Increasing Search Quality with the Semantic Desktop in Proposal Development

Mark Siebert[1], Pierre Smits[2], Leo Sauermann[3], and Andreas Dengel[3,4]

[1] Siemens Business Services GmbH und Co OHG, Munich, Germany
Mark.Siebert@siemens.com
[2] BTU Cottbus, Germany
psmits@informatik.tu-cottbus.de
[3] KM Dept, DFKI[1], Kaiserslautern, Germany
{sauermann, dengel}@dfki.de
[4] CS Dept, DFKI, Kaiserslautern, Germany

Abstract. Quicker response times and less production costs of proposal development require further automation of sales assistant functionality in CRM environments. Automation still struggles with the handling of abstraction and the subjective character of knowledge. Based on the knowledge creation framework the paper outlines and tests the increase of search quality with Semantic Desktop technology. The discussion of peer-to-peer settings and semantic concepts illustrates the influence of individual perspectives on search quality. It reveals first potentials and benefits for process-integration, like semantic CRM and illustrates approaches to increase knowledge worker's productivity.

1 Introduction

Quicker response times to proposal requests and *less production costs for standard proposals* without quality reduction are current CRM requirements within increasing competition and market dynamics in the IT service market. This aligns with Davenports petitions to increased knowledge worker's productivity [5]. He enhances the traditional optimization triangle of processes, IT and people by their physical working environment and their personal networks.

Responding to customer requirements sales managers today either search for similar, existing, and successful standard proposals or ask an assistant to come up with a good draft. Other than proposals in product business, service proposals require a value proposition derived from and designed to the individual customer needs rather a value proposition of the product characteristics.

Existing knowledge management tools or Proposal Automation Tools [15] already support general functionality like document handling and proposal generation. They lack deeper process integration, higher quality of search and respect of the individual characteristics of knowledge. With its new on-demand CRM platform SAP [11] provides a virtual sales assistant guiding the user through the steps within the sales process (e.g. creating a value proposition or analyze competitor's products) and

[1] German Research Center for Artificial Intelligence, http://www.dfki.de/km

U. Reimer and D. Karagiannis (Eds.): PAKM 2006, LNAI 4333, pp. 279 – 290, 2006.

offering personalized information (like my reports, contacts, appointments, tasks, etc.). This provides a sufficient process integration and personalization of the interface but does not support the knowledge creation (e.g. a proposal) sufficiently. Sales assistants though still require extensive search capabilities to manage different information sources and to translate the customer requirements into searchable key words.

Interviews with SBS practitioners name the following reasons:

- Heterogeneous storage paths with wording differences between peers (e.g. Sales Manager uses customer - and Proposal Manager uses organizational language),
- Inexistence of knowledge assets in the central knowledge base and insufficient meta-data (kept on local desktop due to high publishing efforts),
- Insufficient dialog between roles due to communication hurdles misses respect of different perspectives, increases proposal risks and lowers quality.
- Roles are only designed from a process-related point of view defining tasks. Intentions and backgrounds (expert level) are not respected.

Recent research provides with the *Knowledge creation framework – KCF –* [14] a base for understanding the knowledge creation process as a combination of perspective taking and making. It describes six steps to develop knowledge assets (e.g. a management summary of a proposal) from an individual point of view - from gathering and mapping search requirements based on full-text (Receiver), over classifying and categorizing them through ontologies (Interpretator) up to consistent alignment through inferences (Analyzer), content selection (Reconciler), verification (Verificator) and production (Producer). The paper uses the Semantic Desktop prototype "Gnowsis" [7] to explore the possibilities of semantic search and peer-to-peer technology. It supports the three perspective taking steps of the framework in practice at Siemens Business Services: Receptor using full-text search, Interpretator exploring concepts and Analyzer applying inference rules. The perspective making process requires further investigation into reasoning and problem-solving capabilities.

The KCF works on different levels of abstraction. Therefore, Johnson-Laird suggests three levels of representation [10]: first *propositional representations*, which are pieces of information resembling natural language; second *mental models, which* are structural analogies of the world; third *mental imagery*, which are perceptual correlates of models from a particular point of view. Our approach of representing information in a Personal Information Model uses the first level of abstraction along the KCF.

2 Semantic Desktop "Gnowsis" Enabling KCF

In further development of full-text and associative search like Google and LiveLink, semantic technologies like Ontoprise[2] or the brainFiler[3] explore hidden, sub-symbolic relations and make them explicitly usable. Gnowsis [12, 13] additionally investigates peer-to-peer networks in enlarging the central knowledge base like SWAP [6] Based on an index service - Receiver - it combines organizational and individual elements

[2] http://www.ontoprise.com
[3] Developed within research project EPOS at DFKI: http://www.dfki.de/epos

within a Personal Information Model (PIM) [1] – Interpretator - and integrates rule-based (Analyzer) different information sources (e-mail, files, ontologies, web, etc.). Sauermann et al [13] define Semantic Desktop as a device in which an individual stores all her digital information like documents, multimedia and messages. All data is accessible as RDF graph.

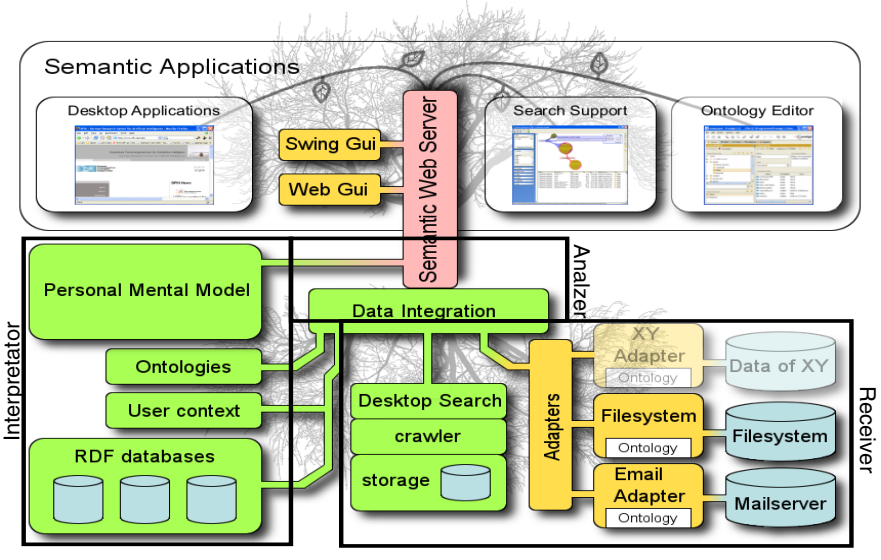

Fig. 1. Semantic Desktop Architecture [13]

Receiver: The Semantic Desktop uses the software brainFiler [8] providing the index service for all clients and determining structure and concepts of the indexed files. The brainFiler therefore creates term vectors of each file and cluster documents initially with their respective origin folder - see [1] - to improve the precision of a search result.

Interpretator: Information can be found in structured and unstructured forms with manual or automated annotations [11]. Information without knowledge structure or meta-data can be retrieved through fulltext search (e.g. Google Desktop). LiveLink uses manual annotations and meta-data to structure the content. Semantic technologies, like Semantic Miner or brainFiler use semi-automated annotations to automatically structure the content. With the automation, they respect the fact that information is stored and retrieved based on the individual preferences and priorities. These preferences lead to different perception of identical content, duplicated information in different categories of folders and information in a one-dimensional folder structure. Due to different expectations about the content of the document, they are hard receivable.

Analyzer: The architectural flexibility allows replacing the brainFiler by any other index engine. BrainFiler and Gnowsis apply rules via the Jena inference engine (http://jena.sourceforge.net/inference/) and SPARQL (http://www.w3.org/TR/rdf-sparql-query/) queries on RDF models to infer from categories. Other semantically enabled products use F-Logic to integrate rules.

3 Increasing Search Quality with the Personal Information Model (PIM)

SBS produces many thousands proposals a year for its whole service portfolio, ranging from outsourcing to solution design and system integration projects. Therein sales teams consist of different roles, like proposal managers and sales managers, with different backgrounds, expert levels and functional tasks. Sales managers feed the first rough information, like request for proposal, together with the approach (e.g. price, competitive environment, etc.) into the proposal factories to gain a first draft story – like a management summary. The result is based on existing information and references, leading to open topics and requirements for further refinement.

Proposal managers work with organizationally pre-designed and re-usable content structures (see Fig. 2, left) in a central repository. Sales managers often work in individual settings (see Fig. 2, right) according their customer requirements and area of responsibility (e.g. sales requirements from public sector are different from e.g. software business) mostly on their desktop. Nevertheless, both work on same or similar documents, value propositions and rely on similar information pools (e.g. LiveLink) with about ten thousands of knowledge objects.

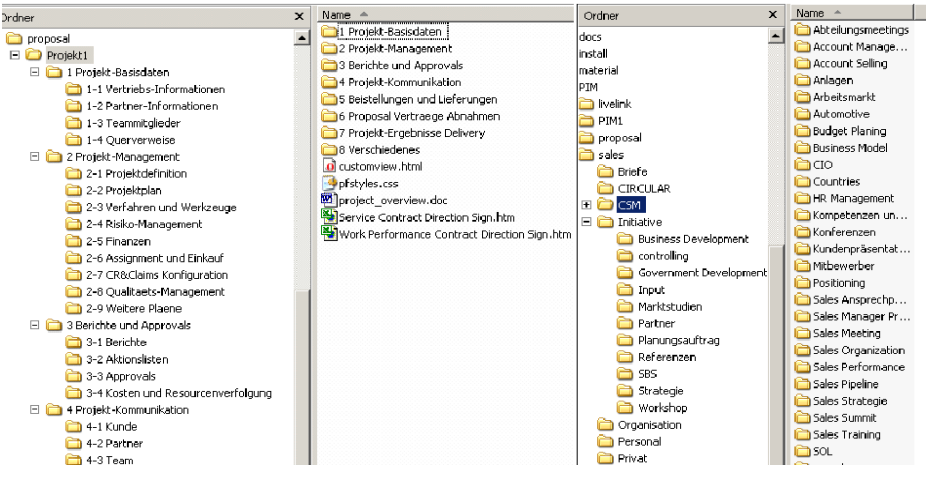

Fig. 2. Folder structure of proposal manager (left) and sales manager (right)

According to Elst et al. [7], a combination of different ontologies represents the organizational setting: information-, customer-, organization-, domain- (customer-, product-) and workflow-related ontologies. Transferred to the SBS example, role (intention) and process (situation) define a knowledge space. It includes beside the customer (e.g. industry), workflow (e.g. templates) and document (e.g. title, creator, and publisher) agreed organisational knowledge domains (e.g. document type or portfolio) and personal information structures (folder structure e.g. customer-, region-, event-based).

Together with the organizational structures, the PIM combines both folder structures (Fig 2) and represents the users' role and expert level as a model of the users' subjective perspective. The folder structures will be represented as PIM ontology in Protégé (http://protege.stanford.edu/) and integrated in Gnowsis. One user might have different roles, which end up in a mixture in his personal workspace structure. To avoid this and not to divert the results of this work, it is assumed that one user has one role and his personal workspace represents the perspective and intention of this role.

However, the model we use (PIM) is not restricted to representing a single role of a person. Instead, using a context-aware approach that activates a certain role of the user when the user is acting in this role is in principle possible. This is based on the subjective view on above ontologies from the organizational setting: the PIM is a personal view on these. It is a mediator between the mental model of the user, and the documents of the company. The current role of the user is modelled using the PIM and RDF, and then the relations of documents and projects to this role can be captured using RDF links. For our experiment, only a very limited PIM was constructed, it is a step towards representing the user's mental model.

In the Analyzer the PIM could be expanded with rules and inferences to further retrieve information about or adjust the display of the search hits. Here, deriving from the SBS proposal requirements, we identified four simple rules:

- If a project was found, determine project leader and author as a contact
- If an entry, fitting the current customer requirements, was found, determine the relevant project(s) as a good reference
- If an entry, fitting the currently required solution, was found, determine the referenced project
- If a project was found, determine the project documents as a possibly good source document for the current task

Those rules are stored in Gnowsis using forward-chaining rules:

```
# Example for reatrieving the project manager as expert contact
(?hit retrieve:item ?project),
(?project rdf:type org:Project) ->
querySparql('
CONSTRUCT {
?x1 org:HasProjectmanager ?m. ?m rdfs:label ?labelm. ?m
rdf:type ?typem.
?x1 retrieve:todoRelateHitTo _:hit .
_:hit rdf:type retrieve:InferedHit .
_:hit retrieve:item ?m .
_:hit retrieve:textSnippet \'Projektleiter\'.
} WHERE { graph ?g {
?x1 org:HasProjectmanager ?m. ?m rdfs:label ?labelm. ?m
rdf:type ?typem.
} }
', ?project).
```

In SBS practice usually the simple rules appear context-related in combination, like: "If current role is sales manager and current task is proposal development and the found document is stored on a desktop of, or written by, a user identified as an expert about the searched topic, determine further details (e.g. assigned project, source documents and co-authors) about the document".

A full implementation in practice would have to model those combinations based on the PIM ontologies and classes. Therefore, for each class and ontology its relation to others has to be pre-defined (e.g. expert-level → show only certain document types).

The Gnowsis web interface, described in [12] is used for accessing the local and peer knowledgebase. It provides a search field for inputs and checkboxes to select peer or local search. The result page shows search results as a headline summary in their respected classes (e.g. concept, document, project, event, persons, etc.) and as detailed list with key word summary. For each item the "browse" and "link" buttons provide additional information, like members or manager of a project, in a popup box.

4 Information Retrieval Analysis Methods to Test Semantic Desktop

Different forms of collaboration exist between sales and proposal managers depending on the combination of peers (different roles and perspectives) as well as availability and similarity of objects (different peers and knowledge objects). They help to outline four use case scenarios, following process operations at Siemens Business Services. The mix between full-texts, concepts and rules in each of the scenarios influence the search results. Their differences will be investigated and discussed postulating that role and perspective will have a specific influence on the search quality and information relevance.

The following chapters will investigate and test the influence of different concepts and rules with key word examples (e.g. Help Desk, Cost Reduction, etc.) from the use case of drafting management summaries for proposal development.

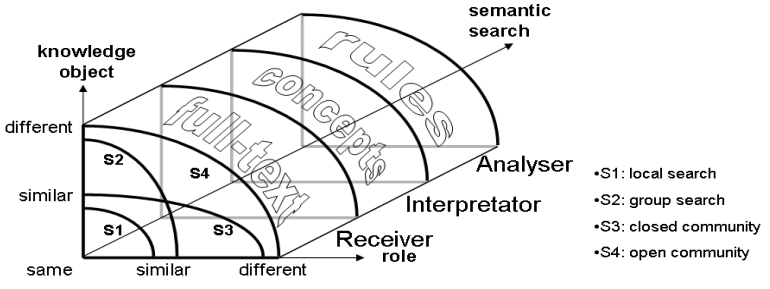

Fig. 3. Test scenarios in dependence of roles and knowledge objects

S1 - Local search: In direct comparison to Google Desktop one role (e.g. sales manager) searches his own desktop and the organizational database, now applying his native structures within the PIM. He can benefit from improvement of the precision and relevance of his results.

S2 – Group search: Similar roles working with mostly different topics and files (e.g. customers), within the same domain (e.g. calculation, trends, solution design, etc.). To optimize the process a similar vocabulary and working structure is organizationally applied. They could benefit from each other offering complementary information, which might influence one's own work.

S3 – Closed community: Different roles working with similar topics and files (e.g. around business development within a customer community). Especially customer communities are weak structured working frames. They tie together people with different roles (sales manager, solution designer, project manager, etc.), all dealing with the same customer. Some documents might be stored centrally but most information will be found on individual desktops, as they are not seen worthwhile sharing from an individual perspective. Everybody still works primarily in his own working space. Some information is shared within the community space. They could benefit from each other offering complementary information and increase of relevance as similar documents and information might be labelled differently in other peers.

S4 – Open community: Different roles working with different topics and files (e.g. Internet, Intranet). Everybody works primarily in his own working space. Information is shared from an organizational point of view. They could benefit from each other offering complementary information and increase of relevance as similar documents and information might be labelled differently in other peers. The size of the database and multiple roles might reduce relevance of search results.

Performing searches with different typical queries from SBS operations like "helpdesk", "call center", "customer centralization", etc., "cost reduction" seems to be a good example for demonstrating the particularities of the Gnowsis search. The results of "cost reduction" will be base for findings and discussion. Language and document relationships have to be considered in manual pre-selection:

1) Core organizational wording is mostly identical in English and German written documents. As no translation functionality is included yet, the test will focus on English search queries.
2) Three principles on how to structure and search for information, are used. First, a document is related to different categories, second, a document is related to characteristic words or phrases and third, a document is related to other documents.

For the test, SBS-internal non-restricted data (102 files) downloaded from the LiveLink knowledgebase is used completed with 23 anonymised SBS management summaries (representing proposal stories) and 30 manually created files (project plans, calculations, contacts and references). These parts represent the basic set of documents. The file structures are part of the user's subjective view derived from directory, email, bookmark etc. [4, 9]. The individual file structures of the respective roles with the given basic set of documents are filled according to different scenarios.

5 Influence of the PIM on Search Quality

Evaluation of retrieval systems uses according to Brünken [4] the following sets as base and measures search quality in recall and precision: M is the set of all relevant objects as part of all system objects, P is the set of retrieved documents and objects (search result), Ma is the set of retrieved relevant documents and objects (relevant search results).

Using these assumptions for PIM, Gnowsis increases search quality [9] through

1) availability of information (**recall**) identifying all of the currently required documents, which are not labelled with similar search key words (e.g. find the

project “customer 1” when searching for a topic called “cost reduction” because the project is related to the topic)

$$r = |Ma| / |M| \text{ with } 0 \; r \; 1$$

2) retrieval of no unnecessary documents (**precision**) respecting users’ role and intention in ranking and selection of appropriate object and assets (e.g. receive results from peers, who all labelled them similar, as higher ranked as those, who are labelled differently by different peers)

$$p = |Ma| / |P| \text{ with } 0 \; p \; 1$$

With this Fig. 4 illustrates the test results with different key words across the scenarios as influence of semantic concepts on recall and precision.

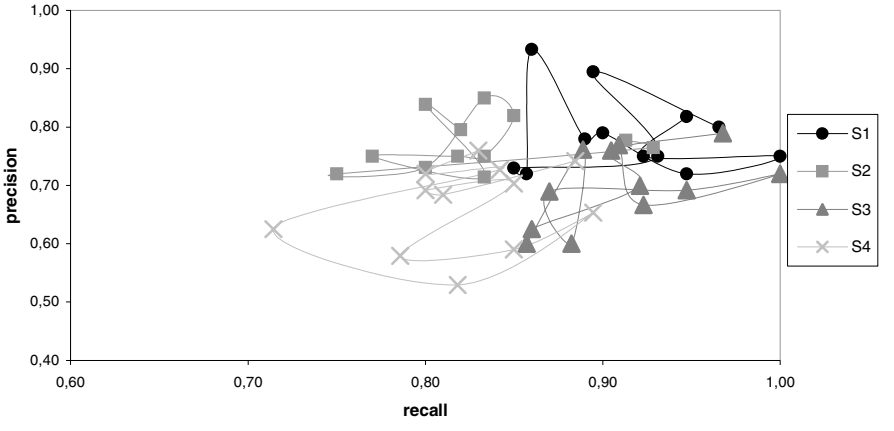

Fig. 4. Influences in precision and recall

Each point represents a key word. The key words are the same across the scenarios. The results form four different areas according to the scenarios. But overall, the tests proofed the trend that the Semantic Desktop increases search quality of existing full-text or associative searches in recall AND precision in an automated form. Full impact could be achieved in S1, whereas the results of S2 and S3 explain that a peer-to-peer approach has either high influence on recall OR precision due to an increased document base. The positive PIM and peer-to-peer effects seem to level each other out in S4. Further research would be required for this context.

S1: Local search

A system, like LiveLink, enables the user to retrieve ranked documents containing the exact search phrase and documents located in a folder containing the required phrase. Important meta-tags (like authors) of the documents are presented as well. The test with the local search of brainFiler for “Business efficiency” and “call-center” shows a higher recall-value (r=1,0) as Livelink (r=0,25), which might be caused by a good document index. However, this recall difference is valid for other key words as well.

If we apply the PIM, the test shows continuously high results in precision but lower results in recall. Reason is that the ontology only allows additional relevant hits but adds a big amount of possible documents.

Using rules within Gnowsis, the tests confirm with few exceptions (RFID, Communication) an increase of recall values with constant precision results. It indicates that rules bring additional value in search quality.

The illustrated effects with higher recall and precision require though a combination of full-text, concepts and rules - retrieving results (receiver: item about "cost reduction"), putting them into context (Interpretator: item is a "customer requirement") and applying rules (Analyzer: customer requirements are assigned to projects, "cost reduction" → "customer 1")

S2: Group Search and S3: Closed community

The value of the influences in S2 and S3 is determined by the search situation, requirements and goals.

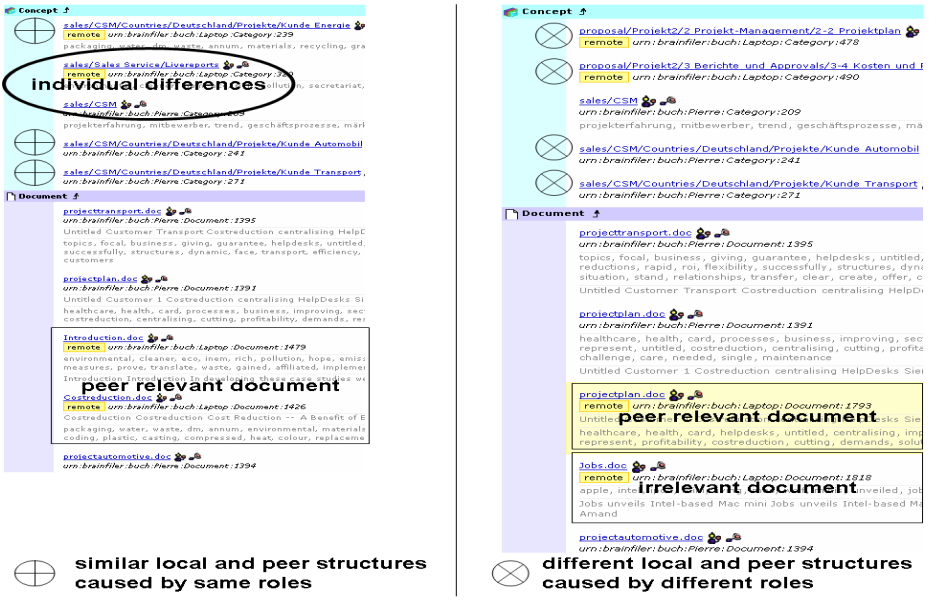

Fig. 5. Results of peer search S2 (left) and S3 (right)

For S3, for instances, it would already be helpful in the collaboration of a proposal and a sales manager to benefit from the knowledgebase of the other role in retrieving more relevant information from the "shared" knowledge space. For S2, for instances, it would already be helpful to receive better documents and rankings using the other's perspective but accepting an increase of irrelevant documents.

Users working in an organizational predefined structure (like proposal managers in S2) can take advantage of the previously mentioned local functionality. They can expand their benefits by searching peers with the same roles, based on similar PIMs

as a comparison among a group of SBS proposal managers showed. Accessing different peers enlarges the total amount of possibly relevant documents and classifications. Fig. 5 illustrates the differences between local and global search. Accessing other (similar) peers returned more relevant documents stored in similar structures. Thus, the tests show that searching peers on similar organized desktops can reveal additional, unknown information compared to local search, having positive effects on precision by "knowing" the perspective of the accessed desktops.

Changing the role (perspective) in S3 means changing the PIM assuming that the same relevant files are shared among the peers (sales and proposal manager). With this, the files are seen within different contexts, experiences and perspectives. This allows the system to either expand the search results as somebody linked the file to a relevant concept or rank them higher as most of the peers linked the file into a homonymous concept. Searching peers increases recall caused by a larger total knowledgebase and consideration of different perspectives/PIMs. Contrary to S2 it might increase recall, due to a smaller set of knowledge objects. Both scenarios illustrate the influence of PIM and the database (peer-to-peer) but show that they are reliant on each other and could only improve significantly one of the search quality indicators recall OR precision. Through the PIM it is possible both to model the role of the user in a way that can be reused across the involved systems (gnowsis and brainfiler). The subjective view that was created by one user (according to classifications that map to his mental model) can be correlated to the views created by others, again based on their role or categorization metadata. In the company setting where people agree to share such metadata, we see that the recall values are increased.

S4: Open community

Searching desktops with different files of users with different roles (sales and proposal managers) additionally could broaden the "mind" of the semantic desktop network e.g. by having access to many classifications of a document. Thus, a required document could be retrieved via several search queries. As S4 is a combination of the previous scenarios, the results are a combination of the list retrieved in the previous situations. It is a combination of all relevant and irrelevant documents. It underscores the expectation from S3 that peer search has still positive impact on recall and precision when searching peers of users with a different role (perspective). The positive impact on precision could be improved by further integration of the roles (abilities, skills, expert level and desktop structure) of each peer user into Gnowsis.

In summary, the role concept and individual perspective are major drivers to increase precision over decentral and different knowledge bases. Further optimization of the balance between recall and precision would be required. Precision depends on the degree of publicity of the searched roles within the local environment. Recall depends on the amount of objects within the knowledge base and increases less with the increased size of the knowledge base.

As the individual preferences are impossible to determine within the given situations they are assumed constant for further optimization of search quality. Additionally other criteria like the amount of knowledge objects is set constant as well, because they are mainly determined by given facts, which can't be influenced by the Semantic Desktop Gnowsis. Integrating the role / perspective into the search process is though the top criteria for improving search quality.

The results show that the semantic desktop can improve precision by considering the role of an individual workspace (S1 and S2). Deeper integration within a social semantic desktop could expand these to an even larger variety of roles (S3 and S4).

6 Conclusion and Summary

The tests proof the positive influence of the Personal Information Model, representing roles and perspectives, on search quality. Qualitative analysis furthermore emphasizes the impact of complete publicity of the roles information model on search quality. Semantic Desktop with this avoids additional editorial or communication efforts, which are required today without semantic search in proposal development.

Semantic search technology and modeling framework enables Johnsons-Lairds's [10] first level of abstraction and are a good base to integrate personal views into process-related, task- and competence-oriented role concepts from an informational point of view respecting the subjective character of knowledge. Future research will have to provide solutions for enabling second and third level of abstraction, like *mental models* and *mental imagery.*

This might add a piece to the shift from a reactive to an *active search support* and information provision, e.g. within the CRM Sales assistant from SAP ("related documents"). Displaying search hits in a process-dependent template structure of a management summary (market trends, business scenarios and customer requirements, compelling events, cost driver, solution, benefits) provides a pre-structured base for further content retrieval from the found documents. Based on the meta-data "document type", e.g. market information, could be shown in the category "market trends".

In total, semantic (handling of abstraction, translation and similarity engine), technology (agents, mash-ups, user observation) and architecture (handle peer-to-peer, content management, workflow systems) are the three elements driving further research in this area leading to better search results and reduced response times and lower process costs.

Acknowledgments

Parts of this work have been supported by a grant from the German Federal Ministry of Education, Science, Research and Technology (EPOS, FKZ ITW–01 IW C01). We gratefully acknowledge Heiko Maus for his review and discussion support.

References

1. Abecker, A.; Bernardi, A.; Maus, H.; Sintek, M. & Wenzel, C. Information Supply for Business Processes - Coupling Workflow with Document Analysis and Information Retrieval, in: Knowledge-Based Systems, Special Issue on AI in Knowledge Management, Elsevier, (2000), 13 , 271-284
2. Bähr, J.T., Dannenmann, P., van Elst, L., Hust, A., Lauer, A., Maus, H., Schwarz, S.: EPOS - Milestone 1. Document D-04-01, DFKI (2003)
3. Berners-Lee, T.: WWW Past&Future(2006) http://www.w3.org/2003/Talks/0922-rsoctbl/
4. Brünken, R.: Automatische Rekonstruktion von Inhaltsbeziehungen zwischen Dokumenten. Benutzeradaptives Informations-Retrieval in Wissensbasen, Shaker, (1998)

5. Davenport, T.: Thinking for a Living, Harvard Business School Press, 2005
6. Ehrig, M., Tempich, C., et al.: Swap: Ontology-based knowledge management with peer-to-peer, in: Izquierdo, E., ed.: 4th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'03), London, World Scientific (2003) 557-562 - http://swap.semanticweb.org
7. Elst, L., Abecker, A., Maus, H.: Exploiting user and process context for knowledge management systems, in: Workshop on User Modelling for Context-Aware Applications at User Modeling, Sonthofen, Germany (2001)
8. Hees, H.: Nextbot knowledge framework. brainbot Technologies AG (2004) (http://brainbot.com/site3/dokumente/nextbot_knowledge_framework_whitepaper.pdf)
9. Holz, H., Maus, H., Bernardi, A., Rostanin, O.: From lightweight, proactive information delivery to business process-oriented knowledge management, in: Journal of Universal Knowledge Management (2), (2005) 101-127
10. Johnson-Laird, P. N.: Mental models:Towards a cognitive science of language, inference, and consciousness. Cambridge, MA: Harvard University Press, 1983.
11. SAP, Video-Presentation CRM on demand system http://www.sap.com/solutions/business-suite/crm/crmondemand/index.epx, March (2005)
12. Sauermann, L.: The gnowsis – using semantic web technologies to build a semantic desktop. Diploma thesis, Technical University of Vienna (2003)
13. Sauermann, L., Bernardi, A., Dengel, A.: Overview and outlook on the semantic desktop, in: Stefan Decker, Jack Park, D.Q., Sauermann, L., eds.: Proceedings of the 1st Workshop on The Semantic Desktop, ISWC 2005 Conference. (2005) (http://www.gnowsis.org)
14. Siebert, M.: Knowledge creation framework - enabling just-in-time information delivery, in: Althoff, K.D., Dengel, A., Bergmann, R., Nick, M., Roth-Berghofer, T., eds.: Wissensmanagement, DFKI, Kaiserslautern (2005) 642-647
15. Wilson, G.: Proposal automation tools, in: Proposal Management Journal, (2001) 67-73