SocioTM – Relevancies, Collaboration, and Socio-knowledge in Topic Maps

Sasha Rudan[1] and Sinisha Rudan[2]

[1] HeadWare Solutions, Norway/Serbia, sasa.rudan [g-m-a-i-l]
[2] Magic Wand Solutions, Serbia, sinisa.rudan [g-m-a-i-l]

Abstract. Topic Maps (*TM*) standard solved a lot of problems in the information overload. With a semantic layer on the top of the existing data pools, TMs provide information interpretation and organization. However, user interaction with technology is still undeveloped and too explicit. This paper introduces *SocioTM* model; an extension of TM paradigm that includes relevancies, collaboration, and *socio-knowledge* (user-specific knowledge/ behaviors). Paper goes through relevancies implementation in SocioTM; relevancies building and population; relevancies interpretation, presentation; and navigation through **SocioTM**. Relevancies are introduced both on topic/ontology level and information (occurrences) level. Paper concludes with collaboration involvement in *SocioTM* building and with migration of socio-knowledge.

Keywords: Topic Maps, relevancies, semantic, ranking, rating, quality, visualization, voting, collaboration, groups, profiles, socio-knowledge, SocioTM, reliefing, socio-potential-low, mountain-view, migration

Maicher, L.; Garshol, L. M. (eds.): *Subject-centric computing. Fourth International Conference on Topic Maps Research and Applications, TMRA 2008, Leipzig, Germany, October 16-17, 2008, Revised Selected Papers.* (Leipziger Beiträge zur Informatik: XII) - ISBN 978-3-941152-05-2

1 Introduction

Topic Maps as a knowledge building and organizing technology is a fairly enough mature and powerful technology. We believe that our research should shift more to design of user interaction with technology; to make it more natural and what is the most important; more implicit. This implies need for components that can monitor and identify user behavior and preferences and be able to migrate it to the another knowledge pool. We would also like to introduce to TM arena a more native support for concepts already exploited in collaborative systems.

TMs generate two problems: *knowledge generalization* and *knowledge redundancy*[1]. Knowledge generalization means that there is no any uniqueness in knowledge representation related to the specific user. Knowledge redundancy is introduced when each user/group wants to have separate knowledge representation (meta-data set) to identify their unique knowledge interpretation. There is a strong need for one unified TM set, but also for keeping personal uniqueness of every user/group.

In order to solve these important problems "***SocioTM HyperReliefing" (SocioTMHR)*** or just shorter ***SocioTM***[2] model is proposed (Figure 1.). The very name of the model makes two important implications: 1) model is intended to be gateway between proprietary applications and TM to be as much as possible integration transparent[3] 2) model should be understood as an integrate part of TM; both in the way of necessity for it and in the relation to knowledge integration and migration of accumulated knowledge. Some features might be simulated using TCML or socio-ontology but there is a need for a standardized and in-box solution. We also need more researches and more generally accepted paradigms in that area.

1 More general speaking; meta-data generalization and meta-data redundancy
2 More details at `www.SocioTM.org`
3 Some system's functionality (not being a part of TM standard) should be accessed through the separate API

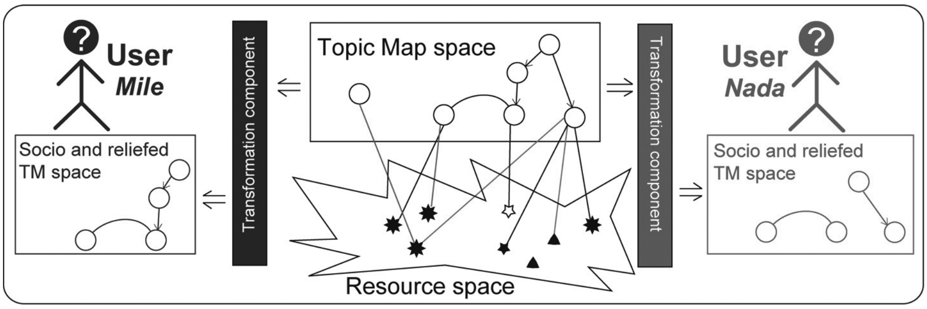

Fig. 1. Global overview of the Socio*TM model*. Different users are presented with different, profiled TM space (with a benefit to collaborative work and impact to the global knowledge)

SocioTM model gives relevancies to each TM-element (topic, association, class, occurrence, etc)[4]. SocioTM model makes us possible creation of relief and fuzzy representation of the knowledge. In this way, much easier usage of information and much better knowledge structure overview is made possible.

After introduction in Chapter 1, Chapter 2 presents relevancies population and creation; Chapter 3 talks about relevancies evolution; Chapter 4 presents all aspects of SocioTM interpretation; Chapter 5 gives a fast overview of SocioTM presentation; followed with Chapter 6 which presents navigation through SocioTM; Chapter 7 is about collaboration within SocioTM and migration of socio-knowledge and finally; Chapter 8 is just an overview of SocioTM implementation with; Chapter 9 as a conclusion of the paper.

1.1 Current state

Topic Map standard introduced roles, associations, scopes, themes, but no mechanism for easily ranking either topics or occurrences. Scope-concept and association-concept are binary-like concepts and more often part of ontology space (hardcoded) than user space. There is a need for more fuzzy and general concept.

To avoid information glut users are interested in browsing on meta-data level wanting to know which topics are more relevant, which path through the topic space will be faster and more effective. That is why users need to be presented with relevancies both on the occurrence and meta-data level and also with each TM-element.

[4] If not explicitly noticed, this research is referring to all kind of TM-elements in general

Here is an illustration of the usage of SocioTM model; users may want to know which Mozart's compositions (topics) are the most popular. Moreover, for composer Marc-Antoine Charpentier they may want to know if he was much less outstanding composer than his the best known piece - *Eurovision* opening hymn.

Recommender systems as ubiquity phenomena and ranking algorithms are already highly researched ([Geroimenko206], [Soboroff2009]) and this paper will not try to go deeper in that direction. It is up to **SocioTM** developer to choose the most appropriate models and algorithms.

1.2 Problem setting

This paper is a part of research on the system called ***KnAlledge***[5] being developed by *HeadWare Solutions*[6] and *Knowledge Federation*[7].

Our research is set in the following context: 1) resource and meta-data space are huge and highly interconnected; 2) interconnections are important for user; 3) no real-time response is required for new knowledge entrance; 4) there are many users that want to get suggestions about presented knowledge; to get structural concept of knowledge they are facing with; 5) they would like to be able to affect knowledge structure locally and preferably even globally; and 6) to be able to migrate with aggregated social-knowledge. Users want to start with preset world (not with tabula-rasa) and then to keep personal memories and make global impact onto that world.

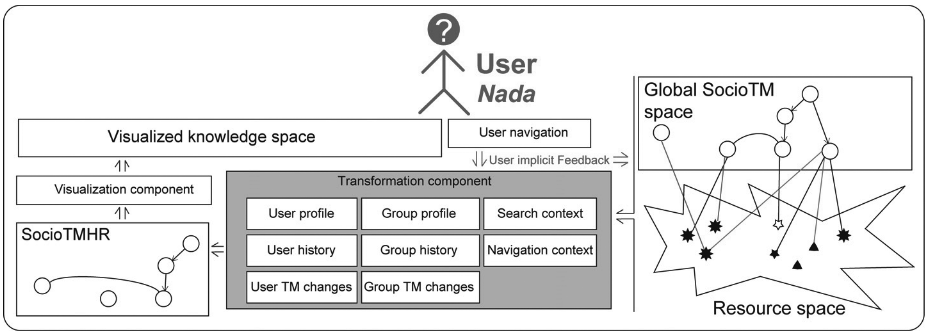

Fig. 2. SocioTM system detailed.

---

5 KnAlledge system is about collaborative knowledge generation, knowledge merging and unified presentation of the whole content (referring to the same topic) as the one filtered and merged content. More details at `www.KnAlledge.com` or `www.Memepolis.com`

6 More details at `www.HeadWareSolutions.com`

7 More details at `www.KnowledgeFederation.org`

2 Relevancies population and creation

This chapter introduces a mechanism for adding relevancies to TM-elements. TM-element population does not have a problem with meta-data-entrance-laziness. Opposite to occurrence-relevance space where we are not aware of user satisfaction with occurrence quality, in TM-element-relevance space we have good methods to monitor user satisfaction. For example, if user navigated in one direction we know that path was good[8]. The same is about a topic; if user accessed some resource which is an occurrence of that topic, etc.

SocioTM assumes static-relevancies (*static-SocioTM*) and dynamic-relevancies (*dynamic- SocioTM*). Static-SocioTM contains *static relevancies* that are persistent over successive use of Topic Map. On the other side dynamic-SocioTM contains *dynamic relevancies* that are being calculated from the static ones intended to present user/search/navigation specific scope of SocioTM.

2.1. Implementation

The easiest way to implement relevancies is by adding weight to each TM-element (similar to weighted graph or ANN (*Artificial Neural Network*) topology). However, better way would be if we could also store *socio-knowledge* (knowledge related to user/group), in a form of already extracted rules (behaviors, preferences, etc) accompanied with accumulated user's actions that are waiting to be processed. Later, accumulated actions could be used for relearning; modifying existing and creating new rules. In this way, we can predict and suggest user's actions and interest but also manage user explicit needs (like ranking specific topic).

Another solution would be to create a special storage (i.e. TM storage) for storing socio-knowledge. In this way we will have users', groups', and global SocioTM. This solution will be elaborated in this paper. Final draft for the socio-knowledge storage's (*SKS*) taxonomy will be provided on the project's web portal.

We also have to balance both with user privacy and collaboration goals. In order to keep user privacy, it is possible either to immediately populate global/group SKS to the response of user activities or to try to extract knowledge and generalize it. In both cases we are using user personal data only at the moment they are already going through the system, so the user is less concerned with the privacy aspect. However, real-time algorithms needed for real-time processing of

8 Relevance of association in relation to the source topic, not a relevance of destination topic

the user activities are both more difficult to develop, and more computational expensive [Linden2003].

For extracting knowledge/patterns and recognizing user behavior, we can use (i.e. ANN) (un)supervised learning methods [Anderson1992]. It is important to recognize users' behaviors in navigation and browsing, but also interests in more general concepts; like specific scopes, topic/occurrence, and association classes. It is also important to recognize users' preferences relating to the way they anticipate data presentation and learn; linear, spiral, etc. In this way system can predict user's navigation, interests and presentation preferable. Opposite to that would be a simple monitoring and recording user actions and then promoting them globally.

User interest in some element needs extremely complicated analysis including lexical understanding of TM content/resources, and it is a part of another, later research.

2.2 User implicit feedback

This is an important aspect of the system. It provides a chance that user's actions permanently changes original TM. In this way we both integrate collected knowledge with information-pool and provide a new user (which does not belong to any group or has own profile) with a chance to use already customized and evaluated knowledge. Some kind of feedback delay and feedback evaluation helps us in providing more globally-approved knowledge.

3 Relevancies evolution

This chapter presents a way of evaluating acquired socio-knowledge to be ready for later interpretation. Relevance evolution is more an offline process compared to the relevancies interpretation.

Opposite to almost real-time response of the new content [Das2007] our solution is more likely to get precise and highly evaluated answers with possibility of offline calculation [Linden2003].

By creating *clusters* of users (using clustering or other unsupervised learning algorithm) we can greatly reduce computation space, and reduce complexity of algorithm from $O(M*N)$ to $O(M+N)$ or even less[9].

---

[9] M presents the number of users and N presents the number of TM-elements

The challenging problem of populating user profile can be overridden by a clustering concept. Our idea is very simple; the starting assumption is that every cluster will have at least one user willing to populate questioner and to build user-profile. On the other side all users from one particular cluster can be identified with cluster-behavior (average of all populated profiles in that cluster); after normalizing cluster-behavior with user-specific behavior. We could even inspire user to confirm/reconsider automatically assigned user-profile; preferable to readjust it.

An important phenomenon developers have to pay attention on is the phenomenon of *promoted elements*. Imagine one element that users gave impression about. Users' reaction will stimulate other users to do the same. For example, if users followed association A_k from topic T_m, another user would probably unjustifiably take suggested A_k when navigating from T_m. This creates avalanche effect. Promoted elements can be handled in the following ways: 1) initial popularity divided with frequency of use, 2) postpone popularity propagation, or 3) evaluate if user was satisfied with suggestion. All these 3 approaches can be combined.

4 SocioTM interpretation

This chapter explains a transformation of static-TM with static relevancies into dynamic-TM with dynamic relevancies. In other words it explains transformation from *collective knowledge* to *scoped/profiled knowledge* (i.e. user's SocioTM). Let us just note here that static-TM is not static in a general meaning. It also evaluates through user-feedback (in this way underlining importance of collective knowledge), by growing population and by changing/voting/authorizing user-preferences.

$$\boldsymbol{TM}_{\mathrm{DYNAMIC}} = \boldsymbol{f}_{\mathrm{USER\text{-}NORMALIZING}}(\boldsymbol{TM}_{\mathrm{STATIC}}) \tag{1}$$

Figure 4 overviews a process of initializing user's SocioTM and process of static/ dynamic normalizing activities during interaction with SocioTM.

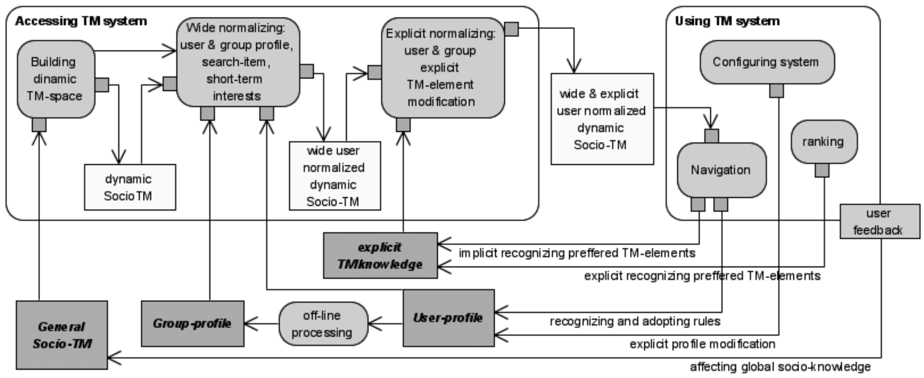

Fig. 4. SocioTM normalizing activities

4.1 Building dynamic SocioTM

There are two major models in building SocioTM: 1) duplicating static SocioTM into dynamic one; 2) generating dynamic SocioTM from the static one *on-the-fly*. No matter which model is used, it will be referred as *dynamic-SocioTM* for the sake of clarity.

4.1.1 Building dynamic SocioTM as a copy of topic space

This model makes a copy of static SocioTM in which all transformations are performed. It is completely safe to make changes against it and algorithms seem to be more efficient and easy to implement. However, for pattern recognition, the next model seems to be more practical, so it should be partially used.

4.1.2 Building dynamic SocioTM on-the-fly

This model does not create a copy, but introduces *mapping-layer* responsible for mapping static SocioTM into dynamic one on-the-fly. This model is more implementation-demanding, but on the other side it is more careful with memory consumption and it gives nice possibility of *user-feedback* implementation.

4.2 Wide normalization

Wide normalizations are all SocioTM normalization activities that have effect on whole SocioTM space and not only on the specific TM-elements.

4.2.1 Normalizing TM with user profile

Each user has a personal profile which represents user's behavior. Personal profile is built and profiled over time, by monitoring user's behavior/interests or by manual user intervention. As we already mentioned, initial user profile is cluster profile.

User profile contains explicit user set of preferences and expectations. It also contains a set of rules learned by (un)supervised learning. Rules can also be offered to user afterwards to fit them more precisely and to be stimulated to create new rules.

4.2.2 Long-term and short-term user interests

Every user has long-term and short-term interests. Long-term interests are recognized through user manual profiling or by monitoring user's behavior over a period of time. However, by avoiding short-term interests we are attracting user's present interests into wrong direction, driven by long-term-interests' suggestion.

4.2.3 Normalizing TM with search-item

Search-item contains in itself a lot of filtering information to provide not only result but also to generate separate *view* on TM. In practice, it is done by normalizing and filtering all TM-elements in TM according to search-item and user profile. How much search-item can help, depends on search-item semantic richness. One important note is that search-item is not only about 1) normalizing SocioTM, but also about 2) cutting-off non-relevant parts of SocioTM space.

4.3 Explicit normalizing with user explicit-socio-knowledge

Explicit-socio-knowledge presents a set of user explicitly modified TM-elements. There is one-to-one association between each record in explicit-socio-knowledge and addressed TM-element. Process of normalization consists of iterating through all records and appropriate modifying every addressed TM-element.

4.4 Normalizing TM through user navigation and time

Time and navigation is a very reach source of implicit knowledge retrieval. As we will see later, by only monitoring user navigation through SocioTM, system is able to ***implicitly*** recognize user behavior, relevancies and expectations.

4.5 Conclusion

The final normalizing function (summing all normalizing activities) looks like:

$$\boldsymbol{f}_{\text{USER-NORMALIZING}} = \boldsymbol{f}_{\text{GROUP-PROFILE}} \text{ x } \boldsymbol{f}_{\text{USER-PROFILE}} \text{ x } \boldsymbol{f}_{\text{ST/LT-INTERESTS}} \text{ x } \boldsymbol{f}_{\text{SEARCH-ITEM}} \text{ x } \boldsymbol{f}_{\text{EXPLICIT-SOCIO-KNOLEDGE}} \text{ x } \boldsymbol{f}_{\text{NAVIGATION}} \text{ x } \boldsymbol{f}_{\text{TIME}} \quad (2)$$

5 SocioTM presentation

5.1 Challenges with Topic Maps presentation

Even if user navigates through meta-data space there is still a huge overload of meta-data at that level but also overload of knowledge in general. This means that our system still have to cope with the problem of visualization/presentation. Relevancies introduction is a try of avoiding that problem, but it introduces new challenges; *view-clipping* and *presenting* the SocioTM.

5.2 View-clipping

When user is browsing SocioTM user should be presented with limited knowledge. The best way is view-clipping related to user tuned *relevance-threshold*. Clipping should go both horizontal and vertical. *Horizontal-clipping* means clipping to the knowledge relevant to the present user interest. If user approaches knowledge border, *socio-potential law* will extend the knowledge in the way it is presented in the next chapter. *Vertical-clipping* includes clipping by the *relevance-threshold* and clipping by the *knowledge-abstraction-level* user is interested at the moment.

5.3 Mountain-View paradigm

Mountain-view paradigm is related to the way of presenting data to the user [Karabeg2002]. Main idea is to use the user's best orientation tool; spatial and time orientation to understand knowledge structure and browse through it. As we will see in the next chapter, relief will continue to change through user navigation, new peaks will appear, and old disappear. Mountain-view paradigm provides user with visual interpretation of knowledge structure stored in SocioTM and it changes with change of user's interests and with user navigation through SocioTM.

6 SocioTM Navigation (socio-potential-law)

There are almost no researches in the area of recommenders, referring to the way of navigating through *data-set* (in our case SocioTM) [Geroimenko206]. The most of them are related to the way of recommending items (in our case TM-elements). In the info glut recommendations/relevancies are also needed for navigation paths.

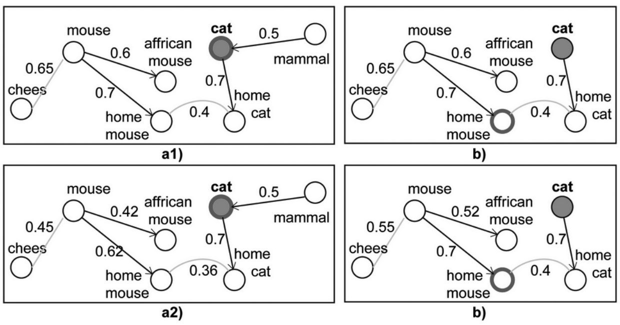

Fig. 5. Socio-potential-law

After dynamic-SocioTM is created SocioTM model can work in the *spatial-time* relevance-domain which means that relevancies are being evaluated and changed over the time and by user-navigation through dynamic-SocioTM. In this domain it is used something we call *socio-potential-law*. The socio-potential-law works similar to the physical force-potential-law[10]; all relevancies are decreasing weighted relatively to the present force-center/epicenter. Additional tensions could be introduced, like search-item origin, etc (Figure 5).

10 It falls in the group of easily convergent Force-based algorithms [Fruchterman1991]

As an implication this law gives us possibility to evaluate relevancies spatially and over time. Spatiality is evaluated by implementing some of the appropriate metric [Bruls2000].

On Figure 5.a1) someone can see that for search-item "cat" (blue circle) dynamic-relevance of every TM-element can calculate by superposition of cumulative metric distances (weights) on path from epicenter to the observed TM-element normalized with static-relevance of the observed TM-element. This makes possible transposition metric to the vertical dimension (*Mountain-View*).

Introduction of additional epicenters gives precedence to other user tensions; like search-item (blue-circle) and user’s present position in SocioTM (red-ring); Figure 5.a2). Through navigation through the SocioTM user relocates her/his tensions and some other topics become more important (Figure 5.b1 and 5.b2).

Just this evaluation of relevancies through spatial-time dimension gives us a chance to make a user-feedback to SocioTM. This is an exciting area for the future research.

Another fascinating manifest we see here is the following: if we have a bare info-pool without any recommendations or relevancies, we can just let users navigate through it; probably with some support of lexical-similarity-recommenders and lexical/tag metrics. Without forcing users to make any explicit recommendations we still can collect amazingly rich cognition about knowledge relations, contexts, relevancies, etc. We believe this area opens us a new horizon of researches in implicit social-knowledge population.

7 Collaboration within Topic Maps

Even some experts debate about meta-data overload; meta-data cannot be overload since they are supposed to help better navigation and filtering information. This means that user does not have to see meta-data but only to use them.

With topic maps and similar technologies we are providing user to navigate/browse/view not only on information level but also on meta-data level. This makes us responsible and concerned about meta-data overload. Our belief is that information overload is not about data itself, but about information presentation and providing information consumer with ability to get overall picture of data and main concepts of knowledge stored in that information pool.

SocioTM provides a better overview and knowledge selection, but at the same time it solves a big collaborative issue by sharing user experience and keeping individual aspect at the same time. In that way, we avoided duplicated socio-semantic space (meta-data redundancy) and also afforded user specific behaviors and expectations (we avoided meta-data generalization).

7.1 Socio-knowledge migration

Very important feature of SocioTM is having social-knowledge, user-profiles, and relevancies separated of TM content. The reason is simple, knowledge about knowledge, and meta-data in general should be reusable and therefore be possible to migrate to the other information pool.

That was the primary reason for introducing separate socio-knowledge storage (SKS) within our system. This makes possible mapping aggregated socio-knowledge to the other TM.

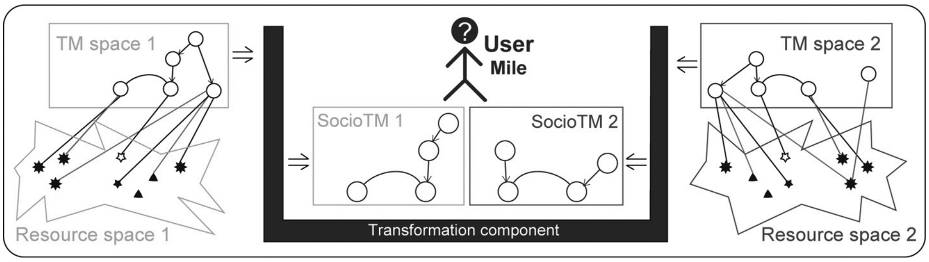

Fig.3 Migration of socio-knowledge

To make social-knowledge migration more efficient there is a need for PRIs/PSIs (Published Subject Identifiers) [Pepper 2008] to map the socio-knowledge. We would like to introduce PRI concept not only at the topic level, but also at the ontology level (which we believe should be easier to negotiate about.)

8 SocioTM implementation

SocioTM taxonomy presented here is just a glimpse of the final draft[11]

- Topic classes: topic_visited, topic_ranking, topic_examined, association_ranking, association_followed, etc
- Association classes: topics_related, etc
- Occurrence classes: relevance_value, visiting_frequency, etc

[11] Final draft would be presented on the SocioTM portal (`http://www.sociotm.org`)

8.1 Introducing SocioTM into the existing systems

Vertical compatibility is always challenging with introduction of new concepts and technologies. If system is build modular than introducing SocioTM should not be a dramatic issue[12]. We are open to the other researchers and developers for possible challenges and help in system modeling.

9 Conclusion

In this paper we presented extension to the Topic Map standard (general enough to be extend to the similar technology) that supports socio-knowledge added on the top of classical TM providing more structural knowledge and knowledge profiled to the user, but also collaborative to the community.

When it comes to standardization problem, we believe that standardization is extremely important to make this concept native and permanent companion of TM.

We can imagine different experts providing their overview/knowledge interpretation to the audience. Users would be able to choose either one or another expert (i.e. music expert) to follow her/his interpretation. In this case we would be able provide on-line, dynamic and living books about the same area and with similar content (the same global and enormous Topic Map) interpreted in a different way.

The new challenge would be to add a contextual reliefing not only to the user-specific-context but also to the search-item-specific-context. At the present moment we see it as memory-demanding issue without easily generalization/pattern recognition approach so we leave it for the later research.

References

1. Karabeg, D.: Designing Information Design. Information Design Journal Vol 11.1, pp. 82-90 (2002).

2. Soboroff I., Nicholas C., “Combining Content and Collaboration in Text Filtering”, IJCAI'99 Workshop on Machine Learning for Information Filtering, Sweden (1999)

3. Geroimenko V., Chen C., “Visualizing the Semantic Web: XML-based Internet and Information Visualization”, (2006) 102-123

[12] Especially if we are using only functionality presented through standard TM gateway

4. Rudan S., "Semantic enrichment of multimedia information and processes", 2005 http://www.sinisarudan.com/computer_sciences/multimedia&semantics/SemanticEnrichmentMM.pdf

5. Lachica R., Karabeg D., Rudan S., "Quality, Relevance and Importance in Information Retrieval with Fuzzy Semantic Networks", TMRA, Germany (2008)

6. Linden G., Smith B., and York J., "Amazon.com Recommendations; Item-to-Item Collaborative Filtering", IEEE Internet Computing, Volume 7, Issue 1 (January 2003) 76-80

7. Das A., Datar M., Garg A., Rajaram S., "Google News Personalization: Scalable Online Collaborative Filtering", Industrial Practice and Experience, Alberta (2007)

8. Bruls M., "Squarified treemaps", TCVG Symposium on Visualization, (2000) 33-42

9. O'Connor M., Cosley D.,"PolyLens: a recommender system for groups of users. Proceeding of the ECSW 2001, Germany (2001)

10. Pampalk E., Goto M., "MusicRainbow: A New User Interface to Discover Artists Using Audio-based Similarity and Web-based Labeling"

11. Sabre J., "'Relevance' in Information Retrieval", 2004

12. Gupta S., Nenkova A., Jurafsky D.,"Measuring Importance and Query Relevance in Topic-focused Multi-document Summarization"

13. Anderson D., McNeill G., "Artificial Neural Networks Technology", http://www.SocioTM.org/docs/ANN_Technology.pdf, 1992

14. Fruchterman, T. M. J., Reingold, E. M, "Graph Drawing by Force-directed Placement ", Software: Practice and Experience, 21(11), 1991

15. Pepper S., "Global data identifier", Published Subjects and global identifiers, Oslo (2008)