004). Pre-built information space: Some observations on the challenges of process-oriented knowledge management. In *Proc. I-KNOW '04.*
[Delphi Group, 2002] Delphi Group (2002). Taxonomy & content classification. Market milestone report.
[Dustdar, 2004] Dustdar, S. (2004). Reconciling knowledge management and workflow management: The activity-based knowledge management approach. In *Proc. I-KNOW '04.*
[Elst et al., 2003] Elst, L. v., Aschoff, F.-R., Bernardi, A., Maus, H., and Schwarz, S. (2003). Weakly-structured workflows for knowledge-intensive tasks: An experimental evaluation. In *IEEE WETICE Workshop on Knowledge Management for Distributed Agile Processes (KMDAP03).* IEEE Computer Press.
[Holz, 2003a] Holz, H. (2003a). An incremental approach to task-specific information delivery in SE processes. In *Proc. 18th IEEE International Conference on Automated Software Engineering (ASE '03).*
[Holz, 2003b] Holz, H. (2003b). *Process-Based Knowledge Management Support for Software Engineering.* dissertation.de Verlag.
[Leake et al., 2000] Leake, D., Bauer, T., Maguitman, A., and Wilson, D. (2000). Capture, storage and reuse of lessons about information resources: Supporting task-based information search. In *AAAI-2000 Workshop on Intelligent Lessons Learned Systems.*
[Mahe and Rieu, 1997] Mahe, S. and Rieu, C. (1997). Towards a pull-approach of km for improving enterprise flexibility responsiveness: A necessary first step for introducing knowledge management in small and medium enterprises. In *Proc. Int. Symposium on Management of Industrial and Corporate Knowledge (ISMICK 97).*
[Maus et al., 2005] Maus, H., Holz, H., Bernardi, A., and Rostanin, O. (2005). Leveraging passive paper piles to active objects in personal knowledge spaces. In *1st Workshop on Intelligent Office Appliances (IOA 05) at the WM 2005, Kaiserslautern. Germany.*
[Reimer et al., 2000] Reimer, U., Margelisch, A., and Staudt, M. (2000). A knowledge-based approach to support business processes. In *Proc. AAAI Workshop on Bringing Knowledge to Business Processes.*
[Resnick et al., 1994] Resnick, P., Iacovou, N., Suchak, M., Bergstorm, P., and Riedl, J. (1994). Grouplens: An open architecture for collaborative filtering of netnews. In *Proc. ACM 1994 Conference on Computer Supported Cooperative Work.*
[Schnurr et al., 1999] Schnurr, H.-P., Staab, S., and Studer, R. (1999). Ontology-based process support. In *Workshop on Exploring Synergies of Knowledge Management and Case-Based Reasoning (AAAI-99).* AAAI.
[Schütt, 2003] Schütt, P. (2003). The post-Nonaka Knowledge Management. *Journal of Universal Computer Science*, 9(6):451–462.
[Wargitsch et al., 1998] Wargitsch, C., Wewers, T., and Theisinger, F. (1998). An organizational memory-based approach for an evolutionary workflow management system - concepts and implementation. In *31st Hawaiian Int. Conf. on System Sciences, HICSS.*

Acknowledgements

Work funded in part by "Stiftung Rheinland-Pfalz für Innovation" (InnoWiss) and BMBF (EPOS, contract number ITW 01 IWC 01).