

Wordnet Consistency Checking via Crowdsourcing

Aleš Horák, Adam Rambousek

Natural Language Processing Centre, Faculty of Informatics, Masaryk University

E-mail: hales@fi.muni.cz, rambousek@fi.muni.cz

Abstract

Large ontologies and semantic networks represent complex multilevel structures, which are incredibly resistant to standard proof checking procedures. Automatic consistency checks can discover system errors such as missing intralingual links, but to find a missing word sense is a difficult task. Standard solutions rely on successive consultations of multiple information sources in a multi-level review process. In this paper, we present a new approach of supplementing such multi-level reviews with engaging the dictionary users in WordNet error corrections and enhancement proposals via systematic crowdsourcing. This approach defines an early release phase with the full dataset published to the target audience followed by a continuous workflow consisting of structured adjustment suggestions obtained from the public users and of the complete editing process by expert reviewers. The review team members are handling prestructured review tasks organized in aggregated forms with correction proposals, the revision management and the appropriate editing of proposed changes. Both the users and reviewers have access to the complete revision history, which allows them to handle repeated proposals responsibly.

Keywords: WordNet, semantic network, ontology, consistency checking

1 Introduction

Long-term development and management of a large ontology or semantic network is a tedious and time-consuming process taking a lot of manual work. Even though automatic ontology consistency checks have been developed since the creation of the first digital semantic databases (Alvez et al., 2008; Tufis & Cristea, 2002; Rath, 1999), full control of the database content is always obtained by manual inspection of the data. A small team of experts cannot completely finish such a process, and new errors are always discovered in the published versions by a broader audience.

A recent example may be the discussion at the WordNet users mailing list¹ initiated by John McCrae in August 2017. He pointed out that the term “*church mouse*” is monosemous in the English WordNet (Fellbaum, 1998):

WordNet Search – 3.1²

Noun

- {02454543} <noun.animal>[05]

S: (n) **church mouse#1**

(church mouse%1:05:00::) (a mouse created by Lewis Carroll)

¹ <https://wordnet.princeton.edu/wordnet/contact/>

² <http://wordnetweb.princeton.edu/perl/webwn>

The further talk by several wordnet experts (including Christiane Fellbaum, the current principal English WordNet coordinator) revealed that there are at least two idiomatic senses of the term missing in the database: a “*poor creature*” (“poor as a church mouse”) and a “*quiet creature*” (“quiet as a church mouse”). The following comments showed that many WordNet developers keep their own list of discovered discrepancies in the published WordNet, and stressed the need for a standard way of reporting them and possibly incorporate the suggestions in the core WordNet.

In the following text, we summarize the current state of WordNet consistency checking approaches and issues, and present a new interface for crowdsourcing, standardizing and speeding up the process of correction of (usually small) errors discovered by the wider public. The discussed tool is developed within the DEB (Dictionary Editor and Browser (Rambousek & Horák, 2016; Horák et al., 2008)) framework used for developing a number of national WordNets.

2 WordNet Development and Issues

With the WordNet concept being the best known and most widespread language ontology approach, now introduced for nearly a hundred languages, some mistakes or questionable content that can appear in the released data are generally unavoidable. Issues in WordNet may be divided into two main categories:

- surface errors – problems with synset description, e.g. spelling errors in literals or definitions,
- structural errors – issues with semantic relations, appropriate literal selection, varying subtrees depth, and granularity, or orphaned synsets.

Two general methodologies defined during the EuroWordNet project (Vossen, 1998) are general used to build new WordNets:

- Expand model – with this approach, Princeton WordNet (or its part) is translated into a new language, keeping the semantic relations mostly intact. Some projects translated the synsets semi-automatically, which may introduce surface errors if the results are not verified thoroughly.
- Merge model – new WordNet is created either from scratch or based on an existing dictionary, which does not contain semantic relations and entries are not grouped to synsets. WordNets utilizing this method tend to contain more structural errors.

Many of the errors may be prevented during the WordNet development phase. The important part is to design and follow detailed guidelines (Pociello et al., 2011; Tufis & Cristea, 2002). Software tools may help significantly. WordNet editing software should check for a range of errors, from spell-checking to semantic relations completeness (Horák et al., 2006). Some projects also use periodical heuristic testing to check recently added or updated synsets (Čapek, 2012).

3 Crowdsourcing in Linguistics

In linguistics and NLP research, crowdsourcing is generally used to manually annotate large datasets with semantic or syntactic information (Grác, 2013), word sense disambiguation (Rumshisky, 2011), or to evaluate the results of automatic tools (Nevěřilová, 2014), but may even help to detect the outbreak of epidemics (Munro et al., 2012).

The results of crowdsourcing experiments in NLP research have been evaluated multiple times, with the results showing that combining annotations by several “unskilled” annotators may result

in cheaper and faster annotation. A study by Snow et al. (2008) found that, on average, voting on four non-expert annotations achieved the equivalent precision as a single expert annotation. Another experiment (Callison-Burch, 2009) evaluated machine translation using crowdsourcing, and concluded that a combination of many non-expert evaluations provides comparable quality to that obtained with experts.

In the field of lexicography, Wiktionary³, a sister project of Wikipedia, is one of the most prominent crowdsourced resources. The goal of Wiktionary is to create a freely available “dictionary of all words in all languages” (Wikipedia, 2017) edited by volunteers. Several analyses (Hanks, 2012; Meyer & Gurevych, 2012; Fuertes-Olivera, 2009) found Wiktionary to be a useful linguistic resource, although the entry quality varies from well-crafted to unreliable.

We have previously applied crowdsourcing principles in various annotation projects, see for example Grác (2013), Nevěřilová (2014), or Kovář (2016). Based on the experience with annotation results, we have decided to develop a tool allowing users to participate in new WordNet updates.

4 Crowdsourcing Tool and Review Process

Czech WordNet (CzWN) was first published as a part of the EuroWordNet and Balkanet projects (Vossen, 1998; Christodoulakis, 2004) and since then CzWN was mostly just maintained. However, there are several versions with various amount of edits, as well as a version semi-automatically extended using a large English-Czech translation dictionary (Blahuš & Pala, 2012). The NLP Centre (the CzWN developer) is currently running a project to integrate all updates to Czech WordNet and publish a new Open Czech WordNet linked to the Collaborative Interlingual Index (Bond et al., 2016).

The Czech WordNet was developed using the Expand model, translating the English WordNet synsets. The most notable example of errors caused by this approach are the synsets containing words that are not exact synonyms, or that are rare in the Czech language, but present in the Czech WordNet because of the translation from English. For example, the English synset *cabriolet:1, cab:2* has the equivalent Czech synset *kabriolet:2, dvoukolový jednosměrný povoz:1, koňská drožka:1 (cabriolet, two-wheeled one-horse cart, horse-drawn carriage)*. Although the translation is correct, this sense of *kabriolet* in Czech is very archaic, and in the current spoken language the only sense used is *the convertible car*. Another problem is the inclusion of multiword expressions in the synset, which may be justified in some cases, but which are not fixed lexical units in the Czech language. However, during the integration we will not have enough resources and lexicographers to check all the synsets and relations in the Czech WordNet. We are receiving reports and emails about issues in the Czech Wordnet, but not always in the exact form, and it is time-consuming for editors to find the right synset and fix the error. A standardized way to report errors would make the whole process much faster and more comfortable.

Based on the feedback from the DEBVisDic users, both viewers and editors, and developers of WordNet-based applications, we have developed a new software tool to enable anyone to report issues in the WordNet data. The list of features was drafted with potential future users in mind, mostly editors of the Czech WordNet and users who use DEBVisDic for WordNet browsing. Although we are testing the tool on the Czech WordNet, it is language-independent and available for all WordNets developed using the DEBVisDic editor.

³ <http://www.wiktionary.org>

The development started in summer 2017, and the first version of the application was released in October. Currently, the prototype is in testing with the Czech WordNet data. We will evaluate the testing phase and user feedback in August 2018.

The application is developed using the client-server model, programmed in Python. The server part is responsible for the suggestion storage in the database and the connection to the DEBVisDic server. The client part is a user web interface, written in JavaScript. All parts of the application are published as open-source and available for download⁴.

The tool is not directly integrated into the DEBVisDic editor, but it uses the DEBVisDic server API to access the WordNet data. It is possible to add new modules for integration with other wordnet editors if they provide API for the synset data update. On the other hand, all available synset representations (the editor, the simplified browser, the API calls) will enable users to move to the error reporting application efficiently.

The users are presented with a data from of the synset they were browsing and they may update any item – change an existing value, add a new one if some part of the synset is missing, or remove an unwanted item. See Figure 1 for an example of the user feedback form. The updates are stored in a separate database as suggestions. Each value (e.g. a gloss or a relation) is stored as a single suggestion.

Road, route [n]

Definition: [an open way \(generally public\) for travel or transportation](#)

Domain: [town_planning](#)

Sumo: [StationaryArtifact](#)

Sumo type: +

Usages [Add](#)

Synonyms [Add](#)

[road 1](#)

[route 2](#)

Relations [Add](#)

[ENG20-01895340-v:route:2](#)

[ENG20-01897936-v:route:1](#)

Figure 1: Reporting an error in wordnet synset

⁴ Source code repository available at <https://github.com/jirkle/DEBVisDic-Report>

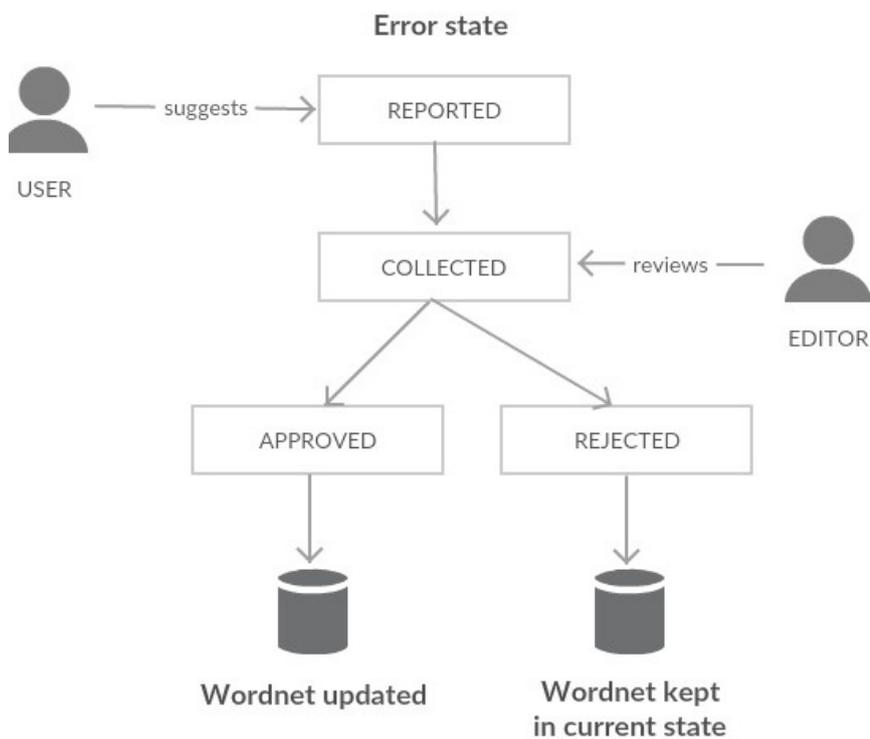


Figure 2: Review process schema

House [n]	16/11/2017	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
<input type="button" value="Add"/> Synonym → home		<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
<u>Car, auto, automobile, machine, motorcar</u> [n]	16/11/2017	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
<input type="button" value="Remove"/> Synonym motorcar →		<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
<input type="button" value="Edit"/> Synonym machine → motorcar		<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
Love, passion [n]	16/11/2017	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
<input type="button" value="Add"/> Usage → love is in the air		<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
Cat, true cat [n]	16/11/2017	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
<input type="button" value="Edit"/> Domain zoology → catology		<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
Cat, true cat [n]	16/11/2017	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
<input type="button" value="Add"/> Relation → ENG20-02352256-n:[n] paw:1		<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>

Figure 3: Administrator’s review of submitted suggestions

Any member of the editing team with access permissions to the given wordnet may browse all user suggestions (or filter them by the reporting user, the information type, or the review status). The editor may approve or reject any single proposal or approve/reject all suggestions for any synset at once. Of course, it is also possible to accept/reject all proposals based on the selected filter. Before deciding, the editor may compare the user feedback with previously approved or dismissed updates for the chosen synset. See Figure 2 for the schema of the review process and the suggestion life cycle, and Figure 3 for an example of the administrator interface for the review of any suggestions.

All the approved suggestions are immediately transferred to the development version of the WordNet database and presented to the users. When a reviewer rejects a user's feedback, the information is kept in the database and future users trying to suggest the same update are notified about the previous refusal.

In future versions, the reporting tool will support more detailed management of user roles with the possibility to provide reliable public users with tools to moderate suggestions and also enable discussion and voting about ambiguous synsets. Based on the prototype evaluation, we will consider extensions to the data presentation, e.g. enable users and developers to use data with suggestions and marking "synset reliability."

5 Conclusion

We have presented a new infrastructure for wordnet consistency checking and error reporting via crowdsourcing. The process covers all the necessary phases of the database enhancement workflow, starting with a structured proposal for an error fix in the WordNet data by a public user, followed by aggregated semi-automatic checks reviewed by a WordNet editor and projecting the correction in the development as well as the stable version of the covered WordNet database.

In the future, we will carry out a thorough public testing of this infrastructure with the Czech WordNet, and finally propagate the interface to all WordNets developed within the DEB (Dictionary Editor and Browser) framework.

Once the tool is thoroughly tested on WordNet data, it will be extended for use with any dictionary in general. The inclusion of the public enhancement proposal capability in the DEB framework will then allow to further unify and generalize the process of aggregating user suggestions to the dictionary content, and offer a straightforward application of the crowdsourced data editing to other dictionary writing applications.

References

- Alvez, J., Atserias, J., Carrera, J., Climent, S., Laparra, E., Oliver, A., and Rigau, G. (2008). Complete and consistent annotation of wordnet using the top concept ontology. In *Proceedings of LREC 2008*.
- Blahuš, M. and Pala, K. (2012). Extending Czech WordNet using a bilingual dictionary. In Christiane Fellbaum et al., editors, *6th International Global Wordnet Conference Proceedings*, pages 50–55, Matsue, Japan. Toyohashi University of Technology.
- Bond, F., Vossen, P., McCrae, J. P., and Fellbaum, C. (2016). CILI: the Collaborative Interlingual Index. In Verginica Barbu Mititelu, et al., editors, *Proceedings of the Eighth Global WordNet Conference*, pages 50–57. Romanian Academy.
- Callison-Burch, C. (2009). Fast, cheap, and creative: evaluating translation quality using Amazon's Mechanical Turk. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 286–295. Association for Computational Linguistics.

- Christodoulakis, D. (2004). *Balkanet Final Report*. University of Patras, DBLAB. No. IST-2000-29388.
- Christiane Fellbaum, editor. (1998). *WordNet: An Electronic Lexical Database*. MIT Press Cambridge.
- Fuertes-Olivera, P. A. (2009). The Function Theory of Lexicography and Electronic Dictionaries: Wiktionary as a Prototype of Collective Free Multiple-Language Internet Dictionary. In Henning Bergenholtz, et al., editors, *Lexicography at a Crossroads*, pages 103–120. Peter Lang, Bern.
- Grác, M. (2013). *Rapid Development of Language Resources*. Ph.D. thesis, Faculty of Informatics, Masaryk University.
- Hanks, P. (2012). Corpus evidence and electronic lexicography. In Sylviane Granger et al., editors, *Electronic Lexicography*, pages 57–82. Oxford University Press, Oxford.
- Horák, A., Pala, K., Rambousek, A., and Povolný, M. (2006). DEBVisDic – First Version of New Client-Server Wordnet Browsing and Editing Tool. In *Proceedings of the Third International WordNet Conference - GWC 2006*, pages 325–328, Jeju, South Korea. Masaryk University, Brno.
- Horák, A., Vossen, P., and Rambousek, A. (2008). A Distributed Database System for Developing Ontological and Lexical Resources in Harmony. In *Lecture Notes in Computer Science: Computational Linguistics and Intelligent Text Processing*, pages 1–15, Haifa, Israel. Springer-Verlag.
- Kovář, V. (2016). Evaluating Natural Language Processing Tasks with Low Inter-Annotator Agreement: The Case of Corpus Applications. In *Tenth Workshop on Recent Advances in Slavonic Natural Language Processing, RASLAN 2016*, pages 127-134. Tribun EU, Brno.
- Meyer, C. M. and Gurevych, I. (2012). Wiktionary: A new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography. In Sylviane Granger et al., editors, *Electronic Lexicography*, pages 259–291. Oxford University Press, Oxford.
- Munro, R., Gunasekara, L., Nevins, S., Polepeddi, L., and Rosen, E. (2012). Tracking Epidemics with Natural Language Processing and Crowdsourcing. In *AAAI Spring Symposium: Wisdom of the Crowd*, volume SS12-06. AAAI.
- Nevřilová, Z. (2014). Annotation Game for Textual Entailment Evaluation. In *15th International Conference, CICLing 2014, Part I*, pages 340-350. Springer, Heidelberg.
- Nevřilová, Z. (2014). *Paraphrase and Textual Entailment Generation in Czech*. Ph.D. thesis, Faculty of Informatics, Masaryk University.
- Pociello, E., Agirre, E., and Aldezabal, I. (2011). Methodology and construction of the basque wordnet. *Language Resources and Evaluation*, 45(2):121–142, May.
- Rambousek, A. and Horák, A. (2016). DEBVisDic: Instant WordNet Building. In *Proceedings of the Eighth Global WordNet Conference, GWC 2016*, pages 25–29.
- Rath, H. H. (1999). Technical issues on topic maps. In *Proceedings of Metastructures 99 Conference*. GCA.
- Rumshisky, A. (2011). Crowdsourcing word sense definition. In *Proceedings of the 5th Linguistic Annotation Workshop*, pages 74–81. Association for Computational Linguistics.
- Snow, R., O'Connor, B., Jurafsky, D., and Ng, A. Y. (2008). Cheap and fast—but is it good?: evaluating nonexpert annotations for natural language tasks. In *Proceedings of the conference on empirical methods in natural language processing*, pages 254–263. Association for Computational Linguistics.
- Tufis, D. and Cristea, D. (2002). Methodological issues in building the romanian wordnet and consistency checks in balkanet. In *Proceedings of LREC2002 Workshop on Wordnet Structures and Standardisation*, pages 35–41.
- Čapek, T. (2012). SENEQA - System for Quality Testing of Wordnet Data. In Christiane Fellbaum et al., editors, *6th International Global Wordnet Conference Proceedings*, pages 400–404, Matsue, Japan. Toyohashi University of Technology.
- Piek Vossen, editor. (1998). *EuroWordNet: a multilingual database with lexical semantic networks for European Languages*. Kluwer.
- Wikipedia. (2017). Wiktionary — Wikipedia, The Free Encyclopedia. www.wiktionary.org [Online; accessed 28-May-2018].

Acknowledgements

This work has been partly supported by the Ministry of Education of CR within the LINDAT-Clarin infrastructure LM2015071 and by the Grant Agency of CR within the project 18-23891S.