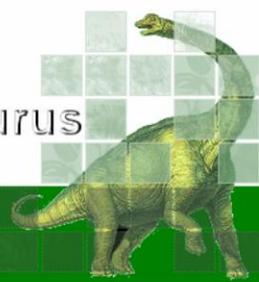


Morphoogle - A Multilingual Interface to a Web Search Engine Morphosaurus

Freiburg University Hospital¹ Jena University² Pontifical University of Paraná³
 MediLOG JULIE HTTP



Philipp Daumke¹, Stefan Schulz^{1,3}, Kornél Markó^{1,2}

Morphoogle is a query translation and expansion tool which allows users to search for biomedical content in different languages in the web. This approach is based on **Morphosaurus**, a system which transforms medical text into a language independent interlingua.

This underlying procedure is named **Morpho-Semantic Indexing (MSI)**, a term normalization methodology developed by the authors, which deals with various morphological processes in different languages. MSI uses a special type of dictionary, whose entries consist of **subwords**, i.e. semantically minimal units. Subwords are grouped into **language independent equivalence classes**, represented by Morpheme identifiers (MIDs). A morphosyntactic parser extracts subwords from texts and assigns MIDs in a three step procedure (cf. Figure 1).

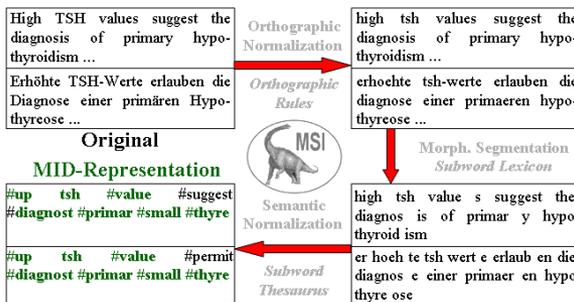


Figure 1: Morpho-Semantic Indexing (MSI)

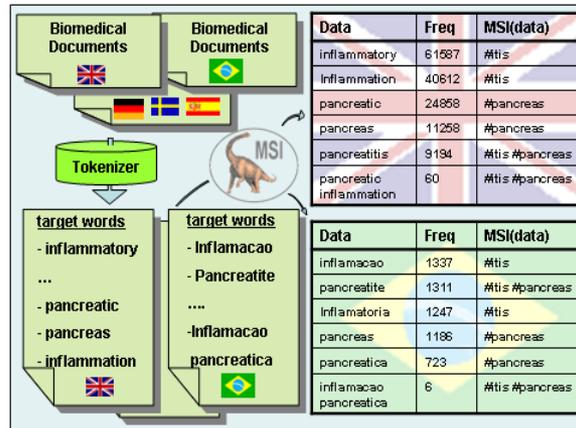


Figure 2: Generation of target word databases

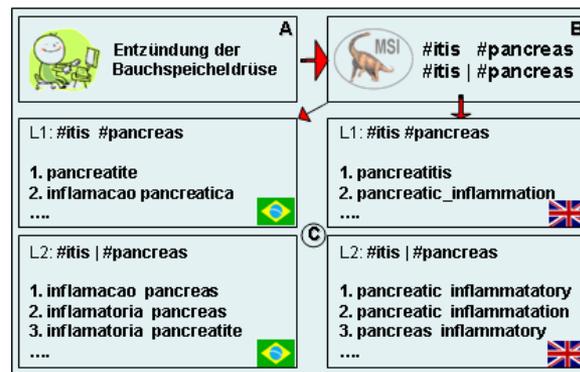


Figure 3: Output of the dictionary (here in Portuguese and English)

Morphoogle

We acquired **domain and language specific corpora** from various medical sources in the WWW. Using a tokenizer we then created large **lists of surface words, bigrams and trigrams of adjacent words** containing their frequencies within these corpora (target words). All target words are translated to a set of MIDs and stored in language specific databases. These databases consist of about 3 M entries each (cf. Figure 2).



1: **Pancreatitis**
 Pancreatitis is **inflammation** of the pancreas, a glandular organ that ...
 Necrotizing pancreatitis (which involves death of **pancreatic** tissue) can lead to ...
www.umm.edu/altmed/ConsConditions/Pancreatitiscc.html - 43K - Cached - Similar pages 0 291| 613 111

2: **Pancreatitis (inflammation) in Dogs**
 Acute pancreatitis is a sudden onset of **pancreatic inflammation**. Causes.
 Multiple factors can contribute to the development of pancreatitis. ...
www.peteducation.com/article_cfm?cls=2&cat=1580&articleid=335 - 40K - Cached - Similar pages 0 311| 3413 31

A user can send a **query** via a web interface. Again, this query is firstly altered to a **set of corresponding MIDs**. This MID set is used to create a list of possible reading variants (partitions) (cf. Figure 3). Each partition consists of one or more subwords which are now **compared to the relevant databases**. All matching records are finally sorted using several heuristics and sent to a search engine (e.g. Google) (cf. Figure 4).

www.morphosaurus.net -> Web Tools -> Morphoogle