Results of the OAEI 2007 Library Thesaurus Mapping Track

Antoine Isaac, Lourens van der Meij, Shenghui Wang, Henk Matthezing

December 5, 2007

1 Test set

The National Library of the Netherlands (KB) maintains two large collections of books: the Deposit Collection, containing all the Dutch printed publications (one million items), and the Scientific Collection, with about 1.4 million books mainly about the history, language and culture of the Netherlands.

– *indexed* – using its own controlled vocabulary. The Scientific Collection is described using the GTT thesaurus, a huge vocabulary containing 35,194 general concepts, ranging from Wolkenkrabbers (Sky-scrapers) to Verzorging (Care). The books in the Deposit Collection are mainly indexed against the Brinkman thesaurus, which contains a large set of headings (5,221) for describing the overall subjects of books. Both thesauri have similar coverage (2,895 concepts actually have exactly the same label) but differ in granularity.

For each concept in the two thesauri, the usual detailed lexical information is provided: preferred labels (each concept has exactly one of them), synonyms (961 for Brinkman, 14,607 for GTT), extra hidden labels (134 for Brinkman, a couple of thousands for GTT) or scope notes (6,236 for GTT, 192 for Brinkman). The language of both thesauri is Dutch,[1] which makes this track ideal for testing alignment in a non-English situation.

The two thesauri also provide structural information for their concepts, in the form of *broader* and *related* links. However, GTT contains only 15,746 hierarchical *broader* links between 35,194 concepts and 6,980 associative *related* links. Within the Brinkman thesaurus, there are 4,572 hierarchical links and 1,855 associative ones. On average, one can expect at most one parent per concept, for an average depth of 1 and 2, respectively.[2] The structural information found in the case is very poor.

For the purpose of the OAEI campaign, the two thesauri were made available in SKOS format. OWL versions were also provided, according to the – lossy – conversion rules detailed on the track page.[3]

[1] A quite substantial part of GTT concepts (around 60%) also have English labels.

[2] Particularly, the GTT thesaurus has 19,752 root concepts.

[3] http://www.few.vu.nl/ aisaac/oaei2007/index.html

2 Evaluation and results

Although ten teams showed some interest in the track, only five submitted preliminary results (in august) and three handed in final results:

- Falcon [3]: 3,697 `skosm:exactMatch` mappings
- DSSim [8]: 9,467 `skosm:exactMatch` mappings
- Silas [9]: 3,476 `skosm:exactMatch` mappings and 10,391 `relatedMatch` mappings.

Two evaluation procedures were chosen, each of them motivated by a potential case of mapping usage, as introduced in [5]. The first one is *thesaurus merging*, where alignment is used to build a new, unified thesaurus from GTT and Brinkman thesauri. Evaluation in such a context requires assessing the validity of each individual mapping, which leads to a "standard" alignment evaluation procedure.

The second usage scenario for the alignment is *annotation translation* from one thesaurus to the other. Here, books are annotated using one thesaurus, and the alignment is used to produce a corresponding annotation using the other thesaurus.

2.1 Evaluation in a thesaurus merging scenario

For this evaluation task, there was no exhaustive reference alignment available. Given the size of the vocabularies, it was impossible to build one, so we had to perform a manual evaluation on participants' results.

Inspired by the anatomy and food tracks of OAEI 2006 [2], we opted for evaluating precision using a reference alignment based on on a lexical procedure[4]. We were also able to produce quantitative measures for coverage, which we define here as the proportion of all good mappings found by an alignment divided by the total number of good mappings produced by all participants and those in the reference. This coverage is different from absolute recall, which is unknown. But we can hypothesize that it is proportional to recall,[5] and in any case it provides an upper bound for it – as the correct mappings found by all participants give a lower bound for the total number of correct mappings.

For manual evaluation, the set of all *equivalence* mappings[6] was partitioned into parts unique to each combination of participant alignments plus reference set (15 parts in all). For each of those parts which were not in the lexical reference alignment, a sample of mappings was selected, and evaluated manually. A total of 330 mappings were assessed by two Dutch native experts.

[4] This makes use of direct comparison between concepts' labels, but also exploits a Dutch morphology database that allows to recognise variants of a word, *e.g.* singular or plural. 3,659 reliable equivalence links are obtained this way.

[5] We are in a situation where the two thesauri cover a same (very general) domain. A good matching shall therefore find links for all the concepts in the two vocabularies.

[6] We did not proceed with manual evaluation of the *related* links, as only one contestant provided with such links, and their manual assessment is much more error-prone.

From these assessments, precision and coverage were calculated with their 95% confidence intervals, taking into account sampling size and evaluator variability. The results are shown in table 1, which identifies clearly Falcon as performing better than both other participants.

Alignment	Precision			Coverage		
DSSim	0.134	±	0.019	0.31	±	0.19
Silas	0.786	±	0.044	0.661	±	0.094
Falcon	0.9725	±	0.0033	0.870	±	0.065

Table 1: Comparison of precision and coverage for the thesaurus merging scenario

A detailed analysis reveals that the Falcon results are very close to the lexical reference, which explains their observed quality. 3,493 links are common to Falcon and the reference, while Falcon has 204 mappings not in the reference – of which 100 are good – and the lexical reference has 166 mappings not in Falcon. DSSim also uses lexical comparisons, but its edit-distance-like approach is more prone to error: we estimate that between 20 and 200 out its 8,399 mappings not in the lexical reference are correct.[7] Silas is the one which succeeds most in adding to the lexical reference: 234 of its 976 "non-lexical" mappings are correct. But it fails to reproduce one third of the lexical reference mappings, therefore its coverage is relatively low.

2.2 Automatic evaluation in an annotation translation scenario

The previous evaluation is oriented towards thesaurus merging case. Assessments of the mappings is usually assumed to be "neutral", the meaning of concepts being primarily derived from their intrisic information and their situation in the thesauri. From an organizational perspective, the evaluation here can be likened to *alignment evaluation* as presented *e.g.* in [11]. However, as shown in [5], there is value in considering more specific application cases, where mapping is deployed in running applications. We explain in the following how we performed an evaluation taking into account specific information needs, more in line with the "end-to-end" approach described *e.g.* in [11].

2.2.1 Evalution scenario

Among the possible scenarios of [5], we chose re-indexing, or *annotation translation*, which fitted previous efforts we made in [12]. In this scenario, aimed at indexers with an intricate expertise of Brinkman or GTT, an annotation translation tool supports the indexing of GTT-indexed books with Brinkman concepts, or vice versa. This is particularly useful if one of the two thesauri (we

[7] Out of the selection of 86 mappings in the set of 8363 mappings unique to DSSim not a single one was evaluated as correct by the human evaluators

have opted here for GTT) is dropped: a huge volume of legacy data has to be converted to the remaining annotation system (*i.e.* Brinkman). In this example case, this requires converting the GTT annotations into equivalent Brinkman annotations.

This evaluation scenario requires building a tool that can interpret the mappings provided by the different participants so as to translate existing GTT book annotations. Based on the quality of the results of the tool for books we know the correct annotations of, we can assess the quality of the initial mapping.

Here we follow the approach introduced in [12]. Out of KB's 2.4 million books, 250,000 actually belong both to KB Scientific and Deposit collections, and are therefore already indexed against both GTT and Brinkman thesauri. For evaluation, the existing Brinkman indices from this dually indexed collection are taken as a gold standard which an annotation translation system must aim to match. That is, for each book in the given corpus, we compare its existing (manually constructed) Brinkman index with the one computed from the GTT-Brinkman alignment produced by participants.

Evaluation settings The one-to-one mappings sent by participants were transformed into mapping rules, i.e.,

$$R : g_r \rightarrow B_r,$$

where the antecedent g_r is one GTT concept and the consequent B_r is a set of Brinkman concepts, to which the GTT concept is mapped.[8] Note that $|B_r| \geq 1$, because some GTT concepts are involved in several different mapping links.

The data set consists of all dually indexed books, 243,887 in total. Each book has both GTT and Brinkman annotations, denoted by G_t and B_t. The real GTT annotation G_t is used to fire the transformed mapping rules.

If the GTT concept of one rule is contained by the GTT annotation of one book, i.e., $g_r \in G_t$, then rule is *fired* for this book. As several rules can be fired for a same book, the union of the consequents of these rules forms the translated Brinkman annotation of this book, denoted as $B_r^{'}$. If this set of translated Brinkman concepts overlaps the real Brinkman annotation of this book, i.e., $B_t \cap B_r^{'} \neq 0$, we consider this book as *matched*.

Evaluation measures

- At the book level, we measure how many books have a rule fired on them, and how many of them are actually matched books, *i.e.*,

$$P_b = \frac{\#books\_matched}{\#books\_fired}, \qquad R_b = \frac{\#books\_matched}{\#all\_books},$$

[8] For Falcon and DSSim, only `exactMatch` mappings were provided, and hence taken into account. Silas provided both `exactMatch` and `relatedMatch`, which enabled different treatment, as we will see in the following.

where $\#all\_books$ is the number of the whole dually indexed books, $\#books\_fired$ is the number of books with a rule fired on them and $\#books\_matched$ is the number of matched books.

- At the annotation level, we measure how many translated concepts are correct, how many real Brinkman annotation concepts are missed and the combined measure of these two, i.e.,

$$P_a = \frac{\sum \frac{\#correct}{|B_r{}'|}}{\#books\_fired}, \quad R_a = \frac{\sum \frac{\#correct}{|B_t|}}{\#all\_books}, \quad J_a = \frac{\sum \frac{\#correct}{|B_t \cup B_r{}'|}}{\#all\_books}$$

  where $\#correct$ is the number of the translated Brinkman concepts which are actually used for the book.

The ultimate measure for alignment quality here is at the annotation level. The Jaccard overlap measure between found concepts and correct ones, i.e., J_a, plays a similar role as the F-measure does in information retrieval. Measures at the book level indicate to some extent users' (dis)satisfaction with the built system. A R_b of 60% means that the alignment does not produce any useful candidate for 40% of the books.

We would like to mention that in these formulas good and bad results are counted on a book and annotation basis, and not on a rule basis. This reflects the importance of different thesaurus concepts: a valid translation rule for a frequently used concept is more important than a valid rule for a rarely used concept. We have chosen this option – referred to as *micro-averaging* [10] – because it suits the real application context better.

Evaluation results Table 2 gives an overview of the evaluation results when we only use the `exactMatch` mappings. Falcon and Silas perform similarly, and much ahead of DSSim. As shown in Figure 1, nearly half of the books were given at least one correct Brinkman concept in the Falcon case, which corresponds to 65% of the books a rule was fired on. At the annotation level, half of the translated concepts are not validated, and more than 60% of the real Brinkman annotation is not found. We already pointed out that the mappings from Falcon are mostly generated by lexical similarity. This indicates that lexical equivalent mappings are not the only solution to the annotation translation scenario. It also confirms the sensitivity of mapping evaluation methods to certain application scenarios.

Participant	$\#rules$	$\#books\_fired$	P_b	R_b	P_a	R_a	J_a
Falcon	3,618	183,754	65.32%	49.21%	52.63%	36.69%	30.76%
Silas	3,208	175,309	66.05%	47.48%	53.00%	35.12%	29.22%
DSSim	9,467	188,165	18.59%	14.34%	13.41%	9.43%	7.54%

Table 2: Performance of annotation translations generated from `exactMatch` mappings produced by three participants

Among the three participants, only Silas generated `relatedMatch` mappings. To evaluate their usefulness for annotation translation, we combined them with the `exactMatch` ones so as to generate a new set of 8,410 rules. As shown in Figure 2, the use of `relatedMatch` mappings increases the chances of having a book given a correct annotation. However, unsurprisingly, precision of annotations decreases, because of the introduction of noisy results.

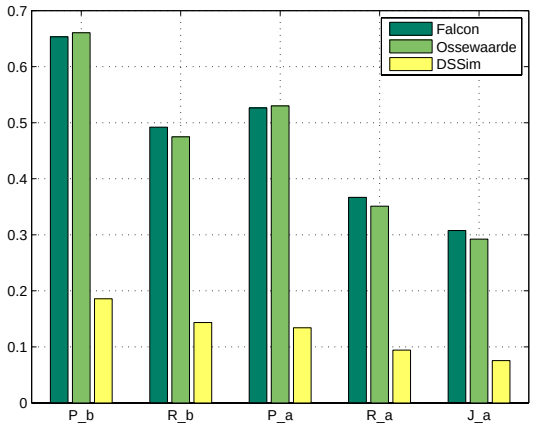

Figure 1: Comparison between all `exactMatch` mappings-generated annotations generated by Falcon, Silas and DSSim

2.3 Manual evaluation in an annotation translation scenario

The automatic evaluation technique used in previous section gives a first large and relatively cheap assessment of participants' results. Yet it is sensitive to indexing variation: several indexers annotating a same book (or a same indexer annotating it at different times) will select slightly different concepts. We decided to perform in the same context a *manual* evaluation to assess the influence of this phenomenon, as well as to validate or invalidate the results of the automatic evaluation. The research questions we want to address here are:

1. *quality of candidate annotations*: what is the quality of the annotations produced using the participants' alignments?
2. *indexing variability*: are evaluators' judgements consistent with the automatic evaluation of annotation translations?
3. *evaluation variability*: are judgements by different evaluators consistent?

For setting the manual evaluation we have partly followed the approach presented in [5].

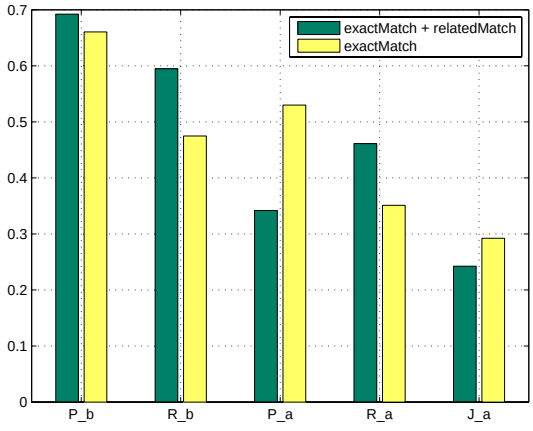

Figure 2: The influence of `relatedMatch` mappings in Silas' case

2.3.1 Data collection

Choice of books A sample of 96 books has been randomly selected among the dually annotated books annotated in 2006.

Computation of candidate annotations On these books, which are annotated by GTT, we applied the annotation translation procedures derived from each participants' results, using only the `exactMatch` links. For each book, the results of these different procedures are merged in lists of candidate concept annotations.

As we wanted some insight on the automatic evaluation based on existing Brinkman annotations, we also included these original annotations in the candidate lists. Finally we added candidates coming from previous alignment experiments [4]. These actually amount to an average of five candidate concepts per book.

Input for evaluators' assessments To collect assessments of the candidate annotations, paper forms were created, one form per book in the sample. Each form constitutes an evaluation task where the evaluator is asked to validate the proposed annotations.

The form – see annex 3 – presents the book's cataloguing information – author, title, year of publication, etc – plus the candidate annotations found for this book.

The evaluator is asked, for each of the candidate concepts, to say whether it would be *acceptable* for an index. This important precision is made to avoid evaluators making too narrow choices: the subject of the book can be unclear, or the thesaurus could contain several concepts equally valid for the book. The

evaluator can also feel that another human indexer could well have selected indices different from his.

For each candidate concept, the evaluator is offered the possibility to refine his choice by specifying whether the candidate is linked to the subject of the book, but not strictly equivalent to (or incompatible with) it. This is done by selecting one of the following links: hierarchical "broader" and "narrower" links, and the associative "related".

Afterwards, the evaluator is asked to select among the candidates the ones he would have *chosen as indices* for the book. He also has the possibility to specify annotation concepts not appearing in the proposed list, in a free-text field.

Pilot evaluation A preliminary version of the form was validated during a pilot evaluation in which two professional indexers were involved – one of them being the manager of Brinkman thesaurus. The experts agreed with the "acceptability" assessment criterion as being relevant from an application perspective. They also found that the average number of candidate concepts was reasonable.[9]

Some layout problems were amended. It was also decided that beyond cataloguing information, evaluators should be provided with the physical books, in order to fit the work setting they are accustomed to.

Evaluators and organization The judges involved in this evaluation are four professional book indexers from the Depot department at the KB, refered hereafter as A, B, C and D. They are all native Dutch speakers. Further, they are member of the same team which produced the data used for the automatic evaluation.

Each of the evaluators assessed the candidates for 96 of the books.[10] Evaluation was performed during sessions of one hour. 10 separated sessions were planned during one week, as the pilot evaluation demonstrated an average of 15 books per hour was reachable.

2.3.2 Evaluation results

Quality of candidate annotations Table 3 presents the results the acceptability assessments, averaged over the four evaluators. It also includes the lexical reference alignment used in section 2.1.

These are significantly and regularly higher than the figures obtained for automatic evaluation. This confirms the dependence of the scenario on the way indexing variability is taken into account in the evaluation setting.

[9] It actually turns out that not including the original Brinkman concepts in the proposed list dramatically decreases the perceived relevance of the candidate annotations, and could have introduced a more harmful bias into the evaluation.

[10] Four books turned out not to be accessble during the time of the evaluation.

Participant	P_a	R_a	J_a	P_a	R_a	J_a
Falcon	74.95%	46.40%	42.16%	52.63%	36.69%	30.76%
Silas	70.35%	39.85%	35.46%	53.00%	35.12%	29.22%
DSSim	21.04%	12.31%	10.10%	13.41%	9.43%	7.54%
Lexical	75.03%	46.61%	42.32%			

Table 3: Performance of `exactMatch` mappings produced by participants and lexical reference, as assessed by manual evaluation (left), compared to automatic evaluation results (right, from Table 2)

Evaluation variability We have computed the average Jaccard overlap measure[11] between the evaluators.

	B	C	D
A	65.72%	57.93%	60.16%
B		63.65%	60.50%
C			53.70%

Table 4: Average Jaccard overlap measure for acceptability assessments between the evaluators

On average, two evaluators therefore agree on 60% of their assessments.

We also measured the agreement between evaluators' individual assessments using Krippendorff's *alpha* coefficient [6], which is a common measure for computational linguistics tasks (word sense disambiguation, summarization, part-of-speech tagging). The results are the following, for each pair of evaluators:

	B	C	D
A	0.61	0.68	0.71
B		0.53	0.60
C			0.62

Table 5: Krippendorff's *alpha* agreement coefficient for acceptability assessments between the four evaluators

The overall *alpha* coefficient – computed over the four evaluators – is 0.62, which, according to standards, indicates a great variability. [7] as cited by [1] mentions for instance that an alpha measure below 0.8 indicates "a pretty low standard", hence undermining the reliability of the assessment according to content analysis standards. This is however to be put into perspective: the tasks usually considered in content analysis – part-of-speech tagging, named entity recognition – are obviously less prone to variability than the one considered here. [1] reports that for tasks like topic marking and word sense tagging – which concern less ambiguous words and pieces of text, compared to books –

[11] Which amounts, for each book, to the number of assessments agreed upon over the number total assessments plus the additional concepts that were selected by the evaluators.

the agreement values can be much lower, leaning towards the 0.67 border sometimes considered as "allowing to draw tentative conclusions" [6]. A more precise comparison is hence required with similar situations, which we could not do here.

Indexing variability As a first measure of indexing variability, we have included the original Brinkman indices in the candidate concepts to be evaluated. Table 6 shows the results of their acceptability assessment.

	P_a	R_a	J_a
Original Brinkman indices	81.60%	66.69%	60.35%

Table 6: Performance of original Brinkman indices, as assessed by manual evaluation

Original Brinkman indices are the results of a careful selection process and do not render all the acceptable concepts for a book. It is therefore no surprise that R_a is relatively low. However, it is very surprising to see that almost one original Brinkman concept out of five is not acceptable. The result show indeed that indexing variability matters a lot, even when the annotation selection criteria are made less selective.

To measure agreement between the indexers involved in our evaluation, we have computed the average Jaccard overlap measure between the indices they chose, as shown in Table 7.

	B	C	D
A	58.85%	53.51%	53.63%
B		60.52%	60.56%
C			46.58%

Table 7: Average Jaccard agreement measure for index selection between the four evaluators

Additionally, the Kripendorff coefficient was computed on the indices chosen by the different evaluators, as shown in table 8. Again, we have quite a low overall agreement value – 0.59. This confirms the high variability of the indexing task.

	B	C	D
A	0.60	0.70	0.63
B		0.51	0.51
C			0.58

Table 8: Krippendorff's *alpha* agreement coefficient for index selection between the four evaluators

3 Discussion

The first comment on this track concerns the *form* of the alignment returned by the participants, especially *wrt.* the type and cardinality of alignments.

First, all three participants proposed alignments using the SKOS links we asked for. However, only symmetric links (`exactMatch` and `relatedMatch`) were used: no participants proposed hierarchical `broader` and `narrower` links. Yet these links are useful for the application scenarios at hand. The `broader` links are useful to attach concepts which cannot be mapped to an equivalent corresponding concept but a more generic or specific one. This is likely to happen, since the two thesauri have different granularity but a same general scope.

Second, there is no precise handling of one-to-many or many-to-many alignments. Sometimes a concept from one thesaurus is mapped to several concepts from the other. This proves to be very useful, especially in the annotation translation scenario where concepts attached to a book should ideally be translated as a whole. As a result, we have to post-process alignment results, building multi-concept mappings from alignment which initially do not contain such links. This processing makes the evaluation of the relative quality of the alignments more difficult for the annotation scenario.

Of course these problems can be anticipated by making participants more aware of the different scenarios which will guide the evaluation. The campaign's timing made it impossible this year, but this is an option we would like to propose for next campaigns.

The results we have obtained also show that the performance of aligners vary from one scenario to the other, highlighting the strengths of different approaches. For the merging scenario, Falcon outperforms the two other participants. While in the translation scenario, Silas, which detects links based on extensional information of concepts,[12] performs as well as Falcon does.

Finally, we would like to discuss the overall quality of the results. The annotation translation scenario showed a maximum precision of 50%, and around 35% for recall. This is not much, but we have to consider that this scenario involves a high degree of variability: different annotators may choose different concepts for a same book. The manual evaluation by KB expert illustrate this phenomenon, and show that, under specific but realistic application conditions, the quality of participant's result is more satisfactory.

This still leaves the low coverage of alignments with respect to the thesauri, especially GTT: in the best case, only 9,500 of its 35,000 concepts were linked to some Brinkman concept. This track, arguably because of its Dutch language context, seems to be difficult. Silas' results, which are partly based on real book annotations, demonstrate that the task can benefit from the release of such extensional information. We will investigate this option for future campaigns.

[12] It is important to mention here that Silas was trained on a set of books which is different from the evaluation set we used.

Acknowledgements

The evaluation at KB could not hav been possible without the commitment of Yvonne van der Steen, Irene Wolters, Maarten van Schie, and Erik Oltmans. This work has also benefited from useful discussion from the following members of the STITCH project team, who contributed to the different articles and meetings which framed this work: Frank van Harmelen, Stefan Schlobach and Claus Zinn.

References

[1] Ron Artstein and Massimo Poesio. Inter-coder agreement for computational linguistics. Survey article, submitted to Computational Linguistics, 2007.

[2] Jrme Euzenat, Malgorzata Mochol, Pavel Shvaiko, Heiner Stuckenschmidt, Ondrej Svab, Vojtech Svatek, Willem Robert van Hage, and Mikalai Yatskevich. Results of the ontology alignment evaluation initiative 2006. In *Proceedings of the First International Workshop on Ontology Matching, collocated with the 5th International Semantic Web Conference (ISWC 2006)*, Athens, GA, US, November 5 2006.

[3] Wei Hu, Yuanyuan Zhao, Dan Li, Gong Cheng, Honghan Wu, and Yuzhong Qu. Falcon-ao: results for oaei 2007. In *Proceedings of the Second International Workshop on Ontology Matching, collocated with the 6th International Semantic Web Conference (ISWC 2007)*, Busan, Korea, November 11 2007.

[4] Antoine Isaac, Lourens van der Meij, Stefan Schlobach, and Shenghui Wang. An empirical study of instance-based ontology matching. In *Proceedings of the the 6th International Semantic Web Conference (ISWC 2007)*, Busan, Korea, November 11-15 2007.

[5] Antoine Isaac, Claus Zinn, Henk Matthezing, Lourens van der Meij, Stefan Schlobach, and Shenghui Wang. The value of usage scenarios for thesaurus alignment in cultural heritage context. In *Proceedings of the First International Workshop on Cultural Heritage on the Semantic Web, 6th International Semantic Web Conference (ISWC 2007)*, Busan, Korea, November 12 2007.

[6] Klaus Krippendorff. *Content Analysis: An Introduction to Its Methodology, chapter 12.* Sage, Beverly Hills, CA, 2004.

[7] Klaus Krippendorff. *Content Analysis: An Introduction to Its Methodology, second edition, chapter 11.* Sage, Thousand Oaks,CA, 2004.

[8] Miklos Nagy, Maria Vargas-Vera, and Enrico Motta. Dssim - managing uncertainty on the semantic web. In *Proceedings of the Second International*

Workshop on Ontology Matching, collocated with the 6th International Semantic Web Conference (ISWC 2007), Busan, Korea, November 11 2007.

[9] Roelant Ossewaarde. Simple library thesaurus alignment with silas. In *Proceedings of the Second International Workshop on Ontology Matching, collocated with the 6th International Semantic Web Conference (ISWC 2007)*, Busan, Korea, November 11 2007.

[10] Fabrizio Sebastiani. Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1), 2002.

[11] Willem Robert van Hage, Antoine Isaac, and Zharko Aleksovski. Sample evaluation of ontology-matching systems. In *Proceedings of the Fifth International Workshop on Evaluation of Ontologies and Ontology-based Tools, 6th International Semantic Web Conference (ISWC 2007)*, Busan, Korea, November 11 2007.

[12] Shenghui Wang, Antoine Isaac, Lourens van der Meij, and Stefan Schlobach. Multi-concept alignment and evaluation. In *Proceedings of the Second International Workshop on Ontology Matching, collocated with the 6th International Semantic Web Conference (ISWC 2007)*, Busan, Korea, November 11 2007.

Appendix 1. Evaluation Form Example

Taak 5

Groot dierenwoordenboek (ook voor kleine dieren)

Informatie over het boek:

PPN: 297491326
3000 Wim Daniëls (1954-)
3011 Jeannette Ensing
4000 Groot dierenwoordenboek (ook voor kleine dieren) / Wim Daniëls ; geèill. door Jeannette Ensing
4030 's-Hertogenbosch : Heinen
4060 152 p
4061 ill
4062 24 cm
4201 Rugtitel: Groot dierenwoordenboek
4601 Woordenboek waarin, naast algemene informatie over verschillende dieren, dierennamen worden uitgelegd. Met zwart-wittekeningen. Vanaf ca. 10 t/ 13 jaar.

Vraag: Zijn de volgende algemene trefwoorden acceptabel als index termen (520X veld) voor dit boek?

— `jeugdboeken ; informatie - biologie`
☐ Acceptabel
als de term niet geheel het boek onderwerp beschrijft,
is de term: ☐ iets te breed ☐ iets te nauw ☐ sterk gerelateerd
☐ Niet acceptabel
als de term toch enigszins gerelateerd is aan het boek onderwerp,
is de term: ☐ te breed, ☐ te nauw, ☐ gerelateerd maar niet voldoende
Eventuele toelichting:

— `biologie ; woordenboeken`
☐ Acceptabel
als de term niet geheel het boek onderwerp beschrijft,
is de term: ☐ iets te breed ☐ iets te nauw ☐ sterk gerelateerd
☐ Niet acceptabel
als de term toch enigszins gerelateerd is aan het boek onderwerp,
is de term: ☐ te breed, ☐ te nauw, ☐ gerelateerd maar niet voldoende
Eventuele toelichting:

— `jeugdboeken ; informatie - taal- en letterkunde`
☐ Acceptabel
als de term niet geheel het boek onderwerp beschrijft,

is de term: ☐ iets te breed ☐ iets te nauw ☐ sterk gerelateerd
☐ Niet acceptabel
als de term toch enigszins gerelateerd is aan het boek onderwerp,
is de term: ☐ te breed, ☐ te nauw, ☐ gerelateerd maar niet voldoende
Eventuele toelichting:

— `dieren`
☐ Acceptabel
als de term niet geheel het boek onderwerp beschrijft,
is de term: ☐ iets te breed ☐ iets te nauw ☐ sterk gerelateerd
☐ Niet acceptabel
als de term toch enigszins gerelateerd is aan het boek onderwerp,
is de term: ☐ te breed, ☐ te nauw, ☐ gerelateerd maar niet voldoende
Eventuele toelichting:

— `liederen`
☐ Acceptabel
als de term niet geheel het boek onderwerp beschrijft,
is de term: ☐ iets te breed ☐ iets te nauw ☐ sterk gerelateerd
☐ Niet acceptabel
als de term toch enigszins gerelateerd is aan het boek onderwerp,
is de term: ☐ te breed, ☐ te nauw, ☐ gerelateerd maar niet voldoende
Eventuele toelichting:

Vraag: Omcirkel Brinkman termen die u zelf gekozen zou hebben.

Vraag: Brinkman algemene termen die ontbreken aan de omschrijving:

Eventueel: Voeg niet algemene Brinkman trefwoorden toe aan de beschrijving:

Commentaar: