



# A Semantic Similarity Measurement Tool for WordNet-Like Databases

Marek Kubis<sup>(✉)</sup>

Faculty of Mathematics and Computer Science,  
Department of Computer Linguistics and Artificial Intelligence,  
Adam Mickiewicz University, ul. Umultowska 87, 61-614 Poznań, Poland  
mkubis@amu.edu.pl

**Abstract.** The paper describes a new framework for computing the semantic similarity of words and concepts using WordNet-like databases. The main advantage of the presented approach is the ability to implement similarity measures as concise expressions in the embedded query language. The preliminary results of the use of the framework to model the semantic similarity of Polish nouns are reported.

## 1 Introduction

Among various applications of WordNet [4], the task of modeling semantic similarity between words has attracted considerable attention over the last two decades. WordNet-based semantic similarity measures, ranging from simple path-length dependent functions [14, 26] and measures that exploit the notion of the least common subsumer<sup>1</sup> [36] to those that utilize information content computed over corpora [10, 16, 27], have been proposed in the literature. These measures have been evaluated within the task of word sense disambiguation [21] and incorporated into natural language processing and information extraction systems [2, 31]. Despite a wide range of applications, the issue of using other wordnets in place of Princeton WordNet as resources for modeling similarity among words appears not to have gained the same level of attention. Our aim is to use PolNet [35] and PIWordNet [17] to model the semantic similarity of Polish nouns. Since we have not found a software package for measuring semantic similarity that could be easily adapted to make use of both Polish wordnets (cf. Sect. 2), we decided to implement our own. Therefore, the goal of this paper is twofold. First, we present WSim: a new tool for determining degrees of semantic similarity using measures computed over WordNet-like databases.<sup>2</sup> Second, we report the preliminary results of the use of WordNet-based similarity measures to model the similarity of Polish nouns. This is, to the best of our knowledge,

---

<sup>1</sup> A joint transitive hypernym of two synsets such that no other joint transitive hypernym of these synsets is placed below it within the hypernymy hierarchy.

<sup>2</sup> Databases that are organized similarly to WordNet [4], called wordnets in the rest of the paper.

the first attempt to apply two wordnets developed for the same language in a shared application-oriented task.

The paper is a revised version of [13]. It presents new, unpublished results of supervised models of semantic similarity built on the values of wordnet-based measures (cf. Sect. 6). Furthermore, it reproduces the experiments from the original paper using a Polish Wikipedia dump from November 20, 2017 instead of the February 6, 2014 dump used previously. Lastly, more detailed results on measuring semantic similarity between English counterparts of Polish nouns are provided (cf. Tables 5 and 9).

## 2 Related Work

WordNet::Similarity [23] is a widely-cited software package that implements a range of WordNet-based semantic similarity measures. This package has become a de facto standard tool for computing similarity scores using WordNet and serves as a reference point for other implementations (e.g., [24]). Unfortunately, WordNet::Similarity operates only on Princeton WordNet and is not able to load wordnets that do not conform to the internal storage format of the wn program distributed with Princeton WordNet [32]. The same restriction applies to the Python interface to WordNet provided by the NLTK toolkit [1]. In addition to Princeton WordNet, the Java reimplementations of WordNet::Similarity by Shima [29], called WS4J, can load the Japanese WordNet [9]. PolNet is not distributed in the Princeton WordNet conformant form and we have not found any tool that could be used to convert it to this format without a vast amount of preprocessing.

A major advance in terms of interoperability is the WordnetTools library [24], which can load any wordnet that is stored in a file conforming to the Wordnet-LMF format [30]. However, at the time of writing, neither PolNet nor PIWordNet had been released in this format. WordnetTools also accepts files in the Global WordNet Grid format [6], but we were unable to load into it the DEBVisDic [8] conformant XML file, which is part of the PIWordNet distribution.

Since exact replication of results using different software packages is not easy to achieve (see [24, sect. 5.3]), we did not want to use separate tools for computing values of similarity measures for the two wordnets (e.g. NLTK for PIWordNet and WordnetTools for PolNet). Therefore, we decided to reimplement WordNet-based semantic similarity measures on top of the WQuery suite [11, 12], which is able to load both PIWordNet and PolNet. An additional advantage of this approach is the ability to modify the similarity measures by revising the concise expressions of the WQuery language (cf. Sect. 4) instead of the Java code of WordnetTools, which, in the case of any changes, would require recompilation. Furthermore, since WQuery (version 0.10) can load wordnets stored in Wordnet-LMF, DEBVisDic [8], and the Princeton WordNet internal format,<sup>3</sup> we acquired the ability to make direct comparisons between the values of similarity measures computed for the lexical databases stored in all of the aforementioned formats.

<sup>3</sup> Through the JWI library [5].

### 3 WSim

As mentioned in the previous section, WSim is built on the WQuery suite. Therefore, before computing the values of similarity measures, the wordnet must be converted into the WQuery database format using the `wcompile` command<sup>4</sup> from the WQuery toolkit. Since both PIWordNet and PolNet are available in the XML files compatible with the DEBVisDic editor [8], the `-t deb` option must be passed to the command

```
wcompile -t deb polnet.xml > polnet.wq
```

With the wordnet in the WQuery format, the similarity of pairs of words (or word senses) can be computed by passing them to the standard input of the `wsim` command, separated by tab characters.

```
wsim polnet.wq < pairs
```

By default, `wsim` determines the similarity of a pair of words by inverting the value of the shortest path length in the hypernymy hierarchy linking the synsets containing the given words; thus for the pair *samochód* (Eng. car) and *rower* (Eng. bicycle) the similarity determined with PolNet is

0.25

WSim implements six semantic similarity measures:

1. inverted length of the shortest path,
2. Wu-Palmer [36],
3. Resnik [27],
4. Jiang-Conrath [10],
5. Leacock-Chodorow [14],
6. Lin [16].

Following [23], we denote these measures by *path*, *wup*, *res*, *jcn*, *lch*, and *lin*, respectively. They can be selected by passing the `-m` option to the `wsim` command. For instance, to compute the Wu-Palmer measure, the command

```
wsim polnet.wq -m wup < pairs
```

must be executed. In the case of information content dependent measures [10, 16, 27] word (or sense) counts can be submitted in a file passed as an argument of the `-c` option, e.g.

```
wsim polnet.wq -m res -c counts < pairs
```

If the counts are distributed along with a wordnet (as is true in the case of Princeton WordNet) the `-c` option can be skipped.

```
wsim wordnet.wq -m res < pairs
```

<sup>4</sup> We assume in the following examples that all commands are invoked in the Linux shell environment.

## 4 Implementation of Measures

The similarity measures are implemented in WSim as functions formulated in the WQuery language [11]. Every function that ends with the `_measure` suffix is interpreted as a similarity measure and is available through the `-m` option of the `wsim` command. For every pair of senses read from the input, the `wsim` command determines their corresponding synsets and passes them to the function indicated by the argument of the `-m` option. In the case of pairs of words, `wsim` returns the maximum of the similarity values computed for every pair of senses of the submitted words.

Let us consider the Wu-Palmer measure as an example. The measure is given by the following formula (cf. [2,36]):

$$\frac{2 * dep(lcs(l, r))}{dist(l, lcs(l, r)) + dist(r, lcs(l, r)) + 2 * dep(lcs(l, r))}$$

where  $l$  and  $r$  are synsets,  $lcs(l, r)$  denotes the least common subsumer of  $l$  and  $r$ ,  $dist$  denotes the distance between two synsets in the hypernymy hierarchy, and  $dep$  returns the distance of a synset from the hypernymy root. The Wu-Palmer measure has the following implementation in WQuery:

```
function wup_measure do
  %l, %r := %A
  %lcs := lcs_by_depth(%l, %r)
  %dl := lcs_dist(%l, %lcs)
  %dr := lcs_dist(%r, %lcs)
  %dlcs := root_dist(%lcs)
  emit 2*%dlcs/(%dl + %dr + 2*%dlcs)
end
```

We will not be discussing WQuery in detail.<sup>5</sup> In order to follow the examples it is enough to understand that arithmetic expressions, variable assignments (`:=`), and function calls (`f(...)`) are interpreted in a manner similar to that of scripting languages such as Python. The arguments are passed to a function in the `%A` variable and return values are passed using the `emit` statement. The main advantage of using WQuery in place of a generic scripting language to implement similarity measures is the ability to use regular expressions over the semantic relation names to denote paths in the wordnet graph. In the case of `wup_measure` the sub-function `lcs_dist` that computes the distance from a synset to its least common subsumer determines the paths from a synset `%s` to its subsumer `%lcs` via the regular expression

$$\%s.\text{hypernym}^*.\%lcs$$

that traverses zero or more times through the `hypernym` relation from the synset `%s` to its subsumer `%lcs`. The `root_dist` function that computes the distance

<sup>5</sup> Interested readers can consult [11].

from a synset to the hypernymy root uses the expression

```
%A.hypernym*[empty(hypernym)]
```

to denote the paths from a synset %A through zero or more hypernymy links to the synsets that do not have hypernyms.<sup>6</sup> We present the complete code implementing these functions below.

```
function lcs_dist do
  %s, %lcs := %A
  emit min_size(%s.hypernym*.%lcs) - 1
end
```

```
function root_dist do
  emit min_size(
    %A.hypernym*[empty(hypernym)]) + 1
end
```

```
function min_size do
  emit distinct(min(size(%A)))
end
```

The `lcs_by_depth` function, which is also called by `wup_measure`, is a built-in function of WQuery that determines the least common subsumers of synsets.

The similarity functions are loaded into WSim at the beginning of execution from a designated directory. Thus, given a correspondence between arguments of `wsim` and function names and the ability to address arbitrary paths in the wordnet graph using the WQuery language, the user can easily experiment with definitions of new measures. For instance, the user can consider a meronymy-based variant of the *path* measure by providing the following function to `wsim`:

```
function mpath_measure do
  %l, %r := %A
  %mpaths := %l.meronym*.^meronym*.%r
  emit 1/min_size(%mpaths)
end
```

## 5 Semantic Similarity Computation Using Polish Wordnets

Given a tool that accepts lexical databases stored in the DEBVisDic editor compatible format, we can compute the values of similarity measures for both Polish wordnets and compare them to the human similarity ratings. In the case of English, the Rubenstein and Goodenough dataset of 65 human-rated noun pairs [28] and its 30-pair subset from Miller and Charles [19] are often used

<sup>6</sup> The synsets satisfying the condition `empty(hypernym)`.

**Table 1.** Correlation coefficients between the human ratings and similarity measure scores determined for PL39 word pairs that occur in both Polish wordnets.

Measure	Pearson’s		Spearman’s	
	PIWN	PolNet	PIWN	PolNet
path	0.6051	0.6421	0.4658	0.6530
wup	0.6322	0.6835	0.6079	0.6902
lch	0.5981	0.6865	0.4658	0.6530
res	0.6026	0.6369	0.6389	0.6539
jcn	0.5358	0.4938	0.6148	0.6700
lin	<b>0.6584</b>	<b>0.7081</b>	<b>0.6520</b>	<b>0.7029</b>

for the purpose of evaluating similarity measures (e.g., [2, 22, 27]). Paliwoda-Pekosz and Lula [20], who translated this dataset into Polish and had it rated, also report the performance of several similarity measures on 39 pairs of the translated nouns covered by version 0.95 of PIWordNet. We refer hereafter to this dataset as PL39 and to the Rubenstein and Goodenough dataset as RG65. For the purpose of our analysis we use version 2.2 of PIWordNet [17] and version 3.0 of PolNet [35]. Furthermore, in order to determine the values of measures that utilize information content (i.e. Resnik, Jiang-Conrath, and Lin), we use word frequencies derived from Polish Wikipedia.

PIWordNet 2.2 and PolNet 3.0 cover 38 and 26 pairs of nouns from the PL39 dataset, respectively. The correlation coefficients between the values of the similarity measures and the human rating of 26 noun pairs common to both wordnets are given in Table 1. It can be seen that, regardless of the correlation type, the Lin measure performs best. The same measure achieves the best results in the case of all 38 word pairs covered by PIWordNet (cf. Table 2). We report the pairs of words from PL39 and the corresponding values of the Lin measure in Table 3.<sup>7</sup> For the purpose of comparison we also computed the correlation coefficients between the human ratings of the RG65 word pairs and similarity measure scores determined using version 3.0 of WordNet. The results obtained for 26 word pairs from RG65, which are English counterparts of PL39 word pairs common to both Polish wordnets, are given in columns 2 and 3 of Table 5. Columns 4 and 5 present the results for 38 pairs of words from RG65, which are counterparts of all PL39 word pairs that occur in PIWordNet. In the case of WordNet, the Leacock-Chodorow measure results in the highest Pearson’s correlation and the Jiang-Conrath and path measures achieve the highest value of Spearman’s correlation coefficient among the analyzed similarity functions.

Given the correlation coefficients for a fixed measure and the same corpus<sup>8</sup> it is tempting to compare the differences between the two wordnets with respect

<sup>7</sup> The pair *środek dnia/południe* is omitted in Table 3, since *środek dnia* occurs in neither PIWordNet 2.2 nor in PolNet 3.0.

<sup>8</sup> In the case of information content-based measures.

**Table 2.** Correlation coefficients between the human ratings and similarity measure scores determined for all PL39 word pairs that occur in PIWordNet.

Measure	Pearson's	Spearman's
path	0.5915	0.5537
wup	0.6896	0.6738
lch	0.6423	0.5537
res	0.6780	0.6866
jcn	0.4419	0.6544
lin	<b>0.7069</b>	<b>0.6941</b>

to the results on the same dataset. However, it must be noted that although the correlation coefficients between human ratings for the 26 nouns from PL39 and measure values induced from PolNet are generally higher<sup>9</sup> than the corresponding coefficients derived for PIWordNet, the results are difficult to interpret due to size of the dataset size and are not significant at the  $\alpha = 0.05$  level according to the Meng, Rosenthal, and Rubin's z-test as implemented by Diedenhofen [3].

## 6 Supervised Similarity Models

Given the values of semantic similarity measures computed for PL39 word pairs, we decided to determine whether supervised models of similarity can be built on the measured values. We developed a range of regression models using the wordnet-based similarity measures as explanatory variables and the similarity score from PL39 as the response variable. The methods of regression we considered are: linear regression (lr), neural networks (nn), regression trees (rt), random forests (rf), and  $\epsilon$ -support vector regression (svr). We used R environment [25] with `stats`, `nnet` [34], `rpart` [33], `randomForest` [15] and `e1071` [18] packages to develop and evaluate the regression models. For the neural network architecture, we chose a multilayer perceptron with one hidden layer and performed a grid search with 5-fold cross-validation on the training set to determine the number of neurons in the hidden layer. In the case of random forests, we performed a grid search with 5-fold cross-validation to determine the number of trees and the minimum size of the terminal nodes. For the support vector regression we examined linear, polynomial,<sup>10</sup> radial basis, and sigmoid kernels and performed a grid search for values of the  $C$ ,  $\gamma$ , and  $\epsilon$  parameters (Table 4).

The models were evaluated using the leave-one-out cross-validation technique (i.e. the similarity of a given pair of words is predicted using the model trained on the similarity scores measured for the other pairs). Table 6 presents the correlation coefficients between the human ratings of the PL39 dataset word pairs and

<sup>9</sup> With the exception of Pearson's correlation coefficient for the Jiang-Conrath measure.

<sup>10</sup> Polynomial kernels of degrees 2 and 3 were considered.

**Table 3.** The values of similarity measures determined for the PL39 word pairs.

Words		lin		rf	
		PIWN	PolNet	PIWN	PolNet
południe	sznurek	0.0000	0.0000	1.0624	1.1758
owoc	piec	0.2268	0.3524	1.4508	1.2489
autograf	wybrzeże	0.0000		0.7924	
auto	czarodziej	0.0000		0.6829	
kopiec	kuchenka	0.1763	0.3671	0.3007	1.4083
azyl	owoc	0.0000	0.0000	1.2264	1.1191
azyl	zakonnik	0.0000	0.0000	0.7361	0.6371
chłopiec	kogut	0.4860	0.6430	1.5495	1.9392
poduszka	klejnot	0.2825		1.8769	
zakonnik	niewolnik	0.7039	0.3311	1.9370	1.0559
azyl	cmentarz	0.4954	0.2848	1.3131	1.5667
wybrzeże	las	0.6917	0.7449	2.1272	2.3904
chłopiec	mędrzec	0.5167		1.0175	
auto	poduszka	0.3290		1.6236	
kopiec	wybrzeże	0.0000	0.0000	0.9607	0.6992
chłopak	czarodziej	0.2250		1.4009	
las	cmentarz	0.0000	0.2892	0.8224	1.1178
jedzenie	kogut	0.7364	0.3480	2.1898	0.5905
wybrzeże	pagórek	0.6107	0.6395	1.6843	1.7047
piec	narzędzie	0.4342	0.5911	1.1178	1.7462
żuraw	kogut	0.6008		2.2085	
cmentarz	kopiec	0.0000	0.0000	0.7922	0.3618
szkło	klejnot	0.3005	0.6592	0.5466	1.1403
żuraw	przyrząd	0.3890		1.5338	
brat	chłopak	0.8419	0.6843	2.9803	1.7356
mędrzec	czarodziej	0.5862		2.2495	
ptak	żuraw	0.7493		2.9402	
ptak	kogut	0.7520	0.7521	3.1408	1.4753
jedzenie	owoc	0.2761	0.8863	0.8367	3.2757
brat	zakonnik	1.0000	1.0000	3.4973	3.2666
piec	kuchenka	0.5393	0.3850	1.8863	0.5378
pagórek	kopiec	1.0000	1.0000	3.5480	3.4066
przewód	sznurek	0.0000	0.5194	0.9194	1.3328
szkło	szklaneczka	0.6077		2.1330	
autograf	podpis	0.9018		3.1604	
narzędzie	przyrząd	0.9794	1.0000	2.6707	2.7127
chłopiec	chłopak	1.0000	1.0000	3.5633	3.1999
auto	samochód	1.0000	0.8629	3.0764	2.9260



**Table 4.** Correlation coefficients between PIWordNet- and PolNet-based measures.

Measure	Pearson's	Spearman's
path	<b>0.8503</b>	0.7344
wup	0.8450	<b>0.8544</b>
lch	0.8369	0.7344
res	0.7682	0.7258
jcn	0.7045	0.7895
lin	0.8015	0.7902

**Table 5.** Correlation coefficients between the human ratings and similarity measure scores determined for the RG65 word pairs which are counterparts of the PL39 pairs.

Measure	26 pairs		38 pairs	
	Pearson's	Spearman's	Pearson's	Spearman's
path	0.7274	0.6351	0.7300	<b>0.6911</b>
wup	0.6795	0.5785	0.7260	0.6749
lch	<b>0.7373</b>	0.6243	<b>0.7678</b>	0.6826
res	0.6598	0.5903	0.7033	0.6521
jcn	0.4310	<b>0.6610</b>	0.3642	0.4315
lin	0.6773	0.5837	0.5652	0.4041

**Table 6.** Correlation coefficients between the human ratings and supervised model scores determined for PL39 word pairs that occur in both Polish wordnets.

Measure	Pearson's		Spearman's	
	PIWN	PolNet	PIWN	PolNet
lin	0.6584	<b>0.7081</b>	<b>0.6520</b>	<b>0.7029</b>
lr	0.5863	0.5779	0.5348	0.4835
nn	0.4833	0.5719	0.3484	0.5232
rt	0.5839	0.6643	0.3546	0.1228
rf	<b>0.6634</b>	0.5535	0.5413	0.3919
svr	0.6411	0.6112	0.5423	0.5574

the similarity scores determined by the supervised models built on the values of wordnet-based similarity measures. For this experiment, we restricted the dataset to 26 word pairs from PL39 that occur in both Polish wordnets. Table 7 reports the results obtained for models built on 38 noun pairs from PL39 that occur in PIWordNet. It can be seen that in both settings the Lin measure outperforms the supervised models, with the sole exception of the random forest model built from the similarity measures determined using PIWordNet for the dataset restricted to 26 common word pairs. Furthermore, even in this exceptional case, the difference

between the correlation coefficients determined for the Lin measure and random forest model is not significant at the  $\alpha = 0.05$  level according to the Meng, Rosenthal, and Rubin's z-test. Similar results can be observed for the word pairs from RG65 which are English counterparts of the pairs of nouns from PL39 (Table 8). The Leacock-Chodorow measure outperforms the supervised models with respect to Pearson's correlation, whereas the Jiang-Conrath and path measures outperform the supervised models with respect to Spearman's rank correlation. This suggests that, in the case of a small dataset, it is worth choosing one of the wordnet-based similarity measures instead of trying to build a supervised regression model on top of them.

**Table 7.** Correlation coefficients between the human ratings and supervised model scores determined for all PL39 word pairs that occur in PIWordNet.

Measure	Pearson's	Spearman's
lin	<b>0.7069</b>	<b>0.6941</b>
lr	0.6497	0.5909
nn	0.5373	0.4269
rt	0.5889	0.4398
rf	0.7012	0.6444
svr	0.6826	0.6337

**Table 8.** Correlation coefficients between the human ratings and supervised model scores determined for the RG65 word pairs which are counterparts of the PL39 pairs.

Measure	26 pairs		38 pairs	
	Pearson's	Spearman's	Pearson's	Spearman's
path	0.7274	0.6351	0.7300	<b>0.6911</b>
lch	<b>0.7373</b>	0.6243	<b>0.7678</b>	0.6826
jcn	0.4310	<b>0.6610</b>	0.3642	0.4315
lr	0.7208	0.5036	0.7367	0.5762
nn	0.4746	0.4044	0.3843	0.3374
rt	0.6971	0.2144	0.6734	0.2831
rf	0.7013	0.4366	0.6993	0.4884
svr	0.7199	0.5781	0.7431	0.6509

**Table 9.** The values of similarity measures computed for the RG65 word pairs which are counterparts of the PL39 pairs.

Words		path	lch	jcn
Noon	String	0.08333	1.2040	6.527e-02
Fruit	Furnace	0.11111	1.4917	6.094e-02
Autograph	Shore	0.10000	1.3863	0.000e+00
Automobile	Wizard	0.07692	1.1239	7.383e-02
Mound	Stove	0.14286	1.7430	6.815e-02
Asylum	Fruit	0.14286	1.7430	6.531e-02
Asylum	Monk	0.09091	1.2910	5.530e-02
Boy	Rooster	0.08333	1.2040	7.266e-02
Cushion	Jewel	0.14286	1.7430	6.944e-02
Monk	Slave	0.20000	2.0794	6.614e-02
Asylum	Cemetery	0.08333	1.2040	5.510e-02
Coast	Forest	0.16667	1.8971	6.276e-02
Boy	Sage	0.16667	1.8971	6.802e-02
Automobile	Cushion	0.16667	1.5404	8.940e-02
Mound	Shore	0.20000	2.0794	1.672e-01
Lad	Wizard	0.20000	2.0794	7.588e-02
Forest	Graveyard	0.11111	1.4917	5.871e-02
Food	Rooster	0.06250	0.9163	6.711e-02
Coast	Hill	0.20000	2.0794	2.187e-01
Furnace	Implement	0.12500	1.6094	7.640e-02
Crane	Rooster	0.12500	1.6094	0.000e+00
Cemetery	Mound	0.09091	1.2910	5.825e-02
Glass	Jewel	0.16667	1.7430	7.163e-02
Crane	Implement	0.20000	2.0794	7.840e-02
Brother	Lad	0.20000	2.0794	8.296e-02
Sage	Wizard	0.16667	1.8971	5.800e-02
Bird	Crane	0.25000	2.3026	0.000e+00
Bird	Cock	0.50000	2.9957	2.681e-01
Food	Fruit	0.10000	1.3863	8.607e-02
Brother	Monk	0.50000	2.9957	6.894e-02
Furnace	Stove	0.10000	1.3863	5.969e-02
Hill	Mound	1.00000	3.6889	1.288e+07
Cord	String	0.50000	2.9957	6.553e-01
Glass	Tumbler	0.50000	2.9957	3.789e-01
Autograph	Signature	0.50000	2.9957	0.000e+00
Implement	Tool	0.50000	2.9957	8.484e-01
Boy	Lad	0.50000	2.9957	2.929e-01
Automobile	Car	1.00000	3.6889	1.288e+07
Midday	Noon	1.00000	3.6889	1.288e+07

## 7 Conclusion

We presented a new framework for semantic similarity computation, using wordnet-based measures. The main advantages of our tool are compatibility with various wordnet database formats and the ability to implement new measures using embedded query language. The framework was employed to model the semantic similarity of nouns using measures derived from two Polish wordnets, PlWordNet and PolNet. The results must be considered preliminary due to the small size of the dataset used for the purpose of evaluation. Nevertheless, this is the first attempt to use both Polish wordnets within the context of a shared task.

In the future, we plan to extend the framework with additional measures (e.g., [7]). We also intend to create a larger evaluation set that will cover the content of PolNet more extensively.

## References

1. Bird, S., Klein, E., Loper, E.: *Natural Language Processing with Python*. O'Reilly Media, Sebastopol (2009)
2. Budanitsky, A., Hirst, G.: Evaluating wordnet-based measures of lexical semantic relatedness. *Comput. Linguist.* **32**(1), 13–47 (2006)
3. Diedenhofen, B.: cocor: Comparing correlations, (Version 1.0-0) (2013). <http://r.birkdiedenhofen.de/pckg/cocor/>
4. Fellbaum, C. (ed.): *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA (1998)
5. Finlayson, M.A.: Java libraries for accessing the princeton wordnet: comparison and evaluation. In: *Proceedings of the 7th Global Wordnet Conference*, Tartu, Estonia, pp. 78–85 (2014)
6. Global WordNet Association: *Global WordNet Grid* (2012). <http://globalwordnet.org/global-wordnet-grid/>. Accessed 20 Sept 2015
7. Hirst, G., St-Onge, D.: Lexical chains as representations of context for the detection and correction of malapropisms, chap. 13, pp. 305–332. In: Fellbaum [4] (1998)
8. Horak, A., Pala, K., Rambousek, A., Povolny, M.: DEBVisDic - first version of new client-server wordnet browsing and editing tool. In: Sojka, P., et al. (eds.) *Proceedings of the Third International WordNet Conference - GWC 2006*. Masaryk University, Brno, Czech Republic (2005)
9. Isahara, H., Bond, F., Uchimoto, K., Utiyama, M., Kanzaki, K.: Development of the Japanese WordNet. In: *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008, Marrakech, Morocco, 26 May–1 June 2008*, European Language Resources Association (2008)
10. Jiang, J.J., Conrath, D.W.: Semantic similarity based on corpus statistics and lexical taxonomy. In: *Proceedings of 10th International Conference on Research in Computational Linguistics, ROCLING 1997* (1997)
11. Kubis, M.: A query language for wordnet-like lexical databases. In: Pan, J.-S., Chen, S.-M., Nguyen, N.T. (eds.) *ACIIDS 2012. LNCS (LNAI)*, vol. 7198, pp. 436–445. Springer, Heidelberg (2012). [https://doi.org/10.1007/978-3-642-28493-9\\_46](https://doi.org/10.1007/978-3-642-28493-9_46)

12. Kubis, M.: A tool for transforming wordnet-like databases. In: Vetulani, Z., Mariani, J. (eds.) LTC 2011. LNCS (LNAI), vol. 8387, pp. 343–355. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-08958-4\\_28](https://doi.org/10.1007/978-3-319-08958-4_28)
13. Kubis, M.: A semantic similarity measurement tool for WordNet-like databases. In: Vetulani, Z., Mariani, J. (eds.) Proceedings of the 7th Language and Technology Conference, pp. 150–154. Fundacja Uniwersytetu im. Adama Mickiewicza, Poznań, Poland, November 2015
14. Leacock, C., Chodorow, M.: Combining local context and wordnet similarity for word sense identification, chap. 11, pp. 265–283. In: Fellbaum [4] (1998)
15. Liaw, A., Wiener, M.: Classification and Regression by randomForest. *R News* 2(3), 18–22 (2002). <http://CRAN.R-project.org/doc/Rnews/>
16. Lin, D.: An information-theoretic definition of similarity. In: Proceedings of the Fifteenth International Conference on Machine Learning, ICML 1998, pp. 296–304. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1998)
17. Maziarz, M., Piasecki, M., Szpakowicz, S.: Approaching plWordNet 2.0. In: Proceedings of the 6th Global Wordnet Conference. Matsue, Japan, January 2012
18. Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F.: e1071: Misc functions of the department of statistics (e1071), TU Wien, R package version 1.6-3 (2014). <http://CRAN.R-project.org/package=e1071>
19. Miller, G.A., Charles, W.G.: Contextual correlates of semantic similarity. *Lang. Cognit. Process.* 6(1), 1–28 (1991)
20. Paliwoda-Pękosz, G., Lula, P.: Measures of semantic relatedness based on wordnet. In: International Workshop For Ph.D. Students. Brno, Czech Republic (2009). ISBN: 978-80-214-3980-1
21. Patwardhan, S., Banerjee, S., Pedersen, T.: Using measures of semantic relatedness for word sense disambiguation. In: Gelbukh, A. (ed.) CILing 2003. LNCS, vol. 2588, pp. 241–257. Springer, Heidelberg (2003). [https://doi.org/10.1007/3-540-36456-0\\_24](https://doi.org/10.1007/3-540-36456-0_24)
22. Pedersen, T.: Information content measures of semantic similarity perform better without sense-tagged text. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT 2010, pp. 329–332. Association for Computational Linguistics, Stroudsburg, PA, USA (2010)
23. Pedersen, T., Patwardhan, S., Michelizzi, J.: WordNet::Similarity: Measuring the Relatedness of Concepts. In: Demonstration Papers at HLT-NAACL 2004, pp. 38–41, HLT-NAACL-Demonstrations 2004, Association for Computational Linguistics, Stroudsburg, PA, USA (2004). <http://dl.acm.org/citation.cfm?id=1614025.1614037>
24. Postma, M., Vossen, P.: What implementation and translation teach us: the case of semantic similarity measures in wordnets. In: Orav, H., Fellbaum, C., Vossen, P. (eds.) Proceedings of the Seventh Global Wordnet Conference, Tartu, Estonia, pp. 133–141 (2014)
25. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2014). <http://www.R-project.org/>
26. Rada, R., Mili, H., Bicknell, E., Blettner, M.: Development and application of a metric on semantic nets. *IEEE Trans. Syst. Man Cybern.* 19(1), 17–30 (1989)
27. Resnik, P.: using information content to evaluate semantic similarity in a taxonomy. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 1, IJCAI 1995, pp. 448–453. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1995)

28. Rubenstein, H., Goodenough, J.B.: Contextual correlates of synonymy. *Commun. ACM* **8**(10), 627–633 (1965)
29. Shima, H.: ws4j - WordNet Similarity for Java (2015). <https://code.google.com/p/ws4j/>. Accessed 28 Aug 2015
30. Soria, C., Monachini, M., Vossen, P.: Wordnet-LMF: Fleshing out a standardized format for wordnet interoperability. In: *Proceeding of the 2009 international workshop on Intercultural collaboration*, pp. 139–146. ACM, New York, USA (2009)
31. Stevenson, M., Greenwood, M.A.: A semantic approach to IE pattern induction. In: *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, ACL 2005*, pp. 379–386. Association for Computational Linguistics, Stroudsburg, PA, USA (2005)
32. Tengi, R.I.: Design and Implementation of the WordNet Lexical Database and Searching Software, chap. 4, pp. 105–127. In: Fellbaum [4] (1998)
33. Therneau, T., Atkinson, B., Ripley, B.: rpart: recursive partitioning and regression trees, R package version 4.1-8 (2014). <http://CRAN.R-project.org/package=rpart>
34. Venables, W.N., Ripley, B.D.: *Modern Applied Statistics with S*. Springer, New York (2002). <https://doi.org/10.1007/978-0-387-21706-2>. <http://www.stats.ox.ac.uk/pub/MASS4>. ISBN 0-387-95457-0
35. Vetulani, Z., Kubis, M., Obreński, T.: PolNet - Polish WordNet: Data and Tools. In: Calzolari, N., et al. (eds.) *Proceedings of the Seventh International Conference on Language Resources and Evaluation, ELRA, Valletta, Malta, May 2010*
36. Wu, Z., Palmer, M.: Verbs semantics and lexical selection. In: *Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics, ACL 1994*, pp. 133–138. Association for Computational Linguistics, Stroudsburg, PA, USA (1994). <https://doi.org/10.3115/981732.981751>