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
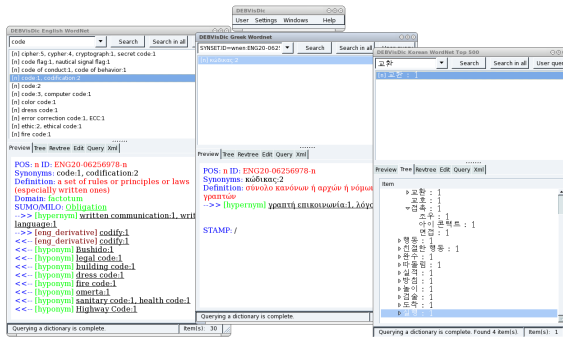   - graphical interface
    - web interface

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
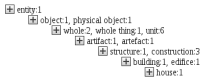
DEBVisDic

- **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
[v] house:1

Preview | Tree | Revtree | Edit | Xml

⊞ entity:1
　⊞ object:1, physical object:1
　　⊞ whole:2, whole thing:1, unit:6
　　　⊞ artifact:1, artefact:1
　　　　⊞ structure:1, construction:3
　　　　　⊞ building:1, edifice:1
　　　　　　⊞ house:1

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

CONCEPT.ID=wneng:ENG20 | Search | Search in all

House

Preview | Tree | Revtree | Xml

ID: House
-->> {subclass} ResidentialBuilding
-->> {subclass} SingleFamilyResidence
Show in Sigma

house:1 | Search | Search in all

[n] house:1
[v] house:1

Preview | Tree | Revtree | Edit | Xml

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
SUMO/MILO: 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, domicile:2
<<-- [mero_part] study:5
<<-- [hyponym] bungalow:1, cottage:1
<<-- [eng_derivative] house:2, put up:7, domicile:2
<<-- [hyponym] residence:2
<<-- [hyponym] ranch house:1
<<-- [hyponym] maisonette:2, maisonnette:2
<<-- [hyponym] farmhouse:1
<<-- [hyponym] chalet:1

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
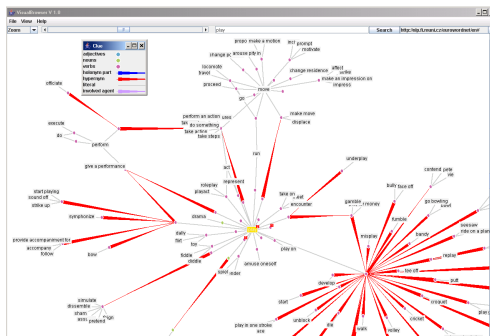
DEBDict

- 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

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

ssč	dictionary - SSJČ
ssč	foreign words dictionary
ssč	dictionary - SSČ
sts	thesaurus
sčm	dict. - SČFI verbal
sčm	dict. - SČFI nonverbal
8v	all dictionaries
	Concise Oxford English D.
	Oxford Dictionary of English
	Oxford Thesaurus of English
	Diderot
	google
	Answers.com
W	Wikipedia
	Seznam Encyklopedie
	CIA World Factbook
	map of Czech Republic
	morph. analyzer ajka

langosta
langouste
langoustine
langra
language
language area

Count: 116

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:

PRALED

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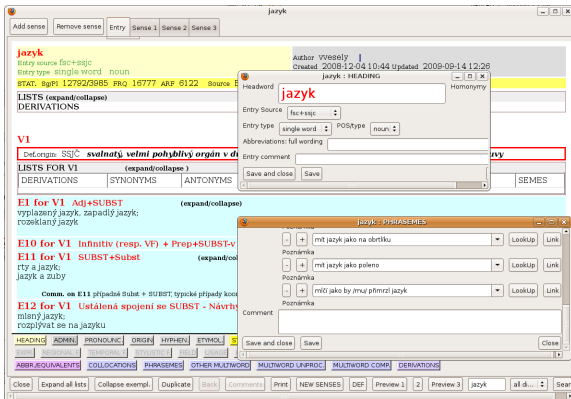
FaNUK

- 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...)

PDEV

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

Czech Sign Language dictionary

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

prezentační forma

video

+video

video:	A_DceraB.flv	autor:	Teiresiás OBZ	...	⌄	odebrat	zobrazit/přidat komentář
typ:	čelní	⌄	copyright:	MU, Středisko Teiresiás	⌄
orig. název:	A_DceraB.flv	zdroj:		...	⌄
dominance:	⊙ LP ⊙ PP	adm. pozn.:	
schváleno	skrýt

video:	B_DceraB.flv	autor:	Teiresiás OBZ	...	⌄	odebrat	zobrazit/přidat komentář
typ:	boční	⌄	copyright:	MU, Středisko Teiresiás	⌄
orig. název:	B_DceraB.flv	zdroj:		...	⌄
dominance:	⊙ LP ⊙ PP	adm. pozn.:	
schváleno	skrýt

SignWriting

+SW ⊙ zrušit primární   **skryté** schválit k publikování   zobrazit/přidat komentář

HamNoSys

HNS editor

skryté schválit k publikování

autor:		...	⌄	zobrazit/přidat komentář
copyright:		...	⌄
zdroj:		...	⌄
adm. pozn.:	

gramatický popis

lexikální kategorie:	podstatné jméno (substant ⌄			
typ:	.	⌄	podtyp:		⌄	-vyjádření plurálu-	⌄
mluvní komp.:	dcera	povinný	⌄	orální komp.:		⌄
text:			⌄

skryté schválit k publikování   zobrazit/přidat komentář

autor:		...	⌄
copyright:		...	⌄
zdroj:		...	⌄
adm. pozn.:	

varianty:	A_DceraC.flv		⌄		-varianta	+varianta
	A_DceraD.flv		⌄		-varianta	

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