Peter Sandrini
University of Innsbruck

From Scratch to the Web: Terminological Theses at the University of Innsbruck

Terminology Diploma Theses (TDT) have been well established in the training of translators and interpreters over the last 20 years. A few aspects, though, changed in the methodology of work and contributed to an important re-evaluation of TDT. New possibilities of terminology management and the overall progress of information technology led to two decisive developments:

a) Terminology Management Systems improved the efficiency and productivity of terminographical work and led to an adaptation of terminographical working procedures.

b) The advent of the Internet and especially the world wide web as a new tool for information and communication provided new possibilities for

- gathering terminological information
- coordinating terminological activities
- publishing TDT at minimal costs

Terminology diploma theses should meet a range of requirements within the framework of our translator training program. The first and foremost objective is a didactic one. After an introductory lecture and a seminar on terminology management systems students should be able to do a terminology project on an independent and autonomous basis. The diploma theses provide a first testing ground for the practical application of their knowledge with the possibility of guidance and correction. Thus a TDT is a test for the application of methodological principles.

In no case should TDT be mere academic examination papers or sterile academic exercises destined to disappear in the university library. Practical relevance is therefore very important and constitutes, among other criteria, part of the evaluation process.

The Translation Department at the University of Innsbruck tries to achieve the practical relevance of terminology work done by students through cooperation efforts with potential users of terminology. A strong commitment to the practical needs of multilingual communication, especially in the regional context where contacts with organisations and companies can be established more easily, constitutes a primary objective of the department. This contacts with the „outside“ are either established on a personal basis by the students themselves, and in these cases they are restricted to one single diploma thesis, or as part of a cooperation between the university department and an outside partner: this could be large and medium sized companies or institutions and organisations with a need for multilingual terminology. In this case several diploma theses in the same subject field are done in cooperation with the partner.

The overall process of compiling a TDT encompasses the following steps:

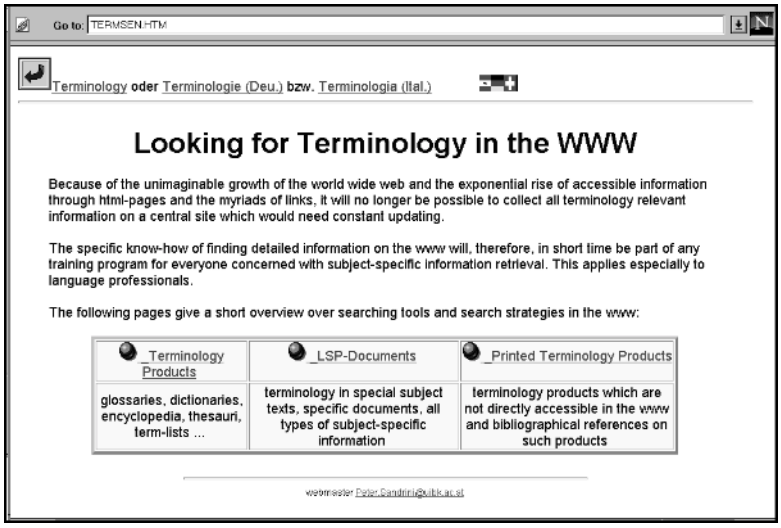
1) choosing of a specific domain and working languages
2) checking existing resources
3) collecting textual material for documentation purposes
4) analysing documents
5) structuring the specific domain (concept systems)
6) compiling terminology records (concept related data, term related data, other types of information)
7) intralingual comparison: homonymy, synonymy
8) interlingual comparison
9) revision
10) printing
11) publication

We should like to comment each step of the overall process of compiling a terminology diploma thesis with regard to the support by terminology management systems and in particular the use of the world wide web at our department.

1) Choosing a particular domain in which students wish to do their terminology work is a task which is very subjective and controlled by personal preferences or contacts. Nonetheless the choice of a domain should not depend solely on the personal preferences of a student but take into account also the practical relevance of a terminology project in this domain: both aspects are extremely important and have to be weighed up. The languages involved depend on the languages chosen as first or second language in the translation training program: TDT are mostly bilingual with some involving three languages and a few four or more languages.

2) Once the domain and the languages have been agreed upon, students should check whether terminology products for this particular subject field and the chosen languages already exist. This is extremely important to avoid the costly duplicating of work. Such a search for glossaries, dictionaries, either on paper in libraries or on floppy disks or even online, represents a very useful exercise for future translators or interpreters since in their professional life they will have to do this particular search all the time.

This is a task for which the web and the whole Internet brought along new tremendous possibilities which integrate traditional means such as a search in libraries or catalogues. The translation department put some pages on the web as a part of its terminology server (http://info.uibk.ac.at/c/c6/c613/termlogy/termsen.html) to give students a guide for this particular task.



The starting page (http://info.uibk.ac.at/c/c6/c613/termlogy/termsen.html) leads to three different possibilities for finding terminology in the world wide web: a) terminology products, i.e. dictionaries, glossaries, termbanks, b) printed terminology products, i.e. bibliographical references to such products, or c) terminology in LSP-texts. The first two are relevant in this phase of the working process and the results of such a search should exclude that specific terminology products exist, or detect specific glossaries which could be analysed whether they comply with the requirements of the needs of users. The web pages guide students in their search and list the different possibilities: specific lists of links to dictionaries or glossaries on the web, terminology databanks accessible online or search engines, either general ones or language specific search engines. To find references to printed dictionaries another web page refers to the online catalogues of publishing houses, of libraries and of university networks, to the web catalogues of national and international standards organisations, to collections of data banks and listings of terminology diploma theses.

3) Collecting textual material for subject specific knowledge acquisition and as a source for terminological information is one of the main tasks in terminography and is done primarily by traditional means, i.e. going to the library searching through catalogues and getting books, journals or articles, by talking to subject field experts who recommend texts, or by contacting practising individuals in companies or organizations.

The world wide web constitutes an immense worldwide textual data base with documents from all subject fields. As such it lends itself excellently as a means for all kinds of textual research. A search for LSP-relevant documents on the www integrates traditional means of collecting sources for terminology. Appropriate search instruments and a suitable search strategy are very important in this context and are outlined on the department's web page „terminology in LSP-texts". Web searches imply web surfing which can be a very time consuming activity. Thus students get an overview over search possibilities which should guide and focus their searching activities.

Catalogues and directories provide a structured and organized hierarchy of categories for browsing by subjects, each category or sub-category collects links to specific web pages. One list of such links opens up a whole new universe of www-based documents on this particular subject field. Such subject trees were initially planned as general catalogues and focussed on the English speaking web space, meanwhile there are quite a few language specific indexes available.

Search engines feature indexes that are automatically compiled by computer programs, such as robots and spiders, that go out over the Internet to discover and collect Internet resources. Searchers can connect to a search engine site and enter keywords to query the index. These keywords should be chosen carefully, a specific term central to the subject field or a combination of two or more such terms would be appropriate for a first try. The more specific the keyword, the better the search result and the more noise (insignificant results)will be avoided.

Obviously the quality and reliability of the textual information on the www should be checked carefully. Some web sites disappear after a short period of time, others are heavily biased or one-sided and represent minority views. A critical look at the authors of the page or the institution which owns the web site helps to assess the quality of the provider. A good criterion for reliability could be represented by the references or links from other subject specific web sites to this site.

4) Phase four of the overall process is characterized by a strong intellectual component in identifying e.g. definitions, synonyms, terms in LSP-documents. This is a task where there could be more support from the machine in future as demonstrated by recent research in the field of cognitive linguistics or term extraction methods. Linguistic text analysis tools can do a variety of quantitive linguistic analyses, only very few, though, can be of use in terminography.

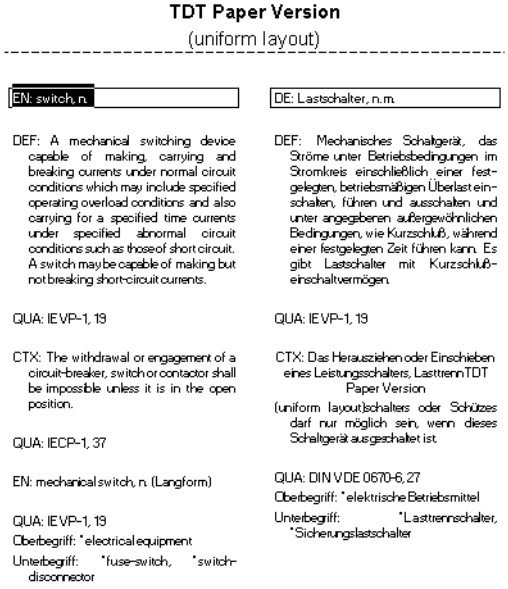
5) New tools in terminography tend to cover more aspects of terminology work that would include phases 4 and 5, provided there is a corpus of machine readable texts. Some research-oriented developments aim at providing appropriate tools for the management of conceptual relations and concept systems, other researchers concentrate on extracting such data on conceptual relation from texts. Again, we hope for significant progress in this respect.

6) Phase six represents the core application of terminology management systems. In an introductory course, students get an overview over different types of terminology management systems and the criteria for evaluating commercial systems. They also learn how to work practically with such a system. At our department we use a commercial TMS which has a relatively dominant position on the market so that students have the chance to get experience with such a system which at high probability they would use again in their professional life.

The same piece of software is used for the management of a terminology database at the department. This implies a standardized entry structure which means that students use a predefined set of data categories for their TDT. The advantages of such a unified database are clear: all TDT can be collected in a single database with standardized data input as well as a uniform printing procedure and a common layout for the paper version. A description of the database and the entry structure is in our web site so that students can go through it from wherever they are. The same database could be used as an example for new TDT, as well as a multilingual terminology resource for translation seminars.

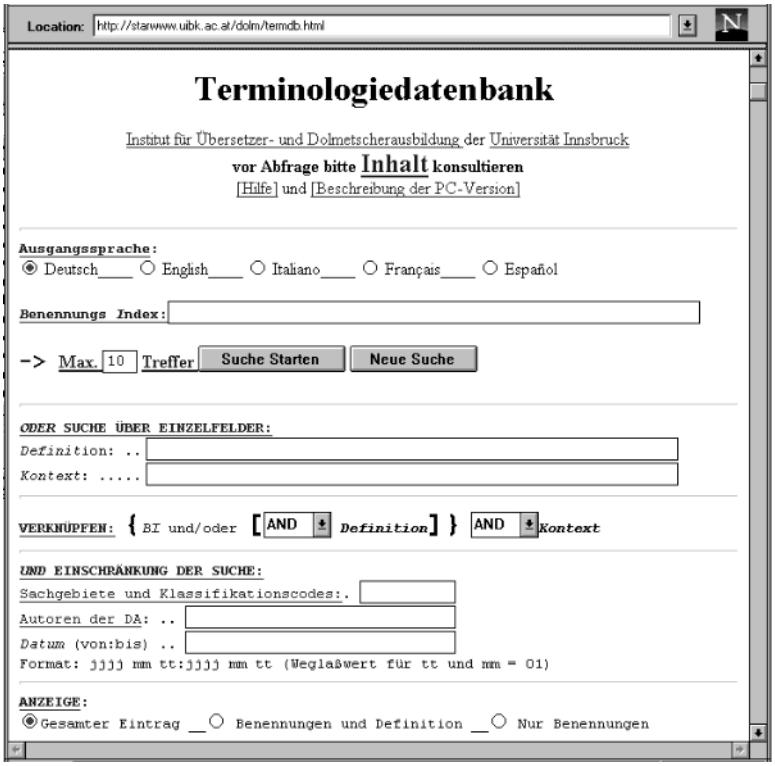
7) 8) These steps in the terminographical process are of fundamental importance in multilingual terminography and require much brainwork and intellectual intervention on the side of students. This is a task where suppport from computers would be very difficult to achieve.

9) 10) Revision and Printing in Innsbruck are supported by the wordprocessing software which processes the terminological records exported from the terminology managament system and brings them into a uniform layout format.



11) At our department we thought that diploma theses should be made available to the general public, first of all to test their practical relevance, secondly to promote terminographical work by giving students the possibility to publish their theses. The active use of their own terminology in translation courses or even outside the university constitutes a strong incentive for students.

The world wide web represented to us an ideal medium for publication since the costs are very small and the potential audience very large. Nonetheless two provisions had to be taken: publication should be done on a voluntary basis and there should not be the possibility to download diploma theses as a whole. To cater for the first case, students are free to sign a declaration of consent after they finish their TDT. To prevent the downloading of a whole TDT restrictions in the query form were build in so that for each query only up to 10 answers are given.



The terminology databank is implemented as a unix database with a web query interface. This web based terminology database was conceived as an additional tool for students and users outside the university. The software used for data input and data management remained the same at the department and at regular intervals the data are exported to the web based term bank.

The combined technology (windows, unix, cgi and dynamic html) has the advantage that students still work with a common TMS under their usual desktop environment. They do not need to bother with programming codes or new database systems and the system allows easy updating. General public access requires no special hardware or software equipment on the side of the user, a www-browser is enough.

At first students were skeptical about publishing their work in the Internet, but now more and more students are willing to give their data, mainly because they see that their TDT will be of practical value to colleagues and outside users.

The terminology data base can be easily integrated into translation classes and it serves as an example of practical terminography in the terminology training.

Since the terminology data base collects theses from completely different subject fields, ranging from cattle breeding to company law and word processing, and these subject fields are restricted to 100-150 entries each, the term bank can never fulfill completely the expectations of practical users. The objective of this term bank is a more didactical one and this is shown clearly by the access statistic figures of the last four months:

month	total	uibk	o. univ.
April 97	340	19%	19,4%
May 97	348	31%	8%
June 97	587	18%	19,4%
July 97	470	8,5%	16,3%

The second column shows the total number of accesses to the term bank for each month, the second shows the percentage of internal users (accesses from within the university of Innsbruck) and the last column gives the number of accesses from other universities in Austria or elsewhere as far as this can be derived from domain names. About 40 % of total users come from universities which shows clearly the didactic use of the term bank.

The world wide web constitutes an excellent communication tool to get information and to offer information. As communication experts, translators and terminologists should be aware of the new possibilities and be able to handle both aspects: the first by controlling search strategies and information retrieval procedures, the second by understanding the fundamentals of HTML and functioning of the web. In our terminology training within the translation department we try to include the world wide web as described in this paper primarily as a support for compiling terminology diploma theses.