SemanticTalk: Software for Visualizing Brainstorming Sessions and Thematic Concept Trails on Document Collections

Chris Biemann, Karsten Böhm, Gerhard Heyer, Ronny Melz

University of Leipzig, Ifi, Department of Natural Language Processing,
Augustusplatz 10/11, 04109 Leipzig
{biem,boehm, heyer, rmelz} @informatik.uni-leipzig.de
http://www.wortschatz.uni-leipzig.de

Abstract: In this demonstration we introduce a technology to support knowledge structuring processes already at the time of their creation by building up concept structures in real time. Our focus was set on the design of a minimal invasive system, which ideally requires no human interaction and thus gives the maximum freedom to the participants of a knowledge creation or exchange processes. The system captures and displays spoken dialogs as well as text documents for further use in knowledge engineer's tools.

1 Introduction

Our goal is to support the communication and mutual understanding between and within groups in two ways: On one hand we provide a visualisation tool for spoken language, having its application in informal creative and usually highly innovative meetings like brainstorming sessions or open space workshops (cf. [Owen, 1998]. In these scenarios the tool does not only serve as an automatic documentation method by arranging the keywords in a meaningful way, but also provides corpus-based associations in order to enrich the conversation with concepts that are related but might have been forgotten by the participants. On the other hand we use the same visualization engine for displaying single documents as trails on a so-called *semantic map*. The idea of a semantic map is heavily relying on the well known concept of geographical maps, in which visual structuring of interesting locations and the emphasizing of relevant paths between them are the key concepts to provide an orientation for the user. Since the location of the concepts on the map is fixed, users can grasp the contents of a document rapidly and compare documents in a visual way.

Both features are part of a prototype implementation called "SemanticTalk" which has been presented to the public in at the worlds largest IT-exhibition CeBit in Hanover, Germany, in spring 2004 and is described in detail in [Biemann et al. 2004c].

2 SemanticTalk – a tool for Innovation Acceleration

SemanticTalk supports two operation modes: a *brainstorming mode* in which a semantic map will be constructed dynamically from the input and a *red-thread mode* that visualizes a communication trail in a pre-calculated map. The input for the system can be obtained from different sources: it may be loaded from text files, directly typed in using a provided input field or spoken text recorded from a headset.

The SemanticTalk user-interface that represents the different views of the semantic map as well as the controls that modify the behaviour of the tool is show in the figure below.

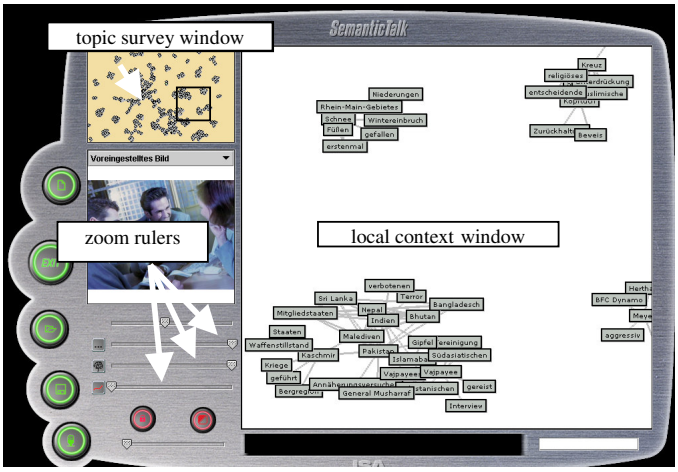

Figure 1: The *SemanticTalk* user interface. The local context window is a zoomed variant of the topic survey window and can be changed with the three zoom rulers. Other controls: New, Load, Save, Exit, Microphone on/off, Freeze graph, Brainstorming / Red thread Switch, Input text bar and scrolling

2.1 Visualization of Associations in Brainstorming Mode

For preparation, a text collection of the domain has to be processed as described in [Biemann et al. 2004a] to provide the terminology and the global contexts to the system. If existent, ontology or some typological system can be loaded in order to assign types to concepts and relations.

The brainstorming session starts with a blank screen. Words of the conversation that are considered to be important (an easy strategy is to use only nouns of a substantial frequency in the domain corpus) are thrown into the visualization. In addition to this, associations obtained from the text collection are shown. It is possible to include an ontology or alike for typed (colored) nodes and edges.

2.2 Visualization of Concept Trails on Semantic Maps in Red Thread Mode

Semantic maps provide a good overview of a domain by visualizing keywords and their relatedness. While they are extracted from a document collection and serve as model for the whole domain, it is possible to display single documents as paths through them. We visualize the content of the broadcast as trajectory in the pre-calculated semantic map. Words of interest from the document are marked in red and are connected in sequence of their occurrence by directed edges. By using a fixed semantic map and dynamically representing input as indicated, it is possible to examine things like coverage of domain, coverage within (sub-)topics, relatedness of document to the map domain and comparison of different documents. The great advantage of this representation lies in displaying what is described in the document, and the same time, what is *not* being dealt with. By using a fixed representation for the domain, the practiced user can grasp the document's content in a few seconds and then decide whether to read it or not.

3 Knowledge Discovery Aspects

The preliminary text processing (see [Biemann et al. 2004a]) is language-independent and fully automatic and learns relations and associations between words. The user is able to discover the knowledge in the text collection by simply talking about the topic he is interested to and watching the associations. Since words and their connections are linked to documents and passages of their origin, the tool can be used for visual Information Retrieval. Other machine learning aspects used for enriching the data are described in [Biemann et al. 2004b]: Ways are described how to assign relation types like antonomy, hyponomy and alike to pairs of words by using statistical measures. Training on a set of relations gives way to rule detection in order to enhance the number of words per typed relation automatically.

References

Biemann, Chr.; Bordag, S.; Heyer, G.; Quasthoff, U.; Wolff, Chr. (2004a): Language-independent Methods for Compiling Monolingual Lexical Data, Springer LNCS 2945

Biemann, C.; Bordag, S.; Quasthoff, U. (2004b): Automatic Acquisition of Paradigmatic Relations using Iterated Co-occurrences, Proceedings of LREC2004, Lisboa, Portugal

Biemann, C.; Böhm, K., Heyer, G., Melz, R. (2004c): Automatically Building Concept Structures and Displaying Concept Trails for the Use in Brainstorming Sessions and Content Management Systems, Proceedings of I2CS, Guadalajara, Mexico

Faulstich, L.; Quasthoff, U.; Schmidt, F.; Wolff, Chr. (2002): Concept Extractor - Ein flexibler und domänenspezifischer Web Service zur Beschlagwortung von Texten, ISI 2002

Jebara, T.; Ivanov, Y; Rahimi, A.; Pentland, A. (2000): Tracking conversational context for machine mediation of human discourse. In: Dautenhahn, K. (Hrsg.): AAAI Fall 2000 Symposium - Socially Intelligent Agents - The Human in the Loop. Massachusetts: AAAI Press.

Owen, H. (1998): Open Space Technology: A User's Guide, 1998, Berrett-Koehler Publishers Inc., San Francisco