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**

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).

**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**

(Mhere in Portuguese and English)