

Global WordNet Tools

Aleš Horák, Adam Rambousek

Faculty of Informatics, Masaryk University
Botanická 68a, CZ-602 00 Brno, Czech Republic
E-mail: hales@fi.muni.cz, xrambous@fi.muni.cz
Web: <http://deb.fi.muni.cz>

SeWe 2008, Svatka

Outline

- 1 Global WordNet Grid
- 2 DEB II Development Platform
- 3 DEBGrid Design and Implementation
- 4 Conclusions and Future Directions

Global WordNet Grid

- the idea first introduced at GWC 2006 in Korea by P.Vossen, A.Pease and Ch.Fellbaum (GWA)
- today:
 - more than 50 national wordnets
 - many of them not freely available due to licence restrictions
 - some are linked to the Princeton WordNet
- in the **Global WordNet Grid**:
 - multilingual database of WordNets linked to a common ontology
 - provides common subset of concepts/synsets in all languages
 - publicly accessible
 - maintained by GWA, software developed by Masaryk University on the DEB II Platform

DEB II Development Platform

DEB – Dictionary Editor and Browser

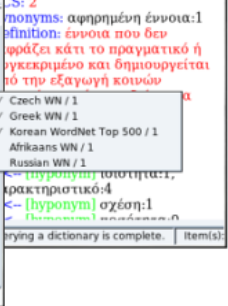
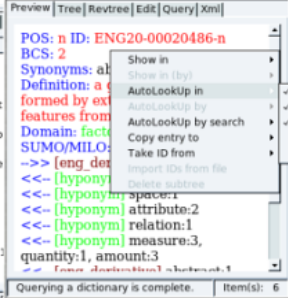
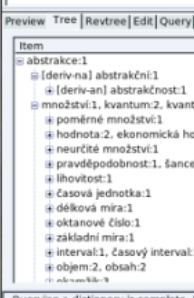
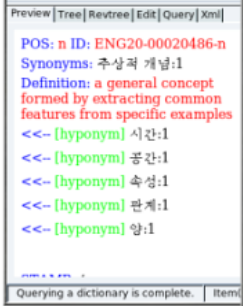
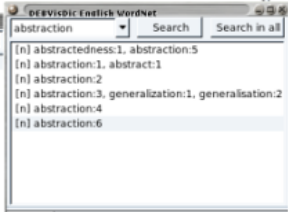
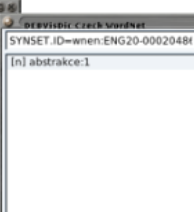
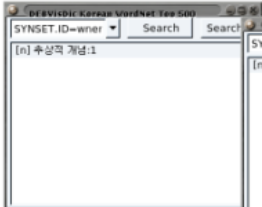
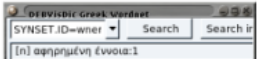
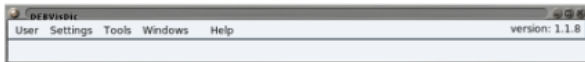
- platform for dictionary writing systems development
 - basically any XML data
 - strict client-server architecture
- server
 - server side modules (servlets in Ruby)
 - database backend (Oracle Berkeley DB XML)
- client
 - simple functionality
 - graphical interfaces (Mozilla/Firefox extensions)
 - web interfaces
- freely available – <http://deb.fi.muni.cz>:
 - 8 servers installed world-wide (Brno, Prague, Amsterdam-UvA, Amsterdam-VU, Poznan, Pretoria, Budapest, Ljubljana)
 - the Brno server – 402 users from 14 domains/countries

Outline

- 1 Global WordNet Grid
 - Global WordNet Grid
- 2 DEB II Development Platform
 - DEB II Development Platform
- 3 DEBGrid Design and Implementation
 - DEBVisDic
 - DEB WordNet Storage
 - DEB WordNet Presentation
- 4 Conclusions and Future Directions
 - Conclusions and Future Directions

DEBVisDic

- reimplementation of WordNet editor **VisDic** (off-line, local processing)
- contains all the main features present in VisDic – advanced **WordNet browsing and editing**
- adds Internet **distributed architecture**
- adds connection to **other resources** (e.g. morphological analyzer, dictionaries, Visual Browser, . . .)
- used for preparation of new Polish, Hungarian, Dutch (Cornetto), Nepalese, Afrikaans, . . . WordNets



Querying a dictionary is complete. Item(s):

Querying a dictionary is complete. Found 12 item(s)

Querying a dictionary is complete. Item(s): 6

Querying a dictionary is complete. Item(s):

DEB WordNet Storage

a WordNet in the Grid can be physically stored:

- i) on the **central server**
 - traditional DEBVisDic setup
 - offers the best performance
- ii) on a DEB server located at the **WordNet owner's institution**
 - the Central Grid server has only the knowledge of which server to contact
 - all queries are dynamically resolved over the Internet
 - can be slower
 - the WordNet owner has full control over the displayed data and access permissions
- iii) a **mixed solution** – some WordNets are stored on central server and some are stored on their respective owners' servers.

DEB WordNet Presentation

each WordNet in the Grid can be presented to the Grid users as:

- a) simple **HTML interface** working in any web browser
 - using XSLT templates, several different views – from a terse view up to a detailed view
 - best for public anonymous access to the Grid
- b) using the **full DEBVisDic application**
 - needs to be installed
 - offers much more complex functionality than the web access
 - the possibility to edit the data (with the access rights)
 - the most advanced environment for working with WordNets
- c) by means of the **DEB API**
 - any external application may query the server and receive WordNet entries (in XML or other form) for subsequent processing

query:

GWG Spanish

New search: [Grid main page](#)

POS: n ID:

ENG20-05903215-n

Synonyms: **palabra:1**

Definition: Unidad de la lengua que los hablantes nativos pueden identificar: "las palabras son las unidades con las que se construyen las frases"; "apenas dijo diez palabras en toda la mañana"

<<< [derived] [léxico:1](#)<<< [hyponym] [forma:3](#)<<< [hyponym] [préstamo:3](#)<<< [hyponym] [derivado:1](#)<<< [hyponym] [término:2](#)<<< [hyponym] [nomenclatura:1,](#)[terminología:1](#)

GWG Czech

New search: [Grid main page](#)[jazyková jednotka:1](#)

POS: n ID:

ENG20-05903215-n BCS: 1

Synonyms: **slovo:1**-->> [hypernym] [jazyková jednotka:1](#)<<< [hyponym] [slovní tvar:1](#)<<< [mero_part] [slabika:1](#)

STAMP: pala 2002/01/05 /

GWG English

New search: [Grid main page](#)[language unit:1, linguistic](#)[unit:1](#)

POS: n ID: ENG20-05903215-n BCS:

1

Synonyms: **word:1**

Definition: a unit of language that native speakers can identify

Usage: words are the blocks from which sentences are made

Usage: he hardly said ten words all morning

Domain: linguistics

SUMO/MILO: [Word](#)-->> [hypernym] [language unit:1,](#)[linguistic unit:1](#)<<< [hyponym] [form:1, word](#)[form:1, signifier:1, descriptor:1](#)<<< [mero_part] [syllable:1](#)

Global wordnet grid - SeaMonkey

File Edit View Go Bookmarks Tools Window Help

http://deb.fi.muni.cz:9000/ Search

query: ENG20-03676067-n search

GWG English

New search:

search | [Grid main page](#)

[entity:1](#)>[object:1](#), [physical](#)
[object:1](#)>[whole:2](#), [whole](#)
[thing:1](#), [unit:6](#)>[artifact:1](#),
[artefact:1](#)>[fabric:1](#), [cloth:1](#),
[material:4](#), [textile:1](#)
 POS: n ID: ENG20-03676067-n BCS:
 1
 Synonyms: **net:6**, **network:3**,
mesh:4, **meshing:2**, **meshwork:1**

Definition: an open fabric of string or rope or wire woven together at regular intervals

Domain: [industry](#)
 SUMO/MILO: [Fabric](#)
 -->> [hypernym] [fabric:1](#), [cloth:1](#),
[material:4](#), [textile:1](#)

GWG Catalan

New search:

search | [Grid main page](#)

[drap:1](#), [roba:2](#), [teixit:1](#),
[tela:1](#)
 POS: n ID: ENG20-03676067-n
 Synonyms: **xarxa:5**

Definition:
 -->> [hypernym] [drap:1](#), [roba:2](#),
[teixit:1](#), [tela:1](#)

Knowledge base browser - Fabric - SeaMonkey

File Edit View Go Bookmarks Tools Window Help

http://sigma.onl Search

Σ Sigma knowledge engineering environment | Home | Graph |

Browsing Interface

KB: SUMO
 Language: EnglishLanguage

KB Term: Show

English Word: Noun Show

Fabric^(fabric)

[Acilan](#), [Aertes](#), [Belford_cord](#),
[Brussels_carpis](#), [Brussels_lace](#),
[Canton_crepe](#), [Canton_fanned](#),
[Carrick](#), [Daxton](#), [Golden_Fleece](#),
[Harris_Tweed](#), [Levanti](#), [Levanti_morocco](#),
[Mackinaw_blanket](#), [Orlon](#),
[Persian_lamb](#), [Russia_leather](#), [Torylene](#),
[Turkish_sued](#), [Ultra suede](#),
[Valenciennois](#), [Valenciennois_lace](#),
[Yelken](#), [Yivella](#), [aba_asetate](#),
[acetate_riven](#), [acrylic](#),
[adhesive_bandage](#), [afghan](#), [alligator](#),
[alpaca](#), [antiscassar](#), [apishamiro](#),
[atrakton](#), [baggins](#), [baize](#), [balbriggan](#),
[band](#), [bandage](#), [banding](#), [bannet](#),
[basket](#), [woave](#), [bath_linen](#), [bath_mat](#),
[bath_sued](#), [bark](#), [baize](#), [beach_sued](#),
[bearskin](#)...

appearance as argument number 1

(disjunct Fabric Stationary/Artifact)	Merge.kif 12150-12150	fabric is disjunct from stationary on
(documentation Fabric EnglishLanguage "Artifacts that are created by weaving together natural or synthetic fibers or by treating the skins of certain sorts of Animals). Note that this Class includes	Merge.kif 12151-12156	%3

Conclusions and Future Directions

- first version of DEBGrid already available with 5 WordNets
<http://deb.fi.muni.cz/debgrid>
- high modularity and configurability
- further development within the KYOTO project
- waiting for other national WordNets to join 😊