

# DEBVisDic – Wordnet editor and browser based on DEB II platform

Adam Rambousek and Martin Kudlej  
Faculty of Informatics, Masaryk University  
Botanická 68a, 602 00 Brno, Czech Republic  
{xrambous, xkudlej}@fi.muni.cz

Project webpage: <http://nlp.fi.muni.cz/projects/deb2/>

Project e-mail: [deb@aurora.fi.muni.cz](mailto:deb@aurora.fi.muni.cz)

## Abstract

In this contribution, we present the new wordnet development tool called DEBVisDic. It is built on the platform for client-server XML databases, called DEB II. This platform is very versatile and allows usage in many applications, from which we concentrate on implementing various dictionary software.

The original wordnet, Princeton WordNet, is one of the most popular lexical resources in the NLP field [1]. One of the projects following it, was Balkanet (2001-4) [2] aiming to develop wordnets for 13 european languages. For this project the browser and editor VisDic has been prepared at the NLP Laboratory at the Faculty of Informatics Masaryk University [4].

Although helping greatly in Balkanet project, VisDic has some disadvantages. Thus we wanted to design a more universal dictionary writing system – this system has been called Dictionary Editor and Browser (further DEB) and its final version named DEB II is used as a main tool for development of several linguistics projects.

## 1 DEB II Development Platform

The DEB II platform follows a strict client-server architecture. Applications within the DEB platform can be divided into the server part (the server side functionality) and the client part (graphical interfaces with only simple functionality). The server part is built from small parts, called *servlets*, which allows a modular composition and reusability of all services.

The clients communicate with servlets using HTTP requests in a manner similar to recently popular concept in web development called AJAX (Asynchronous JavaScript and XML) or using more complex protocol (we decided to use JSON). The data are transported (using plain HTTP) in RDF, generic XML or plain-text formats or are marshalled using JSON.

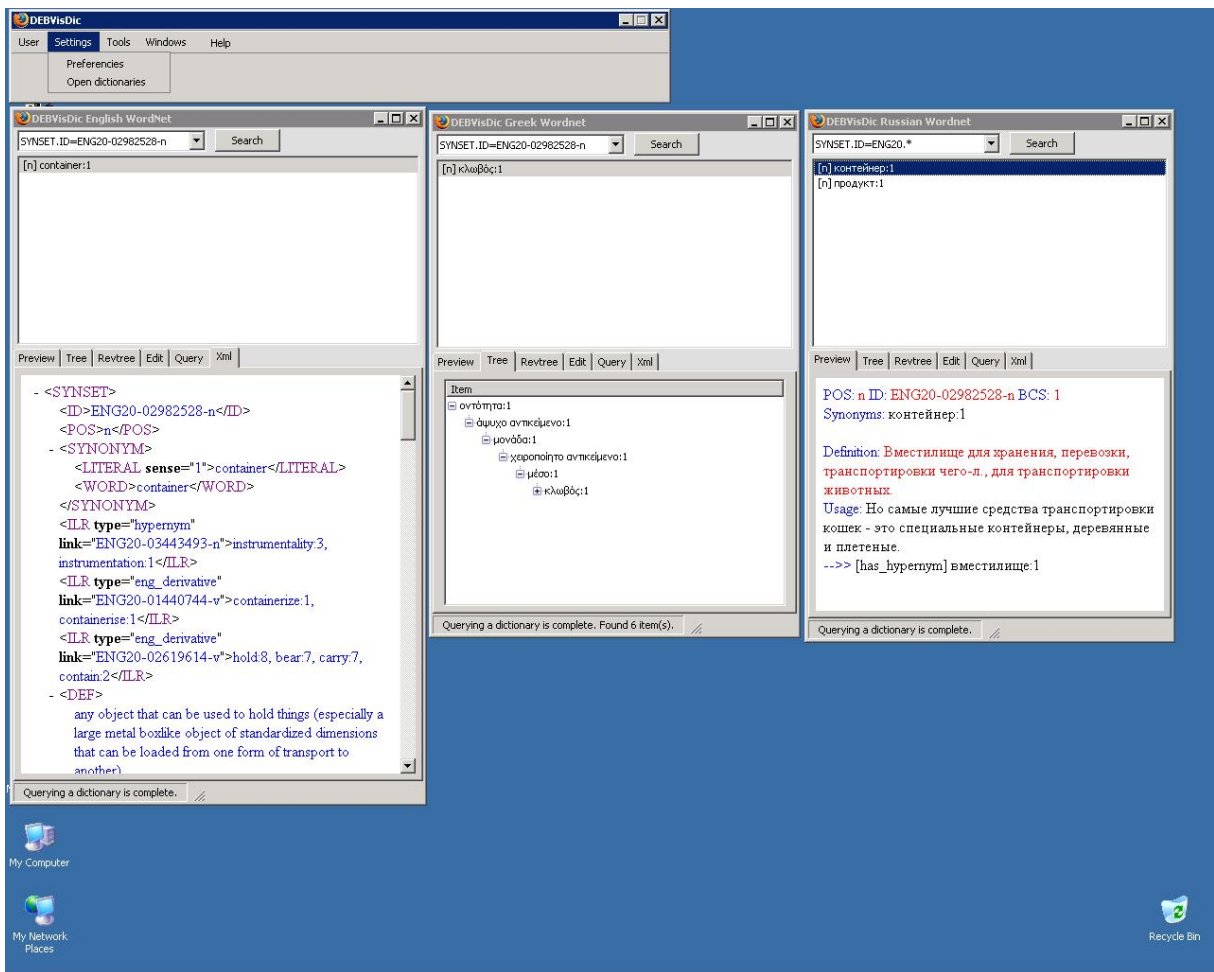


Figure 1: DEBVisDic interface with several open dictionaries

The actual data storage backend on the server side is provided by Berkeley DB XML, which is a native XML database providing XPath and XQuery access into a set of document containers.

The server side of DEB is implemented in the programming language called Ruby. Ruby (originating in Japan) is an object-oriented, interpreted programming language with weak type checking.

Since the client applications are mostly oriented to the graphical user interfaces (GUI), we have decided to adopt the concepts of the Mozilla Development Platform. The Mozilla platform provides a complete set of tools for software development. Firefox web browser is one of the many applications created using this platform. Applications built on the Mozilla platform are working within many operating systems, actually any OS on which Mozilla runs (i.e. officially Windows, Linux, and Mac OS X, unofficially many others).

It's not needed to build clients on Mozilla Platform. Thanks to client-server architecture, client software can be implemented in any programming language or platform. Only requirement is to support HTTP, JSON or SOAP protocol to communicate with server. Client side can vary from simple web page to advanced GUI client in JAVA or .NET.

One of the main benefits of the DEB platform is the homogeneity of the data structure

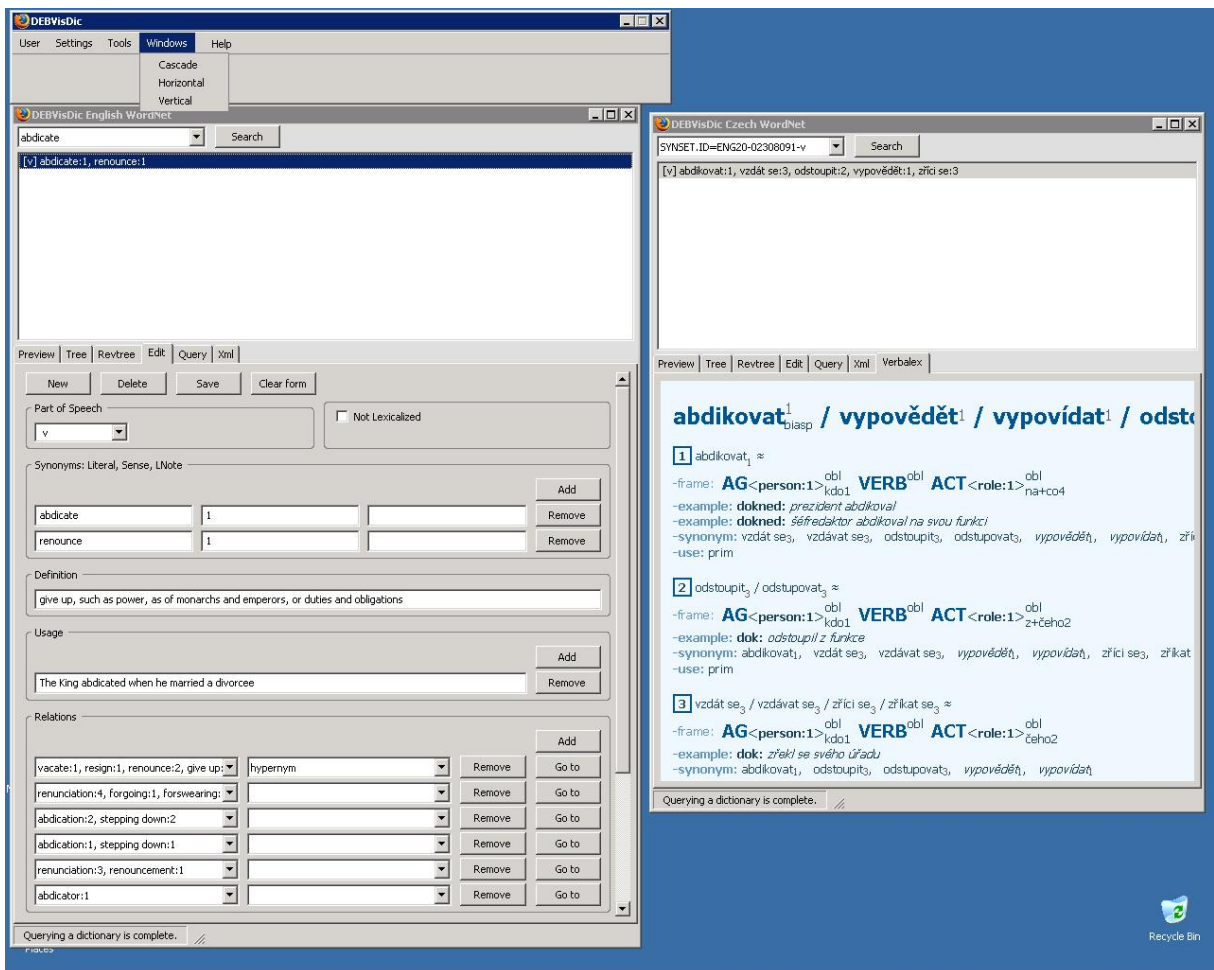


Figure 2: Edit form and Verbalex integration

and presentation, thanks to client-server architecture. Any change in data presentation made on server automatically appear in each client software.

## 2 DEBVisDic

DEBVisDic is one of the clients that are built on the DEB II platform. As the first step of the DEBVisDic we reimplemented the functions of the original VisDic (as presented on GWC 2006 conference [5]) and now we are working on extending the tool with new features for supporting the linguistic work on wordnets.

DEBVisDic uses new versatile interface that allows the user to arrange the work without any limitations. With the help of the DEB platform reusability, many new features can be included that are currently accessible only as separate tools or resources. For example, several electronic dictionaries are available in DEBVisDic and these dictionaries are used in several projects. With the new tool, it's easy to define and look-up relations between literals, for example derivational relations.

Each WordNet dictionary is prepared as a separate extension (module). Dictionaries can use shared functions of the core and add some specific functionality or settings for

the dictionary. Also, modules can provide different view on single dictionary. This way, it's possible to work with more dictionaries at the same time and to define auto look-up between them. If you select entry in one dictionary, it is automatically displayed in linked dictionary (linked entry is found by the entry identifier).

Another example of extensibility is the integration of Verbalex data in Czech wordnet. Verbalex is new lexicon of verb valencies for the Czech language [6]. If you select a verb in DEBVisDic, you can see the corresponding entry from Verbalex without need to use some other tool.

Recently, DEBVisDic was chosen as a base for the Cornetto project [3] editing tool. Aim of the Cornetto project is to build a lexical semantic database for Dutch. Although one of the base part is Dutch WordNet, Cornetto will have much more complex structure. This shows the easy extensibility of DEBVisDic tool for other tasks than a simple WordNet editing.

## Acknowledgements

This work has been partly supported by the Grant Agency of the Academy of Sciences of CR under the project 1ET100300419 and Ministry of Education of CR under the project LC536.

## References

- [1] C. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, Massachusetts, London, England, 1998.
- [2] Balkanet project website, <http://www.ceid.upatras.gr/Balkanet/>.
- [3] Cornetto project website, <http://www.let.vu.nl/onderzoek/projectsites/cornetto/>.
- [4] Aleš Horák and Pavel Smrž. Visdic – wordnet browsing and editing tool. In *Proceedings of the Second International WordNet Conference – GWC 2004*, pages 136–141, Brno, Czech Republic, 2003.
- [5] Aleš Horák, Karel Pala, Adam Rambousek, and Martin Povolný. 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, Brno, Czech Republic, 2006.
- [6] Dana Hlaváčková and Aleš Horák. Verbalex - new comprehensive lexicon of verb valencies for czech. In *Computer Treatment of Slavic and East European Languages*, Bratislava, Slovakia, 2006.