

Providing multilingual subject access through linking of subject heading languages: The MACS approach

Patrice Landry¹

¹ Swiss National Library, Hallwylstrasse 15, 3003 Bern, Switzerland

Abstract. The MACS project aims at providing multilingual subject access to library catalogues through the use of concordances between subject headings from LCSH, RAMEAU and SWD. The manual approach, as used by MACS, has been up to now the most reliable method for ensuring accurate multilingual subject access to bibliographic data. The presentation will give an overview on the development of the project and will outline the strategy and methods used by the MACS project. The presentation will also include a demonstration of the search interface developed by The European Library (TEL).

Keywords: MACS Project, multilingual subject access, LCSH, RAMEAU, SWD, mapping, concordances.

1 Introduction

Providing subject access using several languages in library online catalogues has become an important challenge for many national libraries in Europe. Given that the Web enables anyone to search any catalogue online, subject access using a single language becomes, for many national libraries in Europe, an important obstacle to efficient bibliographic retrieval. Aware of this obstacle and wishing to take advantage of the wealth of networks and the availability of rich and sustainable indexing languages, in 1998 four national libraries established the MACS project (Multilingual Access to Subjects)[1] under the auspices of the Conference of European National librarians (CENL).

Ten years later, the project has achieved this goal of providing the tools for multilingual subject access. It is now operational at the Swiss National Library and the Deutsche Nationalbibliothek where SWD (Schlagwortnormdatei) terms are added to links with RAMEAU, the French indexing subject headings list and LCSH (Library of Congress Subject Headings). The product is in the process of being integrated in The European Library portal [2].

2 MACS approach

Controlled vocabularies offer tremendous search possibilities for expanded use in web based services. Each vocabulary's authority file, for example subject heading lists, can contain upward of 300'000 to 500'000 entries. In controlled vocabularies, each term, name or subject, should have the same form each time it occurs in a bibliographic record or metadata. Relationships are established to ensure that accepted and rejected terms are linked and that users are directed to data identified under the preferred term. In recent years, there have been major developments in making these bibliographic metadata, such as terms (headings), their relationships and descriptions available as resources on the web. The issue is how these resources can be used in a meaningful and efficient way to improve information retrieval in the web environment. One of the solutions used in the context of subject headings has been the development of the SKOS (Simple Knowledge Organisation System) data model that provides a standard process for creating or migrating existing controlled vocabularies to the web.

While this solution offers the potential of being used in automatic matching or linking systems (ontology alignment), the MACS approach [3], [4], [5], [6], [7], [8] of linking terms manually from various subject heading languages offers the possibility of creating reliable and stable sets of linked data. Using as its base 90'000 RAMEAU-LCSH links, the project will gradually add SWD terms to these links and create new links with SWD headings. The manual method does not offer the same quantitative level results as automatic linking systems, but linking initiatives such as MACS create standard based sets of linked metadata that are relevant to large library collections.

The MACS linking methodology allow for links to be established based on an analysis at the terminology level (subject heading), at the semantic level (authority record), and at the syntactic level (indexing). A match or link is considered successful when a concept represented by similar headings in the different SHLs, which are matched manually (intellectually), returns the most closely equivalent results (titles) through subject retrieval. The MACS approach can be summarized by the following principles:

- Equality of languages and SHLs: No language or SHL is used as a pivot; each language is managed autonomously outside of MACS
- Establishment of equivalences between SHLs: Headings are not translated
- Equivalence links conceived as concept clusters: Mapping is done on the basis of concepts represented by SHLs
- Consistency of results: The quality of linking is based on the retrieval of consistent and similar sets of bibliographic records from different sources
- Extendable to other SHLs: Concept clusters can be increased by almost unlimited SHLs

Since March 2007, the Swiss National Library is creating links with SWD headings. This task was successfully integrated into the normal work process of the library and is now part of the workload of indexers. The Deutsche Nationalbibliothek has received funding to hire staff to work exclusively on link creation and work started in April 2009 at the Deutsche Nationalbibliothek in Leipzig. Their target is to

add an additional 45'000 links with SWD in the MACS database. By August 2009, close to 40'000 SWD had already been added to MACS links and it is planned that by 2010, most widely used topical headings used for indexing will have been covered. It is expected that 70'000 links with SWD will then be available and those links will cover 80% of bibliographic records of the Deutsche Nationalbibliothek's collections.

The MACS project has been involved in the TELplus project [9] that is investigating full text indexing and semantic search by providing access to its linking database. The work of TELplus in developing an automatic alignment of vocabularies using the semantic data of controlled vocabularies has provided MACS with a possible complementary linking strategy. The TELplus project has demonstrated that it can produce relatively reliable or relevant links in about 50% of the cases, mostly in the alignment of non ambiguous terms. Their method of evaluating these matches was to use MACS data as a comparison. The major challenge will be to test this methodology in searching different catalogues and finding equivalent results through subject retrieval. Up to now, manual subject headings linking projects has produced the most reliable linked data and it will be interesting to pursue alternate linking strategies.

3 Search Interface

With the growing amount of links in three languages, the task of developing a search interface was also on the MACS project's agenda. The strategy adopted by the MACS partners for realising this goal was to work with The European Library (TEL). Since TEL is a CENL service and that MACS partners are members of the CENL, it was considered to develop a search interface that could be integrated into the TEL portal. Discussions with the TEL office started in 2006 in the context of the EDL Project, one of whose objectives was the development of multilingual capacities of the TEL portal. A first prototype was created in October 2007 and work is presently underway to produce a new version of the search interface (LVAT II). The LVAT II project will address the issue of conducting "exact subject" queries across partner's library catalogues by building a central index using the open source indexing / search engine SOLR/LUCENE. Collections from the MACS partners will then be indexed in this central index. The project will also investigate the integration of TELplusⁱ automatic subject heading alignments in the LVAT II.

Should that search interface project be successful, it will be left to CENL to formally integrate the MACS results in the TEL bibliographic services. This connection to TEL could lead to a permanent management solution to ensure the long term viability of the MACS initiative. The project is now confronted with demands to expand MACS to other languages (national and local) and would need to deal with the financial and management issues related to these demands. The Swiss National Library who has managed the project since its beginning has worked with the MACS partners to ensure funding for current development work. But with the potential of an increasing the number of participants, a new management structure would be needed correctly address these issues.

References

1. <http://macs.cenl.org>
 2. www.theeuropeanlibrary.org
 3. Clavel-Merrin, G.: The Need for Co-Operation in Creating and Maintaining Multilingual Subject Authority Files”, 65th IFLA Council and General Conference, Meeting 155. <<http://WWW.ifla.org/IV/ifla65/papers/080-155e.htm>> (1999)
 4. Freyre, E., Naudi, M.: MACS: Subject Access Across Languages and Networks. In: Subject Retrieval in a Networked Environment: Papers Presented at an IFLA Satellite Meeting Sponsored by the IFLA Section on Classification and Indexing & IFLA Section on Information Technology, OCLC, *Dublin, Ohio, USA, 14-16 August 2001*. Dublin, OH: OCLC (2001)
 5. MacEwan, A.: Crossing language Barriers in Europe: Linking LCSH to Other Subject Heading Languages. *Cataloging & Classification Quarterly* 29 (1/2), 199-207 (2000)
 6. Landry, P.: The MACS Project: Multilingual Access to Subject (LCSH, RAMEAU, SWD). *International Cataloguing and Bibliographic Control* 30 (3), 46-49 (2001)
 7. Landry, P.: Multilingual Subject Access: The Linking Approach of MACS. *Cataloging & Classification Quarterly* 31 (3-4), 177-191 (2004)
 8. Landry, P.: The Evolution of Subject Heading Languages in Europe and their Impact on Subject Access Interoperability. In *New Perspectives on Subject Indexing and Classification: Essays in Honour of Magda Heiner-Freiling*, pp. 249-256. Deutsche Nationalbibliothek, Frankfurt am Main (2008)
 9. www.theeuropeanlibrary.org/telplus
 10. Wang, S., Issac, A., Schopman, B., Schlobach, S., Van der Meij, L.: Matching multilingual subject vocabularies (submitted for publication)
-