The Praxis of Social Knowledge Federation (Draft)

Arnim Bleier, Patrick Jähnichen, Uta Schulze and Lutz Maicher[1],

[1] University of Leipzig, Faculty of Mathematics and Computer Science, Johannisgasse 26, 04103 Leipzig, Germany
{Bleier, Jaehnichen, Uta.Schulze, Maicher}@informatik.uni-leipzig.de

Abstract. There are currently two streams that dominate the research on knowledge federation: The one is the trend towards Linked Data, leading to fine-grained structuring of information that is machine readable; The other is the reuse and co-creation of information that spreads the burden of its creation to the public and enables the availability of large knowledge corpora. In this contribution we outline the design principles and architecture of a prototype platform harnessing the praxis of user behavior to tackle issues of signal noise ratio as well as corrupting bits of information slashing the machine readability of such distributed generated contend.

Keywords: Knowledge Federation, Topic Maps, Maiana

1 Introduction

Most of research on the future of knowledge management tends to assume, that a global leap in technological infrastructure is made. A particular example is the semantic web as envisioned by Berners-Lee (2001), which resonated strongly in academia but has failed to meet the expectations in practice, so far. This semantic web differs from the current version of the web by characteristics such as: strong separation of content and layout, structuring of information and shared vocabularies for the markup as well as machine readability. The motive for the strong response from academia to the Semantic Web (presumably) is that the rolled out vision encourages hope for prototypes that unlock new applications and business models as well as pay off research costs in terms of return on invest. However, that hasn't happened as anticipated, so far. One reason may be, that the assumption of a discrete and global technological leap is unrealistic. Another reason may be the unsolved problem relating to the signal noise ratio in machine-readable content.
The goal of this workshop paper is twofold: First, to outline an approach towards a general praxis of knowledge federation that supports the creation of structured information, as well as it's decentralized verification; Second, to apply this approach to a social knowledge federation platform that is currently under development by the authors.

2 Social Knowledge Federation

2.1 The Social Praxis

Knowledge federation as used in the context of this paper is best characterized by the first part of Karabeg's twofold definition as verb and as a noun: "As an activity, knowledge federation means joining together multiple individual knowledge artifacts under a single identity. This may take any form, ranging from a simple subject-centric organization of those artifacts by using a topic map or a dialog map […], to creating a new artifact from the fragments of existing ones [...], to uniting the individual artifacts under a high-level view ..." [1]. In the praxis of using our platform, individuals create subject centric topic maps and merge them (together with their own annotations) under a new merged topic map, a federated organization of information is then the result of an iterative social merging process.

One characteristic of this social praxis is that actions on the level of the individual are the result as well as the shaping force behind collective structures. In terms of social knowledge creation, it can be observed, that the behavior of an individual user creates information artifacts that in the next step enable their collective reuse and refinement. In other words, the social structures that knowledge artifacts are, are increasingly backed up and encoded by electronic web applications that then allow new social spaces.
Guenther [2] identifies three characteristics determining these new social spaces: The first is content, referring to the information artifacts themselves (i.e. in topic maps and their constructs). The second is the code, incorporating the architecture and enabling the usage in the first place. The third are metadata, referring to the information about the information like usage and provenience.

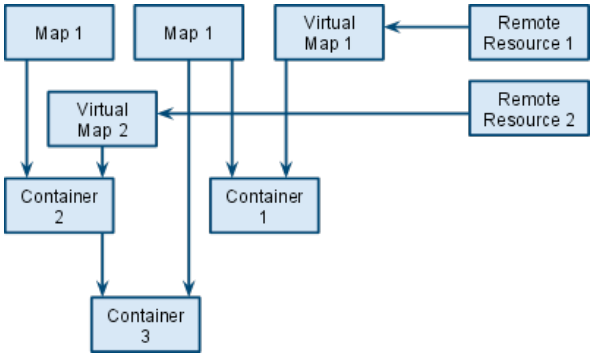


Fig. 1. A tree depiction of the merging of Topic Maps, as well as remote resources to federated virtual maps.

2.2 The User

As we have discussed, social web applications enable the user to rate and relate content. This happens in user communities, i.e. in overlapping, as well as distinct groups, that share an interest in a domain, as well as goals and expectations on how to use the application [3]. This is resulting in different levels of segmentation, ranging from commonly shared and agreed knowledge to highly specialized domains and differing points of view. While some of these groups may be interested in government spending, others may be interested in issues of global warming; and even within a group interested in global warming, different aspects of data may be seen. To support such a kind of knowledge creation and federation, three types of user engagement have to be enabled: the creation of content, the merging of multiple existing content resources to a new one and the annotation of existing content resources. As a result, the users should be enabled to individually and iteratively merge and annotate available information resources. However, to make such a high level federation in

practical scenarios work, we first have to consider some aspects of content generated in a distributed fashion.

2.3 The Signal Noise Ratio

In the traditional media landscape, professional authors and journalists created information resources, commented on them and connected them. The advent of electronic media empowered almost every user to also create and publish content. However, this development enables not only a long tail of highly differentialized niche domains, but also raises questions on how the awareness of crowds is (to be) directed. The resulting problem refers to the signal noise ratio.
Consequently, in the case of community-generated content, it is a requirement to empower the user to distinguish between relevant (for him) and useless content. Let's pose a question first: Is the relevance problem of information not already solved by the linking mechanism of the web? The answer is no, since this web linking mechanism only gives answers to how much attention a particular resource gets, not to how useful it is. Moreover, the shift from low structured and document centric content to semantic structured content requires an even more rigorous filtering criterion than the usefulness, since the information in one document becomes apparent in all documents it is federated with. As an example imagine a scenario in which a "malformed" map is frequently referenced. This malformation leads (for example) to the merging of all the topics in the maps it is merged into. Clearly no one would like to federate this map, yet it is highly linked. Consequently a mechanism is needed, that not only indicates the potential usefulness of a information artifact, but also prevents highly federated resources from being corrupted by a single document.
Keeping the metadata representing the users' assumptions on how useful and trustworthy an information artifact is, and the later utilization of these individual decisions can be used to share the burden of the decision whether to trust a particular resource. Consequently, we intent to apply crowd-sourcing as a mean to harness the individual decisions of the users of a map in Maiana to extrapolate on the relevance and integrity of the information contained. This democratizes the awareness guiding of crowds - useful information bubbles up, the not so useful one bubbles down - and allows the detection of corrupted information artifacts as early as possible.

3 Maiana prototype

3.1 The Architecture

The information artifacts in our platform are conformant to the ISO standard 13250 Topic Maps Data Model, short TMDM [4]. The TMDM is defined by seven constructs and their respective merging rules. These constructs namely are: topic map, association, role, topic, occurrence, name and variant. The merging rules are defined by conditions of equality, hence identity of two constructs, and result, if met, in the merging of both. However, two modes of merging have to be distinguished in the praxis of using Maiana; hard merging, and virtual merging.
The hard merging takes place within a particular map and ensures the internal consistency upon CRUD operations on the constructs of a single map and is carried out by the topic map engine itself, whereas the virtual merging always takes place between whole maps and is triggered by the user's action of merging information located in distinct maps. The technique behind this type of merging is, that instead of merging the actual constructs, virtual constructs are created, that play the role of proxy objects. Those contain a list of references to the actual constructs in the merged maps. As an example, imagine two topics located in separate maps, but both representing the city of Dubrovnik. In case of a virtual merging of two maps, a virtual topic proxy is created, exhibiting the union of properties known about Dubrovnik, but internally handling only references to them. This technique has the benefit of keeping the original information unchanged while enabling the tracking of the provenience of information in the merged virtual maps.

3.2 The Building Blocks of Maiana

MaJorToM: The Merging Topic Maps Engine project (MaJorToM) led by Sven Krosse has the goal to develop a lightweight, merging and flexible Topic Maps engine.[1] The engine provides the persistence layer for Maiana and exposes its functionality in an extended version of Topic Maps API, short TMAPI 2.0 [5].
Hatana: Hatana is the virtual merging engine enabling our platform to dynamically merge topic maps without actually editing any of them. The result(‘s) of the merging process are virtual topic maps called containers. This dedicated merging engine works on the TMAPI 2.0 and exposes the TMAPI 2.0, resulting in a virtual Topic Map that is available for further reuse by components building on the TMAPI 2.0, including merging further merging. An additional useful characteristic of this technique of virtual merging is the track keeping of the provenience of information (Schulze, 2010).
TMQL4J: TMQL4J (Bock, 2010) implements the topic map query language, short TMQL, providing a simple and straight forward way in Maiana to access and query for constructs and topic information organized in topic maps.[2] The TMQL4J query engine is based on TMAPI and thus can also be used on a variety of topic map engines. Besides query optimization and support for the current draft of TMQL [6], TMQL update queries are experimentally enabled and supported in the Maiana

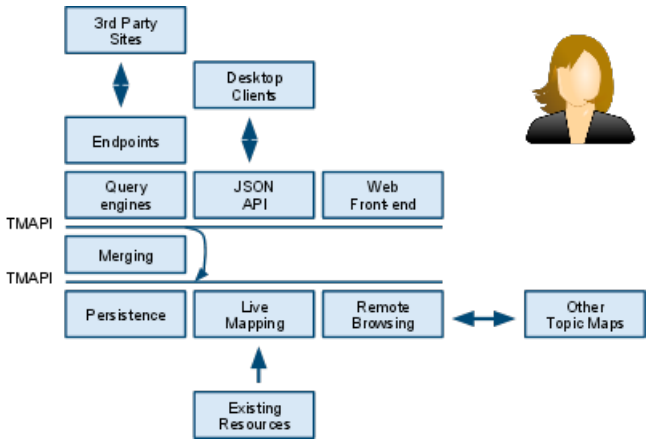
[1] http://code.google.com/p/majortom/
[2] http://code.google.com/p/tmql/

platform.

Nikunau: Nikunau is an implementation of the openRDF Sesame Sail API enabling Linked Open Data Consumers to be integrated with the platform.[3] The Linked Open Data service enables the whole spectrum of Linked Data community driven applications by the means of SPARQL queries as well as RDF and N3 export.

TMCL-Validator: The Topic Maps Constraint Language [6] is a language for the specification of constraints and schemas for topic maps. The TMCL-Validator validates a given topic map against a schema compatible to the current draft.[4]

(J)RTM: JRuby Topic Maps provides the glue between the different components in Maiana as well as the Ruby style language bindings used for the development of the Web Front-end.[5]



[3] http://code.google.com/p/sesametm/wiki/Sesame_Sail_Tmapi

[4] .http://code.google.com/p/tmcl-validator/

[5] http://docs.topicmapslab.de/rtm/

Fig. 2. Depiction of the layers and components used in Maiana.

4 Conclusion and Discussion

In this workshop paper we addressed a prototype platform for social knowledge federation.[6] The contribution of our approach is the combination of subject centric design of a data model based on topic maps with the social practice of gathering and harnessing metadata on their federation. On the topic maps side, Maiana acts as a stable repository for storage and reuse of the information in it, whereas on the social practice side, using Maiana adds value in terms of enabling crowd sourcing to merge meaningful maps but sort out malicious ones as early as possible. Only the synthesis of both properties allows the creation of highly federated views on content created by crowds.

References

1. Karabeg, D.: Knowledge Federation Patterns. In: Karabeg, D.; Park, J.: First International Workshop on Knowledge Federation. CEUR-WS Vol. 552 (2008)
2. Guenther, T.; Schmidt, J.: Wissenstypen im „Web 2.0“ — eine wissenssoziologische Deutung von Prodnutzung im Internet. In: Willems, H.: Weltweite Welten (2008)
3. Schmidt, J.; Pellegrini, T.: Das Social Semantic Web aus kommunikationssoziologischer Perspektive. In: Blumauer, A.; Pellegrini, T.: Social Semantic Web. Springer Berlin Heidelberg (2009)
4. Garshol, L. M.; Moore, G.: Topic Maps - Data Model, ISO 13250, ISO/IEC JTC1/SC34, information technology, (2006)
5. Heuer, L.; Schmidt, J.: TMAPI 2.0. In: Maicher, L.; Garshol, L. M.: Subject-centric Computing. Springer, Berlin (2008)
6. Moore, G.; Bogachev, D.: Topic Maps Constraint Language (Draft), ISO 13250, ISO/IEC JTC1/SC34, information technology, (2010-03-25)

[6] The current version of the Maiana platform is available at http://maiana.topicmapslab.de/