Im Wissensnetz: Linked Information Processes in Research Networks

S. Braun[1], M. Hefke[1], A. Schmidt[1], N. Sevilmis[2]

[1]FZI Research Center for Information Technologies
Haid-und-Neu-Str. 10-14
76131, Karlsruhe, Germany
Email: {aschmidt, hefke, braun}@fzi.de

[2]Fraunhofer Institut Graphische Datenverarbeitung
Fraunhoferstr. 5
64283, Darmstadt, Germany
Email: neyir.sevilmis@igd.fhg.de

Abstract

At first glance, it seems to be quite surprising that so far traditional business process-oriented knowledge management techniques have not been transferred to research in order to improve the efficiency of scientific work on a larger scale. But due to the high variability and unpredictability of scientific work processes, these techniques are not applicable. In fact scientific work processes have to be understood as a network of informal learning processes with a high level of social interaction. For this purpose, we have elaborated the model of a "Knowledge-Added Process" as a new paradigm of process-oriented support. We are developing various models, methods, and tools on the basis of semantic technologies supporting this process. These are bundled in a social semantic desktop and exemplified within three scenarios within the application domain "rapid prototyping".

1 Introduction

In research, traditional (process-oriented) approaches to knowledge management have not been implemented yet. This is due to various reasons. On the one hand, scientific work processes are heavily variable and therefore it is hard to predict the need for knowledge. On the other hand, in research, structures like a common vocabulary are usually in the process of creation rather than already existent. Those structures are usually created within the scope of research cooperation.

For the BMBF research project "Im Wissensnetz" [transl. In the Knowledge Web] we chose a conceptually different approach to face in particular the challenges of application-oriented research: scientific work in and between interdisciplinary projects has to be understood as a network of informal learning processes with the researchers, their individual knowledge, and their mutual interactions constituting the center. In this context, it is important to reduce the barriers of such informal learning processes. The reduction of barriers can be accomplished through improving networking among people who are able to contribute to an overall scientific solution due to their individual competences. Additionally, it is also possible to improve informal learning processes through linking content in order to improve the benefit and reusability of already realized results cross-organizational in new information processes.

For this purpose, we elaborated the model of the so-called "Knowledge-Added Process" (KAP) (section 2). In a next step, we will present three user scenarios for applied scientific work in section 3. They exemplify needed models, methods, and tools supporting and efficiently designing the KAP. In section 4, we will present in detail two of these tools, which are bundled in a social semantic desktop, before we conclude in section 5.

2 Knowledge-Added Process Model

From the analysis of existing (cooperation-) processes, information and knowledge exchanges, and instruments for the preservation of knowledge in the studied application domain “rapid prototyping”, several requirements emerged: flexible composition and support of processes which have to be executed cooperatively, a qualitative improvement of distributed searches in diverse collections of information, methods and technologies for the effective identification and support of multidisciplinary expert teams, and the cooperative creation of jointly usable systems of concepts.

As an initial point, we developed the KAP on the basis of constructivist information processes (cf. [7]). This process depicts a macro-phase model to structure scientific work (cf. Figure 1):

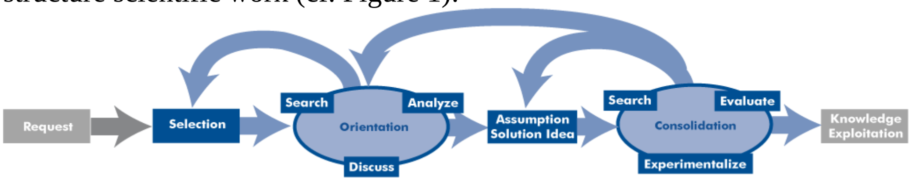

Figure 1: The Knowledge-Added Process

- **Phase I: Request.** The KAP’s initiation results from a concrete request, e.g. a costumer request or a project task.
- **Phase II: Selection**. For handling the incoming request, it is necessary to get an overview at first. Therefore the own know-how is the most important basis. First appropriate contact persons are selected, who can further analyze the problem or to whom the request eventually can be delegated.
- **Phase III: Orientation.** Searching, analyzing and discussing characterize this phase. The given problem is analyzed and information for orientation is searched, e.g. to acquire missing but necessary knowledge or to attain first possible solution ideas. As before, the own know-how is very important here, but also the cooperation with other people, e.g. in discussions. Therefore, it is necessary to identify adequate contact persons inside and outside the own organization.
- **Phase IV: Assumption/Solution Idea.** With the knowledge obtained from the orienting investigations first assumptions about the supposed problem and possible solution ideas are generated.
- **Phase V: Consolidation.** The proposed assumptions and solution ideas are evaluated on different levels and consolidated by focused search, analyzing the results and experimentation.
- **Phase VI: Knowledge Exploitation.** The newly learned knowledge, gained during the whole KAP, is exploited as project results in (in-)formal reports and provides a new resource for following researches in other KAP or is outset for a new process.

The phases within the KAP determine the requirements for the behavior and adaptation of methods and tools, which shall efficiently support scientists in their collaboration and knowledge creation. A social semantic desktop, which is tailored to the special needs of researchers, bundles the developed approaches. Core issues are the usability, and extensive flexibility of the tools, as well as the integration into existing working processes of an individual researcher (cf. [2]). This we want to illustrate in three scenarios in the following.

3 Scenarios

3.1 Scenario I: Linking People

One of the major problems in scientific work is *identifying key persons* and potential *research partners*. Here, expert finder applications are usually provided. Recently, social networking platforms like Xing [10] got very popular, too. But it is not always the expert one wants to contact. Factors like sympathy or trust play a decisive role, too. Missed so far is a qualitative assessment of these contacts in order to balance the "expert status" with the quality of the social relationship towards the potential "expert" and thus to provide relevant results. As a result, a colleague and good friend next door, being somewhat competent in the area, could be a much better result than the ultimate expert viewed as a rival. Therefore, we develop an egocentric network perspective extending the approach of social networking platforms with personal relationships to the fore thus their quality determines the ranking.

S*upport for knowledge transfer and synergetic cooperation* is needed, because the individual know-how has to be acquired in time- and cost-consuming learning processes, and users are willing to rely on explicit opinions of other users, as these opinions represent e.g. a form of guidance for novices. Tools for shared and collaborative information seeking are required where scientists from different domains can cooperatively execute a search, thus the individual knowledge of both can be incorporated into the search. Here, the WIDE tool provides a good basis for simultaneous collaborative and semantically enriched information seeking support (cf. section 4.2). Besides, social bookmarking services like del.icio.us [6] enable collaborative development of subject areas by tagging online resources. However, these tools lack better control from whom to get and with whom to share the findings. A combination of collaborative services with social networking platforms would be more useful; directly linking your contacts and the users in social bookmarking systems.

This further supports *tracking activities and advances of others*. By subscribing their bookmark collection, e.g. as RSS feed, it is possible to gain an insight in their work and interests. Directly tagging people and organizations would be useful especially for tracking people not participating in social networking platforms; further enhanced with a proactive information provision; e.g. by fading in potential contact persons during information seeking or by notifying the user if changes in the profiles of personal (selected) contacts occur.

3.2 Scenario II: Linking Contents

Just as tedious as finding an adequate contact person is finding adequate contentual resources. Users have to access many various data sources with different interfaces, but also the internet with common search engines like Google; here, a *focused search within open and distributed networks* is needed. For this purpose we are developing semantically enriched and adaptive add-ons for search engines. Supported by background knowledge and domain ontologies, it is possible to find out the search context and thus to extend or refine the search in order to reduce irrelevant results and to guide the user. This approach further enables a *context-based faceted search* for worldwide innovations and technologies.

In order to achieve such a semantic-based support, it is also necessary that domain experts build up and maintain underlying ontologies themselves. However, current ontology development methods are geared towards ontology modeling experts and not towards domain experts. Thus, the domain experts have no chance to act on their need for modifications on the ontologies directly—instead they depend on the modeling experts. This leads to frustration in cases when incomplete (e.g. because of too big time delays) or misunderstood contents (e.g. because of communication problems) in ontologies do not help in fulfilling their needs. This is especially crucial for research domains, which are fast changing and where ontologies emerge in the daily work of the users and underlie continuous evolution through collaboration and interaction among the users.

Therefore, we developed the Ontology-Maturing Process, which allows a work-integrated and collaborative view on ontology building (cf. [1, 3]), and which further acknowledges that ontologies cannot be formalized from scratch, but rather mature gradually from informal tags to formal taxonomy hierarchies. This model identifies four characteristic transition phases. It starts with the emergence of ideas from each individual as new concept ideas or informal tags and the consolidation in communities for a common terminology; going through the formalization by relations, which help in creating formal lightweight ontologies; to axiomatization in the last phase resulting in heavy-weight ontologies. The model further pinpoints triggers for technical integration points with actual work processes. In order to enable and foster this ontology maturing process, we are developing lightweight, collaborative and work embedded tools, which lower the barriers to ontology editing for non-modeling experts and will be presented in section 4.

3.3 Scenario III: Creating Added-Knowledge

The third scenario mainly has a future-oriented focus. As aforementioned, information seeking is very iterative and experienced based, but these experiences are not effectively exploited neither on personal nor interpersonal level. Developing, sharing and reusing *know-how* is yet another crucial factor. For instance, identifying the gap between the current state of research and requirements or requests from industry would help to derive future needs, which should be in turn most pressing research issues for universities and research institutes. In this context semi-automatic summarization, structured analysis, clustering and the identification of correlations within and between scientific papers or patents (e.g. based on a shared conceptualization of a domain, represented in the form of an ontology) would be very useful for quickly getting a full-fledged overview on a particular research field, its complex interrelations and so far unsolved research issues. In order to overcome identified gaps between industry needs and current state of research, the next step would be to (technically) support idea generation and to foster innovations. The system-based identification of reusable or even complementary approaches as well as their subsequent recommendation and adaptation (if necessary) would e.g. help researchers to generate more efficiently solutions for new and so far unsolved problems. In this context, an intelligent system could furthermore assist in identifying the potential of approved methods from a specific domain to be applied to another one.

4 The Social Semantic Desktop

Personal Knowledge Management (PKM) approaches like [8] follow the desktop metaphor by integrating different tools for knowledge work ("Semantic Desktop"). We have augmented this idea in the respect that this desktop (and the tools it groups together) needs to be aware of (a) the process phase the user is in and (b) of the individual's social environment (cf. [2]).

In the following we present two tools of the social semantic desktop in detail. These tools focus in particular on the first and second scenario to support scientists in their collaboration and knowledge sharing.

4.1 Semantic Social Bookmarking with SOBOLEO

SOBOLEO (**So**cial **Bo**okmarking and **L**ightweight **E**ngineering of **O**ntologies) combines a web-based lightweight and collaborative ontology editor with an ontology-enabled social bookmarking system (cf. [1, 3]). Its goal is, on the one hand, to support the collaborative exploitation of subject areas, and thus the knowledge transfer and cooperation between the scientists, and, on the other hand, to support the ontology maturing process, thus to facilitate the transition from work processes to ontology building.

Therefore, SOBOLEO enables people working together in one domain to develop a shared vocabulary and a shared index of relevant web resources (bookmarks) whereas the vocabulary is used to organize the bookmarks. That means collected bookmarks can be annotated with concepts from the vocabulary (cf. Figure 2).

Figure 2: Annotating a web resource with concepts

If a needed concept does not exist in the shared vocabulary or is not suitable, the users can either modify and existing concept or create a new one by simply specifying arbitrary tags. These tags are automatically collected and added as so called "prototypical concepts" to the vocabulary. If the users want, they can change to the ontology editor, but due to the automatic addition, they can also continue in their work and consolidate the new tags later by abstracting them into concepts and placing them within the ontology. In this way, we allow on the one hand that new concept ideas are gathered seamlessly when they are occurring and, on the other hand, that the users can define them freely and informally without the usual modeling overhead.

Changing to the ontology editor offers the view in Figure 3. It shows a tree view of the vocabulary on the left hand side of the screen. For structuring the vocabulary we concentrate on taxonomic relations between the concepts. These are easy understandable for non-modelling experts. SOBOLEO uses the SKOS Core Vocabulary [4] as format. It allows connecting the concepts by "broader", "narrower" and "related" links and specifying one preferred label, several alternative (synonymous) labels, and a description for each concept. Thus, the tree view shows the concepts with their preferred labels and their narrower and broader relations.

The new informal tags, which came up during the annotation, are collected in the special branch of "prototypical concepts". The users have now the opportunity to consolidate and place them within the taxonomy, e.g. by drag'n'drop them on other concepts or via the details view in the center of the interface. Within the details view, the users can edit the labels, the relations to other concepts and the description of the concept currently selected in the tree.

The right hand side of the screen provides a message pane that informs about all changes done to the taxonomy by any user. At the same time, it acts as chat window for having a conversation with other users currently editing the taxonomy.

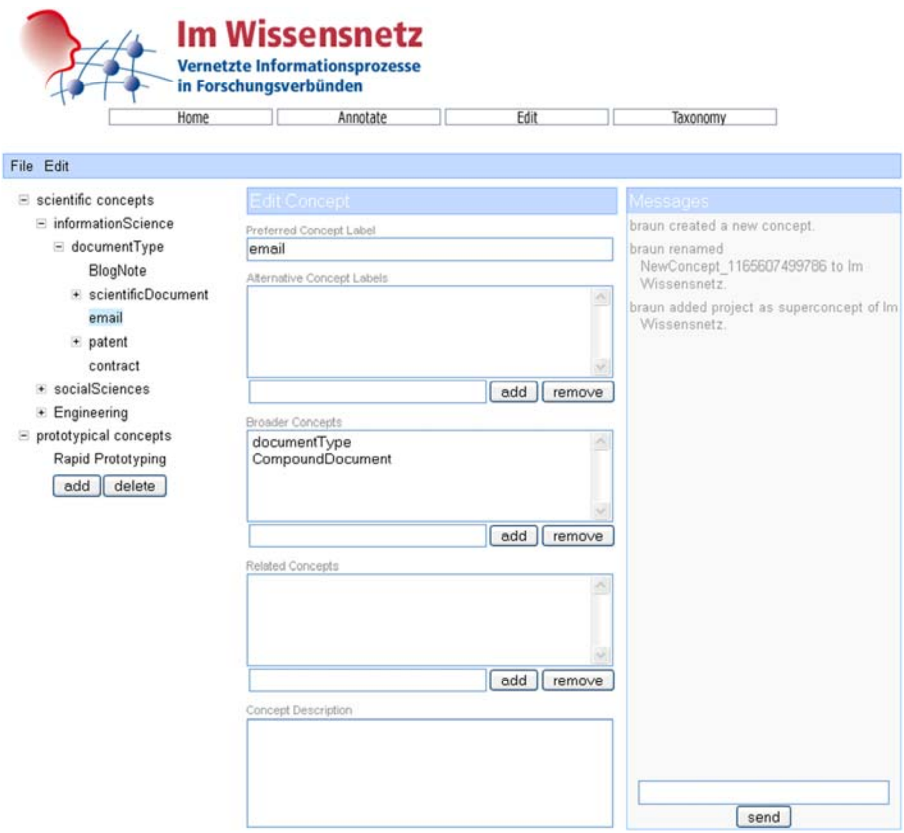

Figure 3: Ontology editor user interface

Besides the collaborative ontology editor, SOBOLEO enables the exploitation of the shared bookmark collection by a semantic search engine or by a taxonomy browser (cf. Figure 4).

The search engine allows for easy search and retrieval of collected bookmarks. It is used by typing search terms into a text field—similar to common internet search engines (cf. left part of Figure 4). The entered search string, however, is analyzed for occurrences of concept labels from the underlying taxonomy. If concepts can be identified, the search engine searches for web resources annotated with these concepts or narrower ones. The results from the semantic search are further combined with the result from a full text search in the contents of all annotated pages. The interface lists all found resources with their title linking to the original page, with annotated concepts, a short excerpt of the page content with highlighted search terms and the exact url. The system may also provide further query refinement and relaxation proposals.

With the taxonomy browser the users can navigate through the taxonomy and associated bookmark documents (cf. right part of Figure 4) and thus discover new interesting resources. Starting from the root concepts, the users can click through the taxonomy concepts. On top, the users see the currently selected concept with its labels and description. Additionally, all its broader, narrower and related concepts are displayed as links for further navigation. Underneath, all resources are listed, which are annotated with the currently selected concept or with one of its narrower concepts. These resources are further ranked by their date they were collected, thus the newest resources appear upmost.

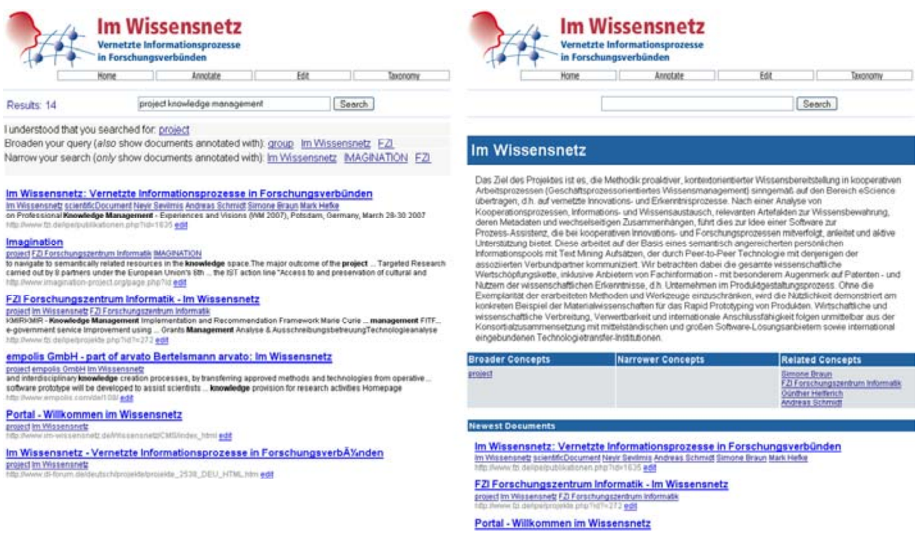

Figure 4: Searching the bookmark collection (left) and browsing the taxonomy and associated bookmarks (right)

4.2 Application Integrated Semantic Annotations

In parallel with the development of SOBOLEO, which is well suited for annotating web resources, we have also concentrated our focus on application integrated approaches in order to be able to annotate and share documents in different file formats that are not only available in the web but also on the local computer. Thus, we have investigated and developed a natural and comfortable interaction technique to support intuitive work centred operations, together with non-invasive knowledge creation and augmentation support, including support for semantic annotation, automated metadata harvesting, that adapt to work conditions aiming to hide the underlying complex machinery from the end-user.

To better support the knowledge creation through semantic annotations in scientific work processes, we have integrated the annotation functionality into the applications effectively used by knowledge workers in their daily work processes. The semantic annotations are based upon collaborative lightweight ontologies shared with SOBOLEO, thus enabling semantic interoperability. To keep the creation of "in place" annotations as simple as possible, knowledge workers can make handwritten notes during reading an article or presentation using concepts from the shared ontology (cf. Figure 5).

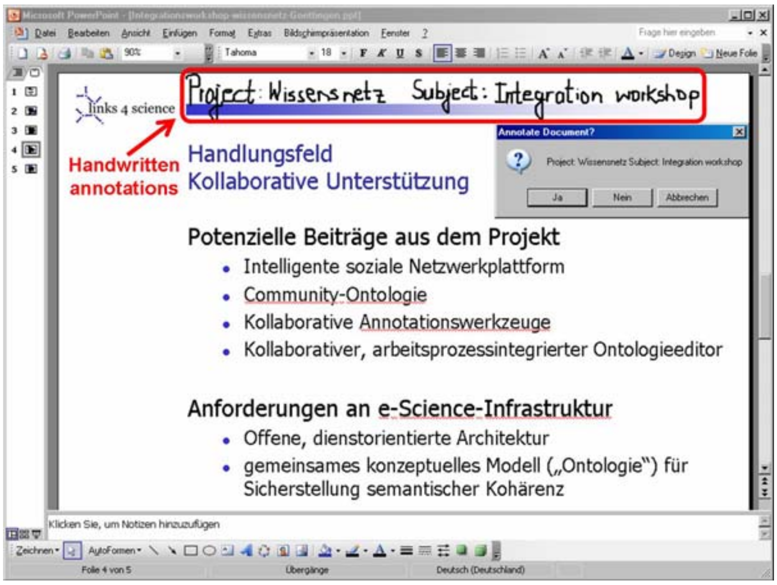

Figure 5: Handwritten annotation inside a MS PowerPoint presentation.

The annotations are based upon tuples, each of which consists of an ontological concept with a corresponding value assignment. For example, in Figure 5 the user made two tuple-annotations: "Project: Wissensnetz" and "Subject: Integration workshop" where "Project" and "Subject" are concepts in the shared ontology, and "Wissensnetz" and "Integration workshop" are corresponding values. Obviously, the user is meaning that this MS PowerPoint document is related to the project "Wissensnetz" and has the subject "Integration workshop". Once a document is annotated, the system recognizes the handwritten input and displays the recognized string in a pop-up window to the user asking for confirmation. The user can then verify the recognized string and confirm or correct. A confirmation entails an automatic metadata harvesting, reasoning, and knowledge synthesis, so that it can all be done in ways that do not interrupt what the user is working on. In doing so, the user's annotations plus some automatically acquired metadata (e.g. the location of the document, the concrete date of the annotation, user name, and document type) are used to automatically create RDF instance graphs. For example, "Project" and "Subject" are defined to be concepts in the shared ontology. Thus, the system automatically identifies them as concepts by asking and interacting with the shared ontology. Both concepts are then automatically instantiated by taking the corresponding tuple values "Wissensnetz" and "Integration workshop" into account, and interlinked with the annotated document. The generated RDF instance graphs are then stored in a commonly used database (using Sesame's RDF store [5]). In this way, we automatically establish a logical combination of shared ontological concepts used in the annotation, which forms the basis for subsequent semantic search.

In case the user used in his tuple-annotations any concepts that do not exist in the shared ontology, SOBOLEO is automatically invoked and users can then create new concepts according to their needs in a natural and simple way (cf. section 4.1).

However, for knowledge sharing, the different kinds of documents augmented with knowledge using common vocabularies have to be searchable and the results have to be presented in a way that is tangible to the user. For this reason, we have integrated the WIDE UI [9], which can be used in combination with the application integrated annotation tool. It operates on the same RDF store. Its user interface provides a graphic front end to the user and supports the development of user queries in an alphanumeric way. Furthermore, it presents the returned results and their relationships (semantics) in a graph-based structure. This graph structure can be navigated by the user in order to explore the returned results and their metadata (cf. Figure 6).

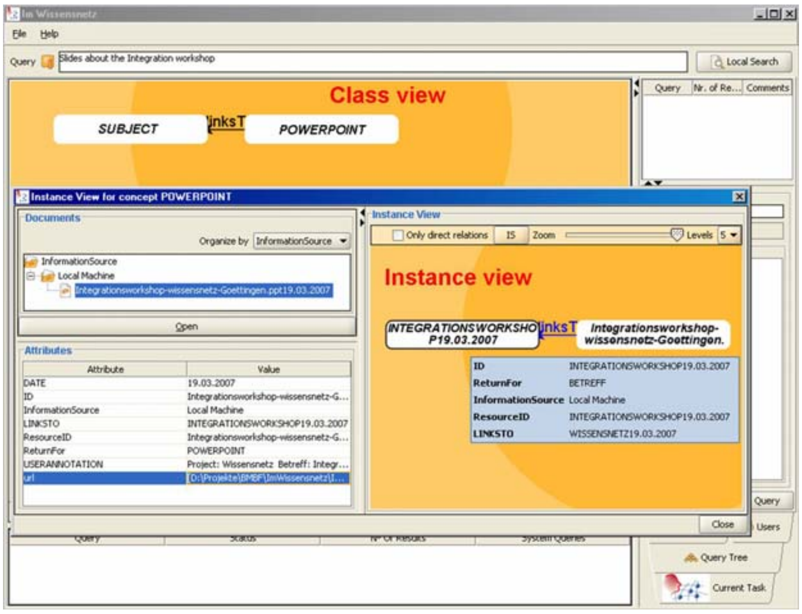

Figure 6: Searching and visualizing results using the WIDE UI.

Figure 6 shows the results for the user query "Slides about the integration workshop". The results are visualized in two different ways: class view and instance view. The class view shows the result graph by using concepts from the shared ontology whereas the instance view visualizes the interlinking between the corresponding instances.

In the "Im Wissensnetz" project we have integrated the presented annotation functionality in MS Word, MS PowerPoint and PDF because these file formats are the most prevalent ones in the project consortium.

The "in place" semantic annotation support is easy to integrate into the user's daily work process and it does not overstrain the user much because humans are used to make handwritten "semantic note taking" in paper-based documents from school age on up. Knowledge can be created and associated with corresponding documents in an easy to use manner. In particular, we attached great value to relief the user from the burden to get familiar with complex third party tools that are designed to make ontology-based annotations in a non-intuitive and invasive way.

5 Conclusion and Future Work

In this paper, we have worked out the model of a “Knowledge-Added Process”, a new paradigm of process-oriented support in e-science. Moreover, models, methods, and semantic-enabled tools are developed. These provide means to support single activities of the process and are bundled in a social semantic desktop. They are illustrated in the context of three selected application scenarios, which have been derived from an exhaustive analysis of the application domain “rapid prototyping”. We presented two of these tools in detail—the SOBOLEO tool for semantic social bookmarking and lightweight ontology engineering and the tool for application integrated semantic annotations—to show their flexibility and integration.

For the future, we plan to synergistically combine state-of-the-art approaches with new developed methods and technologies in order to efficiently support collaborative scientific work processes. Using an iterative process, the developed approaches of the tool providers are validated and evaluated in close collaboration with users within the project “Im Wissensnetz”. Experiences and user feedback are directly flowing into further development of our approaches in order to finally achieve a flexible and adaptive toolset, as well as a process model for enabling collaborative research work, which can be adapted to any other research area.

Acknowledgements: This work was supported by the German Federal Ministry of Research & Education within the project “Im Wissensnetz”.

References

1. S. Braun, G. Nagypal, A. Schmidt, A. Walter, V. Zacharias: “Ontology Maturing: a Collaborative Web 2.0 Approach to Ontology Engineering”. World Wide Web Conference (WWW’07), Workshop on Social and Collaborative Construction of Structured Knowledge (CKC) Banff, Canada, 2007 (to appear)
2. S. Braun, A. Schmidt, M. Hefke: “A Socially-Aware Desktop for e-Science: Supporting Learning in Networked Scientific Processes”. Professional Knowledge Management – Experiences and Visions (WM’07), Workshop Collaborative Knowledge Management (CoKM) Potsdam, Germany, pp. 47-54, 2007
3. S. Braun, A. Schmidt, V. Zacharias: “Ontology Maturing with Lightweight Collaborative Ontology Editing Tools”. Professional Knowledge Management – Experiences and Visions (WM’07), Workshop Productive Knowledge Work (ProKW) Potsdam, Germany, pp. 217-226, 2007
4. D. Brickley, A. Miles: SKOS Core Vocabulary Specification. W3C working draft, W3C, Nov. 2005
5. J. Broekstra, A. Kampman, F. van Harmelen: ”Sesame: A generic architecture for storing and querying RDF and RDF Schema”. International Semantic Web Conference (ISWC) Sardinia, Italy, pp 54-68, 2002
6. del.icio.us. http://del.icio.us (accessed 07/03/20)
7. C. Kuhlthau: “Seeking Meaning: A Process Approach to Library and Information Services”. 2nd edition. Libraries Unlimited, Westport, CT, 2004
8. Sauermann, L.: “The Semantic Desktop – a Basis for Personal Knowledge Management”. International Conference on Knowledge Management (I-Know ’05) Graz, Austria, pp. 294-301, 2005
9. N. Sevilmis et al.: “Knowledge Sharing by Information Retrieval in the Semantic Web”. In: Gómez-Pérez, Asunción (Ed.) u.a.: The Semantic Web: Research and Applications. Proceedings. Berlin, Heidelberg, New York: Springer Verlag, 2005, pp. 471-485 Lecture Notes in Computer Science (LNCS) 3532).
10. Xing. https://www.xing.com (accessed 07/03/20)