
German UDC Translation Project

Aida Slavic, UK (aida.slavic@udcc.org)

Jiri Pika, Switzerland (pika@library.ethz.ch)

Gerhard Riesthuis, The Netherlands (griesth@xs4all.nl)

Chris Overfield, UK (chris.overfield@ntlworld.com)

Since 1990 there have been several UDC editions (electronic or printed) in English, Spanish, Russian and French as publishers from the U.K., Spain, Russia and Belgium have been long standing members of the UDC Consortium. New editions also appeared in countries where UDC was widely used and the translation and publishing of UDC was financially viable, e.g. the Czech Republic, Hungary, Estonia, Lithuania, Romania, Ukraine, Serbia, Croatia, Slovenia, etc.

In many countries where UDC has not been used in the majority of libraries, however, there have been no publishers interested in UDC translation and publication. The costs of translation and the fee for the publishing license that publishers have to pay to the UDC Consortium are often a financial burden on top of the production cost that makes such a publication a financial loss when there is a limited market. Thus, it is not strange that there have been no new editions in, for example, Danish, Finnish, German, Italian, Catalan, Chinese, etc. since the 1970s and 1980s when these editions were subsidized by various national standard, documentation or library organizations (cf. Slavic, 2004).

The problem of new UDC editions in countries where UDC is not widely used was stressed by Miguel Benito (Benito, 2001). The lack of new editions drives away the remaining UDC users in many countries and forces them to change to other systems that provide better support in their respective languages. This problem is compounded by the fact that library schools lacking training materials and easy and cheap access to UDC are discouraged from teaching UDC and this has a further knock off effect on the usage of the UDC.

The lack of new German editions

In 2008 it will be exactly thirty years since the last German edition was published. The lack of an up-to-date German edition was perceived by the UDCC as a much more serious problem as this affects users in Germany, Switzerland, Austria and Liechtenstein. Although libraries using UDC in these countries may not be significant in number they are typically large academic and research libraries with collections that are very relevant on a national and even international level. This is certainly the case with the Eidgenössische Technische Hochschule (ETH) Library in Zürich with a collection of almost 2.5 million documents. Equally important for their regions and respective countries are collections classified by the UDC in the National Library of Liechtenstein, the Austrian University of Linz Library and Graz University of Technology Library, or the German library of the University of Technology Library in Freiburg, the Deutscher Wetterdienst (DWD) in Offenbach am Main, etc. The last German edition available to subject specialists in these libraries is the medium edition in two volumes from 1978 (Dezimalklassifikation, 1978-1985).

Initially, in 1992, the idea was that after the Deutsches Institut für Normung (DIN) ceased its role as publisher of the German edition, the vacuum would be filled in by DACH: a user

club consisting of users in Germany, Austria and Switzerland (Gilchrist, 1993). When it became evident that this would not happen, CEFAL who is one of the UDC Consortium members and the publisher of the French edition proposed producing a much needed German edition. Unfortunately, CEFAL's plans could not be realized.

The disappointment of German UDC users about a new German edition was put forward emphatically by several participants of the Librarian Workshop in conjunction with The 31st Annual Conference of the German Classification Society on Data Analysis, Machine Learning, and Applications held in Freiburg in March 2007. In the discussion that followed the invited talk on UDC (Slavic, Cordeiro & Riesthuis, 2007) it was proposed that the UDCC should act upon this problem as a matter of urgency. The example of a Dewey translation which materialised in three years was mentioned and it was felt that the same work should be even easier for UDC, where good quality editions had already been published in the past (c.f. Heiner-Freiling, 2003, 2006). In a discussion session at the end of the UDC Seminar in June 2007, the urgency of a German translation of the UDC was put forward again. At this point it was suggested that in the absence of a publisher, the best approach could be to initiate a translation project to be organized as an international effort engaging volunteers of institutions and individuals. Once the translation would be available it would be easier to find a publisher.

This proposal was discussed at the UDCC executive meeting in June and the Consortium members agreed that a UDC MRF which already contained around 2000 records in German should be made available for such a project and that institutions such as the ETH should be contacted and offered UDC data and leadership of the project. The subject authority file of the ETH library containing around 60,000 UDC records mapped to descriptors in German, French and English was considered a valuable source of terminology that can be very useful in the translation process. One important decision that came from this meeting was that the UDC Consortium granted the UDC MRF data and translation license providing that once produced, German data would be stored as part of the UDC MRF irrespective of who is going to publish the actual German product. If data stays with UDC Consortium and not with an individual publisher this may be a more reliable solution for future editions.

Following this meeting, an invitation for cooperation in such a project was addressed to the ETH library as it was felt that this institution would be best at coordinating the efforts in German speaking countries. Members of the UDC Editorial team, Jiri Pika, Gerhard Riesthuis and Aida Slavic volunteered to do the preparation for the translation project by the end of 2008. Also Chris Overfield who has previously lent his programming skills to UDC Consortium and is familiar with UDC MRF volunteered to help in the creation of a translation management tool.

Preparations for the UDC German translation

Initial research by J. Pika showed that potential collaborators may be found in institutions in Switzerland, Germany and Lichtenstein. It was, however, felt that in order to start the project some steps ought to be taken in advance. The first was to collect and analyse UDC data that is already available in German and consider a tool that may be needed for such a work. This initial preparation would then provide a sound basis for planning the project, inviting potential collaborators and even looking for possible funding.

Several important parts of the preliminary preparation started in 2007 involving the efforts of the four volunteers mentioned above:

- importing into the UDC MRF the existing electronic data in German
- preparing for the digitization of the relevant parts of the last UDC edition in German
- looking into the possibility of getting permission to use the UDC authority file from the ETH library to harvest vocabularies
- creating a translation tool for the project

The 2006 edition of the UDC MRF contains more than 67,000 records and only 2,300 of these have parallel text in German. When checking the UDCC files archive in September 2007, German UDC data files were found containing around 87,000 UDC numbers, apparently from the German full edition. These files were produced in 1989 and were already formatted for import to the UDC MRF. G. Riesthuis has analysed the files and when comparing them with the current UDC MRF he found that around 30,900 records corresponded with existing data and could be automatically matched while the rest of the files, although containing valuable vocabulary, ought to be reviewed and matched intellectually. In September 2007, G. Riesthuis imported the German data into the UDC MRF which now contains 46% of the text in German.

At the same time J. Pika prepared copies of the last German medium edition from 1978 plus its subject alphabetical index from 1985 and made them ready for digitization. He also provided a sample of the ETH subject authority file so that it could be considered in the process of developing a translation management tool.

The UDC MRF will contain areas that have not changed since the last German edition for which German text will already be present. There are plans to proofread these areas and this will be flagged as a separate task. German text is lacking for over 36,000 classes that either did not exist in the last edition or have existed but have undergone changes in the period from 1978 to 2008.

Translation management tool

The translation management tool being created by C. Overfield is a Windows Forms based .Net application written in C# which uses the UDC MRF database records extracted to a file which is then used as input. Although at the moment it is created for helping with the German translation, this tool is intended as a UDC generic translation management tool. It allows controlled input of text based fields in multiple languages with exports in different formats, the most important one being a UDC MRF compliant export.

The logic of a translation work requires that the tool contains four main parts:

- a UDC MRF source file, available for browsing and navigation;
- an editing area, displaying record by record text in English and fields to fill in with German text;
- easy to use control functions for saving, flagging (to revise), selecting and exporting records;
- supplementary vocabulary help files (e.g. UDC data in German from other German editions or authority files).

Currently, the UDC MRF does not have a subject alphabetical index and this is planned to be created in the next two years. For this reason there will be an open subject alphabetical index field for German. Relevant terminology will be harvested from the subject-alphabetical index from the last German printed edition and we also hope to be able to complement this with descriptors from the ETH subject authority file.

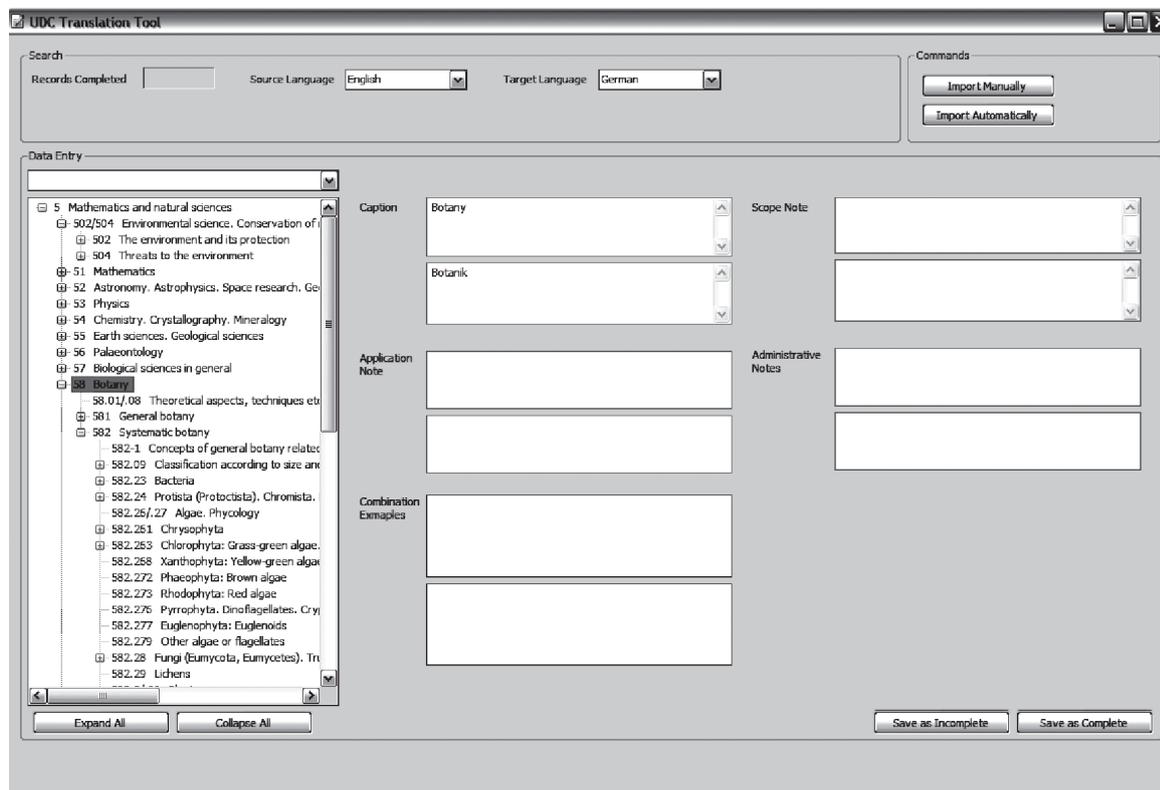


Figure 1 The interface of the translation tool

For the purpose of a translation process which is going to be distributed and collaborative, the UDC MRF will be divided into smaller subject areas and translation tool packages containing specific subject data may be prepared in advance. In the process of translation contributors would have several options for saving and flagging the text for further checking. Completed and saved translations would be easily exported as formatted text and exchanged by email to be compiled centrally.

Summary

Thanks to the volunteers who initiated the idea and the prompt response from the UDC Consortium in giving permission and open hands for the use of the UDC MRF, significant steps towards the UDC German translation have already been made.

We are starting 2008 with 46% of the UDC MRF already in German and with a good progress in the development of the translation management tool. Our main task in 2008 will be to find collaborators: individuals and institutions (libraries, library schools or information and documentation centres) in German speaking countries, willing to join the project and

contribute. Once connections are made and people involved we may consider looking for funding and for potential publishers. If collaborators are found we hope to be able to finalize the majority of work by the end of 2009.

Experience with the UDC German translation will offer important experience, expertise and tools that may help other translation projects in the future.

References

Benito, M. (2001) The UDC in Sweden. *Extensions and Corrections to the UDC*, 21, pp. 23-24.

Dezimalklassifikation : internationale mittlere Ausgabe / herausgeber, DIN Deutsches Institut für Normung E. V. - 2. Aufl. der DK-Handausg. I serie: FID ; 550. Berlin : Beuth, 1978-1985. 2 bind.

Gilchrist, A. (1994) UDC user clubs. *Extensions and Corrections to the UDC*, 15, pp. 29-30.

Heiner-Freiling, M. (2003) DDC German. Paper presented at the 69th IFLA General Conference and Council 1-9 August 2003, Berlin. Available at: http://www.ifla.org/IV/ifla69/papers/137e_trans-Heiner-Freiling.pdf.

Heiner-Freiling, M. (2006) DDC German - the project, the aims, the methods: new ideas for a well established traditional classification system. In: *Moving beyond the presentation layer: content and context in the Dewey Decimal Classification (DDC) System*. Joan S. Mitchell, Diane Vizine-Goetz [Eds.]. Binghamton, NY: The Haworth Information Press, 2006, pp. 147-162.

Slavic, A. (2004) UDC translations: a 2004 survey report and bibliography. *Extensions and Corrections to the UDC*, 26, pp. 58-80. Also available at <http://dlist.sir.arizona.edu/649/>.

Slavic, A.; Cordeiro, M. I.; Riesthuis, G. J. A. (2007) Enhancement of UDC data for use and sharing in a networked environment. Paper presented at the Librarian Workshop in conjunction with The 31st Annual Conference of the German Classification Society on Data Analysis, Machine Learning, and Applications, Freiburg, Germany. Available at: <http://dlist.sir.arizona.edu/2093/>.

[End]