Proceedings of eLex 2011, pp. 76-82

Visualizing sloWNet

Darja Fišer, Jernej Novak
University of Ljubljana, Faculty of Arts, Department of Translation
Aškerčeva 2, 1000 Ljubjana, Slovenia
University of Maribor, Faculty of Electrical Engineering and Computer Science, Institute of Informatics
Smetanova 17, 2000 Maribor, Slovenia
E-mail: darja.fiser@ff.uni-lj.si, jernej.novak1@uni-mb.si

Abstract

With the increasing popularity of semantic lexica such as wordnets that are being developed for more and more languages the need for tools which enable displaying and management of their content has risen as well. Dictionary writing systems or tools for ontology management are not suitable for use with wordnets because they are concept-based and relational on the one hand but less formal and more language-oriented on the other. Several specialized wordnet tools have been developed but it is still very difficult to find an all-in-one solution that would freely available and would enable on-line browsing, editing as well as visualization of wordnet content in a mono- as well as a multilingual setting. The goal of this paper is to close this gap with a light-weight and easily portable, browser-independent wordnet tool called sloWTool which supports easy importing of new wordnets or wordnet-like databases from the standard formats such as LMF and DebVisDic XML. The tool also allows adding external, third-party resources, such as wordnet domain hierarchy, coarse-grained sense clusters, and a database of images that are linked to wordnet synsets.

Keywords: wordnet; wordnet browser; wordnet editor; wordnet visualization

1. Introduction

Wordnets are semantic lexicons that have become increasingly popular in the past decade and have been developed, first for English (Fellbaum, 1998) and then also for a number of other languages (see EuroWordnet, BalkaNet, AsianWordNet), including Slovene, which has been developed automatically from heterogeneous resources, such as bilingual dictionaries, bilingual thesauri and parallel corpora (Fišer, 2009).

Wordnets differ from traditional machine-readable dictionaries with their concept-based organization and an explicit encoding of semantic information. To a certain extent, wordnets are similar to ontologies, commonly used for AI tasks in that they too define a set of semantic relations which interlink concepts into a semantic network but wordnets are less formalized and more language-oriented. This is why neither dictionary writing systems such as TshwaneLex (Joffe and de Schryver, 2004) nor tools for ontology development and maintenance such as Protégé (Noy et al., 2003) are not suitable for wordnets.

While several browsers, editors and visualization tools have been developed to deal specifically with wordnets (e.g. Fellbaum, 1998; Louw, 1997; Horak, 2006), they too are not easy use with a new wordnet due to several reasons: many of them are not publicly available, they might be intended as desktop applications for off-line browsing, do not enable the use of several wordnets in parallel, do not allow for simple editing of synsets, and do not include any visualization options. These obstacles consequently encouraged us to develop our own tool for browsing, editing and visualizing wordnet content which we present in this paper.

The paper is structured as follows: in Section 2 we present sloWNet, Section 3 analyses already existing wordnet tools, in Section 4 we present the features of sloWTool which we developed for browsing, editing and visualization of sloWNet, and then conclude the paper with final remarks and ideas for future work.

2. sloWNet

Slovene wordnet was built automatically in three stages, each time using a different approach according to the resources used for extracting the relevant lexico-semantic information. The first and most straightforward approach relied on an existing Serbian wordnet and then translated the literals into Slovene with a traditional digitized bilingual dictionary (Erjavec and Fišer, 2006). This simple approach lacked automatic disambiguation of polysemous dictionary entries and therefore required a lot of manual cleaning. This was improved in the second approach which was able to assign the correct wordnet sense to a Slovene equivalent by disambiguating it with a word-aligned parallel multilingual corpus and already existing wordnets for several languages (Fišer, 2009). The main contribution of the third and final approach was the extraction of a large number of monosemous specialized vocabulary and multi-word expressions from Wikipedia and its related resources (Fišer and Sagot, 2008). The developed wordnet contained about 17,000 literal which belonged to roughly 20,000 synsets.

Since then, sloWNet has undergone two cycles of manual revision; manually validation of all Base Concept Sets (about 5,000), and editing of all nominal synsets included in the semantic annotation of the corpus (about 1,000, see Fišer and Erjavec, 2010).

The final major step in the development of sloWNet 3.0 is the recent large-scale automatic extension in which we combined all the resources from the previous steps in order to exploit them to their full potential and thereby improve coverage of sloWNet without compromising its quality. First, a model was trained on the existing elements in sloWNet, and then a maximum entropy classifier was used to determine appropriate senses of translation candidates extracted from the heterogeneous resources described above (see Sagot and Fišer, forthcoming).

no. of synsets			no. of literals			no. of (synset, literal) pairs		
	PWN3.0	sloWNet3.0		PWN3.0	sloWNet3.0		PWN3.0	sloWNet3.0
Adj	18,156	6,218	Adj	21,538	5,108	Adj	30,004	12,438
Adv	3,621	453	Adv	4,481	514	Adv	5,580	847
N	82,114	30,911	N	119,034	30,319	N	146,345	55,383
V	13,767	5,337	V	11,531	3,840	V	25,047	14,053
total:	117,658	42,919	total:	156,584	39,781	total:	206,976	82,721
BCS1	1,220	1,220	monosemous	130,208	26,339	avg. synset length	1.76	1.92
BCS2	2,213	2,213	mwe	64,383	9,050	avg. polys.-all	1.51	2.07
BCS3	1,238	1,238	proper names	35,002	2,946	avg. polys.-poly	3.39	4.19
total:	4,671	4,671	non-letter lit.	178	32			

Table 1: A comparison of Princeton WordNet 3.0 and sloWNet 3.0

As Table 1 shows, the current version of Slovene wordnet contains 36% of all the synsets in Princeton WordNet. Nouns are still by far the most frequent, representing more than 70% of all synsets. sloWNet contains all synsets from the Base Concept Sets but also a lot of specialized vocabulary; 66% of all the literals in it are monosemous. The extended sloWNet also contains a lot of multi-word expressions and proper names, which are both mostly nominal. A comparison of the average number of literals per synset and average level of polysemy between sloWNet and PWN is interesting because it can indicate how accurate the automatic population of Slovene synsets was. While average synset length is comparable to PWN, the total average polysemy (2.07 vs. 1.51) and the average polysemy excluding monosemous words (4.19 vs. 3.39) show that Slovene wordnet contains noise that will have to be filtered out in the future.

The fact that sloWNet is somewhat noisy due to the automatic construction process is further indicated by the number of literals in the longest synsets which are, at first glance, quite similar to PWN (see Table 2) but a more careful analysis shows that even though these synsets contain several synonyms, not all of them are correct and should therefore be filtered out in the future. This is even more obvious when the most polysemous literals are searched in sloWNet which are clearly very noisy (see Table 2). The most important source of such errors was the inadequate sense assignment for the most frequent words in the language, such as the verb “to be”, the noun “person”, the adjective “big” and the adverb “very”, and will have to be corrected in the future.

While Princeton WordNet contains glosses for all its 117,658 synsets, sloWNet currently contains only 3,178 definitions for nominal synsets that were extracted automatically from Wikipedia articles. 32,881 PWN synsets are also equipped with at least one usage example which is only the case for the 517 sloWNet nominal synsets that were annotated in the corpus. A focused attempt to providing additional definition and example sentences is planned in the near future.

Domains, on the other hand, are much better represented in sloWNet. 46% of all the synsets in PWN that belong to one of the domains exist in sloWNet as well. Of all 161 domains that are present in PWN, only 4 of them are missing entirely, all of them belonging to the Sports domain hierarchy: Rugby, Soccer, Sub and Volleyball, which is a minor issue since there are only 9 synsets in PW that belong to these four domains. Just like in PWN, the most frequent domain is Factotum and the following three most frequent ones are represented in the same order in both wordnets. There are also many similarities among the ten most frequent domains in the two wordnets (see Table 3).

longest synsets		
POS	PWN 3.0	sloWNet 3.0
Adj	23 (02074929-a)	23 (00148078-a)
Adv	10 (00048739-b)	14 (00004722-b)
N	28 (05559256-n)	20 (05921123-n)
V	25 (01426397-v)	24 (00933821-v)
most polysemous literals		
POS	PWN 3.0	sloWNet 3.0
Adj	27 (heavy)	47 (velik~big)
Adv	13 (well)	13 (zelo~very)
N	33 (head)	70 (oseba~person)
V	59 (break)	757 (biti~to be)

Table 2: A comparison of longest synsets and most polysemous literals in PWN 3.0 and sloWNet 3.0

PWN 3.0	Synsets	sloWNet 3.0	Synsets
Factotum	19,454	Factotum	9,701
Zoology	6,270	Zoology	3,345
Botany	5,998	Botany	2,716
Biology	3,004	Biology	1,512
Gastronomy	2,183	Person	793
Chemistry	2,011	Admin.	790
Medicine	1,999	Chemistry	656
Admin.	1,909	Medicine	625
Anatomy	1,768	Building_ind.	575
Person	1,600	Gastronomy	525
Total	77,701	total	33,126

Table 3: A comparison of synsets belonging to domains in PWN 3.0 and sloWNet 3.0

3. Analysis of existing wordnet tools

Several wordnet tools had already been developed, best known among them being the Princeton WordNet Browser (Fellbaum, 1998), Polaris (Louw, 1997) and Periscope (Cuypers and Adriaens, 1997) for the EuroWordNet, DEBVisDic (Horak, 2006) for BalkaNet, WordNet Editor (Derwojedowa et al., 2008) for Polish and WNBrowser (Tufis, 2008) for English and Romanian wordnets.

Because so many tools already existed, it was our goal was to find the one that would best fit our needs and use it. However, our analysis has shown that it is very hard to find a tool which would enable browsing, editing and visualization all in one, and because it is far from trivial to integrate several tools that were developed for different purposes and with different technologies, specialized tools that offer just one of the desired functionalities were discarded (e.g. PWN Browser).

Also, most tools we analysed are not available under an open-source licence (e.g. Polaris, Periscope) and can therefore not be used in the sloWNet project which is based on the open source initiative. We also had to discard the tools that are platform-dependent and are meant for off-line browsing on desktops (e.g. PWN Browser) because they did not meet the requirements of the sloWNet project as such a limitation significantly undermines the usability of the lexico-semantic resource we are developing. Another technical shortcoming we observed is that it is common for wordnet tools to rely on unstandardized, in-house data formats that make it hard to import third-party lexico-semantic resources such as our wordnet (e.g. WordNet Editor, WNBrowser).

Another serious limitation of the available wordnet tools is that a number of them were developed for use in a monolingual setting and are as such unsuitable for bi- or multilingual scenarios (e.g. PWN Browser). Since the development of sloWNet is based a foreign resource, a cross-lingual comparison of concepts is without any doubt a must-have feature. When comparing options for editing wordnet entries it turned out that they are not present in many wordnet browsers at all (e.g. DEBVisDic), and when they are available, they often require installations of client software or do not support creating accounts directly by users, which makes collaborative work on wordnets difficult (e.g. DEBVisDic). This is a very important feature for the sloWNet project because we wish to use crowdsourcing techniques to validate automatically generated synsets.

Finally, when comparing applications for visualizing semantically related words in wordnet, many use Flash or Java technologies that do not perform well in older browsers and with a slower internet connection. A common problem with these applications is also that they produce overcrowded graphs which are not very informative. Similarly, some applications output a static graph for each query that cannot be further explored (e.g. WNBrowser).

Since the beginning of sloWNet development, we have relied on DEBVisDic, which is probably the most widely used wordnet editor and browser in the wordnet development community. The main reason for the change is its inconvenient collaborative on-line wordnet editing that does not support automatic registration of editors, is only possible in certain versions of Mozilla FireFox and requires installation of client packages on each computer the editor wishes to do their job, which is very inconvenient and prevents people to contribute to improving wordnet content. DEBVisDic also does not have a visualization functionality and does not allow integration of third-party resources.

4. Presentation of sloWTool

The all-in-one wordnet tool we developed tries to take all of the above into account. It incorporates browsing, editing and visualization of wordnet content with hyperbolic graphs and images. It is freely available and based on MySQL and PHP technologies, which makes the tool light-weight and portable. It is browser-independent and allows quick queries. Scripts for automatic database transformations from and into several standardized formats, such as DEBVisDic XML and LMF, are provided so that a wordnet for another language can be imported at any time. The on-line browser is simple to use for non-experts but also enables advanced searching and view settings for expert users that can enter complex search queries and decide which fields to display as well as toggle between a mono- and a multilingual option.

4.1 Technical specifications

The sloWTool is a web server application written in PHP scripting language. The tool is using the CodeIgniter open source web application framework for better transparency and maintainability of the source code. The CodeIgniter is based on the model-view-controller (MVC) development pattern. MVC is a software approach that separates an application's logic from its

presentation. In practice, it permits your web pages to contain minimal scripting since the presentation is separate from the PHP scripting[1]. The web application data is stored in 12 tables in open source MySQL database which takes approximately 100MB of hard drive. Both technologies, PHP and MySQL, are freely available and can be installed on computers with different operating systems (Linux, Mac OS, and Windows).

On the client side, in the web browser, a lot of functionality has been written in JavaScript, a scripting language for browsers. Because the client side is quite JavaScript-intensive we are using a quite few add-ons to help us cope with it. For easier HTML traversing, event handling, animation and asynchronous JavaScript and XML (AJAX) performance, we are using jQuery library. The second important add-on is the visualization plugin Springy[2] which we use to draw force-directed graphs on the HTML canvas because it was our initial desiderata not to use Flash for animating the graphs, only the HTML 5 elements. In addition, we use a window plugin mbContainerPlus[3] which helps us draw nice, movable and resizable widows for customizing the page layout.

Because of intensive use of JavaScript in the web browser we created a fluid web application which works fast and without unwanted page refreshes and interrupts. In addition, the client side of the application is using only browser capabilities for displaying the content of the page, enabling the application to work on all modern browsers that includes computers, tablets and even mobile phones with HTML 5-capable browsers.

4.1 Browsing features

The most basic feature of sloWTool is the wordnet browser which is available in simple and in advanced mode. In the simple mode the user can either display the results for a random word or search for a particular word in the desired language. When the search query is entered in the search field, a list of all synsets containing that word is displayed, including multi-word expressions, so that the user can quickly select the word or phrase they wish to see in more detail.

An example of partial search results for the word "prst" (Eng. "soil" or "finger") are shown in Figure 1. All the instances of the searched word are highlighted. Each sense of the searched word is displayed as a separate entry (synset) with language-independent information such as part of speech, synset ID and domain information shown at the top and synset edit stamp at the bottom of the entry. In the main part of the entry all the language-dependent information is provided in all the selected languages, each appearing in different colour for easier reading. In the example below the results are displayed for Slovene (black) and English (red). The most important part of the entry is the Synonyms field which shows all the words that lexicalize the concept in question (literals). In addition, each entry contains a short Definition, currently available only in English for most synsets. Some synsets also have a Usage example where the literals are used in context. Finally, all the semantic relations for that synset are displayed. In order to examine the semantic network for the searched word, it is possible to follow the related synsets and expand them into a tree.

Figure 1: An example of a Slovene synset in sloWTool

More complex search queries can be entered in the advance search window, where the user can use a combination of conditions in several fields. The example of an advanced search query in Figure 2 will find all the nominal synsets in sloWNet that contain the literal "kot" and have not been manually checked. Searches can also be performed over Definitions, Usage examples and Domains. The standard wilcards can be used as well: * for any number of any character and ? for any one character. The results of the search query and dumps of the entire database can be exported in DEBVisDic XML, LMF and tabular formats.

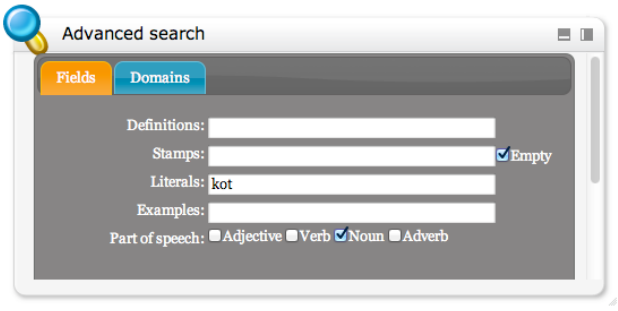

Figure 2: An example of an advanced search query

4.2 Editing features

sloWNet has been developed automatically, which is why synsets need to be manually validated in order to eliminate the noise. We have therefore developed a wordnet editor that is integrated in the browser. We have envisioned two scenarios for editing wordnet content: by

[1] http://codeigniter.com/user_guide/overview/mvc.html
[2] https://github.com/dhotson/springy
[3] http://pupunzi.open-lab.com/mb-jquery-components/mb-containerplus

random visitors of the sloWNet website who spot a mistake in a synset and are willing to correct it immediately, and by a team of lexicographers who perform systematic validation of the developed wordnet.

The first group of users are not willing to invest a lot of effort into the registration process, which is why we enable anonymous editing which does not require a login, making the editing quick and simple. However, anonymous users can only edit literals in synsets, while all the other fields are locked. Also, in order to prevent misuse of the editing option, Captcha tests appear after the maximum number of edits in one session has been exceeded. The changes to synsets that have been suggested by anonymous users are flagged for approval by a database editor, and are only then recorded as such in the database.

The second group of users are lexicographers who log in with a username and password and can edit an unlimited number of synsets, adding changes to all the fields in sloWNet. Because lexicographers are usually carefully selected, approval of the changes they suggest is not required either. Users can edit wordnet content by editing the text in the field (e.g. correcting a mistake in the tedinition), deleting a literal from a synset it does not belong to or by adding a missing literal to an existing synset.

Figure 3 contains an example of a synset which contains an inappropriate literal "pismo" (Eng. letter) for the concept of "alphabetic character" that can be deleted by clicking the Trash button.

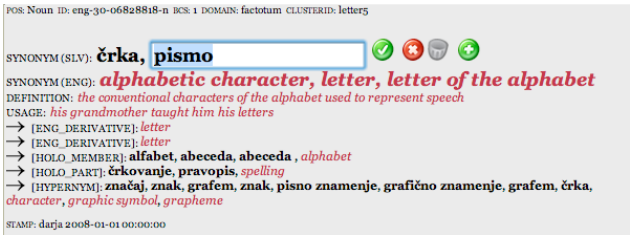

Figure 3: An example of synset editing in sloWTool

4.3 Visualization features

The results of a query are visualized in the visualization window that is displayed next to the results of a search query and can be moved and resized, so that the user can directly compare the dictionary view with the graph view. sloWTool visualizer displays all the synsets containing the searched word as well as their first-order relations. Nodes that share any first- or second- order relations are grouped into a cluster. Individual nodes can be dragged closer together or further apart in order to adjust the graph as desired. Figure 4 contains the results of wordnet visualization for the literal "prst". The search query is displayed in the center of the graph and the blue arrows lead to all its senses in wordnet. Additional information about the meaning of the displayed nodes is provided by following the red arrows to the second level of nodes that are semantically related to the original synsets, this displaying a portion of the wordnet's semantic network. Currently, nodes contain Slovene as well as English literals that belong to the same synset but it will be able to limit the display option to a single language in the future when Slovene wordnet gains in size.

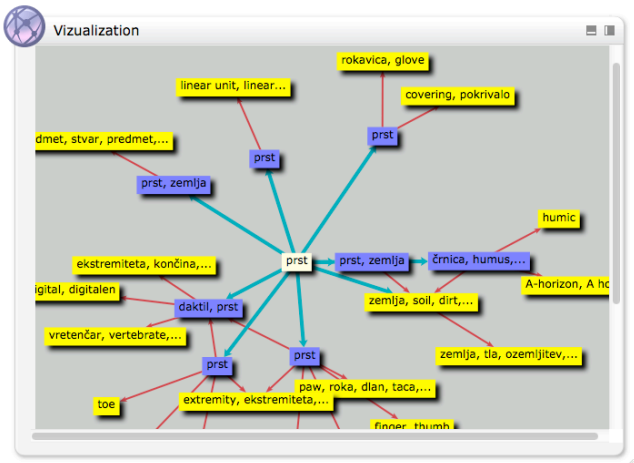

Figure 4: Visualization of wordnet content in sloWTool

4.4 External resources

Apart from developing the browser, editor and visualizer, we have also integrated several external resources into it, which make the tool even more useful. First, in order to enable a comparison between the lexico-semantic inventory in wordnet with actual word usage in context we have integrated the semantically annotated corpus (Fišer and Erjavec, 2010) in the sloWTool that displays the particular senses of the annotated nouns as they are used in context. So far about 5,000 corpus occurrences of 100 most frequent words in the jos100k corpus (Erjavec et al., 2010) have been annotated with approximately 500 different senses. In the future we plan to extend this feature into a platform for annotating all the words in the corpus with wordnet senses.

Second, in addition to sloWNet, we have imported wordnets for English and French in order to be able to compare the lexicalizations of concepts across languages. Plans for incorporating wordnets for other languages are underway.

Next, we have included the WordNet Domains Hierarchy (Bentivogli et al., 2004) which enables the users to look for all the concepts in wordnet that belong to a specific domain, such as Book_Keeping, to its more general parent domain Economy, or to the even more basic domain of Social_Science.

In order to provide a more coarse-grained sense inventory that is sufficient for most users' needs we have grouped the wordnet into meaningful clusters of word senses by mapping wordnet senses to the sense hierarchies of the Oxford Dictionary of English (Navigli,

2006). For example, instead of having to choose between 8 senses of the English word "spirit" in Princeton WordNet, we can use the 3 groups of senses for this word:

- character (2 synsets)
- atmosphere (5 synsets)
- animating force (1 synset)

The clustering is automatic and therefore not without mistakes. Furthermore, it was performed on PWN 2.1, which is why we mapped the clusters to PWN 3.0 that is not perfect either. And, last but not least, clustering was performed on literals, not synsets, which are language-specific and could not be transferred to Slovene as such. This why we conducted an additional sense-oriented grouping of these clusters in order to be able to apply it to Slovene wordnet. Nevertheless, w have already successfully employed the coarse-grained clusters for the extraction of translation equivalents of polysemous words from comparable corpora (see Fišer and Ljubešić, submitted).

Finally, we have enhanced the graph-based visualization module for displaying how words in the wordnet are interlinked by linking the wordnets with an extensive image database called ImageNet (Deng et al., 2009). It contains 12,184,113 images that were carefully selected for their quality and were annotated with 17,624 synsets by humans. Since images are linked to wordnet synsets via word ids, they can be used in other languages as well.

4.5 Availability of sloWTool

The tool is available under the Creative Commons licence of the type Attribution – NonCommercial - ShareAlike. This license lets others remix, tweak, and build upon the tool non-commercially, as long as they credit authors. The new creations of the tool must be available under identical terms. The entire licence can be found on the Creative Commons homepage[4].

The sloWTool full source code is available from Launchpad[5]. Launchpad is a hosting page for free open-source projects. It supports source code hosting using the Bazaar version control system, a bug tracker that allows bugs to be tracked in multiple contexts, a system for tracking specifications and new features, a site for localising applications into different languages, and a community support site.

In order to set up sloWTool, the requirements for the server are a computer that can run web server with the PHP scripting language, support such as Apache and the open source MySQL database. The requirement for running the client part is any modern HTML 5-capable web browser.

5. Conclusions and future work

In this paper we gave an overview of the most important tools for viewing and editing of wordnets and pointed out their shortcomings when trying to use them for sloWNet. We then presented an all-in-one tool we developed ourselves that tries to overcome all the obstacles we ran into with other already existing tools. The first problem with some of the well-known wordnet browsers and editors is that they are not freely available for installation outside the institution where it was developed. Another major issue, especially with the older browsers and editors, is that they have been designed as desktop applications meant for off-line use. Due to our project needs we were also dissatisfied with all the tools that do not support work in a multilingual setting, or tools that enable just one of the desired features.

sloWTool tries to overcome all the identified shortcomings of the available tools and provides a light-weight, easily portable and platform-independent application which is also browser-independent wordnet on the client side. It enables importing of new wordnets or wordnet-like databases from the standard formats such as LMF and DebVisDic XML. sloWTool features include simple browsing and advanced search of wordnet content, anonymous as well as systematic editing of synsets, and a graph-based visualization of the semantic network. The tool also allows adding external, third-party resources, such as wordnet domain hierarchy, coarse-grained sense clusters, and a database of images that are linked to wordnet synsets.

In the future we plan to replace the plain-text wordnet definitions with the Princeton semantically annotated glosses[6] and add links to GeoWordNet[7]. Also, we would like to add wordnets for several other languages for multilingual comparison of wordnet content. And last but not least, we are planning to extend sloWTool to allow assigning wordnet senses to words in the josSENSE corpus.

6. References

Bentivogli, L., Forner, P., Magnini, B. & Pianta, E. (2004). Revising WordNet Domains Hierarchy: Semantics, Coverage, and Balancing. In *Proceedings of the Workshop on Multilingual Linguistic Resources*, COLING'04, Geneva, Switzerland, August 28, 2004, pp. 101-108.

Buscaldi, D., Rosso, P. (2008). Geo-wordnet: Automatic georeferencing of wordnet. In *Proceedings of the 5th International Conference on Language Resources and Evaluation, LREC'2008*, Morocco, Marrakesh.

Cuypers, I., Adriaens, G. (1997). *Periscope: the EWN Viewer. EuroWordNet Project LE4003, Deliverable D008d012*. University of Amsterdam, Amsterdam.

[4] http://creativecommons.org/licenses/by-nc-sa/3.0/
[5] https://launchpad.net/slowtool
[6] http://wordnet.princeton.edu/glosstag.shtml
[7] http://geowordnet.semanticmatching.org/

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR)*.

Derwojedowa, M., Piasecki, M., Szpakowicz, S., Zawislawska, M. & Broda, B. (2008). Words, concepts and relations in the construction of Polish WordNet. In *Proceedings of the Global WordNet Conference, GWA'2008* Seged, Hungary.

Erjavec, T., Fišer, D. (2006). Building the Slovene Wordnet: first steps, first problems. In *Proceedings of the 3rd International WordNet Conference*, Jeju Island, Korea.

Erjavec, T., Fišer, D., Krek, S. & Ledinek, N. (2010). The JOS linguistically tagged corpus of Slovene. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, LREC'10, Malta, May 17-23.

Fellbaum, C. (ed.) (1998) *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Fišer, D. (2009). Laveraging parallel corpora and existing wordnets for automatic construction of the Slovene wordnet. In *Human language technology: challenges of the information society*, (LNCS 5603). Berlin; Heidelberg: Springer, pp. 359-368.

Fišer, D., Erjavec, T. (2010). sloWNet: Construction and Corpus Annotation. In *Proceedings of 5th International Conference of the Global WordNet Association*, Mumbay, India.

Fišer, D., Ljubešić, N. (submitted). Addressing polysemy in automatic bilingual lexicon extraction from comparable corpora.

Fišer, D., Sagot, B. (2008). Combining multiple resources to build reliable wordnets. In *Proceedings of TSD'08*, Brno, Czech Republic.

Horak, A., Pala, K., Rambousek, A. & Povolni, M. (2006). DEBVisDic: First Version of New Client-Server Wordnet Browsing and Editing Tool. In *Proceedings of the 3rd International WordNet Conference*, GWA'2006, Jeju Island, South Korea.

Joffe, D., De Schryver, G.M. (2004). TshwaneLex-A State-of-the-Art Dictionary Compilation Program. In *Proceedings of the Eleventh EURALEX International Congress, EURALEX'2004*, Lorient, France.

Louw, M. (1997). *The Polaris User Manual, Internal Report*, Lermout & Hauspie.

Navigli, R. (2006). Meaningful Clustering of Senses Helps Boost Word Sense Disambiguation Performance. In *Proceedings of COLING-ACL'06*, Sydney, Australia, July 17-21, 2006.

Noy, N.F., Crubézy, M., Fergerson, R.W., Knublauch, H., Tu, S.W., Vendetti, J. & Musen, M.A. (2003). Protégé-2000: An Open-Source Ontology-Development and Knowledge-Acquisition Environment. In *AMIA Annual Symposium Proceedings*. American Medical Informatics Association.

Sagot, B., Fišer, D. (forthcoming). Extending wordnets by learning from multiple resources. In *Proceedings of the LTC'2011 Conference*, Poznan, Poland.

Tufis, D., Ion, R., Bozianu, L., Ceausu, A. & Stefanescu, D. (2008). Romanian WordNet: Current State, New Applications and Prospects. In *Proceedings of the 4th Global WordNet Conference*, GWC'2008, Szeged, Hungary.