Enriching a Keywords Database Using Wordnets – a Case Study

Tomasz Jastrząb
Silesian University of Technology
Gliwice, Poland
Technicenter Sp. z o.o.
Bytom, Poland
Tomasz.Jastrzab@polsl.pl

Grzegorz Kwiatkowski
Silesian University of Technology
Gliwice, Poland
Grzegorz.Wojciech.Kwiatkowski
@polsl.pl

Abstract

In the paper, we study the case of building a keywords database related to the Polish Classification of Activities (PKD 2007). The database enables automatic classification of the companies to the industry branches. The classification is performed based on the company's activity description. We present the initial design of the keywords database and the ways in which wordnets were used to enrich it. Finally, we present the preliminary statistical evaluation of the produced resource.

1 Introduction

The Polish Classification of Activities (PKD 2007) (Council of Ministers, 2007), based on the European Classification of Economic Activities (NACE) (EUROSTAT European Commission, 2006), defines a hierarchical structure of industry branches and activities conducted by Polish companies. It is divided into five levels comprising sections (industries), divisions, groups, classes, and subclasses. There are 21 sections, 88 divisions and 654 subclasses, denoted by symbols consisting of letters (sections), numbers (divisions, groups, classes) or letters and numbers (subclasses).

The Polish Classification of Activities serves as a guideline for governmental institutions such as the Central Statistical Office of Poland. It acts as a source of information for services such as wskaznikibranzowe.pl[1], which publishes the quarterly and yearly financial ratios for the respective industry branches distinguished in PKD 2007. Furthermore, it can be used as a text corpus for different natural language processing tasks. In this paper, we follow the latter possibility. In particular, we use the descriptions of sections, divisions, and subclasses to build a keywords database defining each PKD 2007 section. The keywords database is then used to classify the companies to their industry branches, based on the company's activity descriptions.

Our motivations are as follows. Firstly, we want to help company owners to better describe their activities. Secondly, we want to provide an automatic tool for classifying the company to its industry. Such a tool may support search engines and allow company managers to find their competition easier. Finally, we would like to allow for simpler integration with services such as wskaznikibranzowe.pl, which given the company description, can provide the appropriate financial ratios.

The contributions of the paper are two-fold. First, we present the designed keywords database, which adds new value to the existing lexical and semantic resources. Second, we discuss the ways in which wordnet enriched the database. This way we also evaluate the wordnet in terms of data availability and completeness. We use plWordNet (Maziarz et al., 2016) as the one containing more data than the other Polish wordnet, PolNet (Vetulani, 2014).

The remainder of this paper is divided into four sections. In Section 2 we review the literature pertaining to the applications of wordnets, e.g., for building lexical resources. In Section 3 we describe the process of building and enriching the keywords database. In Section 4 we present the results of the statistical evaluation of the keywords database, while in Section 5 we summarize the paper and provide future research perspectives.

2 Related Works

Wordnets constitute lexico-semantic resources, whose basic building blocks are usually synonym sets (synsets) (Miller et al., 1990; Miller, 1990) or less frequently, lexical units (Maziarz et al., 2016). These blocks are interconnected by means of var-

[1] Available at https://wskaznikibranzowe.pl.

ious relations, such as hypernymy, hyponymy, meronymy, and others.

Wordnets support various natural language processing tasks, which we divide into the following categories:

1. Creation, extension, and enrichment of lexical and semantic resources of different types, including e.g., other wordnets, thesauri, and taxonomies.

2. Text processing tasks, such as word-sense disambiguation, entity linking, sentiment/polarity analysis, and semantic features mapping.

Within the first category of wordnet applications, the primary source of information is the Princeton WordNet (Fellbaum, 1998), which was used to construct various national wordnets. It also plays an important role in the development of multilingual resources such as the EuroWordNet or the MultiWordNet projects (Vossen, 1998; Magnini et al., 1994). Furthermore, thanks to the mapping of Princeton WordNet to the Suggested Upper Merged Ontology (SUMO) (Niles and Pease, 2003) or its use in the creation of the Yago ontology (Suchanek et al., 2007), the Princeton WordNet is used as a reliable link between these ontologies and other wordnets, e.g., plWordNet (Kędzia and Piasecki, 2014). For other projects and resources based on Princeton WordNet, created with the aim of supporting research or providing entertainment, the reader is referred to (Princeton University, 2010).

Wordnets, and in particular the Princeton WordNet, are often combined with other resources such as Wikipedia or Wiktionary to produce new or to improve existing resources. As an example of such a resource, the semantic network BabelNet could be mentioned (Navigli and Ponzetto, 2012). It combines the knowledge (synsets, relations) included in the Princeton WordNet with the data collected from Wikipedia. A similar approach, but using additional resources, was also taken when creating the ConceptNet (Speer and Havasi, 2012). A key feature of the ConceptNet and its relation to Princeton WordNet is that it aligns the Princeton WordNet concepts with other resources making it a part of the Linked Data movement. The idea of linking the Princeton WordNet within the framework of the Linguistic Linked Open Data cloud is also mentioned in (McCrae, 2018). The author focuses on the interconnection between proper nouns included in the Princeton WordNet and Wikipedia articles.

An example of a resource that is based on the Polish wordnets is the integrated wordnet discussed in (Krasnokucki et al., 2017). The resource combines the information included in PolNet and plWordNet by merging the common elements and extending the amount of information available in one of the wordnets with the contents of the other one and vice versa. The use of plWordNet in relation to ontologies is also mentioned in (Postanogov and Jastrząb, 2017), where it is considered as a source of reusable information for building new ontologies.

It is worth to mention that, although usually successful, the use of wordnets as sources of additional knowledge can also end up with a failure. An example of such a case is reported in (Poprat et al., 2008). The authors aimed at using the existing software infrastructure and data formats of Princeton WordNet to create the links between the wordnet and an Open Biomedical Ontology. It turned out that neither the data format nor the software was suitable for biomedical data representation. It mainly suffered from the limited number of relations supported by Princeton WordNet or restrictions regarding the number and format of the created concepts. Finally, the authors claimed that the Princeton WordNet provides a limited coverage of biomedical-specific terms. The limited coverage of required information in wordnets was also mentioned in (Liebeskind et al., 2018). The authors tried to create a thesaurus for Hebrew, based on the Hebrew WordNet, but due to its limited coverage they had to supplement the process by manual labour.

The second category of wordnet applications mentioned before is related to the support of natural language processing tasks. One of the key applications is the use of wordnets for opinion mining as well as sentiment and polarity analysis. Examples of semantic resources created with this purpose in mind include the SentiWordNet (Esuli and Sebastiani, 2006), Q-WordNet (Agerri and Garcia-Serrano, 2010), and plWordNet emo (Janz et al., 2017). The first two resources are based on Princeton WordNet, while the last one is based on plWordNet. The Princeton WordNet was also used e.g., for word-sense disambiguation in text clustering (Wei et al., 2015) as well as for

the document exapansion in information retrieval systems (Agirre et al., 2010).

The use cases of Polish wordnets, and especially the plWordNet, include e.g., the analysis of the amount of emotions-related information covered by plWordNet, which was investiged in (Kwiatkowski and Jastrząb, 2016a; Kwiatkowski and Jastrząb, 2016b). As shown in (Jastrząb et al., 2016; Jastrząb et al., 2017) wordnets can be also used for the semantic features mapping, which in turn can support the valence schema matching.

3 Keywords Database Design

The keywords database construction was based on the XML version of the PKD 2007 document (Główny Urząd Statystyczny, 2007). The keywords database was constructed according to the following steps:

1. Information selection and extraction,
2. Information processing,
3. Keywords extraction,
4. Keywords enrichment.

Of the above steps, the first three were performed based on the source document only, while the last step was performed with the use of wordnets.

The *information selection and extraction* step consisted in choosing the most relevant elements of the XML document. We decided to parse the contents of the following XML elements (the translations in parentheses are added for clarity, since the original names are in Polish):

- `poziom` ("level") – this is the basic element grouping the information on various levels of the PKD 2007 hierarchy;
- `numerPoziomu` ("level number"), `nazwaPoziomu` ("level name") – these two sub-elements of the `poziom` element allowed us to gain the knowledge about document structure and also to filter out the information we considered irrelevant. We decided to use only the elements corresponding to levels 1, 2 and 5, i.e., sections, divisions, and subclasses;
- `element` ("element") – this is the basic element grouping the descriptions of respective sections, divisions, and subclasses;
- `nazwa` ("name"), `symbol` ("symbol") – these two sub-elements of the `element` tag uniquely identify the members of the PKD 2007 hierarchy and can be also used for tracking the relationships between the different levels of the hierachy;
- `opisObejmujeNieobejmuje` ("description includes/excludes") – this element is the most crucial from the perspective of the keywords database construction. It contains the descriptions of the industry branches and company activities included in or excluded from the given level of the hierarchy.

Let us consider the following examples of the document contents. On level 1, we can find a section with symbol *A* and name *Rolnictwo, leśnictwo, łowiectwo i rybactwo* ("Agriculture, forestry and fishing"). On level 2, we can find a division with symbol *01* and name *Uprawy rolne, chów i hodowla zwierząt, łowiectwo, włączając działalność usługową* ("Crop and animal production, hunting and related service activities"). Finally, on level 5, we can find a subclass with symbol *01.41.Z* and name *Chów i hodowla bydła mlecznego* ("Raising of dairy cattle"). The following excerpt of the *description includes/excludes* element for subclass 01.41.Z (note the HTML tags) is an example of the source text used for the keywords database: "<h2>01.41.Z</h2> Podklasa ta obejmuje: chów i hodowlę bydła mlecznego,
produkcję surowego mleka krowiego lub z bawołów.
" ("01.41.Z This subclass includes: raising and breeding of dairy cattle, production of raw milk from cows or buffaloes.") (Główny Urząd Statystyczny, 2007).

Since the keywords database aims to support the assignment of companies to sections only, we used the *name* and *symbol* elements to combine the descriptions of divisions and subclasses with the description of the section. This way we obtained a more detailed description of each section, which constituted the input for the second step of the database construction. Note that from now on, when we speak about section description, we consider the combined descriptions mentioned above.

In the *information processing* step we first removed from the descriptions all the elements that were not words, such as HTML tags, punctuation marks, and digits (we used a set of simple regular expressions to do so). Then, based on the white

signs (spaces, tabulations, new line characters) we divided the text into words. Next, we removed those words that certainly could not become the keywords, such as conjuctions, pronouns, and prepositions. We did it semi-automatically, by removing words of length not greater than three. The process was also complemented by manual verification of the words that remained. Considering the excerpt presented above, the resulting set of words would be {Podklasa, obejmuje, chów, hodowlę, bydła, mlecznego, produkcję, surowego, mleka, krowiego, bawołów}. The number of words was finally reduced by creating a set of unique words describing each section.

The *keywords extraction* step was initialized with the calculation of the edit (Levenshtein) distance between the words describing each section. The aim of this process was to merge similar words together to further reduce their number e.g., the following words could be merged bawół, bawołów, bawoły, bawolę. While calculating the edit distance we temporarily ignored the Polish diacritics, in the sense that characters such as e.g., 'ł' and 'l', were considered to be the same. The reason for omitting the differences resulting from the use of Polish diacritics was again to limit the number of keyword candidates. Based on the obtained Levenshtein distances we merged together the words for which the distance was not greater than three. Additionally, we performed manual verification of the outcomes, to make sure that no undesired merges were made. As a result, for each section i we obtained a set of keyword candidates $K_i = \{\text{word}\}$. For each keyword candidate k, we calculated the following metric:

$$W_k = \sum_{i=1}^{n} x_i \tag{1}$$

where n is the number of sections and x_i is a binary variable such that $x_i = 1$, when $k \in K_i$, and $x_i = 0$, otherwise. Hence, for any keyword candidate k, W_k is an integer from the interval $[1, n]$. The initial set of keywords was established by removing those keyword candidates k for which $W_k \geq 2$. Since the devised set of keywords contained various forms of the same word (resulting from flection), we have manually revised all the keywords producing the set of common word forms. The final set of keywords was constructed after repeating the calculation of the W_k metric, denoted by W'_k, for the set of common word forms and rejecting the words for which the condition $W'_k \geq 2$ was satisfied.

Given the sets of keywords, we performed the *keywords enrichment* step, which involved the use of wordnets and the APT_PL tagger (Pęzik and Laskowski, 2017) used for obtaining lemmas of the keywords[2]. In this step we decided to include synonymy, hypernymy, hyponymy, and cohyponymy relations to expand the sets of keywords for each section. The reason for choosing these relations were as follows. The synonyms represent words which can be used interchangeably in the company's descriptions, so the more synonyms we can gather the better the classification quality should be. Besides, the synonyms are available straightforwardly in the wordnets, since the basic building blocks are synsets. The hypernyms allowed us to gain some knowledge on more general terms describing the concepts represented by keywords, while hyponyms allowed us to get a more detailed view on them. The cohyponyms, although usually incompatible, were chosen to enable a broader view on the given concept. Note that, before introducing the words resulting from any of the relations mentioned above, we verified whether they will not change the value of W'_k to become greater than the assumed threshold value. Words that did not satisfy this condition were rejected.

4 Statistical Evaluation

To assess the database quantitatively we measured the sizes of the resource at the various design stages. In particular, we measured the initial size of the database, calculated at the end of *information processing* step, the sizes after the application of the W_k and W'_k metrics in the *keywords extraction* step and the final size of the database after *keywords enrichment* step. The observed size changes are reflected in Figure 1. As can be observed the use of W_k and W'_k metrics reduced the initial database size almost three times. On the other hand, using the wordnet we managed to increase the number of keywords significantly, since the number of unique synsets added was approximately equal to 50 500, which means over threefold increase in the number of keywords.

The distribution of the number of keyword can-

[2]The tagger enabled us to improve the coverage of keywords by plWordNet, providing the base word forms used also in the wordnet.

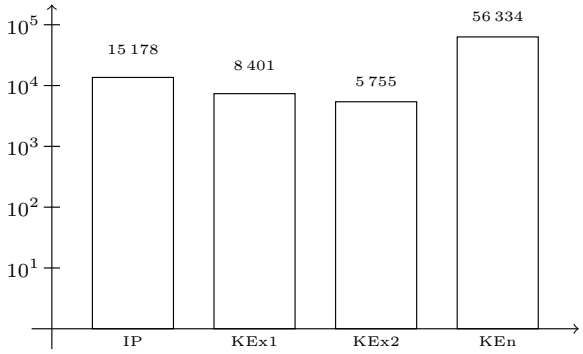

Figure 1: Changes in size of the keywords database after the *information processing* (IP) step, the W_k and W'_k metrics application in the *keyword extraction* (KEx1 and KEx2) step, and the *keyword enrichment* (KEn) step

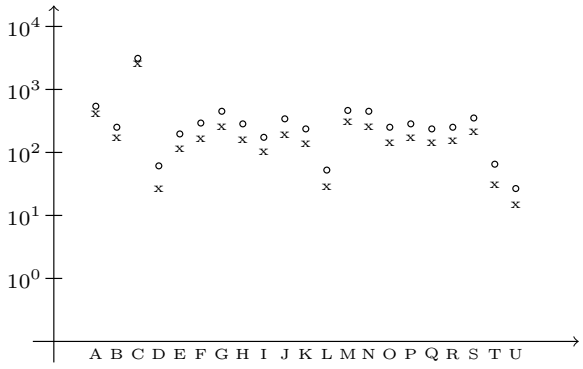

Figure 2: Keyword candidates number distribution in sections A–U of the PKD 2007. The circles represent the number of keyword candidates after W_k metric application, while the 'x' symbols denote the numbers after W'_k metric was used.

didates in the different sections is shown in Figure 2. The figure presents the information on the number of words remaining after the application of W_k (circles) and W'_k metrics ('x' symbols). It can be observed that the sections containing the fewest keyword candidates were D (Electricity, gas, steam, hot water and air conditioning manufacturing and supply), L (Real Estate Activities), T (Households as employers; goods-and-services-producing activities of households for own use), and U (Extraterritorial organisations and bodies). On the other hand, the section described by the largest number of keywords was section C (Manufacturing), which had over 3000 keyword candidates.

We also collected the information on the contribution of plWordNet towards the extension of the keywords database (Table 1). From the table it follows that the hyponymy and cohyponymy (Hyp and CoHyp columns) relations brought the largest number of keywords. Let us also note that the values presented in Table 1 are actually synsets, so the real number of words added to the database is even larger. The value given in the last column (Total) denotes the total number of unique synsets resulting from all four relations considered.

	Syn	**Hpr**	**Hyp**	**CoHyp**	**Total**
A	1187	1907	47 586	16 290	59 279
B	500	967	6131	7783	14 342
C	6695	5416	93 660	39 587	111 731
D	93	295	2683	1756	4595
E	369	840	7582	6874	14711
F	429	831	5854	5921	12 256
G	750	1404	11 674	10 542	22 533
H	472	945	6020	6845	13 364
I	321	680	5296	3929	9720
J	655	1061	59 855	7852	65 784
K	429	871	6815	5322	12 409
L	90	232	633	1623	2553
M	867	1418	47 887	10 784	55 646
N	785	1279	29 135	10 757	38 092
O	420	825	21 760	5529	26 926
P	603	1239	8844	7270	16 621
Q	382	811	5206	7033	12 929
R	478	1042	4078	6590	11 442
S	633	1181	51 828	8763	57 668
T	67	245	1169	886	2287
U	51	156	1184	1525	2825

Table 1: The number of synsets contributed by the synonymy (Syn), hypernymy (Hpr), hyponymy (Hyp), and cohyponymy (CoHyp) relations, and the total number of synsets added to the keywords database

We have observed that around 95% of initial keywords were found in plWordNet, which is a very good result. To further compare the respective sections, we have analyzed the keywords coverage percentage shown in Fig. 3. We noted that sections I, M and S were covered to the least extent. In case of sections I and S the missed keywords were usually quite specific, e.g., they were different hotel types (section I) or abbreviations (section S). In case of section M we noted the problems with the coverage of biomedical terms (see also (Poprat et al., 2008)). On the other end we observed the full coverage of sections T and

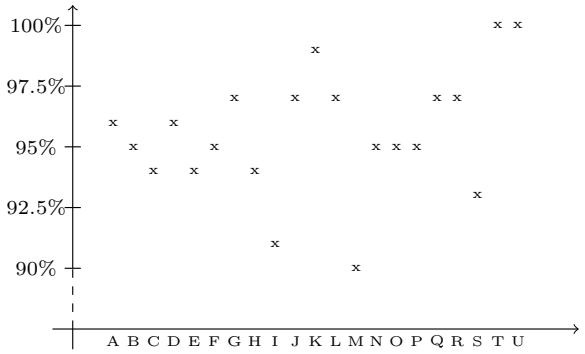

Figure 3: Keywords coverage in plWordNet (expressed as a percentage of the initial number of keywords)

U, which had relatively small number of not-so-specific keywords.

5 Summary

In the paper, we discussed the creation of a new resource related to the Polish Classification of Activities. The designed keywords database has been constructed on the basis of the official documentation related to the PKD 2007 hierarchy. The database was enriched with the use of plWordNet, the largest Polish wordnet. We used the synonymy, hypernymy, hyponymy and cohyponymy relations available in the wordnet. The results of our preliminary evaluation show that plWordNet can be a good source of information related to the activities of Polish companies.

In the future we plan to use the keywords database for the classification of companies to the respective industries given by PKD sections. We want to perform an analysis of multi-word expressions and a word-sense disambiguation step to include only the most relevant terms. Note however, that with the current design the database serves its purpose, because the not-related meanings will not appear in the company's description.

Acknowledgements

The research has been supported by the European Union under the Regional Operational Program of the Śląskie Voivodeship 2014-2020 within the project *Opracowanie zaawansowanych algorytmów automatycznego wspomagania procesów decyzyjnych w przedsiębiorstwach* ("Development of advanced algorithms for automatic decision support in enterprises") awarded to Technicenter Sp. z o.o.

References

Rodrigo Agerri and Ana Garcia-Serrano. 2010. Q-WordNet: Extracting polarity from WordNet senses. In *Proceedings of the 7th conference on Language Resources and Evaluation (LREC'10)*, pages 2300–2305.

Eneko Agirre, Xabier Arregi, and Arantxa Otegi. 2010. Document expansion based on WordNet for robust IR. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING'10)*, pages 9–17.

Council of Ministers. 2007. Regulation of the Council of Ministers of december 24th, 2007. JL No. 251, item 1885.

Andrea Esuli and Fabrizio Sebastiani. 2006. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006)*, pages 417–422.

EUROSTAT European Commission. 2006. Statistical classification of economic activities in the European Community NACE Rev. 2. WE 1893/2006.

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Główny Urząd Statystyczny. 2007. Polska Klasyfikacja Działalności PKD 2007. Available at: `https://www.dane.gov.pl/media/resources/20151019/pkd2007.xml`, last accessed: 07/04/2019.

Arkadiusz Janz, Jan Kocoń, Maciej Piasecki, and Monika Zaśko-Zielińska. 2017. plWordNet as a basis for large emotive lexicons of Polish. In *Proceedings of the 8th Language & Technology Conference (LTC'17)*, pages 189–193.

Tomasz Jastrząb, Grzegorz Kwiatkowski, and Paweł Sadowski. 2016. Mapping of selected synsets to semantic features. In *Beyond Databases, Architectures and Structures. Advanced Technologies for Data Mining and Knowledge Discovery*, volume 613 of *CCIS*, pages 357–367, Cham. Springer.

Tomasz Jastrząb, Grzegorz Kwiatkowski, Paweł Sadowski, and Adam Dyrek. 2017. A comparison of Polish wordnets in the view of semantic features mapping. In *Man-Machine Interactions 5*, volume 659 of *AISC*, pages 375–386, Cham. Springer.

Paweł Kędzia and Maciej Piasecki. 2014. Ruled-based, interlingual motivated mapping of plWordNet onto SUMO ontology. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC-2014)*, pages 4351–4358.

Daniel Krasnokucki, Grzegorz Kwiatkowski, and Tomasz Jastrząb. 2017. A new method of XML-based wordnets' data integration. In *Beyond Databases, Architectures and Structures. Towards*

Efficient Solutions for Data Analysis and Knowledge Representation, volume 716 of *CCIS*, pages 302–315, Cham. Springer.

Grzegorz Kwiatkowski and Tomasz Jastrząb. 2016a. An experimental comparison of Polish wordnets in the context of emotions analysis. In *Badania i Rozwój Mşodych Naukowców w Polsce – Nauki Techniczne i Inżynieryjne*, pages 60–66.

Grzegorz Kwiatkowski and Tomasz Jastrząb. 2016b. A survey of wordnets' applications to sentiment analysis and related problems. In *Badania i Rozwój Mşodych Naukowców w Polsce – Nauki Techniczne i Inżynieryjne*, pages 54–60.

Chaya Liebeskind, Ido Dagan, and Jonathan Schler. 2018. Automatic thesaurus construction for modern Hebrew. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, pages 1446–1451.

Bernardo Magnini, Carlo Strapparava, Fabio Ciravegna, and Emanuele Pianta. 1994. A project for the construction of an Italian lexical knowledge base in the framework of WordNet. Technical Report 9406-15, IRST.

Marek Maziarz, Maciej Piasecki, Ewa Rudnicka, Stan Szpakowicz, and Paweł Kędzia. 2016. plWordNet 3.0 – a comprehensive lexical-semantic resource. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2259–2268.

John McCrae. 2018. Mapping WordNet instances to Wikipedia. In *Proceedings of the 9th Global WordNet Conference*.

George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. 1990. Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235–244.

George A. Miller. 1990. Nouns in WordNet: A lexical inheritance system. *International Journal of Lexicography*, 3(4):245–264.

Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Ian Niles and Adam Pease. 2003. Linking lexicons and ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In *Proceedings of the 2003 International Conference on Information and Knowledge Engineering (IKE 03)*, pages 412–416.

Piotr Pęzik and Sebastian Laskowski. 2017. Evaluating an averaged perceptron morphosyntactic tagger for polish. In *Proceedings of the 8th Language & Technology Conference (LTC'17)*, pages 372–376.

Michael Poprat, Elena Beisswanger, and Udo Hahn. 2008. Building a BioWordNet by using WordNet's data formats and WordNet's software infrastructure – a failure story. In *Software Engineering, Testing, and Quality Assurance for Natural Language Processing*, pages 31–39. ACL.

Igor Postanogov and Tomasz Jastrząb. 2017. Ontology reuse as a means for fast prototyping of new concepts. In *Beyond Databases, Architectures and Structures. Towards Efficient Solutions for Data Analysis and Knowledge Representation*, volume 716 of *CCIS*, pages 273–287, Cham. Springer.

Princeton University. 2010. About WordNet. Related projects. `https://wordnet.princeton.edu/related-projects`.

Robyn Speer and Catherine Havasi. 2012. Representing general relational knowledge in ConceptNet 5. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC-2012)*, pages 3679–3686. European Language Resources Association (ELRA).

Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. YAGO: A core of semantic knowledge unifying WordNet and Wikipedia. In *WWW '07: Proceedings of the 16th International Conference on World Wide Web*, pages 697–706. ACM.

Zygmunt Vetulani. 2014. *Komunikacja człowieka z maszyną*. Akademicka Oficyna Wydawnicza EXIT, Warszawa.

Piek Vossen. 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Netwoks*. Kluwer Academic Publisher.

Tingting Wei, Yonghe Lu, Huiyou Chang, QiangZhou, and Xianyu Bao. 2015. A semantic approach for text clustering using WordNet and lexical chains. *Expert Systems with Applications*, 42(4):2264–2275.