Archives of Control Sciences
Volume 15(LI), 2005
No. 3*, pages 351–362*

Mapping speech streams to conceptual structures

RONNY MELZ, CHRIS BIEMANN, KARSTEN BÖHM, GERHARD HEYER and FABIAN SCHMIDT

We describe a software system that processes textual data and spoken input streams of natural language and arranges the information in a meaningful way on the screen: concepts as nodes, relations as edges. For spoken input, the software simulates conceptual awareness. A naturally spoken speech stream is converted into a word stream (speech-to-text), the most significant concepts are extracted and associated to related concepts, which have not been mentioned by the speaker(s) yet. The result is displayed on a screen as a conceptual structure.

Key words: statistical language processing, co-occurrences, conceptual graphs, speech

1. Introduction

The phenomenon of 'Semantic Priming' has been discovered already by the ancient Greek and is nowadays a well-known and widely studied phenomenon. In principle it states that certain semantically related entities (e.g. words) facilitate or inhibit the recognition of other such entities [10]. For example, 'hospital' clearly belongs to the closer semantic neighborhood of 'doctor' or 'nurse' and hence primes them. By contrast, it would rather inhibit other words like 'teacher', which is associated to still different concepts. An outline of the structure of semantic networks is provided in [11]. According to the central tenet of structuralism, origina-ting in the ideas of Ferdinand de Saussure [4], the meaning of a sign is determined by its relations to other signs only. The discipline of Lexical Semantics deals with words, which are a special case of signs: the meaning of words is revealed through their characteristic relations to other words. 'SemanticTalk' is a software system which actually displays human-like associations. As will be shown in the following sections, co-occurrence analysis provides a framework to automatically extract intuitively appealing word relations from large text corpora.

Explicit semantic context provided in real-time can be of enormous practical use for a broad spectrum of applications, in particular within the settings of conferences, open space workshops or highly innovative team sessions. Another interesting application scenario are interviews which are often used for explicit knowledge elicitation (e.g. in debriefing meetings with employees that leave the company or with domain experts to

The Authors are with Institute of Computer Science, University of Leipzig, Augustusplatz 10-11, 04109 Leipzig. E-mails: {rmelz, biem, boehm, heyer, fschmidt}@informatik.uni-leipzig.de
Received 15.10.2005.

capture individual knowledge). The associated context serves to depict a problem from different points of view, to associate new ideas to those already mentioned or simply to provide an assisting function for instantaneous as well as later reference.

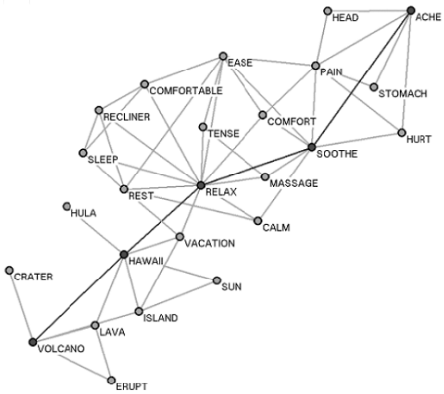

Figure 1. Associative Network. The graph is taken from [11] and has been collected and assembled by Nelson et al. [3]. It shows the con-nection between 'volcano' and 'ache' crossing various other associated concepts

This contribution aims at describing SemanticTalk and explains the mechanisms involved in probabilistic speech processing. The essential principle is as follows: a naturally spoken speech stream is converted into a word stream (speech-to-text), the most significant concepts are extracted and associated to related concepts. These concepts are extracted from a database, which can be conceived as the 'conceptual memory' of SemanticTalk. The result is output as an incrementally, dynamically unfolding conceptual structure on a screen. The graph includes and interlinks concepts mentioned by the speaker(s) as well as automatically associated relevant concepts from the database.

The remainder of this contribution is structured as follows: the architecture of the system is described and each module is introduced in a dedicated subsection. Principles of extraction, association and visualization are described and several application scenarios are sketched in the last section.

2. System architecture

The general architecture of the system consists of the following modules:

1. speech analysis
2. extraction of relevant concepts
3. meaningful connection of concepts
4. association of related but unmentioned concepts
5. calculation and visualization of the conceptual structure.

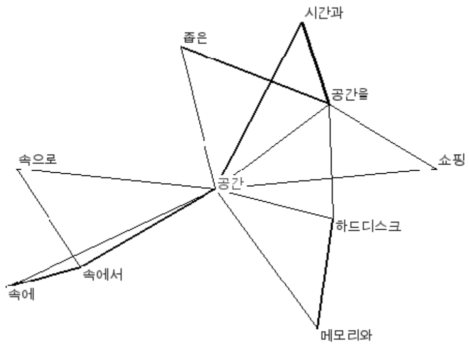

Figure 2. Conceptual structure of ‘space’ in Korean

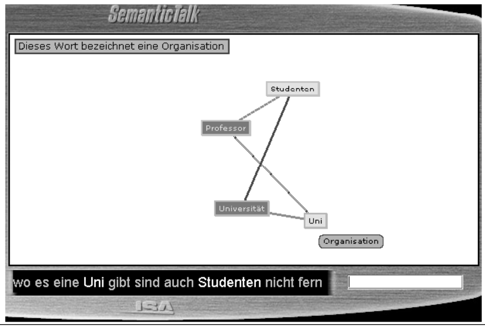

Figure 3. The input stream, monitored at the bottom of the figure, emphasizes the most prominent concepts by a distinguished brighter color. Here, ‘Uni’ and ‘Studenten’ have been extracted. On the main panel, two further concepts are associated to the graph in a second step: ‘Professor’ and ‘Universität’

Although our own research focuses mainly on extraction, connection and association of concepts, each of the five modules will be described briefly. The system is currently designed for the German language, yet it has been shown in [5], that the underlying mechanisms for the analysis are language independent (cf. Fig 2). Language dependencies have therefore been introduced by the speech recognition process that can be easily exchanged for the use of other languages due to the modular architecture of the system.

2.1. Speech analysis

The core formula of statistic speech recognition systems is

$$W = \arg\max P(O|W)P(W)$$

Hence, the most likely sentence out of all sentences in the language L is selected, given some acoustic input O [6]. Both the sentence

$$W = w_1, w_2, w_3, \ldots, o_n$$

as well as the observed input

$$O = o_1, o_2, o_3, \ldots, o_n$$

are assumed to be sequences of discrete symbols, for simplicity. $P(O|W)$ has been termed ‘observation likelihood’ and $P(W)$ ‘a priori’ probability. In speech analysis, they are referred to as ‘acoustic model’ and ‘language model’, respectively. In practice, there are much more important implementation details to consider. To give an example, the logarithm of a probability (‘logprob’) is preferred instead of the probability value P itself to avoid an underflow, that emerges if very low probabilities are multiplied, and secondly, adding logprob costs instead of multiplying probabilities, is generally more efficient. In the SemanticTalk implementation, the commercially available ViaVoice software was used to convert speech to text. It provides the interface between acoustic speech and a stream of words.

2.2. Extraction of relevant concepts

George K. Zipf [8] observed a relationship between words and their occurrence frequencies in sufficiently large, balanced text corpora: The 'conjoining glue' of language are usually short, frequently occurring constituents distinguished by rather syntactic importance. To the other side of the spectrum, most of the very sparsely occurring concepts of a text collection can be assumed to be typing or scanning errors. These spectral insights can be used to tune band pass filters according to desired properties. In our implementation, extraction consists essentially in a matching of the word stream against a database with term frequency information to determine their degree of relevance: A large reference corpus, such as one of the Wortschatz corpora, containing the result of an analysis of a representative document collection is crucial. Irrelevant words to the borders of the spectrum are filtered out. Term frequency in the sense of a maximum likelihood estimation thereof is proportional to first order Markovian unigram frequency: a token's count depending on no preceding state. Both can be derived by a purely automatic corpus analysis. In addition to this probabilistic approach, an extension of the filter may be rule-based: For example, a strategy for German may benefit in particular from the fact, that nouns are generally marked by capitalized initial letters. Nouns usually convey a great deal of the meaning of a text and thus are important to be recognized. Adjusting both the probabilistic and the simple rule-based filters to the extraction task's intention yields a heuristic function that tags the most relevant concepts for further processing. Fig. 3 illustrates, that from the stream 'Wo eine Uni ist, sind auch Studenten nicht fern' (Usually there are students around a university), 'Uni' and 'Studenten' have been extracted as the most prominent concepts of the sentence.

2.3. Connection of concepts

Subsequent to the extraction of the significant concepts from the stream, these concepts are interlinked. The connection is essentially based on the same principles as described in the following 'association' subsection. If two concepts relate to each other and the significance measure indicates a strong relation, they are connected. If an associated concept of a word relates to an associated concept of another concept cluster, they are linked by their neighbors as shows Fig. 3. The underlying data structures from the database ('conceptual memory') are the same for both connection and association, but for an improved conceivability, the background colors of actually spoken concepts and automatically associated concepts differ.

2.4. Association of related concepts

The last step before visualization is to automatically associate relevant information to the concepts already extracted, the so-called ‘explicit semantic context’. The simple example graph of Fig. 3 gives an impression of such a context: ‘Professor’ and ‘Universität’ are linked to ‘Uni’ and ‘Studenten’. To supply a semantically relevant context,

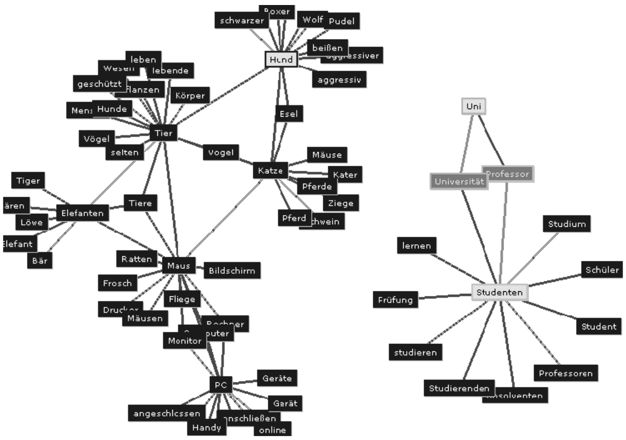

Figure 4. Like in the semantic network of free asso-ciations (cf. Fig. 1), where ‘volcano’ lead to ‘ache’ crossing several other concepts, in this graph ‘dog’ is shown to connect to ‘computer’ across ‘animal’, ‘cat’, and the polysemous ‘mouse’

a common ground of knowledge about frequently used word types and their relations is indispensable. A relational database provides this information - it is compiled from a large text corpus evaluating the local context of a word. The local context of a word is a well-defined, meaningful unit of information (paragraph, sentence, phrase etc.) and a relation between two words is added to the database, if they co-occur frequently. That requires an adequate co-occurrence measure to be defined and an empirically estimated probability distribution function (pdf). Here, a function similar to the G-Test for Poisson distributions is applied. It measures the ‘amount of surprise’, that two words occur together. Two words w_A and w_B, assuming independence, are expected to occur together with the probability $P(w_A \cap w_B) = P(w_A)\Delta P(w_B)$. ‘Occurring together’ requires a restricted language unit wherein this event may take place or not. This could be a window of constant size or - like in our calculations - the sentence s. Assuming that the words only occur once per sentence, the maximum-likelihood estimation for the involved probabilities can be used to calculate the co-occurrences of which we are interested in: $p = P(w_A \cap w_B) = n_A/N \cdot n_B/N$, where n_A is the number of times, token w_A occurred in the corpus, and N is the amount of all sentences in the corpus. For simplicity, the pdf is assumed to be a Poisson distribution with expectation value as well as standard deviation $\lambda = pN$. For the actually observed co-occurrences $k > \lambda$, the pdf $P(X = k)$ is monotonously falling:

$$P(X = k) = \frac{1}{k!}\lambda^k e^{-\lambda}$$

Hence the inverse, $P(X=k)^{-1}$, can be used to rank the co-occurrence significance. The above equation is not efficiently computable. However, since we do not require the significance to have the properties of a probability value, we just can follow the same approach as in speech recognition and use the logprob (where the base of the logarithm merely accounts for an arbitrary constant):

$$sig(w_A, w_B) = \log_e P(X=k)^{-1} = -\ln\left(\frac{1}{k!}\lambda^k e^{-\lambda}\right)$$

which can be further simplified by applying Stirling's formula

$$k! \sim \sqrt{2\pi k}\left(\frac{k}{e}\right)^k$$

Neglecting all non-contributing (independent of k) and vanishing terms results for large k in the 'significance formula'

$$sig(w_A, w_B) \sim k \cdot (\ln k - \ln \lambda - 1)$$

See [7] for a general discussion of co-occurrences and [1] for a discussion of the method implemented here. To further enhance usability and expressiveness, the relations as well as the word types themselves can be tagged. On the one hand, for domains with rather homogeneous entities, the inherent structure of the documents can be employed by means of Information Extraction methods. To give an example, imagine the semantic map of conference papers. The papers feature equal attributes like title, authors, references and share similar keywords. On the other hand, general linguistic information can be added by part-of-speech (POS) tagging or annotation. The latter is the approach used here.

Co-occurrence significance as a measure of word significance differs from term frequency in that it reflects a statistical property of a relation between two words instead of a single word. For specific applications, a domain specific corpus can be added to supply the conceptual memory with more detailed information about specific topics.

2.5. Visualization of the conceptual graph

The visualization is based on TouchGraph (see http://www.touchgraph.com), a freely available graph visualization tool. Concepts - represented as nodes - arrange in a comprehensible manner, depending on their relations - represented as edges. Concepts even dynamically cluster together, which makes the whole network more easily understandable for human users. The force-based TouchGraph-approach preserves much of the graph's symmetry, works efficiently and is incremental in the sense that new nodes can be adjoined to an already existing network. The principle is illustrated in e.g. [2]. Nevertheless, many necessary and convenient features have been implemented on top of the existing functionality taking into account the inherent semantic features of the graph, of which only a few can be mentioned here: various statistically motivated, conceptual scale scrollbars, a color scheme for nodes and edges to map e.g. semantic or syntactic categories, various focus capabilities, or the red-thread-function, which quickly reveals the essential theme coverage of a static conceptual graph.

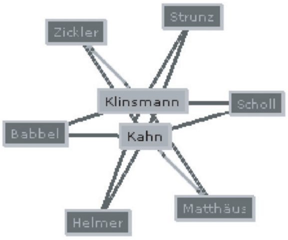

Figure 5. This graph has been generated during a conversation about soccer. 'Klinsmann' (famous German trainer) was recognized by the system as an important concept in this context. This concept being already present on the screen primes 'Kahn' (famous goal keeper). Both concepts together provide a facilitating semantic context to associate still other concepts like 'Matthäus' or 'Babbel'

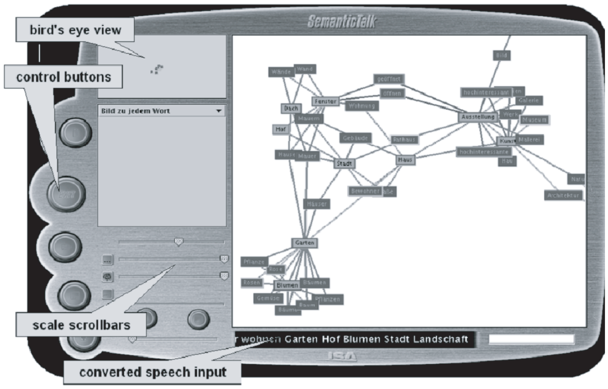

Figure 6. The graphical user interface of SemanticTalk

2.6. Usage scenarios

Generally, SemanticTalk generates and displays conceptual structures. Possible applications of the software are manifold. As has been pointed out in [5], the software features two modes: the free association mode and the red thread mode. Additionally, a temporal dimension can be added to enable trend mining, if the a document splicing function can be provided.

2.7. Free association mode

In free association mode, a conceptual graph is generated from scratch. Members of a discussion directly benefit from the evolving graph: it reflects the covered themes of the discussion by concept clusters and thus assists to keep the discussion focused. Secondly, automatically associated contextual concepts suggest new aspects to be discussed more in detail, reveal the ambiguity of concepts, et cetera. One of the most important features of SemanticTalk is its 'creativity': particularly in brainstorming sessions, new ideas are stimulated when new concepts are associated to the graph. Assume an organization which wanted to extend its production to a new arising, specific market - such as an electronic display manufacturer who wanted to build up a new division for large TFT panels, for example. A brainstorming session with managers, designers, engineers, is scheduled to get straight the concerned issues. Here, automatic associations of general language would be of less interest. But since the applied statistic methods do not require anything more than large amounts of text, it is straightforward to create associations specific for the domain of interest: Usually large amounts of unstructured text are available from the internet or company internal sources, especially for new arising technologies. In the above mentioned example, this text would typically be processed in preparation for a brainstorming session. By co-occurrence analysis, a domain specific database is compiled, which serves to extend or to replace the ‘conceptual memory’ of SemanticTalk during the actual session. The benefit of SemanticTalk in such a setting is the provision of a variety of views (reflected in the uttered individual speech streams) on the underlying document collection that represents the known information. These views potentially contribute to the resolution of conflicts due to different viewpoint, as well as the acceleration of the consensus-finding process by unifying different views into one.

2.8. Semantic maps and red thread mode

A second application is the generation and visualization of static conceptual graphs or semantic maps, which serve as orientation in document collections. Frequently, humans are requested to extend their knowledge to hitherto unknown domains. With the always growing availability of electronic information, it is often difficult to grasp the most important concepts and their relations of document collections quickly. SemanticTalk provides an overview of a collection: related key concepts are extracted and interlinked. As opposed to the more ‘situative’ and dynamic nature of the graphs in free association mode, semantic maps are calculated once-for-all or a collection: important terms are selected by comparing relative frequencies of occurrence to a large general-domain resource, and connected by co-occurrence analysis. By use of this method the keywords arrange themselves in clusters to finally form a map that eases the orientation within the document collection, especially for non-experts. All keywords of one cluster are related to a single topic. Keywords can be part of several clusters, but are painted visualized only once during the visualization process. As a consequence they will be located between the clusters they belong to. While semantic maps 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 (the so called red thread of a conversation). Words of interest from the document are marked in red and are connected in sequence of their occurrence by directed edges. Here we distinguish between two kinds of connections.

1. A connection between two concepts has already been present in the semantic map: the two keywords are semantically related and therefore connected by a red colored edge.
2. The connection has not been present in the semantic map: this indicates a shift of topic, like the introduction of a new piece of information. Such connections are colored orange.

When using spoken input, the colored connections are visualized in real time; the local context window continuously auto-focuses the most recent keywords in the center of the drawing plane. For text file input, the resulting trail is pre-calculated and visualized after processing the whole text file. Words that are contained in the semantic map are marked in red in the input. The input document connects to the semantic map, which gives rise to bidirectional retrieval: the local context window of the semantic map adjusts to the (clickable) red words in the input, and document contexts from the single document and from the underlying collection can be retrieved by simply selecting a word from the semantic map. An example of red thread mode on a semantic map is depicted in Fig. 7.

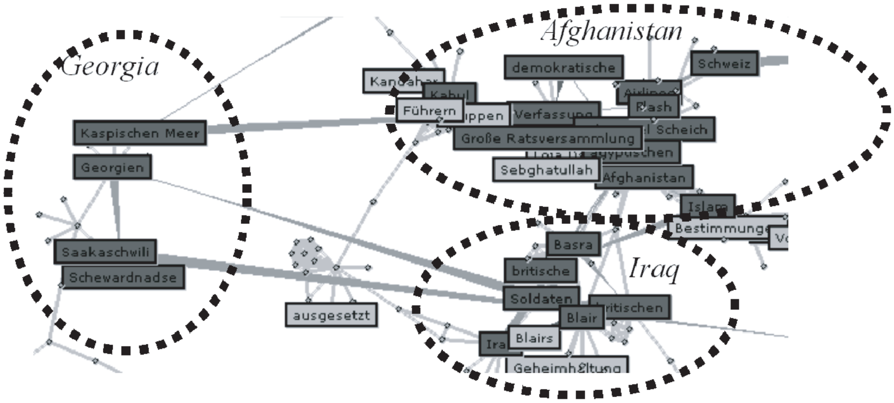

Figure 7. Visualizing a (spoken) news broadcast in a semantic map constructed from newspaper texts of January 4, 2004. Three thematic clusters referring to different political events in Afghanistan, Iraq and Georgia can be identified. Words in red belong to the red thread, long orange connections are topic shifts

By using a fixed semantic map and dynamically representing input as indicated, it is possible to examine

- coverage of domain: the more clusters are visited, the better the document covers the domain,
- coverage within topics: the more words in a cluster are marked in red, the more extensively the document deals with the topic of that cluster,
- relatedness of document and domain: few red nodes indicate non-relatedness,
- contiguity of document: many successive long range connections indicate semantic incoherence,
- visual comparison of different documents by using distinguishing colors.

Beyond co-occurrence analysis, semantic maps may incorporate typed information as obtained by Information Extraction, (see [12]) for a survey), as e.g. carried out for the KnowTech map, a large German conference on Knowledge Management [5]. The generated semantic map describes efficiently the focus of each article and its relatedness to similarly focused contributions by displaying its title, authors, organizations and keywords. The latter serve as a means for connecting related articles and positioning them according to their relatedness. Fig. 8. shows the visualization of a single article.

2.9. Trend mining

Sometimes it is not only interesting to find the most important terms of document collections, but also to track temporal changes of their usage. Designers of products might be interested in following trends of other product families, e.g. the upcoming of retro-style radios might trigger retro-style telephones. To be amongst the first to recognize a trend and to be prepared for it might influence the prosperity of a company to a

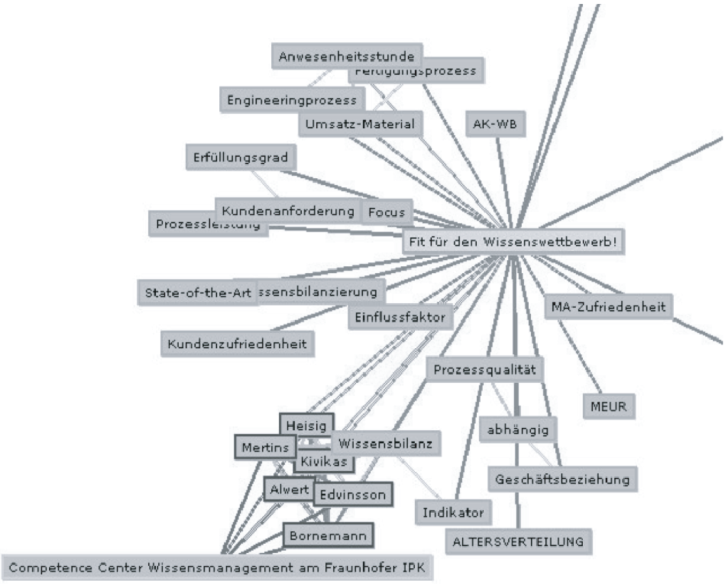

Figure 8. Visualization of a single article in a conference proceedings map. Authors are marked in green, organisations in pink, keywords in grey. The title of this example contribution is highlighted

large extent. When accepting the hypothesis that trends announce themselves by being mentioned increasingly in time-related texts such as daily news, journals or web blogs, a tool to display possible trends is of immense interest. In SemanticTalk, we incorporated a mechanism to merge several time-stamped semantic maps into one combined map. For example, a designer would let the software construct a semantic map for each of several issues of a design journal, and combine them to a time-sensitive map. It is possible to scroll along a time line to see what part of the map originated from which point in time. Furthermore, concepts are marked in different colors representing an increasing, steady or diminishing usage. Still, the identification of trends and the distinction from short-term hypes is left to the user - again the software should rather supports rather than dictate an interpretation of the results. Fig. 9. shows an example taken from daily newspaper sources of April 2005.

3. Closing remarks

SemanticTalk was first presented to public as a prototype implementation at the CeBIT trading fair in Hannover, Germany in March 2004. The final software release has been launched to the market by ISA Informationssysteme GmbH, Stuttgart in December that year and since then a lot of positive user feedback encourages us on our way to further explore end develop the current solution. At the moment, we are enriching the semantics of the XML-based connection to further support the embedding in standard

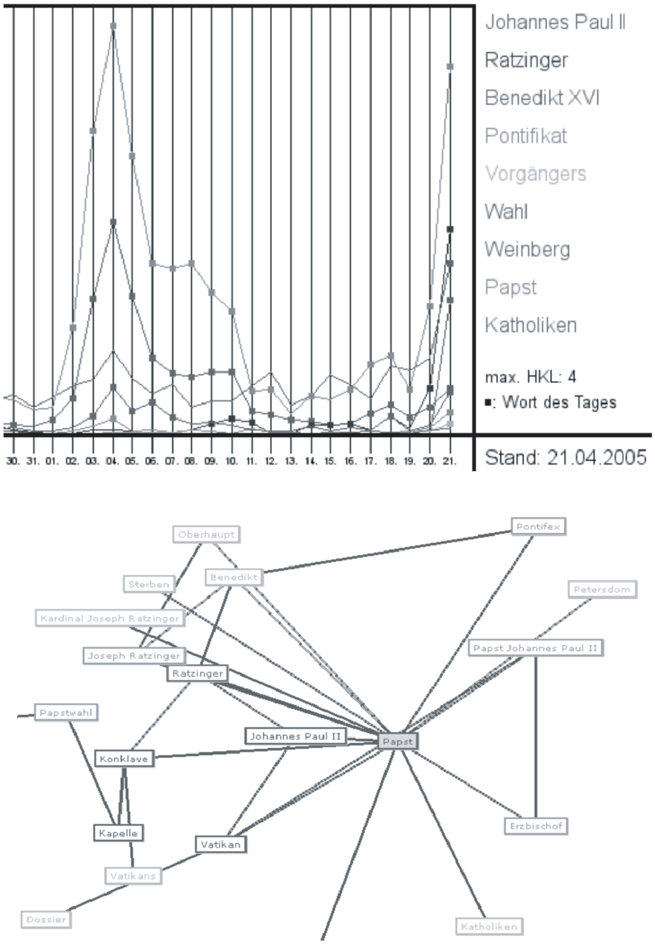

Figure 9. Upper part: the frequency of selected related concepts for the establishment of Pope Benedikt XVI. Lower part: A time-sensitive semantic map for the period of 17th - 21st of April 2005. Diminishing concepts are marked in red, increasing concepts in green, persistent concepts in blue

business environments via Aris. The Aris Process Platform provides integrated tools for designing, implementing and controlling business processes, described by AML (Aris Modeling Language). The aim is to directly generate a business process model of the conceptual graph. Important additional functionality is the recognition of Aris objects and process steps. Vice versa, such models can be loaded and matched against a semantic map of special domain knowledge in Red Thread Mode, for instance. Finally, an adaptation to various languages other than German is planned for the very near future.

References

[1] F. SCHMIDT: Automatische Ermittlung semantischer Zusammenhänge lexikalischer Einheiten und deren graphische Darstellung. Diplomarbeit, Universität Leipzig, 2000.

[2] E. BATTISTA and T. TOLLIS: Graph Drawing - Algorithms for the Visualization of Graphs. Prentice Hall, New Jersey, 1999.

[3] D.L. NELSON, V.M. MCKINNEY, N.R. GEE [and] G.A. Janczura: Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review*, **105** (1998).

[4] F. DE SAUSSURE: Cours de Linguistique Générale. Payot, Paris, (1916).

[5] C. BIEMANN, S. BORDAG, G. HEYER, U, QUASTHOFF and C.WOLFF: Language-independent Methods for Compiling Monolingual Lexical Data. *In Proc. CICLing*, Springer, (2004).

[6] D. JURAFSKY and J.H. MARTIN: Speech and Language Processing. Prentice Hall, (2000).

[7] G. HEYER, U. QUASTHOFF and T. WITTIG: Wissensrohstoff Text. Subject to appear at W3L-Verlag, Dortmund, (2005).

[8] G.K. ZIPF: Human Behavior and the Principle of least Effort. Addison-Wesley Press, 1965.

[9] J.F. SOWA: Conceptual Structures - Information Processing in Mind and Machine. Addison-Wesley, Massachusetts, 1984.

[10] T. HARLEY: The Psychology of Language. Taylor & Francis Group, 2001.

[11] M. STEYVERS and J.B. TENENBAUM: The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. Subject to appear in Cognitive Science, (2005).

[12] R. GRISHMAN: Information Extraction: Techniques and Challenges. *SCIE*, (1997), 10-27.

[13] C. BIEMANN, G. HEYER, H.F. WITSCHEL and F. SCHMIDT: Eine Wissenslandkarte der Knowtech. *In T. Schildhauer N. Gronau, B. Petkoff (Eds)*, Wissensmanagement - Wandel, Wertschöpfung, Wachstum, GITO, Berlin, (2004), s 207-216.