Distributed Knowledge Management in the Absence of Shared Vocabularies (Extended Abstract)

Lutz Maicher
(University of Leipzig, Germany
maicher@informatik.uni-leipzig.de)

Thomas Schwotzer
(University of Technology, Berlin
thsc@cs.tu-berlin.de)

Abstract: Distributed Knowledge Management Systems (DKMS) are often faced to heterogeneous environments associated with the absence of shared vocabularies. DKMS realise Knowledge Flows (KF) between autonomous Knowledge Nodes (KN) as parts of social networks. Schwotzer models the KNs' individual policies for input relevance and output strategy as Knowledge Ports (KP). The used Topic Maps Theory is best suited for the semantic integration of distributed, heterogeneous knowledge. Unfortunately, current implementations base on pure naming approaches in connection with the use of shared vocabularies. Maicher's Subject Identity Measure Approach helps to use Topic Maps for the semantic integration of distributed, heterogeneous knowledge in the absence of these shared vocabularies. Our contribution is the liaison of the Knowledge Port Approach and the SIM Approach. This leads to DKMSs which significantly better deal with the absence of shared vocabularies.

Keywords: Distributed Knowledge Management, Ontology, Topic Maps, Semantic Integration, Knowledge Port Approach, Subject Identity Measure Approach (SIM)

Categories: C 2.4, H 1.2, H 4.2, H 4.3, I 2.4

1 Introduction: Distributed Knowledge Management, Topic Maps and the Absence of Shared Vocabularies

Often the output of classical knowledge management projects is a centralized Knowledge Management System[1] (KMS), which solely can be *accessed* by users in a distributed, decentralised way. These inherently centralized approaches tend to ignore that knowledge is usually distributed in and among complex knowledge-based organizations. Distributed KMSs (DKMS) try to overcome these limitations by splintering a centralized KMS into a network of cooperating Knowledge Nodes (KN). Introduced by Cuel [Cu03] a Knowledge Node is an abstraction of formal (e.g. division) or informal (e.g. community of practice) organizational units which are part of social networks. Knowledge Flows (KF) take place between these nodes.

Schwotzer [Sc04] showed how a DKMS can be modelled with Topic Maps. He introduced Knowledge Ports (KP) which realise Knowledge Flows between KNs. The behaviour of these ports is defined by the current context (interests, location etc.) and

[1] To avoid ambiguities all terminology concerning the Knowledge Port Approach and Topic Maps is capitalised.

the KN's individual policiy for input relevance and/or output strategy. All these parameters are completely described with the help of Topic Maps. Schwotzer's Knowledge Port Approach bases on the exchange of Topic Maps in distributed, heterogeneous environments.

Alternative approaches for the exchange of Topic Maps are proposed (and implemented), too. These are Ontopia's Topic Map Remote Access Protocol [PG04] and Barta's Federated Topic Map Approach [Ba04]. Both technologies are suited for DKMS and suffer from the same shortcomings in semantic integration scenarios discussed below.

For environments where all communication parameters of a KF (contexts and exchange policies) can be described by a shared vocabulary (ontology) the Topic Map theory provide a well defined and implemented fundament for the semantic integration of distributed, heterogeneous knowledge. The underlying theory is called "One Topic for One Subject". Today, it is mainly exploited in Enterprise Information Integration (EII) scenarios, which partially have considerable intersections to challenges in the field of DKMS [Ba04], [Ma04a], [PG04]. This article transfers insights from EII to the field of DKMS.

Problems occur in environments where shared vocabularies are not available, i.e. caused by management problems or not acceptable expenditures. Additionally, the flexibility and newness of knowledge based processes often cause the absence of shared vocabularies. Maicher introduced the Subject Identity Measure Approach (SIM) to allow a Topic Map based semantic integration in the absence of shared vocabularies [MW04, Ma04b]. Though, the SIM Approach completely relies on the Topic Map Theory.

Our contribution is the liaison of the KP Approach and the SIM Approach. Proven by first empirical results, the yielded DKMSs significantly better deal with the absence of shared vocabularies.

In detail, this article makes the following contributions:

- discussion of the power of the Topic Map Theory („One Topic for one Subject") for DKMS and its current limitations in the absence of shared vocabularies,
- discussion of the SIM Approach for the integration of distributed, heterogeneous knowledge in the absence of shared vocabularies and
- the liaison of the KP and SIM Approaches to overcome the discussed limitation in the absence of shared vocabularies.

This extended abstract gives a detailed overview of the full paper.

2 DKMS, Knowledge Flows and the Knowledge Port Approach

Cuel introduced the idea of modelling DKM with the help of Knowledge Nodes [Cu03]. Each KN represents the abstraction of a formal or an informal organisational unit. Individuals, groups (communities of interests) and (virtual) enterprises are represented by KNs which act as *autonomous* entities in dynamic, social networks. Between these KNs several communication channels realise Knowledge Flows. Schwotzer proposes the Knowledge Port Approach for the implementation of these KFs in DMKS [Sc04] (see Figure 1).

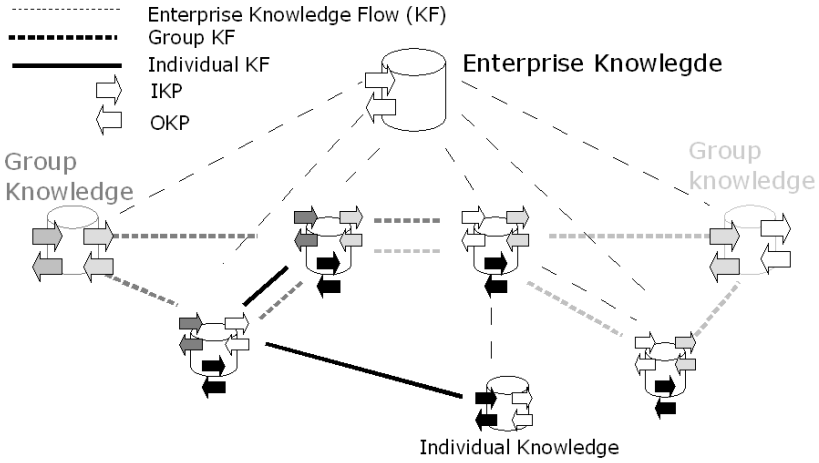

Figure 1 Knowledge Flows in DKMS

The approach bases on Knowledge Ports. These ports describe the knowledge input and output policies for each Knowledge Node context dependently. It is governed by the insight that social systems reduce complexity of input streams from environment based on relevance filters and of output streams based on strategy policies [Wi02, SH03]. Both are described inside the KPs with the help of Topic Maps. A KF between two KPs will be realised if their knowledge exchange policies matches. For further detail see ch. .

3 Topic Map Theory, Topic Map Exchange, and the SIM Approach

The KP Approach exploits the main Topic Map Theory which is called "One Topic for one Subject". In order to understand this theory, the notions of Topic, Subject and their relationship have to be explained. A Topic is "a symbol used within a topic map to represent some subject, about which the creator of the topic map wishes to make statements" [TMDM]. A Subject is "anything whatsoever, regardless of whether it exists or has any other specific characteristics, about which anything whatsoever may be asserted by any means whatsoever" [TMDM]. Shortly, a Topic describes a Subject (anything on which the Topic Map Author chooses to discourse) from the perception of the current Topic Map. This implies that each Topic has to declare its Subject. The Subject is the identity of a Topic on which semantic integration takes place.

While declaring Subjects, important philosophical questions arise: What is identifiable? What constitutes the boundaries of a thing in respect to its identity? Can identity evolve in time? Is identity situational or relative? How do properties of a thing have to change to alter its identity? What about versions and copies? These questions [discussed in detail in Ke78, Ke03] show the limits of pure naming approaches because they hardly handle indefiniteness, openness and ambiguity.

According to the Topic Map Data Model (TMDM) two means for declaring Subjects are implemented. Both are more or less pure naming approaches:

- The *Subject Locator* is used whenever the Subject of the Topic *is* an addressable information resource. In this case, the URI of this resource is used as a Subject Locator. The URI names the Subject.
- Because Subjects can be anything (not only addressable resources) a Topic can declare its Subject with the help of a *Subject Indicator*, too. A Subject Indicator is an information resource which *describes* the Subject. The URI (which names the Subject) of this information resource is called *Subject Identifier*.

To obtain "One Topic for one Subject", two Topics having the same Subjects (a pair of identical Subject Identifiers or Subject Locators) have to be merged. These rules work well if all authors of Topic Maps have made agreements for a shared vocabulary for Subject naming. These agreements are called *Published Subject Indicators* (PSI) [Oasis]. PSIs are published (but not necessarily public) descriptions of Subjects which should be reused by as much Topic Map Authors as possible to obtain a broad interoperability of Topic Maps. Examples in the literature which discuss the merging of distributed Topic Maps (or Topic Maps and RDF documents) exclusively use PSIs [see CPV03, Gr02, Sc04, Ba04]. This is due to the absence of solutions for not-shared vocabularies.

However, in heterogeneous, distributed environments with a high autonomy, the mechanism of PSIs has its shortcomings. PSIs are only used if they are visible to the regarding Topic Map Authors. Additionally, PSIs are faced with the philosophical problems of naming approaches discussed above.

4 The SIM Approach and the Integration of Distributed, Heterogeneous Knowledge Based on Topic Maps

In contrast to the naming approach discussed above Maicher proposed the SIM Approach which follows up a description approach [MW04], [Ma04b]. "Merging beyond the minimal rules [defined in the TMDM] is freely allowed. Most commonly, this will be done by inferring the subject of the topics from their characteristics." [TMDM]. The SIM bases on the assumption that a Subject is indirectly determined by the content of its Topic (and the surrounding of that Topic). In contrast to naming approaches the SIM doesn't name or stringently delimit a Subject. Solely it decides whether two Subjects might be treated as identical in the current context. The calculated level of this "identity" supports humans or machines to decide, whether the correspondent Topics should be merged.

Summarising, *the SIM describes how closely related the Subjects of two Topics are*. If the *SIM* is *1* the regarding Topics definitely represent the same Subject (according to the rules defined in the TMDM). If *0*, the regarding Topics definitely represent different Subjects. All values between *0* and *1* support the decision whether two Topics represent the same Subject.

In the full paper, this chapter additionally discusses the usage of the SIM Approach for integration scenarios in distributed, heterogeneous environments in the absence of shared vocabularies. It describes how the concept of Topic Map Views [see TMRM] can be innovatively used for such integration purposes.

5 The liaison of the Knowledge Port Approach and the SIM Approach for DKMS in the absence of shared vocabularies

Figure 2 sketches a Knowledge Port which is described in full detail in [Sc04]. A KP describes a Knowledge Node's policy of input relevance and/or output strategy for participating in Knowledge Flows. So far all communication parameters (context, partners, objectives and attributes) are defined by PSIs with the help of a Knowledge Exchange Protocol (KEP) [SH03]. While this might be inevitable for the description of a location or a defined person, or for other security related reasons, for the description of the communication objective (the subject of a KF) a PSI based approach often leads to avoidable empty matches. In fact the autonomy of the KNs forces the abandonment of shared vocabularies.

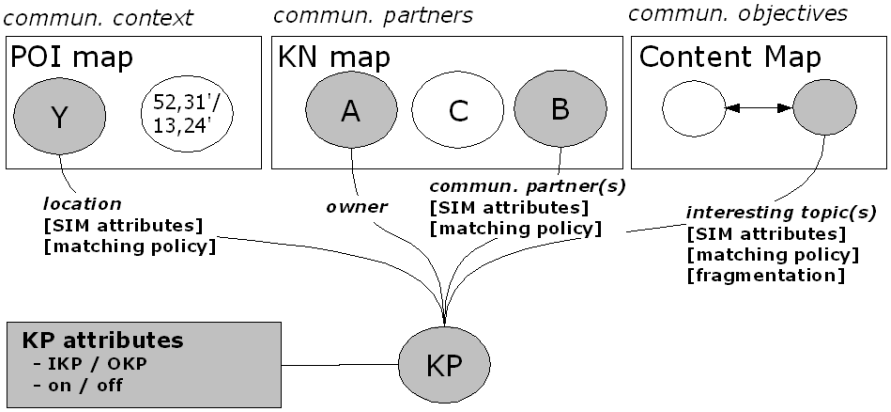

Figure 2 Parameters of a Knowledge Port (extended KEP)

As an extension of the KEP we propose the usage of the SIM Approach for all parameters which are not stringently fixed. Shortly, for each fuzzy parameter the Outgoing Knowledge Port (OKP) submits a (requested) Topic Map Fragment (the environment around the according Topic) which describes the current parameter. The Ingoing Knowledge Port (IKP) receives these fragments and matches it against its own Topic Maps. If the SIM of a Topic Pair exceeds a given threshold, the KF will be ruled by these parameters. In detail, each parameter has the following attributes:

- *SIM attributes.* These attributes define how the SIM is applied, i.e. the parameters of the requested Topic Map Fragment of the opposite KP (single Topic, Topic Map Fragment with given fragmentation depth, normalisation needed) and other parameters for the SIM algorithm.
- *Matching policy.* This policy defines when two Topics from opposite KPs have to be treated as having the same Subject. Additionally, this policy describes the effects to the KF if Subject equality is detected.
- *Fragmentation policy.* This policy defines the amount of knowledge which should be exchanged around the according Topics (especially in the Content Mao) if a KF is instantiated. It should be described with the help of the Topic Map Query Language (TMQL).

This method is similarly usable for other techniques of Topic Map exchange [PG04, Ba04]. Of course, instead of the KEP the according exchange protocols have to be extended.

In the full paper, the empirical results introduced in ch. will be discussed in respect to our proposed extension of the KEP. These results show that for our testbed the number of KFs which will not be instantiated caused by false empty matches can be significantly decreased (near to 0).

6 Summary and Outlook

We showed the advantages of the Topic Map Theory for the modelling of DKMS. Additionally, we discussed its limitations in the absence of shared vocabularies.

Derived from these limitations we introduced the SIM Approach, which abandons from pure naming approaches. Instead a description approach is used to decide whether two Topics might describe the same Subject. In liaison with the Knowledge Port Approach, the SIM helps to implement DKMS for distributed, heterogeneous environments in absence of shared vocabularies. Our empirical results show the positive effect of that liaison. The number of KFs which will not be instantiated caused by false empty matches significantly decrease.

While the first empirical results are encouraging, the proposed liaison of both approaches has to be tested in real life applications. We foresee that the SIM Approach has to be customised for the varying contexts. It is crucial, that it *always* rest on the Topic Map Theory which is the fundament of the semantic integration.

As a long-term vision, the resulting DKMS may help to realise the vision of a cognitive web as the human centric layer of the Semantic Web [Th02]. Additionally, the concept of KFs can be adapted to EII scenarios.

References

[Ba04] Barta, Robert: Virtual and Federated Topic Maps. In: *Proccedings of XML Europe 2004*, Amsterdam (2004).

[CPV03] Ciancarini, P.; Pirruccio, M.; Vitali, F. et al.: Metadata on the Web. On the integration of RDF and Topic Maps. In: *Proceedings of "Extreme Markup Languages 2003"*, Montreal, (2003).

[Cu03] Cuel, Roberta: A New Methodology for Distributed Knowledge Management Analysis. In: *Proceedings of I-KNOW '03,* Graz, (2003), pp. 531-537.

[Gr02] Grønmo, G. O.: Automagic Topic Maps. Available at: http://www.ontopia.net/topicmaps/materials/automagic.html

[Ke78] Kent, W.: Data and reality. Basic Assumptions in Data Processing Reconsidered. North-Holland Publishing, Amsterdam, New York, Oxford, (1978).

[Ke03] Kent, W.: The unsolvable identity problem. In: *Proceedings of "Extreme Markup Languages 2003"*, Montreal, (2003).

[Ma04a] Maicher, Lutz: Enterprise Information Integration with Topic Maps. Submitted to *Doctoral Consortium "Wirtschaftsinformatik"*, Bamberg (2005). Available at: http://www.informatik.uni-leipzig.de/~maicher/forschung3.html

[Ma04b] Maicher, Lutz: "Subject Identification in Topic Maps in Theory and Practice". In: Tolksdorf, Eckstein (eds.): *Berliner XML-Tage 2004*, Berlin (2004).

[MW04] Maicher, Lutz; Witschel, Hans Friedrich: Merging of Distributed Topic Maps based on the Subject Identity Measure (SIM). In: *Proceedings of Leipziger Informatiktage*, Leipzig (2004).

[Oasis] OASIS: Published Subjects. Introduction and Basic Requirements. Available at: http://www.oasis-open.org/committees/download.php/3050/

[PG04] Pepper, Steve; Garshol, Lars Marius: Seamless Knowledge – Spontaneous Knowledge Federation using TMRAP. Presentation at: *XML Europe 2004*, Amsterdam (2004). Available at: http://www.ontopia.net/topicmaps/learn_more.html

[SH03] Schulz, S.; Herrmann, K.; Kalcklösch, R.; Schwotzer, T.: Towards Trust-based Knowledge Management in Mobile Communities. In: *Proceedings of AAAI Spring Symposium on Agent-Mediated Knowledge Management*; Stanford, (2003).

[Sc04] Schwotzer, Thomas: Modelling Distributed Knowledge Management Systems with Topic Maps. In: *Proceedings of I-Know'04,* Graz (2004), pp. 53-60.

[Th02] Thompson, Bryan: The Cognitive Web. Presentation to the Semantic Web Interest Group. Available at: http://www.cognitiveweb.org/publications/

[TMDM] ISO/IEC JTC 1/SC 34: ISO/IEC 13250. Topic Maps – Part 2: Data Model. Latest version available at: http://www.isotopicmaps.org/sam/

[TMRM] ISO/IEC JTC 1/SC34: Topic Maps – Reference Model. Editor's Draft, Revision 3.1. 01.12.2003. Available at: http://www.isotopicmaps.org/TMRM/TMRM-latest-clean.html

[Wi02] H. Willke: Systemtheorie I: Grundlagen. 6th revised edition, Lucius & Lucius, Stuttgart (2000).