

IB047

Nástroje pro psaní slovníků

Pavel Rychlý

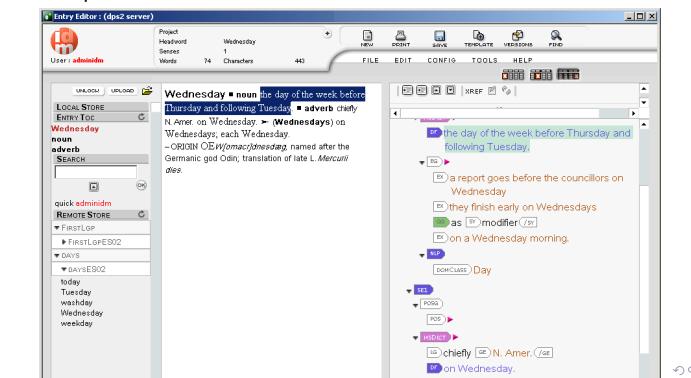
pary@fi.muni.cz

May 13, 2022

Pavel Rychlý IB047

Editory pro velké organizace

- správa uživatelů
- řízení přístupu
- předávání/schvalování



Pavel Rychlý IB047

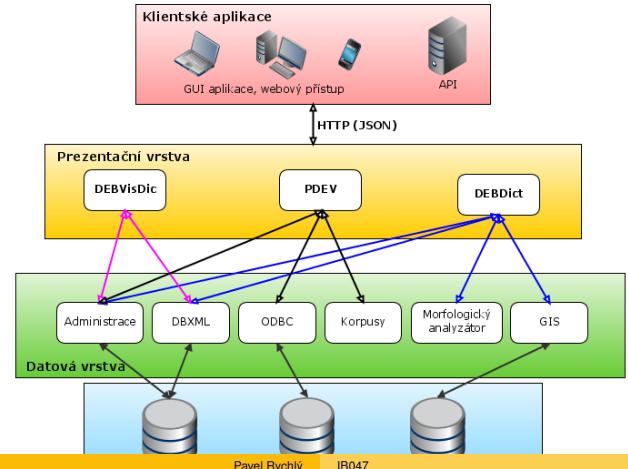
Platforma DEB

DEB – Dictionary Editor and Browser

- platforma pro vývoj slovníkových aplikací
 - všechna data ve formátu XML, Unicode
 - libovolná struktura, jakýkoliv jazyk
- architektura klient-server
- server
 - práce s daty, velká část funkcí
 - databázové úložiště
 - správa uživatelů, slovníků, spolupráce
 - rozděleno do modulů, spojování podle potřeb
- klient
 - omezená funkcionality
 - grafické nebo webové rozhraní

Pavel Rychlý IB047

Platforma DEB



Pavel Rychlý IB047

DEBDict

- prohlížeč slovníků
- 7 slovníků češtiny pro veřejnost, další přístupné jen pro část uživatelů
- napojení na morfologický analyzátor
- data z externích zdrojů
- přes 900 uživatelů (podepsané prohlášení)

Pavel Rychlý IB047

DEBDict

Pavel Rychlý IB047

Praled

Příprava lexikální databáze češtiny v Ústavu pro jazyk český

jazyk

Zdroj pořadí: fsc+sje
Typ hezla: jednotlivé podstatné jméno

Zdroj: EK V1 SUBST+SUBST gen (rozbalit/vložit)

SEZNAMY (rozbalit/vložit) PRQ 15777 ARF 6122 zdroj

SLOVOTVORBA
Dářováná slova

Fundace/motivace

Detrax, leždemy

jazyk

Zdroj pro heslát: fsc+sje
Typ hesla: jednotlivé podstatné jméno

Uzavrk: Plná znění zkratky

Pozn.: k celému heslu

Uložit a zavřít Uložit Zavřít

V1

Fórum definice: SSJC svatný, velmi pohyblivý orgán v dutině ústní (u zvířat v tlámc, zobáku atd.); orgán chuti, mluvy

E1 k V1 Adj+SUBST (rozbalit/vložit)

E2 k V1 SUBST+Adj (rozbalit/vložit)

E3 k V1 SUBST+subj+gen (rozbalit)

E4 k V1 SUBST+SUBST+gen (rozbalit)

E4 k V1 SUBST+SUBST+gen (rozbalit)

Pozn.: k E4 spracovaným substantivum je samoucenitivum

E5 k V1 SUBST+Prep+Subj+SUBST+Subj+ji

Pozn.: k E5 vložit/vložit vložit k poslednímu vložit

ZÁHLED **ADMINT** **VÝSLE** **PŘEDĚL** **DELEN** **EMILY** **STOP**

ZONEKVALIFIKA **SOUZOS** **PRÄZEN** **JEVNICE/VÝRAZ**

VÍCESEL **NEZARAZEN** **KOM-VÍCESEL** **SLOVOTVORBA**

Zavřít Sbalt seznamy Sbalt exempl. Duplikovat Tisk NOVÉ DOKLADY DEF Nahled 1 Nahled 2 Nahled 3 jazyk vlozec Hledat

TeDi – Terminologický slovník

- společný projekt s Fakultou výtvarných umění VUT
 - glosář výtvarných pojmů
 - multimediální prvky
 - nové také Divadelní fakulta JAMU, Agronomická fakulta MU

The screenshot shows the 'Tedi' application window titled 'Edit address (graphic, dedicated) - SeamMonkey'. The menu bar includes File, Edit, View, Go, Bookmarks, Tools, Window, Help. The toolbar includes Home, Back, Forward, Stop, Refresh, and Search. The address bar shows the URL https://apollo.fi.muni.cz:8010/tedi?action=edit&id=adresa_1. The main content area displays the following fields:

- Autor:** ID: **adresa_(grafická_dedikační)-11804343501**
- obor:** **kresba-grafika**
- heslo česky:** **adresa (grafická, dedikační)** (highlighted in red)
- anglicky:** **address**
- německy:** **francouzsky**
- Adresse f.:** **adresse f.**
- varianty:** (empty input field)
- styl, příznak:** **standard, knižní** (highlighted in red)
- definice:** součást značení dila, starých grafických listů obsahující jméno či obchodní označení vydavatele nákladu rytin, leptů atp. Trvale obsažena ve vlastním tisku.
- příklady:** (empty input field)
- nadpojem:** **značení dila**
- podnosem:** (empty input field)

DEBVisDic

The screenshot shows three overlapping windows from the DBpedia-Dic application:

- Top Window:** Title bar says "DBpedia-Dic English WordNet". It has tabs for "code", "Search", and "Search in all". A status bar at the bottom says "Querying a dictionary is complete.".
- Middle Window:** Title bar says "DBpedia-Dic Greek WordNet". It has tabs for "code", "Search", and "Search in all". A status bar at the bottom says "Querying a dictionary is complete. Found 4 items! Items: 1".
- Bottom Window:** Title bar says "DBpedia-Dic German WordNet Top 500". It has tabs for "code", "Search", and "Search in all". A status bar at the bottom says "Querying a dictionary is complete. Found 1 item! Items: 1".

The main content area of each window displays a list of words and their definitions. The "code" tab is selected in all three windows.

Top Window (English):

```

POS: n [ID: EN2G20-06246978-a]
Synonyms: code, codification,2
Definition: a set of rules or principles or laws
Usage: [examples]
Domain: [examples]
SUMO/MIL: Obligation
language1
    >> [eng] derivative: codify,1
    << [hypernym] codify,1
    << [hypernym] Bushido,1
    << [hyperonym] legal_code,1
    << [hyperonym] code_of_civilization,1
    << [hyperonym] dress_code,1
    << [hyperonym] fire_code,1
    << [hyperonym] medical_code,1
    << [hyperonym] sanitary_code,1, health_code,1
    << [hyperonym] Highway_Code,1

```

Middle Window (Greek):

```

POS: n [ID: EN2G20-06246978-a]
Synonyms: κωδικός,2
Definition: οποιαδήποτε κανονώση ή αρχών ή νόμων για την υπαρκεία
Usage: >> [hypernym] γράμμα_επικουρεία,1, λέγκα,1
language1
    >> [eng] derivative: codify,1
    << [hypernym] codify,1
    << [hypernym] Bushido,1
    << [hyperonym] legal_code,1
    << [hyperonym] code_of_civilization,1
    << [hyperonym] dress_code,1
    << [hyperonym] fire_code,1
    << [hyperonym] medical_code,1
    << [hyperonym] sanitary_code,1, health_code,1
    << [hyperonym] Highway_Code,1

```

Bottom Window (German):

```

POS: n [ID: EN2G20-06246978-a]
Synonyms: κωδικός,2
Definition: οποιαδήποτε κανονώση ή αρχών ή νόμων για την υπαρκεία
Usage: >> [hypernym] γράμμα_επικουρεία,1, λέγκα,1
language1
    >> [eng] derivative: codify,1
    << [hypernym] codify,1
    << [hypernym] Bushido,1
    << [hyperonym] legal_code,1
    << [hyperonym] code_of_civilization,1
    << [hyperonym] dress_code,1
    << [hyperonym] fire_code,1
    << [hyperonym] medical_code,1
    << [hyperonym] sanitary_code,1, health_code,1
    << [hyperonym] Highway_Code,1

```

PDEV (CPA)

- editace slovníku vzorů anglických sloves
 - varianty také pro češtinu, italštinu a španělštinu

DEBVisDic

- editor slovníků typu WordNet
 - samostatný modul pro každý jazyk (modifikace)
 - použit pro tvorbu několika wordnetů
 - poskytuje API – napojení externích aplikací
 - možnost rozšíření a modifikací

Pavel Rychlý IB047

Lexonomy

- univerzální editor slovníků
 - slouží i k prezentaci
 - cloud-based = klient v prohlížeči
 - www.lexonomy.eu
 - konfigurace obsahu (struktury) i prezentace
 - integrace se Sketch Engine

A set of small, light-gray navigation icons typically found in Microsoft Word documents, including symbols for back, forward, search, and other document operations.

Budoucnost tvorby slovníků

- předgenerování z korpusu
- anotace
- posteditace

Pavel Rychlý IB047

Examples

A good example must be:

- typical, exhibiting frequent and well-dispersed patterns of usage
- informative, helping to elucidate the definition
- readability
 - intelligible to learners,
 - avoiding gratuitously difficult lexis and structures, puzzling or distracting names, anaphoric references,
 - can be understood without access to the wider context.

Pavel Rychlý IB047

Domains/usage

- *usually spoken, business*
- domain annotation in corpus
- *only in plurals*
- condition specification + threshold
 - values of a structure attribute
 - subcorpus
 - query (tags)

```
=plurals
HR plural
Q1 [lempos="%s" & tag="NN2"]
Q2 [lempos="%s" & tag="NN1"]
RE -n$
```

Pavel Rychlý IB047

Informations from Corpora

- multi-word expressions (MWE)
- collocatons
- thesaurus
- domains
- examples
- translations

Pavel Rychlý IB047

GDEX – Good Dictionary EXamples

- sentence length: 10 – 25 words, longer/shorter penalized
- word frequencies: non common words (top 17,000) penalized
- pronouns and anaphors penalized
- target collocation in the main clause preferred
- whole sentence: beginning with a capital letter and ending with .!?

Pavel Rychlý IB047