Bilingual Information Retrieval with DesIRe and Internet Translation Services

Norbert Gövert*

University of Dortmund, Germany

1 Introduction

DesIRe is the *Dortmund extensible structured Information Retrieval engine*[1]. Its extensibility is based on the implementation of physical data independence; it's query interface consists of datatypes with respective search predicates. This concept enabled us to add bilingual search predicates for the datatypes `Text::English` and `Text::German` (for English and German text, respectively). Our implementation uses free Internet resources for translating topics from English to German and vice versa.

2 Search predicates for bilingual retrieval

Having a system which is extensible w.r.t. datatypes and their respective search predicates we decided to extend the `Text::English` and `Text::German` datatypes by search predicats for bilingual text retrieval. These predicates needed to peform the translation of topics and queries from German to English in case of datatype `Text::English` and vice versa in case of datatype `Text::German`.

For translation of queries we adopted two rather naive, but fully automatic approaches. In both approaches we used free internet resources:

- Approach 1 uses the Babelfish translation service[2] of Altavista. This service allows to translate passages in a source language to a given target language. Besides the translation from German to English and vice versa, Babelfish is capable of various other languages.

---

*<goevert@ls6.cs.uni-dortmund.de>
[1]http://ls6-www.cs.uni-dortmund.de/ir/projects/DesIRe/
[2]http://babelfish.altavista.com/

- Approach 2 uses an ordinary online dictionary for word-by-word translations. We chose the Leo Dictionary service[3] for this purpose. Leo provides for a English / German dictionary with about 223 900 entries. Translations can be done in both directions. Since also composed words and phrases are included in the dictionary, we exploited this by not translating the original topics word-by-word but by interpreting each two neighbouring terms as phrases. Adopting a real naive approach we even didn't take measures in order to tackle the word disambiguation problem.
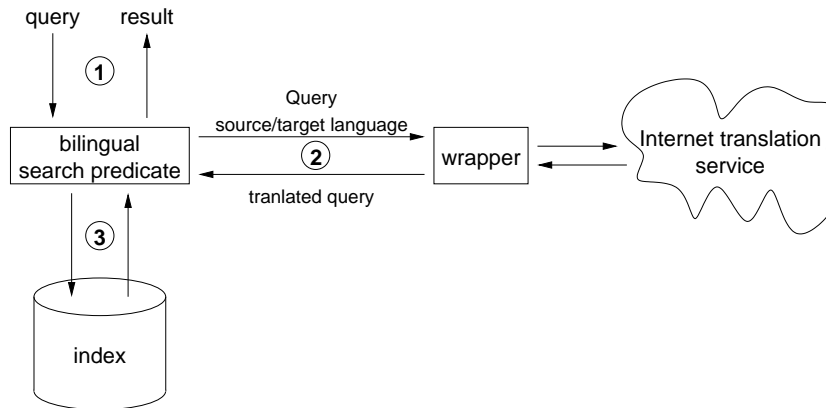
Figure 1: bilingual search predicates

Figure 1 shows the general scheme of our search predicates for bilingual text retrieval. The user gives the query in a source language, which is translated by means of a translation wrapper. The task of the wrapper is to give a uniform interface to free translation resources on the internet: It accepts the query as given by the user plus source and target language and then handles the translation through the service it was implemented for.