# Vocabulary Matching for Book Indexing Suggestion in Linked Libraries – A Prototype Implementation & Evaluation

Antoine Isaac<sup>1,2</sup>, Dirk Kramer<sup>2</sup>, Lourens van der Meij<sup>1,2</sup>, Shenghui Wang<sup>1,2</sup>, Stefan Schlobach<sup>1</sup>, Johan Stapel<sup>2</sup>

<sup>1</sup> Vrije Universiteit Amsterdam
<sup>2</sup> Koninklijke Bibliotheek, Den Haag
{aisaac,lourens,swang,schlobac}@few.vu.nl,{Dirk.Kramer,Johan.Stapel}@kb.nl

Abstract. In this paper, we report on a technology-transfer effort on using the Semantic Web (SW) technologies, esp. ontology matching, for solving a real-life library problem: book subject indexing. Our purpose is to streamline one library's book description process by suggesting new subjects based on descriptions created by other institutions, even when the vocabularies used are different. The case at hand concerns the National Library of the Netherlands (KB) and the network of Dutch local public libraries. We present a prototype subject suggestion tool, which is directly connected to the KB production cataloguing environment. We also report on the results of a user study and evaluation to assess the feasibility of exploiting state-of-the art techniques in such a reallife application. Our prototype demonstrates that SW components can be seamlessly plugged into the KB production environment, which potentially brings a higher level of flexibility and openness to networked Cultural Heritage (CH) institutions. Technical hurdles can be tackled and the suggested subjects are often relevant, opening up exciting new perspectives on the daily work of the KB. However, the general performance level should be made higher to warrant seamless embedding in the production environment-notably by considering more contextual metadata for the suggestion process.

# 1 Introduction

*Motivation* Cultural Heritage (CH) institutions usually own well-described collections of objects, and publishing (descriptions of) their assets to unknown users is an inherent part of their mission. Of course, both aspects are equally crucial for the Web of Data: data with structured meta-data, and the drive to publish data for unforeseen reuse.

This has lead to many CH institutions being at the forefront of developing and applying Semantic Web (SW) technology. Important initiatives for representing knowledge on the SW, such as Dublin Core [1] or SKOS [2], have been driven by expertise and requirements from the CH domain. In turn, many recent projects have largely benefited from using SW techniques to make CH material accessible more easily, e.g., in [3, 4].

Given these success stories, it is no surprise to see yet another fruitful crossfertilisation opportunity between the two fields: *semantic interoperability*. In the SW community, recognition of this problem has lead to exhaustive research in ontology matching [5], the goal of which is to make interoperable the data expressed in different ontologies. Again, CH institutions have strong interest for adapting this technology in their daily routine and for their future vision. Collections from different institutes are indeed described with different knowledge organisation systems (KOS), such as thesauri. In a more and more inter-linked CH world requiring cross-collection applications, alignments between different KOSs become crucial.

Linking vocabularies across Dutch libraries The National Library of the Netherlands (KB) holds numerous books also described by other libraries, many of them having their own way of semantically describing (indexing) the content of books. This leads to human indexers at the KB still going through each book to re-index it in "the KB way," *i.e.*, using their own thesauri, although that book might have already been described by some other institutions.

Such re-indexing process can be streamlined by suggesting indexers relevant concepts from KB thesauri, based on other institutions' descriptions. In the context of the STITCH project, we implemented a prototype that provides such a functionality. For each incoming book already indexed with concepts from the Biblion thesaurus (used by Dutch public libraries), it suggests a new indexing with the Brinkman subject thesaurus used in the KB. The suggestion is based on semantic links between these two thesauri which are derived from generic ontology matching methods [6, 7].

This prototype has been directly connected to the KB production cataloguing environment, used daily by indexers. To the best of our knowledge, KB is one of the first non-academic institutions to use and thoroughly evaluate generic ontology matching technology for a real-life application. This paper describes the index suggestion prototype, an exhaustive evaluation, and some lessons learned, particularly from the perspective of the SW technology that was used.

*Evaluating the SW approach in practice* In order to assess the usefulness in practice, we thoroughly evaluated our proposed methodology and prototype implementation. We had two goals, for which involvement of KB staff at all levels (management, technical and user—indexers) was critical:

- technology transfer. STITCH gathers researchers and practitioners to investigate how SW technology can be used to solve CH interoperability issues. In previous experiments, researchers have played a leading role. Here, KB is sought to actively participate in all steps of the development of a SW-based tool, from data preparation to design and testing.
- feasibility study. This prototyping experiment had to determine whether and to which extent—KB could benefit in the medium term from the techniques employed. The tool must deliver appropriate suggestion performance,

and its design has to fit existing processes seamlessly, both at the conceptual and technical levels. Evaluation is thus key, as is the realistic embedding of the developed tool into the production process.

Findings Our suggestion prototype, on one hand, shows the limitation of the used matching methods to provide high quality suggestions. However, the close collaboration with KB staff gives us much deeper insight into the problem and, importantly, points out potential improvements. In particular, exploiting existing book descriptions for matching thesauri is a valid approach. Only, it may require more thesauri to be linked, in order to make the re-indexing process more efficient.

From that perspective, it is crucial that the prototype demonstrated that SW components can be successfully plugged into the KB production environment. This brings the latter a higher level of flexibility and openness, making it better interoperable with other providers (and consumers) of semantic data.

In Sec. 2 we present the specific case we addressed in our experiment. Sec. 3 then gives a general technical overview of our prototype. Sec. 4 describes our user study and its results, before we conclude.

# 2 Book re-indexing at the KB

To manage and allow access to its collections, KB relies on a careful description of its books. This includes *subject indexing*, a concise and controlled reformulation of book subjects, typically done by assigning concepts from a KOS. Indexing requires trained employees to analyze the content of books and carefully pick the most appropriate concept(s) for describing them. This is labor-intensive, and KB is aiming to assist indexers by using (semi-)automatic techniques.

#### 2.1 Existing work on assisting document description

The first approach to automated document description is to apply natural language processing tools to the textual content (or summary) of documents. In SW research, text analyzers have been used to produce (structured) semantic annotations along formal ontologies that are either pre-defined or learned onthe-fly [8]. Similar techniques have also been deployed to produce document annotations with KOSs that are closer to the ones used in libraries—see [9, 10].

However, these techniques require a sufficient amount of textual data, which is not always available, especially for KB whose books have mostly not been digitized. As a result, one has to find sets of related documents (as in the CHOICE project [10]) or to exploit the limited textual information present in the metadata record (the title, sometimes a summary). KB already experimented with the latter approach, without obtaining convincing results.

We have opted for a different approach, exploiting descriptions resulting from a principled interpretation. Our problem is therefore to bridge across different interpretations, using ontology matching, rather than bridge the semantic gap between uninterpreted text content and controlled indexing.



Fig. 1. (Partial) cloud of related collections and KOSs in KB's environment

#### 2.2 The need for re-indexing at KB

Our proposal is to assist indexers by re-using existing subject descriptions for books, as contributed by other institutions. Quite often, indeed, a same book can be of interest—and thus described—by several actors. KB thus holds numerous books that are also held by other Dutch libraries, by publishers, or even by foreign institutions. For instance, the collections of KB and the Dutch (local) public libraries contain around a quarter of million books in common. Figure 1 shows some collections that are related to the KB ones. As the goals of the holding institutions often overlap, there is a certain amount of redundancy. Sharing and connecting descriptions in such a network would thus generate interesting synergies and productivity gains.

In fact, many libraries already share their descriptions. For instance, book metadata in the KB is stored in a database—OCLC-Pica's *GGC*—which is also used by other libraries for shared cataloguing, in particular, by the Dutch academic libraries and the Dutch public libraries. However, if integration of descriptions at a basic level (authors, publication date, etc.) could be done, these libraries still perform subject indexing with their own KOSs, as depicted in Fig. 1. Indexing is indeed largely application-specific: different collections are gathered for specific information needs, and require descriptions with different granularity or expertise levels, or even in different languages. There can be significant semantic overlap between the KOSs used, but it is not possible to benefit directly from the work done by other institutions. To re-use existing descriptions in the KB application context, they have first to be fit into KB needs.

In this paper, we focus on the specific KB–public libraries case. KB, for indexing the subject of books in its *legal deposit* collection, uses the *Brinkman* subject thesaurus (hereafter called "Brinkman"). This thesaurus has a generic scope and contains more than 5,000 topical concepts. The main subject thesaurus

used in the public libraries is the *Biblion general keyword thesaurus* (hereafter called "Biblion") which contains approximately 18,000 concepts. KB indexers quite often have to *re-index*, with Brinkman concepts, books that have already been indexed with Biblion, while the scope and structure of the two general thesauri overlap quite much. The question is whether this re-indexing can be streamlined by an automatic process providing indexers with candidate subjects.

#### 2.3 Re-indexing requirements

One solution is to create semantic correspondences between the elements of the KOSs at hand, allowing to convert descriptions from one system to the other. This scenario, described in [11], can be regarded as a problem of ontology matching.

In our case, re-indexing requires to implement a translation process from concepts of a Biblion-indexed book into Brinkman ones, so as to yield a Brinkman indexing of that book. Such a functionality can be formalized by the function:

$$f_r: 2^{\mathcal{BI}} \to 2^{\mathcal{BR}}$$

where  $2^{\mathcal{BI}}$  and  $2^{\mathcal{BR}}$  denote the powersets of Biblion and Brinkman concepts.

At first sight, one expects such function to relate Biblion concepts to Brinkman concepts that have semantically equivalent or close meaning. *Variability* issues can however make such simple equivalence associations unfit. First, one same book may be described by two semantically different concepts, even though each of these concepts has an equivalent concept in the other KOS. This may for instance stem from institutions' indexing policies aiming at different description granularity levels. Second, related to the way KOSs are designed, a concept from one vocabulary may be expressed using several concepts from the other KOS, which will be combined together to describe the subject of a book.

To the best of our knowledge, there is no similar work deploying ontology matching for migrating subject indexes from a KOS to another. Ontology matching has been identified as a solution for the more general data migration/translation problem, but it is difficult to find examples of concrete deployments [5]. Also, in more industry-oriented database migration efforts, the focus is rather on translating data from one schema to another—that is, focusing on the relation between (metadata) fields rather than the values that populate them. Finally, in the CH domain, most projects exploiting KOS alignments focus on query reformulation or cross-collection browsing (as in HILT<sup>1</sup> or the SW-based Europeana Thought Lab<sup>2</sup>), which are related but different application issues.

## 3 Prototype design

#### 3.1 General description

The core functionality of our Brinkman subject suggestion tool (SST) is to use available Biblion indexing and other metadata to suggest new Brinkman index-

<sup>&</sup>lt;sup>1</sup> http://hilt.cdlr.strath.ac.uk/

<sup>&</sup>lt;sup>2</sup> http://www.europeana.eu/portal/thought-lab.html

ing. This functionality is coupled to the WinIBW software used to describe books in the production process of the KB acquisition and cataloguing department. WinIBW is connected to the aforementioned GGC cataloguing system, which is also used by the public libraries. The metadata including Biblion indexing can therefore be accessed and exploited seamlessly.

The SST can be activated when an indexer describes a book already indexed with a Biblion subject. A pop-up window presents then the new Brinkman suggestions. The correct suggestions can be selected and automatically inserted within the description of the book currently being edited in WinIBW. After this, the description is saved and stored in GGC.

In the following, we describe further the use and functioning of the SST, as well as some aspects of the suggestion process itself.

#### 3.2 User interface

The SST can be launched for a given book being described in the WinIBW system. It comes as a pop-up window split in two panels, as shown in Fig. 2. This window gathers the existing metadata of the book and suggests new subjects.

The left-hand side displays the existing metadata, using a layout similar to WinIBW. We reproduced this basic layout so as to integrate the prototype as seamlessly as possible in the current process. In particular, our panel keeps to the original description codes—"kenmerkcodes," or KMCs [12]. The only additions are the use of color for spotting different elements, and the access to basic additional information services. For instance, 112X corresponds to the "UNESCO" classification (very general, mostly subject-based classification of books, *e.g.*, "linguistics"), 111X to "KAR" (the "characteristic", roughly corresponding to the type of publication, *e.g.*, "thesis"), 1401 to DGP (intellectual level/target group, *e.g.*, "youth fiction, 7-8 years") and 5061 to NUR (Dutch book trade (topical) classification focused on genres, *e.g.*, "national history"). Further descriptions of these elements can be accessed via online help resources.

The right-hand side of the screen presents the suggestions (bottom). Three colors (blue, purple, red) are used to mirror the confidence level of the suggestions (see Sec. 3.4 for more details). Indexers can select the suggestion(s) they agree with by ticking them. There are also empty boxes for searching new concepts to add to the description ("Ander Brinkman onderwerp opzoeken").

For the user study, indexers are offered the possibility to give a general quality assessment of the suggestions for the book ("tevreden" radio buttons) as well as a comment ("opmerking"). When the screen has been filled, users can save the new indexing and go back to WinIBW by pressing the "STITCH" button.

There is also the possibility to ask for more information on suggestions (the '+' top-right), such as the concepts from the record that triggered them, and their numerical confidence value. This shows more suggestions (with lower confidence). Previously established alignments based on equivalence of concept labels (see Sec. 3.4) can also be requested to provide more suggestions (the 'a' top-right).



Fig. 2. The Suggestion Tool's user interface

#### 3.3 Architecture

Our SST prototype consists of various distributed components, as illustrated in Fig. 3.

Indexers start from the WinIBW cataloguing interface, a Windows client. Simply clicking a button, they launch a VisualBasic script that converts the current book data in XML format and displays it into an Internet Explorer pop-up window.

This window uses an XSLT stylesheet to display the book metadata, as well as querying various services and displaying the information got from them as XML data. To enhance information presentation, the stylesheet triggers access to two business web services—*Bol.com* for book images and *Picarta* for obtaining additional book information. To fulfill its main concept suggestion mission, it accesses two "alignment web services" hosted on machines at the Free University of Amsterdam. Primary suggestions are provided by a SWI-Prolog server that exploits the statistical rules described in Sec. 3.4. Additional (lexical) alignments (for the 'a' option) are accessed via a SPARQL endpoint [13] on top of a Sesame<sup>3</sup> RDF store.

At this stage, indexers can add one or more subjects, which they find with the help of a vocabulary service (JavaScript/Ajax). When they are finished,

<sup>&</sup>lt;sup>3</sup> http://openrdf.org



Fig. 3. The Suggestion Tool's architecture

clicking on the STITCH button sends back the created metadata into WinIBW (JavaScript/VBScript).

For the purpose of evaluation, existing book metadata, suggestions and user input (including the evaluation-specific satisfaction level and comments) are also stored in XML (VBScript). A dedicated XLST stylesheet enables the redisplaying of that information for each considered book, as well as the creation of tables with all gathered data.

Access to concept information is done via a generic SKOS vocabulary service<sup>4</sup> we implemented in STITCH. To streamline management of KOSs in our architecture, we have opted for converting them to the SKOS standard to represent KOSs in RDF [2]. The service allows to retrieve SKOS data for concepts (labels, documentation, semantic relationships to other concepts) and enables searching for concepts in a given KOS, *e.g.*, by means of autocompletion. It thus provides the basic means to explore vocabularies.

Crucially, the service can be used to access distributed RDF KOS sources, published via either a SPARQL endpoint or other instances of the service that are loaded with the corresponding SKOS files. This allows to seamlessly extend our prototype to deal with unforeseen KOSs, when these are published in SKOS on the Web of Data. Our SST could then be articulated with more collections, even ones already linked at the semantic level—see Fig. 1.

#### 3.4 Brinkman subject suggestion rules

The suggestion process relies on statistical techniques to find association between the elements found in the available book metadata and the Brinkman concepts to

<sup>&</sup>lt;sup>4</sup> http://stitch.cs.vu.nl/repository

be suggested for indexing. For the sake of brevity, we limit ourselves here to the description of what the suggestion tool outputs and which information sources it uses, and only give an overall description of the suggestion rules creation process.

Metadata used for creating and applying suggestion rules In a first version of the suggestion tool, we tried to build and exploit an alignment between Biblion (LTR field) alone and Brinkman. The idea was to provide suggestions based on simple, ready-to-use individual "translation rules" from Biblion to Brinkman, as obtained from lexical comparison of their labels.

However, as previously mentioned, our case comprises books which are shared between the KB and the public library collections. Many of these books (nearly 238K) are thus indexed with both Biblion and Brinkman. Following the reindexing evaluation approach of [11], we used the existing Brinkman subjects of these books as a gold standard with which we could automatically compare new suggested subjects. Based on the low performance of this preliminary automatic evaluation, we decided to drop these first suggestion rules. They are only accessible via the advanced 'a' option mentioned above, as a complementary source of suggestions.

Matching techniques using the *extension* of concepts, namely the way they co-occur in book descriptions, seemed much more promising. From a theoretical perspective, they fit well the application scenario, which is about reproducing indexing practices instead of trying to grasp the "intrinsic meaning" of concepts. Practical evaluations in the case of other KOSs—*e.g.*, for the Library track of the Ontology Alignment Evaluation Initiative [14] and for STITCH research [15]—have further demonstrated that these techniques indeed perform well for reindexing.

As a matter of fact, the dually indexed corpus that we used for automatic evaluation above could also be used as a learning set for deriving statistical associations between Biblion and Brinkman concepts. In order to fit the application scenario better and to take into account potential indexing- and vocabularyrelated granularity mismatches, we sought to find suggestions based on *combinations* of one, two or three Biblion concepts. We also extended the matching process by considering the values found in *several* metadata fields. This provides more information to elicit finer-grained suggestion rules. Our prototype suggestion tool uses as input the following four metadata fields (and their associated vocabularies of controlled values) used in books with Biblion indexing:

- LTR Biblion concepts,
- AUT main authors of books,
- KAR (coded) "characteristic" and
- DGP (coded) intellectual level/target group.

**Establishment of suggestion rules** The most important source of information for the SST is the set of books which are described by both Brinkman and Biblion. In order to analyse the reliability of the rules, we extracted a third of this set as a new test set for automatic evaluation. Suggestion rules are established by looking for statistical associations between a source combination of metadata values (including at least one Biblion concept, and optionally, one or two values for the AUT, KAR, DGP fields mentioned above) and target Brinkman concepts. We adapted previous work on instancebased thesaurus matching [6,7] to fit the re-indexing scenario. Given a source combination of metadata values, noted as  $C_i$ , the probability that a Brinkman concept  $B_j$  should be suggested is calculated as

$$p(C_i \to B_j) = \frac{|C_i \cap B_j|}{|C_i|} \tag{1}$$

where  $|C_i \cap B_j|$  is the number of books which are described by both  $C_i$  and  $B_j$ and  $|C_i|$  is the number of books which are described by  $C_i$ . In order to reduce the bias from the sparseness of the data, we took a variation of (1):

$$p(C_i \to B_j) = \frac{|C_i \cap B_j|}{|C_i|} - 0.46 \times \frac{2}{|C_i| + 1},$$
(2)

which gives the best performance in practice—as observed by automatic evaluation over the third of the dataset which we extracted for testing.

In this way, we established a set of 1.5 million suggestion rules, each with a "confidence level"—in fact, the probability for a suggestion to be correct determined by formula (2). Table 1 presents some of them. The left-hand side of each rule corresponds to a combination of metadata values found in the training set; the right-hand side corresponds to the Brinkman suggestion. The rightmost column gives "correct books," the number of books in our test set that had both the source combination and the target concept, and "total," the number of books that matched the source combination.

Compared in the second compared	Confidence	Correct books	
Source combination $\rightarrow$ target concept	level	/ Total	
DGP:Jeugd fictie; vanaf 13 jaar			
(youth fiction; from 13 year)	0.995	182/182	
+ KAR:Stripverhaal (comics)			
$\rightarrow$ BTR:stripverhalen (comics)			
LTR:Reisgidsen (travel guides) + LTR:Spanje (Spain)	0.082	50/50	
$\rightarrow$ BTR:Spanje ; reisgidsen (Spain; travel guides)	0.982	50/50	
LTR:Brabantse dialecten (Brabant dialects)	0.967	27/27	
$\rightarrow$ BTR:Nederlandse dialecten (Dutch dialects)	0.301	21/21	
LTR:Liefde (love) + AUT:Jeanette Winterson			
$\rightarrow$ BTR:Romans en novellen ; vertaald	0.540	1/1	
(novels and novellas; translated)			
LTR:Bouwkunde (building engineering)			
$\rightarrow$ BTR:leermiddelen ; bouwtechniek	0.196	25/123	
(learning material; building engineering)			

Table 1. Subject suggestion rules

Working the suggestion rules When a book's metadata contains a suitable source combination (including Biblion subject indexing, and possibly, AUT, KAR and DGP annotations), the SST suggests a list of Brinkman subjects, each with a confidence value. This allows to rank the suggested concepts, as well as to classify them in broader confidence categories indicated by specific colors, as mentioned in Sec 3.2. Thus, blue suggestions correspond to a confidence level higher than 0.54—all suggestions based on concept combination (co-)occurring in a single book have this measure. Purple suggestions have a confidence level between 0.1 and 0.54, and the red suggestions are between 0.02 and 0.1.

#### 4 User study

#### 4.1 Aim and settings

Our user study aims to determine the extent to which the SST could be used at the KB. We are also interested in users' suggestions for improving the tool, or applying it in alternative ways.

We gathered six indexers of the *deposit* cataloguing team (out of a total 16). They were all trained indexers, with experience ranging between 10 and 30 years—five of them having at least 20 years experience. The tests were originally planned for a period of three weeks, but this was lengthened to 6 weeks after it appeared that too little suitable material could be gathered in that time span.<sup>5</sup> In total, 284 books were evaluated.

The evaluation task was seamlessly integrated in the daily work of the evaluators. As, when describing books with WinIBW, they recognized a book for which Brinkman suggestions could be done, they activated the tool, and then resumed to their "normal" tasks after that book had been indexed.

As already described, the SST interface required the evaluators to select which were the correct suggestions, or complete these with non-suggested concepts when required. It was also equipped with radio buttons with which evaluators could indicate their level of satisfaction, and a free-text field where they could enter their comments on the suggestions for the book at hand.

Before the test, the evaluators were given in a group briefing on the task and SST's features. We also gave each of them a questionnaire to fill at the end of the test period—see Appendix. After the test period, we organized a (group) evaluation interview. There, the evaluators where given an extra question list we devised after analyzing the input from the first questionnaire, so as to get more precise (and quantitative) feedback on specific points—see Appendix. The list also provided a sound structure to guide the interview, during which the evaluators had to answer it.

<sup>&</sup>lt;sup>5</sup> Currently, KB indexes books very shortly after their publication; the number of books indexed by Biblion *before* being indexed by KB turned thus to be smaller than we hoped. Luckily, our committed evaluators pretty soon decided themselves to collect all suitable books from their entire department, beyond the ones attributed to them *via* the normal KB process.

#### 4.2 Suggestion Performance

Using the correct indexing selected or added by the evaluators, we could measure the performance of our concept suggestions in terms of precision and recall. The 284 books evaluated were together given 4,021 Brinkman subject suggestions, for a total 468 correct concepts. Table 2 gives the results for each of the three suggestion classes mentioned in Sec. 3, where the "non suggested" category lists the number of individual concept indices that could not be found by the system and had to be added by indexers.

Suggestion class	# suggestions	# correct	precision	recall
blue	308	224	72.7%	47.9%
purple	1,188	127	10.7%	27.1%
red	2,525	28	1.11%	5.98%
non suggested		89		19.0%

Fable	2.	Suggestion	performance
-------	----	------------	-------------

Looking at the "blue" row, we achieve a recall of 47.9% with a precision of 72.7%. Taking the blue, purple and red suggestion results together, we can achieve a recall of 81% if we accept a much lower precision of one in ten correct suggestions (9.4%).

These results confirm our expectations regarding the categorization of suggestions in terms of precision classes. Interestingly enough, they are also consistent with the results of the automatic evaluation over the testing set, mentioned in Sec. 3.4—see Appendix.

#### 4.3 User feedback

**User satisfaction** We evaluated the more subjective aspect of user satisfaction based on the following sources:

- the general satisfaction level given to each book's suggestions;
- the comments given for each book;
- the answers given to the first and second questionnaires;
- the informal feedback obtained during the group interview.

At the level of books, the satisfaction level seemed first relatively high. Out of the 264 books for which this information was filled, 193 were given a "++" or a "+", the two best marks available out of five. However, this positive appreciation does not correlate with the perceived global usefulness of the tool, as it emerged from latter comments and questionnaires. Indeed, when asked if they would continue using the tool *as it is*, only two evaluators answered positively.

A first source of complaints is the robustness of the tool: the vocabulary service hosted in Amsterdam crashed a couple of times. But the evaluators could abstract from this, and they generally appreciated the functionality and the design by the prototype.

Rather, the most important issue was clearly the perceived quality of individual concept suggestions: the tool did not hinder evaluators' work, yet they expected more accurate suggestions. The current prototype displays quite many concepts, many of which were considered useless. At low confidence levels, suggestions can be based on just a insignificant overlap that is not meaningful from an indexing perspective, as for "Cross-culturele psychologie" (cross-cultural psychology) and "groepsdynamica" (group dynamics). While including them still brings some correct suggestions, the "serendipity" effect we wanted to test was not considered as an advantage, in such a controlled indexing process. The basic ranking of suggestions, based on their confidence level and expressed by the color distinctions, was judged a very positive feature. Yet evaluators still felt they were compelled to examine all suggestions, even the least certain ones. This suggests finer strategies have to be devised concerning the way suggestions are presented to the user, depending on their confidence level. A higher confidence threshold might therefore be required for a suggestion to be displayed among the first choices.

Our prototype also suggests that better suggestions could (and should) be obtained, in absolute. The way suggestions are obtained is indeed judged very promising, as testified by reactions collected on individual suggestions. Suggesting new indexing based on previous ones often successfully managed to bridge the "indexing gap" between the two collections at hand, *e.g.*, for "Perzische taal" (Persian language) and "Iraanse taal- en letterkunde" (Iranian language and literature). In fact in some cases it provided with suggestions which evaluators would not have anticipated, but that turned out to be correct.

**Feedback for improvement** As a token of interest, the evaluators strongly suggested to have the same approach applied to other situations, i.e., not only books indexed with Biblion. They hinted that the tool could also exploit metadata fields like the ones using the aforementioned NUR or other elements of Fig. 1, such as the two lists of Persons as subjects and Corporations, which have a strong connection with topical concepts. As a matter of fact, these extra sources may be used in the first place to improve the performance of the Biblion-based process, they confirmed us. Exploiting co-occurrences of a larger set of concepts would allow to grasp some interesting patterns. But they might also be sufficient to do useful suggestions when a Biblion concept is missing. Finally, showing their awareness of the semantic connections between the different indexing systems in their environment—and their potential value—they suggested that concepts from other vocabularies than Brinkman could be suggested. In particular, the UNESCO classification, when it is not yet given for a book, would be a useful and easy target to aim at. Other elements of Fig. 1 such as the two lists of Persons as subjects and Corporations, would be valuable too.

In fact, as they felt that suggestion process was similar to the way they perform indexing themselves, our evaluators also brought ideas to investigate in the longer term, to make that similarity even higher. For instance, using "negative rules" such as "LTR:Zussen (sisters)  $\rightarrow$  BTR:Zussen, unless KAR:roman is present"; or setting up a more interactive suggestion process that actualises the list of suggested Brinkman concepts once a first one has been selected.

A last issue, which evaluators felt was clearly more urgent to tackle, is the incompleteness of the data that is used for the suggestions and of the set of concepts that can be suggested. In many cases, they expected more suggestions to be made, considering the available metadata that could be used as a learning set—which they have rather extensive knowledge on. This can be explained by our selecting only two thirds of the dually indexed books as a learning set. But the original dump of records itself dates back to 2006, and is therefore relatively obsolete. Similarly, the SKOS versions of the vocabularies which we exploit both for the suggestion and the vocabulary services (esp. to allow users to search for additional concepts using auto-completion) are mere snapshots. They were already slightly outdated at the time of our experiment, which was detected by indexers. Update mechanisms should be set up to exploit the most recent data.

# 5 Conclusion

As regards our initial aims, the main results are:

- SW technology uptake: despite a few robustness issues that should be solved in a next implementation phase, the main technical goals have been met. Exploiting Semantic Web technologies, we could develop a prototype with features both innovative and relevant from a domain perspective, and which can be integrated into the KB production process.
- feasibility of the re-indexing approach: our evaluation showed strong interest from the book indexers. The quality of the suggestions must clearly be improved before they start using such a tool. However, the prototype has validated the principle of using previous indexing to suggest new one, as well as most of our design decisions. Further, the user study provided the development team with clear and feasible solutions to start improving on the current prototype, wrt. both interface and suggestion quality.

An important point that motivates the adoption of SW solutions for the case at hand is their genericity and flexibility. We could well have developed a prototype with similar functionality, using more traditional techniques, on top of the same collections and KOSs. In fact, some of the key benefits of SW solutions, such as relying on unambiguous identifiers for resources, are already available for use in the current application context. However, our technological choice guarantees high flexibility while not requiring much additional implementation effort—off-the-shelf RDF stores can be deployed easily and accessed by simple scripts using SPARQL. Especially, the prototype can be adapted so as to use other metadata (and their associated KOSs) to derive more precise suggestion rules, which is one of the most important directions for future improvement.

As a matter of fact, our tool could be seamlessly deployed with completely different collections and vocabularies, provided the associated metadata is also made available in the appropriate RDF format—which was one of the bottlenecks we identified as we proceeded with our own data conversion effort. Considering the growing interest of CH institutions for Open Linked Data<sup>6</sup> developments, e.g., [16–18], one can be optimistic about it. For instance, as shown in Fig 1, collections that are indexed by the subject heading lists of the American Library of Congress and the French National Library (resp., LCSH and RAMEAU) overlap with the KB one. Our prototype could be extended to have KB benefit from the work that is being done in these libraries. In that specific case we could even seek to exploit existing semantic mappings between the two aforementioned KOSs, which SW techniques made much easier to publish and access.<sup>7</sup>

Indeed, we believe that an important step has been made at the KB towards embracing the Open Linked Data vision. In the library domain, as in many others, institutions will not adopt new technologies unless they have been demonstrated with scenarios bringing clear value to them. Here, we have been able to show that a library can benefit from having other libraries' data being accessible, even for a back-office, mission-specific process like indexing.

In turn, our prototype confirmed that there is value for KB making accessible its own data and vocabularies. If other institutes could benefit from KB descriptions in their own metadata creation process as well, this would boost the productivity of the entire network. In fact, as reflected in Fig. 1, this experiment was also the opportunity to identify in a clearer way:

- from a technical perspective, the various vocabularies that can be connected at the semantic level to enhance the indexing process;
- from a more institutional perspective, the collection holders that form together a neighbourhood of peers in the network, which can benefit from the virtuous circle mentioned in the previous paragraph.

This may drastically enhance synergies, and make common adoption of these new technologies an even more realistic perspective. Other institutes in the KB neighbourhood, such as the Dutch association of public libraries, are actually already considering moving to RDF and SKOS [19]. The time is now relatively near, when a library cloud such as the one in Fig. 1 will lend itself to treatments that allow the participating institutions to benefit from each other's efforts.

Acknowledgements This work is funded by the NWO CATCH and EU eContentPlus programmes (STITCH and TELPlus projects). We are especially grateful to the team of KB indexers who participated the evaluation and provided crucial feedback.

#### References

- 1. Dublin Core Metadata Initiative: Dublin Core Metadata Element Set. DCMI Recommendation, http://dublincore.org/documents/dces/ (2008)
- Miles, A., Bechhofer, S.: SKOS Reference. W3C Proposed Recommendation, http://www.w3.org/TR/skos-reference/ (2009)

<sup>&</sup>lt;sup>6</sup> http://linkeddata.org/

<sup>&</sup>lt;sup>7</sup> See http://id.loc.gov and http://stitch.cs.vu.nl/rameau.

- Hyvönen, E., Mäkelä, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., Junnila, M., Kettula, S.: MuseumFinland – Finnish Museums on the Semantic Web. Journal of Web Semantics 3(2) (2005)
- 4. Wielemaker, J., Hildebrand, M., van Ossenbruggen, J., Schreiber, G.: Thesaurusbased search in large heterogeneous collections. In: Proc. 7th International Semantic Web Conference (ISWC 2008), Karlsruhe, Germany (2008)
- 5. Euzenat, J., Shvaiko, P.: Ontology Matching. Springer Verlag (2007)
- Isaac, A., van der Meij, L., Schlobach, S., Wang, S.: An empirical study of instancebased ontology matching. In: Proc. 6th International Semantic Web Conference (ISWC 2007), Busan, Korea (2007)
- Wang, S., Isaac, A., van der Meij, L., Schlobach, S.: Multi-concept alignment and evaluation. In: Proc. 2nd ISWC Workshop on Ontology Matching (OM 2007), Busan, Korea (2007)
- 8. Buitelaar, P., Cimiano, P., eds.: Ontology Learning and Population: Bridging the Gap between Text and Knowledge. IOS Press (2008)
- Golub, K.: Automated subject classification of textual web documents. Journal of Documentation 62(3) (2006) 350–371
- Gazendam, L., Malaisé, V., de Jong, A., Wartena, C., Brugman, H., Schreiber, G.: Automatic annotation suggestions for audiovisual archives: Evaluation aspects. J. Interdisciplinary Science Reviews 34(2-3) (2009) 172–188
- Isaac, A., Wang, S., Zinn, C., Matthezing, H., van der Meij, L., Schlobach, S.: Evaluating thesaurus alignments for semantic interoperability in the library domain. IEEE Intelligent Systems 24(2) (2009) 76–86
- OCLC: Kenmerkcodes in titelrecords. http://oclcpica.org/?id=13&ln=nl&par= p-kmc (in Dutch, accessed in August 2009)
- Clark, K.G., Feigenbaum, L., Torres, E.: SPARQL Protocol for RDF. W3C Recommendation, http://www.w3.org/TR/rdf-sparql-protocol/ (2008)
- Euzenat, J., Isaac, A., Meilicke, C., Shvaiko, P., Stuckenschmidt, H., Svab, O., Svatek, V., van Hage, W.R., Yatskevich, M.: Results of the Ontology Alignment Evaluation Initiative 2007. In: Proc. 2nd ISWC Workshop on Ontology Matching (OM 2007), Busan, Korea (2007)
- Schopman, B., Wang, S., Schlobach, S.: Deriving concept mappings through instance mappings. In: Proc. 3rd Asian Semantic Web Conference (ASWC 2008), Bangkok, Thailand (2008)
- Malmsten, M.: Making a Library Catalogue Part of the Semantic Web. In: Proc. International Conference on Dublin Core and Metadata Applications (DC 2008), Berlin, Germany (2008)
- 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)
- Neubert, J.: Bringing the "Thesaurus for Economics" on to the Web of Linked Data. In: Proc. WWW Workshop on Linked Data on the Web (LDOW 2009), Madrid, Spain (2009)
- Vereniging Openbare Bibliotheken: Informatiearchitectuur Openbare Bibliotheken. http://www.debibliotheken.nl/content.jsp?objectid=22642 (in Dutch, 2008)

#### Appendix

An online annex to this paper (http://www.few.vu.nl/~aisaac/iswc2009/) features the 1st and 2nd user study questionnaires and the results of the preliminary automatic evaluation of suggestions using existing dually indexed books.