Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 4851–4858
Marseille, 11–16 May 2020

Aligning Wikipedia with WordNet: a Review and Evaluation of Different Techniques

Antoni Oliver
Universitat Oberta de Catalunya (UOC)
Barcelona (Catalonia-Spain)
aoliverg@uoc.edu

Abstract

In this paper we explore techniques for aligning Wikipedia articles with WordNet synsets, their successful alignment being our main goal. We evaluate techniques that use the definitions and sense relations in Wordnet and the text and categories in Wikipedia articles. The results we present are based on two evaluation strategies: one uses a new gold and silver standard (for which the creation process is explained); the other creates wordnets in other languages and then compares them with existing wordnets for those languages found in the Open Multilingual Wordnet project. A reliable alignment between WordNet and Wikipedia is a very valuable resource for the creation of new wordnets in other languages and for the development of existing wordnets. The evaluation of alignments between WordNet and lexical resources is a difficult and time-consuming task, but the evaluation strategy using the Open Multilingual Wordnet can be used as an automated evaluation measure to assess the quality of alignments between these two resources.

Keywords: WordNet, Wikipedia, sense alignment

1. Introduction

In this paper we address the task of creating alignments between Princeton WordNet (PWN 3.0) synsets and English Wikipedia articles. In principle the alignment of monosemic words should be relatively easy, as a simple comparison of the word forms would be sufficient. The problem arises with the granularity of senses involved when aligning polysemic words, i.e. words attached to several senses. One given word may be considered monosemic in one of the resources and polysemic in the other resource; or a given word may even be considered monosemic in both resources but refer to different senses in each resource. The task of aligning words that are polysemic in both resources is even more difficult.

Wikipedia[1] is a collaborative multilingual encyclopaedia that uses the wiki format. All Wikipedia's contents can be downloaded as XML dumps containing the code in wiki format. Extracting the content and the associated information from the Wikipedia articles is a difficult task, but several parsers are available. For example, WikiExtractor [2], a Python script that can extract the text and other information from articles and give an XML dump; or JWPL[3] (*Java Wikipedia Library*), an open-source Java-based API that gives access to all the information contained in Wikipedia. From the Mediawiki download page we can also download SQL table dumps containing information from Wikipedia articles such as their titles, categories and interlanguage links.

DBpedia[4] (Lehmann et al., 2015) is a project aiming to extract structured content from the Wikipedia project. This structured information can be downloaded in several database file formats. In our experiments we used DBpedia and the following tables were processed: labels, short_abstracts, long_abstracts, article_categories and interlanguage_links.

The paper is organised as follows: in the next section we present a series of alignments between WordNet and Wikipedia. First we present the existing alignments, and then we present new alignments we have created using different strategies. In the same section we also present the existing gold standards that can be used for evaluation. We follow with the experimental part aiming to evaluate all the presented WordNet-Wikipedia alignments. This section starts with the process of creation of a new gold and silver standards. A novel evaluation method using the Open Multilingual Wordnet is presented in this same section. Then the methodology of creation and evaluation of the new alignments is presented. The paper finishes with a conclusion and future work section.

2. Aligning WordNet and Wikipedia

2.1. Existing alignments

The alignment of lexical resources is an active area of research and a lot of works in this area are available. Gurevych (Gurevych et al., 2016) offers an excellent introduction to this field. Some previous works on the alignment of WordNet and Wikipedia offer open access to the alignment files created. We included these available WordNet-Wikipedia alignments in our evaluations.

- **Similarity-based alignment of WordNet and Wikipedia** (Niemann and Gurevych, 2011): this alignment technique uses the Personalized PageRank combined with a word overlap measure, with the idea of creating a gold standard and using machine learning techniques. As almost all Wikipedia articles refer to nouns, it focuses on this POS. The authors reported a bug in the extraction process, that some Wikipedia page titles had codification errors, and that some symbols were replaced by a question mark.

[1] www.wikipedia.org
[2] https://github.com/attardi/wikiextractor
[3] https://www.ukp.tu-darmstadt.de/software/jwpl
[4] http://wiki.dbpedia.org/

After cleaning these cases, a total of 31,362 pairs were obtained.

- **Graph-based alignment of WordNet and Wikipedia** (Matuschek and Gurevych, 2013): the algorithm used in this technique works in two steps: (a) the initial construction of graphs using several parameters: semantic relations (using hyperlinks between Wikipedia articles) and the so-called linking of monosemous lexemes or monosemous linking t (a lexeme is a combination of lemma and POS found in the gloss of a given sense and also in the gloss of another sense); (b) the alignment is created computing the Dijkstra distance for each candidate alignment and selecting the one with the shortest distance. There are a total of 42,314 pairings in this dataset.
- **Alignment from Babelnet version 2.5** (Navigli and Ponzetto, 2010): the distribution files for Babelnet version 2.5 make it easy to construct its mappings. A total of 47,956 mappings are available.
- **Samuel Fernando mappings** (Fernando and Stevenson, 2012): these mappings are developed in 3 steps: the generation of candidate articles so as to reduce the search space, using title matching and information retrieval; the selection of the best mappings, where both text and title similarities are used; and the refining of the mappings, using an overview of the mappings and information about the link structure in Wikipedia. This set offers a total of 36,677 pairings.

2.2. New alignment techniques

We also used some classic word-sense disambiguation (WSD) algorithms to create alternative alignments. We call these alignments *new* because they are not found in the set of existing alignments, although some of them use well-known techniques. All these algorithms are based on the comparison of the definitions in PWN 3.0 and Wikipedia. We must bear in mind that Wikipedia does not provide proper definitions; therefore we are considering the following elements as definitions:

- *Short abstracts*, as given in DBpedia. These are created using the first sentences of an article, with a maximum length of 500 characters.
- *Long abstracts*, as given in DBpedia. These are usually created using the full text of the first section of the article.
- The *full text* of the article.

One of the major drawbacks in the comparison of WordNet and Wikipedia definitions is the big difference in their lengths. WordNet definitions are usually very concise. For some synsets a set of examples is available, and we are using these examples as a part of the definitions. Wikipedia articles' short abstracts are much longer than PWN 3.0 definitions and examples (over five times longer), and the contrast is of course even greater with the long abstracts and full texts.
We implemented the following algorithms to create alignments that we could then use in evaluations:

- **Most frequent sense** (M.F.S): for each article title in Wikipedia we made a relation between its meaning ID and the most frequent sense given by PWN 3.0. The precision values obtained with this technique can be considered a baseline.
- **Lesk algorithm using lemmata** (Lesk L.): we used the classic Lesk algorithm (Lesk, 1986) counting the lemmata common to both the PWN 3.0 definition and Wikipedia. We only took into account the lemmata of open class words (nouns, verbs, adjectives and adverbs). We avoided counting the lemmata corresponding to the very high frequency verbs *be, have* and *do*. For PWN 3.0 we also took into account the other variants for the same synset.
- **Adapted/Extended Lesk algorithm** (Banerjee and Pedersen, 2002) (A. Lesk L.): we extended the list of lemmata from PWN 3.0 definitions using the lemmata from the definitions in synsets related to the given PWN synset, based on the WordNet relations. The variants of the related synsets were also taken into account. This extension was only made on the PWN 3.0 side.
- **Lesk algorithm using lemmata extended with relations** (Lesk L.R.): we used the Lesk algorithm, using lemmata and counting the related words common to both PWN 3.0 and Wikipedia, for which we used categories as related words.
- **Lesk senses** (Lesk S.): we used an idea similar to the Lesk algorithm but using senses instead of lemmata. We have all the definitions sense-tagged using Freeling and UKB, so we were able to compare the sense tags (that are in fact PWN synsets) in the PWN 3.0 definitions and the Wikipedia definitions. Once again, we avoided counting the very high frequency verbs *be, have* and *do*. The PWN synset is also added in the list of synsets in the PWN definition.
- **Adapted/Extended Lesk Senses** (A. Lesk S.): we combined the ideas behind the Adapted/Extended Lesk algorithm technique and the Lesk senses technique. We extended the list of senses in the PWN definitions by adding the senses from the synsets related to the given PWN synset using the WordNet relations. The related synsets were also taken into account.

2.3. Existing gold standards for evaluation

To our knowledge, there are two gold standards for the evaluation of alignments between Wikipedia and WordNet:

- The well-balanced reference dataset put forward by Niemann and Gurevych (Niemann and Gurevych, 2011). This dataset was created from a sample of 320 noun synsets, yielding 1,815 sense pairs that were manually annotated as correct or incorrect. This provided a total of 227 pairings between WordNet synsets and Wikipedia articles.
- The dataset provided by Wolf and Gurevych (Wolf and Gurevych, 2010), created by randomly sampling 14

nouns that are polysemic in WordNet and for which at least one Wikipedia article was retrieved. This resulted in 297 pairings, which were manually annotated as correct or incorrect, giving a total of 22 correct pairs.

We combined these two resources and obtained a total of 249 pairings between WordNet synsets and Wikipedia articles. In the rest of the paper we will refer to this gold standard as *gold standard P* (P for *previous*).

3. Experimental part

In this section we start by explaining our process for creating a new larger gold standard that can be used to evaluate alignment techniques. Then we explain a new strategy for evaluating alignments. For this new strategy, instead of using a gold standard, we used the alignments we had created to generate wordnets in other languages, and these newly created wordnets were evaluated by comparing them against the wordnets available in Open Multilingual Wordnet[5] (Bond and Kyonghee, 2012). We then offer a detailed explanation of the new alignments' creation process. The section finishes with the presentation of the evaluation results for both the existing alignments and for the newly created Wikipedia-Wordnet alignments. The evaluation is performed in two ways: using our new gold standard and using our Open Multilingual Wordnet comparison strategy.

3.1. Creation of the new gold and silver standard

To create a new gold standard we used all the alignments available. We classified the pairs from these alignments into 4 groups, taking into account how many of the existing alignments contain the given pairing between the PWN synset and the Wikipedia article:

- *4-commons*: pair in all 4 existing alignments.
- *3-commons*: pair in 3 existing alignments.
- *2-commons*: pair in 2 existing alignments.
- *1-commons*: pair in 1 alignment.

We manually reviewed 10% of the pairs in the 4-commons, 3-commons and 2- commons subsets; and 5% of the pairs in the 1 commons subset. As a result, we obtained a gold standard with 3,927 manually reviewed correct pairings between PWN synsets and Wikipedia articles.

The gold standard obtained has 3,927 manually reviewed correct pairs. In the rest of the paper we will refer to this gold standard as *gold standard A*.

We also created a larger set using all the manually revised correct pairs and all the pairs in the 4-commons and 3-commons groups. As the precision for the 3-commons subset is 98.55, we can assign this precision to this larger set. This set has a total of 24,083 pairs. As this set has pairing obtained automatically with no manual revision, it can be considered a silver standard. In the rest of the paper we will refer to this set as *silver standard B*. In table 1 the number of pairs in each gold standard are shown.

[5]http://compling.hss.ntu.edu.sg/omw/

	Pairs
gold standard P	249
gold standard A	3,927
silver standard B	24,083

Table 1: Size of the gold standards.

To test the utility of the gold and silver standards we performed an evaluation task using the three above-mentioned gold and silver standards. We evaluated the existing alignments described in 2.1. along with an alignment created using the most frequent sense. In the latter alignment we related each Wikipedia article title (after normalizing it) with its most frequent sense in WordNet (if present).

Tables 2, 3 and 4 show the precision, recall and F_1 values for each alignment evaluated using the gold standards P, A and B. These results lead us to very interesting conclusions. First, we can observe the importance of which set is used to evaluate the alignment results, as the precision and recall (and consequently F_1) values obtained are very different depending on the gold or silver standard used. Second, the gold standard A and silver standard B agree on which alignment is best, both in terms of precision (Samuel) and in terms of recall and F_1 (Babelnet 2.5). The smaller gold standard P disagrees on which is best in terms of precision (Similarity, but with a very small difference), but agrees on which is best in terms of recall and F_1. If we compare the ability to rank each alignment in terms of precision, we can observe that gold standard A and silver standard B both rank them in the same order (Samuel, Similarity, Babelnet 2.5, Graph and then MFS). If we observe the ranking of F_1, we can observe that all three standards agree on the order (Babelnet 2.5, Samuel, Graph, Similarity and then MFS).

As a conclusion of this limited experiment we can see that the size of the gold or silver standard can have a strong influence on precision, recall and F_1. Upwards of a certain size (for example gold standard A and silver standard B), the standard's ability to rank alignments is not altered.

3.2. Evaluation using Open Multilingual Wordnet

We have come up with a new strategy for evaluating alignments. First we use the alignments to create wordnets in 15 other languages: Catalan (Gonzalez-Agirre et al., 2012), Greek (Stamou et al., 2004), Basque (Pociello et al., 2011), Finnish (Lindén and Carlson., 2010), French (Sagot and Fišer, 2008), Galician (Gonzalez-Agirre et al., 2012), Croatian (Oliver et al., 2015; Raffaelli et al., 2008), Indonesian (Mohamed Noor et al., 2011), Italian (Toral et al., 2010), Japanese (Isahara et al., 2008), Polish (Piasecki et al., 2009), Portuguese (de Paiva and Rademaker, 2012), Slovene (Fišer et al., 2012), Spanish (Gonzalez-Agirre et al., 2012) and Swedish (Borin et al., 2013), for which there are also wordnets already available in the Open Multilingual Wordnet (OMW) (Bond and Kyonghee, 2012). We then evaluate the newly created wordnets by comparing them against the existing OMW wordnets in the other languages. This strategy has the advantage of enabling us to evaluate many more alignments than we could with the gold

	P	R	F_1
Most frequent sense	39.64	53.01	45.36
Similarity	90.00	46.99	61.74
Graph	75.82	55.42	64.04
Samuel	89.51	51.41	65.31
Babelnet 2.5	83.78	62.25	71.43

Table 2: Evaluation using gold standard P.

	P	R	F_1
Most frequent sense	44.49	73.34	55.38
Similarity	83.16	43.65	57.25
Graph	75.04	54.34	63.03
Samuel	93.67	72.32	81.86
Babelnet 2.5	84.56	80.34	82.40

Table 3: Evaluation using gold standard A.

	P	R	F_1
Most frequent sense	50.93	83.60	63.30
Similarity	94.26	61.83	74.68
Graph	89.89	75.97	82.35
Samuel	98.23	86.57	92.03
Babelnet 2.5	91.61	92.59	92.09

Table 4: Evaluation using silver standard B.

standard strategy. As a drawback we must bear in mind that just because a given alignment produces a correct target language variant, it does not necessarily mean that the pairing is correct; consider the following two examples.

```
00053097-n - paired with - Farewell
(Bob Dylan song)
```

This pairing will be evaluated as incorrect, as the Wikipedia interlanguage links would tell us that the Italian variant for this synset is *farewell*[6], but OMW gives the Italian translation *addio, commiato, distacco*. So in this case the evaluation strategy based on OMW would succeed and evaluate this as incorrect.
However, in other cases this strategy fails.

```
00097504-n - paired with - Execution
(computing)
```

This pairing would be evaluated as correct using the French and Italian Wikipedias, as the target language variants derived would be *exécution* and *esecuzione*, respectively. These target language variants would be confirmed by OMW, even though the correct OMW synset for the computing sense of *execution* is 13477462-n (not 00097504-n). As we use several languages for the evaluation this drawback can be minimized. For each language we get a precision value and we get the average precision. We are not able to calculate a real recall value, as we do not know the full number of pairs existing between WordNet and Wikipedia. We calculated a simulated recall value considering that all the noun synsets in WordNet should have a valid alignment to Wikipedia (that is, there should be 82,115 valid pairs). This is of course not true, but having a recall value is useful for comparisons between alignment techniques. We can calculate an average recall value using the 15 languages. With the average values we can calculate the F_1 value.

[6]The corresponding title in the Italian Wikipedia is *Farewell (Bob Dylan)*, but the information between the brackets is ignored

3.3. Creation of the new alignments

3.3.1. Database

To process all the data from WordNet and Wikipedia and to perform our evaluations of the alignments, we created a SQLite database with the following tables:

- *entry*: containing the meaning identifier (the offset-pos for WordNet and the DBpedia identifier for Wikipedia), the word, the part-of-speech, the language and the source (the name of the lexical resource). Each Wikipedia article title is used as a word, and the titles are always capitalized and sometimes includes extra information in brackets, so in the database we included three fields for the name: *word* (as it appears), *word_n* (deleting the information in brackets) and *word_nmin* (deleting the information in brackets and converting the full word to lower case).
- *definition*: the definition corresponding to a given meaning ID. For WordNet we also included the examples. For Wikipedia, as already mentioned, we experimented with three elements considered as definitions: the short abstract; the long abstract and the full article.
- *tagged_definition*: the POS and sense tagged definition. To tag the definitions we used the Python API of the Freeling 4.0 analyzer (Padró and Stanilovsky, 2012). This analyzer uses the PWN 3.0 synsets to sense tag the texts, and the UKB algorithm (Padró et al., 2010) to disambiguate them. For Wikipedia articles we only tagged the short abstracts, longs abstracts and full texts for articles where the title (after normalization and lower casing) matched the lower cased variants in PWN-3.0.
- *relation*: for WordNet this table stores meaning IDs for words that have some form of semantic relation. For Wikipedia no such relations can be extracted, so we used the set of categories associated with any given article as semantic relations.
- *translation*: this table stores the translations for each meaning ID given by the lexical resources. For Wikipedia, we used the translations given by the Open Multilingual Wordnet project[7] (Bond and Kyonghee, 2012).

3.3.2. Alignment algorithms

All the algorithms we developed use the database to retrieve the information they need. The algorithms therefore give a list of alignments formed by: the PWN synset; the Wikipedia article title; a score for the best candidate indicating the number of elements matching in PWN and the

[7]http://compling.hss.ntu.edu.sg/omw

Wikipedia definition (depending on the strategy, these elements may include lemmata, senses, related words, etc.); and the difference in this score between the first candidate and the second candidate. Here is an example, obtained using short abstracts and the Lesk Lemmata technique:

```
00023773-n - paired with -Motivation 10 8
```

This means that PWN 3.0 synset *00023773-n* is aligned with the Wikipedia article titled *Motivation* and they have 10 lemmata in common, whereas the second candidate (not shown in the results) has only 2 lemmata in common.

3.3.3. Evaluation of the alignment techniques

Table 5 gives the full results from our gold standard A evaluations made on both the existing alignments and the newly created alignments. Regarding precision, note that all the results are above the Most Frequent Sense results (considered the baseline). The tables shows the precision, recall and F_1 values, as well as the number of pairings. Regarding the existing alignments, the best precision value is obtained in the Samuel alignment (93.21%), whereas the best F_1 value is obtained with the Babelnet 2.5 alignments (82.37). Regarding the newly created alignments, the best precision is obtained with the Lesk Senses technique using the short abstracts with a minimum score of 4 and a minimum difference of 2 (90.87%). The best F_1 value is obtained with the Adapted Lesk Lemmata alignment using articles' full texts and with a minimum score of 2 and a minimum difference of 1 (81.86). If we compare the existing alignments with the newly created alignments using this evaluation methodology, best results are obtained with the existing alignments.

Table 6 shows the full results from the evaluations made using the Open Multilingual Wordnet strategy. Regarding the existing alignments, the best precision value (76.53%) and the best F_1 value (18.42) are obtained from the Babelnet 2.5 alignment. Regarding the newly created alignments, the best precision is obtained with the Lesk Senses alignment using the short abstracts with a minimum score of 4 and a minimum difference of 2 (79.32%). The best F_1 value is obtained with the Adapted Lesk Lemmata alignment using the full article text and with a minimum score of 2 and a minimum difference of 1 (21.97). If we compare the existing alignments with the newly created alignments, we can observe that both the best precision and the best F_1 values are obtained with two of the newly created alignments. It is worth noting that the recall and F_1 values from the gold and silver standards and OMW evaluations' results are not comparable, as they were calculated using different bases. Note that the precision of one of the existing alignments (Graph-based alignment) is lower than that of the Most Frequent Sense alignment (the baseline).

4. Conclusions and future work

In this paper we have presented evaluations of some existing alignments between WordNet and Wikipedia and of some newly created alignments. The evaluations were performed using two strategies: one using a newly created gold standard, and the other using OMW wordnets for other languages. These evaluation strategies give different results; the alignment considered best is different depending on the strategy used. The strategy based on the gold and silver standards has the advantage that precision and recall values (thus also F_1 values) can be obtained. The evaluation based on the Open Multilingual Wordnet has the advantage that it uses much more evidence, but in some cases can mistake an incorrect pairing for one that is correct.

As a conclusion, we can point out that the newly created alignments give similar results in comparison with the existing alignments. Another important conclusion is that the evaluation of alignments between WordNet and Wikipedia is still an open research task. The newly created gold and silver standards are the largest to our knowledge, but a larger one would be desirable. These gold and silver standards can be freely downloaded[8]. We think that the new evaluation strategy using OWM wordnets can be useful for future research, as the wordnets in OMW are steadily increasing, both in the number of available target language wordnets and in the size of the wordnets available.

In future research we plan to use Wikipedia redirection information. We also plan to use a similar approach for the alignment of WordNet with other lexical resources, such as Wiktionary and Omegawiki.

Acknowledgments

This research has been carried thanks to the project deepReading RTI2018-096846-B-C21 (MCIU/ AEI/ FEDER, UE) .

5. Bibliographical References

Banerjee, S. and Pedersen, T. (2002). An adapted lesk algorithm for word sense disambiguation using wordnet. In *Computational linguistics and intelligent text processing*, pages 136–145. Springer.

Bond, F. and Kyonghee, P. (2012). A survey of wordnets and their licenses. In *Proceedings of the 6th International Global WordNet Conference*, pages 64–71, Matsue, Japan.

Borin, L., Forsberg, M., and Lönngren, L. (2013). Saldo: a touch of yin to wordnet's yang. *Language Resources and Evaluation*, 47(4):1191–1211.

de Paiva, V. and Rademaker, A. (2012). Revisiting a Brazilian wordnet. In *Proceedings of the 6th Global WordNet Conference (GWC 2012)*, Matsue.

Fernando, S. and Stevenson, M. (2012). Mapping wordnet synsets to wikipedia articles. In *LREC*, pages 590–596.

Fišer, D., Novak, J., and Tomaž. (2012). sloWNet 3.0: development, extension and cleaning. In *Proceedings of 6th International Global Wordnet Conference (GWC 2012)*, pages 113–117. The Global WordNet Association.

Gonzalez-Agirre, A., Laparra, E., and Rigau, G. (2012). Multilingual central repository version 3.0: upgrading a very large lexical knowledge base. In *Proceedings of the 6th Global WordNet Conference (GWC 2012)*, Matsue.

Gurevych, I., Eckle-Kohler, J., and Matuschek, M. (2016). *Linked Lexical Knowledge Bases: Foundations and Applications*. Morgan & Claypool Publishers.

[8]https://sourceforge.net/projects/pwnalign

Isahara, H., Bond, F., Uchimoto, K., Utiyama, M., and Kanzaki, K. (2008). Development of the Japanese WordNet. In *Sixth International conference on Language Resources and Evaluation (LREC 2008)*, Marrakech.

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., et al. (2015). Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web*, 6(2):167–195.

Lesk, M. (1986). Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *Proceedings of the 5th annual international conference on Systems documentation*, pages 24–26. ACM.

Lindén, K. and Carlson., L. (2010). Finnwordnet — wordnet påfinska via översättning. *LexicoNordica — Nordic Journal of Lexicography*, 17:119–140. In Swedish with an English abstract.

Matuschek, M. and Gurevych, I. (2013). Dijkstra-wsa: A graph-based approach to word sense alignment. *Transactions of the Association for Computational Linguistics*, 1:151–164.

Mohamed Noor, N., Sapuan, S., and Bond, F. (2011). Creating the open Wordnet Bahasa. In *Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation (PACLIC 25)*, pages 258–267, Singapore.

Navigli, R. and Ponzetto, S. P. (2010). Babelnet: Building a very large multilingual semantic network. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 216–225. Association for Computational Linguistics.

Niemann, E. and Gurevych, I. (2011). The people's web meets linguistic knowledge: automatic sense alignment of wikipedia and wordnet. In *Proceedings of the Ninth International Conference on Computational Semantics*, pages 205–214. Association for Computational Linguistics.

Oliver, A., Šojat, K., and Srebačić, M. (2015). Automatic expansion of croatian wordnet. In *In Proceedings of the 29th CALS international conference âœApplied Linguistic Research and Methodologyâœ*, Zadar (Croatia).

Padró, L. and Stanilovsky, E. (2012). Freeling 3.0: Towards wider multilinguality. In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, Istanbul, Turkey, May. ELRA.

Padró, L., Reese, S., Agirre, E., and Soroa, A. (2010). Semantic services in freeling 2.1: Wordnet and UKB. In *Proceedings of the 5th International Conference of the Global WordNet Association (GWC-2010)*.

Piasecki, M., Szpakowicz, S., and Broda, B. (2009). *A Wordnet from the Ground Up*. Wroclaw University of Technology Press. (ISBN 978-83-7493-476-3).

Pociello, E., Agirre, E., and Aldezabal, I. (2011). Methodology and construction of the Basque wordnet. *Language Resources and Evaluation*, 45(2):121–142.

Raffaelli, I., Bekavac, B., Agić, Z., and Tadić, M. (2008). Building croatian wordnet. In Attila TanÃ¡cs, et al., editors, *Proceedings of the Fourth Global WordNet Conference 2008*, pages 349–359, Szeged.

Sagot, B. and Fišer, D. (2008). Building a free French wordnet from multilingual resources. In European Language Resources Association (ELRA), editor, *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco.

Stamou, S., Nenadic, G., and Christodoulakis, D. (2004). Exploring Balkanet shared ontology for multilingual conceptual indexing. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, page 781â"784, Lisbon.

Toral, A., Bracale, S., Monachini, M., and Soria, C. (2010). Rejuvenating the italian wordnet: upgrading, standardising, extending. In *Proceedings of the 5th International Conference of the Global WordNet Association (GWC-2010)*, Mumbai.

Wolf, E. and Gurevych, I. (2010). Aligning sense inventories in wikipedia and wordnet. In *Proceedings of the 1st Workshop on Automated Knowledge Base Construction, Grenoble, France*, pages 24–28. Citeseer.

	P	**R**	$\mathbf{F_1}$	**Pairings**
EXISTING ALIGNMENTS				
Similarity	82.92	43.52	57.08	31,362
Graph	74.65	54.06	62.71	42,314
Samuel	**93.21**	71.96	81.22	36,677
Babelnet 2.5	84.53	80.32	**82.37**	47,956
Most Frequent Sense	37.70	77.13	50.71	122,720
SHORT ABSTRACTS min score=2 min diff=1				
Lesk Lemmata	84.14	63.51	72.38	35,002
Adapted Lesk Lemmata	80.90	79.48	80.18	47,839
Lesk Lemmata Relations	83.98	64.86	73.19	36,120
Lesk Senses	81.58	67.89	74.11	38,310
Adapted Lesk Senses	78.01	79.76	78.87	49,793
SHORT ABSTRACTS min score=4 min diff=2				
Lesk Lemmata	88.77	23.35	36.98	10,582
Adapted Lesk Lemmata	86.45	52.48	65.31	26,350
Lesk Lemmata Relations	87.88	25.11	39.06	11,495
Lesk Senses	**90.87**	24.34	38.40	11,390
Adapted Lesk Senses	86.47	53.07	65.77	26,860
LONG ABSTRACTS min score=2 min diff=1				
Lesk Lemmata	83.31	68.37	75.10	38,728
Adapted Lesk Lemmata	79.85	82.33	81.07	51,225
Lesk Lemmata Relations	83.10	69.62	75.77	39,721
Lesk Senses	83.55	60.81	70.39	33,756
Adapted Lesk Senses	77.94	82.68	80.24	52,734
LONG ABSTRACTS min score=4 min diff=2				
Lesk Lemmata	87.46	30.91	45.68	14,536
Adapted Lesk Lemmata	85.28	61.09	71.19	31,812
Lesk Lemmata Relations	86.89	32.75	47.57	15,461
Lesk Senses	89.89	21.29	34.42	10,013
Adapted Lesk Senses	85.50	60.68	70.99	31,515
FULL ARTICLE min score=2 min diff=1				
Lesk Lemmata	81.04	77.62	79.29	47,790
Adapted Lesk Lemmata	78.05	86.05	**81.86**	58,761
Lesk Lemmata Relations	80.89	78.13	79.48	48,355
Lesk Senses	79.93	77.77	78.83	48,505
Adapted Lesk Senses	76.36	85.71	80.77	59,270
FULL ARTICLE min score=4 min diff=2				
Lesk Lemmata	85.16	50.42	63.34	25,860
Adapted Lesk Lemmata	81.90	77.87	79.83	45,352
Lesk Lemmata Relations	85.07	51.67	64.29	26,667
Lesk Senses	86.64	47.06	60.99	23,277
Adapted Lesk Senses	82.78	75.363	79.04	42,935

Table 5: Evaluation figures using the gold standard A strategy.

	P	**R**	**F$_1$**	**Pairings**
EXISTING ALIGNMENTS				
Similarity	66.56	5.18	9.62	31,362
Graph	51.23	6.90	12.16	42,314
Samuel	76.26	9.63	17.10	36,677
Babelnet 2.5	**76.53**	10.47	**18.42**	47,956
Most Frequent Sense	60.34	11.56	19.41	122,720
SHORT ABSTRACTS min score=2 min diff=1				
Lesk Lemmata	74.42	9.14	16.28	35,002
Adapted Lesk Lemmata	71.80	11.44	19.73	47,839
Lesk Lemmata Relations	73.30	9.48	16.79	36,120
Lesk Senses	74.96	9.53	16.92	38,310
Adapted Lesk Senses	72.07	11.31	19.55	49,793
SHORT ABSTRACTS min score=4 min diff=2				
Lesk Lemmata	77.90	3.31	6.35	10,582
Adapted Lesk Lemmata	75.30	7.79	14.11	26,350
Lesk Lemmata Relations	77.15	3.67	7.01	11,495
Lesk Senses	**79.32**	3.62	6.92	11,390
Adapted Lesk Senses	76.18	7.81	14.16	26,860
LONG ABSTRACTS min score=2 min diff=1				
Lesk Lemmata	73.41	10.14	17.81	38,728
Adapted Lesk Lemmata	70.55	12.07	20.62	51,225
Lesk Lemmata Relations	72.44	10.42	18.22	39,721
Lesk Senses	74.74	8.83	15.79	33,756
Adapted Lesk Senses	70.88	11.92	20.41	52,734
LONG ABSTRACTS min score=4 min diff=2				
Lesk Lemmata	77.42	4.76	8.96	14,536
Adapted Lesk Lemmata	74.00	9.31	16.54	31,812
Lesk Lemmata Relations	76.57	5.09	9.55	15,461
Lesk Senses	78.65	3.27	6.27	10,013
Adapted Lesk Senses	74.86	9.14	16.29	31,515
FULL ARTICLE min score=2 min diff=1				
Lesk Lemmata	69.02	12.01	20.45	47,790
Adapted Lesk Lemmata	66.04	13.18	**21.97**	58,761
Lesk Lemmata Relations	68.61	12.13	20.61	48,355
Lesk Senses	69.94	11.69	20.03	48,505
Adapted Lesk Senses	66.69	12.98	21.73	59,270
FULL ARTICLE min score=4 min diff=2				
Lesk Lemmata	72.84	8.19	14.72	25,860
Adapted Lesk Lemmata	68.82	12.16	20.67	45,352
Lesk Lemmata Relations	72.44	8.44	15.12	26,667
Lesk Senses	75.36	7.45	13.57	23,277
Adapted Lesk Senses	70.21	11.72	20.08	42,935

Table 6: Evaluation figures using the Open Multilingual Wordnet strategy.