

Towards Knowledge Discovery through Context Explication

Stefan Kremer
University of St. Gallen
Stefan.Kremer@unisg.ch

Stefan Smolnik
University of Paderborn
Stefan.Smolnik@notes.upb.de

Lutz Kolbe
University of St. Gallen
Lutz.Kolbe@unisg.ch

Abstract

More and more knowledge does play a decisive role as a factor of production besides the classic factors work, raw materials, and capital. Companies get a competitive advantage from a lead of knowledge and the capability to transform superior knowledge into market-driven business processes. The consistent orientation towards customers and their processes requires the customization of intra-corporate processes and systems. Customer process oriented portals support the collaboration of customers, employees, and suppliers. They integrate companies' systems and provide transparent access to information objects stored in these systems. One key problem is to find relevant information objects in continuously growing and distributed systems. Necessary conditions for the core processes of knowledge identification and knowledge use are mechanisms of navigation and linking as well as functionalities for extensive searches and investigations. A decisive challenge is to make knowledge available at the right time and the right place. Therefore, the implicit context of information has to be explicated to fill 'knowledge gaps'. In this paper, we will introduce a continuum of context explication, which comprises different relations between information objects and its contexts by means of their degree and ease of context explication. Furthermore, we will evaluate different approaches for knowledge discovery in customer process oriented portals, provide patterns when to apply which approach, and present two small cases for knowledge discovery in such portals.

1. Introduction and overview

1.1. Challenge

Due to the extensive use of information technology, a lot of organizations have gained access to vast internal and external information repositories. Intranets, content management systems, and enterprise portals have become popular and give employees the possibility to discover knowledge, enshrined in information objects, e.g. electronic documents [1], [2]. Yet nonetheless, efficiency

is not on an appropriate level when dealing with information and trying to find the right content. "Information management must begin by thinking about how people use information..." [3], otherwise a practical usage is not possible.

Even though organizations nowadays have access to an astonishing number of different information repositories, we observe major deficits in the process of knowledge discovery such as:

- *Lack of information:* Finding information objects about certain topics might sometimes be frustrating, because although users might know that some information objects already exist, they cannot be discovered.
- *Overload of information:* Knowledge discovery might also be time consuming if too many information objects are found with no or minor relevance.

Looking at these deficits, the control of semantic, i.e. the meaning of terms, availability of explicated context, and orderly classification of information objects utilized in business environments is one possible key to success in order to minimize the deficits [4], [5]. Different technological approaches – based on different degrees of context explication – have already been proposed to address the problem of lack versus overload of information, e.g. search and classification engines. Nevertheless, selection criteria to support organizations to choose the appropriate solution are hardly available. Thus, we will give a comprehensive overview of several distinct approaches, their underlying principle, benefits and constraints, and provide selection criteria for the appropriate degree of context explication, fulfilling specific organization's needs.

1.2. Objective and research approach

The overall objective of this paper is to propose a continuum of context explication, which comprises different relations between information objects and their contexts by means of their degree and the ease of context explication. We will show, that the continuum allows conclusions for organizations to select the appropriate

approach for implementing knowledge discovery in portals.

Literature study results in a comprehensive overview of related work particularly in knowledge management literature and shows the gap between research that deals with the context of an information object and its explication. Desk and action research, e.g. prototypical implementations of our conceptual approaches, lead us to logically deduced conceptualizations. We use case research to deduce and validate our conceptualizations. Case research used for this effort is particularly suitable for problems where research and theory are in early stages of formulation [6]. Thus, research and descriptive processes are influenced by results from action research [7], [8] with several corporate partners, namely results from workshops and projects with research partners conducted in the last two years. The findings are currently tested and expanded with further partners.

1.3. Structure of the article

Section 2 deals with related work in the area of knowledge discovery and portals. It also delimits the main terms, necessary for understanding the research field.

Challenges for knowledge discovery are illustrated in section 3 and three major approaches for knowledge discovery within portals are introduced by providing a chronological outline through different development phases. These three approaches – full-text searching, attribute-based searching and, topic maps – are each explained by their characteristics, benefits, and constraints in regard to context explication.

Knowing the capabilities and constraints of these distinct approaches, we then propose a continuum of context explication in section 4, providing selection criteria and advice for choosing the appropriate solution, based on given pre-requisites.

Section 5 gives two examples of successfully applying the continuum to a daily work scenario. Based on different pre-requisites, we have chosen and implemented different solutions for knowledge discovery at our institutes. In the completing section 6, we draw a conclusion and propose directions for further research.

2. Related work and delimitation of terms

In this section, we will introduce the theoretical background and delimit the most relevant terms of our paper. We will point out related work and explain the differences to our approach.

2.1. Knowledge and context

Within literature, many definitions of knowledge can be found (cf. for example [9], [10], [11], [12], [13], [14], [15]; some of these references also provide detailed

discussions of the differentiation of the terms data, information, and knowledge and furthermore, they discuss different types of knowledge and their classification). From our point of view, knowledge comprises both information and person-specific aspects like experiences, values, and insights. An important characteristic of knowledge and difference to information is the strong affinity of knowledge to activities (cf. [9]). Individuals act and react due to their experiences and intrinsic attitudes. Knowledge is much more than transformed information and therefore, cannot be represented in the form of information objects or data. Polanyi developed a concept for implicit knowledge, which he described as follows [14]: “We can know more than we can tell.” We follow Polanyi’s basic concept of knowledge that the implicit and the explicit dimension of knowledge are complementary parts [14]. All knowledge contains both dimensions. Therefore, it is not possible to speak of pure explicit or pure implicit knowledge resp. of the transition of one part into the other.

Many of the above referenced definitions of knowledge have context as an important facet in common. One form of transition from information to knowledge is contextualization. Dey and Abowd [16] define context as “any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.” Similarly, Sowa [17] describes context in its nonlinguistic meaning as “situation, environment, domain, setting, background, or milieu that includes some entity, subject, or topic of interest.”

A differentiation of context types is described by Klemke in a level-based approach [18]. On the first level, he identifies the following context dimensions: organizational, domain/content based, personal, and physical. These dimensions are specified in more detail on the second level, e.g. the organizational dimension is subdivided into a process and a structure component. In spite of the common assumption that context only consists of implicit information the above definitions allow context to be either explicitly or implicitly. In our paper, we expose that the explication of the implicit context of information objects supports the creation of new knowledge and we describe different approaches of how to achieve this. Klemke suggests a holistic understanding of context with several dimensions (cf. above) and an integrated architecture to trace and maintain context models [18]. In addition, within literature, several different characteristics are considered as context and different approaches are used to model these contexts, e.g. workflow process context is modeled by workflow management systems [19] or organizational structures are modeled by enterprise ontologies. In contrast, we focus directly on information objects and

their contexts and provide approaches to discover, explicate, and use these contexts in different situations.

2.2. Knowledge discovery

Within literature, many knowledge management activities, methods, or modules have been discussed. Lai and Chu suggest an integrated knowledge management framework, which comprises the activities initiation, generation, modeling, repository, distribution and transfer, use, and retrospect [11]. Davenport and Prusak differentiate between determine requirement, capture, distribute, and use [9]. A pragmatic approach for the organization wide management of knowledge is presented by Probst, Raub, and Romhardt [20]. This approach comprises six core processes and two pragmatic modules: identification, acquisition, development, distribution, use, and preservation of knowledge as well as objectives and performance measurement of knowledge. In addition, more or less similar classifications of knowledge management activities are presented, for example, in Nonaka and Takeuchi [10]; Arthur Andersen [21], and Alavi [22]. All these approaches have a method for the identification or use of knowledge in common, either implicitly or explicitly. Normally, unused knowledge, which has to be uncovered by appropriate methods, is available within organizations, and then, it can be utilized for the respective organization. Knowledge discovery methods foster an increase of transparency of knowledge in organizations and support users finding relevant information objects. They are a necessary precondition for the core processes of knowledge identification and knowledge use [20] and therefore, they improve organizational use of existing individual and common knowledge, and contribute to the process of knowledge generation, i.e. the development or collection of new knowledge [23].

2.3. Portals

Since about 1998, portals are discussed as an integration concept for user access to personalized information and applications [24]. Even if there are many different descriptions of portals (cf. [25], [26], [27], [28], [29]) in our focus portals are defined as web-based, personalized, and integrated access systems to internal and external applications and information repositories. Portals support knowledge-oriented processes by providing users with a graphical frontend-integration of backend systems. This is achieved by comprehensive services like integration, personalization, and administration. Knowledge discovery methods in portals are supported mainly by navigation and search mechanisms [30], [31]. Especially, the role of search mechanisms to increase transparency of knowledge in organizations and support users finding relevant

information objects is significant, as the following section shows.

3. Knowledge discovery within portals

As stated above, search and retrieval plays a vital role in the concept of portals. There are special challenges for knowledge discovery in portals (cf. [32], [33]):

- *Heterogeneous structures and formats:* Information objects are stored in multiple more or less structured formats, with different classifications and consisting of different languages. But portal users need a consistently structured view upon all available information objects.
- *Distributed and redundant data:* Organizations have information objects residing partly redundantly in a variety of sources, e.g. e-mail systems, content management systems, and file server systems. Knowledge discovery has to offer various mechanisms to connect these repositories to the portal to provide a consolidated view to users.
- *Protected content:* The role of each individual portal user dictates which information objects that individual is able to access. With knowledge discovery, navigation entries and search results have to be filtered to display only the information objects accessible to the user, i.e. a secure access has to be provided.

A fundamental service to support knowledge discovery methods in portals must be able to address these challenges [34]. The following section illustrates three major approaches for knowledge discovery in portals by providing a chronological outline through different development phases.

3.1. Full-text searching

Classical full-text searching is an established retrieval approach since the early 90's [35]. A search engine is an information technology component of a portal that acts as a central instance between the user's information need and available information objects, stored in one or multiple repositories. Users transform their information need into a search query and enter it in a search field provided within a portal. In order to respond to the user queries, search engines index each information object, representing it by a set of weighted words. The search engine compares the entered terms with the previously indexed information objects and sends a result list back to users.

The benefits for users are:

- *Speed:* Searching with one single query against the content of multiple repositories is faster than

- searching with separated queries in each application itself.
- *Ease of use*: Nowadays, full-text searching is well known and most users have certain experiences [36].
- *Pre-structuring unnecessary*: Due to the fact that relevant terms are indexed automatically no human driven intervention is necessary.

As the verbalization of an information need into a single query has a strong impact on the quality of search results, the main constraints of full-text searching result from semantic issues:

- *'Wrong' or too many results*: Receiving search results pointing to information objects with no or low relevance is time consuming for users [37]. Here, the major challenge for users is to anticipate the correct terminology, i.e. to match the author's terminology.
- *Spelling important*: A user's query should be orthographically correct.

As shown, full-text searching relies on the information object's content, because the information object's context is exclusively contained in the information object itself. Explication has been done neither by the authors at the creation-time of the information object nor by a system afterwards.

3.2. Attribute-based searching

To overcome the above illustrated constraints of pure full-text searching, attribute-based searching came up in the mid 90's [37]. This approach is based on an explication model of context by storing the context of an information object in metadata, i.e. in data about the data [38]. The metadata are stored with the information object itself and can be viewed and retrieved by users and applications. Common attributes of metadata associated with information objects include the author, the date of publication, the source of publication, etc. With attribute-based searching, knowledge discovery in portals allows structured queries on context explicated in information objects' metadata ([39]). Nowadays, several metadata standards exist, e.g. the Dublin Core Metadata Element Set, which proposes 15 fields resp. attributes to describe a document ([40], [41]). The major benefits of attribute-based searching are:

- *Reduced result set*: Compared to full-text searching, it is less time consuming for users to retrieve relevant information objects.
- *Controlled vocabulary*: Users can choose standard terms out of drop-down lists.
- *Personalization capabilities*: Search queries can be enriched automatically by personalization information (user attributes, e.g. roles, language, and organizational unit).

But there are also certain constraints:

- *Maintenance of controlled vocabulary*: Even if using this approach is less time consuming for users to find relevant information objects, certain resources are needed at the creation-time of an information object. Human intervention is necessary to classify information objects with appropriate context attributes.
- *Metadata are stored with the information object itself*: As terms may change over time, re-classification seems to be necessary. Alternatively, re-classification can be avoided by mapping old to new terms on a separate layer, e.g. with a customized thesaurus.

As mentioned, attribute-based searching is based on context explication. The information objects contain content and explicated context, both maintained by the information object's author at creation-time or during maintenance changes.

3.3. Topic maps

The initial ideas behind topic maps – which date back to the early 90's – arose from the need to model intelligent electronic indexes (of books), tables of contents, glossaries, thesauri, or cross references in order to be able to merge them automatically. But during several years of discussions and evolutionary development cycles, the topic map model has developed into something much more powerful that is no longer restricted to simply modeling indexes. The 1999 adopted ISO standard ISO/IEC 13250 Topic Maps defines a model and architecture for the semantic structuring of link networks. Topic maps establish an associative network between subjects, which represent information objects, and provide navigation paradigms to enable its search. By applying topic maps to large sets of heterogeneous information repositories, reusable structured semantic link networks are created above those resources [42]. The key concepts of topic maps are topics, which represent real world subjects, occurrences of topics, and relationships between topics (topic associations). In addition, the topic map standard provides the extended concepts of scope, public subject, and facets. For a comprehensive introduction and reference refer to Rath and Pepper [42] and ISO/IEC 13250 [43].

Topic associations describe the relationships between topics. They are completely independent of the information object itself and therefore, they represent the essential value-add of the topic map. The addition of topic associations to the concept of topics enables topic maps to model networks of information. Topic maps organize information repositories within a new knowledge space, by relating them to topics, and associating those topics, in a structured way. Furthermore, they enable heterogeneous sets of information repositories to be used in an integrated way, by interrelating them using a unifying conceptual

framework. Another characteristic of topic maps is that they are well suited to represent ontologies. Thus, they facilitate a way of describing a shared common understanding, e.g. about the kinds of objects and relationships, which are being talked about [44]. The link mechanism between topics and occurrences provides a means for ‘bridging the gap’ between knowledge representation and the field of information management [45]. Eventually, the human brain always remembers memorized things in a specific context [46]. The basic way of thinking is the association. Topic maps support this way of thinking by pointing to related themes while a user looks at a specific theme.

To summarize, topic maps particularly provide the following benefits:

- *Creation of knowledge structures:* Applying topic maps to information repositories generates knowledge structures. Topic maps form structured semantic link networks above great sets of information repositories.
- *Creation of meta layers:* Transparent access to information objects by searching and navigating in knowledge structures, i.e. in a meta layer above the information objects. Modifications of the meta context do not affect information objects or their descriptors. Searching in topic maps can be compared to searching in knowledge structures.
- *Discovery of new context:* Added value by creation of new knowledge through the discovery of new contexts.
- *Support of human thinking:* Topic associations support the basic way of thinking by interrelating themes.

Some basic constraints are:

- *Effort of topic map creation and maintenance:* High human effort is needed to define, create, and maintain topic maps. Persons who deal with managing topic maps need expertise both in the concepts and paradigms of topic maps and in the specific domain to which the topic map applies.
- *New search paradigm:* Users have to learn to use the search concepts of topic maps. They are familiar with known concepts of full-text or attribute-based searching and their flat result sets. Using topic maps, they have to adopt the associative way of thinking.

Topic maps provide strong concepts and paradigms to discover and explicate contexts of information objects. The explicated context is not part of the information object and can even be used without the information object. The process of context explication must be supported by special skilled persons. Groups and roles of such ‘knowledge workers’ are introduced in detail in Smolnik and Nastansky [47]. These knowledge workers need expertise in managing topic maps as well as in the specific domain to which the topic map applies. Neither

authors nor users need to provide metadata or descriptions at any time.

4. The continuum of context explication

As pointed out in our motivation, context has been recognized as being an important aspect to consider when looking at the meaning of information. In section 3, we have presented three approaches how to find information objects and how to recognize, represent, and use contextual information within portals. Even though these approaches have the support of users in finding relevant information objects in common, they focus on contextual information in different ways and varying intensity. As one main result of our research, we introduce the continuum of context explication, which focuses on data, information objects, and knowledge as basic subjects in portals, their different embodied degree of context explication, and the presented approaches to find and use information objects and contextual information (v. Table 1). We define five approaches depending on a different degree of context and its ease of explication: three approaches relate to information objects and the chronological development of search methods, the other two approaches regarding data and knowledge form a logical extension in the transition of data via information towards knowledge. Furthermore, we provide selection criteria and advice for choosing the appropriate solution, based on given pre-requisites.

Data approach

Data are meaningless symbols without content and context. Therefore, there is no context to explicate. Depending on the quantity of data and the considered domain several methods for the transition of data to information objects or even to domain-specific knowledge can be mentioned. For example, in the research domain of knowledge discovery in databases and data mining, the identification of patterns in large structured data sets results in the non-trivial extraction of implicit, previously unknown, and potentially useful knowledge [48]. Processed and conceptualized data like documents created by authors are defined as information.



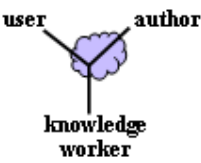
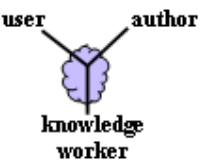

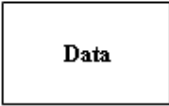

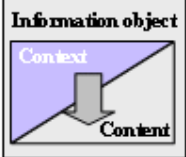
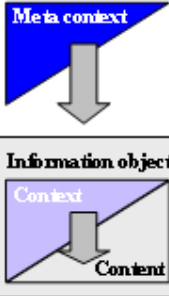
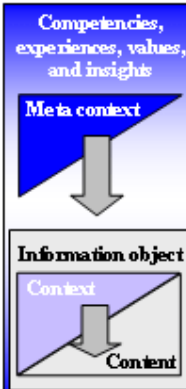
The data approach applies to a scenario with the following characteristics:

- No or little interaction with users, authors, or knowledge workers
- Large structured data sets
- Possibly automatic data generation or collection

Information approach

A simple information object contains some kind of content, e.g. a text, an audio annotation, or a spreadsheet. Even though the information object provides no explicit context like descriptors or other contextual information, it inherently contains context. The context is interwoven with the content and difficult to conceptualize. Methods to find requested information objects have to rely on the

Table 1. Continuum of context explication

Data approach	Information approach	Descriptor approach	Meta context approach	Knowledge approach
				
Large structured data sets	Unstructured information objects	Semi-structured information objects Information objects with descriptors	Large sets of heterogeneous information repositories	Information objects in person-specific contexts
				
Pattern identification	Full-text searching	Attribute-based searching Resource Description Framework	Topic maps	Action: Communication Cognition Construction
Ease of context explication				
Context				

content and cannot access contextual information. An exemplary method is full-text searching as described in section 3.1. Common full-text search engines use indexed contents of information objects to respond to a query and do not access contextual information. Therefore, the effort to the explication of context can be neglected as neither authors nor users provide or use explicit contextual information.

The information approach applies to a scenario with the following characteristics:

- Many users, who are inexperienced with enhanced searching approaches or who are unwilling to use them
- Authors are not trained in describing their information objects

- Numerous unstructured information objects

Descriptor approach

Information objects are often enriched with metadata, i.e. they consist of content and explicit contextual information. Examples are Microsoft Word documents, Adobe PDF documents, or semi-structured documents in a groupware-based office environment. Compared to the information approach, information objects do not only contain implicit context information but also explicit context information. As presented in section 3.2, a standard to formulate contextual information is the Dublin Core Metadata Element Set, which proposes specific attribute classes of an information object. Another concept of structuring and providing metadata is the Resource Description Framework (RDF). RDF is

resource-oriented; its main objective is the description of resources and their relation to other resources. The description mostly inheres in the resource. In contrast to the information approach, some effort is necessary to enrich an information object with explicit contextual information. At creation time, authors have to provide this information. In addition, software systems try to maintain some of the contextual information. The benefit of attribute-based searching as a retrieval method of information objects is dependent on the quality of provided explicit contextual information (v. section 3.2). If the metadata are wrong, misleading, or incomplete attribute-based searching will provide insufficient result sets. Otherwise, attribute-based searching provides more accurate results, which fit to some degree to the context of users.

The descriptor approach applies to a scenario with the following characteristics:

- Authors are trained and skilled in describing their information objects
- Information objects contain descriptors
- Numerous semi-structured information objects

Meta context approach

Extending the descriptor approach, information objects cannot only be described by metadata that inhere in the information object but also by subjects, concepts, or themes that form contextual information in a meta layer above the information objects and are not necessarily explicitly stored within the information object. Topic maps provide strong paradigms to discover, maintain, navigate, and visualize this meta context and therefore, they explicate the context of an information object (v. section 3.3). Semantic relations between information objects are expressed by associating topics. This semantic network links the explicated contextual information of different information objects and thus, discovers new contexts. By discovering these new contexts, users are supported in creating new knowledge. They associate known information objects in a new way with other information objects. To achieve these benefits of explicated and new contexts substantial effort has to be invested to define, create, and maintain a topic map. This effort is disproportional higher than the definition of metadata in the descriptor approach. In the latter case, authors or software systems explicitly declare contextual information. Authors know what they publish and can easily describe their information objects. In the meta context approach, knowledge workers are needed to provide and maintain a topic map. The benefit for users depends on the quality of the knowledge worker's work. If the space of meta contexts covers the whole domain of interest and contains rich and various topic associations users will easily be able to explore the search domain and will be enabled to discover new contexts and therefore, to leverage and enhance their knowledge.

The meta context approach applies to a scenario with the following characteristics:

- Knowledge workers, who are familiar with topic map concepts and the domain of interest
- Manageable domains of interest
- Possibly existing taxonomies for the domains of interest
- Users are experienced in searching and navigating topic maps
- Large sets of possibly heterogeneous information repositories

Knowledge approach

So far, we have only focused on the human factor in very specific perspectives like authors defining metadata of information objects or knowledge workers developing topic maps. The human factor plays a decisive role in the transition of information to knowledge. We differentiate two dimensions of the human factor: Firstly, competencies, experiences, values, and insights, form a rich, person-specific context. This context is a characteristic of the implicit dimension of knowledge and hardly to explicate [14] (cf. section 2.1). Within this context, a highly individual and subjective meaning is assigned to an information object. Secondly, active involvement of users is a necessary precondition to transfer information into knowledge. This active involvement comprises actions like communication, construction, or more intrinsic cognition. If users experience an 'I see!' event as a result of some action knowledge is created.

Characteristics of the knowledge approach are:

- Competencies, experiences, values, and insights
- Information objects in person-specific contexts
- Creation of knowledge by human actions, e.g. cognition of information objects

5. Small cases and lessons learned

In this section, we present two small cases, derived from prototypical implementations at our institutes. They illustrate the benefits and constraints of the approaches to discover information objects presented in section 3. The first case meets the selection criteria of both the information and the descriptor approach of the continuum of context explication introduced in section 4. The second case is an example of a solution addressed by the meta context approach. Motivated by the theory of participatory action research [7], our selection is based on significance and available information to achieve an appropriate reliability and validity [49].

5.1. A combination of full-text and attribute-based searching at the IWI

Within the Institute of Information Management (IWI) at the University of St. Gallen, we have four chairs, each

with two or more competence centers. Competence centers are led by project managers who are responsible for achieving the objectives of the competence center. Each competence center produces a lot of information objects, e.g. lecture materials, presentations, and publications. These materials are stored in different systems, e.g. file server, groupware-based office environments, or web content management systems.

From a terminological point of view, all information objects have one thing in common to contextualize the content: they all deal with specific topics, e.g. knowledge management, enterprise application integration, business networking. As most information objects are semi-structured and the responsibilities for maintaining meta-data are manageable, the continuum of context explication (cf. section 4) led us to an hybrid approach: to reduce both the maintenance effort for a controlled vocabulary of a pure attribute-based approach (cf. section 3.2) and the risk of a misspelled full-text search (cf. section 3.1), we have chosen a combination of both. Within an internal project, we declared 'topic' to be the most important descriptor to contextualize the content of an information object for storage and further retrieval. Relevant topics have been locally collected from all competence centers and stored in a single parameter database. Regarding a specific topic, e.g. portal, the following contextual information is embodied in our definition framework: Item (preferred term for topic), Assigned to (responsible competence center), Status (draft, active, or frozen), Synonyms (similar terms or different languages) and Description (description of the term). A document history supports the traceability of modifications [50]. These topics are used to classify information objects, e.g. within our team databases or literature- and publication applications.

Currently, there are about 350 topics overall, owned by 30 competence centers. About 11,000 documents are classified with these topics for further retrieval through intranet- and extranet-portals. Using this approach for about one year, we are able to derive the following success factors:

- *Simplicity*: A 'lean' context explication framework with only few dimensions reduces workload for information object classification, which increases user acceptance.
- *Mixture of centralization and decentralization*: A few simple centralized 'rules' for topic definitions are helpful, e.g. naming conventions. Detailed decisions of certain terms are made instead decentralized by responsible team members, reducing coordination overhead.

5.2. Topic maps at the GCC

As pointed out in [47], groupware-based office systems provide an excellent environment for organizational knowledge management. Within the

Groupware Competence Center (GCC) of the University of Paderborn, the GCC K-Pool (GCC Knowledge Pool) is used in almost every facet of operative work. It is a groupware-based repository for several kinds of information objects. In particular, books, conference information, links, media objects, contributions and articles, and software information are maintained in the GCC K-Pool. The different information objects are enriched with numerous descriptors: categories are used to set information objects in different contexts, keywords describe the information objects in detail, and publishing information gives further explanations. Even though there are a lot of semantic relations between the information in these databases, it is hardly possible to navigate between them or to identify knowledge structures. The capabilities for accessing information objects are restricted to basic full-text searching and navigating through context sensitive views and categories. Full-text indexes are not sufficient for searching information, and structures, like document types or taxonomies, are sometimes too constraining to qualify or to categorize information objects [51]. Furthermore, the usage and the scope of these techniques are limited to a single database.

Using the generic approach of applying topic maps to groupware-based organizational memories as described in Smolnik and Nastansky [47], the search and navigation concepts presented in section 3.3 can be applied to the GCC K-Pool. In detail, we define a topic map template, which comprises topic types and association types and describes the skeletal structure of the topic map. Typical topic types are "author", "title", "publisher", or "place"; typical association types are "author writes title", "publisher publishes title", or "publisher is located in place". Software agents automatically create and maintain the topic map, which is applied to the GCC K-Pool.

The GCC K-Pool topic map facilitates the creation of knowledge structures and meta layers, the discovery of new contexts, and supports the users' cognitive capabilities (cf. sections 2.1, 2.2, and 3.3). Users can search and navigate the GCC K-Pool topic map in several ways. A text-oriented web browser interface provides an easy and familiar access. Additionally, users can explore the GCC K-Pool topic map by using two visualization tools: The GCC K-Viewer, a two-dimensional approach with auto-layout capabilities for restructuring the topic map visualization, and the Sky Surfer, a three-dimensional approach with extensive navigation and search functionalities. These different topic map visualization approaches are described in detail in Smolnik, Nastansky, and Knieps [52].

The team of the GCC consists of highly skilled researchers, who are familiar with the topic map concepts and who know all subjects of daily business. Users and authors are supported by a lightly distinctive taxonomy. These preconditions meet the criteria that are required for the meta context approach of the continuum of context

explication (cf. section 4) and form an excellent environment for a deployment of topic maps. We use this approach since several months and are able to observe the following main results:

- *Understanding of work contexts:* With interrelating themes resp. information objects users understand more easily how subjects of work are related. They are able to explore the domain of interest in an intuitive way.
- *Low maintenance:* Once configured and created, the topic map is updated automatically. Software agents insert new information objects, topics, and topic associations and delete outdated ones.

6. Conclusions and future areas of research

As shown, the context of information objects and context explication play an important role in the area of knowledge discovery and portals. As there are several possible approaches, the real task for knowledge discovery begins with the selection of the appropriate solution of context explication. Therefore, we have illustrated three approaches – full-text searching, attribute-based searching, and topic maps – each explained by their characteristics, benefits, and constraints with a focus on context explication and aligned them into our continuum of context explication. Successful application of the suggested selection criteria has been shown by the two reference implementations, realized at our institutes.

To enrich our proposed model of context explication, we see at least four areas of future research:

- Firstly, we will further evaluate the distinctness of situations for applying the continuum's elements by adopting GCC's topic map framework on IWI's content. Even if the pre-conditions are different, this might lead to some insights into the degree of exchangeability of our different approaches.
- Secondly, we have to prove, if certain patterns for transition can be found, e.g. from topic maps to full-text searching or full-text searching to attribute-based searching.
- Thirdly, we would like to extend our continuum with advices concerning the implicit dimension of knowledge by including the explication of skills and skill management.
- Fourthly, we will validate and expand our findings with further external partners.

Thus, we envision that knowledge discovery through context explication provides a comprehensive framework to support knowledge management processes productively.

References

- [1] Latham, L., *Web Content Management and Portals: Who's Doing What?*, Research Note SPA-13-9670, GartnerGroup, Stamford, 2001.
- [2] Web/CKS, *Web Portals: Present & Future Positioning in the European Marketplace*, 1999.
- [3] Davenport, T.H., Harris, J.G., Kohli, A.K., *How Do They Know Their Customers So Well?*, in: *MIT Sloan Management Review*, 42 (2001) 2, p. 63-73.
- [4] Dale, A., *Designing Taxonomies at Unilever*, in: *Knowledge Management Review*, Vol. 3 (2001) Issue 6, p. 30-34.
- [5] Felber, H., Budin, G., *Terminologie in Theorie und Praxis*, Gunter Narr Verlag, Tübingen, 1989.
- [6] Benbasat, I., Goldstein, D.K., Mead, M., *The Case Research Strategy in Studies of Information Systems*, in: *MIS Quarterly*, 11 (1987) 3, p. 369-386.
- [7] Whyte, W.F., Greenwood, D.J., Lazes, P., *Participatory Action Research: Through Practice to Science in Social Research*, in: Whyte, W.F., *Participatory Action Research*, Sage Publications, Newbury Park, CA, 1991, p. 19-55.
- [8] Gummesson, E., *Qualitative methods in management research*, Sage Publications, London, 2000.
- [9] Davenport, T.H., Prusak, L., *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press, Boston, 1998.
- [10] Nonaka, I., Takeuchi, H., *The Knowledge-Creating Company - How Japanese Companies Create the Dynamics of Innovation*, Oxford University Press, New York, New York, 1995.
- [11] Lai, H., Chu, T.-h., *Knowledge Management: A Review of Theoretical Frameworks and Industrial Cases*, in: *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, 2000.
- [12] Murray, P.C., *Information, knowledge, and document management technology*, in: *Knowledge Management Briefs*, Volume 1, Issue 2, 1996.
- [13] Sveiby, K.E., *The New Organizational Wealth: Managing & Measuring Knowledge-Based Assets*, Berrett-Koehler, San Francisco, 1997.
- [14] Polanyi, M., *The Tacit Dimension*, Routledge & Kegan Paul, Gloucester, 1966.
- [15] Biggam, J., *Defining Knowledge: An Epistemological Foundation for Knowledge Management*, in: *Proceedings of the IEEE Computer Society Press*, 2001.
- [16] Dey, A.K., Abowd, G.D., *Towards a Better Understanding of Context and Context-Awareness*, College of Computing, Georgia Institute of Technology, Atlanta, GA, 1999.
- [17] Sowa, J.F., *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Brooks/Cole, Pacific Grove, 2000.
- [18] Klemke, R., *Context Framework - an Open Approach to Enhance Organisational Memory Systems with Context Modelling Techniques*, in: *Proceedings of the Third International Conference on Practical Aspects of Knowledge Management (PAKM 2000)*, 2000.
- [19] Wargitsch, C., Habermann, F., *IMPACT: Workflow-Management-System als Instrument zur koordinierten*

Prozessverbesserung: Anforderungen, Universität Erlangen-Nürnberg Bereich Wirtschaftsinformatik, Nürnberg, 1998.

[20] Probst, G.J.B., Raub, S., Romhardt, K., Wissen managen - Wie Unternehmen ihre wertvollste Ressource optimal nutzen, Gabler, Wiesbaden, 1999.

[21] Andersen, A., The Knowledge Management Assessment Tool: External Benchmarking Version, 1996.

[22] Alavi, M., KPMG Peat Marwick U.S.: One Giant Brain, Harvard Business School, 1997.

[23] Güldenber, S., Wissensmanagement und Wissenscontrolling in lernenden Organisationen: ein systemtheoretischer Ansatz, 1996.

[24] Bristow, P., Dickinson, C., Duke, S., Henry, S., Makey, P., Enterprise Portals: Business Application and Technologies, Butler Group, East Yorkshire, 2001.

[25] Röhrich, J., Schlögel, C., cBusiness - Erfolgreiche Internetstrategien durch Collaborative Business, Addison-Wesley, München, 2001.

[26] Kalakota, R., Robinson, M., M-Business, McGraw-Hill Publishing Company, New York, 2001.

[27] Davydov, M.M., Corporate Portals and e-Business Integration, McGraw-Hill, New York, 2001.

[28] Schwarz, J., Mass Customization von Prozessen durch Unternehmensportale, in: Information Management & Consulting, Vol. 15 (2000) Issue 2, S. 40-45.

[29] Dias, C., Corporate Portals: a literature review of a new concept in information management, in: International Journal of Information Management, (2001) 21, p. 269-287.

[30] Fleisch, E., Österle, H., Das Tor zur IT-Welt: Thesen zum erfolgreichen Portaleinsatz, in: Computerwoche extra, (2001) 2, S. 28-31.

[31] Puschmann, T., Collaboration Portale - Architektur, Integration, Umsetzung und Beispiele, Dissertation, Universität St. Gallen, Difo-Druck, Bamberg, 2003.

[32] Raghavan, P., Information Retrieval for Enterprise Content, in: UPGRADE (The European Online Magazine for the IT-Professional), III (2002) 3, p. 5-8.

[33] Baeza-Yates, R., Schäuble, P., Retrieving Information: A Discipline with a Tradition, in: UPGRADE (The European Online Magazine for the IT-Professional), III (2002) 3, p. 3-4.

[34] Andrews, W., Visionaries Invade the 2003 Search Engine Magic Quadrant, Gartner, Inc., Stamford, 2003.

[35] Rappoport, A., Web Search Engines: Users Surprising the Experts, <http://www.searchtools.com/slides/baychi2002/index.html>, 2002.

[36] Gordon, M., Pathak, P., Finding information on the World Wide Web - The retrieval effectiveness of search engines, in: Information Processing and Management, Vol. 35 (1999), p. 141-180.

[37] Cathro, W., Metadata: An Overview, <http://www.nla.gov.au/nla/staffpaper/cathro3.html>, 1997.

[38] Berners-Lee, T., Web architecture: Metadata Tim Berners-Lee, 1997, <http://www.w3.org/DesignIssues/Metadata>, 1997.

[39] McGovern, G., Why metadata is important, http://www.gerrymcgovern.com/nt/2001/nt_2001_10_01_metadata.htm, 2001.

[40] Dublincore, Dublin Core Metadata Initiative, <http://dublincore.org/>, 2003.

[41] Baeza-Yates, R., Ribeiro-Neto, B., Modern Information Retrieval, Addison Wesley, New York, 1999.

[42] Rath, H.H., Pepper, S., Topic Maps: Introduction and Allegro, in: Proceedings of the Markup Technologies 99, 1999.

[43] ISO, ISO/IEC 13250, International Organization For Standardization, Geneva, <http://www.iso.org/iso/en/CatalogueListPage.CatalogueList?ICS1=35&ICS2=240&ICS3=30>, 2002.

[44] Wrightson, A., Topic Maps and Knowledge Representation, Ontopia AS, 2001.

[45] Pepper, S., Navigating haystacks and discovering needles, in: Markup Languages: Theory and Practice, Volume 1, Issue 4, 1999.

[46] Goldfarb, C.F., Prescod, P., XML Handbook - Anwendungen, Produkte, Technologien, Addison-Wesley Verlag, München, 2000.

[47] Smolnik, S., Nastansky, L., K-Discovery: Using Topic Maps to identify Distributed Knowledge Structures in Groupware-based Organizational Memories, in: Proceedings of the Proceedings of the 35th Hawaii International Conference on System Sciences, IEEE Computer Society Press, Los Alamitos CA, Washington, Brussels, Tokyo, 2003, pp. 10.

[48] Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., From Data Mining to Knowledge Discovery: An Overview, in: Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R., Advances in Knowledge Discovery and Data Mining, AAAI Press (American Association for Artificial Intelligence), Menlo Park, CA, 1996, p. 1-36.

[49] Yin, R., Case Study Research. Designs and Methods, SAGE Publications, London, 1994.

[50] Kremer, S., Kolbe, L.M., Brenner, W., Do you know your terms? - A procedure model for Terminology Management, accepted for: Proceedings of the European Conference of Information Systems (ECIS), 2003.

[51] Biezunski, M., Newcomb, S.R., Topic Maps and XTM - A Manager's Overview, in: Proceedings of the Proceedings of the Knowledge Technologies 2001 (KT2001), 2001.

[52] Smolnik, S., Nastansky, L., Knieps, T., Mental Representations and Visualization Processes in Organizational Memories, in: Proceedings of the Proceedings of the 7th International Conference on Information Visualization (IV03) - International Symposium on Knowledge Domain Visualization (IV03-KDViz), IEEE Computer Society Press, Washington, Brussels, Tokyo, 2003.