

DEB Platform Utilization in Lexicographic Processes

Aleš Horák, Adam Rambousek

Natural Language Processing Centre
Faculty of Informatics, Masaryk University

`hales@fi.muni.cz`

`xrambous@fi.muni.cz`

`http://deb.fi.muni.cz`

`deb@fi.muni.cz`

Introduction

- DEB – Dictionary Editor and Browser
- dictionary writing systems platform
 - data stored in flexible XML format, Unicode
 - [client-server architecture](#)
 - all dictionary types (mono/multilingual, encyclopaedia, terminology, ontology, ...)
 - modular building blocks
- server
 - most of the functionality
 - data storage
 - user management, dictionary management
 - providing API interface
- client
 - lightweight applications
 - graphical interface
 - web interface

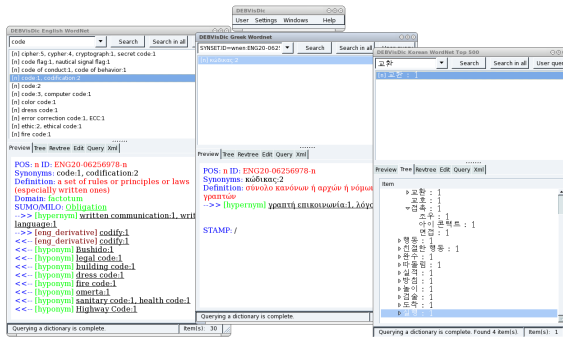
DEB projects

- used in 22 international projects, both research and industrial
- 11 server installations all over the world
- NLPC DEB server in Brno used by more than 1300 users



DEBVisDic

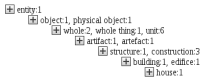
- **wordnet** and semantic network editor and browser
- more than **30 wordnet development projects** all over the world
- web application, works in all major browsers



house 1 Search Search in all

[n] house:1
[M] house:1

Preview Tree Revtree Edit Xml

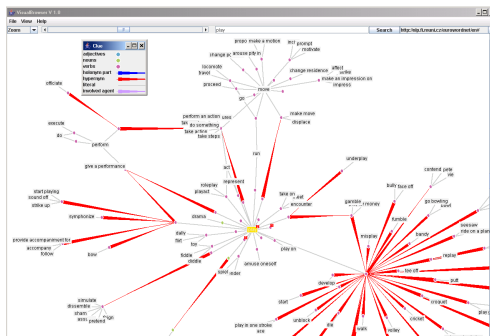


Querying a dictionary is complete. Found 7 item(s). Item(s): 2

CONCEPT.ID=wmg.ENG20	Search	Search in all	house 1	Search	Search in all
House			[n] house:1 [M] house:1		
Preview	Tree	Revtree	Xml	Preview	Tree
<p>ID: House --> [subclass] ResidentialBuilding --> [subclass] SingleFamilyResidence Show in Sigma</p>			<p>POS: n ID: ENG20-03413667-n BCS: 1 Synonyms: house:1 Definition: a dwelling that serves as living quarters for one or more families Usage: he has a house on Cape Cod Usage: she felt she had to get out of the house Domain: building_industry SUMOMILO: House --> [hypernym] building_1.edifice:1 --> [hypernym] dwelling_1.home:2.domicile:2.abode:2.habitation:2.dwelling.house:1 --> [eng_derivative] house:2.put up:7.domiciliate:2 <<- [mero_part] study:5 <<- [hyponym] bungalow:1.cottage:1 <<- [eng_derivative] house:2.put up:7.domiciliate:2 <<- [hyponym] residence:2 <<- [hyponym] ranch.house:1 <<- [hyponym] maisonette:2.maisonnette:2 <<- [hyponym] farmhouse:1 <<- [hyponym] chalet:1</p>		

DEBVisDic-based applications

- **Cornetto** (Vrije Universiteit Amsterdam) – editor and browser of Dutch lexical-semantic database
- backend database and API for wordnet and ontology storage in EU-FP7 project **KYOTO**
- **Visual Browser** – semantic network visualization



DEBDict

- general **multiple-dictionary browser**
- public access to 7 digitized Czech dictionaries
- other electronic resources (WordNet, encyclopaedia, morphological analyser, geographical information system etc.)
- dictionaries with limited access (eg. research project groups)
- dictionary portal

chrome://debdict - DEB Dictionary Browser - Mozilla Firefox

Oxford Dictionary of English | lan | Search | search in definition | Settings

dictionary - SSJC
foreign words dictionary
dictionary - SSC
thesaurus
dict. - SCFI verbal
dict. - SCFI nonverbal
all dictionaries
Concise Oxford English D.
Oxford Dictionary of English
Oxford Thesaurus of English
Diderot
google
Answers.com
Wikipedia
Seznam Encyklopedie
CIA World Factbook
map of Czech Republic
morph. analyzer ajka

langosta
langouste
langoustine
langra
language
language

Count: 116

User: xrambous@apollo.fi.muni.cz | news | DEBDict 1.7.6

language

► **noun 1.** [mass noun] the method of human communication, either spoken or written, consisting of the use of words in a structured and conventional way: *a study of the way children learn language* | [as modifier] *language development*.

NLP:
■ a non-verbal method of expression or communication: *body language*.

NLP:
2. a system of communication used by a particular country or community: *the book was translated into twenty-five languages*.

NLP:
■ [Computing] a system of symbols and rules for writing programs or algorithms.

NLP:
3. [mass noun] the style of a piece of writing or speech: *he explained the procedure in simple, everyday language*.

NLP:
■ the phraseology and vocabulary of a particular profession, domain, or group: *legal language*.

NLP:
■ [usu. as *bad/foul/strong language*] coarse or offensive language.

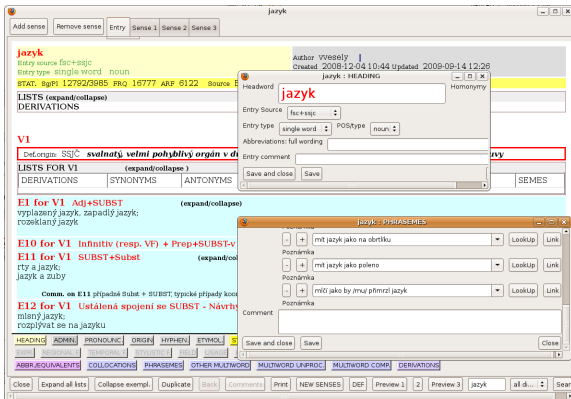
NLP:

—**PHRASES speak the same language** understand one another as a result of shared opinions or values.

NLP:
Oxford Middle English from Old English *langua*, based on Latin *lingua*

PRALED

- complex application for building new Czech lexical database
- Czech Language Institute, Academy of Sciences



FaNUK

- Family Names in UK, University of the West of England
- to be published by Oxford University Press
- data extraction from many sources (archives, census, genealogy...)

The screenshot displays the FaNUK interface for the name 'Smith'. At the top, it shows search filters: 'me', 'var', 'PR', 'Progress: R', 'GBLoc', and 'IreLoc'. Below this, the 'headword' is 'Smith', with a 'source' field and a 'variants' table. The 'variants' table lists various spellings and their frequencies:

variant	frequency
Smtha	64/3
Smeth	308/8
Smyth	8901/13211
Smythe	1779/113
Smye	89/12
Smy	748/5
Smtho	45/

Below the variants table, the 'sense' section is expanded to show the 'Occupational name' sense. The text reads: "from Middle English <i>smith</i> 'smith' (Old English <i>smið</i>), denoting a worker in metal, especially iron, such as a blacksmith or farrier. Early examples are also found in the Latin form <i>Fabr</i>. See also <i>smith</i>. Post-medieval bearers may belong under (2)."

There are also 'archive EB' sections for 'England' and 'Scotland' with associated text and links.

PDEV

- Pattern Dictionary of English Verbs, University of Wolverhampton
- Corpus Pattern Analysis, tightly connected with corpora

PatternDictionaryofEnglishVerbs

Search Browse: complete verbs (1094) | work-in-progress verbs (171) | not yet started verbs (3842) | low frequency verbs (886) | all verbs (6793)

follow

sample size: all4374 patterns: 11

No.	%	Pattern / explanation	Concordance
1	0.55%	[Human 1 Animal 1 Vehicle 1] follow [Human 2 Animal 2 Vehicle 2] <small>[Human 1 Animal 2 Vehicle 2] moves in the same direction as that selected by [Human 2 Animal 2] or the driver of [Vehicle 2]</small>	None Explanation:
2	0.19%	[Human 1 + Student Disciple] follow [Human 2 + Teacher] <small>[Human 2 + Student Disciple] studies and is influenced by or tries to practice the teachings of [Human 2 + Teacher]</small>	None Explanation:
3	1.63%	[Human institution] follow [Command Rule Plan Document] <small>[Human institution] acts in accordance with [Command Rule Plan] expressed in [Document]</small>	None Explanation:
4	6.42%	[Event 1] follow [Event 2] <small>[Event 2] happens after and typically as a consequence of [Event 1]</small>	None Explanation:
5	2.84%	[Eventuality 1] follow [EO OBJ] from [Eventuality 1] <small>[Eventuality 2] is a necessary consequence of [Eventuality 1]</small>	None Explanation:
6	27.42%	[it] follow [EO OBJ] from [Eventuality 1] that [CLAUSE] <small>[Eventuality 2] expressed by [CLAUSE] as a necessary consequence of [Eventuality 1] (either expressed in a main phrase or one that was previously stated).</small>	None Explanation:
7	52.27%	as follows <small>as follows are listed here</small>	None
8	5.97%	uses [Human 1 Institution 1] follow (suit) <small>[Human 2 Institution 2] does the same as [Human 2 Institution 2] has done</small>	None
9	2.58%	uses [Human 1] follow [EO OBJ] in [Outsteps of] [Human 1] <small>[Human 1] engages in similar activity to one previously engaged in by [Human 2]</small>	None
10	0.17%	primary uses [Human] follow [Eventuality] (up) <small>[Human] receives [Eventuality] and finds out more about it</small>	None Explanation:
11	0.02%	primary uses [Human] follow [EO OBJ] (up) (on) [Eventuality] <small>[Human] receives [Eventuality] and finds out more about it</small>	None

Pattern follow 1

Interpret: English | Concordance | Copy | Delete | Save | Save&Close | Close

Printer: [] | Comments: [] | Data Viewer: []

Subject

Human 1 Role: Least

Animal 1 Role: Least

Vehicle 1 Role: Least

Verb

follow no object no subject

Object

Human 2 Role: Least Causal opt

Animal 2 Role: Least Causal opt

Vehicle 2 Role: Least Causal opt

Primary inflection

Show [Human 1 | Animal 1 | Vehicle 1] moves in the same direction as that selected by [Human 2 | Animal 2] or the driver of [Vehicle 2] | idiom | pr | concord

Czech Sign Language dictionary



- coordinated by Teiresias Centre, Masaryk University
- several groups working on different parts
- video recordings

The screenshot shows the web interface of the 'Výkladový slovník českého a českého znakového jazyka online'. The page features a blue header with the title and navigation links for 'úvod' and 'kontakt'. On the left, there are two small video thumbnails and a list of links: 'Účastník', 'Přihlášení', 'Vyhledávání', and 'Kontakt'. The main content area is divided into several sections: a large video player showing a sign language user with a play button, a 'Signifikativ' section with a diagram of hand positions, a 'Hlasový systém' section with a '3D avatar' button, a 'Gramatický popis' section with a play button and a 'odvozeno ze znaku skřít podobenosti' label, and a 'Význam 1' section with a video thumbnail and a 'Příklady použití' section with another video thumbnail.

prezentační forma

video

+video

	video: <input type="text" value="A_DceraB.flv"/>	autor: <input type="text" value="Teiresiás OBZ"/>	<input type="button" value="odebrat"/>	<input type="button" value="zobrazit/přidat komentář"/>
	typ: <input type="text" value="čelní"/>	copyright: <input type="text" value="MU, Sředisko Teiresiás"/>	zdroj: <input type="text"/>	
	orig. název: <input type="text" value="A_DceraB.flv"/>	adm. pozn.: <input type="text"/>		
	dominance: <input type="radio"/> LR <input checked="" type="radio"/> PR			
	schváleno <input type="button" value="skrýt"/>			
	video: <input type="text" value="B_DceraB.flv"/>	autor: <input type="text" value="Teiresiás OBZ"/>	<input type="button" value="odebrat"/>	<input type="button" value="zobrazit/přidat komentář"/>
	typ: <input type="text" value="boční"/>	copyright: <input type="text" value="MU, Sředisko Teiresiás"/>	zdroj: <input type="text"/>	
	orig. název: <input type="text" value="B_DceraB.flv"/>	adm. pozn.: <input type="text"/>		
	dominance: <input type="radio"/> LR <input checked="" type="radio"/> PR			
	schváleno <input type="button" value="skrýt"/>			

SignWriting

+SW zrušit primární **skryté**

HanřkoSys

HNS editor **skryté** autor:

copyright:

zdroj:

adm. pozn.:

gramatický popis

lexikální kategorie:

typ: podtyp: -vyjádřeni plurálu:

mluvní komp.:

text:

varianty: -varianta +varianta

-varianta

skryté

autor:

copyright:

zdroj:

adm. pozn.:

DEB advantages

- team cooperation, authentication and authorization
- data management
- modular, flexible
- corpora connection (SketchEngine)
- Linked Data, API
- free (GPL)

Thank you for your attention!

More information:

<http://deb.fi.muni.cz>
deb@fi.muni.cz