

Matching Multi-lingual Subject Vocabularies

Shenghui Wang^{1,2}, Antoine Isaac^{1,2}, Balthasar Schopman¹, Stefan Schlobach¹,
and Lourens van der Meij^{1,2}

¹ Vrije Universiteit Amsterdam

² Koninklijke Bibliotheek, den Haag

{swang,aisaac,baschopm,schlobac,lourens}@few.vu.nl

Abstract. Most libraries and other cultural heritage institutions use controlled knowledge organisation systems, such as thesauri, to describe their collections. Unfortunately, as most of these institutions use different such systems, unified access to heterogeneous collections is difficult. Things are even worse in an international context when concepts have labels in different languages. In order to overcome the multilingual interoperability problem between European Libraries, extensive work has been done to manually map concepts from different knowledge organisation systems, which is a tedious and expensive process.

Within the TELplus project, we developed and evaluated methods to automatically discover these mappings, using different ontology matching techniques. In experiments on major French, English and German subject heading lists Rameau, LCSH and SWD, we show that we can automatically produce mappings of surprisingly good quality, even when using relatively naive translation and matching methods.

1 Introduction

Controlled knowledge organisation systems, such as thesauri or subject heading lists (SHLs), are often used to describe objects from library collections. These vocabularies, specified at the semantic level using dedicated relations—typically *broader*, *narrower* and *related*—can be of help when accessing collections, *e.g.*, for guiding a user through a hierarchy of subjects, or performing automatic query reformulation to bring more results for a given query.

However, nearly every library uses its own subject indexing system, in its own natural language. It is therefore impossible to exploit the semantically rich information of controlled vocabularies over several collections simultaneously. This greatly hinders access to, and usability of the content of The European Library [1], which is one of important problems to address in the TELplus project [2]. A solution to this issue is the semantic linking (or *matching*) of the concepts present in the vocabularies. This solution has been already investigated in the Cultural Heritage (CH) domain, as in the MACS [3], Renardus [4] and CrissCross [5] projects. MACS, in particular, is building an extensive set of manual links between three SHLs used respectively at the English (and American), French and German national libraries, namely LCSH, Rameau and SWD.

These links represent most often equivalence at the semantic level between concepts and can, *e.g.*, be used to *reformulate* queries from one language to the other. For example, an equivalence link between **Sprinting**, **Course de vitesse** and **Kurzstreckenlauf** will allow to transform a query for sprints, that would only give results in the British Library catalogue, into equivalent queries that will have matching results in the French and German catalogues, respectively.

A crucial problem is the cost of building such manual *alignments* of vocabularies. While some reports mention that around 90 terms may be matched per day by a skilled information professional dealing with concepts in a same language [6], the vocabularies to match often contain hundreds of thousands of concepts. In this paper we will show that automatic matching methods can be viable instruments for supporting these efforts in an effective way.

Methodology. We have implemented four straightforward methods for ontology alignment, two based on lexical properties of the concept labels, and two based on the extensions of the concepts, *i.e.* the objects annotated by them. The simplest approach lexically compares the labels of concepts without translating them; a more sophisticated version uses a simple translation service we could deploy out-of-the-box. A simple extensional method makes use of the fact that all three collections have joint instances (shared books), which can be determined by common ISBN numbers. Finally, we extend our previous work of matching based on instance similarity [7] to the multilingual case.

Research questions. In line with TELplus objectives, which include establishing practical solutions as well as guidelines for the participating libraries, the research questions we want to address in this paper are (i) whether an automatic vocabulary matching approach is feasible in a multilingual context, and (ii) what kind of matching approach performs best.

Experiments. To answer the questions above we pairwise applied our four matching methods to the three SHLs Rameau, LCSH and SWD, for which the MACS project has over the years gathered significant amounts of methodological expertise and reference alignments which we can compare newly produced ones with. This comparison is possible because our case comes with large amounts of collection-level data from the libraries using these vocabularies.

Results. The general results show a relatively good precision of all four methods with respect to reproducing manual mappings from the MACS project. Even stronger, the lexical methods produce a high coverage of those mappings, which indicates that the use of such very simple algorithms can already support the creation of such links significantly. Interestingly enough, the extensional mappings produce results that are non-identical to the MACS mappings, which indicates that intensional and extensional semantics (*i.e.*, the meaning of a concept attached by people and its actual use in the libraries) differ significantly.

Structure of the paper. In Section 2 we describe the context of our research, and the problem we started out to solve. Section 3 describes the matching methods in more detail; our experiments are summarised in Section 4, before we conclude.

2 Problem and Context

The TELplus project aims at adding content to The European Library, but also at improving access and usability, notably by investigating full-text indexing and semantic search.

One crucial issue is that collections—and their metadata—come in different languages, which hampers the access to several of them at a same time. A first solution to this issue relies on using *cross-language information retrieval* methods over the different collections at hand, as currently investigated, *e.g.*, in the Cacao project [8]. This approach is promising for cases where full-text content is available, and standard retrieval of documents is sought. However it may require some complement for the cases where only structured metadata records are available, or when one aims at semantically enriched or personalised access to CH objects, as in the Europeana Thought Lab demo [9].

As collection objects are described by controlled knowledge organisation systems, another promising solution is to identify the semantic links between such systems. Currently the TELplus and MACS projects are collaborating so that manually built MACS mappings between the LCSH, Rameau and SWD SHLs can be used to provide automatic query reformulation to bring more results from the three corresponding national library collections.

Meanwhile, automated alignment methods have been investigated in other domains, such as anatomy [10] and food [11]. Particularly, the mapping between AGROVOC and the Chinese Agricultural Thesaurus [12] exemplifies the difficulties of multilingual matching. In the Semantic Web community, efforts related to Ontology Matching [13], like the OAEI campaigns [14], have already dealt with the library and food domains, as well as with multilingual settings. In such context, our work is to investigate the feasibility and potential benefit of applying automated matching techniques for semantic search in TELplus.

3 Ontology Matching Methods Applied

3.1 SKOS Lexical Mapper

Many lexical mappers—that is, mappers exploiting lexical information such as labels of concepts—are only dedicated to English. To palliate this, we have adapted a lexical mapper first developed for Dutch [15] to French, English and German languages. It is mostly based on the CELEX [16] database, which allows to recognise lexicographic variants and morphological components of word forms. This mapper produces *equivalence* matches between concepts, but also *hierarchical* (broader) matches, based on the morphological (*resp.* syntactic) decomposition of their labels.

The different lexical comparison methods used by this mapper give rise to different confidence values: using exact string equivalence is more reliable than using lemma equivalence. Also, the mapper considers the status of the lexical features it compares. It exploits the SKOS model for representing vocabularies [17], where concepts can have *preferred* or *alternative* labels. The latter ones can be approximate synonyms. For two concepts, any comparison based on them is therefore considered less reliable than a comparison based on preferred labels. The combination of these two factors—different comparison techniques and different features compared—results in a grading of the produced mappings, which can be used as a confidence value.

Our mapper only considers one language at a time. To apply it in a multilingual case, we translated the vocabularies beforehand. For each vocabulary pair (*e.g.*, Rameau and LCSH), we translate each vocabulary by adding new labels (preferred or alternative) that result from translating the original labels, using the Google Translate service [18]. We then run the mapper twice, once for each language of the pair. In the Rameau-LCSH case, the translation of Rameau to English is matched (in English) to the original LCSH version, and the translation of LCSH in French is matched (in French) to the original Rameau version. The obtained results are then merged, and we keep only the *equivalence* links.

3.2 Instance-Based Mapping

Instance-based matching techniques determine the similarity between concepts by examining the extensional information of concepts, that is, the instance data they classify. The idea behind such techniques, already used in a number of works like [19], is that the similarity between the extensions of two concepts reflects the semantic similarity of these concepts. This is a very natural approach, as in most ontology formalisms the semantics of the relations between concepts is defined via their instances. This also fits the notion of literary warrant that is relevant for the design of controlled vocabularies in libraries or other institutes.¹

Using overlap of common instances. A first and straightforward method is to measure the *common extension* of the concepts—the set of objects that are simultaneously classified by both concepts [19,21]. This method has a number of important benefits. Contrary to lexical techniques, it does not depend on the concept labels, which is particularly important when the ontologies or thesauri come in different languages. Moreover, it does not depend on a rich semantic structure; this is important in the case of SHLs, which are often incompletely structured.

The basic idea is simple: the higher the ratio of co-occurring instances for two concepts, the more related they are. In our application context, the instances of a

¹ As Svenonius reportedly wrote in [20] “As a name of a subject, the term Butterflies refers not to actual butterflies but rather to the set of all indexed documents about butterflies. [...] In a subject language the extension of a term is the class of all documents about what the term denotes, such as all documents about butterflies.”

concept c , noted as $e(c)$, are the set of books related to this concept via a **subject** annotation property. For each pair of concepts, the overlap of their instance sets is measured and considered as the confidence value for an equivalence relation. Our measure, shown below, is an adaption of the standard Jaccard similarity, to avoid very high scores in the case of very few instances: the 0.8 parameter was chosen so that concepts with a single (shared) instance obtain the same score as concepts with, in the limit, infinitely many instances, 20% of which co-occur. This choice is relatively arbitrary, but this measure has shown to perform well on previous experiments in the library domain for the STITCH project [21].

$$overlap_i(c_1, c_2) = \frac{\sqrt{|e(c_1) \cap e(c_2)| \times (|e(c_1) \cap e(c_2)| - 0.8)}}{|e(c_1) \cup e(c_2)|}$$

Note that one concept can be related to multiple concepts with different confidence values. In our experiments, we consider the concept with the highest confidence value as the candidate mapping for evaluation.

Using instance matching. Measuring the common extension of concepts requires the existence of sufficient amounts of shared instances, which is very often *not* the case. However, as instances—in our cases, books—have their own information, such as authors, titles, etc., it is possible to calculate the similarity between them. Our assumption is that similar books are likely to be annotated with similar subject headings, no matter they are from different collections or described in different languages.

The instance matching based method first compares books from both collections. For each book from Collection A, i_a , there is a most similar book from Collection B, i_b . We then consider that i_a shares the same subject headings as i_b does. In other words, i_a is now an instance of all subject headings which i_b is annotated with. This matching procedure is carried out on both directions. In this way, we can again apply measures on common extensions of the subject headings, even if the extensions have been enriched artificially.

There are different ways to match instances. The simplest way is to consider instances as documents with all their metadata as their feature, and apply information retrieval techniques to retrieve similar instances (documents). We use the tf-idf weighting scheme which is often exploited in the vector space model for information retrieval and text mining [22]. Obviously, the quality of the instance matching has an important impact on the quality of concept mappings. One may argue that the whole process is questionable: it is in fact one aim of this paper to investigate it.

To apply such a method in a multilingual context, automated translation is crucial. We take a naive approach, using the Google Translate service to translate book metadata, including subject labels. The translation was done offline on a word-to-word level. We created a list of all unique words in each book collection. Batches of words were sent via an API to the Google Translate service. Every word for which we obtained a translation was stored in a translation table. During the instance matching process, we translate every word of that instance by looking it up in the translation table and replacing it with the translation if available. We then calculate book similarity scores within a same language.

4 Experiments and Evaluation

4.1 Data Preprocessing and Experiment Settings

The three SHLs at hand, namely, LCSH in English, Rameau in French and SWD in German, have been converted into the standard SKOS format [17]—see [23] for an overview of how this was done for LCSH. Collections annotated by these three SHLs, respectively, are gathered from the British Library (BL), the French National Library (BnF) and the German National Library (DNB).

In order to apply instance-based matching methods, the link between each book record and its subjects, *i.e.*, concepts from the SHL SKOS conversions, should be properly identified. Instead of using unique identifiers of subjects, librarians often use simple *string* annotations. This introduces many issues, for example, using the lower case version of a concept label, or using alternative labels instead of the preferred ones. This has to be addressed by using a simple string look-up and matching algorithm to identify the book-concept links, using the concept labels found in the SKOS representations.

Furthermore, it is also necessary to tackle the pre-coordination issue. Librarians often combine two thesaurus concepts into a single complex subject to annotate books, *e.g.*, `France--History--13th century`. Some of these combinations are so often used that they are included into the subject vocabulary later, while some are generated only at annotation time. In our data preprocessing step, we applied the following strategy: if the subject combination cannot be recognised as an existing concept, it is then separated into single (existing) concepts, and the book is considered to be annotated by each of these concepts.

We are well aware that this choice is certainly not neutral: hereby, a concept’s extension, beyond the instances simply annotated by it, also contains the instances indexed with a compound subject that includes it, if this combination is not an existing concept in the vocabulary. However, it also brings more instances for concepts, which is very important for the statistical validity of the instance-based methods we employ here. Indeed this is made even more important by the low number of annotations we identified from the collections. Not every concept is used in the collections we have, *cf.* Tables 1 and 2. This issue, which is mostly caused by the SHLs being designed and used for several collections, will cause a mapping coverage problem for the instance-based methods, which we will discuss later.

Another related, important decision we made is to restrict ourselves to match only individual concepts, or combinations that are reified in the vocabulary files. This drastically reduces the problem space, while keeping it focused on

Table 1. Size of SHLs and number of concepts used to annotate books in collections

| | Total concepts | Concepts used in collection |
|--------|----------------|-----------------------------|
| LCSH | 339,612 | 138,785 |
| Rameau | 154,974 | 87,722 |
| SWD | 805,017 | 209,666 |

Table 2. Three collections and identified records with valid subject annotations, *i.e.*, there is at least one link between these books and one SHL concept

| | Total records | Rec. with valid subject annot. | Individual book-concept links |
|---------|---------------|--------------------------------|-------------------------------|
| English | 8,430,994 | 2,448,050 | 6,250,859 |
| French | 8,333,000 | 1,457,143 | 4,073,040 |
| German | 2,385,912 | 1,364,287 | 4,258,106 |

Table 3. Common books between different collections

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

the arguably more important part of the potential book subjects. In fact this is rather in line with what is done in MACS, where very few mappings (up to 3.8% for SWD mappings) involve coordinations of concepts that are not already in the vocabularies.

A last step of preprocessing we need is identifying the common books in two collections. The ISBN number is a unique identifier of one book. By comparing the ISBNs in both collections, we found three dually annotated datasets between the three pairs of SHLs, as shown in Table 3. The amount of common books is extremely small compared to the size of collections. This is not unexpected, but certainly causes a serious problem of concept coverage for the simple instance-based method that completely relies on these common instances.

4.2 Comparing with MACS Manual Mappings

Method. The MACS manual mappings were used as reference mappings. This gold standard is however not complete, as MACS is still work in progress. Table 4 gives the concept coverage of mappings between each pair of vocabularies.

Obviously, there is a serious lack in terms of concept coverage if using MACS as a gold standard. For example, only 12.7% of LCSH concepts and 27.0% of Rameau concepts are both involved in MACS mappings and used to annotate books in the collections we gathered. The situation is much worse for the other two pairs of thesauri, where only 1 to 3% concepts are both considered by MACS and used to annotate books.

To perform a relatively fair evaluation on our matchers’ *accuracy*, that is, taking into account the sheer coverage of the MACS gold standard, we separated the generated mappings as “judgeable” and “non-judgeable.” A mapping is judgeable if at least one concept of the pair is involved in a MACS manual mapping, otherwise, it is not judgeable—that is, no data in MACS allows us to say whether the mapping is correct or not. We measure *precision* as the proportion of the correct mappings over all generated and judgeable mappings.

Table 4. Simple statistics of MACS manual mappings and concepts involved

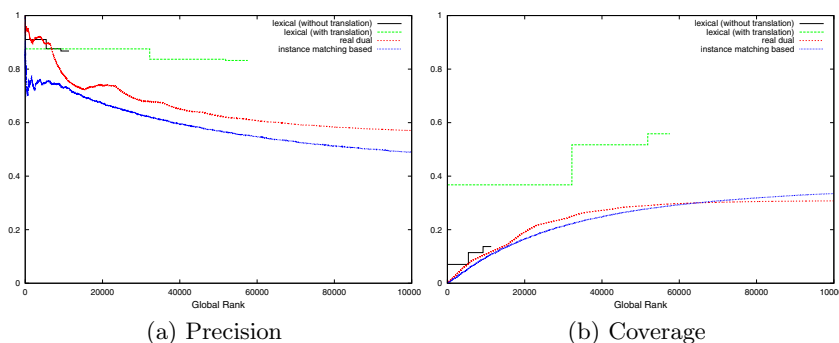
| | Total MACS mappings | Concepts involved |
|---------------|---------------------|-----------------------------------|
| LCSH – Rameau | 57,663 | 16.4% of LCSH and 36.1% of Rameau |
| LCSH – SWD | 12,031 | 3.2% of LCSH and 1.4% of SWD |
| Rameau – SWD | 13,420 | 7.8% of Rameau and 1.6% of SWD |

To measure the *completeness* of the found alignments, we would need to compute *recall*, that is, the proportion of the correct mappings over all possible correct mappings. Unfortunately it is very difficult to get all correct mappings in practice. Manual alignment efforts are time consuming and result in a limited amount of mappings if the two vocabularies to align are big. Despite the lack in concept coverage for MACS, we decided that these manual mappings were still useful to exploit. Indeed, measuring how well we can reproduce manual mappings with the help of automatic tools is valuable *per se*. As a proxy for completeness, we thus measure the *coverage* of MACS, that is, the proportion of MACS mappings we find in the automatically produced alignments.

As already hinted, our matchers return candidate mappings with a confidence value—based on linguistic considerations or the extensional overlap of two concepts. This allows us to rank the mappings, and, moving from the top of the ranked list, to measure the above two measurements up to certain ranks.

Results. Fig. 1 gives the performance of four different mapping methods on the task of matching LCSH and Rameau. Here the x-axis is the global rank of those mappings—by “global ranking,” we take the non-judgeable mappings into account; however, they are not considered when computing precision. Note that our lexical mapper provides three confidence levels. Mappings with the same value are given the same rank; they are therefore measured together.

The lexical method applied on non-translated LCSH and Rameau gives a very limited amount of mappings: in total, 86% of these mappings are in MACS, but they only represent 13% of the MACS mappings. By naively using Google

**Fig. 1.** Performance of different methods on matching LCSH and Rameau

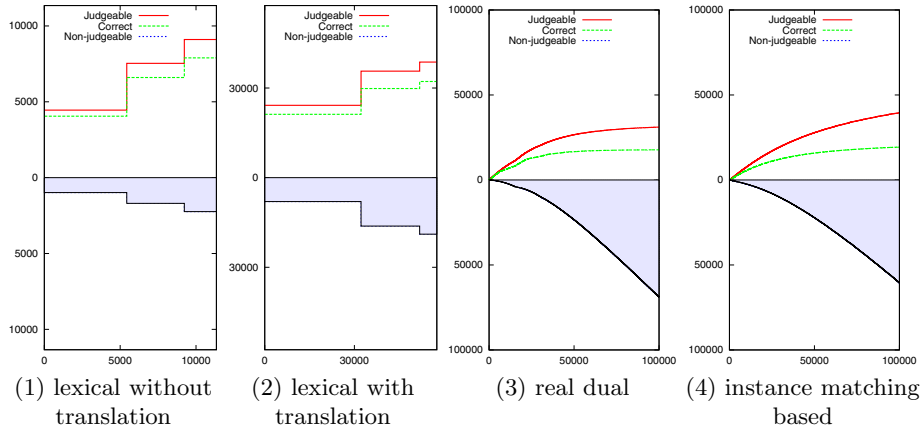


Fig. 2. The distribution of mappings generated by different methods — LCSH *vs.* Rameau

Translate, the automated lexical method already recovers 56% of MACS mappings, while the precision decreases by 3%. The main reason for this precision decrease is that the translation is not perfect nor stable. For example, in SWD the German name **Aachen** occurs in several subject headings. However, it was sometimes (rightly) translated to the French name **Aix-la-Chapelle** and in other cases it was not translated at all.

From Fig.1, the precision and coverage of the first 7K instance-based mappings generated from the real dually annotated dataset (1% of total book records in two collections) are similar to the lexical method on non-translated thesauri. Gradually, the precision decreases and the coverage increases, and both level after approximately 60K mappings.

The sheer amount of instances inevitably influences the performance of this method. Another possible reason is that instance-based methods focus on the extensional semantics of those concepts, *i.e.*, how they are used in reality. Some mappings are not really intensionally equivalent, but they are used to annotate the same books in two collections. For example, according to MACS, the Rameau concept **Cavitation** is mapped to the LCSH concept **Cavitation**; however, our instance-based method maps it to another LCSH concept **Hydraulic machinery**, because they both annotate the same books. Such mappings could therefore be very useful in the query reformation or search applications, and of course would require further evaluation. This also indicates that the intensional and the extensional semantics (*i.e.*, the meaning of a concept attached by people and its actual use in the libraries) may differ significantly.

The method based on instance matching performed worse here. The loss of nearly 10% in precision could have two reasons: 1) the translation of book information is not good enough; 2) the similarity between books is calculated purely based on weighted common words, where we ignore the semantic distinction between different metadata field, which could potentially help to identify similar

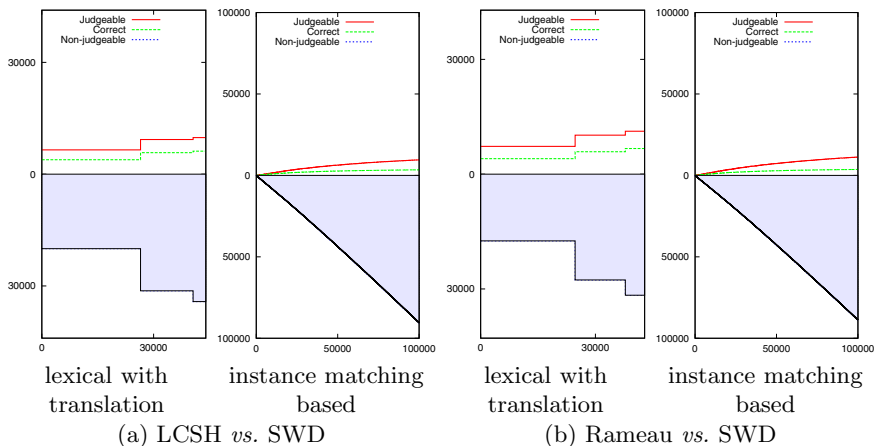


Fig. 3. Coverage issue for the LCSH-SWD and Rameau-SWD cases

books. Meanwhile, by matching similar instances on top of using real dually annotated books, we gradually include new concepts, which increases coverage.

As introduced earlier, not every mapping can be evaluated, as neither of their concepts are considered in MACS before. For example, up to rank 50K, 29% of MACS mappings (16,644) between LCSH and Rameau are found, and the precision is 63%. However, only less than 26K mappings were actually judgeable. Fig. 2 compares the distribution of different kinds of mappings of each method, where the shaded area shows the amount of non-judgeable mappings. The coverage issue for the SWD-related cases is more serious, even for lexical mappings, as shown in Fig. 3. Among those non-judgeable mappings, we expect to find valid ones given the precision of the judgeable mappings around them.

We carried out manual evaluation of non-judgeable mapping samples. For instance-based mappings, we first ranked them based on their confidence values, and then chose every 10th mapping among the first 1000 mappings, every 100th mapping from 1000 to 10,000 mappings, and every 1000th mappings from 10,000 to 100,000 mappings. For lexical mappings, we took 50 random mappings within each of the three confidence levels. In all cases, we kept for manual evaluation only the mappings that are not judgeable according to MACS. Depending on the actual sample size, the corresponding error bar was also calculated.

Fig. 4 shows the precision of the manual evaluation proportionally combined with that from comparing with MACS reference mappings. For the LCSH–Rameau case, the precision, which is consistent with Fig 1 (a), indicates that our methods indeed provide a significant amount of valid, and more importantly, complementary mappings to MACS manual mappings. For the LCSH–SWD and Rameau–SWD cases, the global precision is also comparable with the one obtained using MACS alone. It also confirms that all methods perform worse in these two cases, which we can relate to the fact that LCSH and Rameau headings are quite similar in the way they are designed and used, and less similar to

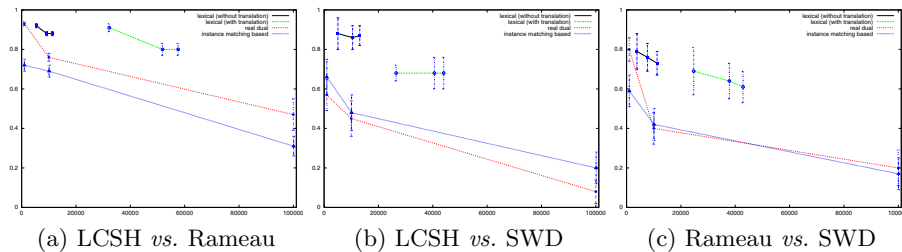


Fig. 4. Overall precision combining MACS and manual evaluation

SWD. Finally, the performances of the two instance-based methods cannot be distinguished anymore. Yet, due to the small sample size, it is impossible to say whether this is due to statistical uncertainty, or to that fact that the method using instance matching would suffer less from a very small overlap of collections. We will investigate more this aspect in the future.

5 Conclusion

We have explored methods to automatically align multilingual SHLs in a realistic setting, using different techniques known from ontology matching, exploiting labels of concepts or book-level information. We also reported various problems with applying existing techniques and the possible (and simple) solutions to them. In particular, we use the Google Translate service to address the translation problem. We are well aware that other, more sophisticated approaches exist, but deploying them requires thorough expertise in the use of linguistic resources, which makes it more difficult to assess the *practical* feasibility of such an approach.

These experiments on the major French, English and German SHLs, Rameau, LCSH and SWD, show that we can automatically produce mappings of surprisingly good quality, even when using quite naive translation and matching methods. The lexical methods produce a relatively high coverage of the MACS manual mappings, which indicates that the use of such very simple algorithms can already support the creation of manual mappings significantly. The instance-based mapping methods provide mappings that are nearly complementary to manual ones. This is more interesting, as it indicates that, while each can be useful in different applications, the intensional and extensional semantic links are significantly different. More efforts would now be required to turn these findings into an appropriate methodology to assist manual alignment, or to evaluate to which extent imperfect mappings can still benefit to multilingual collection access for end users in the TELplus case. Our results also identify different directions to improve the performance of methods, and we will continue reporting our efforts in this area.

Acknowledgements. This work is funded by the NWO CATCH and EU eContentPlus programmes (STITCH and TELplus projects). Patrice Landry, Jeroen

Hoppenbrouwers and Geneviève Clavel provided us with MACS data. Various people at Library of Congress, DNB, BnF and TEL Office have provided or helped us with SHL and collection data, incl. Barbara Tillett, Anke Meyer, Claudia Werner, Françoise Bourdon, Michel Minguam and Sjoerd Siebinga.

References

1. <http://www.theeuropeanlibrary.org/>
2. <http://www.theeuropeanlibrary.org/telplus/>
3. Landry, P.: Multilingualism and subject heading languages: how the MACS project is providing multilingual subject access in Europe. *Catalogue & Index: periodical of CILIP Cataloguing and Indexing Group* 157 (to appear, 2009)
4. Day, M., Koch, T., Neuroth, H.: Searching and browsing multiple subject gateways in the Renardus service. In: *Proceedings of the 6th International Conference on Social Science Methodology*, Amsterdam, The Netherlands (2005)
5. Boterham, F., Hubrich, J.: Towards a comprehensive international Knowledge Organisation System. In: *7th Networked Knowledge Organization Systems (NKOS) Workshop at the 12th ECDL Conference*, Aarhus, Denmark (2008)
6. Will, L.: Costs of vocabulary mapping, <http://hilt.cdlr.strath.ac.uk/Dissemination/Presentations/Leonard%20Will.ppt>
7. Schopman, B., Wang, S., Schlobach, S.: Deriving concept mappings through instance mappings. In: *Proceedings of the 3rd Asian Semantic Web Conference*, Bangkok, Thailand (2008)
8. <http://www.cacaoproject.eu/>
9. <http://www.europeana.eu/portal/thought-lab.html>
10. Zhang, S., Bodenreider, O.: Experience in aligning anatomical ontologies. *International journal on Semantic Web and information systems* 3(2), 1–26 (2007)
11. Lauser, B., Johannsen, G., Caracciolo, C., Keizer, J., van Hage, W.R., Mayr, P.: Comparing human and automatic thesaurus mapping approaches in the agricultural domain. In: *Proceedings of the International Conference on Dublin Core and Metadata Applications*, Berlin, Germany (2008)
12. Liang, A.C., Sini, M.: Mapping AGROVOC and the Chinese Agricultural Thesaurus: Definitions, tools, procedures. *The New Review of Hypermedia and Multimedia* 12(1), 51–62 (2006)
13. Euzenat, J., Shvaiko, P.: *Ontology Matching*. Springer, Heidelberg (2007)
14. <http://oaei.ontologymatching.org/>
15. Malaisé, V., Isaac, A., Gazendam, L., Brugman, H.: Anchoring Dutch Cultural Heritage Thesauri to WordNet: two case studies. In: *ACL 2007 Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2007)*, Prague, Czech Republic (2007)
16. <http://www.ru.nl/celex/>
17. Isaac, A., Summers, E.: *SKOS Primer*. W3C Group Note (2009)
18. <http://translate.google.com/>
19. Vizine-Goetz, D.: Popular LCSH with Dewey Numbers: Subject headings for everyone. *Annual Review of OCLC Research* (1997)

20. Svenonius, E.: *The Intellectual Foundation of Information Organization*. MIT Press, Cambridge (2000)
21. Isaac, A., van der Meij, L., Schlobach, S., Wang, S.: An empirical study of instance-based ontology matching. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) *ASWC 2007 and ISWC 2007*. LNCS, vol. 4825, pp. 253–266. Springer, Heidelberg (2007)
22. Salton, G., McGill, M.J.: *Introduction to Modern Information Retrieval*. McGraw-Hill, New York (1983)
23. Summers, E., Isaac, A., Redding, C., Krech, D.: LCSH, SKOS and Linked Data. In: *Proceedings of the International Conference on Dublin Core and Metadata Applications*, Berlin, Germany (2008)