Proposal for the 6th European NKOS Workshop 2007

Results from a German terminology mapping effort: intra- and interdisciplinary cross-concordances between controlled vocabularies

Philipp Mayr, Vivien Petras, Anne-Kathrin Walter
GESIS Social Science Information Centre (GESIS-IZ), Bonn, Germany
philipp.mayr|vivien.petras|anne-kathrin.walter@gesis.org

In 2004, the German Federal Ministry for Education and Research funded a major terminology mapping initiative[1] at the GESIS Social Science Information Centre in Bonn (GESIS-IZ), which will find its conclusion this year. The task of this terminology mapping initiative was to organize, create and manage 'cross-concordances' between major controlled vocabularies (thesauri, classification systems, subject heading lists) centred around the social sciences but quickly extending to other subject areas. Cross-concordances are intellectually (manually) created crosswalks that determine equivalence, hierarchy, and association relations between terms from two controlled vocabularies. Most vocabularies have been related bilaterally, that is, there is a cross-concordance relating terms from vocabulary A to vocabulary B as well as a cross-concordance relating terms from vocabulary B to vocabulary A (bilateral relations are not necessarily symmetrical). Till August 2007, 24 controlled vocabularies from 11 disciplines will be connected with vocabulary sizes ranging from 2,000 – 17,000 terms per vocabulary. To date more than 260,000 relations are generated.

A database including all vocabularies and cross-concordances was built and a 'heterogeneity service' developed, a web service, which makes the cross-concordances available for other applications. Many cross-concordances are already implemented and utilized for the German Social Science Information Portal Sowiport (www.sowiport.de), which searches bibliographical and other information resources (incl. 13 databases with 10 different vocabularies and ca. 2.5 million references).

In the final phase of the project, a major evaluation effort is under way to test and measure the effectiveness of the vocabulary mappings in an information system environment. Actual user queries are tested in a distributed search environment, where several bibliographic databases with different controlled vocabularies are searched at the same time. Three query variations are compared to each other: a free-text search without focusing on using the controlled vocabulary or terminology mapping; a controlled vocabulary search, where terms from one vocabulary (a 'home' vocabulary thought to be familiar to the user of a particular database) are used to search all databases; and finally, a search, where controlled vocabulary terms are translated into the terms of the respective controlled vocabulary of the database. For evaluation purposes, types of cross-concordances are distinguished between intradisciplinary vocabularies (vocabularies within the social sciences) and interdisciplinary vocabularies (social sciences to other disciplines as well as other combinations).

Simultaneously, an extensive quantitative analysis is conducted aimed at finding patterns in terminology mappings that can explain trends in the effectiveness of terminology mappings, particularly looking at overlapping terms, types of determined relations (equivalence, hierarchy etc.), size of participating vocabularies, etc.

This project is the largest terminology mapping effort in Germany. The number and variety of controlled vocabularies targeted provide an optimal basis for insights and further research opportunities. To our knowledge, terminology mapping efforts have rarely been evaluated with stringent qualitative and quantitative measures. This research should contribute in this area.

For the NKOS workshop, we plan to present an overview of the project and participating vocabularies, an introduction to the heterogeneity service and its application as well as some of the results and findings of the evaluation, which will be concluded in August.

[1] http://www.gesis.org/en/research/information_technology/komohe.htm