

Leveraging Passive Paper Piles to Active Objects in Personal Knowledge Spaces^{*}

Heiko Maus, Harald Holz, Ansgar Bernardi, Oleg Rostanin

DFKI GmbH – German Research Center for AI, Kaiserslautern, Germany
<firstname>.<lastname>@dfki.de

Abstract. This paper presents a novel prototype for an intelligent office appliance which results from an integration of three state-of-the-art office applications/appliances: a workflow system, a document classification system, and a multi-functional peripheral. The resulting system allows for leveraging an office worker’s papers to her personal knowledge space in order to realize an pro-active and context-sensitive information support within knowledge-intensive tasks and processes.

1 Motivation

Office work, from an abstract point of view, comprises a wide variety of activities which are concerned with paper handling, information processing, thinking, and decision making. Although sometimes despised as bureaucratic overhead, office work is nevertheless at the heart of knowledge and value creation and the indispensable basis for innovation and progress. The ever-increasing percentage of office work in today’s economic activities, ranging from standard administrative processes to complex and unique decisions, clearly illustrates this significance.

Modern office appliances offer valuable support for routine activities in office work. In spite of such support, it is claimed that knowledge-intensive office work has not reached satisfying increases in productivity in recent years (cf. [4]). 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 effective, light-weight information support within knowledge-intensive processes and work environments by realizing just-in-time knowledge delivery in agile knowledge workflows (cf. [2]). The prototype results from an integration of three state-of-the-art office applications/appliances: (i) a workflow system, (ii) a document classification system, and (iii) a multi-functional product.

2 Leveraging Paper into Personal Knowledge Spaces

Considering the development of office appliances, a clear trend towards integration can be observed. This is exemplified in the success of multi-functional products (MFP, e.g. as developed by Ricoh) that combine a copier, (colour) printer,

^{*} Work funded in part by “Stiftung Rheinland-Pfalz für Innovation” and BMBF.

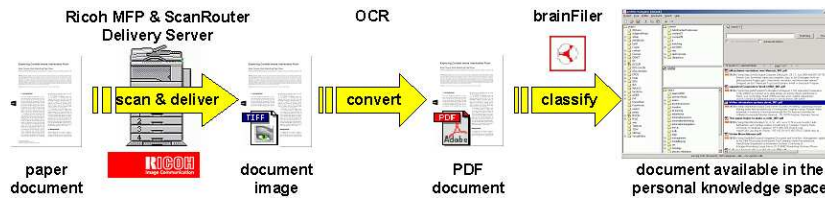


Fig. 1. From passive paper piles to active documents in the personal knowledge space

scanner, and fax machine, featuring a built-in hard drive, network access, and a minimal web server for administrative tasks and document management. Moreover, Ricoh's ScanRouter Delivery ServerTM allows for delivering and storing a scanned document in various formats (TIFF, JPG, etc.) in the user's inbox at the delivery server, Ricoh's own document management system, any network drive, or sending it as an attachment to a user-defined e-mail-account.

Such an MFP is the starting point to transfer piles of paper into an electronic format in which it can be handed over to a document classification system, as depicted in Fig. 1. The introduction into the knowledge worker's personal knowledge space is realized with brainFilerTM, a system that we develop together with brainbot technologies as part of the EPOS-project (<http://www.dfki.de/epos>).

BrainFiler realizes a personal document management environment allowing multicriterial classification of documents, search functionality such as boolean search and document similarity evaluation, and incorporation of remote brainFiler instances. BrainFiler enables a user to build a personal knowledge space by allowing to import native structures such as e-mail folders, bookmarks, and file directories together with contained documents. The imported structures are shown as trees (usually interpreted as an *is-a* hierarchy). The nodes (interpreted as concepts) get their meaning by a document term-similarity vector generated by assigned documents. A user is now able to construct a personal information space by creating new structures, making relations between concepts (a concept can have multiple parents), and assigning documents to several concepts. 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.

Technically, the introduction of paper documents in the personal knowledge space is done by handing over the scanned, transformed (as PDF), and delivered document to brainFiler. BrainFiler in turn analyzes the document and suggests those of the user's concepts which fit the best, based on the learned information structures of the user. The user can accept one or more proposed concepts as classification for the document, freely choose others, or simply ignore the assignment task.

3 Task-Specific Document Delivery

In order to be able to proactively provide users with access to relevant documents stored within their personal knowledge space, we make use of an explicit repre-

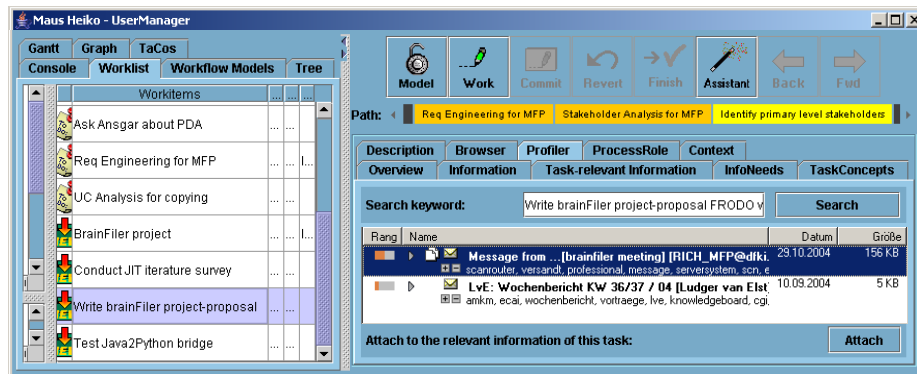


Fig. 2. Task-specific, proactive document provision from the personal knowledge space within FRODO TaskMan

sentation of a knowledge worker’s current tasks. The basis of this approach is the use of an application which provides a to-do list (or: task list) for knowledge workers that allows to manage their current tasks, e.g. such as in MS Outlook or 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 system – is used as scratchpad to jot down things to remember with regard to the task. Fig. 2 shows a screenshot from the FRODO TaskMan¹ workflow system that we used to realize the prototype: the left-hand pane shows the user’s to-do list, with the task “Write brainFiler project-proposal” currently being selected.

As the list of a knowledge worker’s current tasks reflects the work contexts he is currently involved in, it can be assumed that the majority of business e-mails received and sent, as well as all other documents handled by the worker are related to one or more of the tasks on his to-do list. However, current technology does not adequately support the organizing and filing of documents around tasks. Instead, knowledge workers are forced to manually browse to (or search for) the documents relevant to the task at hand, resulting in a considerable waste of time (and hence: money) [1].

In order to address this problem, we extended this to-do list by a component that displays the results of brainFiler’s classifications of documents with regard to the current task. Technically, a corresponding concept is automatically created for each task, containing a file with the task name and description, as well as documents (including e-mails) that the knowledge worker manually associated with the task (e.g. by simple drag-and-drop). This extends the knowledge worker’s personal knowledge space with a process-oriented view and yields the benefit of providing the knowledge worker with immediate access to the heterogeneous set of all documents related to a given task. However, what is most important, any newly “incoming” documents are automatically analyzed by the

¹ <http://www.dfki.de/frodo/taskman>

component and tentatively associated with all of the worker's current tasks that the new document seems to be related to, by making use of the brainFiler's classification suggestions. For example, Fig. 2 shows 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.

Currently, our prototype can cope with two different ways on which a document can be "incoming": the document can be sent by e-mail, or a document scanned and delivered via the MFP as explained in the previous section.

4 Conclusion

We presented a prototype of an intelligent office appliance which brings together research and state-of-the-art office tools and devices in order to realize 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 [2][3], which require considerably more modeling effort on behalf of the user.

The prototype presented here is currently being development within the competence center "Virtual Office of the Future"² established at DFKI and will be evaluated as part of a distributed software development case study in the next year. Based on the positive evaluation results for our process-embedded information support [2], 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

1. Taxonomy & content classification. Market milestone report, Delphi Group, 2002.
2. Ludger van Elst, Felix-Robinson Aschoff, Ansgar Bernardi, Heiko Maus, and Sven Schwarz. 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, 2003.
3. Harald Holz. *Process-Based Knowledge Management Support for Software Engineering*. dissertation.de Verlag, 2003.
4. Peter Schütt. The post-Nonaka Knowledge Management. *Journal of Universal Computer Science*, 9(6):451–462, 2003.

² <http://www.ricoh.rlp-labs.de/>