

European Network for Biodiversity Information (ENBI)



Document with recommendations on how to further develop multilingual access to biodiversity data

Best practice guidelines for building and maintaining multilingual websites and services

Work package: Wp 11: Multilingual Access

Principal contractor: Dept. of Marine Ecology, Leibniz-Institut für Meereswissenschaften, IfM-GEOMAR

Deliverable: D11.9

Deliverables included in this report:

D11.3, D11.4, D11.5, D11.6, D11.7 & D11.8

Date: 15 December 2005

ENBI report No: WP11_D11.9_12/2005

Compiled by: Bernd Ueberschär & Rainer Froese, Düsternbrooker Weg 20, 24105 Kiel, Germany. Contact: e-mail: bueberschaer@ifm-geomar.de

Version: 1.0

ENBI is an EC supported Thematic Network. Contract No: EVK2-CT-2002-20020. Duration: 36 months; start 01/01/2003 end 31/12/2005. Co-ordinator: University of Amsterdam, The Netherlands. Contact: www.enbi.info

Thematic Network: European Network for Biodiversity Information (ENBI)

"Overcoming language barriers on the Internet"

Best practice guidelines for building and maintaining multilingual websites and services

**Report from ENBI-WP-11:
Strategies & Techniques to realize Multilingual Access to European Biodiversity Sites
through a user-friendly interface on the World Wide Web**

Final Technical Report WP-11
Represents Deliverables 11.5 and 11.9 in
combination with Deliverables 11.3, 4, 6 to 11.8



B. Ueberschär, Rainer Froese & Sven Mohr,
Leibniz-Institute for Marine Science, Kiel, Germany.
In cooperation with the MT-Team of the European Commission
Directorate-General for Translation,
and with support of a European Translation Team.
Contact: e-mail: bueberschaer@ifm-geomar.de

Preface:

This report is a comprehensive compilation and amalgamation of the results from ENBI Work Package 11 in Cluster IV of the ENBI-project with the title:

"Multilingual Access to European Biodiversity Sites"

In order to make transparent where and which deliverables are used as the major elements of this report we briefly outline the publication and appearance of deliverables from WP-11 in general and in this document.

Deliverables WP-11, No., Title***D11.1: Two workshops on multilingual matters and how to translate dictionaries conducted.***

1st Workshop on "Multilingual Access to European Biodiversity Sites" conducted at the Institute of Marine Research, Kiel, Germany, 06. - 07. October 2003.

2nd Workshop on "Multilingual Access to European Biodiversity Sites" conducted at the Institute of Marine Research, Kiel, Germany, 20. – 21. September 2004.

D11.2: Report from 1st and 2nd workshop on how to translate dictionaries

Both workshop reports are stored at the ENBI-WP-11 website

<http://www.enbi.linguaweb.org/Downloads.htm> and can be downloaded as word document or PDF-File.

Results from both workshops created the basis of this report and are specifically components of the delivered customized dictionaries and of ANNEX A-C of this report.

D11.3 Beta version of translation service

A trial version of machine translation was applied to FishBase, the chosen trial system for implementation of multilinguality to biodiversity sites. This Systran trial system was subsequently replaced by the EU-Systran translation engine. (www.fishbase.org, various pages, e.g. species summaries)

D11.4: "How to proceed" workshop conducted

The third and final WP-11 "How to proceed" workshop was conducted at the 19th of January 2005 in Chania, Crete.

D11.5: Report from "How to proceed" workshop

There was no formal report delivered from this workshop, since major results of this workshop formed the basis for the strategies and techniques delivered with this report on best practice guidelines for building and maintaining multilingual websites and services. This deliverable was combined with deliverable D11.9 (see below).

D11.6: Online service for quality translation of biodiversity web pages

Static and machine translation (MT) services developed jointly with the EU-Translation Department were applied to selected systems such as FishBase, an online database on fish, www.fishbase.org, (various pages, e.g. species summaries) and to OBIS (Ocean Biogeographic Information System, www.iobis.org). The MT technique applied represents the current "State of the Art" for machine translation of website content.

In accordance with the dissemination plan as discussed at the final "How to proceed" workshop, the WP-11 manager now encourages other systems to apply the strategies & techniques to realize multilingual access. The structure of this report was designed to serve as a major instrument to convince database managers that the implementation of multilinguality to their systems is feasible. In summary, chapter 4, 5 & 6 in this report describe in detail the necessary steps for quality online translation of biodiversity web pages.

D11.7: Recommendations and prototype for biodiversity glossaries

A successful public information system will have mostly non-specialist users that are not familiar with many of the terms used, e.g., in biology. While the first goal is to avoid technical jargon as much as possible, the need to explain terms such as biodiversity, habitat, or longevity remains. In the context of multilingual access, such terms as well as the definitions need to be translated into the various languages. Machine translation in this context has fair prospects but is still to be used with caution, because out-of-context translation of terms and definitions may be misleading. Chapter 3 of this report provides a list of available online glossaries that can be considered for implementation into biodiversity information systems. Some of these glossaries provide multilinguality.

A recommendation of how different glossaries can be implemented in a website is given at <http://filaman.uni-kiel.de/search/>. This page also offers machine translation of definitions into several European languages (FishBase glossary). The online translation is based on the online service offered by the European Commission (Chapter 3).

D11.8 Recommendations and prototype for access to biodiversity information through common names.

Common names of species have a long history, far exceeding that of scientific names. Although common names sometimes are employed as market names and often apply to several species or differ for various reasons, most (non-academic) users start a database search for more information by entering the common name for a species. Moreover, the limited vocabulary of non-native English speakers often restricts their knowledge to a localized common name; thus it is not sufficient to supply the common names in English only, they should be provided in as many languages as possible. One of the tasks in WP-11 was to suggest ways to enhance user access to species information through common names in different languages. In cooperation with the Species 2000 project, the FAO lists of common names in English, German, Dutch, Spanish, Portuguese, French, Greek, Italian were encoded for the Catalogue of Life database (about 10,000 names from FAO, 5,000 names from ITIS, several hundred additional names from books; among them 10,000 for Fish and 5000 others). The compilation is available on the most recent Catalogue of Life CD-ROM and soon in the Internet (<http://www.sp2000.org/>). From there, users will find an improved access to the designated data provider (e.g. FishBase for fish) species information through common names. In Chapter 2 of this report, recommendations are depicted how to get access to common names and for potential common names providers how to submit more common names to the Catalogue of Life.

D11.9 Document with recommendations on how to further develop multilingual access to biodiversity data.

This report represents the major output of WP-11 and presents the recommendations how to further develop multilingual access to biodiversity sites. It is a comprehensive document on strategies & techniques to realize multilingual access to European biodiversity sites through a user-friendly interface on the World Wide Web and combines all other deliverables from WP-11.

ENBI and MT: What are the Implications?

The Internet has proven to be a huge stimulus for MT, with hundreds of millions of pages of text and an increasingly global — and linguistically diverse — public. What role will MT play in bridging language barriers in accessing biodiversity information in the Internet? There is no doubt, that the application of MT is needed to assist in getting information from a database in a foreign language, one that the user does not well understand and the activities conducted in ENBI can be part of the stimulus which might help to push forward the accuracy of MT for scientific websites. Since biodiversity information is rather science related results in it's nature, those resources in the Internet are predestined for MT, because their presentation can follow the simple rules how to simplify text in order to assist MT (as submitted with this report e.g. for SYSTRAN). The more texts are "standardised", the more the text is unexciting and uncreative,

the more accurate the MT output will be (and eventually the less correction by post-editing is required). Machine translation works best on standardised input, which is most appropriate for structured databases.

The ENBI project has realized, for the first time, website translation "on the fly" for a selected biodiversity information system and, in cooperation with the EU-MT Department, has given the subject much thought and tested the limits of the technology effectively. It is our hope, that other public information systems will make use of this innovative technique to present their content to a broader community of non-speaking English users. The presented report conveys all necessary strategies and techniques to help to overcome language barriers for biodiversity information systems in the Internet.

Constraints with the realisation of the tasks in WP-11

As common with such new approaches, some constraints were encountered which could not be solved completely in the given time frame of WP-11.

The connection to the EU-Translation Service for Website "on the fly" translation for biodiversity information encountered a serious delay. At the date of commencement of ENBI, this service was not yet established and it took more than two years from there to further technical developments to establish this service (major part of the problem were security reasons on the MT server software) for external use. The trial system for MT (FishBase) is now connected via PHP script and works well, however testing and improvement of the "real-time" translation of FishBase content could not yet be completed.

Further, in accordance with our dissemination plan, we would like to encourage now other systems to apply the strategies & techniques to realize multilingual access. For this reason we need to tighten now the contact to the EU-translation department, to develop strategies how other systems can get access to the EU-MT service and how other systems can participate in further developing specialized dictionaries. This effort will probably also require at least one working visit of the contact person for ENBI at the Directorate-General for Translation.

The financial resources of the ENBI project did not allowed paying for the implementation of the dictionaries into the EU-MT system, thus we are dependant on the capacity of the encoder team, which deals by courtesy with this extra work.

Content of this Document

1) Introduction	6
1.1) The need to make websites multilingual	6
1.2) Solutions come in several steps	6
1.3) ENBI WP-11: Strategies & techniques to achieve multilingual access	7
D11.8 Recommendations and prototype for access to biodiversity information through common names.....	8
2) Common names.....	8
2.1) Why is it important to have common names?	8
2.2) Where do I get common names?.....	8
2.3) How do I submit more common names?.....	8
D11.7: Recommendations and prototype for biodiversity glossaries.....	9
3) Glossaries	9
3.1) Why is it important to have glossaries?.....	9
3.2) Where do I get online glossaries?.....	9
D11.6: Online service for quality translation of biodiversity web pages	10
4) Manual translation of ‘static’ content	10
4.1) How do I implement static translation?	10
4.2) Where do I get look-up tables for static translation?	13
4.3) How do I maintain static translation?	14
5) Machine Translation.....	14
5.1) Pros and cons of machine translation	14
5.2) When do I need machine translation?.....	14
5.3) How does it work?	15
5.4) Where do I get machine translation?	18
5.5) How can I improve machine translation?	18
6) How to connect your system to the EU-MT-service	18
7) Contact information and version history	33
ANNEX A: Rules for preparing English source-text for machine translation	34
ANNEX B: Text-Editing Examples from FishBase	39
ANNEX C: Some additional considerations: Translator's Notes (workshop outcome)	42
ANNEX D: Glossary of terms used in the context of machine translation.	45

What are the key objectives of this project?

The project's main focus was the development of best-practice guidelines for the design of multilingual (and therefore essentially also multicultural) Web-based information. Specifically, the following key objectives were considered:

- i. To inventory existing strategies, techniques, and tools for multilingual text production and gain knowledge about their suitability for aiding in the creation and maintenance of multilingual Websites;
- ii. To improve the access to biodiversity information through common names;
- iii. To develop a web portal (prototype) for biodiversity glossaries in several languages;
- iv. To develop "best-practice guidelines" for building and maintaining multilingual Web sites and services;
- v. To report on practical experience with the creation and use of several multilingual websites.

1) Introduction

This document is designed to facilitate the implementation of multilingual access to biodiversity portals in the Internet. It provides the necessary framework for website translation and is supposed to help to set realistic expectations with regards to machine translation.

1.1) The need to make websites multilingual

Although English is the language of globalisation and dominates the Web, there is a need to present Internet content in other languages in order to reach more users. 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.

The Internet can be counted in hundreds of millions of pages, and it is growing at a very high rate. It is expected that the non-English speaking web users will soon outnumber the English-speaking users. If one aims at a national and global audience, it is no longer enough to translate local websites only to English. In 2006, the Web is expected to have one billion users and 70% of them will not be able to communicate in English. Research has shown that those Internet visitors who are addressed in their own language will stay at a site twice as long. Many owners of commercial Internet sites are aware of this issue and present their content bi- or multilingual. Most of these multilingual presentations are based on manual (static) translation. This service provided by wealthy companies has raised the expectation of users, who increasingly expect all relevant web services to be multilingual in presentation and functionality.

The European Union has paid attention to this demand and is placing considerable emphasis on the development of translation tools. This project therefore aimed at cooperation with the EU's Translation Department in order to make use of their experience and to develop new tools jointly to translate websites online.

1.2) Solutions come in several steps

The design and maintenance of multilingual websites requires tools and procedures well beyond what is needed for monolingual websites. Implementation of multilinguality for biodiversity web sites can be realized in several steps as outlined below.

Vernacular species names in many languages are the first step towards a multilingual information system because most users do not know scientific names or common names in other than their native language. Being able to search by non-English common names already greatly increases the number of potential users, even if the results are presented in English.

The second step is to offer glossaries to explain technical and scientific terms related to biodiversity and beyond, in English and in as many other languages as possible.

The third step is manual translation of 'static' web page content, such as titles, labels, drop-down menus, choice-lists, foot notes, help texts, etc. This approach uses placeholders (variables) in the source code of the web page and has these replaced by the appropriate terms in the selected language, derived from a look-up table.

The fourth step is manual translation of larger texts. This is, however, very demanding and may not be a viable option, considering that every change in the source document requires new translation.

The fifth step is the implementation of machine translation of larger texts. This is especially appropriate if content is huge and frequently changing, such as free-text fields with information on biology, morphology or distribution for thousands of species as for example in the trial database FishBase which was used as a trial system. Note, however, that the quality of machine translation depends from several conditions to obey and may be poor and sometimes even misleading if not appropriately applied.

1.3) ENBI WP-11: Strategies & techniques to achieve multilingual access

This document is an output of the EC supported European Network of Biodiversity Information (ENBI), a joint project of 65 European institutes in 24 countries. In that context, it was considered important to provide access to biodiversity information in the major languages of the European Community (French, Portuguese, Spanish, Dutch, German, Italian and Greek). The usefulness of a multilingual service was demonstrated by FishBase, the global information system on fish: The request for information in FishBase showed a considerable increase from developing countries since the first version of machine translation for selected resources (search page, species summary page) was implemented (Fig.1).

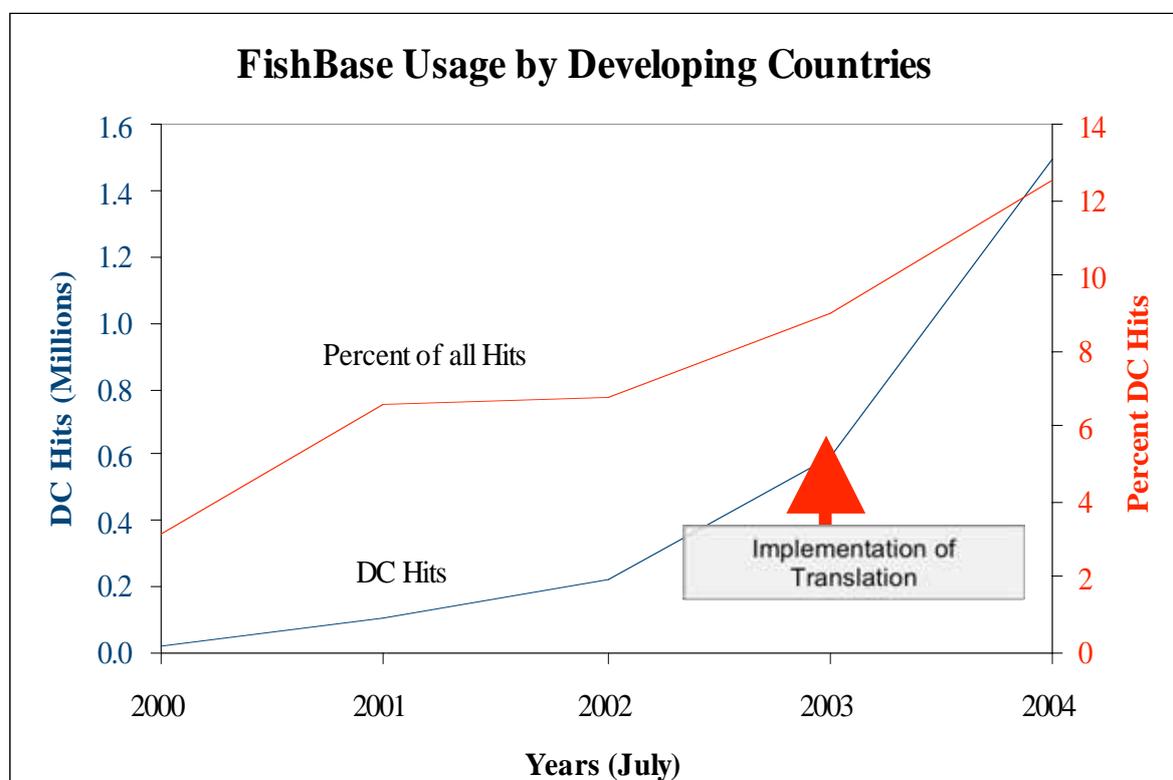


Fig. 1: Substantial increase in hits to FishBase from developing countries (DC) since a preliminary translation service from English into 7 other major European languages was established in November 2003.

D11.8 Recommendations and prototype for access to biodiversity information through common names.

2) Common names

2.1) Why is it important to have common names?

Biodiversity web sites typically contain information on species. What most people know about most species are their common names, and statistics on how user access information in FishBase confirms that most people enter common names in their respective language, even if they are capable of communicating in English. Scientific names are used only by a minority. Thus, an information system for the general public must allow access by common names in as many languages and scripts (Cyrillic, Chinese, Arabic, etc) as possible.

2.2) Where do I get common names?

The "Catalogue of Life" program produces an annual CD-ROM, which, in its 2005 edition, contained several hundred thousand common names for many of the 527,000 species covered. For further information please check <http://www.sp2000.org/> for general information and <http://annual.sp2000.org/2005/search.php> for the online Annual Checklist. The "Catalogue of Life"- CD-ROM can be ordered from Species 2000, e-mail: sp2000@sp2000.org. The underlying database is in MySQL and experienced users will be able to extract names for use in their own biodiversity websites, such as is done by www.iobis.org and www.gbif.org.

2.3) How do I submit more common names?

While the "Catalogue of Life" Annual Checklist is currently the largest compilation of common names, it is far from being complete, and actually there are no common names for the majority of species. Those users who want to contribute lists of scientific names with their respective common names may send such compilations as e.g. Excel files to sp2000@sp2000.org, who will then included new names into the next version of the Annual Checklist. See an example of such a file in Table 1.

Common Names of <i>Gadus morhua</i> [n=165]			
Common Name	Used In	Language	Type
	Russian Fed	Russian	Vernacular
Almindelig torsk	Denmark	Danish	Vernacular
An trosc	Ireland	Gaelic, Irish	Vernacular
Atlantic cod	Canada	English	Vernacular
Atlantisk Torsk	Denmark	Danish	Vernacular
Bacalao	Can Newfounl	English	Vernacular
Bacalao	Spain	Spanish	Vernacular
Bacalao del Atlántico	Spain	Spanish	FAO
Bacaleau	Can Newfounl	English	Vernacular
Bacalhau	Portugal	Portuguese	Vernacular
Bacalhau-do-Atlântico	Portugal	Portuguese	Vernacular
Bacallà	Spain	Catalan	Vernacular
Baccalao	Can Newfounl	English	Vernacular
Baccale	Can Newfounl	English	Vernacular
Baccalo	Can Newfounl	English	Vernacular
bakajar	Croatia	Croatian	Vernacular
bakal	SerbiaMontenegro	Serbian	Vernacular
bakalaj	SerbiaMontenegro	Other	Vernacular
bakalar	Croatia	Croatian	Vernacular

Table 1: Example of a list of common names for *Gadus morhua* (Atlantic Cod) as contained in FishBase. (Note: the list shown here is not complete).

D11.7: Recommendations and prototype for biodiversity glossaries

3) Glossaries

3.1) Why is it important to have glossaries?

A successful public information system will have mostly non-specialist users that are not familiar with many of the terms used, e.g., in biology. While the first goal is to avoid technical jargon as much as possible, there remains the need to explain terms such as biodiversity, habitat, or longevity. In the context of multilingual access, such terms as well as the definitions need to be translated into the various languages. Machine translation in this context has fair prospects but is still to be used with caution, because out-of-context translation of terms and definitions may be misleading.

3.2) Where do I get online glossaries?

There are a number of online biodiversity glossaries accessible in the Internet, some even in several languages (Table 2). For example, FishBase has developed a glossary with more than 2500 terms which are useful for the entire aquatic range, with translations to French, Spanish, Portuguese and Russian. The glossary can be accessed from other websites through a simple URL, such as <http://www.fishbase.org/Glossary/Glossary.cfm?TermEnglish=abyss>.

An example of how different glossaries can be implemented in a web site is given at <http://filaman.uni-kiel.de/search/>. This page also offers machine translation of definitions into several European languages (FishBase glossary). The online translation is based on the online service offered by the European Commission.

In addition, some popular search engines like Google may sometimes help to find annotated information about unknown terms and abbreviations.

A glossary of terms used in the context of machine translation is attached as ANNEX C.

Glossary Resource	URL	Content
FishBase Glossary	http://www.fishbase.org	Glossary on definitions related to ichthyology, taxonomy, ecology, conservation, population dynamics, genetics, oceanography, geography and related disciplines.
Multilingual ENBI Glossary	http://filaman.uni-kiel.de/search/	ENBI Glossary Portal with the option on machine translation from English into 7 languages (Trial version).
EEA Glossary:	http://glossary.eea.eu.int/EEAGlossary/	Multilingual Glossary of the European Environmental Agency
Encyclopedia Britannica:	http://www.britannica.com/search?query=teste&ct=&fuzzy=N	Highly authoritative source.
EPA Terms of Environment:	http://www.epa.gov/epahome/search.html	Glossary of the U.S. Environmental Protection Agency
Tree of Life:	http://tolweb.org/tree/home.pages/search/TOL?taxon=yolk&Submit2=Find&select=partially&taxonClade=	Glossary of biological terms
LingInfo:	http://www.mblogic.net/glossary/default.asp?txtLang=english&txtSearch=yolk&field_term=&field_abbr=&field_def=	Glossary of molecular biology terms and abbreviations
FAO Fisheries Glossary	http://www.fao.org/fi/glossary/default.asp	A multilingual glossary on typical terms from fishery

Table 2: A list of online glossaries related to information on Biodiversity.

D11.6: Online service for quality translation of biodiversity web pages

4) Manual translation of 'static' content

Manual translation refers to a human translating a given text and entering the result into the computer by hand. This translation is then thought to be static in the sense that the content and its translation will not change very often. This approach is suitable for web page content such as titles, labels, menus, choice fields, foot notes, and other text that is part of the 'frame' of a web page. It is less suited for short-lived text information such as a *What's new?* section or for huge amounts of non-standardized text information such as may be derived from free-text fields of a database.

4.1) How do I implement static translation?

- (1) First, take stock of who will do translation for your web site, and how much time these translators will be willing to spend on it. Note that there will be two phases of manual translation: a first *high-workload-limited-duration* phase where every term and phrase needs to be translated in all languages and stored in respective look-up tables before the launch of the multilingual service. Then a second *low-workload-fast-response-indefinite-duration* phase where new or altered terms and phrases need to be translated as soon as possible over the foreseen lifetime of the service.
- (2) Depending on what support you can realistically expect from your translators, decide which parts of your website will be translated. This will typically be the most-used pages aiming at a broader audience. In contrast, pages that will only be used by specialists or members of your project who are likely to be fluent in English will not need translation.

- (3) For those sections of the database that need translation, decide whether manual or machine translation is more suitable. For some pages an amalgamation of manual and machine translation may be useful.

For static translation, use "placeholders" for titles, headers, labels, choice fields and notes in the source code. A placeholder is a variable that will be replaced by a text string from the look-up table in the requested language (see below). A placeholder control is used to reserve the respective space in the page when displayed. Only one lookup-table is needed for various pages.

An example of a look-up table from FishBase is given below (Table 3). A look-up table is a tabular data file containing the translated terms in the various languages. Such a table can be a mix of words, terms and even whole sentences, such as used e.g. in the search page of FishBase (<http://www.fishbase.org/search.php>)

eng_term	port_term	ger_term
Conservation	Conservação	Arterhaltung
Geography and climate	Geografia e Clima	Geographie und Klima
Hydrography	Hidrografia	Hydrographie
spines	espinhos	Stacheln
soft rays	raios moles	Weichstrahlen
Uses and Status	Usos e Status	Nutzung und Status
Freshwater	Água doce	Süßwasser
Reported	Reportado	Berichtet
Potential	Potencial	Potential
Species used in fisheries	Espécies haliêuticas	Fischerei Arten
Species used in aquaculture	Espécies utilizadas em aquicultura	Aquakultur Arten
Species exported live	Espécies exportadas vivas	Lebend exportierte Arten
Species used as sport fish	Espécies utilizadas na pesca desportiva	Sportfisch Arten
Endemic species	Espécies endêmicas	Einheimische Arten
Threatened species	Espécies ameaçadas	Bedrohte Arten
Protected species	Espécies protegidas	Geschützte Arten
Available key information	Informação-chave disponível	Wichtige Informationen
food	alimento	Nahrung
Occurrence	Ocorrência	Vorkommen
Hits/month million	visitas/mês milhão	Hits/Monat Millionen
How to cite FishBase	Como citar a FishBase	Wie man FishBase zitiert
Cite FishBase itself as Mirrors	Cite a FishBase deste modo Servidores	FishBase wird zitiert als Server
Note: Lists may be incomplete. Some lists may be very long and will take time to load. loading may take 2-4 mins.	Nota: As listagens podem estar incompletas. Algumas listagens podem ser longas e, portanto, demorar a carregar. carregamento pode durar 2-4 minutos	Anmerkung: Listen können unvollständig sein; manche Listen sind sehr lang und brauchen Zeit zum Laden. Laden kann 2-4 Minuten dauern
References Citing FishBase	Referências citando a FishBase	Referenzen die FishBase zitieren

Note: Use the Back button of your browser to return to FishBase.	Nota: Use o botão PARA TRÁS do seu Browser para retornar à FishBase.	Anmerkung: Mit 'Zurück' des Browsers zurück zu FishBase.
Fish Forum	Fish Forum	Fisch Forum

Table 3: Example of a look-up table from FishBase. Translation from English (source text) to Portuguese and German is shown. Please note that only a subset of the table is shown here.

- (4) Assemble a team of reliable translators for the manual translation into the desired languages and secure their long-term commitment (see also item 4.4). Note that the translators need to be familiar with biodiversity and biology, otherwise the quality of the translation will suffer or the translators will need too much time. Also, the translators need to see the terms in their respective context: the same term may need different translations in different contexts. As much as possible, such ambiguities should be avoided in the source text. The above glossaries may also prove useful in this process, and there are additional Internet sources such as the European Terminology Database at <http://europa.eu.int/eurodicautom/Controller> which can assist in proper translation of special terms.
- (5) Apply a script (e.g. in PHP), which handles the technical linking of terms, phrases sentences etc. and the placeholders. The following PHP-script demonstrates how look-up tables are connected to place holders in FishBase (Table 4). The script is specific for the conditions in FishBase, but can be fitted to other databases as well.

```

Script from FishBase: Connection of look-up tables and placeholders
<?php
// Language setting
if ( !isset($_SESSION['Language']) ) {
    session_register('Language'); // Set default language
    $Language = $_SESSION['Language']='English';
} else { // Set language selected by the user
    $Language = $_SESSION['Language'];
}

if ( isset($_GET['lang']) ) {
    $Language = ucfirst(strtolower($_GET['lang']));
}

if ( !(strcmp($_SESSION['Language'], $Language)==0) ) {
    // Save language selected by the user
    $_SESSION['Language'] = $Language;
}

if ($Language != "English") {
    if ($Language == 'Chinese') {
        $qry = "select Label_Index.term_".
            $Language.",term_English, ".
            "term_Chinese_u from Label_Index";
    } else {
        $qry = "select Label_Index.term_". $Language.
            ",term_English from Label_Index";
    }
}

$s = $conn->query($qry);

while( $row = $s->fetch_assoc() ) {
    $term_English = $row['term_English'];

    if ($Language == "Chinese") {
        $term_Chinese_u = trim($row['term_Chinese_u']);
        $vterm=($term_Chinese_u=="")?$row['term_English']:$term_Chinese_u;
    } else { // $Language is not "Chinese"

```

```

    $tx = "term_" . $Language ;
    $vartemp = $row["$tx"];
    if ( $vartemp == ' ' ) {
        $vterm = $row['term_English'];
    } else {
        $vterm = $vartemp;
    }
}

if ($term_English == "Note: FishBase is also available ".
    "on CD-ROM/DVD, see the FishBase ".
    "homepage or the FishBase book ".
    "for more information. FishBase ".
    "was created with the help of ".
    "many partners and with support ".
    "from the EC and other donors") {
    $fishbaseis_alsoavailable_u = $vterm;
}

if ($term_English == "To give due credit to the ".
    "original authors, please cite ".
    "data taken from FishBase by ".
    "Main Ref. and/or Data Ref. ".
    "of the respective record") {
    $togive_duecredit_u = $vterm;
}

$xterm = trim($term_English);

$mirrors_u=strcasecmp("Mirrors",$xterm)==0?$vterm:$mirrors_u;
$Glossary_u=strcasecmp("Glossary",$xterm)==0?$vterm:$Glossary_u;
$Collection_sites_by_1_u=strcasecmp("Collection sites by 1-degree
EAG",$xterm)==0?$vterm:$Collection_sites_by_1_u;
$Collection_sites_by_2_u=strcasecmp("Collection sites by 2-degree
EAG",$xterm)==0?$vterm:$Collection_sites_by_2_u;
$Collection_sites_by_10_u=strcasecmp("Collection sites by 10-degree
EAG",$xterm)==0?$vterm:$Collection_sites_by_10_u;

.
.
.
}
$s->close();
}
?>

//how to display the actual fields
if ($Language != "English") {
    echo $mirrors_u . ":";
} else {
    echo "Mirrors:";
}

if ($Language != "English") {
    echo $Glossary_u;
} else {
    echo "Glossary";
}
}

```

Table 4: How to connect look-up tables and placeholders with the output page in a database such as FishBase using PHP.

4.2) Where do I get look-up tables for static translation?

The complete FishBase look-up table can be downloaded in Excel-format from http://www.enbi.linguaweb.org/Downloads/fb_lookup_table.xls and can be used as a starting point for other web portals. Translations of country and ecosystems names may prove especially useful. Other terms are fish- or context-related and may not be suitable for other

portals. The look-up table prepared for OBIS (Ocean Biogeographic Information System, <http://www.iobis.org/Welcome.htm>) in 8 languages is also provided at http://www.enbi.linguaweb.org/Downloads/obis_lookup_table.xls).

4.3) How do I maintain static translation?

Since every change in your website now needs translation, a list-server-type email approach of sending out new terms to all translators needs to be established; the mail will contain the term(s) to be translated, a link that on click shows the respective context, and a link that opens the translation table for direct remote data entry. If it is to function in the long term, this has to be a quick and easy procedure for your webmaster in creating the respective email as well as for the translators entering the translated terms. For FishBase, such a *translation-by-email* system is under development; please contact Bernd Ueberschär (bueberschaer@ifm-geomar.de) or Rainer Froese (rfroese@ifm-geomar.de) for more information.

5) Machine Translation

5.1) Pros and cons of machine translation

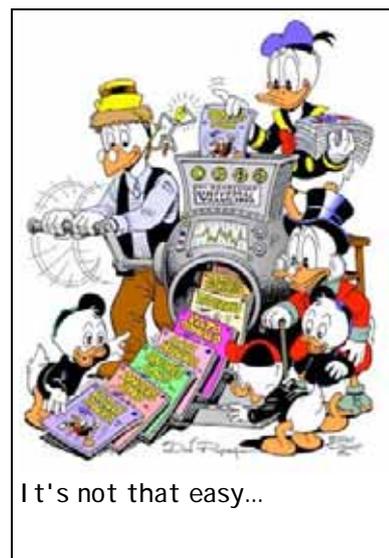
Machine translation (MT) has still a long way to go to match the quality provided by professional human translators. The main advantage of MT is high speed and low cost. Machine translation works best if the source text is of a technical nature, with simple phrases, standardized vocabulary, and dedicated dictionaries for use by the translation software. If a telegraphic style or a complex syntax is used, or if there are typographical errors, the results of MT will be poor. Users have to be made aware that they are looking at text translated by a machine, so that they are prepared to lower their expectations, bear with 'unusual' phrases, and watch out for misleading translations.

What users can expect from machine translation was expressed by Brian Garr (former IBM Computer Department) as follows:

"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."

5.2) When do I need machine translation?

Static translation of labels and choice-fields goes a long way and several such fields can be connected or joined with numbers to form pseudo-sentences that will look acceptable in most languages. However, if your information system contains several thousand records with free-text fields, manual translation is not an option and machine translation is then the only possible method. The FishBase species summary page is an example of applying both, manual and machine translation. Manual translation is applied to titles, labels and choice fields, including some connected sentences; machine translation is applied for free-text fields on distribution, morphology and biology of 29,000 species of fishes. Figure 3 & 4 below show an example for a summary page in English and subsequently translated to German. Much of the text is drawn from the translation table (look-up table), but the fields *Verteilung*, *Morphologie* und *Biologie* are machine-translated.



The best intervention to achieve reasonable translation quality is the standardization of the source text, following the rules which were designed e.g. for the SYSTRAN software to aid in the preparation of English text for machine translation (see following paragraphs and ANNEX

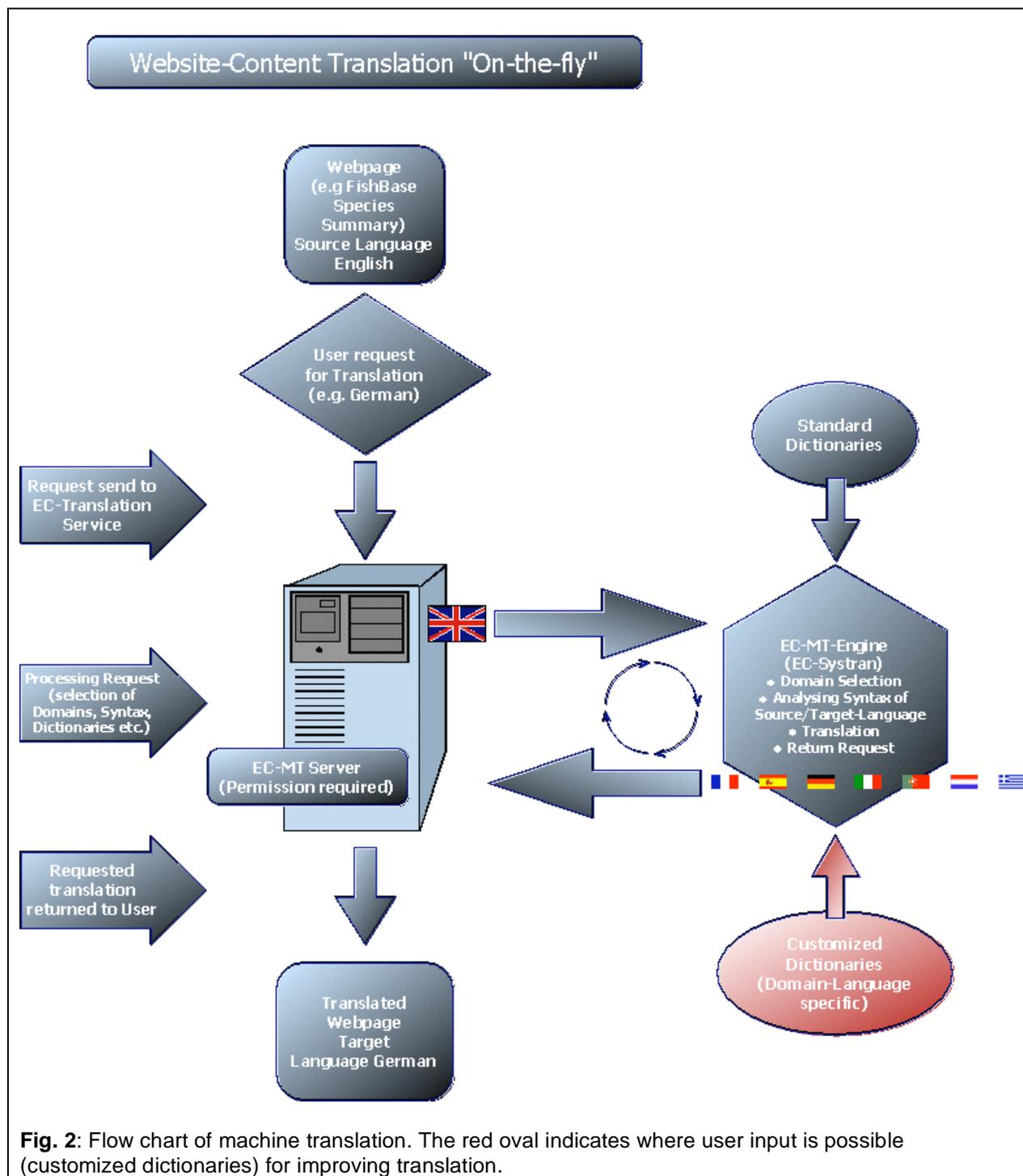
A). Another strong measure for improving translation quality is to provide system-specific dictionaries (see 5.4).

5.3) How does it work?

The translation of a text from a source language into a target language is done by translation software such as SYSTRAN with the aid of a system of customized dictionaries. The translation engine of the software knows the grammar of the various languages and tries to select appropriate dictionaries for, e.g., business or technical documents. Please note, that translations between similar languages are usually better.

Figure 2 depicts a flowchart of the processes involved: A user wants to see the species summary page for *Salmo salar* in another language. The request is transmitted to the FishBase server, which replaces titles, labels, choice fields, etc. with the respective, manually translated terms and forwards the text of the Distribution, Morphology and Biology fields to the machine translation service of the European Commission. This software processes the text in just a few seconds by

- (1) analysing the syntax of the source;
- (2) analysing the nature of the source text and assigning domains or categories such as Business, Biology, or Engineering (can also be provided with the translation request);
- (3) consulting the standard dictionaries and, under consideration of the domains or categories, the special or customized dictionaries;
- (4) and translating and returning the text;



<u><i>Salmo salar</i></u> Linnaeus, 1758	
Family:	Salmonidae (Salmonids), subfamily: Salmoninae
Order:	Salmoniformes (salmons)
Class:	Actinopterygii (ray-finned fishes)
FishBase name:	Atlantic salmon
Max. size:	150 cm TL (male/unsexed; Ref. 7251); 120 cm TL (female); max. published weight: 46.8 kg (Ref. 41037); max. reported age: 13 years
Environment:	benthopelagic; anadromous (Ref. 51243); freshwater; brackish; marine; depth range - 10 m
Climate:	temperate; 2 - 9°C; 72°N - 37°N, 77°W - 61°E
Importance:	fisheries: highly commercial; aquaculture: commercial; gamefish: yes
Resilience:	Medium, minimum population doubling time 1.4 - 4.4 years (K=0.29-0.76; tm=3-5; tmax=14; Fec=8,000)
Distribution:	Atlantic Ocean: temperate and arctic zones in northern hemisphere (Ref. 51442). In western Atlantic Ocean distributed in coast drainages from northern Quebec in Canada to Connecticut in USA (Ref. 5723). In eastern Atlantic Ocean distributed in drainages from the Baltic states to Portugal (Ref. 51442). Landlocked stocks are present in Russia, Finland, Sweden and Norway (Ref. 6439) and in North America (Ref. 1998). Appendix III of the Bern Convention (protected fauna; except at sea).
Morphology:	Dorsal spines (total): 3-4; Dorsal soft rays (total): 9-15; Anal spines : 3-4; Anal soft rays : 7-11; Vertebrae : 58-61. Fusiform body (Ref. 51442). Mouth extends only to area below rear of eye and has well developed teeth (Ref. 51442). Vomerine teeth weak (Ref. 7251). Caudal fin with 19 rays (Ref. 2196). Little scales (Ref. 51442). Adults are blue-green colored with a silvery coating and a few spots in salt water; no spots under lateral line (Ref. 37032, Ref. 51442). During reproduction period, in fresh water, it loses the silvery guanine coat and becomes greenish or reddish brown mottled with red or orange, certainly the males (Ref. 37032, Ref. 51442). Few black spots on body, caudal fin usually unspotted and adipose fin not black bordered. Juveniles have 8 to 12 blue-violet spots on the flanks with little red spots in-between (Ref. 51442). Also Ref. 3137.
Biology:	Amphihaline species, spending most of his life in fresh water (Ref. 51442). Young remain in freshwater for 1-6 years, then migrate to the ocean and remain there for 1-4 years before returning to freshwater. It grows up at sea on the continental plate west of Greenland (Ref. 51442). Adults then return to the river of their origin to spawn (Ref. 51442). It returns to sea after spawning, but a lot of adults already die (Ref. 51442). Active during the day. Juveniles feed mainly on aquatic insects, mollusks, crustaceans and fish; adults at sea feed on squids, shrimps, and fish (Ref. 51442). Adults in freshwater which are approaching the reproductive stage do not feed (Ref. 30578, Ref. 51442). Growth in freshwater is slow while in the sea is very rapid. Life history of the salmon can be read from the growth zones in the scales (Ref. 35388). Several lake populations are landlocked. Marketed fresh, dried or salted, smoked, and frozen; eaten steamed, fried, broiled, cooked in microwave, and baked (Ref. 9988). Prefers cool temperature (Ref. 37032).
Red List Status:	, 01-Aug-1996 (Ref. 53964)
Dangerous:	harmless
Coordinator:	
Main Ref:	Page, L.M. and B.M. Burr. 1991. (Ref. 5723)



Fig. 3: FishBase Summary Page about *Salmo salar* in the standard Language of FishBase (English)

<u><i>Salmo salar</i></u> Linnaeus, 1758	
Familie:	Salmonidae (Lachse), subfamily: Salmoninae
Ordnung:	Salmoniformes (Lachsfische)
Klasse:	Actinopterygii (Strahlenflosser)
FishBase Name:	Atlantic salmon
Max. Größe:	150 cm TL (Männchen/unbestimmt; Ref. 7251); 120 cm TL (female); max. veröff. Gewicht: 46.8 kg (Ref. 41037); max. veröff. Alter: 13 Jahre
Lebensraum:	benthopelagisch; anadrom (Ref. 51243); süßwasser; brackwasser; seewasser; Tiefenbereich - 10 m
Klimazone:	gemäßigt; 2 - 9°C; 72°N - 37°N, 77°W - 61°E
Bedeutung:	Fischereien: hoch kommerziell; Aquakultur: kommerziell; Sportfisch: ja
Widerstandsfähigkeit:	mittel, Verdopplung der Population dauert 1,4 - 4,4 Jahre. (K=0.29-0.76; tm=3-5; tmax=14; Fec=8,000)
Verbreitung:	Atlantischer Ozean: mäßige und arktische Zonen in nördlicher Hemisphäre (Ref. 51442). In westlichem atlantischem Ozean, der in Küstengewässern aus Nordquebec in Kanada an Connecticut in USA verteilt wird, (Ref. 5723). In östlichem atlantischem Ozean, der in Entwässerungen aus den baltischen Staaten an Portugal verteilt wird, (Ref. 51442). Landumschlossene Bestände liegen in Rußland, Finnland, Schweden und Norwegen vor (Ref. 6439) und in Nordamerika (Ref. 1998). Anhang III der Konvention von Bern (geschützte Fauna; außer auf See).
Morphologie:	Rückenflossenstacheln (insgesamt): 3-4; Rückenflossenweichtstrahlen (insgesamt): 9-15; Afterflossenstacheln 3-4; Afterflossenweichtstrahlen: 7-11; Wirbelzahl: 58-61. Spindelförmiger Körper (Verweis 51442). Öffnung verlängert nur auf Bereich unterhalb der Rückseite des Auges und hat gut entwickelte Zähne (Verweis 51442). Vomerine Zähne schwach (Verweis 7251). Schwanzflosse mit 19 Strahlen (Verweis 2196). Wenig Skalen (Verweis 51442). Die Erwachsenen sind gefärbt mit einer silbrigen Schicht und einigen Punkten im Salzwasser blaugrünes; keine Punkte unter seitlicher Linie (Verweis 37032, Verweis 51442). Während der Wiedergabeperiode im Süßwasser, verliert sie den silbrigen guanine Mantel und wird grünlich oder gesprinkelt mit Rotem oder Orange, zweifellos die Männer rötlichbraunes (Verweis 37032, Verweis 51442). Wenige Schwarzpunkte auf Körper. Schwanzflosse unspotted normalerweise und nicht Schwarzes der fetthaltigen Flosse eingefärbt. Jugendliche haben 8 bis 12 blau-violette Punkte auf den Flanken mit kleinen roten Punkten in-between (Verweis 51442). Auch Verweis 3137.
Biologie:	Amphihaline-Arten, die den größten Teil seines Lebens in Süßwasser verbringen, (Ref. 51442). Junge bleiben im Süßwasser für 1-6 Jahre, also wandern zum Ozean aus und bleiben dort für 1-4 Jahre vor dem Zurückkehren zum Süßwasser. Es wächst auf See auf der kontinentalen Platte westlich von Grönland (Ref. 51442). Erwachsene kehren dann zum Fluß ihres Ursprungs hervorzubringen zurück (Ref. 51442). Er kehrt zum Meer nach dem Hervorzubringen zurück, aber viele Erwachsene sterben schon (Ref. 51442). Aktiv während des Tages. Jugendliche fressen hauptsächlich Wasser-Insekten, Mollusken, Krebstiere und Fische; auf See Erwachsene fressen Kalmare, Garnelen und Fische (Ref. 51442). Erwachsene im Süßwasser, die sich dem reproduktiven Stadium nähern, speisen nicht (Ref. 30578, Ref. 51442). Das Wachstum in Süßwasser ist langsam, während im Meer sehr schnell ist. Lebensgeschichte des Lachses kann von den Wachstumszonen in den Maßstäben gelesen werden (Ref. 35388). Mehrere Seebevölkerungen sind landumschlossen. Vermarktetes frisch, getrocknet oder gesalzen, geräuchert und eingefroren; gegessen gedämpft, gebraten, gebacken, gekocht in Mikrowelle und gebacken (Ref. 9988). Zieht kühle Temperatur vor (Ref. 37032).
Rote Liste:	, 01-Aug-1996 (Ref. 53964)
gefährlich:	harmlos
Koordinator:	
Hauptref:	Page, L.M. and B.M. Burr. 1991. (Ref. 5723)



Fig. 4: FishBase Summary Page about *Salmo salar* in the target Language (here German), translation conducted by MT

5.4) Where do I get machine translation?

Several machine translation services are available in the Internet such as Systran (<http://www.systransoft.com/index.html>), LINGUATEC (<http://www.linguec.de>), and BabelFish (<http://babelfish.altavista.com>). They allow online translation by submitting words, phrases and even whole paragraphs. Another option is to submit a URL request translation for an entire web page. These options are useful for trial purposes, but do not allow any user interactions or the implementation of special dictionaries and the resulting translation is of highly variable quality. The professional version of these services can be hired for websites, but the fee for regular usage can amount up to €25,000 per year, depending on the number of translation requests.

The EC developed its own machine translation service based on the SYSTRAN engine (EC-SYSTRAN) that has been running since the 1970s. This service is supposed to be available free of charge for European non-profit projects or initiatives. Using this service requires an individual agreement with the EU-MT Service, you can make contact with the responsible ENBI contact person in the MT-Department. However, it is advisable to get in touch with the WP-11 manager (bueberschaer@ifm-geomar.de) for further instructions before trying to contact the EU-MT Department.

5.5) How can I improve machine translation?

Source text:

Preparation for machine translation starts with the source text. Unfamiliar word combinations and sentence constructions and telegraphic style with incomplete syntax result in poor translation. The more uncreative or simple a text, the better the translation results. Also, flagging of words or phrases that should not be translated is possible. ANNEX A presents rules for source text preparation.

Customized dictionaries:

Dedicated dictionaries for different types of texts can considerably improve translation. Translation programs make use of a master dictionary, which holds only standard vocabulary for each language. But many terms have different meaning in different contexts. For example, the word "stock" refers to a population of commercially exploited fishes in the context of fisheries biology, but it refers to ownership certificates in the context of the stock market. To avoid these ambiguities, domains or categories are used which select the proper dictionaries. Several such dictionaries are already available, e.g. with the EC Translation Service, but may need to be improved for the purpose at hand. For the purpose of translating free text fields in FishBase, three new dictionaries on "Morphology", "Distribution" and "Biology" were compiled, each containing more than 1000 terms and phrases. The dictionaries needed hand-encoding by the technical team in the EC-MT-department (noun, proper noun, verb, adjective etc.) which is time consuming. If creation of new dictionaries is envisioned, financial support for this work needs to be considered.

The dictionaries were compiled for FishBase, however, they contain many terms which are of general significance for information systems dealing with biodiversity. They can form the basis for other dictionaries and can be downloaded in Excel-Format at <http://www.enbi.linguaweb.org/Downloads.htm> ("Morphology" and "Biology" are at present under revision, download comes soon).

6) How to connect your system to the EU-MT-service

The technical connection to the EC translation server is realized with a PHP or PERL script in accordance with the advisements of the MT-Department. Table 5 & 6 provide an example for FishBase. However, since each user of the EU-MT service needs a personal account and password, please contact Bernd Ueberschär (bueberschaer@ifm-geomar.de) or secure permission from the EC-MT service before considering such connection.

```
<?php
/**
 *
 * @author Sven Mohr
 * @version 0.1
 * @module fb.functions
 */

/**
 *
 */

class cDocumentInformation {
public $sourceLanguage;
public $targetLanguage;
public $terminologyDomains;
public $userDictionary;
public $isConfidential;
public $handleCelexEncoding;
public $allCoding;
public $textType;
public $refid;
}

function output_header($title,$charset)
{
echo '<?xml version="1.0"?>'. "\n";
echo '<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" ';
echo ' "http://www.w3.org/TR/xhtml1/DTD/transitional.dtd">';
echo '<html>'. "\n";
echo '<head>'. "\n";
echo '<title>'.$title.'</title>'. "\n";
echo '<meta http-equiv="Content-Type" content="text/html;';
echo ' charset='.$charset.'" />'. "\n";
echo '</head>'. "\n";
}

function start_body()
{
echo "<body>\n";
}

function end_body()
{
echo "</body>\n";
echo "</html>\n";
}

function remove_html_tags($txt)
{
return preg_replace("/<[\!\/\?]*?[^>]*?>/si", '', $txt);
}

function debug_out($msg)
{
{
if (isset($_GET["debug"])) {
echo '<p style="background-color: #ffffee;font-family: TreBuchet MS; font-size: 12px;border: 1px solid #eee;">';
echo $msg. "\n";
echo '</p>';
}
}
}

function debug_out_var($msg)
{
{
if (isset($_GET["debug"])) {
echo '<pre style="background-color: #ffffee;font-family: Courier New; font-size: 12px;border: 1px solid #eee;">';

```

```
var_dump($msg);
echo '</pre>';
}
}

function info_out($msg)
{
echo '<p style="background-color: #eefff; font-family: TreBuchet MS; font-size: 12px; border: 1px solid #eee;">';
echo $msg."\\n";
echo '</p>';
}

function error_out($msg)
{
echo $msg."\\n";
}

function get_glossary_info($word) {
global $fb_svr, $fb_user, $fb_pw, $fb_db, $fb_port;
$conn = new mysqli($fb_svr, $fb_user, $fb_pw, $fb_db, $fb_port);

$query="select * from fbapp.GLOSSARY where Glossary.TermEnglish = '$word' or
Glossary.TermFrench = '$word' or Glossary.TermSpanish = '$word' or
Glossary.TermPortuguese = '$word' or glossary.TermRussian = '$word' or
glossary.TermCyrillic_u = '$word'";
debug_out("Query : ".$query);
$Detail = $conn->query($query);
debug_out("Detail->num_rows : ".$Detail->num_rows);

while( $row = $Detail->fetch_assoc() )
{
$detail['desc']=$row["English"];
}

$Detail->close();
return $detail;
}

function get_detail_info($SpecID) {
global $fb_svr, $fb_user, $fb_pw, $fb_db, $fb_port;
$conn = new mysqli($fb_svr, $fb_user, $fb_pw, $fb_db, $fb_port);

$query = "SELECT FAMILIES.Family,
FAMILIES.Order AS DetailField3,
FAMILIES.Class AS DetailField4,
Species.Aquarium,
Species.DemersPelag,
Species.Importance,
Species.GameFish,
Species.UsedforAquaculture,
Species.Genus AS DetailField5,
Species.Species AS DetailField6,
Species.Author AS DetailField7,
Species.AuthorRef as AuthorRef,
Species.FBname AS DetailField8,
Species.comments,
Species.Fresh AS DetailField12,
Species.Brack AS DetailField13,
Species.Saltwater AS DetailField14,
Species.Entered AS DetailField16,
Species.Modified AS DetailField17,
```

```
Species.Expert AS DetailField18,  
STOCKS.StockDefs,  
stocks.IUCN_id,  
STOCKS.IUCN_Code,  
stocks.IUCN_Assessment,  
Stocks.IUCN_Group,  
stocks.IUCN_DateAssessed,  
  
STOCKS.pHMin,  
STOCKS.pHMax,  
STOCKS.dHMin,  
STOCKS.dHMax,  
  
stocks.genbankid,  
stocks.rfeid,  
stocks.figis_id,  
Species.Famcode as DetailField22,  
Species.SpeciesRefNo as DetailField23,  
Families.FamilyExpert as DF24,  
FAMILIES.CommonName,  
families.CommonName_Dutch,  
families.CommonName_French,  
families.CommonName_Spanish,  
families.CommonName_German,  
families.CommonName_Russian,  
families.CommonName_Italian,  
families.CommonName_Portuguese,  
families.CommonName_Swedish,  
families.CommonName_Greek_u,  
  
STOCKS.Envtemp,  
  
STOCKS.NorthernMost,  
STOCKS.NorthSouthS,  
  
STOCKS.SoutherMost,  
STOCKS.NorthSouthN,  
  
STOCKS.WesternMost,  
STOCKS.WestEastW,  
  
STOCKS.EasternMost,  
STOCKS.WestEastE,  
  
STOCKS.TempMin,  
STOCKS.TempMax,  
  
species.coordinator,  
Species.subfamily,  
Species.length,  
Species.lengthfemale,  
Species.LTypeMaxM,  
Species.MaxlengthRef,  
Species.weight,  
Species.weightfemale,  
Species.MaxweightRef,  
Species.longevitywild,  
species.anacat,  
species.migratref,  
species.DepthRangeShallow,  
species.DepthRangeDeep,  
species.UsedasBait ,  
species.LTypeMaxF,  
species.DangerousRef,  
species.Dangerous,
```

```

species.profile

FROM (Species LEFT JOIN FAMILIES ON Species.FamCode = FAMILIES.FamCode)
LEFT JOIN STOCKS ON Species.SpecCode = STOCKS.SpecCode
WHERE (((Species.SpecCode)=" . $SpecID . ") AND
((STOCKS.Level)='Species in general' Or (STOCKS.Level)='subSpecies in
general'))";

$Detail = $conn->query($qry);
debug_out("Detail->num_rows : ".$Detail->num_rows);

while( $row = $Detail->fetch_assoc() )
{

$detail['rfeid'] = $row['rfeid'];
$detail['genbankid'] = $row['genbankid'];
$detail['figis_id'] = $row['figis_id'];
$detail['profile'] = $row['profile'];
$detail['DangerousRef'] = $row['DangerousRef'];
$detail['Dangerous'] = $row['Dangerous'];

$detail['IUCN_Code'] = $row['IUCN_Code'];
$detail['IUCN_id'] = $row['IUCN_id'];
$detail['IUCN_Assessment'] = $row['IUCN_Assessment'];
$detail['IUCN_DateAssessed'] = $row['IUCN_DateAssessed'];
$detail['IUCN_Group'] = $row['IUCN_Group'];

$detail['LTypeMaxF'] = $row['LTypeMaxF'];
$detail['comments'] = $row['comments'];
$detail['StockDefs'] = $row['StockDefs'];

$detail['NorthernMost'] = $row['NorthernMost'];
$detail['NorthSouthS'] = $row['NorthSouthS'];
$detail['SoutherMost'] = $row['SoutherMost'];
$detail['NorthSouthN'] = $row['NorthSouthN'];
$detail['WesternMost'] = $row['WesternMost'];
$detail['WestEastW'] = $row['WestEastW'];
$detail['EasternMost'] = $row['EasternMost'];
$detail['WestEastE'] = $row['WestEastE'];

$detail['Envtemp'] = $row['Envtemp'];
$detail['TempMin'] = $row['TempMin'];
$detail['TempMax'] = $row['TempMax'];

$detail['DetailField22'] = $row['DetailField22'];
$detail['DetailField16'] = $row['DetailField16'];
$detail['DetailField17'] = $row['DetailField17'];
$detail['DetailField18'] = $row['DetailField18'];
$detail['DetailField4'] = $row['DetailField4'];

$detail['coordinator'] = $row['coordinator'];
$detail['DF24'] = $row['DF24'];

$detail['Aquarium'] = $row['Aquarium'];
$detail['DemersPelag'] = $row['DemersPelag'];
$detail['Importance'] = $row['Importance'];
$detail['GameFish'] = $row['GameFish'];
$detail['UsedforAquaculture'] = $row['UsedforAquaculture'];
$detail['family'] = $row['Family'];
$detail['DetailField3'] = $row['DetailField3'];

$detail['DetailField5'] = $row['DetailField5'];
$detail['DetailField6'] = $row['DetailField6'];
$detail['DetailField7'] = $row['DetailField7'];

```

```

$detail['CommonName'] = array();

$detail['CommonName']['English'] = $row['CommonName'];
$detail['CommonName']['Dutch'] = $row['CommonName_Dutch'];
$detail['CommonName']['French'] = $row['CommonName_French'];
$detail['CommonName']['Spanish'] = $row['CommonName_Spanish'];
$detail['CommonName']['German'] = $row['CommonName_German'];
$detail['CommonName']['Russian'] = $row['CommonName_Russian'];
$detail['CommonName']['Italian'] = $row['CommonName_Italian'];
$detail['CommonName']['Portuguese'] = $row['CommonName_Portuguese'];
$detail['CommonName']['Swedish'] = $row['CommonName_Swedish'];
$detail['CommonName']['Greek'] = $row['CommonName_Greek_u'];
$detail['subfamily'] = $row['subfamily'];
$detail['DetailField8'] = $row['DetailField8'];

$detail['length'] = $row['length'];
$detail['lengthfemale'] = $row['lengthfemale'];
$detail['LTypeMaxM'] = $row['LTypeMaxM'];
$detail['MaxlengthRef'] = $row['MaxlengthRef'];
$detail['weight'] = $row['weight'];
$detail['weightfemale'] = $row['weightfemale'];
$detail['MaxweightRef'] = $row['MaxweightRef'];
$detail['longevitywild'] = $row['longevitywild'];
$detail['anacat'] = $row['anacat'];
$detail['migrateref'] = $row['migrateref'];

$detail['pHMin'] = $row['pHMin'];
$detail['pHMax'] = $row['pHMax'];
$detail['dHMin'] = $row['dHMin'];
$detail['dHMax'] = $row['dHMax'];

$detail['DetailField12'] = $row['DetailField12'];
$detail['DetailField13'] = $row['DetailField13'];
$detail['DetailField14'] = $row['DetailField14'];
$detail['DepthRangeShallow'] = $row['DepthRangeShallow'];
$detail['DepthRangeDeep'] = $row['DepthRangeDeep'];

$detail['UsedasBait'] = $row['UsedasBait'];
$detail['DetailField23'] = $row['DetailField23'];

//start trasferred...
$detail['Entered'] = $row['DetailField16'];
$detail['Modified'] = $row['DetailField17'];
$detail['Expert'] = $row['DetailField18'];
$refAuthor = "";
If ($row['DetailField23'] != "")
{
$GetRefDetails = $conn->query("select Author, Year from Refrens where refno
= " . $row['DetailField23']);
while( $row = $GetRefDetails->fetch_assoc() )
{
$refAuthor = $row['Author'];
$refYear = $row['Year'];
}
$GetRefDetails->close();
}
//end transferred...

$detail['refAuthor'] = $refAuthor;
$detail['refYear'] = $refYear;
}

$Detail->close();
return $detail;
}

```



```

}
Else
{
if ($Language == "English")
{
$concatinfo = "
<a href='../Glossary/Glossary.cfm?termEnglish=dorsal fin' style='text-
decoration : underline;color : Black;'>Dorsal</a>
<a href='../Glossary/Glossary.cfm?termEnglish=spine' style='text-decoration
: underline;color : Black;'>spines</a>
(total): ";
}
else
{
$concatinfo = $dorsal_spines_u . " (" . $total_u . "): ";
}
$concatinfo_eng = "
<a href='../Glossary/Glossary.cfm?termEnglish=dorsal fin' style='text-
decoration : underline;color : Black;'>Dorsal</a>
<a href='../Glossary/Glossary.cfm?termEnglish=spine' style='text-decoration
: underline;color : Black;'>spines</a>
(total): ";
}

// $row['']
if ($row['dorsalspinesmin'] != "")
{
$concatinfo = $concatinfo . $row['dorsalspinesmin'];
$concatinfo_eng = $concatinfo_eng . $row['dorsalspinesmin'];
}
if ($row['dorsalspinesmax'] != "")
{
//if (#Right(concatinfo, 1)# is " " >
If (substr($concatinfo, strlen($concatinfo)-1) == " " )
{
$concatinfo = $concatinfo . $row['dorsalspinesmax'];
$concatinfo_eng = $concatinfo_eng . $row['dorsalspinesmax'];
}
else
{
$concatinfo = $concatinfo . "-" . $row['dorsalspinesmax'];
$concatinfo_eng = $concatinfo_eng . "-" . $row['dorsalspinesmax'];
}
}
//<!-- start new l-->
if ($row['dorsalspinesmin'] == $row['dorsalspinesmax'])
{
/* to php
$concatinfo = #ReplaceNoCase(concatinfo, dorsalspinesmin&"-
"&dorsalspinesmin, dorsalspinesmin , "one" )#>
$concatinfo_eng = #ReplaceNoCase(concatinfo, dorsalspinesmin&"-
"&dorsalspinesmin, dorsalspinesmin , "one" )#>
*/
$concatinfo = str_ireplace($row['dorsalspinesmin'] . "-" .
$row['dorsalspinesmin'], $row['dorsalspinesmin'], $concatinfo);
$concatinfo_eng = str_ireplace($row['dorsalspinesmin'] . "-" .
$row['dorsalspinesmin'], $row['dorsalspinesmin'], $concatinfo);
}
//<!-- end new l-->
}

if ($detail_DetailField4 != "Elasmobranchii (sharks and rays)")
{
if ($row['dorsalsoftraysmin'] != "" || $row['dorsalsoftraysmax'] != "")
{
If ($concatinfo != " " )
{
if ($Language == "English")
{
$concatinfo = $concatinfo . ";

```

```

<a href='../Glossary/Glossary.cfm?termEnglish=dorsal fin' style='text-decoration : underline;color : Black;'>Dorsal</a>
<a href='../Glossary/Glossary.cfm?termEnglish=soft ray' style='text-decoration : underline;color : Black;'>soft rays</a>
(total): ";
}
else
{
$concatinfo = $concatinfo . " ; " . $dorsal_soft_rays_u . " (" . $total_u .
"): ";
}
$concatinfo_eng = $concatinfo_eng . " ;
<a href='../Glossary/Glossary.cfm?termEnglish=dorsal fin' style='text-decoration : underline;color : Black;'>Dorsal</a>
<a href='../Glossary/Glossary.cfm?termEnglish=soft ray' style='text-decoration : underline;color : Black;'>soft rays</a>
(total): ";
}
else
{
if ($Language == "English")
{
$concatinfo = "
<a href='../Glossary/Glossary.cfm?termEnglish=dorsal fin' style='text-decoration : underline;color : Black;'>Dorsal</a>
<a href='../Glossary/Glossary.cfm?termEnglish=soft ray' style='text-decoration : underline;color : Black;'>soft rays</a>
(total): ";
}
else
{
$concatinfo = $dorsal_soft_rays_u . " (" . $total_u . "): ";
}
$concatinfo_eng = "
<a href='../Glossary/Glossary.cfm?termEnglish=dorsal fin' style='text-decoration : underline;color : Black;'>Dorsal</a>
<a href='../Glossary/Glossary.cfm?termEnglish=soft ray' style='text-decoration : underline;color : Black;'>soft rays</a>
(total): ";
}
If ($row['dorsalsoftraysmin'] != "")
{
$concatinfo = $concatinfo . $row['dorsalsoftraysmin'];
$concatinfo_eng = $concatinfo_eng . $row['dorsalsoftraysmin'];
}
If ($row['dorsalsoftraysmax'] != "")
{
/*
cf = php
#Right(concatinfo, 1)# ===== substr($concatinfo, strlen($concatinfo)-1)
*/
If (substr($concatinfo, strlen($concatinfo)-1) == " " )
{
$concatinfo = $concatinfo . $row['dorsalsoftraysmax'];
$concatinfo_eng = $concatinfo_eng . $row['dorsalsoftraysmax'];
}
}
else
{
$concatinfo = $concatinfo . "-" . $row['dorsalsoftraysmax'];
$concatinfo_eng = $concatinfo_eng . "-" . $row['dorsalsoftraysmax'];
}
}
}

//<!-- start new 2-->
If ($row['dorsalsoftraysmin'] == $row['dorsalsoftraysmax'])
{
$concatinfo = str_ireplace($row['dorsalsoftraysmin'] . "-" . $row['dorsalsoftraysmin'], $row['dorsalsoftraysmin'], $concatinfo);
}

```

```

$concatinfo_eng = str_ireplace($row['dorsalsoftraysmin'] . "-" .
$row['dorsalsoftraysmin'], $row['dorsalsoftraysmin'], $concatinfo);
}
//<!-- end new 2-->
}

If ($row['analfinsspinesmin'] != "" or $row['analfinsspinesmax'] != "")
{
If ($concatinfo != "" )
{
if ($Language == "English")
{
$concatinfo = $concatinfo . ";
<a href='../Glossary/Glossary.cfm?termEnglish=anal fin' style='text-
decoration : underline;color : Black;'>Anal</a>
<a href='../Glossary/Glossary.cfm?termEnglish=spine' style='text-decoration
: underline;color : Black;'>spines</a>: ";
}
else
{
$concatinfo = $concatinfo . "; " . $anal_spines_u . " ";
}
$concatinfo_eng = $concatinfo_eng . "; <a
href='../Glossary/Glossary.cfm?termEnglish=anal fin' style='text-decoration
: underline;color : Black;'>Anal</a> <a
href='../Glossary/Glossary.cfm?termEnglish=spine' style='text-decoration :
underline;color : Black;'>spines</a>: ";
}
else
{
if ($Language == "English")
{
$concatinfo = "
<a href='../Glossary/Glossary.cfm?termEnglish=anal fin' style='text-
decoration : underline;color : Black;'>Anal</a>
<a href='../Glossary/Glossary.cfm?termEnglish=spine' style='text-decoration
: underline;color : Black;'>spines</a>: ";
}
else
{
$concatinfo = $anal_spines_u . " ";
}
$concatinfo_eng = "
<a href='../Glossary/Glossary.cfm?termEnglish=anal fin' style='text-
decoration : underline;color : Black;'>Anal</a>
<a href='../Glossary/Glossary.cfm?termEnglish=spine' style='text-decoration
: underline;color : Black;'>spines</a>: ";
}
}
If ($row['analfinsspinesmin'] != "")
{
$concatinfo = $concatinfo . $row['analfinsspinesmin'];
$concatinfo_eng = $concatinfo_eng . $row['analfinsspinesmin'];
}
If ($row['analfinsspinesmax'] != "")
{
//If (#Right(concatinfo, 1)# is " " >
If (substr($concatinfo, strlen($concatinfo)-1) == " " )
{
$concatinfo = $concatinfo . $row['analfinsspinesmax'];
$concatinfo_eng = $concatinfo_eng . $row['analfinsspinesmax'];
}
else
{
$concatinfo = $concatinfo . "-" . $row['analfinsspinesmax'];
$concatinfo_eng = $concatinfo_eng . "-" . $row['analfinsspinesmax'];
}
}
}
//<!-- start new 3-->

```

```

If ($row['analfinsspinesmin'] == $row['analfinsspinesmax'])
{
$concatinfo = str_ireplace($row['analfinsspinesmin'] . "-" .
$row['analfinsspinesmin'], $row['analfinsspinesmin'] , $concatinfo );
$concatinfo_eng = str_ireplace($row['analfinsspinesmin'] . "-" .
$row['analfinsspinesmin'], $row['analfinsspinesmin'] , $concatinfo );
}
//<!-- end new 3-->
}

If ($detail_DetailField4 != "Elasmobranchii (sharks and rays)")
{
If ($row['araymin'] != "" or $row['araymax'] != "")
{
If ($concatinfo != " " )
{
if ($Language == "English")
{
$concatinfo = $concatinfo . ";
<a href='../Glossary/Glossary.cfm?termEnglish=anal fin' style='text-decoration : underline;color : Black;'>Anal</a>
<a href='../Glossary/Glossary.cfm?termEnglish=soft ray' style='text-decoration : underline;color : Black;'>soft rays</a>: ";
}
else
{
$concatinfo = $concatinfo . "; " . $anal_soft_rays_u . ": ";
}
$concatinfo_eng = $concatinfo_eng . ";
<a href='../Glossary/Glossary.cfm?termEnglish=anal fin' style='text-decoration : underline;color : Black;'>Anal</a>
<a href='../Glossary/Glossary.cfm?termEnglish=soft ray' style='text-decoration : underline;color : Black;'>soft rays</a>: ";
}
else
{
if ($Language == "English")
{
$concatinfo = "
<a href='../Glossary/Glossary.cfm?termEnglish=anal fin' style='text-decoration : underline;color : Black;'>Anal</a>
<a href='../Glossary/Glossary.cfm?termEnglish=soft ray' style='text-decoration : underline;color : Black;'>soft rays</a>: ";
}
else
{
$concatinfo = $anal_soft_rays_u . ": ";
}
$concatinfo_eng = "
<a href='../Glossary/Glossary.cfm?termEnglish=anal fin' style='text-decoration : underline;color : Black;'>Anal</a>
<a href='../Glossary/Glossary.cfm?termEnglish=soft ray' style='text-decoration : underline;color : Black;'>soft rays</a>: ";
}
}

If ($row['araymin'] != "")
{
$concatinfo = $concatinfo . $row['araymin'];
$concatinfo_eng = $concatinfo_eng . $row['araymin'];
}
If ($row['araymax'] != "")
{
//If (#Right(concatinfo, 1)# is " " >
If (substr($concatinfo, strlen($concatinfo)-1) == " " )
{
$concatinfo = $concatinfo . $row['araymax'];
$concatinfo_eng = $concatinfo_eng . $row['araymax'];
}
else
{

```

```

$concatinfo = $concatinfo . "-" . $row['araymax'];
$concatinfo_eng = $concatinfo_eng . "-" . $row['araymax'];
}
}
}
//<!-- start new 4-->
If ($row['araymin'] == $row['araymax'])
{
$concatinfo = str_ireplace($row['araymin'] . "-" . $row['araymin'],
$row['araymin'] , $concatinfo );
$concatinfo_eng = str_ireplace($row['araymin'] . "-" . $row['araymin'],
$row['araymin'] , $concatinfo );
}
//<!-- end new 4-->
}

If ($row['Vertebraetotalmin'] != "" || $row['Vertebraetotalmax'] != "")
{
{
If ($concatinfo != "" )
{
if ($Language == "English")
{
$concatinfo = $concatinfo . ";
<a href='../Glossary/Glossary.cfm?termEnglish=vertebrae' style='text-
decoration : underline;color : Black;'>Vertebrae</a>: ";
}
}
else
{
$concatinfo = $concatinfo . "; " . $vertebrae_u . ": ";
}
$concatinfo_eng = $concatinfo_eng . ";
<a href='../Glossary/Glossary.cfm?termEnglish=vertebrae' style='text-
decoration : underline;color : Black;'>Vertebrae</a>: ";
}
}
else
{
if ($Language == "English")
{
$concatinfo = "<a href='../Glossary/Glossary.cfm?termEnglish=vertebrae'
style='text-decoration : underline;color : Black;'>Vertebrae</a>: ";
}
}
else
{
$concatinfo = $vertebrae_u . ": ";
}
$concatinfo_eng = "<a href='../Glossary/Glossary.cfm?termEnglish=vertebrae'
style='text-decoration : underline;color : Black;'>Vertebrae</a>: ";
}
}
If ($row['Vertebraetotalmin'] != "")
{
{
$concatinfo = $concatinfo . $row['Vertebraetotalmin'];
$concatinfo_eng = $concatinfo_eng . $row['Vertebraetotalmin'];
}
}
If ($row['Vertebraetotalmax'] != "")
{
{
//If (#Right(concatinfo, 1)# is " " >
If (substr($concatinfo, strlen($concatinfo)-1) == " " )
{
$concatinfo = $concatinfo . $row['Vertebraetotalmax'];
$concatinfo_eng = $concatinfo_eng . $row['Vertebraetotalmax'];
}
}
else
{
$concatinfo = $concatinfo . "-" . $row['Vertebraetotalmax'];
$concatinfo_eng = $concatinfo_eng . "-" . $row['Vertebraetotalmax'];
}
}
}
}
}

```

```

//<!-- start new 5-->
If ($row['Vertebraetotalmin'] == $row['Vertebraetotalmax'])
{
$concatinfo = str_ireplace($row['Vertebraetotalmin'] . "-" .
$row['Vertebraetotalmin'], $row['Vertebraetotalmin'] , $concatinfo );
$concatinfo_eng = str_ireplace($row['Vertebraetotalmin'] . "-" .
$row['Vertebraetotalmin'], $row['Vertebraetotalmin'] , $concatinfo );
}
//<!-- end new 5-->
}

If ($row['addchars'] != "" )
{
If ($concatinfo != "" )
{
$concatinfo = $concatinfo . ". " . $row['addchars'];
$concatinfo_eng = $concatinfo_eng . ". " . $row['addchars'];
}
else
{
$concatinfo = $row['addchars'];
$concatinfo_eng = $row['addchars'];
}

If ($with_ch == 1)
{
$qry = "SELECT morphdat.* FROM morphdat
WHERE (
((morphdat.StockCode)=" . $getMorph_StockCode . ") and
((morphdat.females)=" . $getMorph_females . ") and
((morphdat.males)=" . $getMorph_males . "))";
$$s4 = $conn_chi->query($qry);
while( $row_2 = $$s4->fetch_assoc() )
{
$concatinfo = $concatinfo . "<br><font size='2' style='background-color : "
. $color_ch . "; '>" . $row_2['addchars_u'] . "</font>";
}
$$s4->close();
}
}

} //if ($ctr == 1)
$debug_lcc++;
debug_out("Loop counter: ".$debug_lcc);

} //while

$getMorph->close();

return $concatinfo;

}

function lang_abbrev($longlang)
{
switch ($longlang) {
case "Spanish" : $s='es'; break;
case "English" : $s='en'; break;
case "German" : $s='de'; break;
case "Greek" : $s='el'; break;
case "French" : $s='fr'; break;
case "Italian" : $s='it'; break;
case "Dutch" : $s='nl'; break;
case "Portuguese" : $s='pt'; break;
default: $s=NULL;
}
}

return($s);

```

```

}

function translate( $dst,$txt)
{
$c = new cDocumentInformation();

$c->sourceLanguage="en";
$c->targetLanguages=array($dst);
$c->terminologyDomains=array("AG", "BZ", "AGB");
$c->userDictionary=NULL;
$c->isConfidential=false;
$c->handleCelexEncoding=false;
$c->allCoding=false;
$c->textType=NULL;
$c->refid="inBody_0";

$wsdl='http://filaman.uni-kiel.de/trans.wsdl';
$client = new SoapClient($wsdl, array (
"trace" => 1,
"exceptions" => 0));

$text=$txt;
$text="(9331) ".$text." (EOT)";
$text=utf8_encode($text);
$client->translateArrayOfDocuments(
array(
userName => "xxxxxxx",
userPassword => "xxxxxxx",
authenticationDomain => "xxxxxxx"
),
array($c)
,array($text));

$data=$client->__getLastResponse();

$outtext= preg_replace('/.+\\(9331\\) (.+) \\(EOT\\).+/s','$1',$data);

return $outtext;
}

```

Table 5: PHP-Functions for FishBase and ECMT access.

```

<?php
/**
 *
 * @author Sven Mohr
 * @version 0.1
 * @module fb.translation
 * @lastchange 23.11.2005
 */
session_start();

//include ("application.php");
include ("db_connect.php");
include ("fb_functions.php");

// Set Language
if ( !isset($_SESSION['Language'] ) ) {
    session_register('Language');
    $Language = $_SESSION['Language']='English'; // Set default language
} else {
    $Language = $_SESSION['Language'];
}

if ( isset($_GET['lang'] ) ) {
    $Language = ucfirst(strtolower($_GET['lang']));
}

```

```

if ( !(strcmp($_SESSION['Language'], $Language)==0) ) {
    $_SESSION['Language'] = $Language;
}

/* =MAIN= */

if (!isset($_GET["id"]) )
    die(" No SpecId specified ");
else
    $SpecId=$_GET["id"];

$detail=get_detail_info($SpecId);

output_header("Translation (ECMT) Species Summary for
".$detail['CommonName']['English'], "utf-8");
start_body();
debug_out("Language : ".$Language);
debug_out("Requested translation for Species (id): ".$SpecId);
debug_out("Calling get_detail_info");

debug_out_var($detail);

$morphology=get_morphology_data($SpecId);

if ($dst=lang_abbr($Language)) {
    debug_out("Translating to ".$dst);
    if ($detail["StockDefs"]!='') {
        $distribution_l=translate($dst, "Distribution");
        $distribution_t=translate($dst, remove_html_tags($detail["StockDefs"].
""));
    }
    if ($detail["comments"]!='') {
        $biology_l=translate($dst, "Biology");
        $biology_t=translate($dst, remove_html_tags($detail["comments"].
" "));
    }
    if ($morphology!='' && strlen($morphology)>2 ) {
        $morphology_l=translate($dst, "Morphology");
        $morphology_t=translate($dst, remove_html_tags($morphology."
"));
    }
} else {
    info_out($Language." not supported by ECMT service");
}
debug_out("Distribution: ".$detail["StockDefs"]);

debug_out("Distribution(t): ".$distribution_t);
debug_out("Biology: ".$detail["comments"]);

debug_out("Biology(t): ".$biology_t);
info_out("Translation provided by ECMT");
echo '<table>';
if (isset($distribution_l)) {
    echo '<table>';
    echo '<tr><td style="background:#C6D9D7;font-weight:
bold">Distribution</td><td
style="background:#fff">'.$detail["StockDefs"].'</td></tr>';
    echo '<tr><td style="background:#C6D9D7;font-weight:
bold">'.$distribution_l.'</td><td style="background:
#FFF">'.$distribution_t.'</td></tr>';
}
if (isset($morphology_l)) {
    echo '<tr><td style="background:#C6D9D7;font-weight:
bold">Morphology</td><td
style="background:#fff">'.$morphology.'</td></tr>';
    echo '<tr><td style="background:#C6D9D7;font-weight:
bold">'.$morphology_l.'</td><td style="background:
#FFF">'.$morphology_t.'</td></tr>';
}
if (isset($biology_l)) {

```

```

echo '<tr><td style="background:#C6D9D7;font-weight: bold">Biology</td><td
style="background:#fff">'. $detail["comments"]. '</td></tr>';
echo '<tr><td style="background:#C6D9D7;font-weight:
bold">'. $biology_l. ':</td><td style="background:
#FFF">'. $biology_t. '</td></tr>';
}
echo '</table>';

info_out("Page created by Sven Mohr");
end_body();
?>

```

Table 6: Translation of Biology, Morphology and Distribution fields via EC-MT Service

7) Contact information and version history

For further information, advise and in case of any query on multilingual services for data resources in the Internet please contact the manager of WP-11 in ENBI, Bernd Ueberschär (bueberschaer@ifm-geomar.de) or Rainer Froese (rfroese@ifm-geomar.de).

Version	Date	Author	Purpose
0.1	10/10/05	B. Ueberschär	Created
0.2	01/12/05	B. Ueberschär, Rainer Froese Sven Mohr, EU- MT-Department	Extension Combination of all deliverables
0.9	15/12/05	B. Ueberschär	Improvements
1.0	06/01/06	B. Ueberschär	Final version for distribution

ANNEX A: Rules for preparing English source-text for machine translation

The below rules are provided for preparing English source text for machine translation. Some rules are specific to the SYSTRAN software, but most rules will apply in general.

- **Rules for the Use of Articles**
- **Rules for Lists**
- **Rules for Phrase Structure**
- **Rules for Punctuation**
- **Rules for Formatting**
- **Other Rules**

Many of these rules address the use of correct grammar and punctuation, as English presents many opportunities for ambiguity and the degree of precision necessary for machine translation is much greater than that needed for human understanding.

Rules for the Use of Articles

The large challenge in machine translation is the resolution of homographs, words that are written the same way but that have different meanings and often are different parts of speech (for example, noun and verb or adjective and verb). In English, the placement of articles often helps to clarify how a word is being used. Articles are also useful for clarity in many other contexts.

The rules listed below apply mostly to English as source language.

A1. Use articles to reduce the ambiguity caused by homographs.

Instead of: empty file

Use: empty the file

Or: the empty file

A2. Use articles to clarify the function of the present and past participles.

Instead of: moving car

Use: the moving car

Or: moving the car

A3. Use articles and punctuation to clarify the part of speech of a word.

Instead of: Check the lighting, electrical, and navigation systems.

Use: Check the lighting, the electrical, and the navigation systems.

Or: Check the lighting, and the electrical and navigation systems.

Rules for Lists

The use of articles is especially useful in the identification of list items. Also, it aids greatly in translation to have each item of the list stated in the same manner. The rules listed below apply only to the English language.

L1. Use articles at the beginning of each item in a list.

Instead of: the brake and tail lights

Use: the brake light and the tail lights

L2. Use articles to show that only a certain item in a list is modified.

Instead of: the brake pedal and accelerator

Use: the brake pedal and the accelerator

L3. Write list items as clauses or full sentences whenever possible; don't mix sentences, words, and phrases.

Instead of: Rotate the wheels, lubricate, clean head.

Use: Rotate the wheels, lubricate the joints, and clean the head.

Or: Rotate the wheels. Lubricate the joints. Clean the head.

Rules for Phrase Structure

Correct and concise English grammar is essential to good machine translation, as is absolute clarity. The rules listed below apply only to the English language.

S1. Repeat nouns or noun phrases instead of using pronouns; avoid particularly the pronoun "it."

Instead of: Wash the car, clean the windshield, and then wax it.

Use: Wash the car, clean the windshield, and then wax the car.

S2. Put phrases as close as possible to the nouns that they modify.

Instead of: Engine cover for sale by elderly gentleman with a few bolts missing.

Use: Engine cover with a few bolts missing for sale by elderly gentleman.

S3. Use the word "to" or an auxiliary verb to indicate the infinitive form of a verb and to distinguish it from a finite form of the verb.

Instead of: Rotate the wheel to clean and then lubricate the head.

Use: Rotate the wheel to clean it and to lubricate the head.

Or: Rotate the wheel to clean it, then lubricate the head.

S4. Arrange sentences to minimize ambiguity.

Instead of: Remove the bolts holding the assembly with the left hand.

Use: Remove the bolts, which hold the assembly, with the left hand.

Or: To remove the bolts, hold the assembly with the left hand.

S5. Don't leave out subordinate clause markers (that, which, who, etc.).

Instead of: Make sure you select the proper tool.

Use: Make sure that you select the proper tool.

Rules for Punctuation

Punctuation is essential for dividing sentences into their logical components, allowing for their correct interpretation. The rules listed below apply only to the English language.

P1. Separate two main clauses with a comma followed by "and," or with a semicolon.

Instead of: Check the figures, verify the test results.

Use: Check the figures, and verify the test results.

Or: Check the figures; verify the test results.

P2. Set off subordinate phrases and subordinate clauses with commas.

Instead of: After you have checked the lights, the brakes, and the steering make a report.

Use: After you have checked the lights, the brakes, and the steering, make a report.

P3. Put commas after all prepositional phrases that begin sentences.

Instead of: During the landing personnel should remain seated.

Use: During the landing, personnel should remain seated.

P4. Use commas to set off embedded clauses.

Instead of: Wolf's analysis (which supports this conclusion) is scholarly and detailed.

Use: Wolf's analysis, which supports this conclusion, is scholarly and detailed.

P5. Use parentheses where the material enclosed in the parentheses does not have a close logical relationship to the sentence.

Instead of: Burke's discovery, see page 23, supports this conclusion.

Use: Burke's discovery (see page 23) supports this conclusion.

P6. Limit the use of the slash*.

Instead of: at the beginning/end

Use: at the beginning or at the end

*Unless they have been entered into your user dictionary (UD) prior to translation, two words with a slash between them will most likely be marked as a single Not-Found Word. The untranslated word pair will then be transferred as-is to the translated text.

P7. Hyphenate phrases that modify other words or phrases.

Instead of: man eating shark

Use: man-eating shark

P8. Do not insert hyphens to break words that fall at the end of a line; do not hyphenate, or use "soft" hyphens instead.

Instead of: If you hyphenate the words that fall at the end of a line to try to get an even right margin, then the program will look for each half of the word in the SYSTRAN dictionaries or UD's. Depending on the word fragments, it will either list both parts of the word as separate Not-Found Words or assign incorrect * or partial definitions to the word.

Use: If you hyphenate the words that fall at the end of a line to try to get an even right margin, then the program will look for each half of the word in the SYSTRAN dictionaries or UD's.

Depending on the word fragments, it will either list both parts of the word as separate Not-Found Words or assign incorrect or partial definitions to the word.

*In-correct will be translated as the word "in" and the word "correct," instead of as "not correct."

P9. Avoid the use of a dash as a punctuation mark, use other punctuation instead.

Instead of: Because the data were incorrectly analyzed – the reason for which will be discussed later – the wrong conclusions were drawn.

Use: Because the data were incorrectly analyzed (the reason for which will be discussed later), the wrong conclusions were drawn.

Or: Because the data were incorrectly analyzed, the reason for which will be discussed later, the wrong conclusions were drawn.

Rules for Formatting

SYSTRAN requires certain indicators in order to identify the ends of sentences and of paragraphs. If these indicators are not in place in the corpus to be translated, the program may not be able to determine this information and, as such, it will not provide accurate translation.

The rules listed below are useful for most language pairs.

F1. Use two spaces at the end of all sentences and after a colon.

Instead of: These are the critical areas: development, production and marketing.

Use: These are the critical areas: development, production and marketing.

F2. Use one space after abbreviations, commas, and semicolons.

Instead of: Mr. Smythe

Use: Mr. Smythe

F3. Use the word wrap or the soft return feature of your word processor for all sentences within a paragraph, instead of inserting hard returns.

There is no need to press ENTER at the end of each line. It may cause your text to be broken in strange places, and it will convince the translation program that each line is a paragraph.

Instead, just keep on typing, and your software will automatically wrap the text around to fit your page, whatever size it may be.

F4. Use a hard return at the end of each paragraph.

F5. Use two hard returns after all titles and headings, unless they end in a punctuation mark.

F6. Use indents, tables, and tabs instead of many spaces.

Other Rules

There are a number of rules that do not fit comfortably into any of the other five categories, and these rules are offered here. Most of these rules are useful for many language pairs.

O1. Use abbreviations consistently.

Instead of: a 3 min. min.

Use: a 3 min. minimum.

Or: a 3 minute min.

O2. The use of different fonts helps to draw the eye, but the program has no way of identifying fonts. Use other methods to set off text for machine translation. Set off names, such as key names, icon names, and functions; by punctuation, usage, or case rather than by font alone.

Instead of: Press enter.

Use: Press "ENTER."

Or: Press the ENTER key.

O3. Single words or acronyms that are not to be translated should be preceded by a period.

Instead of: I work for the CORE.

Use: I work for the .CORE.

Imperative Translations

These are sample sentences and their translations for each Imperative Translation Mode available for English into French, German, Italian, Portuguese and Spanish. There are no imperative options available for English Target translations.

English Into French

Sample English sentence using the imperative:

Push the button.

Default Translation:

Poussez le bouton. Formal / Plural Informal

Option:

Pousser le bouton. Infinitive

English into German:

Sample English sentence using the imperative:

Turn the knob.

Default Translation:

Drehen Sie den Hebel. Formal

Options:

Den Hebel drehen.

Drehe den Hebel.

Dreht den Hebel.

Infinitive Singular Informal Plural Informal

English into Italian:

Sample English sentence using the imperative:
Shut the door.

Default Translation:
Chiuda il portello. Subjunctive

Option:
Chiudere il portello. Infinitive

English into Portuguese:

Sample English sentence using the imperative:
Open the door.

Default Translation:
Abra a porta. Subjunctive

Option:
Abrir a porta. Infinitive

English into Spanish:

Sample English sentence using the imperative:
Verify the position.

Default Translation:
Verifica la posición. Imperative

Option:
Verificar la posición. Infinitive

ANNEX B: Text-Editing Examples from FishBase

The examples below apply specifically to English source text translated to German and with one example for French, but may also be relevant for other language pairs. Below are some examples of original text and edited text, with the different translation results. Subject nouns and verbs have been inserted. Even when there is no ambiguity as described above, inserting words can still improve the arrangement of a translation, making it clearer. Please note: the text in bold italics represents the original text in FishBase species summaries.

(1)

Diagnosis: Head large without deep occipital groove.

→

Diagnose: Ohne tiefe occipital Rinne großer Kopf.

Diagnosis: head **is** large and without deep occipital groove.

→

Diagnose: der Kopf ist groß und ohne tiefe occipital Rinne.

(Although not in the English text, the article “der” was inserted by the translation engine)

(2)

Gill membranes fused to the body and isthmus. Superior trunk and tail ridges continuous, inferior trunk and tail ridges discontinuous, lateral trunk ridge confluent with inferior tail ridge. Brood area of male located under trunk.

→

Kiememembranen, die zum Körper und Isthmus durchgebrannt werden. Überlegenes kontinuierliches, untergeordnetes Kabel der Kabel- und Endstückkanten und unterbrochene, seitliche Kabelkante der Endstückkanten, die mit untergeordneter Endstückkante zusammenfließend sind. Brutgebiet von Mann, das sich unter Kabel befindet.

Gill membranes **are** fused to body and isthmus. Superior trunk and tail ridges **are** continuous, inferior trunk and tail ridges **are** discontinuous, and lateral trunk ridge **is** confluent with inferior tail ridge. Brood area of male **is** located under trunk.

→

Kiememembranen werden zu Körper und Isthmus durchgebrannt. Überlegene Kabel- und Endstückkanten sind kontinuierliches, untergeordnetes Kabel, und Endstückkanten sind unterbrochen, und seitliche Kabelkante ist zusammenfließend mit untergeordneter Endstückkante. Das Brutgebiet des Mannes befindet sich unter Kabel.

Note on definite/indefinite articles. In most of the examples, omission of articles does not really pose a problem for understanding the text, so these could be ignored to save time. However, they can also have an impact. In the following example, insertion of articles improves understanding:

The gill membranes **are** fused to the body and isthmus. **The** superior trunk and tail ridges **are** continuous, **the** inferior trunk **and** tail ridges are

discontinuous, **and the** lateral trunk ridge **is** confluent with **the** inferior tail ridge. **The** brood area of **the** male **is** located under **the** trunk.

→

Die Kiememembranen werden zum Körper und Isthmus durchgebrannt. Die überlegenen Kabel- und Endstückkanten sind kontinuierlich, die untergeordneten Kabel- und Endstückkanten sind unterbrochen, und die seitliche Kabelkante ist zusammenfließend mit der untergeordneten Endstückkante. Das Brutgebiet des Mannes befindet sich unter dem Kabel.

(The translation remains unsatisfactory because the terms are not translated in a 'Morphology' context)

(3)

Distribution: Eastern Atlantic: Norway and Greenland south to Morocco, and from Mauritania to Guinea (Mauritanian Upwelling Region). Seasonally present from Morocco to Mauritania along the edge of the continental shelf.

→

Verteilung: Ost-Atlantik: Norwegen- und Grönland-Süden nach Marokko und aus Mauretanien nach Guinea (mauritanische Region Upwelling). Saisonale Gegenwart aus Marokko nach Mauretanien den Rand des Kontinentalsockels entlang.

Distribution: Eastern Atlantic: **species ranges** from Norway and Greenland **in the North** to Morocco **in the South**, and from Mauritania to Guinea (Mauritanian Upwelling Region). **It is** seasonally present from Morocco to Mauritania along the edge of the continental shelf.

→

Verteilung: Ost-Atlantik: die Art reicht von Norwegen und Grönland im Norden bis zu Marokko den Süden und von Mauretanien bis zu Guinea (mauritanische Region Upwelling). Sie liegt saisonal von Marokko nach Mauretanien den Rand des Kontinentalsockels entlang.

(4)

Feeds on crustaceans, mostly shrimps and shore crabs; fishes, mostly gobies, flatfish, young herring and sand eels.

→

Alimentations sur les crustacés, principalement crevettes et crabes côtiers; poissons, principalement gobies, poissons plats, jeunes harengs et équilles.

It feeds on crustaceans, mostly shrimps and shore crabs; fishes, mostly gobies, flatfish, young herring and sand eels.

→

Il nourrit sur les crustacés, principalement crevettes et crabes côtiers; poissons, principalement gobies, poissons plats, jeunes harengs et équilles.

(Verb correctly used instead of noun, but still not quite right – should be se nourrit de. For German, a verb is in fact used, but not for French or Spanish, hence a French example!)

(5)

Leaps out of the water when hooked. Utilized fresh and frozen; can be fried, broiled and baked.

→

Sprünge aus dem Wasser wenn eingehackt. Benutzt frisch und eingefroren; können Sie gebraten, gebraten und gebacken werden.

It leaps out of the water when it is hooked. **The fish** is utilized fresh and frozen; **it** can be fried, broiled and baked.

→

Es springt aus dem Wasser, wenn es eingehackt wird. Der Fisch ist benutzt frisch und eingefroren; er kann gebraten, gebraten und gebacken werden.

(6)

Biology: Gregarious. Forms schools.

→

Biologie: Gesellig. Die Formschulen.

Biology: gregarious. **The species** forms schools.

→

Biologie: gesellig. Die Art bildet die Schulen.

Punctuation

Resilience: Medium, minimum population doubling time 1.4 - 4.4 years (K=0.16; tm=3-4; tmax=16; Fec=200,000)

→

Beweglichkeit: Mittleres, Bevölkerungsminimum, das Zeit 1,4 - 4,4 Jahre verdoppelt, (K=0.16; tm=3-4; Tmax=16; Fec=200,000)

The program thinks that *medium* applies to *population*; a colon after *time* would also make it clearer that *minimum population doubling time* is one unit.

Resilience: medium. **Minimum population doubling time:** 1.4 - 4.4 years (K=0.16; tm=3-4; tmax=16; Fec=200,000).

→

Beweglichkeit: Medium. Bevölkerungsminimum, das Zeit verdoppelt,: 1.4 - 4,4 Jahre (K=0.16; tm=3-4; Tmax=16; Fec=200,000).

The result is not brilliant, now there is a noun instead of an adjective after *Beweglichkeit*, but at least the different “packets” of information are kept together.

For information, the 3 punctuation marks **;**, **:** and **.** all mean “end of translation segment, start analyzing a new segment”. This means that for phrases like *Climate: temperate*, the adjective *temperate* is not associated with the noun *Climate*.

Consequently, in the translation, the adjective ending may not agree with the noun gender (the default is masculine ending). This is only a minor nuisance, since the text will still be understandable, and it may not worth trying to edit the text in order to change it.

Summary: Punctuation can help, but writing full sentences is the most useful. Other tips are: use the active rather than passive voice, avoid long sentences and lots of sub-clauses.

ANNEX C: Some additional considerations: Translator's Notes (workshop outcome)

The following paragraphs are compiled from some informal notes taken by the translation team as an outcome of the ENBI-workshops on "Machine Translation of Biodiversity Information Systems".

Machine translation is still a challenge and results are sometimes not really predictable and often cannot be standardized (because of the many differences between natural target languages). Thus the following notes from brainstorming on translation issues from ENBI-translation workshops reflect individual practical experience on machine translation and may help to understand some of the difficulties (and possible improvements and solutions) which are connected with machine translation. Potential user of MT-technique may find useful hints for their specific language.

- Telegraphic style ruins the machine translation. A major finding of the exercises in the workshop was, the way FishBase texts are written, a subject or verb is often missing, causing the program to misinterpret the context and/or the information. It is suggested to revise the free text in FishBase for more complete sentences.
- Context sensitivity is another major problem for translation. Terms have different meaning when appearing in another context. That is why domains or categories are an important and powerful feature in machine translation technology. For the information systems that are treated as trials in ENBI-WP-11, special categories will host the translated list of terms (e.g. Fisheries, Environment, Biology) and the translation engine will be advised to access those resources first for proper translation of the specific words when they appear in the related context (e.g. "stock" has a different meaning in relation to the category "Fisheries" compared to the category "Business").
- Since the English language is often not sufficiently precise, it is obviously required to replace English source text to facilitate machine translation. Some words have two meanings in English but not in other languages. An example is "to feed" which means both "to offer food to somebody else" and "to eat". This results in very ambiguous translations. The easiest thing to do would be to substitute the problem term with a simpler one (or: rather more precise one) in the English source text (there is no option for an English-English dictionary), in this case feed replaced by "eat" would result into a precise translation. Some of those problems in that context are liable to be associated with the lack of subjects in sentences. "Feeds" is an example of homography, where "Feeds on XXX" could indicate a noun or verb.
- How to apply context-related translation. As mentioned, many words have a different meaning depending of the context in which they are used. Often the translator does not know what translation will apply to an individual word which appears in the translation list without any context. Thus, at least for the lists of words from FishBase, it makes sense, to check the context in which those word appear. This can be done either with the "search" function in the word-document delivered along with the lists and which contains all free text paragraphs from FishBase, or with e.g. the Internet search engine "Google" (other search systems may be applied as well). Just type the word in combination with FishBase and the result will be the species summary, mostly at the first position on the lists of search results, where the word can be considered in the real context. This is also a useful procedure for unknown words. In summary, when proposing a translation, it should be taken into account whether a proposed translation will work in all circumstances in the texts.
- The part of speech of a word is important information for the translation engine. It is helpful if the translator deliver, along with the translation, the respective part of speech of the term, such as noun, proper noun, adjective, or verb. This attribute helps the

technical team of the EC-MT department with the (manual) encoding process for the user dictionaries.

- Sometimes the English language uses two or more words for a term, which is in another language only one word. On the other hand, English terms could not be translated in only one word, because there's no proper word for it, e.g. Great Britain (2 words) – in Dutch: Groot-Brittannië (1 word). The same applies e.g. to "raker" in "gill-raker" or "mid-water" (gill-raker and mid-water should be one entry). Expressions (combination of several words) are certainly welcome as it gives the computer a clue as to how to translate a word in context. Another example in French is "rendre" -> "make" but "rendre un avis" -> "give an opinion". Wherever we consider a group of words as being one semantic concept, then we have to keep them together (noun noun, verb object, verb preposition object, etc.). One other example is "to feed on". In general, expression coding is quite powerful and good results are possible.
- How to treat Family names. Some families do not show English translations in FishBase, only Latin names are given which cannot be translated. In those cases, it is advisable to keep the Latin name in. Also, some English family names don't have corresponding translation into other languages yet. In that case, keep the Latin name and add the common family name in English in parenthesis. This is acceptable for the translation engine.
- "Latinized" words in other languages than English. Not all "latinized" words can be translated. In English, e.g. "Cnidarians" (from the Latin word Cnidaria), have no matching word in other languages, e.g. Dutch has no "latinized" words. In that case the Latin word has to be applied as translation. However it is finally the responsibility of the translator how to manage this item in his/her language, with a recommendation that the common name is given in the translation.
- Translation of common names. Many common names in English have a corresponding common name in other languages. However, sometimes the translation might be misleading. On the other hand, translation is supposed to be as complete as possible because many users may not be able to understand any English common names. So, it would be useful to show the Latin name plus the sounding common name in parentheses. The following rules have to be obeyed: If we enter "Cragon" = "Cragon (Sandgarnele)" then that translation will always appear. But we could solve the problem in the source text: the first time we use the Latin term in English, and we put the English common name in parentheses too. Then in the terminology file, we have to indicate the English common name plus its translations.
Example: In the terminology file: "tursiops truncatus = tursiops truncatus" + "bottlenose dolphin = dauphin souffleur". In the text, first to mention: "tursiops truncatus (bottlenose dolphin)" which should be translated as "tursiops truncatus (dauphin souffleur)"; thereafter just "tursiops truncatus" which will be translated simply as "tursiops truncatus".
- Capital letters: for the sake of automatic coding, it's better to use lower case for entries unless they are really proper nouns or the translation in the target language requires it (e.g. German). If a word can appear as lower or upper case, just enter the lower case. For the automatic dictionary, entering only a word in upper case instructs the machine to match only an upper case version of the word in the source text. If entering "Football" -> "Rugby", that entry would not be matched if the source text contained only "football". If entering "football -> rugby" in the dictionary, it should work for "football" or "Football".
- Abbreviations can certainly harm the translation if they're not recognized by the system: the "." can be interpreted as an end of sentence. So it would be helpful to

include abbreviations which are used frequently. Abbreviations are entered with their final dot when necessary, like aff., sp. Ref., because it may confuse the machine, e.g.: "(Ref. 2834) Status of protection ..." is translated in "(statut Réf. 2834) de protection".

- Some words are doubled with singular and plural: do we have to enter forms of terms, the singular and plural (once needed in relation to the source text)? Automatic coding is expected to guess that the word is a noun, recognize the word in the plural in English (assumes) and then apply an appropriate plural ending in the translation language. This is less of a problem than correctly recognizing "amphihaline" because in the absence of other information, the system will assume that an entry is a noun, and since most English words add -s in the plural, the plural form is easy to recognize. On the translation side, pluralization is not so easy and there is no guarantee that the correct ending will be applied, but it should be obvious it's a plural. It does not help to enter both singular and plural forms. It was tried in tests, and it didn't help: the plural form seemed to be overruled by the singular form. So just put the singular form.
- In the distribution file it is necessary for combinations of country names such as "New Caledonia" to be kept together as an expression, instead of having "New" translated in one line and "Caledonia" in the next one. WP-11 has distributed a corrected file "Distribution" to the translation consortium, which includes almost any country in the world in the correct form. Other systems can make use of this dictionary of countries, regions etc.
- Some English common names refer to groups of species from several families and there may be no corresponding term in other languages: this may render translation impossible. An example is Basslets.
- Do bold letters have an impact on translation? Words in bold shouldn't pose a problem. However, the formatting is not always respected in the translation.
- Options for terms to be added to the given list: If the translation partner considers one term not explained enough in one translated form, you may add a term to the dictionary, example "Aral" appears in the dictionary and may relate to "Aral pubfish" in Fishbase, in other positions it appears as Aral sea (more often). So, just add the English term "Aral Sea" and the translation into your language. Please mark the added terms in the dictionary to notify the coordinator of the dictionaries!
- The adjectives are to be entered with masculine gender, we assume that the EC-MT will automatically make the changes, or should we give indications? For instance, does the system know if "amphihaline species" will be translated in "espèce amphihaline" when we have entered amphihaline 'amphihalin' as the masculine only? A noun, adjective, or verb should always be entered in its base form - so "amphihalin" in French for the example. Automatic coding is then supposed to guess that the word is an adjective in English and to apply the correct ending in French when the adjective is associated with a feminine or plural noun. So just leave it in the basic form. Of course the guessing is limited and doesn't always resolve the part of speech correctly - in a test, "amphihaline" was assumed to be a noun: "amphihaline species" became "espèce d'amphihalin". So if an adjective like that is generally associated with a particular noun, best to include the phrase e.g. "amphihaline species" -> "espèce amphihaline" to ensure the correct translation.

ANNEX D: Glossary of terms used in the context of machine translation.

A useful glossary of terms and phrases used in the world of machine translation, specifically for the SYSTRAN translation facilities.

Acronym: In SYSTRAN terminology, a word, all upper case, formed from the initial letters of other words or parts of a series of words, such as *WAC* for *Women's Army Corps*.

Adjective: A word used to modify a noun by limiting or qualifying. In English, it can be distinguished by one of several suffixes, such as *-able*, *-ous*, *-er*, and *-est*, because it directly precedes a noun or noun phrase, such as *red* in *a red door*, or because it is preceded by a form of *to be*, such as *the door is red*.

Adverb: A word that modifies a verb, an adjective, or another adverb, such as *brightly* in *The sun shines brightly*.

Article: A word used to signal a noun and to specify its application. In English, the definite article is *the*, and the indefinite articles are *a* and *an*.

Auxiliary verb: A verb, such as *have*, *can*, or *will*, that accompanies the main verb in a clause and helps to make distinctions in aspect, mood, tense, and voice.

Clause: A group of two or more words which contains a verb. One or more clauses make up a complete sentence.

Direct Mode: An interactive mode of translation that allows the translator to enter text to be translated in the SYSTRAN source text window, and to see the translation in the SYSTRAN target text window. See File Mode.

DNT: An acronym for Do Not Translate.

DNT List: A list of words created by the translator that will not be translated by the SYSTRAN Translation System.

Do Not Translate: A translation option that allows the translator to choose to leave certain words untranslated in the Source text.

Document Type: A translation option that allows the translator to specify the general style of the text to be translated.

Embedded clause: A subordinate clause that is embedded in the middle of a main clause. For example, *The camera, which I bought yesterday, is already broken*.

Expression: In SYSTRAN terminology, a noun phrase consisting of more than one noun or any combination of nouns and adjectives, where the Head Word is a noun, such as *lug nut*, *Library of Congress*, and *red letter day*. Complete sentences are not valid SYSTRAN expressions; neither are any phrases that contain verbs.

Expression Dictionary: The SYSTRAN dictionary that contains expressions. See Stem Dictionary.

File Mode: A non-interactive mode of translation that allows the translator to send a file to be translated and to choose the translation options for it. See Direct Mode.

Finite verb: A verb form that is limited in tense, person, and number, such as *goes* in *He goes*.

First person: A pronoun that refers to the speaker. For example, *I* in *I see*, or *We* in *We are*. Also, a verb form that refers to the speaker. For example, *am* in *I am*, or *are* in *We are*. Verb inflection rarely indicates person in English, but in other languages, it often does.

Head Word: In SYSTRAN terminology, the noun in a noun phrase which may change in the plural, such as *nut* in *lug nut*, *Library* in *Library of Congress*, and *day* in *red letter day*. The Head Word is not necessarily the same as the Principal Word.

Homograph: In SYSTRAN terminology, one of two or more words that have the same spelling and are different parts of speech (for example, noun and verb, or adjective and verb). For example, *head* as in *head west* or *on the head*, and *light*, as in *the light box* and *light the match*.

Imperative: A verb form that is used to express an order or command. For example, *Eat* in *Eat your vegetables*.

Infinitive: A verb form that is the ordinary dictionary-entry form. In English, it is often used with “to” as in *He wants to eat*. It may also occur without “to”, for example, *get* in *I made them get in line*, or with auxiliary verbs such as “must” as in *We must leave*.

Inflection: An alternation of the form of a word by adding suffixes without changing the basic meaning or part of speech, as in making *rugs* from *rug*, or by changing the form of a base word, as in making *ate* from *eat*.

ISO639: ISO standard codes for the representation of names of languages.

Main clause: A clause that can stand alone as a complete and correct sentence. For example, *It was raining*.

NFW: The acronym for Not-Found Word.

Not-Found Word: A word or string that occurred in the text, but not in the dictionaries used in translation.

Not-Found Word List: An alphabetical list of all words in a text that were not found in the dictionaries used for translation.

Not-Found Word Marker: A mark that the translator can select, which appears in the translation to indicate a Not-Found Word.

Noun: A word that is used to name a person, place, thing, quality, or action, for example, *house*, *flammability*, or *movement*.

Noun phrase: A phrase that functions as a noun and that has a noun as its head word.

Parse: To break (a sentence) down into its component parts of speech with an explanation of the form, function, and syntactical relationship of each part.

Parser: In SYSTRAN terminology, the module of the computer program that performs syntactic or semantic analysis of the Source text.

Participle: A verb form that can be used with an auxiliary verb. It can also function as an adjective or a noun. See Past Participle and Present Participle.

Past participle: A verb form that indicates past or completed action. It can be used with an auxiliary verb as in *The cake was baked yesterday*, or as an adjective, as in *baked beans*.

Phrase: A sequence of two or more words which express an idea. See also, Noun Phrase, Verb Phrase, and Prepositional Phrase.

POS: The acronym for Part of Speech.

Preposition: A word placed before a noun or noun phrase, indicating the relation of that noun or noun phrase to a verb, an adjective, or another noun or noun phrase, such as *at, by, in, to, from, and with*.

Prepositional phrase: A phrase that consists of a preposition plus the noun or noun phrase that it governs, such as *at the park, in reference to your letter, or from Mars*.

Present participle: A verb form expressing present action, formed in English from the infinitive plus *-ing*, which can be used with auxiliary verbs as in *He is baking a cake*; as an adjective, *the baking rack*; or as a noun, as in, *the act of baking*.

Pronoun: A word that functions as a substitute for a noun or a noun phrase and designates persons or things asked for, previously specified, or understood from the context. For example, *I, it, that, and which*.

Proper noun: A noun belonging to the class of words used as names for individuals or places. For example, *Clinton* or *Boston*.

Run-time dictionary: The binary indexed dictionary file used during the translation process.

SDM: The acronym for SYSTRAN Dictionary Manager.

Second person: A pronoun that refers to the listener. For example, *you* in *You see* or *You are*. Also, a verb form that refers to the listener. For example, *are* in *You are*. Verb inflection rarely indicates person in English, but in other languages, it often does.

Semantic: Having to do with meaning.

Sentence: A grammatical unit that is syntactically independent and has a subject that is expressed or understood and a predicate that contains at least one finite verb.

Source: The language of the original text, before translation.

Subjunctive: A verb form that indicates possibility, doubt, or desire rather than fact. For example, *were* in *I wish it were true*, or *start* in *I suggest you start immediately*.

Subordinate clause: A clause that modifies or expands on other clauses. A subordinate clause modifies or expands on another clause. It cannot stand alone, as *that he gave* in *The account that he gave was true*.

Subordinate clause marker: A word which indicates that a clause is a subordinate clause, such as *that, which, or who*.

Subordinate phrase: A phrase that modifies or expands on other phrases.

Syntactic: Having to do with sentence structure.

Target: The language into which the text is translated.

Third person: A pronoun that refers to neither the speaker nor the listener. For example, *He* in *He is*, *She* in *She is*, *it* in *it is*, or *They* in *They are*. Also, a verb form that refers to neither the speaker nor the listener. For example, *is* in *He is* and *She is*, or *are* in *They are*. Verb inflection rarely indicates person in English, but in other languages, it often does.

Specialized Dictionary: A translation option that allows the translator to choose appropriate subject areas for the document.

UD: The acronym for User Dictionary.

User Dictionary: A dictionary used in translation that has been created by the user for the purpose of tailoring translations to his specific needs. Definitions in a UD override the ones in SYSTRAN's dictionaries.

Verb: A word that expresses existence, action, or occurrence.

Verb phrase: In SYSTRAN terminology, a phrase or other construction used as a verb. Verb phrases are not acceptable entries in a UD.