Knowledge Discovery Methods: An Exploration of Different Approaches

Stefan Smolnik

Dept. of Business Computing 2
University of Paderborn
Warburger Str. 100
D-33098 Paderborn
Stefan.Smolnik@notes.upb.de

Abstract: Coping with complex and dynamic environments forces companies to invest in the resource knowledge for the maintenance of their operational reliability; they have to create and to renew the resource knowledge continuously. Elementary activities of knowledge management comprise collecting, organizing, connecting, sharing, and applying of knowledge. A necessary prerequisite before the outcome of knowledge can be deployed and activities can be utilized is the availability of knowledge. Thus, creation and discovery of knowledge play a decisive role to increase the availability of knowledge assets of a company, the growth of corporate knowledge, the capacity to act, and finally, the effectiveness of companies. This paper contributes to the understanding of knowledge management through an exploration of different approaches of knowledge discovery methods. A detailed literature study forms the basis of this contribution and thus, it results in a selection of several approaches of different research domains, which seems to be appropriated for our purpose. In detail, we focus on following main approaches: developmental psychology and epistemologies, like the ways small children learn to discover the world (following Piaget) or inquiring systems (following Churchman), artificial intelligence methods, like data mining, the discovery of general world knowledge in texts, and the creation of knowledge structures and reusable structured semantic link networks. The findings comprise delimitations of specific paradigms, characteristics, and fields of application of knowledge discovery methods as well as similarities and interdependencies. We argue that the exploration of these methods is a necessary prerequisite to fill "knowledge gaps", as the importance of the resource knowledge as a success factor increases continuously.

1 Introduction and overview

Ever more, knowledge plays a decisive role as a factor of production besides the classical 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 and decisions. A company's ability to improve organizational use of implicit and explicit knowledge, in other words, its ability to deploy its organizational memory, contributes to the process of knowledge generation, i.e. the development or accumulation of new knowledge, and helps the company to realize its objectives. A necessary precondition for the core processes of knowledge identification and knowledge use are knowledge discovery methods, e.g. mechanisms of navigation and linking, as well as functionalities for extensive searches and investigations are needed to explore and to use complex knowledge offer.

In general, and depending on the structure of an organizational memory information system, different methods of knowledge discovery can be considered. Piaget's researches in developmental psychology and genetic epistemology, e.g. the discovery that small children exhibit as they learn about the world, lead to the finding, that the growth of knowledge is a progressive construction of logically embedded structures superseding one another by a process of inclusion of lower less powerful logical means into higher and more powerful ones up to adulthood [Pi81], [PI69]. Therefore, children's logic and modes of thinking are initially entirely different from those of adults. Churchman's idea was to look at different epistemologies that the history of philosophy has brought forth as designs for inquiring systems, i.e. systems that would be capable of learning. Cognition is the creation of knowledge and understanding: An individual discovers what it could be or will be, and how the world could be or should be [Ch71]. An approach, which is originated in the research field of artificial intelligence (AI), is known as *knowledge discovery in databases* (KDD). Statistical analysis as in data mining is used to discover unsuspected relationships in often large sets of structured data [HMS01]. Schubert, a researcher in natural language understanding, points out that there is a largely untapped source of general knowledge in texts, lying at a level beneath the explicit content. This knowledge consists of relationships implied to be possible in the world, or, under certain conditions, implied to be normal or commonplace in the world [Sc02]. The ISO standard *ISO/IEC 13250 Topic Maps* defines a model and architecture for the semantic structuring of link networks. By applying topic maps to large sets of heterogeneous information resources, knowledge structures and reusable structured semantic link networks are created in a meta layer above those resources.

In our contribution, we will present and evaluate the mentioned methods of knowledge discovery. We will outline and delimit their specific paradigms, characteristics, and fields of application. In doing so, we attempt to clarify the meaning and methods of knowledge discovery, and thus provide strong conceptual foundations for the discipline of knowledge management, particularly for the core process of knowledge identification in organizational memory information systems.

2 Theoretical framework and delimitation of terms

2.1 Knowledge

Within literature, many definitions of knowledge can be found. In this contribution, we want to give a brief overview of some important definitions, because this term is vital for our further elaborations. In subsequent sections, we refer to the following definitions and explanations.

Davenport/Prusak define knowledge as follows [DP98]: “Knowledge is a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information. It originates and is applied in the minds of knowers. In organizations, it often becomes embedded not only in documents or repositories but also in organizational routines, processes, practices, and norms.” Therefore, knowledge comprises both information and person-specific aspects like experiences, values, and insights. Nonaka/Takeuchi understand knowledge as the flow of information in combination with intrinsic attitudes and values of the receiver. Thus, information as well as knowledge can be seen as context-specific [NT95] and it can be deduced that knowledge is strongly linked to individuals within organizations. Davenport/Prusak point out that skills based on knowledge, like to organize, to select, and to judge, probably depend more on values and intrinsic attitudes of an individual than on explicit available information and logic [DP98].

Even though a differentiation between data, information, and knowledge is important, we only give a brief definition of data and information. Data characterizes several objective facts of events or processes [DP98]. Thus, it is meaningless symbols and it is machine readable. Information is a flow of messages or processed data [LC00]; it is data, which has an effect [DP98]. For more detailed discussions of these terms refer to Davenport/Prusack [DP98] or Lai/Chu [LC00].

An important characteristic of knowledge and difference to information is the strong affinity of knowledge to activities (cf. [DP98]). Individuals act and react due to their experiences and intrinsic attitudes. Murray states [Mu96]: “All doing is knowing and all knowing is doing (...) We admit knowledge whenever we observe an effective behavior in a given context.” Thus, in the sense of semiotics, knowledge contains a pragmatic dimension.

Knowledge is much more than transformed information, which can be presented in the form of information or data. Sveiby points out [Sv97]: “Explicit knowledge in the form of facts is therefore ‘metaphorically speaking’ only the tip of the iceberg.” This statement indicates the distinction between implicit and explicit knowledge, which is often discussed in literature. This differentiation was first defined by Polanyi, who developed a concept for implicit knowledge, which he described as follows [Po97]: “We can know more than we can tell.” For further discussions of different types of knowledge refer to Biggam [Bi01] and Lai/Chu [LC00]. In particular, Nonaka/Takeuchi discuss some philosophical approaches for the definition of knowledge and classifications of different knowledge types [NT95].

2.2 Knowledge discovery

Within literature, many knowledge management activities, methods, or modules are discussed. Lai/Chu suggest an integrated knowledge management framework, which comprises the activities initiation, generation, modeling, repository, distribution and transfer, use, and retrospect [LC00]. Davenport/Prusak differentiate between determine requirement, capture, distribute, and use [DP98]. A pragmatic approach for the organizationwide management of knowledge is presented by Probst/Raub/Romhardt [PR99]. 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/Takeuchi [NT95], Arthur Andersen [Ar96], and Alavi [Al97]. All these approaches have implicitly or explicitly a method for the identification or use of knowledge in common. Normally, unused knowledge is available within organizations, which has to be uncovered by appropriate methods, and therefore, it could be utilized for organizations. Knowledge discovery methods enable an increase of transparency of knowledge in organizations and support users finding relevant knowledge. They are a necessary precondition for the core processes of knowledge identification and knowledge use [PR99], 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 [Gü98].

3 Exploration of knowledge discovery methods

Depending on different scenarios in general as well as on the structure of an organizational memory information system, different knowledge discovery methods can be considered. We have exemplarily chosen discovery methods which are representative for different research fields. Beyond this selection, there are several other methods, we do not focus on. We begin our exploration with an approach of Piaget, which describes how small children learn to discover the world. Piaget's theory provides an epistemological explanation of how individuals learn and how knowledge is constructed. This is followed by a description of Churchman's inquiring systems, which provide a philosophical basis and new perspective for the organizational knowledge management and organizational learning. In the subsequent sections, actual knowledge discovery methods are presented. Knowledge discovery in databases resp. data mining, text analysis and taxonomies, as well as topic maps provide approaches to discover knowledge in different kinds of organizational memories.

3.1 The discovery of small children

Piaget defines genetic epistemology as "… attempting to explain knowledge, and in particular scientific knowledge, on the basis of its history, its sociogenesis, and especially the psychological origins of the notions and operations upon which it is based." [Pi70] Piaget's results are based on the full range of contemporary mathematical knowledge, an extensive empirical base of observation of the learning of very young children and reports of observations of older children and a general knowledge of the development of knowledge in history. We focus on Piaget's theory of children's cognitive development and, in particular, on the discovery processes they pass. Piaget differentiates four stages of intellectual development (following [Pi81], [At02a]):

1. Sensori-motor (birth-2 years): During the sensori-motor stage, children learn to use basic motor and sensory skills to interact with the world. These skills and a basic ability to interpret them are the child's basic elements. Children establish a first cognitive orientation, a cognitive environment including concrete objects in an external reality from which they differentiate themselves. They recognize themselves as agents of action and begin to act with intentions, e.g. they press a doll to receive a sound. Furthermore, they achieve object permanence, i.e. they realize that things continue to exist even when these things are no longer present to children's sense. Piaget sub-classifies this stage into 6 phases, which form the process until sensori-motoric intelligence occurs. For a comprehensive description of these phases refer to Piaget [Pi81] and Piaget/Inhelder [PI69].
2. Pre-operational (2-7 years): During the pre-operational stage, children learn to use language and to represent objects by images and words. They classify objects by a single characteristic, e.g. they group together all the blue blocks regardless of their shape or all the round blocks regardless of their color. The internalization of action takes place by means of not reversible thinking. In addition, the thinking of children is still egocentric, i.e. they have difficulties to take up the viewpoint of others.
3. Concrete-operational (7-11 years): During the concrete-operational stage, children begin to think logically about objects and events. They classify objects according to several characteristics and can order them in series along a single dimension such as size. Number, mass, and weight are conserved. Even though formal thinking is not existent, simple operations regarding classes and relations are performed. Operations describe the process of elaborating something in mind. For instance, children in the sensori-motor and pre-operational stages have to act, and try things out in the real world; in contrast, older children and adults can do more in their minds.
4. Formal operations (11 years and up): During the formal operations stage, children start to think logically about abstract propositions and test hypotheses systematically. They become concerned with the hypothetical, the future, and ideological problems. Their thinking is propositional, content-free and logic.

Piaget was very interested in knowledge and how children come to know their world. As children develop and continuously interact with the world around them, knowledge is invented and reinvented. According to Piaget, knowledge is not merely transmitted verbally but must be constructed and reconstructed by individuals who want to discover. For children to know and to construct knowledge of the world, they act on objects, and it is this action which provides knowledge of those objects; the mind organizes reality and acts upon it. In autonomous activities, children discover relationships and ideas in classroom situations that involve activities of interest to them. Understanding is built up step by step through active involvement. Piaget concludes in his theory that scientific knowledge is a continuous process of construction and reorganization, and not a thing of a moment or something static [Pi70].

3.2 Inquiring systems

"Cognition creates knowledge." [Ch71] Churchman emphasizes the importance of activities to create knowledge. Knowledge depends on individuals and not on information sets. Organizations, which are supported by information technologies (IT), can be considered as inquiring systems since creation and discovery of knowledge are essential for their efficiency in competitive environments [CCP98]. Churchman discusses the viewpoints of the philosophers Leibnitz, Locke, Kant, Hegel, and Singer in the context of inquiring systems. The purpose of these systems is the creation of knowledge. We give a brief summarization of the aspects and characteristics of each philosophical inquiring system and of the IT support that is needed by organizations based on them, following Courtney/Croasdell/Paradice [CCP98]:

- Leibnitzian inquiring system: These systems are closed systems without access to external environments. They operate based on given axioms and may fall into competency traps based on decreasing returns from the heuristics embedded in the inquiring processes. [Ma00] Processes of formal logic and sentence generator lead to fact nets, tautologies, and contingent truths. In terms of IT support, expert systems result in simple error detection and correction, and suggested courses of action. [CCP98]
- Lockean inquiring system: Based on consensual agreement, these systems aim to reduce equivocality embedded in the different interpretations of the world-view [Ma00]. By means of communication and assigning labels to elementary observations, taxonomies (cf. section 3.4) are received. Equivocality reduction in organizations is achieved by networking and the use of databases. [CCP98]
- Kantian inquiring system: These systems attempt to give multiple explicit views of complementary nature and are best suited for moderate ill-structured problems [Ma00]. They enable the construction of models from theories, the interpretation of data, and the choice of best suited models. Facilitators to get integrated, timely knowledge are the internet, the model base, and knowledge bases. [CCP98]
- Hegelian inquiring system: A synthesis of multiple, completely antithetical representations is the basis of these systems. The representations are characterized by intense conflict because of the contrary underlying assumptions. [Ma00] Dialectic and the construction of thesis and antitheses lead to synthesis. Dialectic engines support conflict resolution, renewal, enlarged perspectives, and new strategic directions. [CCP98]
- Singerian inquiring system [CCP98]: Specifications of steps to be followed in resolving disagreements among members of a community build a system of measures. Strategies of agreement result in new standards, exoteric knowledge, and simplistic optimism. In particular, expert systems and the internet support the creation of new measures and exoteric knowledge.

The presented inquiring systems provide a philosophical and pragmatic framework to address the critical processes of discovery of new knowledge and renewal of existing knowledge for better design of organizational knowledge management systems. Different methods depending on the inquiring system enable the creation and discovery of different kinds of knowledge and could be supported by information technologies (v. tab. 1, cf. [CCP98]).

	Leibnitz	*Locke*	*Kant*	*Hegel*	*Singer*
Knowledge discovery method	*Formal logic* *Sentence generator*	*Assign labels to inputs* *Communication*	*Construct models from theories* *Interpret data* *Choose best model*	*Construct theses, antithesis* *Dialectic*	*Strategy of agreement* *Sweeping-in*
Knowledge	*Fact nets* *Tautologies* *Contingent truths*	*Taxonomy*	*Fact Nets*	*Synthesis*	*New standard* *Exoteric knowledge* *Simplistic optimism*
IT support	*Expert systems*	*Databases* *Networking*	*Internet* *Knowledge bases* *Model base*	*Dialectic engines*	*Expert systems* *Internet/WWW*

Table 1. Summary of inquiring systems (following Courtney/Croasdell/Paradice [8])

An approach of combining decision support systems and inquiring systems is presented by Hall/Paradice/Courtney. They come to the conclusion that this combination enables organizations to design a comprehensive knowledge management system that will fully support learning within organizations [HPC01].

3.3 Knowledge discovery in databases and data mining

Fayyad et al. define knowledge discovery in databases (KDD) as follows [F96]: "Knowledge discovery in databases is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data." This term originates in the artificial intelligence (AI) research field. The definition of KDD comprises many aspects of knowledge discovery in the sense of section 2.2. The identification of patterns in large structured data sets results in knowledge. The process of KDD involves different interactive and iterative steps (v. fig. 1): selection, preprocessing, transformation, data mining, and interpretation resp. evaluation [FPS96], [ES00]. The complete KDD process is described in more detail in Brachman/Anand [BA96].

The core process of knowledge discovery is data mining. Hand/Mannila/Smyth define [HMS01]: "Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner." Thus, in the context of data mining, knowledge can be interpreted as discovered relationships between data sets, which have a novel meaning for an individual. The most important tasks of data mining, thus the process of seeking relationships within data sets, are (v. fig. 2) [FPS96]:

- Clustering and deviation detection: Clustering is a task to identify a finite set of categories or clusters to describe data. Deviation detection focuses on discovering data items, which belong to no category or cluster [ES00].
- Classification: A learning function maps data items into one of several predefined classes. Given are training items, which are already classified and which form a basis for the learning function.
- Dependency modeling: A model describes significant dependencies between variables. Association rules support the identification of coherencies within data sets.
- Summarization: The goal of summarization is to find a compact description of a subset of data. Attribute values are generalized and the number of data items is reduced [ES00].

KDD resp. data mining enables the non-trivial extraction of implicit, previously unknown, and potentially useful knowledge from large data sets. A variety of techniques is used to identify decision-making knowledge in data sets, and extracting it in such a way that it can be put to use in areas such as decision support, prediction, forecasting, and estimation. The data is often voluminous but of low value as no direct use can be made of it; it is the hidden knowledge in the data that is useful and knowledge discovery methods enable its extraction.

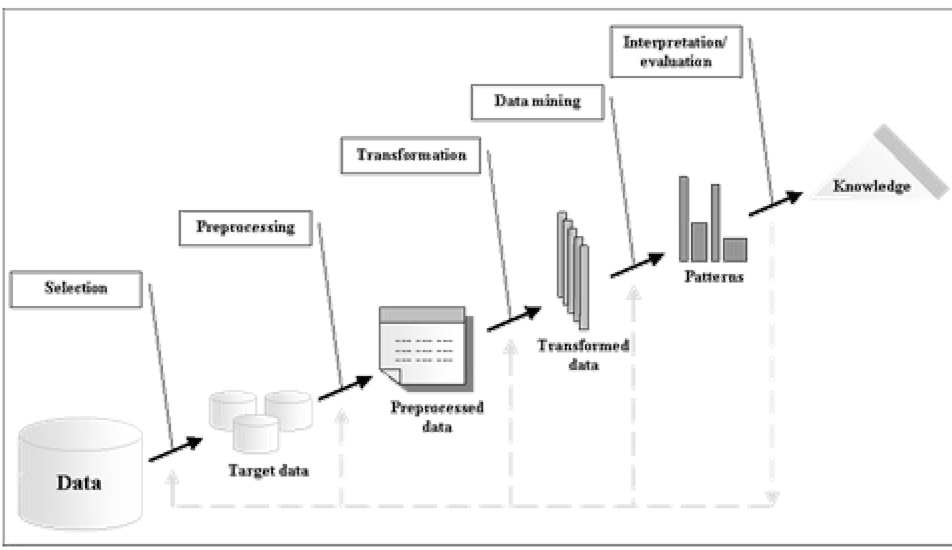

Figure 1. Steps of the KDD process [FPS96]

An example of discovering knowledge in large data sets is presented by McCown/Milligan. They describe a method "... by which users can intelligently and automatically navigate and sift through acquired data and identify information content." [MM00]. In particular, they address large organized databases like data warehouses to provide users a navigation framework. This approach enables the easy and intelligent search for requested data and provides a method to visually discover relationships within data sets and therefore, to create knowledge.

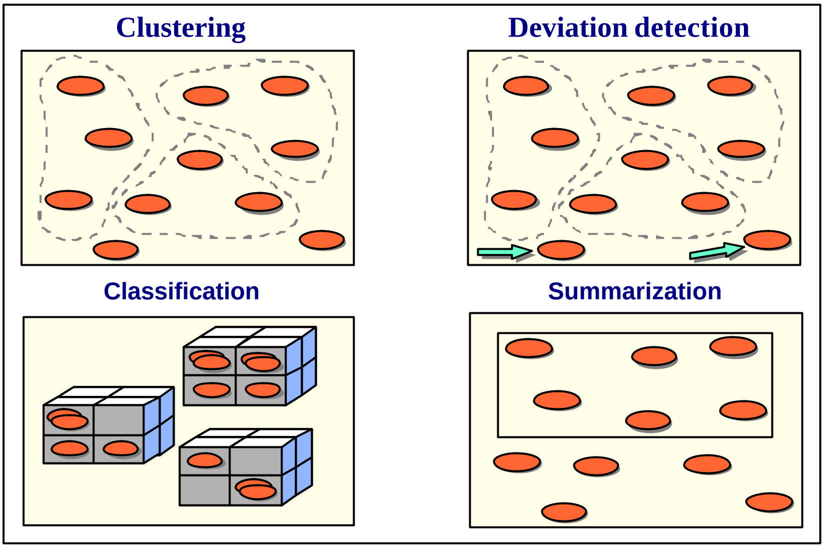

Figure 2. Data mining tasks [ES00]

3.4 Deriving general world knowledge from texts and taxonomies

Texts potentially contain great assets of knowledge. However, text represents factual information in a complex, rich, and opaque manner [NN01]. Standard statistical data mining methods, like those described in section 3.3, couldn't be used for an analysis of texts in contrast to an analysis of numerical and structured data. In general, several technologies have been developed to work with textual data. Information retrieval methods extract keywords and enable the search of text. Clustering and classification based on the analysis of keyword distributions help to organize multiple texts and result in an overview of relevant topics. In our contribution, we focus on the discovery of knowledge to extract interesting information from content. Nasukawa/Nagano differentiate several kinds of knowledge that "... can be extracted from textual data, such as linguistic knowledge for natural language processing (NLP) and domain-specific lexical and semantic information that may be stored in a database." [NN01] Schubert goes one step further: He thinks "… that there is a largely untapped source of general knowledge in texts, lying at a level beneath the explicit assertional content." [Sc02] This knowledge consists of experiences, values, and relationships, which are normal, commonplace, or well known in the world. For instance, from the sentence "She drove to London for shopping." can be followed that a female person, who obviously can drive a car and possibly owns a car, lives nearby London, that London is a place to go shopping, etc. To derive this general world knowledge from texts, Schubert suggests two steps [Sc02]: Firstly, general propositions are derived from noun phrases and clauses. In contrast to mechanism of compositional semantic interpretation, found meanings are abstracted by simplifying and generalizing them before they are combined to a phrase. Modifiers and unimportant conjuncts are deleted as well as individual terms are generalized to types, e.g. "a red, fast car built in Italy" would be transformed into "a car". Secondly, stronger generalizations, based on the nature and statistical distribution of the claims found in the first step, are derived. For instance, step 1 results in simple variants of the first output proposition like "a female may drive a car". From this, it could be followed that only people drive cars. For a deeper insight and secondary literature refer to Schubert [Sc02]. The described methods focus the knowledge discovery within a text. Moreover, Kazi/Ravin present an approach to discover relations between themes, concepts, or subjects across multiple texts [KR00]. They address the challenge of many different contexts by minimizing the need of context comparisons.

In a broader context, these different methods and approaches to identify important linguistic and semantic elements (understanding phrases or the parts of speech) in texts belong to the research field of taxonomy. Taxonomies are structures of categories or topics to which documents, thus texts, can be assigned [Le01]. Another method to discover knowledge in heaps of unorganized information are ontologies. This term was introduced by researchers of artificial intelligence (AI) to facilitate knowledge sharing and reuse [Fe01]. An ontology is a data construct that reflects the structure of a body of knowledge by including categories, vocabulary, and information about relationships; an ontology is generally assembled by humans and then used to analyze documents [Le01]. Ontologies create a common and shared understanding of some domain and this understanding can be communicated between people and application systems [Fe01]. They capture consensual knowledge, i.e. it is not restricted to some individual, but accepted by a group [Fe01], and thus, they are an important and basic method to enable knowledge discovery.

3.5 Topic maps: knowledge structures and structured semantic link networks

The ISO standard *ISO/IEC 13250 Topic Maps* defines a model and architecture for the semantic structuring of link networks. By applying topic maps to large sets of heterogeneous information resources, reusable structured semantic link networks are created upon those resources [RP99].
The key concepts of topic maps are topics, occurrences of topics, and relationships between topics (topic associations). A *topic* is a construct that represents a real world subject and in this sense a topic can be everything: a theme, a concept, a subject, a person, an entity, etc. (v. fig. 3). A concrete topic is an instance of a topic type. Therefore, a topic and a topic type form a class-instance relationship. At the same time a topic type is also a topic. A topic may be linked to one or more real information objects, like a report, a comment, a video or a picture, that are considered to be relevant to the topic in some way. Such information objects are called *occurrences* of the topic (v. fig. 3). Generally, an occurrence is not part of a topic map. The link mechanism itself depends on the underlying system. *Topic associations* describe the relationships between topics (v. fig. 3). They are completely independent of the real information object and represent the essential value-add of the topic map. This concept leads to some conclusions: A concrete topic map can be applied to different information repositories.

Seen from the other side, different topic maps can be applied to one information repository and therefore, they can provide different views e.g. for different users. Furthermore, topic maps are interchangeable and they can be merged. Generally, topic associations are not one-way relationships. They are symmetric as well as transitive and thus, they have no direction. In addition, the topic map standard provides the extended concepts of scope, public subject, and facets, which are not described here. For a comprehensive reference refer to Rath/Pepper [RP99] and ISO/IEC 13250 [Is00].

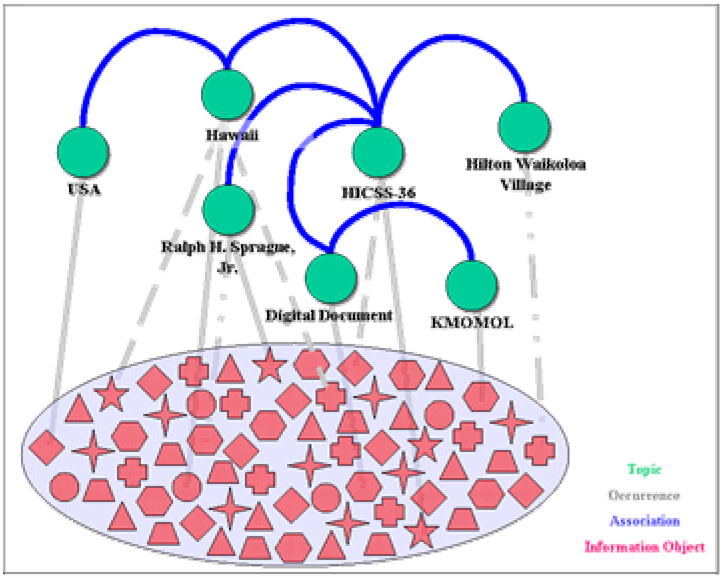

Figure 3. Key concepts of topic maps

The addition of topic associations to the concept of topics enables topic maps to be able to model networks of information. Topic maps organize information resources into a new knowledge space, by relating them to topics, and associating those topics, in a structured way. Furthermore, they enable heterogeneous sets of information resources to be used together, by interrelating them using a unifying conceptual framework. Another characteristic of topic map 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. [Wr01] The link mechanism between topics and occurrences provides a means for "bridging the gap" between knowledge representation and the field of information management [Pe99].

The human brain always remembers memorized things in a specific context [GP00]. 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. Topic maps establish an associative network between information objects which represents subjects, and provide navigation paradigms to enable its search. Therefore, topic maps embody semantic relations between subjects and enable the discovery of knowledge structures in heterogeneous sets of information resources. An example of using topic maps to identify distributed knowledge structures in groupware-based organizational memories is presented in Smolnik/Nastansky [SN02]. They focus on a technical approach by means of using topic maps in groupware environments. Our contribution can be seen as a theoretical foundation for the practical results provided by their study.

A complementary approach, which addresses the same objectives and fields of application as topic maps, is called *Resource Description Framework* (RDF). As opposed to the subject-oriented topic maps, which form a semantic layer above information objects, RDF is focused on resources, i.e. on information objects itself. Thus, there are fundamental differences in conception and implementation. A detailed introduction as well as a comparison to topic maps can be found in Moore [Mo01].

4 Findings

The presented methods of knowledge discovery have different origins and apply to different contexts and scenarios. We have summarized several methods, algorithms, and activities as well as resulting knowledge types in table 2.

From Piaget we learn that the growth of knowledge is an active construction process in which individuals build increasingly differentiated and comprehensive cognitive structures through their own activities. Individuals try to adapt to their environment and in doing so, they organize thinking into knowledge structures. In the context of organizational memory information systems, this process of creating knowledge structures can be carried out by topic maps (cf. [SN02]). They enable the creation of knowledge structures in those systems and support thinking of individuals by establishing associative networks. Furthermore, important implications of Piaget's research are the power of exploration and discovery as well as the role of interest to improve the growth of knowledge. However, he underemphasizes the influences of social, historical, and cultural forces. To improve and to support the organizational knowledge management, companies have to overcome these limitations and have to establish adequate working environments.

In their studies about inquiring organizations, Courtney/Croasdell/Paradice have worked out the implications of inquiring systems and the support of information technologies for them, on the development of new knowledge and insights that potentially influence behavior, thus, on organizational learning [CCP98]. The inquiring systems described by Churchman provide a different basis and framework for knowledge discovery methods and therefore, for the organizational knowledge management, e.g. classifications, found in Locke's inquiring system, are also used in KDD and data mining, and result in taxonomies. Similarly, document and text analysis methods yield in taxonomies. In addition, experiences and relationships can be found in texts, like as results of Piaget's theory, data mining, or topic maps.

In a nutshell, knowledge discovery methods combine activities on different kinds of objects. These objects can be real world objects like dolls or cars, standards, procedures, organizational memories, large sets of structured data, documents and texts, or potentially large sets of heterogeneous information resources. The activities comprise cognitive processes like to gain experiences and insights, the construction of intrinsic values, personal qualities, and intangible assets, the association of objects and the construction of semantic networks of relations. Several knowledge discovery methods have some activities in common. For instance, classification mechanisms are found in Piaget's approaches, in inquiring systems, and in KDD resp. in data mining. Ontologies or association mechanisms are activities of text analyses, data mining, and topic maps.

	Discovery of children	*Inquiring systems*	*KDD and data mining*	*Text analysis and taxonomies*	*Topic maps*
Methods, algorithms, and activities	*Observation* *Cognitive orientation* *"Learning by doing"* *Classification* *Internalization* *Operation* *Thinking*	*Formal logic* *Communication* *Classification* *Construction* *Interpretation*	*Clustering* *Deviation detection* *Classification* *Dependency modeling* *Summarization*	*Derivation* *Abstraction* *Simplification* *Generalization* *Ontologies*	*Associating* *Structuring* *Modeling* *Ontologies*
Resulting knowledge types	*Experiences* *Real world objects and events* *Acting and thinking processes* *Causes and effects*	*Fact nets* *Tautologies* *Taxonom ies* *Synthesis* *New standards* *Exoteric knowledge*	*Valid, novel, useful, and understandable patterns* *Unsuspected relationships*	*Experiences, values, and relationships* *Normal, commonplace, and well known knowledge* *Taxonomies*	*Reusable structured semantic link networks* *Semantic relations*
Fields of application	*Epistemological foundation*	*Philosophical foundation*	*Large sets of structured data*	*Documents and texts*	*Large sets of heterogeneous information resources*

Table 2. Overview of knowledge discovery methods

Therefore, they support the development, collection, and renewal of knowledge in organizations. Thus, they form a fundamental basis of organizational knowledge management.

5 Summary and Conclusions

Based on a comprehensive literature study, we have presented an exploration of different knowledge discovery methods. While this study contributes to our understanding of knowledge discovery, there are some limitations. It is important to note that our findings are based on our own selection of knowledge discovery methods. Additional research should be conducted to include other methods, like different approaches in developmental psychology and epistemology, or the discovery of process knowledge. Furthermore, we intrinsically focus on the methods and do not consider the concrete integration into organizational knowledge management environments in detail, or social or cultural aspects that come along with this integration.

However, we provide a strong conceptual foundation for the organizational knowledge management with the clarification of the meaning of knowledge discovery and the exploration of several methods. We argue that several methods comprise similar activities and result in several types of knowledge. Historical and theoretical approaches establish a basis and framework for actual knowledge discovery methods used in organizational memory information systems.

Focusing on the mentioned limitations is an important area for further research. Future work could extend the variety of knowledge discovery methods and therefore, provide a broader basis for the findings of this study. In addition, these results could be applied to organizational memory information systems in general, or to the specific case of a groupware-based organizational memory information system (cf. [SN02]). In doing so, a concrete case study should be conducted, which could lead to further insights and practical conclusions.

This study supports the understanding of the role of knowledge discovery methods. The presented methods enable the identification and generation of knowledge, and support companies' ability to explore and assess their knowledge bases. Exploring different methods of knowledge discovery will potentially be an important research field of organizational knowledge management as the importance of the resource knowledge as a success factor increases continuously.

References

[Al97] Alavi, Maryam: KPMG Peat Marwick U.S.: One Giant Brain, Harvard Business School, Case 9-397-108, July 11, 1997.

[Ar96] Arthur Andersen and the American Productivity and Quality Center: The Knowledge Management Assessment Tool: External Benchmarking Version, Winter 1996.

[At02a] Atherton, James S.: Learning and Teaching: Piaget's developmental psychology; http://www.dmu.ac.uk/~jamesa/learning/piaget.htm; United Kingdom, 2002.

[At02b] Atwood, Michael E.: Organizational Memory Systems: Challenges for Information Technology, Proceedings of the 35th Annual Hawaii International Conference on System Sciences, IEEE Computer Society Press, Los Alamitos, January 2002.

[Bi01] Biggam, J.: Defining Knowledge: An Epistemological Foundation for Knowledge Management, Proceedings of the 34th Annual Hawaii International Conference on System Sciences, IEEE Computer Society Press, Los Alamitos, January 2001.

[BA96] Brachman, Ronald J.; Anand, Tej: The Process of Knowledge Discovery in Databases; in: Fayyad, Usama M.; Piatetsky-Shapiro, Gregory; Smyth, Padhraic; Uthurusamy, Ramasamy: Advances in Knowledge Discovery and Data Mining, American Association for Artificial Intelligence, Menlo Park, California, USA, 1996.

[Ch71] Churchman, C. West: The Design of Inquiring Systems: Basic Concepts of Systems and Organization, Basic Books, New York, 1971.

[CCP98] Courtney, James F.; Croasdell, David T.; Paradice, David B.: Inquiring Organizations, in: Australian Journal of Information Systems, Fall, 6(1):3-14, 1998.

[DP98] Davenport, Thomas H.; Prusak, Laurence: Working Knowledge – How Organizations Manage What They Know; Harvard Business School Press, Boston, Massachusetts, 1998.

[ES00] Ester, Martin; Sander, Jörg: Knowledge Discovery in Databases - Techniken und Anwendungen, Springer-Verlag, Berlin, Heidelberg, 2000.

[F96] Fayyad, Usama M.; Piatetsky-Shapiro, Gregory; Smyth, Padhraic; Uthurusamy, Ramasamy: Advances in Knowledge Discovery and Data Mining, American Association for Artificial Intelligence, Menlo Park, California, USA, 1996.

[FPS96] Fayyad, Usama M.; Piatetsky-Shapiro, Gregory; Smyth, Padhraic: From Data Mining to Knowledge Discovery: An Overview; in: Fayyad, Usama M.; Piatetsky-Shapiro, Gregory; Smyth, Padhraic; Uthurusamy, Ramasamy: Advances in Knowledge Discovery and Data Mining, American Association for Artificial Intelligence, Menlo Park, California, USA, 1996.

[Fe01] Fensel, Dieter: Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce, Springer-Verlag, Berlin, Heidelberg, 2001.

[GP00] Goldfarb, Charles F.; Prescod, Paul: Das XML-Handbuch - Anwendungen, Produkte, Technologien; 2., aktualisierte und erweiterte Auflage, Addison-Wesley Verlag, München, 2000.

[Gü98] Güldenberg, S.: Wissensmanagement und Wissenscontrolling in lernenden Organisationen: ein systemtheoretischer Ansatz; DUV: Wirtschaftswissenschaft, Edition Österreichisches Controller Institut, Zugl.: Wien, Wirtschaftsuniv., Diss., 1996 2. durchges. Auflage, DUV, Deutscher Universitäts-Verlag GmbH, Wiesbaden, 1998.

[HPC01] Hall, Dianne J.; Paradice, David B.; Courtney, James F.: Creating Feedback Loops to Support Organizational Learning and Knowledge Management in Inquiring Organizations, Proceedings of the 34th Annual Hawaii International Conference on System Sciences, IEEE Computer Society Press, Los Alamitos, January 2001.

[HMS01] Hand, David; Mannila, Heikki; Smyth, Padhraic: Principles of Data Mining; The MIT Press, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, 2001.

[Is00] ISO/IEC: International Standard ISO/IEC 13250:2000(E) - Information technology - SGML Applications - Topic Maps; International Organization for Standardization, Geneva, Switzerland, 2000.

[KR00] Kazi, Zunaid; Ravin, Yael: Who's Who? Identifying Concepts and Entities across Multiple Documents, Proceedings of the 33th Annual Hawaii International Conference on System Sciences, IEEE Computer Society Press, Los Alamitos, January 2000.

[LC00] Lai, Hsiangchu; Chu, Tsai-hsin: Knowledge Management: A Review of Theoretical Frameworks and Industrial Cases, Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, IEEE Computer Society Press, Los Alamitos, January 2000.

[Le01] Letson, Russell: Taxonomies put Content in Context, in: Transform Magazine, December 2001, Manhasset, NY, USA, 2001.

[Ma00] Malhotra, Yogesh: Knowledge Management and Virtual Organizations, Idea Group Publishing, Hershey, USA, 2000

[MM00] McCown, Steven H.; Milligan, Charles: A Navigation System for Personalized Databases "StarMap", Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, IEEE Computer Society Press, Los Alamitos, January 2000.

[Mo01] Moore, Graham: RDF and Topic Maps – An Exercise in Convergence; in: Proceedings of the XML Europe 2001, Berlin, 2001.

[Mu96] Murray, Philip C.: Information, knowledge, and document management technology; in: Knowledge Management Briefs, Vol. 1, Iss. 2, May/June 1996.

[NN01] Nasukawa, Tetsuya; Nagano, Tohru: Text analysis and knowledge mining system, in: IBM Systems Journal - Vol. 40, No. 4 "Knowledge Management", New York, USA , 2001.

[NT95] Nonaka, Ikujiro; Takeuchi, Hirotaka: The Knowledge-Creating Company – How Japanese Companies Create the Dynamics of Innovation; Oxford University Press, New York, Oxford, 1995.

[Pe99] Pepper, Steve: Navigating haystacks and discovering needles; in: Markup Languages: Theory and Practice, Vol. 1 No. 4, September, 1999.

[Pi81] Piaget, Jean: Intelligence and Affectivity – Their Relationship during Child Development, 1954, in: Brown, T. A.; Kaegi, C. E (Editors): Annual Reviews Monograph, Palo Alto, CA, 1981.

[PI69] Piaget, Jean; Inhelder, Bärbel: The Psychology of the Child, Basic Books, New York, 1969.

[Pi70] Piaget, Jean: Genetic Epistemology, Columbia University Press, New York, 1970.

[Po97] Polanyi, Michael: The Tacit Dimension, 1966; in: Prusak, Laurence: Knowledge in Organizations, Butterworth-Heinemann, Boston, 1997.

[PR99] Probst, G.; Raub, S.; Romhardt, K.: Wissen managen: Wie Unternehmen ihre wertvollste Ressource optimal nutzen; 3. Aufl., Gabler, Wiesbaden 1999.

[RP99] Rath, Dr. Hans Holger; Pepper, Steve: Topic Maps: Introduction and Allegro; in: Proceedings of the Markup Technologies 99, Philadelphia, USA, 1999.

[Sa98] Sander, Martin: Entwicklung eines prozeßorientierten Knowledge Management Systems im Intranet einer internationalen Großbank, Masters Thesis, University of Paderborn, Department of Business Computing 2, Paderborn, 1998.

[Sc02] Schubert, Lenhart K.: Can we derive general world knowledge from texts?, to appear at the Human Language Technology Conference (HLT 2002), San Diego, CA, March 24-27, 2002.

[SN02] Smolnik, Stefan; Nastansky, Ludwig: K-Discovery: Using Topic Maps to Identify Distributed Knowledge Structures in Groupware-based Organizational Memories, Proceedings of the 35th Annual Hawaii International Conference on System Sciences, IEEE Computer Society Press, Los Alamitos, January 2002.

[Sv97] Sveiby, Karl Erik: The new organizational wealth: Managing & measuring knowledge-based assets; Berrett-Koehler Publishers, San Francisco, CA, 1997.

[Wr01] Wrightson, Ann: Topic Maps and Knowledge Representation; White Paper, Ontopia AS, February 2001.