Augmenting the Knowledge Bandwidth and Connecting Heterogeneous Expert Communities through Uncovering Tacit Knowledge

Jasminko Novak,
Monika Fleischmann,
Wolfgang Strauss,
Martin Schneider

Fraunhofer Institute for Media Communication
MARS Exploratory Media Lab
Schloss Birlinghoven
D-53754 Sankt Augustin, Germany
Tel: +49/2241/14-3437
jasminko.novak@imk.fhg.de

Michael Wurst,
Katharina Morik

University of Dortmund,
Artificial Intelligence Dept.,
D-44221 Dortmund, Germany
Tel. 0231 - 755 5101
wurst@ls8.cs.uni-dortmund.de

Jürgen Ziegler,
Christoph Kunz

Fraunhofer Institute for Industrial Engineering,
Competence Center Software Technology and Interactive Systems
Nobelstraße 12
D-70569 Stuttgart, Germany
Tel. 0711/970-2334
juergen.ziegler@iao.fhg.de

Abstract

This paper presents an approach to exploring how implicit knowledge structures in different communities of experts can be discovered, visualised and employed for semantic navigation of information spaces and construction of new knowledge. The focus is on the developed conceptual model and system for creation and collaborative use of personalised learning knowledge maps. The developed prototype combines semantic text analysis with machine learning and interfaces for visualising relationships and creating new knowledge structures. Application scenarios include automatic generation of personalised knowledge portals, collaborative semantic exploration of complex information spaces and construction of shared ontology networks. The real-world testbed and context of development is the Internet platform netzspannung.org that aims at establishing a knowledge portal connecting digital art, culture and information technology.

1. Introduction

Our basic point of departure is that implicit knowledge structures shared by a group of users (tacit knowledge) are crucial for communication and shared use of knowledge. So a critical element in developing a model for building a knowledge community that connects experts from different fields of profession becomes the following question:

How can existing, but not explicitly formulated knowledge structures, and shared assumptions of a given community or a group of experts be discovered, visualized and made usable for cooperative discovery of knowledge in heterogeneous information pools?

In formulating a practical approach to addressing these issues we introduce following constraints and definitions. We relate the notion of "knowledge discovery" [4] to supporting the discovery of semantic contexts and relationships in an information pool which is either 1) too big or too fast growing to be scanned and categorised manually, or 2) consists of too heterogeneous content to impose one fixed categorisation structure, or 3) serves different user groups with heterogeneous interests.

In this paper we present a model for expressing implicit knowledge structures of individuals and groups of users and for using this as a means for semantic navigation and discovery of relationships in heterogeneous information spaces. We describe the implemented prototypes and outline ways for collaborative use of the system. We also illustrate the application of the developed system for creation of a knowledge portal connecting different professional communities on the example of netzspannung.org.

2. Basic concept: personalised learning knowledge maps

We relate our approach to the approach that argues that knowledge consists above all of a very personal, difficultly articulable and partly unconscious component, usually referred to as implicit or tacit knowledge [12]. Accordingly, a key to the communication and use of existing, and creation of new knowledge, lies in the transformation of implicit knowledge to explicit structures perceivable und usable by others.

As a concrete context in addressing these issues we take the process of information seeking and exploration of an information space. This can be understood as a process in which the users' interaction with information reflects their existing knowledge and produces new structures.

We introduce the metaphor of a knowledge map in order to describe a model for representation of information spaces in which the individual information items are structured according to possible meanings and semantic relationships. The central question is how such maps can be constructed based on user interaction with information, and how they can be employed for collaborative discovery and sharing of knowledge.

As one concrete solution we developed the following model. As a starting point for his actions the user is presented with an initial semantic structure generated autonomously by the system (e.g. by using methods for machine clustering). This is the system-generated knowledge map.

On one hand the user can use this structure to navigate and inspect the information pool. On the other, he can express his personal point of view by rearranging the system-generated structure (e.g. by moving objects between groups, creating new groups, adding relationships etc.).

In this way the meaning of users' actions is contextualised by being referred to the structure generated and hence „understood“ by the system. Accordingly, the users actions can be taken as a reference for constructing a new knowledge structure that reflects the users personal point of view.

The user-defined structure can now be learned by the system (e.g. employing supervised learning methods) and formalised, in a way which allows it to be applied as a user-defined „template“ for semantic structuring of an arbitrary information space. In this way the implicit knowledge of a given user that has been revealed through his actions on the initial semantic structure has been acquired and can be visualised or applied to new situations.

Developing a concrete realisation of the described model involves integrating and extending methods from several research fields. For autonomous machine-based generation of knowledge maps based on semantic relationships one can refer to work on document clustering and semantic networks such as [11], [9], [15], [1]. For extracting semantics from user interactions with information, to work on machine learning and collaborative filtering such as [13], [6],[3], [7]. And for developing an intuitive user interface for visualisation of relationships and large data sets we refer to work on information visualisation and focus+context interface techniques such as [14], [16], [10], [2].

3. The knowledge map system

The described model of personalised learning knowledge maps is realised in the following way. A knowledge map consists out of four main visualisation models: Content Map, Dimension Map, Index Map and Source Map (Fig. 1).

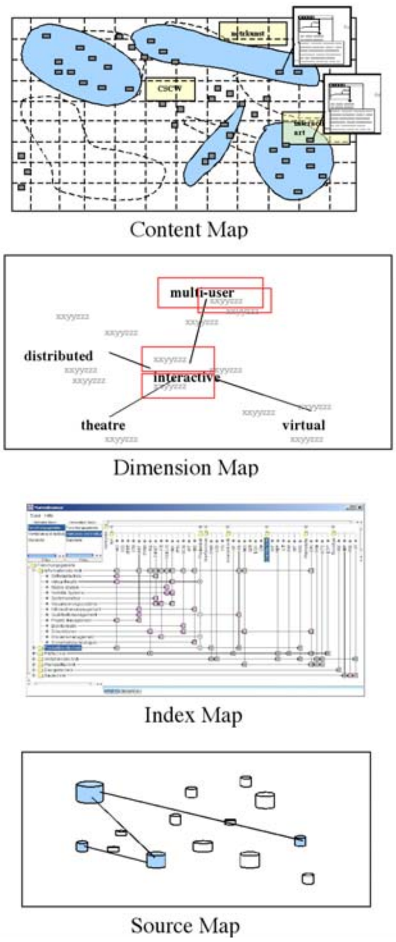

Fig 1. Main elements of a Knowledge Map

The Content Map provides an overview of the information space structured according to semantic relationships between information items. In the first realisation the ContentMap visualises clusters of related documents and offers insight into implicit relationships between their content. This is the main context for users exploration and interaction with information.

The Dimension Map provides an insight into the criteria that have determined the semantic structuring in the Content Map. These criteria are a kind of semantic axes that define a given structuring out of a variety of possibilities. In the case of structuring the ContentMap based on analysis of word usage in texts, the Dimension Map is a WordMap. It shows the sets of words that determined the parameterisation of documents and groups these word sets based on contexts of their appearance in the documents themselves.

The IndexMap visualises the hierarchical structures of the information space. It allows the user to create and visualise explicit semantic relationships between information items and to formalise them into an ontology.

The SourceMap displays semantic relationships between different data sources that emerge as a result of the relationships between their content[1].

As a concrete realisation of this model the current system prototype implements following modules:

- a text-analyser for encoding semantic properties of texts into a vector space model suitable as input for clustering and learning agents
- a Kohonen clustering agent for system-generated semantic clustering of information
- an SVM[2] classification agent for learning the user-defined classification of information into
- the LSM[3] Browser for visualizing and exploring the information space based on the results of system-generated (Kohonen) or user-defined clustering (SVM)
- the LSM Editor for user restructuring of the Content Map
- the Matrix Browser for visualization of hierarchical structures and creation of named relationships as user-defined ontologies

[1] The Source Map also serves as the interface for the dynamic data adapter for user-oriented semantic integration of heterogeneous data sources. This issue is not discussed here.

[2] Support vector machines [8]

[3] Learning Semantic Maps.

The interplay of these modules makes it possible to: 1) generate an automatic structuring of a large document pool based on semantic analysis of document contents, 2) provide a visualisation and interface for navigating along this structure, 3) capture the personal points of view of different users, and 4) make them available to others in form of personal maps or by connecting them into a shared ontology. The following sections describe the individual steps of this typical use scenario.

3.1 Exploring contexts and relationships through system generated maps

By invoking the LSM Browser (Fig. 3) the user is provided with a visualisation which presents the documents grouped into clusters of possibly related content (ContentMap). The basis for this is the combination of statistical text-analysis with a Kohonen clustering agent [11][9] which groups similar documents together and positions the resulting clusters on a 2D grid. Highly correlated texts are close to each other, while little correlated texts are far from each other.

By displaying the distribution of all items and their grouping in semantically related clusters the Overview Mode of the LSM Browser gives a quick, general impression of the information pool. The semantic space of each cluster is described by a number of keywords. One kind of keywords is extracted from the data records as entered by the user, while the other is generated by the server side text-analysis.

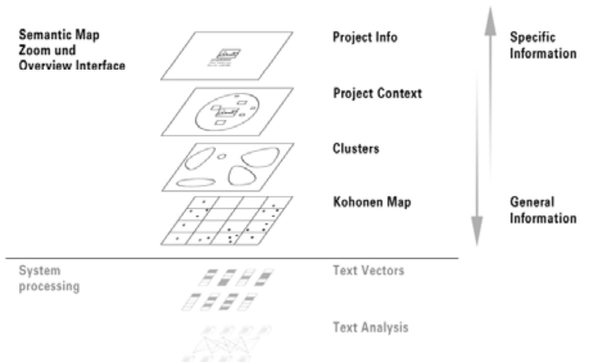

Fig. 2. Information Layers of the LSM Browser

A detailed investigation of the relationships in the presented information space is supported by the Zoom Mode, which displays a set of objects within a given semantic distance from a selected object. Keeping the visual distance between the objects in direct proportion to their semantic similarity allows the user to develop a more accurate impression of possible relationships between an individual object and its “semantic neighbours”.

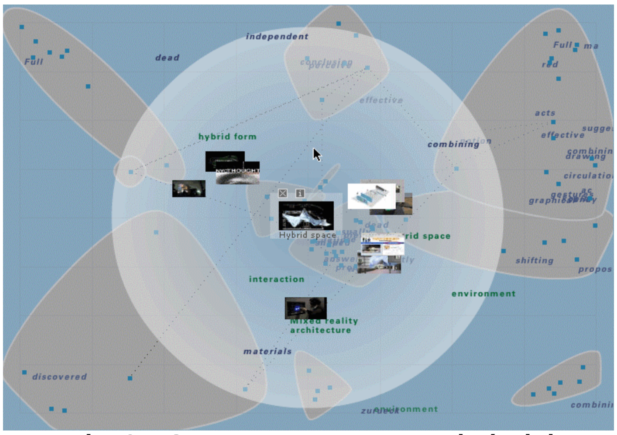

Fig. 3. LSM Browser: Zoom Mode (Bright-Circle), Overview-Mode (Background)

By using the LSM Browser the user can now explore possible contexts and relationships in the given document pool. In doing so he will discover relevant documents as well as construct a personal interpretation of the way they should be put into relation. In order to express this personal point of view he can rearrange the map structure by moving documents between clusters, deleting or creating new clusters as well as assigning labels describing the cluster contents. In the current implementation this is made possible by a separate interface, the LSM Editor (Fig. 4).

3.2 Capturing and learning a user specific knowledge structure

By editing a map in the LSM editor the user defines a set of classes (Fig. 4). The idea of learning a map is to find a decision function, which allows to assign new documents to these classes automatically. There are several possibilities to find such a function. Currently we use statistical text analysis to transform documents into word vectors. The text classification algorithm builds a decision function based on these word vector representations. To assign a new document automatically to a class, the document is first transformed into a word vector and then passed to the decision function.

The algorithm we use for generating a decision function is based on "support vector machines" [8]. This method has not only proven to lead to low error rates, but also to be very efficient. Both issues are of special importance in the described model of creating personalised knowledge maps. Normally the user does not provide many examples by editing a map in the LSM Editor. Even so the resulting decision function should be sufficiently accurate. The second important issue is speed, as the user normally stands by, while the system learns a map. Using the SVM, learning takes less than a second on a common PC, which is by far acceptable waiting time.

3.3 Applying the personal map

After a map has been learned, it can be applied to any single document or document source provided by the system. Figure 4 shows a screenshot of the LSM Editor demonstrating this.

The map on the right side is structured by the user. It contains the user-defined classes "acting", "media" and "story". These three classes are learned and after this applied to the whole pool of documents. The structure of the resulting map (on the left) is the same as the structure of the user defined map, but there are additional documents in each class. These documents were automatically assigned by the learned decision function. Applying a learning semantic map to a document pool allows the user to gain a personalized view of the documents in this pool, as only documents relevant to the users LSM are chosen and assigned to its clusters.

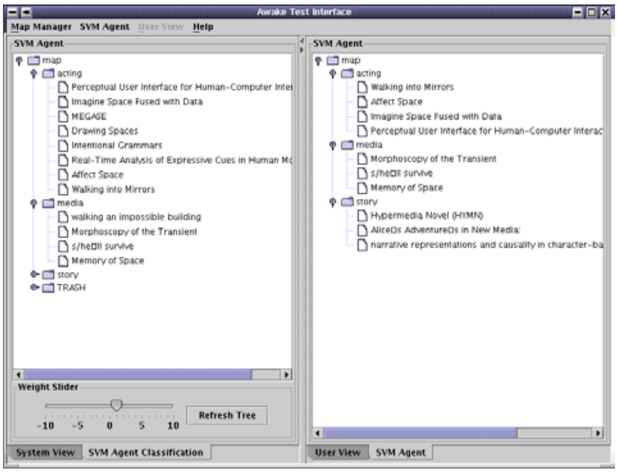

Fig. 4. LSM Editor

3.4 Creating a personal ontology

As a basis for enabling the user to extend her personal map into a formalised personal ontology, the Matrix Browser provides two visualizations of the correlation between personal maps and indexed resources. One shows a subset of indexed resources on the vertical axis and the conceptual structure of a users map on the horizontal one (Fig. 5). Already existing resources in the map are shown by marking a cell of the matrix. By adding additional relations between resources and concepts, the personal map can be extended, or by adding additional (hierarchical) concepts, it can be refined.

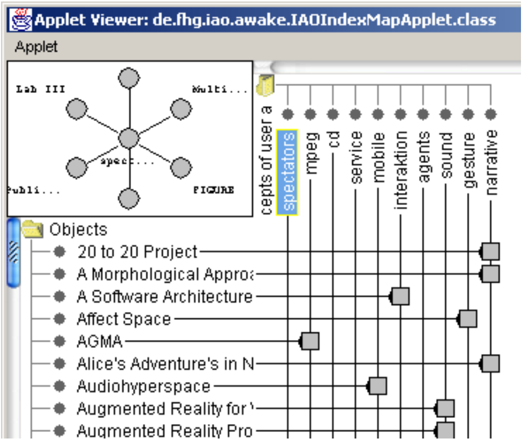

Fig. 5. Matrix Browser

Moreover, additional types of relationships between resources and concepts or between the concepts themselves can be added (e.g. "cd" "is a part of" "media"), in order to enlarge the map structure to an ontology. The other view uses the axis of the matrix to display maps of two different users in relation to each other. Having a relationship between two folders is an indicator for sharing the same resources and thus possessing a semantic link. In this way the relationships discovered by using the map visualisations and editing tools presented in earlier chapters can be extended and explicitly formalised into a user-defined ontology that can span different maps.

4. Collaborative generation and use of knowledge

Using the system in the described way a group of users (e.g. a community of experts from different fields) can create a collection of personal knowledge maps, which they can share with each other. For example, a given user can employ a map of another user to automatically classify an information pool with that users "knowledge". A typical use case here would be the use of expert knowledge to navigate in information pools or thematic fields not well known to oneself.

Another scenario that we are investigating is the collaborative contextualisation of search results. Currently the system contextualises keyword based search on personal maps. But the interesting question is which map is the "best" to be chosen as a context for displaying given search results for a given user?

Furthermore, given a community of users with an open policy for knowledge sharing, by using our system the users can build their own maps as extensions of the maps of others. In this way they actively extend knowledge structures of other users and connect them with their own. Adding procedures for tracking these connections and map versioning together with an underlying ontology is a way for generating an ontology network based on contexts of actual use. This is an important issue that we are investigating in further work.

5. Application to netzspannung.org

The practical test bed and first application context of the described work is the Internet platform netzspannung.org. netzspannung.org aims at establishing a knowledge portal that provides insight in the intersections between digital art, culture and information technology [5]. The basic requirement for realising such a knowledge portal that connects a community of experts from different fields is: a continually evolving information pool needs to be structured and made accessible according to many different categorization schemes based on needs of different user groups and individual contexts of use (e.g. artists and researchers, expert users and interested public).

Typical netzspannung.org users are experts and professionals such as artists, researchers, designers, curators and journalists. One scenario for using the functionalities implemented by the described knowledge map system is exploring possible relationships between work in different topics or fields of profession (e.g. for finding projects from different disciplines relevant for one's own work).

An example of this is the test deployment of the LSM Browser prototype as information access interface to the netzspannung.org information pool containing the submissions of the cast01 conference and the competition of student projects digital sparks[4].

Another scenario is comparing sets of projects against one's own personal point of view and against views of other experts (e.g. for discovering contexts and hidden assumptions). And finally the general public could use the knowledge of the expert community by employing knowledge maps of experts as a guide for navigating the information space of netzspannung.org.

6. Ongoing work and development

One focus of current work is the deployment of collaborative methods for learning user-defined map structures. Although text processing methods have shown to be a sufficient starting point for learning user-defined maps, they are not fully satisfying. The main reasons for this are the heterogeneity of the underlying text documents and the relatively small amount of training data provided by the users. As an effect, some documents, which would be relevant in a given context, are not

[4] http://netzspannung.org/cast01/proceedings

retrieved by the system. To overcome this problem, we are working on employing collaborative methods to classify and cluster documents. These methods make use of the whole set of maps edited by the users of the system. They are used to calculate collaborative similarities between documents, such that documents, which often appear in the same cluster, are supposed to be similar. In this way, even documents which are based on very different texts (e.g. written in different languages) can be found to be relevant to each other.

The other important issue is the improvement of the interface with visualisation models for large data volumes and focus and context handling with non-linear, semantic zooming. A special challenge here is incorporating a model, which allows multiple non-linear zoom focuses. This model is currently being implemented as a java-based client that will allow the user to apply the functionalities of the system in one integrated interface.

7. Acknowledgements

The work described in this paper has been undertaken within the projects *AWAKE – Networked Awareness for Knowledge Discovery*[5] and *netzspannung.org – an Internet Media Lab*, both financed by the German Federal Ministry for Education and Research.

Thanks go to Boris Müller for his participation in the work on the LSM Browser during his year at Fraunhofer IMK and to Kresimir Simunic for his help with the early prototype of the Kohonen clustering.

8. References

[1] Becks, A.;Sklorz, S.; Jarke, M.:A Modular Approach for Exploring the Semantic Structure of Technical Document Collections, ACM Proc. of AVI 2000, Palermo, May, 2000.

[2] Bederson, B.B. et al.;. Pad++: A zoomable graphical sketchpad for exploring alternate interface physics. *J. Vis. Lang. Comput.*,7:3, Mar. 1996.

[3] Chalmers, M.; Paths and Contextually Specific Recommendation; DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries, 2001

[4] Fayyad, U. M.; Piatetsky-Shapiro, G.; and Smyth, P.; From data mining to knowledge discovery: An overview, In Advances in Knowledge Discovery and Data Mining, AAAI Press and the MIT Press, Chapter 1, 1-34.

[5] Fleischmann, M., Strauss, W.; Novak, J. et al.: netzspannung.org : an internet media lab for knowledge discovery in mixed realities, in Proc. of cast01 // living in mixed realities, Fraunhofer IMK; 2001; ISSN 1618-1379

[6] Herlocker, K.; Konstan, J.; Reidl, J.; Explaining Collaborative Filtering Recommendations, in Proc. of CSCW 2000, December 2-6, 2000, Philadelphia, Pennsylvania, USA

[7] Joachims, T. et al.; Web Watcher: A Tour Guide for the World Wide Web, IJCAI, 1997

[8] Joachims, T.; Text Categorization with Support Vector Machines: Learning with Many Relevant Features, Proceedings of ECML-98, 10th European Conf. on Machine Learning, 1998

[9] Kohonen, T.; Kaski S.; et al.; *Self Organization of a Massive Document Collection*, IEEE Transactions on Neural Networks, Vol. 11, No. 3, May 2000

[10] Leung Y. K. and Apperley M. D., A Review and Taxonomy of Distortion-Oriented Presentation Techniques, ACM Transactions on Computer-Human Interaction 1:2, 1994.

[11] Lin, X.; Soergel, D.; Marchionini, G.; *A self-organizing semantic map for information retrieval*, in Proc. of 14th ACM/SIGIR Conf. Research and Development in Information Retrieval, 1991

[12] Nonaka, I.; Takeuchi, H.; The Knowledge-Creating Company, Oxford University Press, London, et al 1995

[13] Resnick, P. et al; GroupLens: An Open Architecture for Collaborative Filtering of Netnews; Proc. of ACM 1994 Conference on CSCW, Chapel Hill, North Carolina, 1994

[14] Robertson, G.G.; Mackinlay. J.D.; The document lens, In *UIST '93*, , Atlanta GA, USA, Nov. 1993., ACM Press.

[15] Warren Sack, "Conversation Map: An Interface for Very Large-Scale Conversations," Journal of Management Information Systems (Winter 2000-2001)

[16] Sarkar, M. et al.; Stretching the Rubber Sheet: A Metaphor for Viewing Large Layouts on Small Screens, ACM Symposium on User Interface Software and Technology (UIST), 1993

[5] http://awake.imk.fraunhofer.de/