

Library Track

Antoine Isaac Shenghui Wang

November 15, 2009

1 Library

This task, organized in the context of the TELplus¹ project, focuses on a case for which the MACS² project established a (partial) manual gold standard. Participants of this task had to create pairwise alignments between three large subject heading lists in different languages. The required alignments links were SKOS relations. This task is similar, from a methodological perspective, to the OAEI 2008 Library track. It uses however a different dataset.

1.1 Data set

The vocabularies to align are:

- LCSH, the Library of Congress Subject Headings, available as linked data at <http://id.loc.gov>. Contains around 340K concepts, including 250K general subjects.
- RAMEAU, the heading list used at the French National Library, available as linked data at <http://stitch.cs.vu.nl/rameau>. Contains around 150K concepts, including 90K general subjects.
- SWD, the heading list used at the German National Library. Contains 800K concepts, including 160K general subjects.

The concepts from the three vocabularies are used as subjects of books. For each concept, the usual SKOS lexical and semantic information is provided: preferred labels, synonyms and notes, broader and related concepts, etc. The three subject heading lists have been represented according to the SKOS model, but an OWL version has also been made available. Note that even though two of these vocabularies are available online as RDF data, we have provided dumps for the convenience of participants.

We have also made available a part of the MACS manual mappings between these vocabularies, which can be used as a *learning set*. However, none of the participants asked for it.

¹<http://www.theeuropeanlibrary.org/telplus>

²<http://macs.cenl.org>

1.2 Evaluation and results

Only one team handed in final results: TaxoMap, which produced results as listed in Tab. 1.

Type of relation	LCSH-RAMEAU	RAMEAU-SWD	LCSH-SWD
exactMatch	5,074	1,265	38
broadMatch	116,789	17,220	0
narrowMatch	48,817	6,690	0
relatedMatch	13,205	1,317	0

Table 1: Taxomap results

We follow the dual evaluation approach of the previous 2008 Library Track, which featured a “thesaurus merging” evaluation (based on a post-hoc partial gold standard) and a “re-indexing” one (assessing the use of mappings for translating subject annotations from one thesaurus to another). The main difference is that the first evaluation method has now been replaced by comparing to an already existing partial reference alignment (the MACS one), avoiding to manually assess the participant’s results.

1.2.1 Comparing with partial reference alignment (MACS)

As no participant used the training set we provided, we use the complete MACS mappings as gold standard. In the version we received,³ this reference alignment comprised 87,183 LCSH-RAMEAU mappings, 13,723 RAMEAU-SWD mappings, and 12,203 LCSH-SWD mappings.

Table 2 shows the results when taking into account all mappings that belong to a certain relation selection. For a given relation selection, the token “–” means that no extra relation was provided at that level, hence the results are identical to the ones of the previous selection level. Prec. refer to the usual precision; Cov. refers to the coverage, that is, the percentage of MACS mappings which were found in the evaluated alignment.

TaxoMap links evaluated	LCSH-RAMEAU		RAMEAU-SWD		LCSH-SWD	
	Prec.	Cov.	Prec.	Cov.	Prec.	Cov.
exactMatch	72.1	5.7	27.1	1.4	44.4	0.03
eM + broadMatch	3.6	6.9	2.3	1.9	–	–
eM + bM + narrowMatch	2.8	7.3	1.8	2.0	–	–
all relations	2.7	7.5	1.9	2.2	–	–

Table 2: Results for comparison with MACS (percentage) – using all mappings.

Table 3 shows the results obtained when selecting only the “best” available mapping for one concept (that is, the one with the highest confidence measure), and discarding the others.

³MACS is still currently adding manual mappings to this reference set.

TaxoMap links evaluated	LCSH-RAMEAU		RAMEAU-SWD		LCSH-SWD	
	Prec.	Cov.	Prec.	Cov.	Prec.	Cov.
exactMatch	78.7	5.7	39.5	1.4	44.4	0.03
eM + broadMatch	22.0	6.0	13.5	1.6	–	–
eM + bM + narrowMatch	14.4	5.9	10.8	1.6	–	–
all relations	13.4	5.8	10.9	1.7	–	–

Table 3: Results for comparison with MACS (percentage) – using only the best mapping for each concept.

Collection	Books with subject annotation
English	2,448,050
French	1,457,143
German	1,364,287

Table 4: Collections and books with subject annotations

1.2.2 Results for the re-indexing scenario

The second usage scenario is based on an *annotation translation* process supporting the re-indexing of books indexed with one vocabulary, using concepts from the mapped vocabulary (see [?]). Here we use book annotations from the British Library (using LCSH), the French National Library (using RAMEAU) and the German National Library (using SWD), see Tab. 4.

For each pair of vocabularies A-B, this scenario interprets the mappings as rules to translate existing book annotations with A into equivalent annotations with B. In the case at hand, the book collections have a few books in common (cf. Tab. 5), which are therefore described according to two vocabularies. Based on the quality of the results for those books for which we know the correct annotations, we can assess the quality of the initial correspondences.

Evaluation settings and measures. For each pair of vocabularies A-B, the simple concept-to-concept correspondences sent by participants were transformed into more complex mapping rules that associate one concept from A with a set of concepts from B – as some concepts are involved in several mappings.

The set of A concepts attached to each book is then used to decide whether these rules are *fired* for this book. If the A concept of one rule is contained by the A annotation of a book, then the rule is fired. As several rules can be fired for a same book, the union of the consequents of these rules forms the translated B annotation of the book.

On a set of books selected for evaluation, the generated concepts for a book are then compared to the ones that are deemed correct for this book. At the annotation level, we measure the precision (Prec.), how many translated concepts are correct over the annotation produced for the books on which rules were fired, the recall (Rec.), how many correct Brinkman annotation concepts are found for all books in the evaluation set, and the Jaccard overlap measure (Jac.) between the produced annotation and the correct one.

Collection pair	Common books
French–English	182,460
German–English	83,786
German–French	63,340

Table 5: Common books between different collections

In the formulas used, results are counted on a book and annotation basis, and not on a rule basis. This reflects the importance of different thesaurus concepts: a translation rule for a frequently used concept is more important than a rule for a rarely used concept.

Results. Tab. 6 shows the results when taking into account all mappings that belong to a certain relation selection.

TaxoMap links evaluated	LCSH-RAMEAU			RAMEAU-SWD			LCSH-SWD		
	Prec.	Rec.	Jac.	Prec.	Rec.	Jac.	Prec.	Rec.	Jac.
exactMatch	22.3	6.1	5.5	14.2	3.1	2.4	1.3	0.003	0.002
eM + broadMatch	2.1	7.8	1.5	2.3	3.6	1.1	–	–	–
eM + bM + narrowMatch	1.2	9.2	1.0	0.8	3.9	0.5	–	–	–
all relations	1.1	9.3	0.9	0.7	4.0	0.5	–	–	–

Table 6: Re-indexing evaluation results (percentage) – using all mappings.

Tab. 7 shows the results obtained when selecting only the “best” available mapping for one concept, and discarding the others.

TaxoMap links evaluated	LCSH-RAMEAU			RAMEAU-SWD			LCSH-SWD		
	Prec.	Rec.	Jac.	Prec.	Rec.	Jac.	Prec.	Rec.	Jac.
exactMatch	22.8	5.8	5.3	14.2	1.9	1.7	1.2	0.002	0.002
eM + broadMatch	10.2	6.0	4.9	6.9	2.0	1.7	–	–	–
eM + bM + narrowMatch	7.2	4.5	3.3	5.9	1.9	1.5	–	–	–
all relations	6.4	4.0	2.9	5.8	1.9	1.5	–	–	–

Table 7: Re-indexing evaluation results (percentage) – using all Taxomap mappings.

1.3 Discussion

The setting for this year’s library task clearly shows the limits of current matching tools. The case at hand, mostly because of its size and its multilingual aspect, is extremely difficult to handle. The performance of TaxoMap, from this perspective, should be regarded as a significant achievement, as they were the only ones to manage to ingest hundreds of concepts and return alignments between them.

The results of TaxoMap, which could not apply its usual partition approach, and uses to a great extent automatic translation, are not very good. More precisely, they

are especially weak when relations other than strict equivalence are considered, highlighting the value of being able to sort mapping results using the type of relation or the strength of the confidence measure granted to mappings—options which are both offered by TaxoMap. Both precision and coverage/recall are low for the non-equivalence mappings, even though they bring a huge number of potential matches. The translation could give better results for the equivalent mappings, at the cost of coverage of course.

It is worth mentioning that as last year, the results for the comparison with a reference mapping and the re-indexing evaluation largely differ, showing that mappings have a different relevance depending on the application scenario. Mappings based on translation will perform obviously better for scenarios where the intension of concepts matters, rather than for cases where their actual usage in book collections should be carefully taken into account.

2 Acknowledgements

We thank Patrice Landry, Genevieve Clavel and Jeroen Hoppenbrouwers for the MACS data. For LCSH, RAMEAU and SWD, respectively, The Library of Congress, The French National Library and the German National Library. The collection of the British Library was provided by the The European Library Office.