

## GeneralNews – An Interactive Metabrowser

### Abstract

GeneralNews (<http://www.generalnews.de><sup>1</sup>) is a meta-browser that substitutes in real-time the words on websites. The substitutions may be synonymous to, more abstract, or more specific than the original expressions. By these variations, new descriptions of the world and descriptions of new worlds are created.

### 1 GeneralNews: The Idea

The system takes an existing web page - in principle, an arbitrary one, but the idea works best with news - and replaces words in real-time, as desired by the user. Specifically, the user can change the level of abstraction with a graphical slider (see Figure 1) and thereby investigate the emerging 'space of possibilities' around a text.

- Synonyms preserve the original meaning - more or less! - and indicate the variety of linguistic expressions.
- Abstractions (hypernyms) lead to generalizations and possibly open a bigger picture. When taken to the extreme, though, they trivialize the content.
- Specifications (hyponyms) create similar but alternative worlds to the original text.

The system relies on an electronic lexicon (WordNet from Princeton University<sup>2</sup>), which contains the various kinds of related words, such as synonyms, hyper- and hyponyms. The system displays these entries when the user activates the detail mode. The system then analyzes the average level of abstraction for each original text and indicates the value on the scale.

A German version of the system, using GermanNet<sup>3</sup>, is currently in preparation. Furthermore, we are gradually improving the coverage of the system by adding linguistic analysis, moving from PoS-tagging and a simple morphology module to a more elaborate one that can deal more properly with verb inflections.

### 2 Background: Reflection of Language

GeneralNews reflects language. When we speak or write, the manifoldness of the world is reduced to the linearity of language. The mechanism of GeneralNews enriches this linearity again with the semantic variety of possible other descriptions. These variations are displayed as animated text on the screen.

### 3 Text-mediated Reality

When applied to news-sites, GeneralNews alters the text-mediated reality, thereby roughly keeping the same meaning. With extreme abstractions, however, the content is trivialized, since only abstract objects relate to each other. On a slightly lower level of abstraction, the general structures of the content can result. In the other, specific extreme, we get various alternative stories about the original incident. Concerning the selection of information and its strategic presentation, GeneralNews is exposed to the mechanisms of mass media communication.

As a special application called ArtAbstracts, working exclusively on the website of the ZKM (Center for Media and Arts, Karlsruhe), the system changes the interpretations of art-pieces (meaning-preserving substitutions and abstractions) and refers to alternative 'artworks' (specifications).

#### 4 Abstraction – a Powerful Invention

Thinking in alternatives is important for progress since we develop new ideas by changing existing descriptions. Here, abstraction as a thought pattern plays a major role, even though it is often used in the pejorative sense: ‘too abstract’. Yet, a plea for abstraction would include the following topics:

**Language and the structure of knowledge** According to many cognitive theories, we group our concepts in taxonomies. With these hierarchies, properties from the generic terms are inherited to the specific terms. This also allows us to derive conclusions about the unknown. With the aid of abstractions we organize our knowledge and create a big picture.

**Science** According to our notion of science, we need the understanding of general laws and structure, in order to create good, i.e. probable, prophecies. Most science is the attempt to formulate these laws - and thereby abstraction captures the variety of single cases and exceptions. Thus abstraction helps to organize the relevant information within the abundance of data.

**Creative and artistic processes** Our actions may be described as moving up and down between abstract, diffuse and intuitive goals and concrete actions following these intentions.

The ability to abstract can only partially be automated with the contemporary means of computational linguistics and artificial intelligence. GeneralNews is the computational attempt to generalize the incoming information purely on the basis of lexical knowledge. Empirical, causal or even strategic insights can not be implemented by that method. Yet, as a meta browser GeneralNews aims at triggering our fantasies - via abstraction.

#### 5 Implementation

Technically, GeneralNews is implemented as a relocating HTTP proxy for HTML documents. HTML documents are requested through the GeneralNews proxy by prepending their origin URL by the URL of the GeneralNews proxy.

Thus, assuming the proxy being located at *http://www.generalnews.de/proxy/* for instance, the document at *http://www.ihf.com/* is requested through the proxy under the URL *http://www.generalnews.de/proxy/www.ihf.com/*. The proxy requests the document from its origin URL, processes its content, and sends the processed document in its response. The processing of the document consists of these functions:

**Compositional references** All relative references to external resources that compose the document by means of transclusion – such as images, style sheets, script files – are resolved relative to the origin location of the document such that they are not requested through the proxy. Absolute references to the resources are not affected. This processing results in the resources being loaded directly from their origin location and not through the proxy, besides assuring that relative references that consist of absolute path names can be correctly resolved in the relocated document. Excepted from this processing are relative references in framesets, which are treated like navigational references, as described below.

**Navigational references** All navigational references to other documents i.e. hyperlinks, image-map areas, and form actions, as well as references in frame elements, are changed to point through the proxy. This causes hyperlinks in the relocated document to link to a relocated document as well, that is, following links from GeneralNews processed pages leads again to GeneralNews processed pages.

## GeneralNews

**Lexical filtering** Words in the text on the filtered page are looked up in WordNet. Data about the synsets as well as hypernyms and hyponyms are inserted as Javascript data structures into the filtered document. These data structures are associated with the occurrences of the words in the text via HTML element ids, which are added to the HTML text.

**User interface** User interface code in HTML and Javascript is added to the filtered document as HTML text inserted into the *body* element and references to CSS style sheet and a JavaScript script document as a *link* and a *script* element respectively that are inserted into the *head* element. Initialization code for the user interface and the text manipulation code is placed in the *onload* eventhandler of the *body* element such that the user interface is initialized after

loading the filtered document in the browser. This added Javascript code uses the lexical information provided in the data structures generated in the lexical filtering step to affect the text display of the page in the browser.

The GeneralNews proxy is written in Perl<sup>4</sup> and runs as mod perl<sup>5</sup> application under the Apache http server<sup>6</sup>. The WordNet database<sup>7</sup> is stored in a MySQL<sup>8</sup> relational database and is accessed using the DBI and DBD::mysql<sup>9</sup> Perl modules.



Figure 1: The slider blended into a filtered web page. It can be interactively dragged to any convenient position in the browser window. Shown is the front page of the International Herald Tribune<sup>[10]</sup>



Figure 2: The slider blended into the filtered web page in different modi of operation, shown as clippings from figure 1. From top: the slider operated to highest abstraction, to most specificity, and the detail view of one of the words. The gap on the scale marks the average abstraction level of the text measured as the average length of the WordNet hypernym graph starting at each word in the text.

**Credits**

**Idea, Production:** Daniela Alina Plewe

**Software:** Steffen Meschkat

**Computational Linguistics:** Manfred Stede, Peter Tauter, Uwe Küussner

**Layout, Interface:** Daniela Alina Plewe, Daniel van Alphen

**Website:** Karl Gampper

**Partner:** ZKM, Center for Art and Media

University of Potsdam - Institute for Linguistics

**Lexical Database:** WordNet, GermaNet

**Contact:** Daniela Alina Plewe,

mail@danielaplewe.de,

+49-30-6141497 or +49-172-3116388.

### Online Ressources

- THE APACHE SOFTWARE FOUNDATION (2004). mod perl Homepage. <http://perl.apache.org/>, accessed April 2004 .
- THE APACHE SOFTWARE FOUNDATION (2004). Apache HTTP Server Project. <http://httpd.apache.org/> [accessed April 2004].
- GERMANET-PROJEKT (2003). GermaNet Homepage. Universität Tübingen, Seminar für Sprachwissenschaft. <http://www.sfs.uni-tuebingen.de/lsd/> [accessed April 2004].
- FIELDING, R. ET AL. (1999). Hypertext Transfer Protocol -- HTTP/1.1. The Internet Society, Internet Engineering Task Force (IETF), Network Working Group, Request for Comments Nr. 2616, June 1999. <http://www.ietf.org/rfc/rfc2616.txt> [accessed April 2004].
- INTERNATIONAL HERALD TRIBUNE (2004). International Herald Tribune. IHT Online Homepage. <http://www.ihf.com/> [accessed April 2004].
- LIPPAN, R. (2003). DBD-mysql. <http://search.cpan.org/dist/DBD-mysql/> [accessed April 2004].
- MILLER, G. A. ET AL. (2004) WordNet - a Lexical Database for the English Language. <http://www.cogsci.princeton.edu/wn/> [accessed April 2004].
- MYSQL AB (2004). MySQL Database Server Homepage. <http://www.mysql.com/> [accessed April 2004].
- THE PERL FOUNDATION (2004). The Perl Directory. <http://www.perl.org/> [accessed April 2004].
- PLEWE, D. A. (2004). GeneralNews Website. <http://www.generalnews.de/>, [accessed April 2004].
- RENNIE, J. (2003). WordNet::QueryData Module. <http://ai.mit.edu/people/jrennie/WordNet/>, <http://search.cpan.org/dist/WordNet-QueryData/> [accessed April 2004].
- WORLD WIDE WEB CONSORTIUM (W3C) (2003). Naming and Addressing: URIs, URLs, ... . <http://www.w3.org/Addressing/> [accessed April 2004].
- WORLD WIDE WEB CONSORTIUM (W3C) (2004). HyperText Markup Language (HTML) Home Page. <http://www.w3.org/MarkUp/> [accessed April 2004].