Hydra for Web: A Browser for Easy Access to Wordnets

Borislav Rizov
Institute for Bulgarian Language
Bulgarian Academy of Sciences
body@dcl.bas.bg

Tsvetana Dimitrova
Institute for Bulgarian Language
Bulgarian Academy of Sciences
cvetana@dcl.bas.bg

Abstract

This paper presents a web interface for wordnets named *Hydra for Web* which is built on top of Hydra – an open source tool for wordnet development – by means of modern web technologies. It is a Single Page Application with simple but powerful and convenient GUI. It has two modes for visualisation of the language correspondences of searched (and found) wordnet synsets – single and parallel modes. Hydra for web is available at: http://dcl.bas.bg/bulnet/.

1 Introduction

As the wordnets of the world are growing in number, implementations, applications, and the complexity of the information encoded in their relational format, wordnets data need tools for flexible but also readily accessible and easy to comprehend visualisation. Further, the tools used for creation of wordnets and visualisation of the lexical-semantic information also have to consider the relational character of the wordnet data in order to give the users access to most of the rich data without further complications and without much hidden information, especially the information concerning the relations between the synonym sets and concepts these synonym sets encode.

In the last decades, a number of web interfaces for browsing wordnet databases have been developed, with Wordvis, Mexidex, etc. among the most often used. Additionally, many web tools (mainly dictionaries) which use wordnet (especially the English wordnet) as a database for definitions and information about synonyms are available (e.g., Bee Dictionary; LookWAYUp; a2zDefined; cozyenglish, among others). Although based on wordnet, these dictionaries do not provide access to all the information about the relational organisation of the data which is one of the most valuable information in the wordnet.

There are popular user interfaces that visualise wordnet relationships as graphs. Wordvis (Vercruysse and Kuiper, 2013), for instance, do not support a parallel view of two or more wordnet language databases. Besides, being based on modern visual technological solutions, WordVis still prevents the whole needed information to be readily accessible, especially for wordnet developers. There are also tools that support parallel view as graphs such as Visual Browser (Neverilova, 2005) that can process wordnet synsets from a DEB server storage to convert them into RDF notation for visualisation (Horak et al., 2008). In the DEB platform environment, all the wordnets are stored on a DEBVisDic server; the client application supports a core module and individual modules for wordnets, so different data structure, workflow, external sources, etc. can be defined for each wordnet. The DEBVisDic was used as a basis for several multilingual projects including the Global Wordnet Grid (Horak et al., 2008). he web interface is very complicated though it is really useful for wordnet developers and for tasks involving heavy linking between wordnets, ontologies, and other lexical and semantic resources.

2 User interface and functionalities

The Hydra for web tool [1] is a web interface GUI implementation for wordnet that uses as backend the freely accessible open source modal logic tool for wordnet development Hydra (Rizov, 2008; Rizov, 2014).[2] The interface presented in this section is dependable on most of the functionalities of the

[1] Hydra for web can be checked at http://dcl.bas.bg/bulnet/.

[2] Hydra is freely available at http://dcl.bas.bg/en/hydra.html and through the META-SHARE repository at the Institute for Bulgarian Language: http://metashare.ibl.bas.bg/repository/search/.

Hydra which uses a convenient relational model to present and manage linguistic resources with relational structure.

Hydra for web is designed as Single Page Application that supports two modes – a Single Wordnet mode and Parallel Wordnets mode. Currently, it allows users to make queries into a wordnet database containing the Princeton wordnet (PWN) 3.0 (Fellbaum, 1999), the Bulgarian wordnet (BulNet) 3.0 (Koeva et al., 2004), and the Romanian wordnet (RoWN) (Tufis et al., 2013); the SentiWordnet data is in process of deployment. Thus, the web tool allows for searching into the databases of different language wordnets with a single query.

Hydra for webs interface is currently available in English, Bulgarian, and Romanian. The names of the relations and other elements were manually translated into Bulgarian (the part-of-speech – pos – and language markers – en, bg, ro – are still kept in English), and (partly) in Romanian. However, as the interface supports internationalisation, it is possible for other languages to be used.

The window has a top panel for switching the wordnets to be viewed which currently allows for the options of a Single mode – only selected synset is visualized, and the three pairs of wordnets in the Parallel wordnets mode, namely BulNet vs. PWN, BulNet vs. RoWN, and RoWN vs. PWN. However, the tool has functionalities that can further allow users to select and query any wordnet in the database.

The search panel is also present in both Single and Parallel wordnets modes. It allows searching for an exact match of a word string – a single word such as **[dance]** as shown on Fig. 1) or a multiword unit, e.g., **[barn dance]** – see Fig. 2.

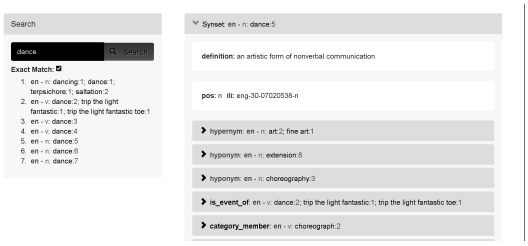

Figure 1: Hydra for web – exact match search

The non-exact match search returns any synset where the searched word string is found, as shown on Fig. 2 where the search for **[dance]** returns 24 different synsets from the Princeton word-

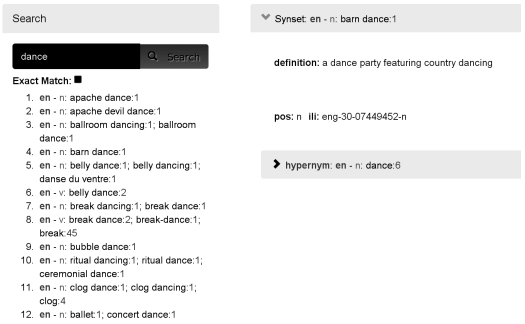

Figure 2: Hydra for web – non-exact match search

net database, among them two-word units such as **[apache dance:1]**, three-word units such as **[apache devil dance:1]**, and a hyphenated word string such as **[counter-dance:1]**, etc.

To limit the results shown, the search respects word (string) boundaries, i.e., the user can search only for whole words but not parts of the words (e.g., **[dance]** but not **[danc]** as this would return more than one hundred results – an option that is otherwise available in the Hydra software for the purposes of wordnet development. This also means that searching for the string **[dance]** will not return **[dancer]** or **[dancing]** although this word string is only part of the derived word.

2.1 Single wordnet mode

The layout in a single wordnet mode consists of two panels, namely a search panel to the left, and the synset view panel of the selected word to the right of the screen (as shown on Fig. 3 for the Princeton wordnet).

When searching for a word string in a Single wordnet mode, the search returns the synsets that contain the searched literals in all the languages in the database. The right panel displays the synset selected (e.g., the search for **[canis]** on Fig.3 returns all synsets with **[canis]** in English, Bulgarian, and Romanian wordnets).

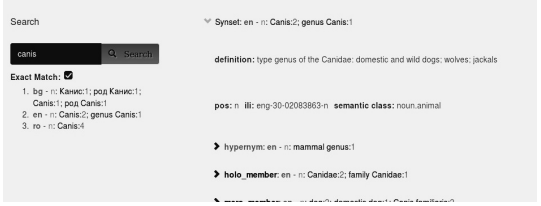

Figure 3: Hydra for web – single mode

2.2 Parallel wordnets mode

The parallel wordnets mode consists of three panels, with the second and the third panel visualising the parallel wordnets – see on Fig. 4. The two wordnet panels show the correspondences of the synset in the selected language. In this way, the user can search for a word in English, e.g., **[dog]** and with the selection of the synset **[dog:2, domestic dog:1, Canis familiaris:2]**, she can access the parallel synsets in the Bulgarian wordnet (BulNet) **[kuche:1, Canis familiaris:1]**, and in the Romanian wordnet (RoWN) – **[caine:1]**, as shown on 4. This option is very useful for fast checking the translation equivalents.

Figure 4: Hydra for web – parallel wordnets mode

On a small width (mobile), the responsive layout orders the panels successively – the search panel, then the synset views (see Fig. 5).

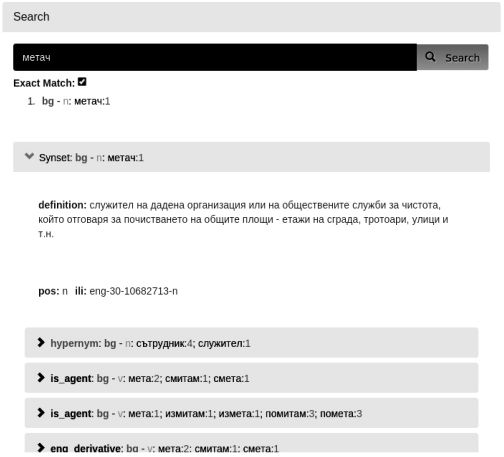

Figure 5: Hydra for web – small width (mobile) version

2.3 Synset visualisation

The elements of the synset structure are visualised in a predefined order, as shown on Fig. 6.

The literals in a synset are shown first - such as **[sweep:4, broom:2]**. The definition comes second

Figure 6: Hydra for web – ordering of synset elements

as shown by *'sweep with a broom or as if with a broom'*.

Relations that are most closely connected to the meaning of the synset are below the definition (these are usage examples in the extralinguistic relations USAGE: *'Sweep the crumbs off the table'* and *'Sweep under the bed'* on Fig. 6). The part-of-speech (pos) and the interlanguage index (ILI) are next, with hypernym(s) and hyponym(s) following, and all other relations – verb_group, is_agent, etc. (SNOTE coming at the bottom).

The relations are visually distinguished by their colour (in addition to their ordering in the synset structure) – this can be seen on the web. The part-of-speech and usage links are given in orange, the hypernyms are in blue, the hyponyms – in green, while the other relations are coloured in white.

The information in the relations are processed according to the synsets ILI. Thus, the current synset and the synset with the same ILI are marked with the arrow bullet turning red. The resulting visualisation in Hydra for web is shown on Fig. 7 where the verb synset **[sweep:4, broom:2]** is linked to the noun synset **[sweeper:3]** via the relations *is_agent* and *eng_derivative* that are both marked by the same arrow bullet – light-coloured on the Figure (on the web, the arrow turns red). The same is true for the same noun synset **[sweeper:3]** and the verb synset **[sweep:5]**. The same notification appears on the synsets in the parallel wordnet – the Bulgarian wordnet on this Figure (and if these synsets are available in the parallel wordnet).

The visualisation is recursive in a sense that every relation that leads to a synset (hypernym, holo_part, etc.) is expandable in the same way as the root one. The data like pos, ILI, etc. are available immediately, while the relations are loaded by means of AJAX query, but without blocking the

UI.

Figure 7: Hydra for web – selection of elements

3 Implementation

Hydra for web is implemented by means of modern web technologies and libraries. Its source code is relatively small, straightforward and it is easy to maintain and extend *Hydra for web* with new features.

Hydra for web is built with Node.js[3] and Express[4]. It is a single page application and uses one of the most popular HTML, CSS and JS frameworks – Bootstrap[5].

Hydra for web is themed in Slate from Bootswatch[6]. Bootstrap makes easy the GUI to be responsive, and so it is mobile friendly.

For the html rendering, the very clean and elegant JADE template engine[7] is used.

Many of the tasks in the GUI are solved in the client with the use of Knockout.js[8] framework. It uses declarative bindings, dependency tracking and provides automatic UI refresh.

The wordnet data retrieval is made by means of the Wordnet Service. The retrieval uses AJAX and is completely asynchronous (non-blocking).

3.1 Wordnet Service for wordnets

Wordnet service is a RESTful web service written in Python and Twisted[9]. The service uses the Hydra API to extract the information from the wordnet database.

The services API provides requests for searching and extracting the objects from the database (synsets, literals, and texts). It is also useful for retrieving the neighbours of a particular wordnet object by all the relations (hypernyms, hyponyms, antonyms, etc.) and its correspondent synsets in the other languages.

3.2 Hydra library

Hydra is implemented in Python, using the platform independent GUI library Tkinter. The data is managed by a MySQL server. The program allows users to query any number of wordnets simultaneously. Individual wordnets can be synchronized, allowing simultaneous visualisation of the equivalent synsets in different languages.

The program allows concurrent access by multiple users. The changes in the database are available to all users right after they are made and this option is very useful for simultaneously working wordnet developers. The important thing in our case is that it provides API for wordnet data extraction and manipulation which is at the heart of the Wordnet Service for Wordnet.

4 Applications

The most obvious application of Hydra for web is for queries into different wordnets, as well as for viewing parallel wordnet resources. Such parallel data can be used for comparative lexical and other linguistic studies. It highlights the links between words and concepts.

Hydra for web is a convenient tool for using wordnet from every place, computer, phone or other device with internet connection.

One obvious application, alongside the wordnet databases behind it, is as a multilingual dictionary.

Searching and return of single words and multiword units may help in building certain models for text identification and categorisation, word sense disambiguation, etc.

The list of results (single words and multiword units) returned contains also information about other (synonym) words and the part-of-speech of the resulting words.

[3]Node.js is a JavaScript runtime: https://nodejs.org/

[4]Web application framework for Node.js http://expressjs.com/

[5]http://getbootstrap.com/

[6]https://bootswatch.com/

[7]http://jade-lang.com/

[8]http://knockoutjs.com/

[9]https://twistedmatrix.com/

References

Christiane Fellbaum. 1999. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Ales Horak, Karel Pala, and Adam Rambousek. 2008. The Global Wordnet Grid Software Design. *Proceedings of the Fourth Global Wordnet Conference*, Szeged, Hungary, 194–199.

Svetla Koeva, Tinko Tinchev, and Stoyan Mihov. 2004. Bulgarian Wordnet – Structure and Validation. *Romanian Journal of Information Science and Technology*, 7(1–2):61–78.

Zuzana Neverilova. 2005. Visual Browser: A Tool for Visualising Ontologies. *Proceedings of I-KNOW'05*, Graz, Austria 453–461.

Borislav Rizov. 2008. Hydra: A Modal Logic Tool for Wordnet Development, Validation and Exploration. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, ELRA.

Borislav Rizov. 2014. Hydra: A Software System for Wordnet. *Proceedings of the Seventh Global Wordnet Conference*, Tartu, Estonia, 142–147.

Dan Tufis, Verginica Barbu Mititelu, Dan Stefanescu, and Radu Ion. 2013. The Romanian Wordnet in a Nutshell. *Language Resources and Evaluation*, December 2013, 47(4): 1305–1314.

Steven Vercruysse and Martin Kuiper. 2013. WordVis: JavaScript and Animation to Visualize the WordNet Relational Dictionary. *Advances in Intelligent Systems and Computing*, 179: 137-145.