DEB II Platform Deployment
Current Applications

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

RASLAN 2007

Outline

Outline

Outline

Outline

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 functionality (servlets in Ruby)
    - database backend (Oracle Berkeley DB XML)
- client
    - only simple functionality
    - graphical interfaces (Mozilla/Firefox extensions)
    - web interfaces
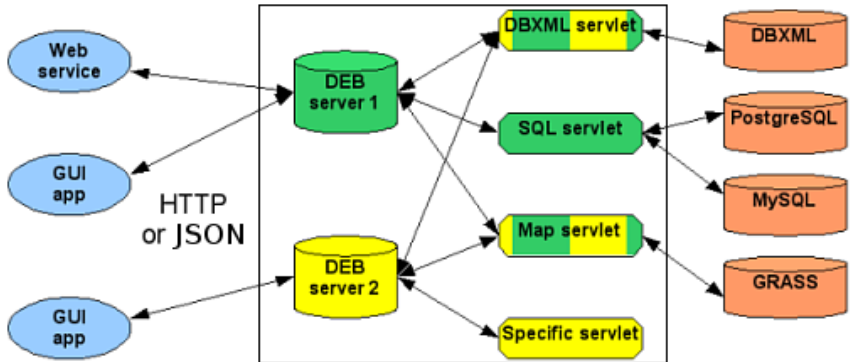
DEB uses AJAX-like (Asynchronous JavaScript And XML)
communication

Assets of the DEB Platform

- data stored on the server
- functionality mostly on the server, lightweight clients
- good team cooperation
- authentication and authorization tools provided by the server
- different interfaces over the same data structure, reusable by many clients
- homogeneity of the data structure and presentation

Outline
DEB II Development Platform
Current DEB projects
Dictionary Administration Interface
Conclusions and Future Directions

DEBDict
DEBVisDic
DEB CPA
PRALED
DEB TeDi

DEBDict

- viewer of different dictionaries
    - 7 Czech dictionaries
    - 3 English dictionaries (limited use)
    - CIA World Factbook
- connection to Czech morphological analyzer (ajka)
- connection to an external website
- geographical system, map of Czech Republic
- multilingual user interface

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

- dictionary - SSJČ
- foreign words dictionary
- dictionary - SSČ
- thesaurus
- dict. - SČFI verbal
- dict. - SČFI 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

▶ **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:
– ORIGIN Middle English: from Old French langage based on Latin lingua

Outline
DEB II Development Platform
Current DEB projects
Dictionary Administration Interface
Conclusions and Future Directions

DEBDict
DEBVisDic
DEB CPA
PRALED
DEB TeDi

DEBVisDic

- reimplementation of WordNet editor VisDic
- contains all the main features present in VisDic – advanced WordNet browsing and editing
- connection to other resources (e.g. morphological analyzer, dictionaries)
- preparation of new Polish, Hungarian, Slovenian, Nepalese and Afrikaans WordNets

Cornetto – combining Dutch WordNet with lexical semantic database

Global WordNet Grid – multilingual database of (small) WordNets linked to a common *ontology*, public web access

Cornetto ID

aanbieden

[Search] [Search in all]

aanbieden:1, VERB, LU=r_v-7, SY=d_v-10;, MC
aanbieden:2, VERB, LU=c_545132, SY=;, MC
aanbieden:3, VERB, LU=d_v-192, SY=d_v-9; -1.0, -C
aanbieden:4, VERB, LU=d_v-400504, SY=d_v-11; -1.0, -C

Preview | **Edit** | **Xml**

[Delete] [Save] [Map current LU to current Synset]

Current link

LU ID   [*] r_v-7 verb, aanbieden:1, laten weten dat het beschikbaar is ▼

Synset ID   [*] d_v-10 [verb] aanbieden:2, aanbieden:1 | | (van zaken) zich vertonen, zich voor ▼

score   33.9

status ☑

Comment

Alternate mappings: LU ID, Synset ID, status

[*] r_v-7 verb, aanbieden:1, laten weten dat het beschikbaar is ▲▼

d_v-8 [verb] bieden:4, brengen:7, lenen:5, offreren:1, presenteren:4, aanbieden:1 | | ▲▼

72.75   ☑ status   ☐ [Delete]

[*] r_v-7 verb, aanbieden:1, laten weten dat het beschikbaar is ▲▼

[*] d_v-10 [verb] aanbieden:2, aanbieden:1 | | (van zaken) zich vertonen, zich voordoen ▲▼

33.9   ☑ status   ☑ · [Delete]

[*] r_v-7 verb, aanbieden:1, laten weten dat het beschikbaar is ▲▼

d_v-11 [verb] aanbieden:4 | | ▲▼

Querying a dictionary is complete.   Item(s): 4

Outline
DEB II Development Platform
Current DEB projects
Dictionary Administration Interface
Conclusions and Future Directions

DEBDict
DEBVisDic
DEB CPA
PRALED
DEB TeDi

DEB CPA

- Corpus Pattern Analysis editing tool, see e.g. (Hanks 2004)
- integrated with corpora and Word Sketch Engine
- CPA ontology editing

Outline
DEB II Development Platform
Current DEB projects
Dictionary Administration Interface
Conclusions and Future Directions

DEBDict
DEBVisDic
DEB CPA
PRALED
DEB TeDi

PRALED

- authoring system for Czech lexical database used by Institute of Czech Language
- very complex data structure
- data from Frequency Dictionary of Czech
- integrated with corpora
- connected with DEBDict for dictionary lookup

Add sense | Remove sense | Headword | Sense 1 | Sense 2

!fenykl
Type single noun Zpracov
<z latiny> Source ASCS Hyphen. fe-ny-kl
Deklinace skl. Číslo sg. 1 pl. Rod muž. neživ. Podtyp PARAD Vytvoř

KOMPONENTY U VÍCESL. SLOVOTVORBA

V1 bot.

silně aromatická bylina s drobnými dělenými lís

E k V1

doc.txtype=POP,doc.opus=tropicto již od starověku
obecného , který se pěstuje pro nať semena a použív

!fenykl: GEOGRAPHIC

Region	Dialect
⊞ Čechy	⊞ Čechy
⊞ Morava	⊞ Morava
⊞ Slezsko	⊞ Slezsko

Note

Save&close | Save | Close

!fenykl: MAIN INFO
Headword **!fenykl**   Homonymy

Source

Typ hesla single   PoS noun

Save&close | Save | Close

officinale) v pozadí a načechranými listy < fenyklu> (

RONYMA HYPONYMA KOHYPONYMA SÉMY

V2

kořen z nažek fenyklu obecného

E k V2

doc.txtype=POP doc.opus=zahr98sklenic a prokládáme několika kolečky křenu . Zalijeme nálevem , přidáme < fenykl> a

MAIN INFO | ADMIN. | PRON. | ORIGIN | OBOR | EXPR. | GEOGRAPHIC | TIME | KPI | HYPH. | ETYM. | STAT. | GRAM. | VAR.

ABBREV | MULTIWORD | PHRASES | COMPONENTS | WORDFORM

Close | Test | Notes | Print | nový | Preview | Preview 2 | fenykl | Diderot | Hledat

Outline
DEB II Development Platform
Current DEB projects
Dictionary Administration Interface
Conclusions and Future Directions

DEBDict
DEBVisDic
DEB CPA
PRALED
DEB TeDi

DEB TeDi – Terminological Dictionary

- joint project with Faculty of Fine Arts, Brno University of Technology
- dictionary of terms concerning art for FFA students
- 1500 entries by the end of the year, continuation uncertain
- web interface, different for editing and browsing

https://apollo.fi.muni.cz:8010/tedi?action=edit&id=adresa_(g 🔍 Search

🏠 Home 📁 Bookmarks

Autor: ID: **adresa_(grafická_dedikační)-11804343501**

obor   kresba-grafika

heslo česky   anglicky

adresa (grafická, dedikační)   address

německy   francouzsky

Adresse f.   adresse f.

varianty   +

styl. příznak   standard, knižní

definice   součást značení díla, starých grafických listů obsahující jméno či obchodní označení vydavatele nákladu rytin, leptů atp. Trvale obsažena ve vlastním tisku.

příklady   +

nadpojem   značení díla   +

podpojem   např. pinxit, invenit, delineavit, incisit,   +

morfologie   -y ž. z něm., původně z fr.   zobrazit

Uložit   Návrat bez uložení

Done

Outline

1 DEB II Development Platform

2 Current DEB projects
- DEBDict
- DEBVisDic
- DEB CPA
- PRALED
- DEB TeDi

3 Dictionary Administration Interface

4 Conclusions and Future Directions

Introduction

- DEB server deployed in different organizations
    - Institute of Czech Language, University of Amsterdam, Adam Mickiewicz University, Hungarian Academy of Sciences
- easy administration using web interface
- common functions for all services
- same principles as DEB server
    - all the data in XML format, stored in DB XML
    - implemented in Ruby

Basic Management Features

- dictionary management
  - attributes – name, DB XML database, XML schema, XSL templates. . .
  - specific attributes, for example related dictionaries for WordNets
  - journaling
- services
  - groups of dictionaries
- user accounts
  - HTTP authentication mechanism
  - service level or dictionary level
  - immediate changes in all services

name	file	class	root tag	key
Czech Wordnet	wncz6.dbxml	WordNet	SYNSET	/SYNSET/ID

ment-equality   syno-gslo
syno

code	name	file	class	root tag	key	indexes		xslt (nam
wncz	Czech Wordnet	wncz6.dbxml	WordNet	SYNSET	/SYNSET/ID	new:	node-element-equality	new:

LITERAL	node-element-substring
WORD	node-element-equality
WORD	node-element-substring
ILR	node-element-equality

ID	node-element-equality
ID	node-element-substring
LITERAL	node-element-equality
LITERAL	node-element-substring
WORD	node-element-equality
WORD	node-element-substring
ILR	node-element-equality
ILR	edge-element-equality
POS	node-element-equality
POS	edge-element-equality
BCS	node-element-equality
BCS	node-element-substring
SNOTE	node-element-equality

preview
vb
single
xml

code	name	file	class	root tag	key	indexes		xslt (nam
wnen	English Wordnet	wnen6.dbxml	WordNet	SYNSET	/SYNSET/ID	new:	node-element-equality	new:

| ID | node-element-equality |

File   Edit   View   History   Bookmarks   Tools   Help

https://████████████/admin?action=listusers&edituser=adam   G▾ Google

login	name	organization	password
	email	address	comment
adam	Adam Rambousek	Faculty of Informatics	
	xrambous@fi.muni.cz	Botanická 68a Brno	

cpa ☐
* cpa ☐
debdict ☑
* cia r☐ cod11 ☐ diderot r☐ gslov ☐ ode ☐ ote ☐ scfin r☐ scfis ☐ scs
debvisdic
* diderot ☐ scfin ☐ scfis ☐ ...s ☐ ... ... wne
wngre ☐ wnhf ☐ wnpol ☐ w rus ☐
praled ☐
* cia ☐ diderot ☐ gslov ☐ praled ☐ ... ... ssc ☐ ssjc

lmin ☐

save
new pass
del user

SSJČ dictionary - SSJČ
SCS foreign words dictionary
SSČ dictionary - SSČ
SCS thesaurus
šfm dict. - SČFI verbal
šfn dict. - SČFI nonverbal
av all dictionaries
D Concise Oxford English D.
D Oxford Dictionary of English
D Oxford Thesaurus of English
D Diderot
google google
A Answers.com
W Wikipedia
S Seznam Encyklopedie
C CIA World Factbook
map of Czech Republic
morph. analyzer ajka

SCS foreign words dictionary
šfm dict. - SČFI verbal
šfn dict. - SČFI nonverbal
D Diderot
google google
A Answers.com
W Wikipedia
S Seznam Encyklopedie
C CIA World Factbook
map of Czech Republic
morph. analyzer ajka

all users (169)

abican (Aleš Bičan) acerna (An...a Černá) ... (Ágnes Gya... ...cova (Agáta Karčová)
anneu (Anne Urbschat) aoravcova (Adri... ...n) arum (An... ...i) asufliar (Alexandra
Šufliarska) bchocholova (Bronislava ...ho... ...) birkhahn... na Bozděchová, CSc.)
chudomelova (Věra Chudomelová) con... ... arum (CPA... ...05 (CPA05) cpa06 (CPA06)
cpa07 (CPA07) cpa08 (CPA08) cpa09 ... ... majeurakova (Daniela
Majeuráková) duzi (Marie Duží) dvasilis ... ...kova (Dáša Z... rjavec (Tomáš Erjavec)
fcermak (František Čermák) fcvrcek (Fra... ...gerald) fstí... ajdošová) gazi (Gabriel Gaži)
gkotze (Gideon Jozua Kotzé) golanova (Hana Goláňová) gryciuk (Joanna Gryciuk) gszarv... ...rikova (Helena Ivoríková)
hkonecna (Hana Konečná) hlava (Jaroslava Hlaváčová) hlavsem (skupinový učet do 1.10.20... ...va (Hana Palová) hradec (Jan
Hradec) hruskova (Jitka Hrušková) inemcova (Ivana Němcová) isvobodova (Ivana Svobod... ...n Bosák) ichromy (Jan Chromý)

Advanced Management Features

- XML import and export (backup)
- entry locking, sequence identifiers
    - usage depends on application
- file upload (XML for import, XSLT, web files...)
- client package or web service generation according to XML schema
    - only basic functions
    - basis for custom modifications

mars

headword	mars
def	creamy planet filled with caramel	[+]
synonym	[▼] [+]
grammar_note	
@pos	n ▼
@hypernym	planet ▼
pronounce	
origin	
author	emls3d

[Save]   [Back to list]

Outline

1. DEB II Development Platform

2. Current DEB projects
   - DEBDict
   - DEBVisDic
   - DEB CPA
   - PRALED
   - DEB TeDi

3. Dictionary Administration Interface

4. Conclusions and Future Directions

Conclusions and Future Directions

- several client applications under development are used in running projects
- high modularity and configurability
- versatile base for implementation of individual and powerful dictionary writing tools
- 7 servers installed (Brno, Prague, Amsterdam-UvA, Amsterdam-VU, Poznan, Johannesburg, Budapest)
- the Brno server – 233 users from 14 domains/countries