



ENBI -WP-11: Multilingual Access to European Biodiversity Sites through a user-friendly interface on the World Wide Web



Why to make Websites multilingual?

Although English is the language of globalisation, at present it is estimated that 85% of the Internet's content is in English, but about 45% of Internet users today cannot read English at all (on a global scale). It is expected that the non-English speaking web users will soon outnumber the English-speaking users. In 2006, one expects the Web to reach one billion users and about 70% of them will be non-English speaking.

Most web user prefer to be addressed in their native language, at least at the top-level pages of services no matter how flawed and error-ridden it may be, rather than to struggle to understand a foreign language text. User, who are addressed in their own language, will stay at a site twice as long (in this context see Fig. 1).

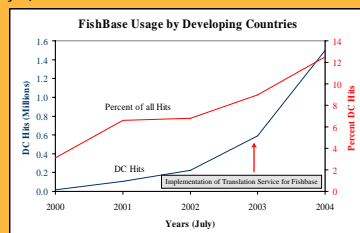
Multilingual Information Systems in the Internet

Increasingly, the expectation of user is that on-line databases should be multilingual and searchable in their own language and that the information should be translated and summarised into their own language. The European Union pays attention to this demand and was placing considerable emphasis on the development of tools for information access for all members of the community. Since Internet content is a very dynamic issue, manual translation only is hardly an option here, specifically for sites which have naturally a dynamic content with many information being updated in short intervals. Global information systems such as FishBase (www.fishbase.org) are a typical example for those dynamic sites. The wider availability of those kind of databases and information resources (particularly on the Internet) has led to the need for multilingual search and access devices with in-built translation modules (e.g. for translating search terms and/or for translating abstracts or summaries). The use of Machine Translation (MT) in this wider context is clearly due for rapid development in the near future.

Introduction WP-11, Specific Task: "Multi-lingual Access to Biodiversity Sites in the Internet"

- General strategies and techniques were developed to feature user-friendly multilingual access to European biodiversity sites through the World Wide Web (at present in major European languages: Dutch, French, German, Greek, Italian, Portuguese, and Spanish).
- Traditional "manual" translation is not an option for dynamic and/or websites with varying content. Rather, machine translation on demand has to be applied. The quality of machine translation varies greatly. Results can be tremendously improved if specialized dictionaries are available for the topic in question and avoiding idiomatic style and too complex syntax in the source text.
- The European Union is one of the longest users of MT and it is probably the largest user of MT. The Translation Service of the European Commission (EC-MT) has developed its own machine translation system, starting in the 1970s and building on the Systran engine, currently supporting 8 European languages and 18 language pairs. The service basically provides machine translation of documents for registered users.
- EC-MT Department has recently added machine translation (on-the-fly) of web pages to its services. This service has been implemented as a trial for selected pages in FishBase (species summary pages) featured from EU-MT Department. The quality of MT considerably depends on the customized activities. In order to improve the translation of specific words and terms, ENBI-WP 11 has created special biodiversity dictionaries which are integrated in the machine translation service of the European Commission.
- Other European biodiversity web sites can avail of this service by showing a "Translate" link on their pages (about the specific conditions and details, please contact the coordinator of WP-11, Bernd Uberschär, e-mail bueberschaer@ifm-geomar.de).

Fig. 1: FishBase Usage before and after implementation of the translation service in Developing Countries (Data from Logfile analysis).



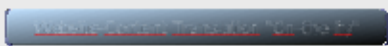
About Machine translation

The principle of machine translation: a translation of a document, from a source language into a target language, is made on the basis of a system of dictionaries and linguistic programs (e.g. SYSTRAN).

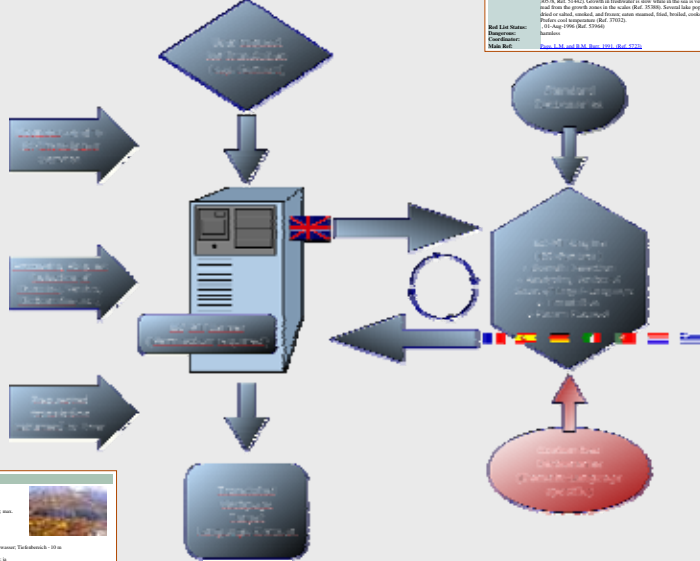
Machine translation quality

Machine translation quality mainly depends on the kind of documents (with typing errors, telegraphic style or complex syntax, the result will be poor) language similarities and on specific dictionaries available.

"Machine translation is a viable technology that can have good value. You just need to set your expectations properly so you get the most out of it. Machine translation can open an inexpensive crack in linguistic barriers that would otherwise require costly human translation to scale".



FishBase Summary Page about *Salmo salar* in the standard Language of FishBase (English)



FishBase Summary Page about *Salmo salar* in the target Language (here German), translation conducted by MT.

How does it work? Flowchart on Machine Translation

The plot: A user wants to see an English species summary in FishBase on *Salmo salar* in his own language. She or he is clicking on a link in the page and indicating the translation into the desired language, in this case German. The request is transmitted to the EC-MT server and processed by

- analysing the syntax of the source and target language, applying proper translation under consideration of the rules of the target language;
- analysing the nature of the source text and identifying assigned Domains (a Domain is specifying the "nature" of the source text, e.g. from Business, Biology, Engineering etc. the source text provider usually sets a proper Domain);
- consultation of the standard dictionaries and, under consideration of the Domains which are identified for this species summary (e.g. Biology, Marine Science), comparison and correction with terms from the customized dictionaries;
- returning the translated page to the users display
- and all of this needs only a few seconds!