Generating Semantic Contexts from Spoken Conversation in Meetings

Jürgen Ziegler
Interactive Systems
University Duisburg-Essen
47048 Duisburg, Lotharstrasse 65
+49 203 379-1430
ziegler@interactiveSystems.info

Zoulfa El Jerroudi
Interactive Systems
University Duisburg-Essen
47048 Duisburg, Lotharstrasse 65
+49 203 379-1430
eljerroudi@interactiveSystems.info

Karsten Böhm
Computer Science Department
University of Leipzig
04109 Leipzig, Augustusplatz 10-11
+49 341 9732340
boehm@informatik.uni-leipzig.de

ABSTRACT
SemanticTalk is a tool for supporting face-to-face meetings and discussions by automatically generating a semantic context from spoken conversations. We use speech recognition and topic extraction from a large terminological database to create a network of discussion topics in real-time. This network includes concepts explicitly addressed in the discussion as well as semantically associated terms, and is visualized to increase conversational awareness and creativity in the group.

Categories and Subject Descriptors
H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces -- *computer-supported cooperative work, synchronous interaction, theory and models.*

General Terms
Human Factors, Algorithms, Experimentation.

Keywords
Semantic context, conversational awareness, real-time speech-recognition, intelligent assistance.

1. INTRODUCTION
Meetings and face-to-face discussions are still one of the most frequent forms of cooperative work and are often targeted at creative, innovation-oriented types of tasks. While there is a wide range of meeting support tools addressing aspects like collecting ideas, editing shared representations or supporting the moderator's tasks (see e.g. [2],[9]), these tools require participants to explicitly manipulate the system which may negatively interfere with the creative process. As a consequence from this and other factors, physical whiteboards, flipcharts and pens are, aside from electronic presentations, still the prevalent instruments for joint creative activities.

IUI'05, January 10–13, 2005, San Diego, California, USA.
ACM 1-58113-894-6/05/0001

The objective of the work reported here is to provide a more light-weight, casual form of supporting meetings. Our approach focuses on supporting awareness of the semantic context of the discussion which is generated without explicit user action from the spoken conversation in the group. The relevance of various context aspects such as content, communication and relations among participants has been discussed extensively in the CSCW literature (for an overview, see e.g. [1]. Our work is targeted at enhancing conversational awareness [3] by making the content of a conversation visible at a semantic level. With this approach, we aim at stimulating creativity, especially by supporting associative thinking, while keeping the discussion focused. By automatically documenting key concepts of the discussion, a smooth transition from the creative, explorative phases of a meeting towards more structured, operative group activities should be supported.

In the following we introduce the system SemanticTalk, which generates a semantic context automatically from the spoken conversation in a meeting. The system uses speech recognition and real-time extraction of topic structures in conjunction with an information visualization component. We will start by describing in a brief scenario how the system is used in a group setting.

2. SCENARIO
The functionality of SemanticTalk can best be illustrated by a brief scenario:

A group of investment consultants are discussing future developments in financial services, gathering ideas for a new financial product. All participants are equipped with small headset microphones. Their individual statements are transformed into text by continuous speech recognition and integrated in a coherent text stream. Relevant terms are extracted from that text stream and relations between the terms are detected. Subsequently, the result is visualized for the group as a network (map) of topics (see Figure 1). In addition to the terms actually spoken, semantically associated terms are activated and integrated in the topic structure.

Figure 1 shows part of a topic map, which has been generated in the course of the conversation (German terms in Figure 1 are translated in the text). The terms shown in white boxes, such as

"financial product" or "Eastern Europe", were actually spoken in the discussion. The system automatically associates semantically related terms (shown in grey boxes), such as names of Eastern European countries. This allows participants to subsequently address these associated themes and discuss, for instance, the opportunities of financial products targeted at those countries. In the course of the conversion, the topic map grows dynamically.

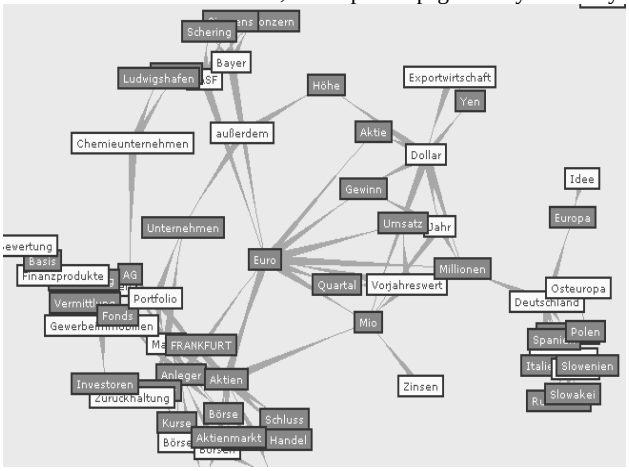

Figure1. Automatically generated topic map

3. ARCHITECTURE AND TOPIC EXTRACTION

SemanticTalk combines existing technologies for speech recognition and for the analysis of large unstructured text corpora (text mining) with a topic extraction mechanism and a visualization component. Figure 2 shows an overview of the architecture of the prototype system.

The speech inputs of several speakers are first processed individually by commercial speech recognition software (VoicePro) and transformed into a text stream. After a certain time interval, the texts are sent to a server component where they are integrated and serialized. Term extraction is performed using a static, large-scale reference database of German terminology, which is described later. A first-level filtering is done by extracting nouns. At the next level, an adjustable threshold allows to define a relevance level terms must to have for the domain of discourse. In the next step, associations between terms selected from the text stream are determined based on term relations in the underlying database. If no additional terms can be associated with a certain term within a defined time frame, the term is discarded. Furthermore, the initial network is expanded by determining additional associated terms (not present in the original text), which are integrated in the resulting term structure if the strength of the association is above a certain threshold.

The resulting network is displayed and continuously updated in a graph visualizer running as an applet. Terms and their semantic associations are retrieved from a database which constitutes a large reference corpus of the German language [6]. In order to build up these semantic references, word collocations in large document collections (about 10 million phrases mainly from newspaper articles) were determined automatically. Collocations are pairs of words, which appear sufficiently often together in a certain text segment (e.g. a phrase or directly beside each other). Semantic references are derived from phrase collocations and have a significance value, which is calculated from the number of words involved, the number of appearances of the collocations and the size of the corpus. The resulting graph reflects the context, in which the particular word appears in the underlying corpus.

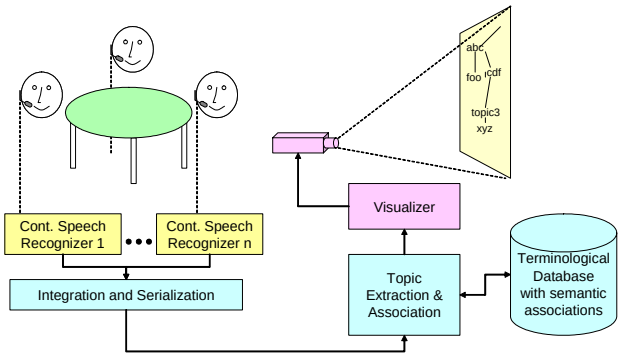

Figure 2. Overview of the SemanticTalk architecture

While the basic terminological database represents general linguistic data (comparable to the general linguistic competence of native speakers), it can be extended by domain specific terminology which can be derived by processing relevant sets of documents. In this way the system can be adapted and tuned for the specific discourse domain the group addresses

4. APPLICATION POTENTIAL

While SemanticTalk is still an exploratory prototype there is a wide range of applications that could benefit from the approach. Analysts discussing business processes could come up with a first draft of relevant concepts, tasks and relations. Software engineers discussing requirements with their customer could use an initial topic map to derive pertinent functions or object classes. A marketing campaign team could use the system to explore the associative space around a new product.

An important potential lies in tightly coupling retrieval and recommender functions with a creative group discussion. The topics addressed in the conversation can immediately be used for formulating queries or establishing a context for retrieval operations. Recommender functions could propose relevant documents or background information (e.g. existing patents in a product definition activity) that make the group discussion more focused and effective.

SemanticTalk could also be used for supporting the joint development of domain models and ontologies. It offers a possibility to integrate different inputs and build up an initial, informal concept structure from the conversation and possibly additional documents. The resulting material can be directly used for subsequent restructuring and refinement which is an important step towards more integrated content engineering processes.

5. RELATED WORK

While there exists a substantial body of work on text classification and topic or metadata extraction from multimedia data, research on real-time topic extraction from speech in group meetings has so far been very limited. Jebara et al. [7] describe a system that classifies the current focus of the conversation according to a limited number (12) of pre-defined topics. The

classifier is trained by initially providing training sets of documents associated with each topic.

Real-time analysis of spoken conversation is also reported in DiMicco & Bender [4]. Their focus is on facilitating equal participation in group discussions by visualizing the contributions of each participant. Terms and sentence fragments are associated with a fixed number of categories by a classification component also based on machine learning. Galley et al. [5] have investigated techniques for segmenting meeting transcripts into topically related units, but without naming these topics. Kazman et al. [8] presented a system which automatically indexes videotaped meetings so that they may be queried like a database, whereas this paper is primarily addressed at the indexation of distributed meetings.

None of the above approaches, however, builds up a concept structure by relating the spoken terms and visualizing them in real-time during the meeting. Furthermore, no extraction of semantically related terms is reported.

6. EVALUATION AND PERSPECTIVE

Due to the exploratory nature of the current prototype, a full evaluation with respect to factors like applicability in real group settings or usability is not yet feasible. Nonetheless, the feasibility of the approach could be demonstrated in a number of initial small evaluations, using both predefined conversation scenarios as well as unconstrained discussions. Participants as well as observers mostly considered the derived topic structure relevant for the actual content of the conversation. New terms could be picked up and integrated into the discussion. The performance was in general good enough to achieve a sufficient synchronization between the flow of the discussion and the update of the topic map displayed. One must note, however, that we have so far only tested the system with shorter discussions (several minutes, support of realistic meetings lasting up to several hours is still a subject for further research.

Continuous speech recognition, especially when talking to other people, is still a very error-prone process. We estimate the accuracy of the speech recognizer in our scenarios to be only about 60-80%. The resulting topic structure, however, proved to be surprisingly well even for texts spoken fast and casually. Because of the subsequent matching against the term data pool, recognition errors were of much less consequence than they would have been when simply dictating text. Therefore, the semantic context generated helps to disambiguate the recognition-based input. We plan to investigate this aspect further in future research.

A restriction that is not likely to be removed in the foreseeable future is the speaker dependency of the speech recognition, requiring up-front training. For practical use, however, this problem can be alleviated, for instance, through server-based speech profiles.

We plan to continue our work along several lines. On the one hand, work will be done in the core component, to provide typed relations between the terms extracted. For this purpose, typed relations are generated in the underlying vocabulary database.

On the other hand, there will be research on visualizing and manipulating the topic structures more effectively. One goal is to provide means for displaying and handling very large structures and for supporting meetings of realistic duration.

This involves improved techniques for showing the dynamic development of the focus of the conversation, possibly in several threads. A second goal is a better integration of manual group activities like mind-mapping with automated techniques. This also includes ways to interactively edit and refine the generated topic maps. Finally the integration of existing agendas, ontologies or topic structures is important for realistic applications.

7. ACKNOWLEDGEMENT

The authors would like to thank the German Ministry of Research and Education (BMBF) for partly funding this work in the project INVITE, and the following project members for their valuable contributions: Wolfgang Beinhauer (Fraunhofer IAO), Reinhard Busch (linguatec GmbH) and Christian Raether (ISA Informationssysteme GmbH).

8. REFERENCES

[1] Borghoff U. and Schlichter J. *Computer-Supported Cooperative Work. Introduction to Distributed Applications*, Springer-Verlag, Heidelberg, 2000.

[2] Conklin, J. and Bergman, M. L. gIBIS: A hypertext tool for team design deliberation. In P*roceedings of Hypertext ´87*, ACM Press, Chapel Hill, 1987, 247-251.

[3] Dey, A. K., Salber, D., Abowd, G. D. and Futakawa M. The Conference Assistant: Combining context-awareness with wearable computing, In *Proceedings of the 3rd International Symposium on Wearable Computers ISWC '99* (October 1999), IEEE Computer Society, San Francisco, 1999.

[4] DiMicco, J. M. and Bender, W. Second Messanger: Increasing the Visibility of Minority Viewpoints with a Face-to-face Collaboration Tool. In *Procceedings of Conference on Intelligent User Interfaces*. (Funchal, Madeira, Portugal, January 2004), ACM Press, Funchal, 2004.

[5] Galley, M., McKeown, K., Fosler-Lussier, E. and Jing, K. Discourse Segmentation of Multi-Party Conversation. In *Proceedings of ACL-2003*, Sapporo, Japan, 2003, 562-569.

[6] Heyer, G., Quasthoff, U. and Wolff, C. Automatic Analysis of Large Text Corpora - A Contribution to Structuring WEB Communities In *Proceedings of Second International Workshop*, (IICS 2002), Springer Verlag, Kühlungsborn, 2002.

[7] Jebara, T., Ivanov, Y., Rahimi, A., and Pentland, A. Tracking conversational context for machine mediation of human discourse. In *AAAI Fall 2000 Symposium - Socially Intelligent Agents - The Human in the Loop*, AAAI Press, Massachusetts, 2004.

[8] Kazman, R., Hunt, W. and Mantei M. Dynamic Meeting Annotation and Indexing. In *Proceedings of the 1995, Pacific Workshop on Distributed Multimedia Systems*, (Honolulu, HI, March 1995), Idea Group Publishing. Honolulu, 1995, 11-18.

[9] Streitz, N., Geißler, J., Haake, J. and Hoi, J. Dolphin: Integrated meeting support across Liveboards, local and remote desktop environments. In *Integrated publication and information systems*. GMD IPSI, Darmstadt, 1994.