Multilingual Information Retrieval with the UNL

University of Geneva - UNDL Foundation - UNITAR

One of the major problems that information retrieval systems face on the internet is the linguistic diversity of the information sources. Most of the current search engines are unable to find documents that are not in the same language as the user query language. Providing users with a multilingual information retrieval system is more and more needed because
• there is no common international language
• users can often read and understand documents in several different languages
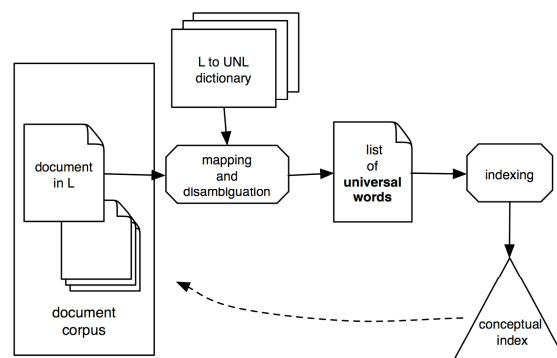• documents in non-dominant languages are currently hidden to most users

System Features

• users can express queries in any one of the available languages

• the system retrieves all the relevant documents, in all languages.

Conceptual indexing
(instead of language dependent word indexings, as Google and others do)

• create an language-independent index of the concepts that appear in the documents.

• one entry for each concept (represented by a UNL universal word),

• associates the concept with the set of documents in which it occurs

Querying