

Anita Nuopponen

Terminology Online: From Term Banks to the World Wide Web

Introduction

For terminologists, computers have always been natural tools and expectations for their technical development was high very early on. The computers are seen as tools to store and retrieve terminological information as well as as means of coordination.[1] In the seventies these could be done, but there were many shortcomings. Retrieval systems were complicated. There were no possibilities for graphical information and limited possibilities for systematic representation. Co-operation consisted of collecting information and distributing it as paper documents. In 1970 Eugen Wüster was convinced that there would soon appear on the horizon, as *fata morgana*, computers that contained all terms and dreamed of a global "Blitzwörterbuch" accessible from every corner of the world [1]. He had right.

Today there are terminology data banks that Wüster dreamed about but in addition, specialist knowledge and terminology from almost any field is available also in the World Wide Web (WWW). In this article I am going to discuss these developments and their meaning for terminology work and distribution of terminological information. First I shall, however, give some historical background of the use of computers for storing and distributing terminological information.

Terminological data bases proper

In 1978 there was fourteen terminological data banks in the world [2] and several others were founded after that. Hvalkof divides these term banks into translation, standardisation, language policy and linguistic research oriented data banks.[2] As to the contents of these terminological data bases proper, they are not only collections of terms, definitions and translation equivalents, but have a fixed structure for the term records. Term records include various data among other things about term (grammar, collocation, use, synonyms), concept (definition, relations to other concepts), translation equivalents (degree of equivalence etc.) and about the origin of the term record (date, author etc.). These data bases have an advanced retrieval system often giving a wide range of search options.[3]

In the first phase the terminological data was stored on magnetic tapes or other off-line storage system. Later on, data bases were locally available online and the next step was to offer them available by modem, mostly for charge or with a special permission. At the moment there are several terminological data banks that can be accessed by direct dial, Datapak or Internet. They are protected by a password and charged according the connect time (e.g. Tapa in Finland, TERMDOK in Sweden[4]), while eg. EURODICAUTOM[5] is freely available in Internet. CD-ROM technology is also adapted for distribution of the information stored in the terminological data banks (e.g. TERMDOK CD-ROM, TERMIUM, B.Q.T - Banque de terminologie du Québec[6]).

Common to these terminology data banks proper is also that they are institutionally organized enterprises involving often also production of printed glossaries. Exchange of data between the term banks has also been successful since eighties and common exchange formats have been created [7]. All in all there is much work behind these data bases.

Terminological data bases in the Internet

When the Gopher system, and later on the WWW, were introduced in Internet, new possibilities opened for the terminological data banks and the terminological cooperation. There are already different kinds of terminological activities online, but the established terminological data banks are not the first ones to be offered through this forum. Many of these data bases are supposed to support themselves and they can't be offered free, which is the case with most of the material in the WWW. At the moment the WWW makes it possible to charge for the services offered online.

The term bank of the European Union, EURODICAUTOM, could be reached free over Internet very early. It has also received an experimental WWW gateway as an unfunded project at the University of Frankfurt[8]. The data base has not been transferred to a WWW-data base, but instead, the researchers are developing a link, an interface, between the the original data base and the WWW. Here are still many problems to be solved and many of the features of the database can't be offered through the WWW interface.[8]

In addition to the world wide availability, terminologists have been dreaming for a long time of being able to add picture, video, animation and sound in a terminological data bank[9]. Drawings, pictures, figures etc. have always been an essential component in printed technical or other specialized vocabularies, because they can replace long definitions. Multimedia technology enables now to combine also these elements to the vocabularies. However, at the moment there are not yet many multimedial data bases, but it is very probable that there will be soon.

Terminological sources in the WWW

Even though the term banks proper are still missing, there are already several dictionaries and glossaries for online searching. There are plenty of general language dictionaries[10] but the amount of special language glossaries and vocabularies[11] is growing even more rapidly. This could be noticed during over one year's time that I have been collecting links to them. The first ones online were dictionaries like Webster[12] and Roget Thesaurus[13]. Most of the WWW is in English and so are most of the glossaries and dictionaries, too. A big problem for many other languages is the restricted default character set that covers only the needs of the English language.

As soon as the Gopher and the WWW became popular, glossary activities started to flourish online. Both private and cooperative efforts are made to publish both general

language dictionaries and special language vocabularies online. They are compiled by one person, university institutions, research institutes or research groups, associations or companies. In addition to these there are also projects that exist only in Internet. Cooperation is often conducted on voluntary bases by those interested in the same field. As one example I could mention *Free On-Line Dictionary of Computing* (FOLDOC)[14]. For very specialized subject fields producing an online glossary seems to be a way to connect the geographically widely scattered experts.

Specially for otherwise "hidden" terminological efforts the WWW seems to offer an excellent publishing media. Many glossaries that are collected in research projects or during courses remain often hidden from other users. They might be distributed as xerox copies or printed as an enclosure of the project reports. One group of usually hidden glossaries is the internal companies glossaries. Companies are, however, more and more starting to provide a glossary of their central terminology on their home pages.

Some commercial hard copy dictionaries are online in electronical form like the Dictionary of Cell Biology[15] published by Academic Press Limited 1995. It is free to "individuals for purposes of personal study only" and information from this service "may not be saved, printed, copied or re-transmitted in any form" without consent from the publisher. This dictionary server like many other services that are available free online is apparently ment to bring feedback from the users for further development of a commercial electronic version. It is predicted that Internet will be the main development platform for all kind of electronic services and software, so why not for dictionaries! Some commercial publishers give only a free sample of their printed glossaries or CD-ROM versions in the WWW (e.g. Multilingual Dictionary of the Horse)[16].

How to find terminological data

There are already hundreds of glossaries online and new ones appear all the time. It is an effort to find all the relevant information in the Internet. Search engines (Alta Vista, Yahoo, Lycos etc.) can be used for this purpose but there are also several collections of links to the online sources in the WWW[17]. These collections, however, are not always very well organized and seldom make any difference between general and special languages or classify glossaries according to the field. They might mix full texts, language courses, touristic information etc. between the links to online dictionaries or hide the dictionaries between other kind links.

I started my own collection of links to online dictionaries and glossaries and concentrate mostly on special languages. These collections belong to an information service called *Terminology Forum*[18] that contains different kind of information about terminological activities:

- terminology centres, international cooperation, terminological standardization
- terminological research and development, terminology projects in subject fields
- teaching in terminology science

- terminological articles online, bibliographies
- online terminological data bases and glossaries, general language dictionaries
- conferences and seminars
- related fields etc.

The amount of users of this service is growing rapidly. The most frequently visited is the collection of general language dictionaries. I am updating the collections quite continuously and often with the help from the users who send information about new dictionaries and glossaries they have found or compiled. Follow-up of the links would be very important but I have been concentrating so far mostly on collecting new links. Addresses are changing now and then and documents are withdrawn from the WWW e.g. for copy-right reasons.

Evaluation

From user's, e.g. translators or text writers, point of view it is problematic not only to find the right glossary but also to evaluate it. Evaluation of the reliability is a problem that is accounted also anywhere else in the WWW. What we can do in the WWW is to try to find out information about the origin of the documents. Often it is not an easy task to retrace the editors or other background forces or to find description about the principles or methods used. There is mostly an e-mail address to a person, but surprisingly often other - or even all - information is forgotten or is to be found somewhere else than on the dictionary or glossary page.

Many of the authors still plan their document collection in the same way as a printed publication, i.e. they take it for granted that everybody visiting their pages enter through a certain home page. Different kind of search methods, however, can lead the information seeker to any of the pages without any warning or hint to an "official entrance". For my collection I have started to retrace the authors and information about their expertise in the subject field. I try to give names and links not only to the glossaries but also to the authors and institutes behind them. The information seeker can then faster decide how reliable the glossary is. Because nobody is controlling the quality of the documents that are published in the WWW, the decision on the reliability lies on the information seeker.

Conclusion

The future as predicted by Wüster has arrived and the networked computers can be used to store and retrieve terminological information world wide without any bigger time delays. The WWW holds a great potential for terminological activities, not only for storing the results of a project or access to the otherwise hidden sources, but also for cooperation and development. In order to make it easier to access information clearing houses like Terminology Forum are needed. They are virtual libraries or bookshelves where information sources from all over the world are ordered and could also be reviewed and commented.

References

1. Wüster, Eugen (1970). Die internationale Terminologie im Dienste der Informatik. In: *Monda Lingvo-Probl.*, vol. 2.1970, 138-144. p. 143.
2. Hvalkof, Sonja (1985: 8). *Etude comparative des donnees terminologiques des banques de terminologie DANTERM, B.T.Q., EURODICAUTOM, NORMATER, O.F.L. et SIEMENS*. Handelshøjskolen i København.
3. As an example see <http://info.uibk.ac.at/c/c6/c613/termlogy/anleit.html>, where the terminological data bank in Institut für Übersetzer- und Dolmetscherausbildung at the University of Innsbruck is described.
4. See more: <http://info.rbt.no/nordguide/1995/ms/ms.83>.
5. See: <http://www2.echo.lu/echo/databases/en/eu92.html>
6. See more: http://www.olf.gouv.qc.ca/txt_btq.htm
7. See more: Reinke, Uwe (1993) *Der Austausch terminographischer Daten*. IITF/Termnet, Wien.
8. See: <http://www.uni-frankfurt.de/~felix/eurodicautom.html>.
9. Nordterm 1983. Seminarium: Datorstött terminologiarbete 3-5.5.1993. Swedish Centre for Technical Terminology, Sweden.
10. See e.g.: <http://www.uwasa.fi/comm/termino/diction.html>.
11. See e.g.: <http://www.uwasa.fi/comm/termino/special.html>
12. See e.g.: <http://c.gp.cs.cmu.edu:5103/prog/webster>.
13. ARTFL Project's hypertext version:
http://humanities.uchicago.edu/forms_unrest/ROGET.html
14. Available: <http://wombat.doc.ic.ac.uk/foldoc/editing.html>.
15. Available: <http://www.mblab.gla.ac.uk/~julian/Dict.html>.
16. Multilingual Dictionary of the Horse: <http://www2.zone.ca/~bouletjc/>
17. See <http://www.uwasa.fi/comm/termino/others.html>
18. Terminology Forum: <http://www.uwasa.fi/comm/termino/>