

Added Value of Sociofact Analysis for Business Agility

Uwe V. Riss¹, Johannes Magenheim², Wolfgang Reinhardt², Tobias Nelkner², Knut Hinkelmann³

¹ SAP Research

Vincenz-Priessnitz-Str. 1
76131 Karlsruhe, Germany
uwe.riss@sap.com

² University of Paderborn

Computer Science Education Group
Fuerstenallee 11
33102 Paderborn, Germany

{jsm,wolle,tobin}@uni-paderborn.de

³ FHNW University of Applied Sciences Northwestern Switzerland

School of Business
Riggenbachstrasse 16
4600 Olten, Switzerland
knut.hinkelmann@fhnw.ch

Abstract

The increasing agility of business requires an accelerated adaptation of organizations to continuously changing conditions. Individual and organizational learning are prominent means to achieve this. Hereby learning is always accompanied by the development of knowledge artifacts. For the entire of learning and artifact development the term knowledge maturing has been introduced recently, which focuses on these three manifestations of knowledge: cognifacts, sociofacts, and artifacts. In this paper we will focus on sociofacts as the subject-bound knowledge manifestation of social actions. Sociofacts are rooted in respective cognifacts play an independent role due to their binding to collective actions and subjects. These are particularly difficult to grasp but play a decisive role for the performance of organizations and the collaboration in there. The presented paper approaches the notion of sociofacts, discusses them on a theoretical level and establishes a first formal notation for sociofacts. We use the case of a merger between two companies to describe the advantages of sociofact analysis for such process. Some sociofact related problems during a merger are described and possible solutions are presented. We identify technical approaches for seizing sociofacts from tool-mediated social interaction and discuss open question for future research.

Introduction

The increased agility of business due to globalization requires an acceleration of the adaptation of organizations

to changing conditions. These requirements make new approaches to individual and organizational learning indispensable that allow employees to update and extend their knowledge at a pace that they cannot achieve by traditional human resource development any longer. However, such learning processes must always be accompanied by the development of knowledge artifacts that take a mediator role. From a technical point of view this mediator role has been accomplished by knowledge management systems while the actual learning process has been supported by e-learning applications. Meanwhile the integration of both approaches has attracted considerable interest and is denoted by the term *knowledge maturing* (Schmidt 2005).

Knowledge maturing encompasses individual and organizational capabilities as well as the development of so-called knowledge artifacts, that is, all forms of static or dynamic knowledge representations such as documents, ontologies etc. The knowledge maturing process comprises three manifestations of knowledge: cognifacts, sociofacts, and artifacts. While cognifacts as individual and artifacts as material manifestations of knowledge have been in the focus of scientific interests for quite a long time, investigations on the nature and possible management of sociofacts are still rather limited and have been considered more intensively only recently (Nelkner, Magenheim, Reinhardt 2009). Nevertheless they play a decisive role for collaboration and learning in organizations.

Talking about cognifacts, sociofacts, and artifacts we have to keep in mind that these three concepts are partially overlapping. Every sociofact is rooted in respective cognifacts; often sociofacts are closely related to corresponding codified rules, which are artifacts;

cognifacts and artifacts influence each other, that is, people learn from artifacts and make their knowledge manifest in other artifacts. However, there are also clear differences. The existence of a codified rule does not necessarily mean that people follow it while living rules are not always codified.

Mergers and Acquisition

Agility is the ability of enterprises to quickly adapt themselves to changes in an increasingly dynamic environment and to seize opportunities as they emerge. This agility relies on knowledge about products, processes, services, capabilities, internal structures etc. which have to be adapted in order to react on changes like mergers and acquisition, innovation, new regulation, etc.

Hereby sociofacts play an important role as knowledge in an enterprise is distributed among people interacting with each other when performing tasks and solving problems and where knowledge matures through people's interaction.

In the following we will mainly refer to the example of a merger or acquisition in order to illustrate the importance of sociofacts in business agility. Mergers and acquisitions are a special challenge. Two formerly independent organizations have to be integrated with dramatic changes for the interactions of the people. This means that existing sociofacts are migrated to new ones. Being conscious about the existing sociofacts would allow to establish new interactions and thus to influence the emergence of new sociofacts.

Mergers and acquisitions of companies are a strategy often found to improve financial performance. Economy of scale and scope, increased market share, achieving synergies or cross selling are often mentioned motives. Acquisition refers to a purchase of a firm by another where usually the buyer is the larger one. A merger, however, happens when two firms agree to go forward as a single new company. Besides the legal and financial changes mergers and acquisitions have dramatic consequences on the organization, the corporate culture and the communication.

Mergers and acquisitions mean reorganization; departments have to be merged, new teams have to be formed and business processes have to be re-engineered. On the management level, mergers and acquisitions destroy leadership continuity. According to (Krug and Shill 2008), target companies of an acquisition lose 21 percent of their executives each year for at least 10 years following an acquisition, more than double the turnover experienced in non-merged firms. Such turnover is partly to be expected; there can only be one CEO, CFO, et cetera at a time. If the businesses of the merging companies overlap, then such turnover is also to be expected on lower levels of the hierarchy. Reorganization means also to decide, who should work in which group.

Identification of sociofacts could deliver important basic data. Besides personal skills and competences, the role of a

person in communication and collaboration networks can be important decision criteria. But also the topics of sociofacts are important to identify in order to be aware of the main protagonists and to support maturing processes in essential knowledge areas. It can also mean to avoid the same work twice. The reorganization and employee turnovers require new paths of communication. Questions like *'Who is the appropriate person to talk to in order to solve a problem?'* arise. This requires the establishment of sociofacts, bringing together people from both companies being engaged in similar topics.

Theoretical Framework

In order to comprehend and support knowledge maturing within communities, organizations and companies we need a theoretical framework that offers us different perspectives on these processes and allows us to identify relevant categories.

Based on the theory of Symbolic Interactionism (Blumer 1969) the distinction between cognifacts and sociofacts as individual and social internal knowledge representations in contrast to artifacts as external knowledge representation has been introduced by (Nelkner, Magenheimer, Reinhardt 2009). We can describe them preliminarily in the following way: artifacts are the only kind of the three knowledge representations that possess a clear material manifestation. They are the reproducible physical or digital results of externalization processes, which occur as documents, digital media or written laws and can serve as an external memory (Keil-Slawik 1992). Software and Software tools may also be regarded as artifacts as they are coded with symbols according to syntactic rules. Cognifacts refer to individual capabilities of intelligent agents as parts of their knowledge. As such they are internal and therefore not directly observable. They include competencies, specific expertise, skills and so on. As explained by Newell (1982) in his seminal and influencing paper, intelligent agents can be described on the knowledge level. Finally, sociofacts describe all kinds of collective knowledge including culture, social structures such as families or social communities, norms but also shared understanding of concepts and theories about the world. They are essential for any kind of social activity such as collaboration or communication. Thus, they become manifest in specific social action and activities. On the other hand they are internal representations of mutual expectations and common understanding of individuals who might be members of a group or an organization. Due to limited resources we cannot explain all details of what we communicate and therefore require some knowledge about the way others will understand what we say. If we work together we must know what we can expect from others and what others expect from us without detailed description of all particularities. In this way sociofacts are the bases of our social intelligence and close to the concept of the *'Generalized other'* within the theory of Symbolic Interactionism.

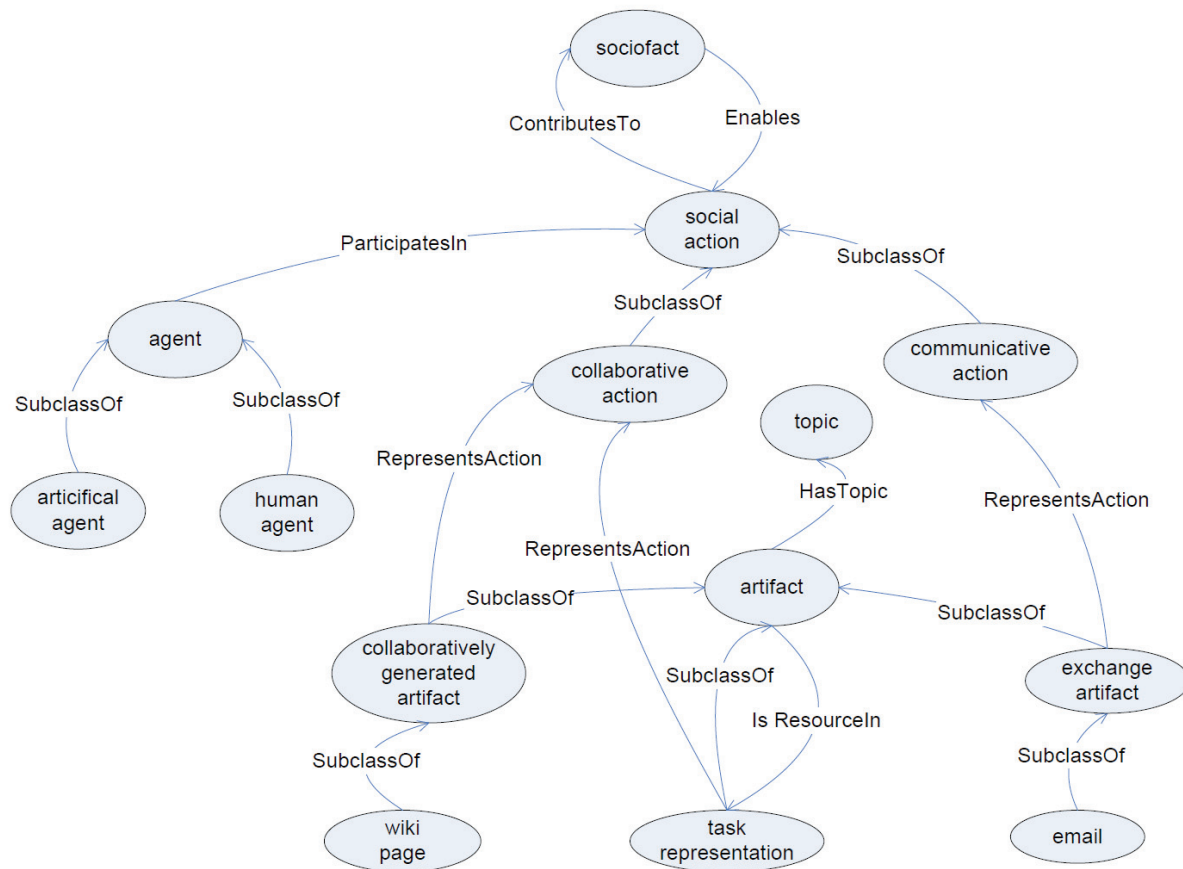


Figure 1: Graphical representation of relevant concepts and relations for a sociofact ontology

While artifacts are directly accessible – even by machines – the access to and description of cognifacts and sociofacts is much more complicated. However, if we understand them as capacities to perform specific actions (Newell 1982, Riss 2005, Riss *et al.* 2009), we realize that we can recognize them on the basis of their actualizations in individual and collaborative action, respectively. Newell (1982) explains that an observer treats an agent on the knowledge level, as he ascribes knowledge and goals to him. The observer knows the agent’s goals and knowledge alongside with his possible actions and his environment. For example, the individual capability of understanding Spanish becomes manifest in the action of writing a Spanish email while the knowledge about motors becomes manifest in an accomplished task that consists in building or repairing such a machine. But writing a book individually or repairing a bicycle by oneself are externalizations of an individual’s knowledge and therefore indicators for individual cognifacts. This type of an individual’s capacity to act is neither related to interpersonal communication nor to social interaction or collaboration. Contrary to this individual activity, sociofacts always include the capacity to act in a social context. Thus, social interaction and collaboration are

relevant indicators for sociofacts within a group, community, company or organization and can be regarded as the external observable part of existing sociofacts.

Following this concept means that repairing a car’s engine together in a team is the observable part of a sociofact. Handling over the appropriate tools and materials for the repair in the appropriate order even without oral communication indicates that the members of the repair team share a common knowledge about the technical concept of an engine and the required steps of its repairing. Furthermore, they also know about the written and even more important unwritten rules of cooperation in a successfully operating team. Similar examples for existing sociofacts can be identified by regarding surgery teams or working groups in a company. The sociofact-related capacity to act has a social dimension and always includes activities as their externalization in a social context. These activities are goal-oriented and of deliberate intention. For instance, delivering an email to a specific target group mostly aims at providing information to the group members and at initializing some reaction like responding to this mail or starting some other activities. Thus, a network of activities between involved persons may occur.

In order to refine the theoretical concept of a sociofacts we have to learn from this examples but can also refer to e.g. activity theory (Kaptelinin and Nardi 2006), that the duality of individual internal representation (capacity for social action) and its externalization (real action of an individual in a social context) is one of the fundamental characteristics of a sociofact. Talking about sociofacts we have to consider both levels of a sociofact as an entwined entity, where only the external level, the action performed, is observable.

Regarding the structure of sociofacts we find certain similarities to cognifacts, but sociofacts are even more complex. Both, cognifacts and sociofacts, possess persons as carriers and are further distinguished by a specific content or topic. In contrast to cognifacts, sociofacts are - in a process of abstraction - assigned to group activities or even described as a property of an organization. Statements like 'this company produces high-quality cars' or 'this department most successfully operates according to the company spirit' indicate the existence of related sociofacts. As a long term outcome of a merger it might be expected that positive attributes allocated to one of the companies may also emerge in the new common brand mark and failures of one of the companies in the past may fade away in the general public.

Generally sociofacts are more difficult to grasp than cognifacts since they concern groups and their coordinated activities. Nevertheless, we can identify individuals as the carriers of sociofacts inasmuch their actions contribute to organized social activities.

Sociofacts additionally do not only imply capabilities of topic-specific communication and related social interaction, but also effect the creation of associated artifacts. Therefore, as a result of that kind of social activity, e.g., delivering emails in a working group, we also generate artifacts, which allow us to analyze the topic of sociofact-related communication and the resulting interaction that is establishing a specific workflow in a department of a company. Thus, so-called Artefact-Actor-Networks (AANs; Reinhardt, Moi, Varlemann 2009) may indicate the existence of related sociofacts and can be used to describe them. We will discuss this later on.

Moreover, sociofacts do not only play an enabling but also a restrictive role. For instance, they exclude certain ways of communication and collaboration due to explicit and implicit rules of dealing with hierarchy or boundaries of departments. In many cases, persons who belong to a higher hierarchy level should not be contacted from persons of a lower level without being requested to do so. Communicating with employees from another department directly via email may be regarded as inappropriate but has to be organized by distinguished responsible 'communicators'. These could be unwritten rules in an organization that are followed by most of the employees. In this way sociofacts help to reduce the complexity within an organization and enable efficient action. They are also very important in the merger example: If there exists a different culture to deal with hierarchy in both companies

due to written or unwritten rules, it is crucial for a successful transformation of departments with similar functions to a unified common unit to overcome this barrier and resolve communication problems. In this respect, sociofacts are associated to formal structures of communication and interaction in a group, community or organization including communication barriers. However, we also have to consider, that interaction and communication is always related to content. Thus, topics and formal aspects of action are closely linked to each other and both are properties of a sociofact. Identifying those types of communication barriers and resolve them would be an important contribution to a successful merger and also could be regarded as knowledge maturing on the sociofact level.

Due to their partially hidden character, the identification of sociofacts requires that we observe the individual activities and objects with which they are concerned.

Before we come to the question how sociofacts can be identified, we first have to understand which role they play in daily life. This appears to be difficult since we are often not even aware of sociofacts in our life. For example, if I ask my neighbor to open the door it is clear that I mean that he opens the door so that I can go through although I do not explicitly mention this. As we know from research on linguistics, in particular speech acts (Searle 1969) this is far from obvious. However, we immediately realize if we cannot rely on sociofacts. For example, when we are in a foreign country and culture whose language and customs we do not know, we will rely on communication with hands and feet, what can lead to misunderstandings and disputes. The effect is that communication possibly breaks and we do not achieve our goals. Similar problems occur in everyday situations if experts with different backgrounds talk to each other and use terms in a different way. Consequently sociofacts become manifest in successful communication and collaboration, that is dependent on "mutual knowledge, mutual beliefs, and mutual assumptions" (Clark and Brennan 1991) but it is difficult to identify sociofacts in an automatic way.

This is a particular challenge for the identification of sociofacts. Here it is not sufficient that persons share some specific content. For instance, if two persons are interested in Semantic Web issues but have never spoken to each other, there is a certain probability that they understand Semantic Web in a different sense and have communication problems. In this case we would not say that they share a *common* understanding of 'Semantic Web'. Nevertheless, the fact that they are discussing on the topic 'Semantic Web' may generate a common ground after some time. The communication on this topic itself represents a sociofact, but we have to state a different degree of common understanding. In general the communication between or a common history of both parties is another precondition for a sociofact. This communication can even be indirect. For instance, if a researcher A has read most of researcher B's scientific publications and vice versa, they are usually able to

communicate even if they have never met in person. On the other hand, if persons communicate a lot but do not share a specific topic there is no sociofact to be identified.

Nevertheless, we have to consider different levels of shared common understanding of a sociofact. This degree of common understanding may also be an indicator for knowledge maturing on the sociofact-level.

Reconsidering our merger case we can identify a need of fostering the degree of shared understanding on relevant topics between employees in those departments which prospectively have to cooperate or will be unified in a new common organizational structure. Tools, which support people's co-operative reflection on semantics and their negotiations on meaning in a specific topic area, would be an essential technical contribution to knowledge maturing during a merger with regard to sociofacts (Braun, Kunzmann, Schmidt 2010)

The problem becomes even more complex if we go to rather widespread sociofacts as they appear in societal norms and culture. Here it is rather the common background such as family and education that determines the sociofact and not only a specific topic. Two persons with similar education already possess a certain basis for communication. Belonging to a specific societal subculture may influence a person's normative orientation and thereby its social activities. Furthermore, a type of societal behavior may be assigned to the whole sub-cultural group according to a set of social activities that is performed by a majority of the group members.

For the current purposes, to observe and foster knowledge maturing in organizations on a sociofact level, however, we will mainly concentrate on topic-specific sociofacts since these are more easily to handle by means of IT. For the IT-based identification of sociofacts we require sets of indicators that provide sufficient evidence for the existence of sociofacts. Since we have no direct access to sociofacts we always depend on indicators. The more indicators we find the higher the probability of the existence of a sociofact and the clearer its specification becomes.

In order to identify these indicators we firstly summarize what we have learned so far from the different examples and theoretical assumptions discussed above and we also try to refine and categorize these assumptions.

Thus, sociofacts are multidimensional entities, which:

- ...have a group of people as carrier when regarded from an abstract perspective. But zooming in and getting the more specific close-up view leads to the disclosure of different individuals who are interacting within the group as their members and who are the carriers of internal representations of sociofacts.
- ...have an internal representation in people's mind as a capacity to act in a social context. This demands a social dimension of the intended action.
- ...have internal representations of mutual expectations, common understanding and shared

values of individuals who are members of a group or e.g. an organization. This concept includes 'unwritten' normative orientations (e.g. you should always meet the expectations of your supervisor) and regulating norms for actions (e.g. don't communicate directly with a person from a higher level in the organizational hierarchy).

- ...are related to a target group and mostly actions; they are therefore goal-oriented additionally to the social dimension of their intended action.
- ...are related to topics and include a different degree of shared topics and common understanding of those topics.
- ...have an external representation, observable as social interaction and as activities of individuals within a group.
- ...imply a double duality: firstly a duality of internal representation of social interaction (capacity to act) and an externally observable manifestation of this interaction (performance of action); secondly a duality between the associated topics and the formal structure of actions.
- ... cannot be generated or extracted but emerge in social interaction and thus can be realized through personal involvement in the sociofact.
- ...actualize themselves in action.

Sociofacts become manifest in activities in different types and degrees of complexity. They may become manifest in only one single action between two participants (A writes an email to B). However, they can affect lots of similar activities over a certain period of time including several participants (email exchange between various participants $P_1 \dots P_n$ related to the same topic) or even may determine the way of communication of a core group using different media (email, intranet, microblogs, wiki, communicator chats...) with different participants (depending on the used media) and even changing topics over the time elapsed.

Complex sociofacts enable activities, which lead to observable processes like e.g. workflows and task sequencing. A single social activity of one or between two persons may be regarded as an update of a sociofact. Often repeated activities or even activities, which only rarely take place, can be represented by tasks or task patterns (Schmidt and Riss 2009). These are part of the external representation of existing sociofacts. As the social interaction within a group that demonstrates a sociofact is topic related and bound to a certain communication medium as a data carrier we can describe Artefact-Actor-Networks, which may be analyzed in order to identify sociofacts. The analysis of changes in an AAN along a timeline may help us to gain information about knowledge maturing on a sociofact level (e.g. with regard to a higher degree of common understanding of a topic within a group, changes in internal communication structures of the group, changes in external activities regarding organizational

hierarchy, changes in used vocabulary, coherence of produced artifacts).

This theoretical approach to the concept of a sociofact has similarities to Vygotsky's conception of mediation (Vygotsky 1978) and Engeström's activity theory (Engeström 1987) with its relevant mutually depended categories of subject, rules, community, mediating artifacts, division of labor and objects, which represent the outcomes of the activities in the dimensions of sense and meaning. In this paper there is no space for a comparison between the two approaches. We just have to consider, that the sociofact approach is not an adaption or extension of the activity theory but an approach that is based on the common roots of socio-psychological theories, that state the duality of internal (higher mental processes) and external representation of activities (social action). (Leontjev 1981).

Formal Approach to Sociofacts

We have characterized sociofacts as capabilities to perform social and collaborative actions and have provided a sociofact ontology (Table 1). However, we face the central problem that we have no direct access to a sociofact since it describes a kind of disposition or capacity that only becomes manifest in social action. From a phenomenological point of view sociofacts appear as patterns in the actions that they help to enable. Ontologically they are related to these social actions, which furthermore mediate relationships to groups of agents who partake in the action and to sets of weighted topics via the artifacts that are associated with the action. These topics describe what we have called the content of the sociofact. The concept of an agent includes human individuals as well as intelligent artificial agents.

However, this description is still too vague to come to a feasible description. We actually have to make use of the described structural properties of sociofacts. According to the ontological description our first guess for a sociofact s is to associate it to a vector (s_i) where the i -th component of the vector describes the weight of the (social) action a_i in the sociofact s . On the basis of this vector we will later determine a more comprehensible representation of a sociofact as a graph. However, our actual aim is not a description of sociofacts in terms of actions but in terms of persons and topics in order to make it comprehensible for a user. In the following we will explain how we can come to such a description.

The relation between an intelligent agent – in the following denoted as person - and topics is given by the respective actions that belong to components of the sociofact vector. Fortunately the structure of an action is much more transparent than that of a sociofact and we can approximately describe it as

$$(1) \mathbf{a} = \mathbf{a}(\mathbf{P}, (\mathbf{z}_i), \mathbf{t}_1, \mathbf{t}_2)$$

where \mathbf{P} is a graph of persons \mathbf{p}_i with weighted edges \mathbf{e}_{ij} , (\mathbf{z}_i) a vector that describes the weight of the respective topic for the sociofact, \mathbf{t}_1 the start time and \mathbf{t}_2 the end time of the action. The topics are associated with the action via the related artifacts. This representation of actions can be obtained from the analysis of artifacts that are related to actions. Let us explain this by the following example:

1. *Email*: An email is an artifact that represents a communicative action. It has a sender \mathbf{p}_0 and a set of recipients \mathbf{p}_i with $i=1, \dots, n$. In a simplified way the edges can be defined as connections between the sender and the recipient for which we assume $\mathbf{e}_{0i} = 1$ and 0 else. We could argue that also the introduction of weak (e.g., 0.1) relationships between the recipients makes sense. Topics (\mathbf{z}_i) might be extracted from the mail body. For sake of simplicity we can further assume that $\mathbf{t}_1 = \mathbf{t}_2$ is the time of sending the email.
2. *Task*: As well a task is an artifact that represents a collaborative action. It has a task owner \mathbf{p}_0 and a set of collaborators \mathbf{p}_i with $i=1, \dots, n$. The weights of the edges \mathbf{e}_{ij} might be determined by the respective contribution to the task but could also be assumed as $\mathbf{e}_{ij} = 1$ for all collaborators for sake of simplicity. Topics (\mathbf{z}_i) might be extracted from the task body and partially from the associated documents. The timestamps \mathbf{t}_1 and \mathbf{t}_2 describe the start and end time of the respective tasks.

Similar assumptions could be made for documents, wiki pages and other artifacts that describe social actions. The timestamps are important for the temporal analysis of sociofact since we have to assume that a sociofact changes in the course of time. This temporal dimension of the action goes beyond the ontological description.

The crucial question now is how we can determine sociofacts from such actions. As we explained above we have to analyze the similarity between different actions, where we define this similarity in a way that it indicates the existence of a sociofact. This will of course require further assumptions and approximations. Before we start this we will shortly explain how we proceed when we have identified such similarities that form a symmetric matrix $\mathbf{A} = (\mathbf{a}_{ij})$. A usual proceeding in such cases is a spectral analysis of this matrix (He *et al.* 2001). Here we interpret an eigenvector (s_i) as the associated vector of the sociofact s . Such spectral analysis makes sense since subspaces that do not possess any 'sociofact similarity' – as we are going to define it – lead to separate eigenspaces. For the time being we ignore the mentioned time aspect simply for sake of simplicity.

The associated vectors now open up an opportunity to represent the sociofact in terms of persons and topic. For example, for the people graph of the sociofact s we can set:

$$(2) \mathbf{e}_{ij} = \sum_k s_k \mathbf{e}_{ij}^{(k)}$$

where $\mathbf{e}_{ij}^{(k)}$ describes the weight of the relationship between person i and person j in the k -th action that belongs to the

component s_k . In a similar way we can also determine a topic vector for the sociofact:

$$(3) \mathbf{z}_i = \sum_k s_k \mathbf{z}_i^{(k)} + \mathbf{w} \sum_k \sum_l \sum_p \sum_q s_k \mathbf{z}_p^{(k)} s_l \mathbf{z}_q^{(l)} \delta_{pqk}$$

where $\mathbf{z}_i^{(k)}$ describes the weight of the i -th topic in the k -th action, while the additional term describes higher order effects resulting from a relation between the topics \mathbf{z}_p and \mathbf{z}_q that might be associated with another topic \mathbf{z}_i (e.g., a super topic). In this way we can take mediated relationships between actions and topics into account, for example, if an action is related to a topic via one or more subtopics. The function δ_{kij} describes this relation and is 1 if there is a relation between these three actions and 0 else. The factor \mathbf{w} describes the influence of these higher order terms and has to be adapted empirically.

The crucial step in this approach is the determination of the sociofact-related similarity measure between actions. Here we have to express which factors indicate the existence of a sociofact. The most obvious factor might be the simultaneous appearance of the same persons and topics in two actions. Thus we can say describe

$$(4) \mathbf{a}_{ij} = \left(\sum_k \mathbf{p}_k^{(i)} \mathbf{p}_k^{(j)} \right) \left(\sum_k \mathbf{z}_k^{(i)} \mathbf{z}_k^{(j)} + \mathbf{w}' \sum_k \sum_p \sum_q \mathbf{z}_p^{(i)} \mathbf{z}_q^{(j)} \delta_{pqk} \right)$$

where $\mathbf{p}_k^{(i)}$ stands for the weight of intelligent agents k in action \mathbf{a}_i (and analogously for j) and $\mathbf{z}_k^{(i)}$ for the weight of topic k in action \mathbf{a}_i (and analogously for j) while \mathbf{w}' is another weight factor and δ_{pqk} as in (3). The additional topic term with the weight \mathbf{w}' includes indirect correspondence of topics analogously to (3). Equation (4) expresses that two actions i and j are similar if there is an overlap of involved persons and, at the same time, an overlap of involved topics (including relations between topic such as subtopic or topic similarity relations). Here it is assumed that persons and topics are independent, which is not always the case. Moreover, equation (4) only concerns topic related aspects that only represent one type of possible sociofacts. The inclusion of other sociofact types requires a corresponding extension of the similarity measure (4).

It is to be remarked that the described formalization of sociofacts still leaves various details open and shall only indicate what a possible technical realization could look like. Especially the details of the sociofact oriented similarity measure will require some refinement to cover a broader spectrum of sociofacts. It might also be considered if spreading-activation techniques (Anderson and Pirolli 1984) could be applied in this case.

Temporal Dimension of Sociofacts

As described above, sociofacts describe a concept related to social activities of individuals within a group. Since the nature of these actions often changes in the course of time, sociofacts have obviously a temporal dimension that needs to be considered. It is relevant for

understanding, (technically) observing but also supporting sociofact maturing. Regarding the formal consideration of sociofacts we can take the temporality into account by increasing the weight of recent actions and decreasing the weight of older actions in the respective formulas.

This change of the considered actions results in a quasi-continuous development of the corresponding sociofacts. This allows us to monitor certain measures that have been introduced to foster sociofact maturing. For example, if we identify missing communication on a certain topic during the merger we can encourage an exchange and monitor the resulting influence on the communication behavior.

Technical Approaches to Sociofact Identification

As mentioned above, sociofacts emerge in the tool-mediated and physical interaction of people in communities. Identifying sociofacts thus means to follow interaction paths. Tools can be any mean that supports the goal-oriented collaboration between people and thus be of analogue or digital fashion. Tools that are most often used for communication and collaboration of people in an organization are face-to-face chats and formal meetings, email, documents and document management systems, audio and video conferencing, CRM and HR as well as task management systems, and more recently also organizational networking software, expert recommendation, people tagging and social bookmarking systems, wiki software and tools for sharing short status updates. Whereas it is relatively hard to access the interactions and topics of face-to-face interactions or phone calls, the social interactions using other media are easily accessible and their content can be analyzed automatically. The individual actions of workers that interact in knowledge, their interactions with fellow workers and the trails those activities leave in the physical and digital can be used for sociofact analysis.

In order to analyze organizational networks, it is necessary to model the possible connections between knowledge holders and objects of potential value for sociofact analysis. The theoretical model of Artefact-Actor-Networks (AANs; Reinhardt, Moi, Varlemann 2009) combines classic social network theory with so-called artifact networks that emerge from explicit and semantic connections between artifacts in a digital space and can be used as facilitating model for sociofact analysis. The formal description and the analysis of Artefact-Actor-Networks could be realized according to the formal description of sociofacts and as a specific conceptualization of this formal approach described above.

In AANs we abstract from a single person and consider their accounts with any IT tool as the actors. Doing this we can separate a user's actions in an email space (email actor) from its actions in a document repository (DMS actor) or a social bookmarking system (bookmarking actor). Moreover, this segregation of actors permits the analysis of individual and community performance on different layers

of communication and collaboration. The abstraction of different layers is also realized in the artifact domain, where we can distinguish mail artifacts from, documents and bookmark artifacts.

ART² relations on the other hand describe how artifacts are connected to each other. The Dublin Core metadata standard (DCMI 2010) and the SIOC project (SIOC 2010) provide valuable relations that can be used to describe

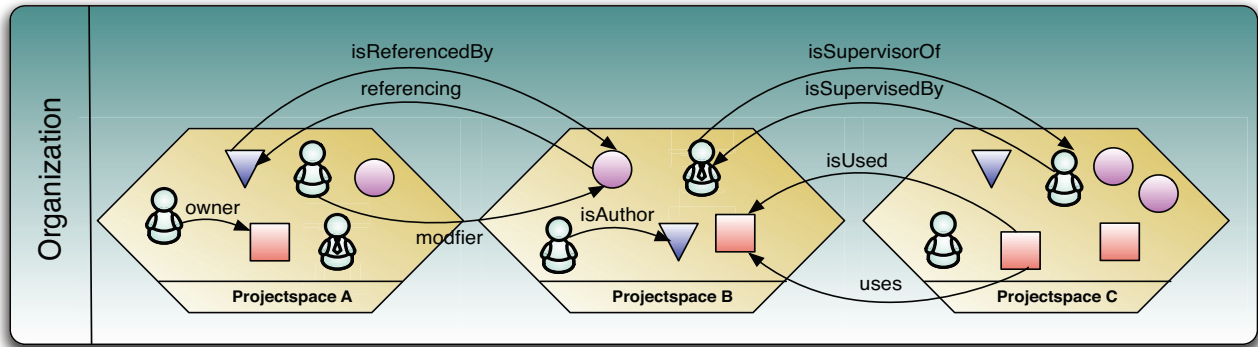


Figure 2: Example of AAN relations in an organization

As sociofacts emerge in the social interaction of people with artifacts, which are related to certain topics, the analysis of the contents of artifacts is the first step of AAN creation and a necessary precondition for the sociofact analysis. The artifact analysis can be accomplished using methods and tools from text mining and natural language processing (NLP; Allen 2003), e.g. Latent Semantic Analysis (Landauer and Dumais 1997), Probabilistic Topic Models (Steyvers and Griffiths 2007) or WordNet (Fellbaum 1998).

In order to not only store artifacts with their respective topics, the relations between artifacts and actors in the digital space need to be modeled and stored in a network.

Relations in Artefact-Actor-Networks

In AANs we discern three different types of relations between objects in the networks: first, relations that exist between artifacts (ART² relations), second, relations between actors (ACT²-relations) and finally, relations between artifacts and actors (AA relations). Each relation type can be used for specific analysis of the network and provides a different view on the collaborative interactions space within an organization. ACT²-relations exist between two actors in an AAN and describe the nature of personal relationships. They characterize simple connections, friendships or organizational hierarchies and can also be used to model the tools with which people are connected. ACT² relations could refer to the fact that someone is in the others address book (*isInAddressBookOf*) or on the buddy list of his organizational communicator (*isOnBuddyListOf*) but also that two people work in the same team (*isTeammateOf*) or that on person is the supervisor of another person (*isSupervisorOf*). ACT² relations can also model organizational hierarchies, its organizations and sub-organizations as well as reporting hierarchies when incorporating the Org ontology¹.

artifact relations: two artifacts can directly reference each other (*references*) or be a derivate of the other (*isVersionOf*). Especially when dealing with codified rules, one rule can replace (*replaces*) another or require some other artifact (*requires*). ART² relations can also be used to store series of interactions in tool-mediated communication when storing which artifact is a reply of which other (*replyOf*) or a forward to another (group of) actor(s) (*forwardOf*). Moreover, the version history of artifacts can be expressed using the ART² relations *nextVersion* and *previousVersion*.

AA relations connect actors with their respective artifacts. For each artifact there are several actors with a semantic relation to the artifact as it is always created. Modified, reused or recommended by different actors. Therefore, AA relations expose information about the connectivity between artifacts and actors. The main AA relations thus are *creatorOf*, *publisherOf*, *modifierOf*, *discussantOf*, and *forwarderOf*.

Besides the three relations mentioned earlier, artifacts and actors are always connected via common topics that are extracted from the artifacts and attributed to the enlisted actors (also to persons and groups on a higher level of aggregation).

In order to map actions with and within a tool to an according AAN, it is necessary to be able to access the respective tool via a rich interface. This interface has to provide detailed information about the artifacts and actors that are involved in the tool, e.g. the headers and the text content of an email, folder structures that may exist on the mail server and user accounts associated with an email. For an inclusion in AANs it is also important that the three types of relations between artifacts and actors described above can be extracted via the tool's interface. See (Reinhardt *et al.* 2010) for a detailed description of a Java-based prototypical framework that implements the model of Artefact-Actor-Networks in the context of social media community analysis.

¹ <http://www.epimorphics.com/public/vocabulary/org.html>

Technical solutions considering the temporal dimension

As described with the model of Artefact-Actor-Networks, many indicators for social relationships, artifact relationships and artifact-actor relationships can be mined from the certain databases. These are mainly independent from a temporal attribute. However, this temporal dimension can be helpful for three aspects:

1. Find unknown sociofacts in order to understand organizational strengths and weaknesses in unknown processes.
2. Processes are defined and it is investigated if the organizational reality is optimal regarding them.
3. The change of processes over time might be detectable.

For all points the dynamic change of the AAN over time needs to be evaluated. Therefore, traces of activities have to be identified and evaluated. Data mining techniques are needed, which would be quite similar to process mining techniques. Finding unknown sociofacts automatically would be very difficult. If this is a matter of interest, this has to be done interactively and could be supported by visualizations which allow for example to summarize activities regarding a certain topic in a specific period of time.

Recognizing processes might be easier, as patterns of sociofacts are available and could be detected. The big problem is in most cases the 'noise' in most of the available data as it is hard to differentiate between relevant and irrelevant artifacts. But the AAN provides the acting persons, the artifacts, and the receiver of artifacts, so that a pattern of a sociofact over time might be relatable if certain activities, their sequence and the expected entities occur. A software solution could relate activities to certain process steps and can mark irregularities if certain steps have not been finished in the expected way.

The third aspect of detecting changes in processes is probably the most hardest to be solved. A change in processes relates to (natural) sociofact maturing as processes (defined or not) have changed for whatever reason. In order to detect this, it is first important to find indicators, which enable us to show changes of sociofacts. These could be for example:

- change of artifacts in a certain period of time
- change of the normal use of software
- change of communication styles, e.g. more telephone, less email
- change of used vocabulary in the tag-cloud

Although such changes give us a hint to sociofact maturing, they do not address a specific question or problem but are symptoms of changing artifacts. Therefore, it is necessary to find valid indicators for each case. The identification process then can be supported by mining techniques, which specifically include the occurrence of these symptoms into the weighting function for the sociofact determination.

Topic Identification

As already mentioned above we can apply text-mining methods to identify the topics of sociofacts by analyzing the artifacts exchanged between the actors. During the analysis process, the text content of the artifacts is split into parts of speech, named entities are recognized and relevant keywords are extracted.

Important text mining tasks in this artifact analysis are concept/entity extraction, text classification/categorization, and text clustering.

Entity extraction can be performed using a thesaurus or a lexical database like WordNet (Fellbaum 1998). In WordNet words of English language are performed into sets of synonyms. Thus, if different terms of such a synonym set occur in the artifacts they can be mapped to one topic. Furthermore, hierarchies of extracted topics can be resolved and made accessible in appropriate data structures (Blei *et al.* 2004).

In contrast to entity extraction, where topics are extracted from the content of artifacts directly, text classification groups artifacts into one or more categories. The categories represent given topics of interest usually labeled with a noun or other short natural language expression (Sebastiani 2005a, 2005b). Automatic classification seeks to determine which category (or categories) a given artifact belongs to. Automated classification applies rules that assign documents to classes based on the terms occurring in the document. In machine learning-based text classification, the decision criterion of the text classifier is learned automatically from training data (Manning, Raghavan, Schütze 2008).

Text clustering is a special form of unsupervised classification. In contrast to text classification no predefined set of categories is given. Clustering may thus be seen as the task of identifying categories by grouping a given set of documents (Sebastiani 2005a).

Because of the temporal dimension of sociofacts, clustering seems to be less appropriate. If a predefined set of topics is given, we can apply text classification. With no predefined set of topics, term extraction can be applied.

Examples of Sociofact Analysis in the Case of a Merger

The two central aspects of sociofact analysis are communication networks and contents or patterns of this communication. In the following we want to consider some examples of this kind and discuss properties of such sociofacts that could support the agility of processes in a merger of two companies.

A typical example of a sociofact has concerned a topic that a group of people deals with, where the sociofact results from communication by the establishment of a common understanding. The first task in the analysis consists in the identification of such topic. Here a topic is not necessarily an individual keyword but rather a vector of weighted keywords. Since topics and communication

about them can appear in discussion forums or email exchange we have to analyze the respective media for the frequency of specific topics and their importance. Additionally to those methods mentioned in section *Technical approaches to sociofact identification* standard topic detection methods such as (Garcia *et al.* 2009) might be applied. After determining the most important topics we have to identify the persons, who have been involved in the respective communication, for example, by analyzing the sender and recipients of the respective emails or the contributors of a discussion thread and storing them in an appropriate Artefact-Actor-Network. The involved people are then related to this sociofact representation and the relations between the contributors are determined based on the exchange of information between them as far as the communication can be broken down to two-sided relation, e.g., as in email communication. In this way we obtain an AAN that is related to a certain topic. In the literature we also find approaches to identify author-topic models, e.g. (Dietz 2006).

For the merger case it is important to identify topics which are addressed by unconnected or only weakly connected user groups, mainly if these groups belong to different organizational parts of the merging organization. To this end we have to identify sociofacts for a given topic (e.g. product innovation), the topics of which are similar but for which the corresponding groups that are assigned to this topic are significantly separated. The analysis would consist in identifying such sociofacts and exploring their properties such as network structure, coherence or the application of organizational network analysis (ONA; Merrill *et al.* 2007). Following the sociofact analysis, contacts between the two groups would need to be established and the monitoring whether such initiatives lead to a sustainable integration of both sociofacts would need to be started. The latter point requires the consideration of the temporal development of sociofacts, which can help users to better understand the dynamics of groups dealing with specific topics.

For a merger to be truly successful, it is inevitable that the communication practices in the new company adapt to a common system of rules and regulations. Sociofact analysis can help fostering the identification of differing adherence to hierarchies, oppositional collaboration structures as well as the identification of knowledge transaction roles such as gatekeepers or boundary spanners.

Assuming we identified sociofacts in the product innovation departments of the two merger companies, the analysis of the sociofacts' properties can reveal many interesting social facts (cf. Durkheim 1982) that will be helpful in transforming two separate companies in a merged one. For example, if the analysis focuses on the adherence to hierarchies in communication processes, we would consider all tool-mediated communication between the employees in the innovation departments and take into consideration their position in the departments' hierarchy. Doing this, we will gain awareness about the real communication structure within the departments. In this

analysis, we could find out that the innovation department in company A is communicating in strict hierarchies (meaning employees only communicate with people on an hierarchical level above or under their own) whereas the respective department in company B is communicating across hierarchical boundaries. With this information we would know that we would need to place special focus on the establishment of explicit rules of inter-hierarchical communication in the merged innovation department.

In the same sociofacts we could explore the collaboration structures in collectively created documents. If for example both departments use wikis for documenting their innovation processes and innovation development strategies, the analysis of the collaboration and interactions within the wikis helps to understand the existing culture for collaboration within the two departments.

Task Patterns and their Relation to Sociofacts

The concept of task pattern has been introduced to grasp and describe agile aspects of task and process execution (Schmidt and Riss 2009) and represents a prototype of an artifact that allows for the inference of existing sociofacts so that we will consider it in some more detail. A task pattern helps users to execute specific types of tasks (e.g. 'Prepare Business Trip') by providing information sources, involved persons and possible sub-activities that the user can add to their individual task for further use. The user decides which offering she accepts and can add additional information source that she may add to the task pattern later. Task patterns provide a structure that makes them suitable for sociofact analysis. They include topics that are determined by the respective task, as well as person, who are involved in this task, and finally every task represents a social action or even a network of actions.

In this way the task pattern describes common knowledge how to perform a specific type of task and thus represents a sociofact. To be precise, the task pattern itself is an artifact while the sociofact is formed by its use and refers to the shared experience of its users. In particular it is the existence of the sociofact that enable users to contribute their own information sources to a task pattern since they can assume a common understanding of how their contribution is used within this kind of task. In this way task patterns contribute to the formation of sociofacts while at the same time they represent these sociofacts.

Discussion

We have presented artifacts, sociofacts and cognifacts as complementary concepts that describe manifestations of knowledge. When we consider them in more detail we realize, however, that a clear separation is not always possible. To which degree an action is individual (and therefore related to a cognifact) or part of a social action (and related to a sociofact) is generally difficult to decide. Even an action that a person executes alone might be based on social knowledge while in every social activity the

individual contributors set their own pattern. The border between the internal representation of sociofacts and cognifacts is fluid.

The same holds for the border between sociofacts and artifacts. The latter can be used in various ways. Even a document can be interpreted in different ways and the understanding of artifacts always depends on both cognifacts and sociofacts. Nevertheless, the distinction of standard sociofacts, cognifacts, and artifacts is clear enough to make effective use of them. In our pragmatic approach we exclude more philosophical considerations regarding the relationship between capacities to act, intended action and the finally performed action. We are also not discussing the question if a capacity to perform a specific action should be an abstract concept, which exist per se without being bound to a specific person.

It has to be noted that one might deny the existence of social knowledge as it appears in sociofacts and only accept individual knowledge as a capability of the individual's brain. One might accept that people can share knowledge but ignore the influence of the awareness of this sharing on the respective action. However, complex collaborative actions such as the production of a plane require more than independent individual competencies but a specific capacity of the group, team or organization. This is expressed by the fact that we define sociofacts as capacities to social actions even if their execution relies on individual capabilities.

Due to this reference to action we focus on the observable, external representations of sociofacts. These are characterized by actions of individuals associated with a certain topics. In order to identify sociofacts by means of IT systems, the relevant entities and their mutual relations are described above. The given formulas allow us to have a rather technical view on the concept of sociofacts. This formal concept provides a basic access to (semi-) automatic mining and support of sociofact maturing. Obviously, not all relations that can be described by this formula are sociofacts, which makes their identification much harder. Single actions can be described as well but are not necessarily related to a sociofact.

We are also aware of the fact that the introduced technical approach is not able to directly access the objectives, which are assigned to an identified sociofact. The identification of sociofacts always relates to an objective or question: Which goal do we pursue by considering sociofacts? What are the relevant aspects of the sociofacts in this respect? How can we handle them? Although we proposed a way to identify sociofacts, we realize that the handling of sociofacts is always context – dependent. Moreover, we also find various types of sociofacts such as communication patterns or focuses on specific topics.

The communication patterns give us a hint about the structures of communication within an organization. Furthermore, sociofacts are time-dependent. For instance, we can determine how the structure of collaboration changes and which topics are in the center of interest in the

course of time. Such analysis might help to realize a specific organizational strategy.

Another challenge is that end users need to be provided with appropriate visualizations of sociofacts and related information, in order to deal with the internal complexity of sociofacts and the high amount of underlying data.

We have also to consider the meaning of different media for the formal description of sociofacts. This also concerns the question how different media have to be valued in this description. For instance, a contribution in Twitter might be weighted in another way than a wiki contribution.

It becomes clear that the complexity of detecting and working with sociofacts in terms of all technically observable traces of actions is very high. For example in AANs we obtain millions of edges and nodes if we include the various media used by a high number of employees in an organization. Another reason for complexity results from tracing sociofacts and all their components over time.

Beside the technical view on sociofacts, the theoretical conceptualization needs to be refined, for example, considering intelligent agents or including refined similarity measures (e.g. to take communication patterns into account). The first issue leads to the question how sociofacts can be distinguished from artifacts and which requirements an artificial agent must fulfill to contribute to a sociofact.

Even if we assume that we are able to technically deal with the landscape of sociofacts within an organization, we also have to cope with privacy concerns. The current approach relies on data mining techniques based on as many as available externalized representations of sociofacts, as derived from email, wiki or documents. Users have to be motivated to participate in a collaborative bottom up approach, which provides a direct added value to them. This can be achieved for example by sociofact related recommendations. Alternatively it could be possible to use anonymized data or on the base of a case-related permission request.

Conclusion and Outlook

In this paper we have introduced sociofacts as complementary concept to artifacts and cognifacts. Sociofacts describe a topic related goal-oriented set of social actions within a community. We have concentrated on possibilities of ICT based sociofact mining and developed a set of expressions to formally describe relevant aspects of sociofacts. Based on the idea of AANs we described relevant properties of sociofacts, like communication structures. In order to observe sociofact maturing, which will be a key challenge in future, we have to monitor them over time. Based on an example of two organizations that are merging, we could show the relevance, usefulness and the added value for mining and monitoring sociofacts. Finally, we concluded our study with the discussion of open issues, to direct further research.

For the future it would be interesting to investigate how AI technologies can further contribute to sociofact management. Since the identification of sociofacts is a rather complex task – due to the rich network of actions, persons, and topics – AI methods seem to be rather promising to address these challenges.

Acknowledgement

This work was co-funded by the European Commission under the Information and Communication Technologies (ICT) theme of the 7th Framework Programme (FP7) within the Integrating Project MATURE (contract no. 216356).

References

- Andersen, J. R.; Pirolli, P. L. 1984. Spread of activation. *Journal of Experimental Psychology* 10(4):791-798.
- Allen, J.F. 2003. Natural language processing. In *Encyclopedia of Computer Science*. 1218-1222, Chichester, UK: John Wiley and Sons Ltd.
- Blei, D. M.; Griffiths, T. L.; Jordan, M. I.; & Tenenbaum, J. B. 2004. Hierarchical topic models and the nested Chinese restaurant process. In *Advances in Neural Information Processing Systems 16*. Cambridge, MA: MIT Press.
- Blumer, H. 1969. *Symbolic interactionism : perspective and method*. Englewood Cliffs, N.J.: Prentice-Hall.
- Braun, S.; Kunzmann, C.; Schmidt, A. 2010. People Tagging & Ontology Maturing: Towards Collaborative Competence Management. In D. Randall, P. Salembier (eds.) *From CSCW to Web 2.0: European Developments in Collaborative Design*. Berlin, New York: Springer.
- Clark, H.H.; Brennan, S.E. 1991. Grounding in communication. In *Perspectives on socially shared cognition*. American Psychological Association.
- DCMI 2010. Dublin Core Metadata Initiative <http://dublincore.org/>, retrieved on 2010-11-13.
- Dietz, L. 2006. Exploring Social Topic Networks with the Author-Topic Model. In *Proceedings of ESWC '06*. 54-60.
- Durkheim, E. 1982. *The Rules of the Sociological Method*, New York: Free Press.
- Engeström, Y. 1987. *Learning by Expanding*. Helsinki: Orienta-Konsultit Oy.
- Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, Mass.: MIT Press.
- García, R. D.; Berlea, A.; Scholl, P.; Böhnstedt, D.; Rensing, C.; Steinmetz, R. 2009. Improving Topic Exploration in the Blogosphere by Detecting Relevant Segments. In *Proceedings of I-KNOW'09*. 177-188.
- He, X.; Ding, C. H. Q.; Zha, H.; Simon, H. D. 2001. Automatic topic identification using webpage clustering. In *IEEE International Conference on Data Mining*, 195-202.
- Kaptelinin, V.; Nardi, B. 2006. *Acting with technology: activity theory and interaction design*. Cambridge, Mass.: MIT Press.
- Keil-Slawik, R. 1992. Artifacts in Software Design. In C. Floyd et al. (eds.) *Software Development and Reality Construction*. 168-188, Berlin/New York: Springer.
- Krug, J. A.; Shill, W. 2008. The big exit: executive churn in the wake of M&As. *Journal of Business Strategy* 29(4): 15-21.
- Landauer, T. K.; Dumais, S. T. 1997. A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- Leontiev, A. N. 1981. *Problems of the development of the mind*. Progress: Moscow.
- Manning, C.D.; Raghavan, P.; Schütze, H. 2008. Chapter 13 of *Introduction to Information Retrieval*, Cambridge University Press.
- Merrill, J.; Bakken, S.; Rockoff, M. Gebbie, K.; Carley, K. 2007. Description of a method to support public health information management: Organizational network analysis. *Journal of Biomedical Informatics* 40(4):422-428.
- Nelkner, T.; Magenheimer, J.; Reinhardt, W. 2009. PLME as a Cognitive Tool for Knowledge Achievement and Informal Learning. In A. Tatnall & A. Jones (Eds.): *WCCE 2009, IFIP AICT 302*, 378–387.
- Newell, A. 1982. The Knowledge Level. *Artificial Intelligence* 18(1):87.
- Reinhardt, W.; Moi, M.; Varlemann, T. 2009. Artefact-Actor-Networks as tie between social networks and artefact networks. In *Proceedings of the 5th International ICST Conference on Collaborative Computing 2009*.
- Reinhardt, W.; Varlemann, T.; Moi, M.; Wilke, A. 2010: Modeling, obtaining and storing data from social media tools with Artefact-Actor-Networks. In *Proceedings of the 18th Intl. Workshop on Personalization and Recommendation on the Web and Beyond*.
- Riss, U. V. 2005. Knowledge, Action, and Context: Impact on Knowledge Management. *LNCS 3782*: 598.
- Riss, U.V.; Witschel, H. F.; Brun, R.; Thönssen, B. 2009. What Is Organizational Knowledge Maturing and How Can It Be Assessed? In *Proceedings of I-KNOW 09*. 28-38.
- Schmidt, A. 2005. Knowledge Maturing and the Continuity of Context as a Unifying Concept for Knowledge Management and E-Learning. In *Proceedings of I-KNOW 05*.

Schmidt, B.; Riss, U. V. 2009. Task Patterns as Means to Experience Sharing. *LNCS 5686*: 353-362.

Searle, J. R. 1969. *Speech acts: an essay in the philosophy of language*. Cambridge, UK: Cambridge University Press.

Sebastiani, F. 2005a. Text categorization. In L.C. Rivero, J.H. Doorn and V.E. Ferragine (eds.), *The Encyclopedia of Database Technologies and Applications*, Hershey, N.J., Idea Group Publishing, 683-687

Sebastiani, F. 2005b. Text categorization. In A. Zanasi (ed.), *Text Mining and its Applications*, WIT Press, Southampton, UK, 109-129.

SIOC 2010. The semantically-interlinked online communities (SIOC) project. <http://sioc-project.org/>, retrieved on 2010-11-13.

Steinberger, M.; Griffiths, T. 2007. Probabilistic Topic Models. In *Latent Semantic Analysis: A Road to Meaning*. Laurence Erlbaum.

Vygotsky, L. S.: 1978, *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press,