

Challenges for Information Access in Multi-Disciplinary Product Design and Engineering Settings

Dirk Ahlers*, Mahsa Mehrpoor†, Kjetil Kristensen†, John Krogstie*

* *Department of Computer and Information Science*

† *Department of Engineering Design and Materials*

NTNU – Norwegian University of Science and Technology

Trondheim, Norway

{dirk.ahlers|john.krogstie}@idi.ntnu.no,

mahsa.mehrpoor@ntnu.com, kjetil.kristensen@kristensenconsulting.com

Abstract—In any larger engineering setting, there is a huge number of documents that engineers and others need to use and be aware of in their daily work. To improve the handling of this amount of documents, we propose to view it under the angle of a new domain for professional search, thus incorporating search engine knowledge into the process. We examine the use of Information Retrieval (IR), Recommender Systems (RecSys), and Knowledge Management (KM) methods in the engineering domain of Knowledge-based Engineering (KBE). The KBE goal is to capture and reuse knowledge in product and process engineering with a systematic method. Based on previous work in professional search and enterprise search, we explore a combination of methods and aim to identify key issues in their application to KBE. We list detected challenges, discuss information needs and search tasks, then focus on issues to solve for a successful integration of the IR and KBE domain and give a system overview of our approach to build a search and recommendation tool to improve the daily information-seeking workflow of engineers in knowledge-intense disciplines. Our work contributes to bridging the gap between Information Retrieval and engineering support systems.

Keywords—Industry, Manufacturing, CAx, Manufacturing design and product lifecycle management, Information Retrieval, Information Access, Case Study, Information and Knowledge Management, Enterprise Search.

I. INTRODUCTION

Information retrieval and access is an important research field. It is encountered in Web search engines, but it is also an important issue in professional settings as enterprise search. Here, search is also part of knowledge management. Effective and efficient search remains a key element of engineering work, within and especially across different engineering disciplines in large, multidisciplinary design and engineering projects. Search and interactions are closely related activity categories. Interactions have been defined as “the searching, coordinating, and monitoring required to exchange goods and services” [1]. At an individual level, interactions peak at nearly 80% for interpersonal knowledge workers, subject matter experts, executives, managers, and supervisors. As products and services are typically becoming more complex, the interaction overhead becomes substantial,

and effectively managing this overhead becomes critically important [2]. A large-scale study of industrial companies identifies searching for information (defined as the search barrier: people are unable to easily find what they seek) as one of four main barriers to collaboration [3]. Search is happening in all sorts of specialized fields based on combinations of application domain, documents, document bases, users, etc. This has led to the development of vertical search for, e.g., 3D models, music, images, location information, patents, genomes, within peer-to-peer networks, desktop or enterprise federated search, and many more.

We examine the application of Information Retrieval methods in a Knowledge-based Engineering (KBE) setting. Obviously, all engineering is based on knowledge, but the approach of KBE is to systematically make the knowledge explicit throughout the design process and the lifecycle of products [4], [5], [6], [7]. The goal of KBE is the capture and systematical re-use of product and process engineering knowledge and to enable knowledge sharing between engineers to reduce design time, increase collaboration, and improve workflows. It is a powerful tool for information and knowledge management tailored to the engineering domain. The main use case for KBE is reducing design time by allowing engineers to prototype and analyse more concepts in the early stage of a project, before the more detailed planning phase. It can achieve this by automating repetitive tasks with knowledge-based rules for faster design of components. KBE can lead to reported time-savings in the design process of up to 80%. Thus, the main aspects of a KBE system are utilisation, acquisition, storage, and retrieval of knowledge, secondary aspects concern updates, maintenance, and feedback loops [4]. In this high-level view, it has similarities to existing knowledge management and software development tools with an added deep integration into the workflow. However, there are some important differences and challenges. Within a KBE system, the design knowledge of building blocks, design elements, processes, etc. is captured in a knowledge base, which is used by engineers to design and develop products or parts of them. This typically includes

different types of documents and relations between them. KBE systems contain highly formalized design rules. They integrate into CAx systems, and can be used to directly generate 3D geometries, structural analyses, and other relevant documents and artefacts. KBE rules are captured as libraries, but the underlying knowledge of documentation is usually stored outside of the system. It may be linked from the rules, but usually not in a formalised and easy-to-use way. The knowledge can be rules and regulations, industry standards, best practices, lessons learned, experience, assumptions, design alternatives, design decisions, etc. in different types of documents or databases. The application domain involves specialised tasks, documents, and workflows, all in a professional search environment for domain experts who already have considerable tool support for large parts of the workflow. In our work, we particularly aim to explore the use of Information Access, Information Retrieval, Recommender Systems, Process Ontologies, and Knowledge Management methods to a KBE project on engineering design [8], [7], [9], [10]. We will discuss challenges, and potential solutions in Section III after introducing the scenario in Section II. We give a system overview in Section IV before Section V discusses future work and Section VI concludes the paper.

II. APPLICATION SCENARIO

Our industrial use-case is based on an engineering company which provides oil and gas services, especially designing one-off offshore process plants, commonly known as oil platforms. They apply KBE methods during the process development in order to improve the product design development. Process plant design is engineering of the facilities that use various processes (chemical, mechanical, etc.) to produce or refine products and/or resources (e.g., crude oil, gas, drilling mud). A process plant design project consists of a team of users mainly in engineering roles and a large amount of data that is stored in different data sources and formats as conceptually shown in Fig. 1. Users with various roles and level of expertise such as layout designer, civil or structural engineer, process designer, developers, knowledge engineer, etc. are interacting with each other. Each of them has certain tasks and deals with information in different formats to perform their assigned tasks. Their tasks can be related to the tasks of other disciplines and all parts can cross the system boundaries of the company, having to integrate external factors. Within each concept of users, products, and documents, entities are related to each other and also to other concepts, which in this case is simplified to only large inter-concept links. For example, individual documents can be related to individual products or users in variable ways.

Some engineering building blocks are already modelled as KBE rules in the KBE system and are available in a hierarchical structure inside a library, which offers simple keyword filters. This means that once a certain design

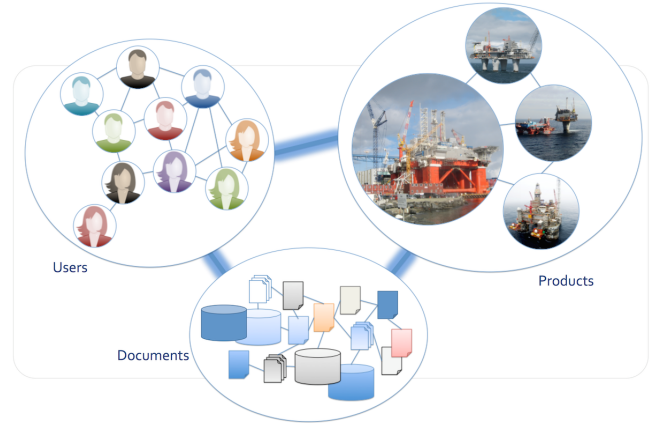


Figure 1. Conceptual view of inter-domain connections (adapted from [9])

has been chosen, it is easy to instantiate, for example, a staircase to fit the particular parameters of the project. Two points should be noted here. First, the decision for a certain approach is done outside of the system, based on available background and foreground knowledge. Second, the design decisions leading to the defined rules are not captured inside the KBE system. Both have to be documented appropriately to be available for re-use at this higher conceptual level. In the Information Access scenario, users need to also find this associated knowledge to adapt their design process. One example task is to retrieve and understand the design decisions leading to a certain set of rules for structural components. This is a sort of reverse engineering of a completed design intended to learn about possible re-use. The user wants to retrieve similar cases from the system or gather alternatives to certain design decisions. The system will enable users to not only search for similar structures, but for cases where similar problems had to be solved, which can give hints towards high-level alternatives. In other tasks, users need to design familiar products but with unfamiliar constraints, such as a different set of governing standards (e.g., different regulations for US waters than for the North Sea). Users can be aiming to learn something from a different discipline, such as a structural designer learning about safety requirements. Similarly, users can be supported in using an automated design process for a particular product. This is especially important for rare or first-time tasks, often encountered by engineers that have less specific work experience, a usual effect of personnel turnover. In these cases, the system can provide support for procedural design processes and workflow decision-making. We can draw an analogy to software engineering. In a library, some parts of the code are self-explanatory, others should be commented, but without a higher-level description of the library, it is difficult for someone else to use it by retracing the decisions for certain approaches and estimating whether their current case is similar enough to allow reuse.

In a knowledge-management system, existing documents will, to a large extent, be already organised in some domain-specific way, including classification, authorship, dates, etc. Typically, some relations between different documents are present in the system itself, as hierarchy relations, topic groups, workflows, or related documents. However, we see two issues. The first is that not all such relations are captured (or are easy to capture automatically) and the second is that relations may not be obvious from the knowledge base, but only emerge during the task of a user who is trying to find documents as they relate to his task and might need to be mined. Also, rarely is the data situation that simple and there rather is a disjoint knowledge management. On the one side, there is the formal KBE system, and on the other side, documents just informally reside in a file system [11], [12], [13]. In this general case, an engineer will work with very heterogeneous data from many sources as well as utilise multiple tools such as CAx, KBE, document storage, desktop search, Web search, etc. [6].

III. CHALLENGES FOR KNOWLEDGE MANAGEMENT AND RETRIEVAL

Developing vertical or enterprise search [14] for knowledge workers is a complex task that should accommodate and enhance the existing workflows [15]. We are working towards support by a recommendation system as part of semantic search [8]. We include context-aware search and recommendation approaches to focus on the current situation [16]. This does not only include contextual views as the contextual “halo” of a document [17] but also to see each document in the context of others in the document store and in the context of work tasks and user interaction. Intermediate steps are understanding of information needs and development of professional search based on domain-specific features. In a first step, we describe and develop the specifics and particularities for search in the KBE setting. Our main identified issues in the KBE domain with respect to IR are:

- Limited information space: domain-specific information, in-house solution, company documents in a closed library.
- Knowledge separation: Documents are not available in one single format and may be distributed over multiple sources.
- Inconsistent metadata: Documents can be precisely fully annotated, but can also range up to completely missing or misleading metadata.
- Insufficient connections between documents: Relations between documents may not be sufficiently annotated. Also KBE rules and directly supporting documents may not be linked.
- Sparsity of both document space and user interactions.
- Searchers are domain experts and engineers, who perform search tasks with high complexity.

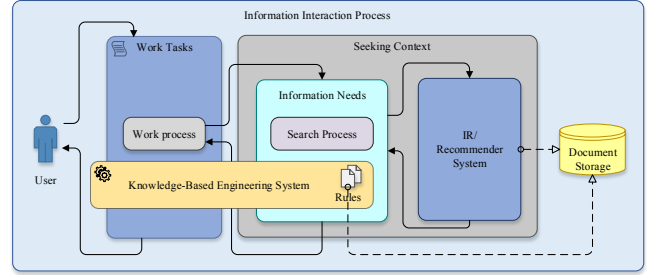


Figure 2. System view from the Information Access perspective with KBE integration (adapted from [16])

A. Information Needs and Search Tasks

Looking a longer way back, we find a description of a knowledge worker’s routine in the Memex: “Wholly new forms of encyclopedias will appear, ready made with a mesh of associative trails running through them, ready to be dropped into the memex and there amplified. [...] The chemist, struggling with the synthesis of an organic compound, has all the chemical literature before him in his laboratory, with trails following the analogies of compounds, and side trails to their physical and chemical behavior.” [18] This example has often been interpreted towards systems like Wikipedia, but it also shows a strong support for professional knowledge interaction. Applied to specific organisational knowledge on product development, the interactions are more complex; an engineer works through available documents, makes new connections between knowledge objects, and strengthens the links that are relevant for the specific tasks. The trail analogy is a sort of personalisation to incorporate implicit or explicit relations between documents. During our initial investigations it turned out that a majority of work is done completely within the KBE system and its library. This means that most of the common tasks are easily done in KBE as is its value proposition. Thus, the remaining tasks are increasingly non-standard and naturally more complex and difficult, which suggests a higher need for search and retrieval. In the following, we work from a user requirements perspective. Related work [16], [19] describes how users’ information needs can be addressed by seeking, actively and passively searching, and using and processing information. For this, the information sources and the users’ information access to these sources has to be understood as well as the processing steps users take to answer their need.

We use the model of Information Behaviour [16] and extend it with the KBE integration as seen in Fig. 2 which expands from the system-centric view of IR and encompasses the underlying information needs and tasks of the user. These can be understood as nested contexts and can be useful in understanding the goals of users in our professional search scenario. In this model, the interaction with a search system is only the innermost part of the overall process that

aims at helping the user to continue on a work task. The work task defines the executed search task, which is in turn defined by the context of the information needs. In detail, there are six main questions to answer (adapted from [20]):

- Information need (what is the user looking for?)
- Work task context (why is she looking for this?)
- Knowledge state (what is the user’s background knowledge on the topic?)
- Ideal information (what would the ideal answer be like?)
- System support (should this be answered inside the IR system, inside the KBE, or by out-of-band means (asking colleagues etc.))?)
- Query (which are the central search terms to express the situation and the need?)

It is clear that to understand and subsequently support this process, we need to explore the context and the tasks [15], [16]. This means to include the roles, tasks, background, and organizational context of the user as well as the documents’ context. Also other work in professional enterprise search makes the case for understanding the search process first [17], [21]. The spectrum of information access can range from external knowledge seeking such as simply knowing where blocks are in the KBE or asking the responsible expert over simple document shares up to intelligent semantic search and recommendation systems. In the following, we focus on navigation and search modes that can improve the management, navigation, and search tasks.

B. Users’ Information Seeking Behaviour and Contextual Features

Users’ information seeking behavior [19] and their context are important factors when designing information access solutions. Each user has a certain role and task and thus some objectives while looking for information. Various variables influence and form these objectives [15], [22], [23]. Fig. 3 describes domain-specific contextual variables from the scenario that influence the behaviour of the user: role, task, level of expertise, information sources, project stage, and intention. Additional variables or variances can be considered based on the professional settings. For example, the current task is significant and since task progression has different stages, information behaviour varies from the first stage of task construction to the last stage of task completion [24]. In addition, users’ intention in each stage of task progression varies from learning about a subject in the early stages, finding facts and procedural information in the middle stages and finally evaluation methods and report preparation in the late stages. These variant contextual variables impact on users’ information seeking and lead to different patterns of information seeking behavior [25]. Users in professional settings tend to often use verbal communication with colleagues, personal information stores, and internal reports in order to find required information

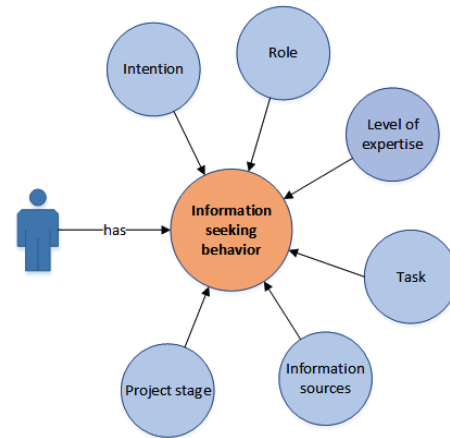


Figure 3. Information seeking behaviour and contextual features (adapted from [25])

which makes it very challenging to track their information seeking behaviour in the main information sources [26]. In our approach we use surveys and interviews and later capture user interaction with the system to try to classify these information seeking behaviours and most importantly, to translate them into better contextual search results.

C. Issues of Personalisation in Professional Search

One way to deal with huge amounts of structured and semi-structured data is personalisation of the search. In fact, this is a standard way to inform the ranking, filtering, or recommendation in multiple systems. It can also be a viable way in professional search, for example by using the query history of users or basing the personalisation on background knowledge or user profiles [20], [22]. The main aspect is interaction of the users to arrive at a more complex model. Then, not only can the system count how many users read a certain document, but it can also derive additional information based on which users access certain documents, for example grouping documents that are more often read by engineers versus documents used by project managers or controllers.

However, there is some issue with this approach. For example, collaborative filtering and relevance feedback may suffer from a permanent cold start problem due to the sparsity of the data. Using the example of patent search, [27] notes that most professional search has insufficient scale for purely statistical machine learning, either by the number of documents, searchers, similar tasks etc. Several improvements are suggested, such as the use of taxonomies, classification, ontology representations, or other semantic annotations with a possibility of semi-automated approaches. This runs consistent with an assumed long tail distribution of relevant documents and searchers. Some overview documents may be regularly accessed by many involved people, others only by a few specialised engineers once in a while.

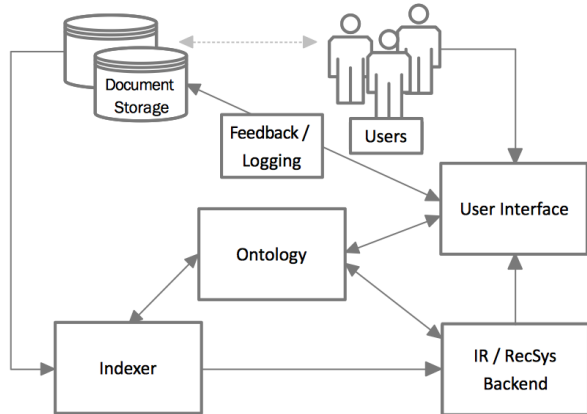


Figure 4. System framework for the search and recommendation tool

Another issue is that personalisation goals can change a lot during a work task or in quickly changing work tasks. In these cases, too high a dependency on user or history data can reduce the quality of the results for a task-oriented goal. We therefore focus less on collaborative filtering and other machine-learning methods relying on large data volumes, but take the opportunities presented by the context and domain-specific information we can gather in the scenario to follow a hybrid approach.

IV. SYSTEM OVERVIEW

For our search and recommendation tool, we propose a system framework that is represented in Fig. 4. Documents from multiple sources are ingested and semantically indexed with the help of a domain ontology [10]. The system then is able to provide tailored results to users. Their context is modelled also by way of the ontology and enriches their queries. Their interaction with the system is captured to model and refine their interests and the relevant documents based on their current role and task. We are adopting available open source advanced search engines and indexers to index and annotate the documents along with employing recommender toolkits to enable the system to understand and predict the relevant documents to a user’s working situation.

We use an ontology as a semantic tool to classify different disciplines, roles and tasks of the engineering project and to help in navigating the documents semantically. The reference ontology is work in progress and is being refined with context features of users and documents as well as document metadata. This is necessary to ensure that context features defined for users can be translated into meaningful document features and queries. There are multiple aspects to a recommendation system. In a first step, the system can learn from the documents themselves. In such a case of content-based recommendation, the system derives the similarity of documents based on their content and metadata. In a next step, context information is taken into account. As discussed

before, this hybrid approach is necessary. Therefore, the adaptation and personalisation of the system is based on semantic analysis and annotation of documents and the context of users, but is refined by the interaction of users with the system, using both explicit and implicit feedback. Because of the special type of documents that may include many binary documents that cannot be analysed with text-based retrieval methods, we also aim to use features of the filesystem [11] to compute similarity features. A challenge is to integrate these varied and partially sparse features into a consistent grouping and ranking of documents.

V. FUTURE WORK

A major aspect of our future work is a refinement of the system. One part is a deeper analysis of the documents sources to learn how to best analyse and present them. The other part is the user side, where we need a running system to then evaluate and improve the system based on feedback from the domain experts. Our work focused both on recommendation and semantic navigation and search support. Evaluations will show which of these modes is best suited in which situations to the users. Additional document and history-based features could in combination be used to, e.g., extract certain workflows or best practices, similar to process mining [28].

While this paper has mostly explored the knowledge use, there is also a strong relation to knowledge capture. From that side, an important aspect is data quality [29] which needs to be explored for two reasons. First, the document base needs to contain the required knowledge; second, the metadata and the relations of documents needed to infer context need to be available and of a sufficient, to be determined, quality [30], and the used methodologies tuned to the availability of features in the documents.

VI. CONCLUSION

We have introduced a challenging field for professional search at the intersection of Information Retrieval, Recommender Systems, and Knowledge-Based Engineering: manufacturing engineering information access. To the best of our knowledge, this is the first work that explicitly makes the connection between these fields and works towards an integration. Previous work in this field has mostly worked to support navigation by enabling context-aware features only in the context of individual documents, less so in using the context of the entire document collection as a knowledge base, or integrating the context of users. As the field of patent search has demonstrated, there is a strong and ongoing need for specialised search solutions in many professional areas. The paper has shown that the manufacturing engineering topic contains many worthy challenges from a search and recommendation perspective. We have examined the use of content-based personalisation [8], user profiles and background knowledge connected to a reference ontology [10],

semantic content-based similarity and filesystem features, and developed an overall robust framework for an integrated Information Access system for the KBE domain.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n°284613 for the LinkedDesign project¹.

REFERENCES

- [1] B. C. Johnson, J. M. Manyika, and L. A. Yee, "The next revolution in interactions," *McKinsey Quarterly*, no. 4, 2005.
- [2] K. Kristensen and B. Kijl, "Collaborative performance: Addressing the ROI of collaboration," *International Journal of e-Collaboration*, vol. 6, no. 1, 2011.
- [3] M. T. Hansen and N. Nohria, "How to build collaborative advantage," *MIT Sloan Management Review*, vol. 46, no. 1, pp. 22–30, 2004.
- [4] G. L. Rocca, "Knowledge based engineering: Between AI and CAD," *Advanced Engineering Informatics*, vol. 26, no. 2, pp. 159 – 179, 2012.
- [5] W. J. C. Verhagen, P. Bermell-Garcia, R. E. C. van Dijk, and R. Curran, "A critical review of knowledge-based engineering: An identification of research challenges," *Adv. Eng. Inform.*, vol. 26, no. 1, pp. 5–15, Jan. 2012.
- [6] M. Bertoni and K. Chirumalla, "Engineering 2.0: leveraging a bottom-up and lightweight knowledge sharing approach in cross-functional product development teams," in *i-KNOW 2010*, 2010.
- [7] C. Kalavrytinis and O. I. Sievertsen, "A knowledge-based engineering approach for offshore process plant design," in *ICE 2014*. IEEE, 2014.
- [8] M. Mehrpoor, A. Gjørde, and O. I. Sivertsen, "Intelligent services: A semantic recommender system for knowledge representation in industry," in *ICE 2014*. IEEE, 2014.
- [9] D. Ahlers and M. Mehrpoor, "Semantic Social Recommendations in Knowledge-Based Engineering," in *SP 2014: Workshop on Social Personalisation at Hypertext 2014*, 2014.
- [10] M. Mehrpoor, J. A. Gulla, D. Ahlers, K. Kristensen, S. Ghodrati, and O. I. Sivertsen, "Using Process Ontologies to Contextualize Recommender Systems in Engineering Projects for Knowledge Access Improvement," in *ECKM2015*, 2015.
- [11] D. Ahlers and M. Mehrpoor, "Everything is Filed under 'File' – Conceptual Challenges in Applying Semantic Search to Network Shares for Collaborative Work," in *Hypertext '15*. ACM, 2015.
- [12] C. A. N. Soules and G. R. Ganger, "Connections: Using context to enhance file search," in *SOSP '05*. ACM, 2005.
- [13] S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins, "Stuff I've Seen: A System for Personal Information Retrieval and Re-use," in *SIGIR '03*. ACM, 2003.
- [14] X. Liu, H. Fang, C.-L. Yao, and M. Wang, "Finding relevant information of certain types from enterprise data," in *CIKM'11*. ACM, 2011.
- [15] P. Petcu and R. Dragusin, "Considerations for the Development of Task-Based Search Engines," in *IRPS @ ECIR2013*, 2013.
- [16] P. Ingwersen and K. Järvelin, *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer, 2005.
- [17] T. Russell-Rose, J. Lamantia, and M. Burrell, "A taxonomy of enterprise search and discovery," in *EuroHCIR*, 2011.
- [18] V. Bush, "As We May Think," *The Atlantic Monthly*, vol. 176, no. 1, pp. 101–108, 1945.
- [19] T. D. Wilson, "Human information behavior," *Informing Science*, vol. 3, no. 2, pp. 49–56, 2000.
- [20] S. Verberne, M. Sappelli, D. R. Sørensen, and W. Kraaij, "Personalization in Professional Academic Search," in *IRPS @ ECIR2013*, 2013.
- [21] P. Hansen, "Task-based Information Seeking and Retrieval in the Patent Domain: Processes and Relationships," Ph.D. dissertation, University of Tampere, 2011.
- [22] P. Brusilovsky and D. W. Cooper, "Domain, task, and user models for an adaptive hypermedia performance support system," in *IUI '02*. ACM, 2002.
- [23] T. Gu, X. H. Wang, H. K. Pung, and D. Q. Zhang, "An ontology-based context model in intelligent environments," in *CNDS'04*, 2004.
- [24] P. Vakkari, "Task-based information searching," *ARIST*, vol. 37, no. 1, 2003.
- [25] L. Freund, "Exploiting task-document relations in support of information retrieval in the workplace," Ph.D. dissertation, University of Toronto, 2008.
- [26] M. Hertzum and A. M. Pejtersen, "The information-seeking practices of engineers: searching for documents as well as for people," *Information Processing & Management*, vol. 36, no. 5, pp. 761 – 778, 2000.
- [27] J. Tait, "Issues and Non-Issues in Professional Search," in *IRPS @ ECIR2013*, 2013.
- [28] W. Van Der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011.
- [29] J. Krogstie, "Evaluating Data Quality for Integration of Data Sources," in *The Practice of Enterprise Modeling*, ser. LNBP. Springer, 2013, vol. 165, pp. 39–53.
- [30] D. Ahlers and J. Krogstie, "Document and corpus quality challenges for knowledge-management in engineering enterprises," *JDIQ*, 2015, to appear.

¹<http://www.linkeddesign.eu/>