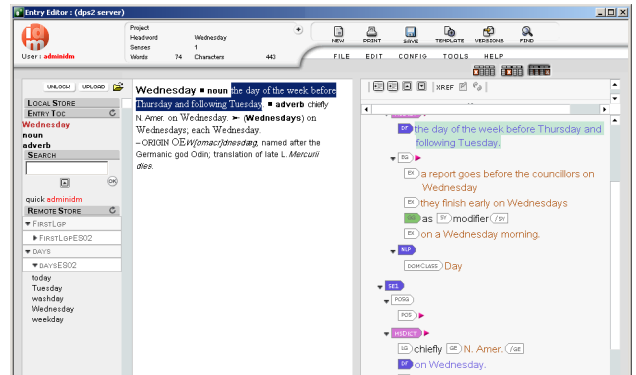


Editory pro velké organizace

- správa uživatelů
- řízení přístupu
- předávání/schvalování



IB047 Nástroje pro psaní slovníků

Pavel Rychlý

pary@fi.muni.cz

May 13, 2022

Pavel Rychlý IB047

Pavel Rychlý IB047

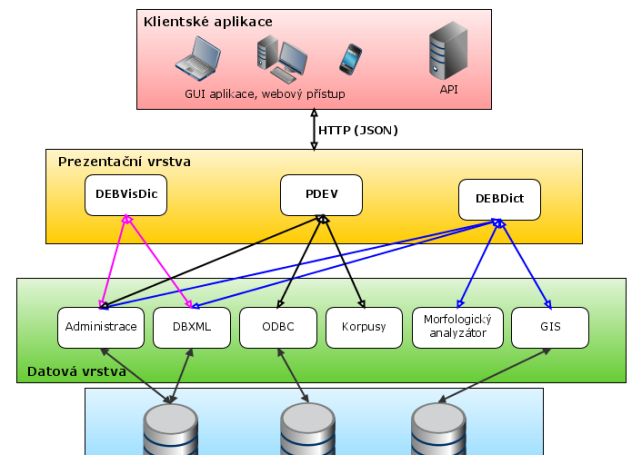
Platforma DEB

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 funkcionalita
 - grafické nebo webové rozhraní

Pavel Rychlý IB047

Platforma DEB



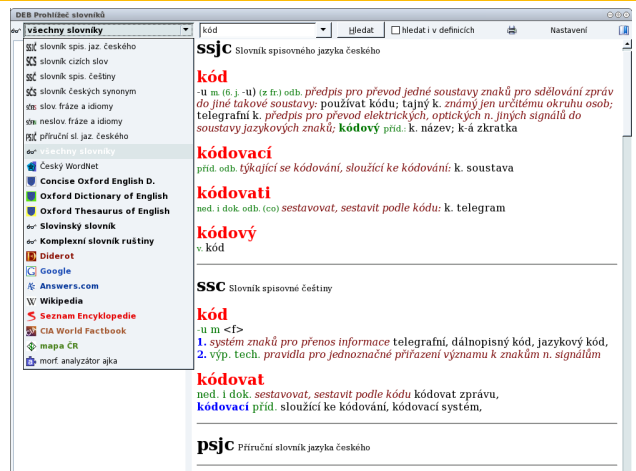
Pavel Rychlý IB047

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

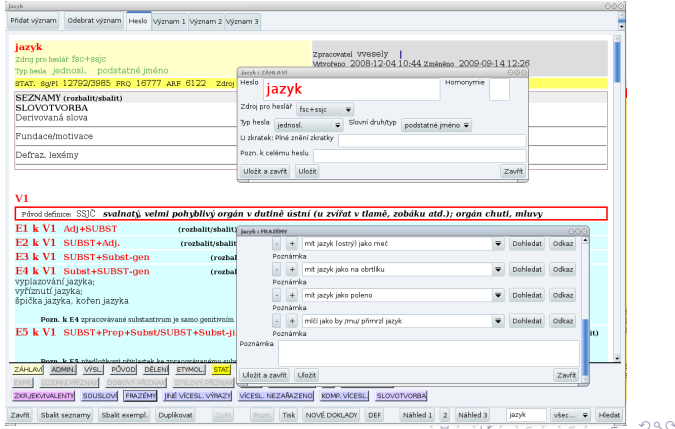
DEBDict



Pavel Rychlý IB047

Praled

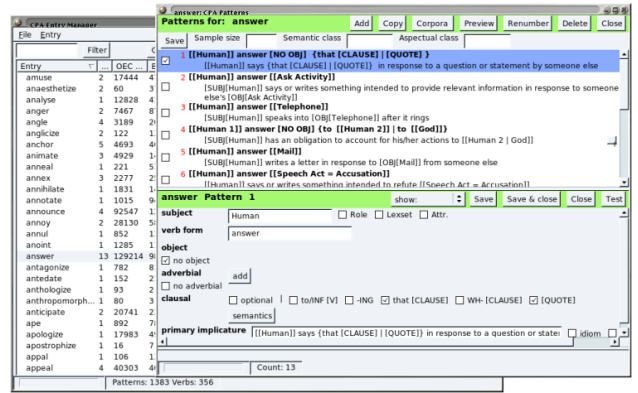
Příprava lexikální databáze češtiny v Ústavu pro jazyk český



Pavel Rychlý IB047

PDEV (CPA)

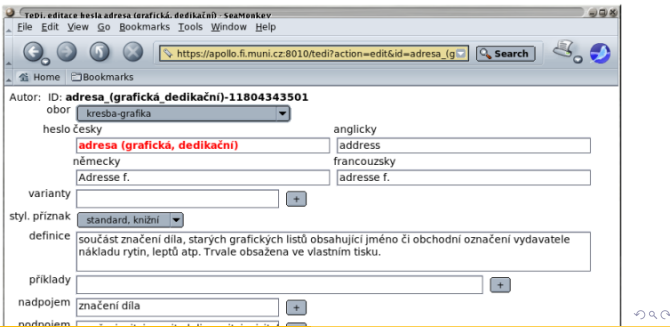
- editace slovníku vzorů anglických sloves
- varianty také pro češtinu, itaštinu a španělštinu



Pavel Rychlý IB047

TeDi – Terminologický slovník

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



Pavel Rychlý IB047

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

DEBVisDic

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

Pavel Rychlý IB047

Budoucnost tvorby slovníků

- předgenerování z korpusu
- anotace
- posteditace

Pavel Rychlý IB047

Informations from Corpora

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

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

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

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