IB047

Nástroje pro psaní slovníků

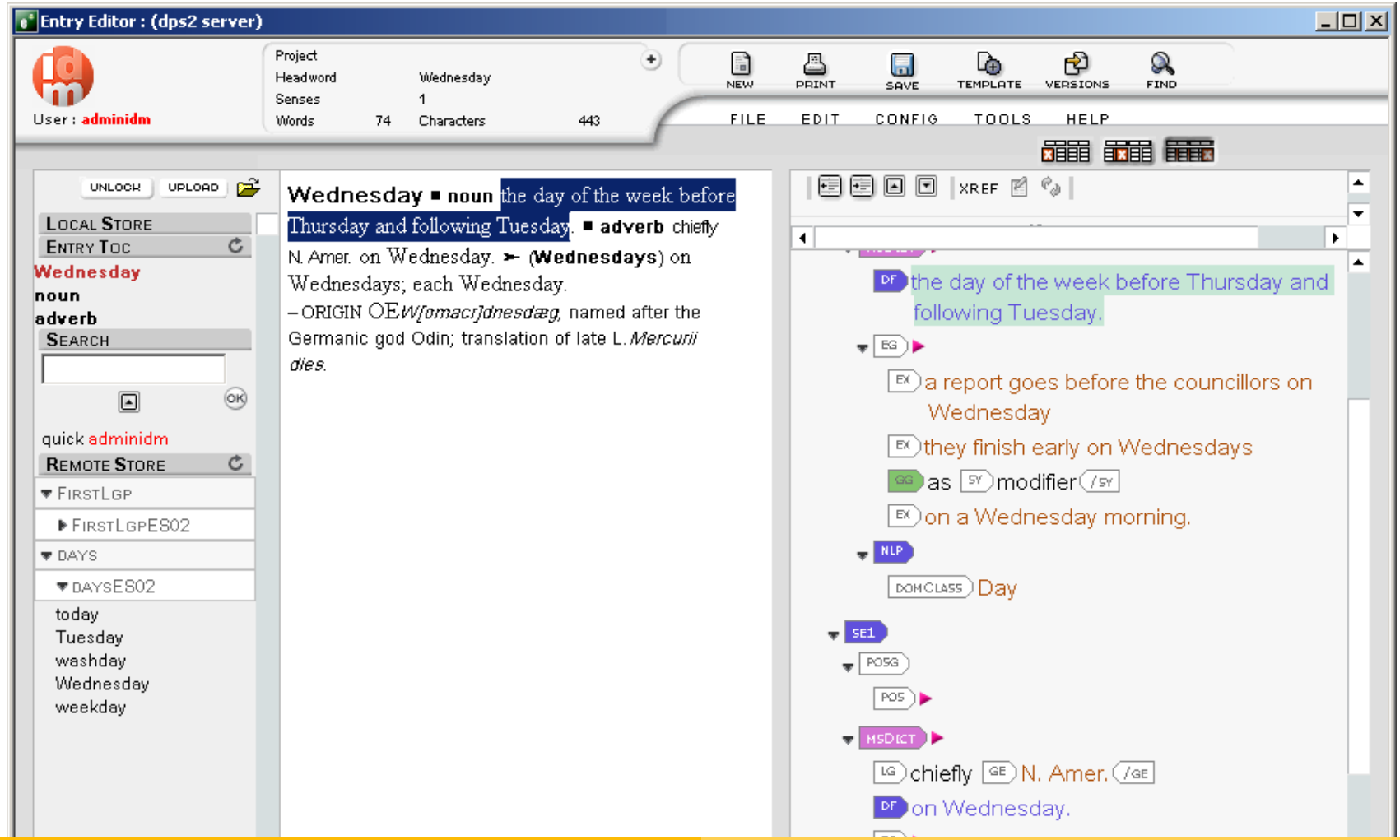
Pavel Rychlý

pary@fi.muni.cz

May 13, 2022

Editory pro velké organizace

- správa uživatelů
- řízení přístupu
- předávání/schvalování

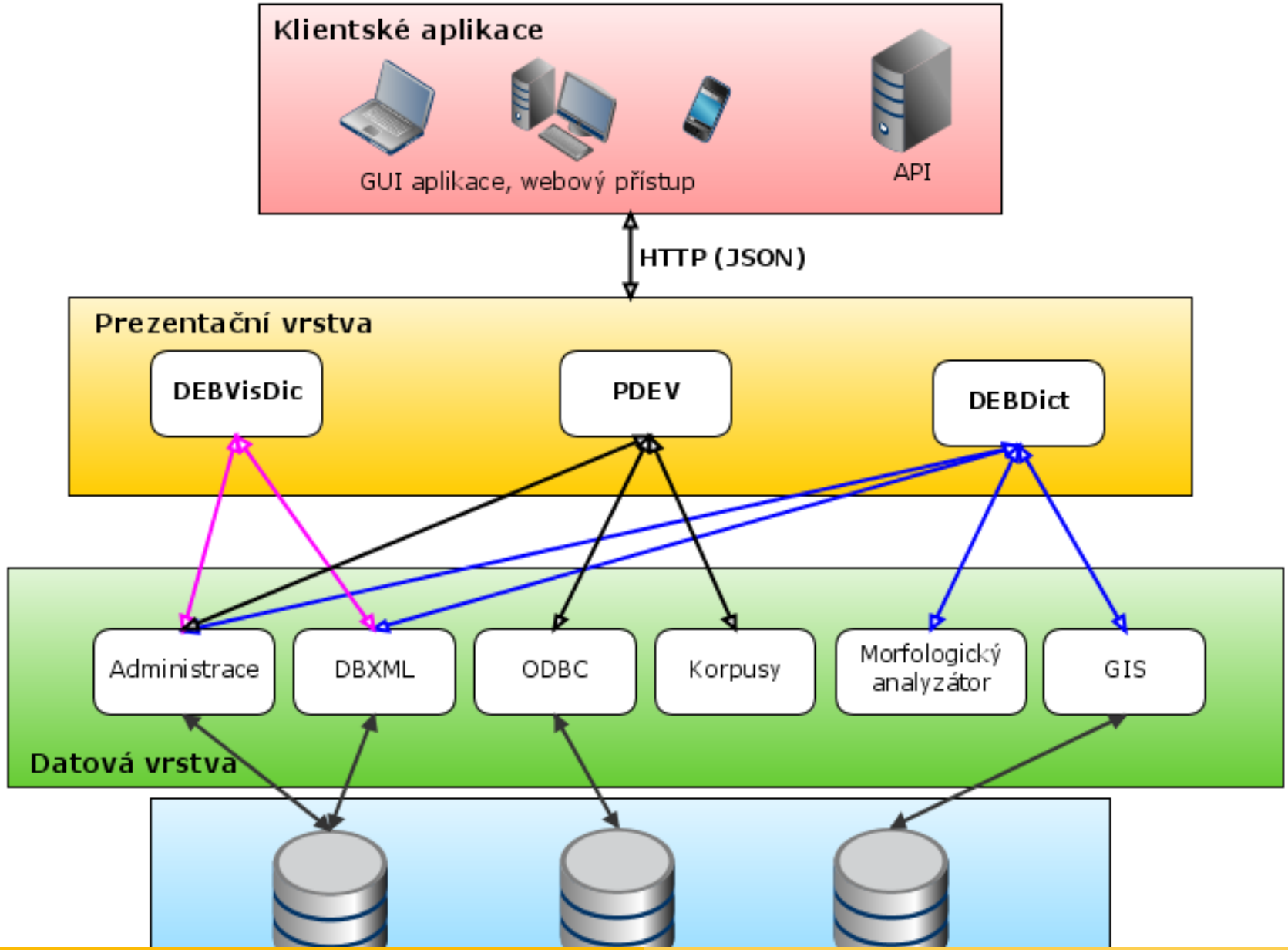


Platforma DEB

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

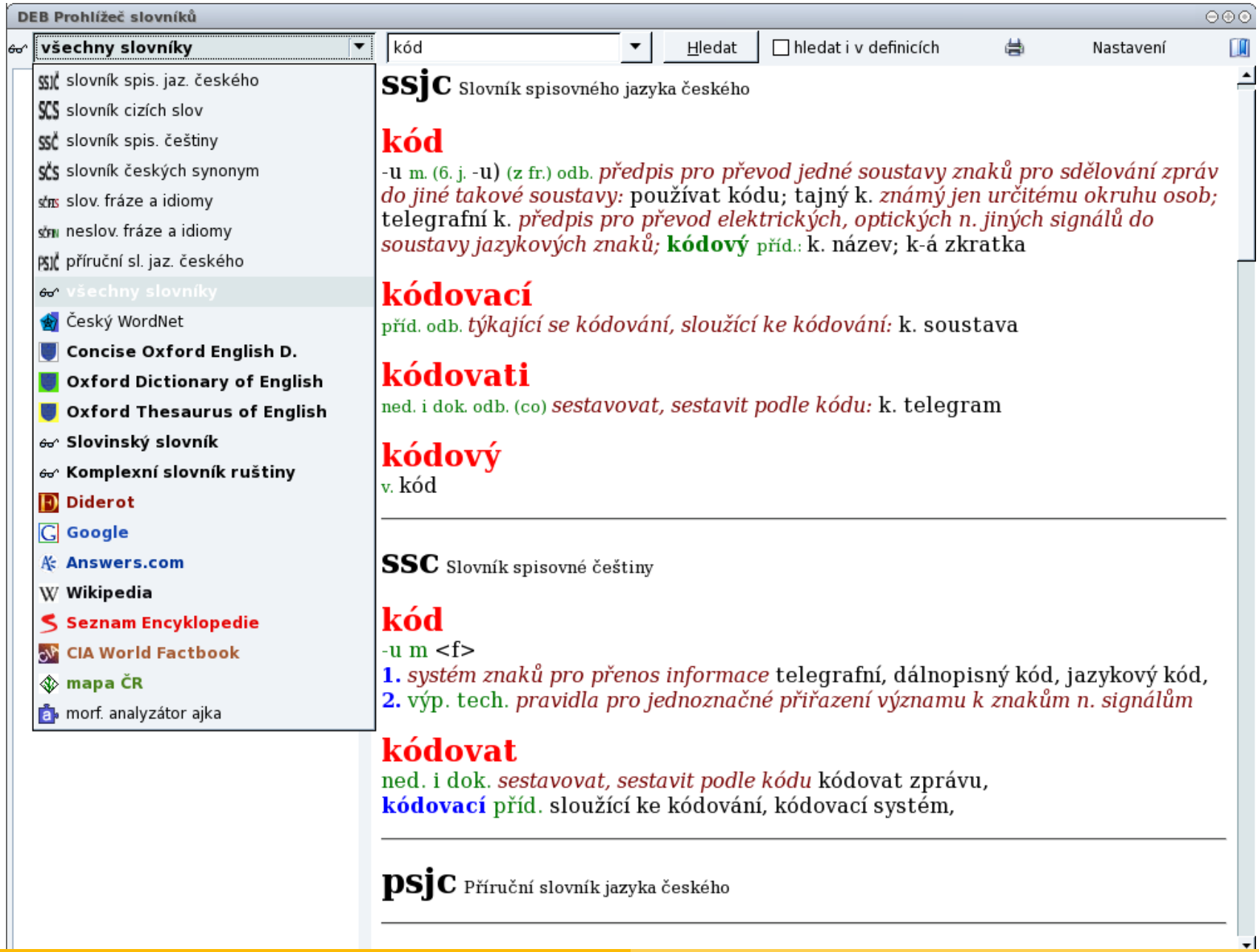
Platforma DEB



DEBDict

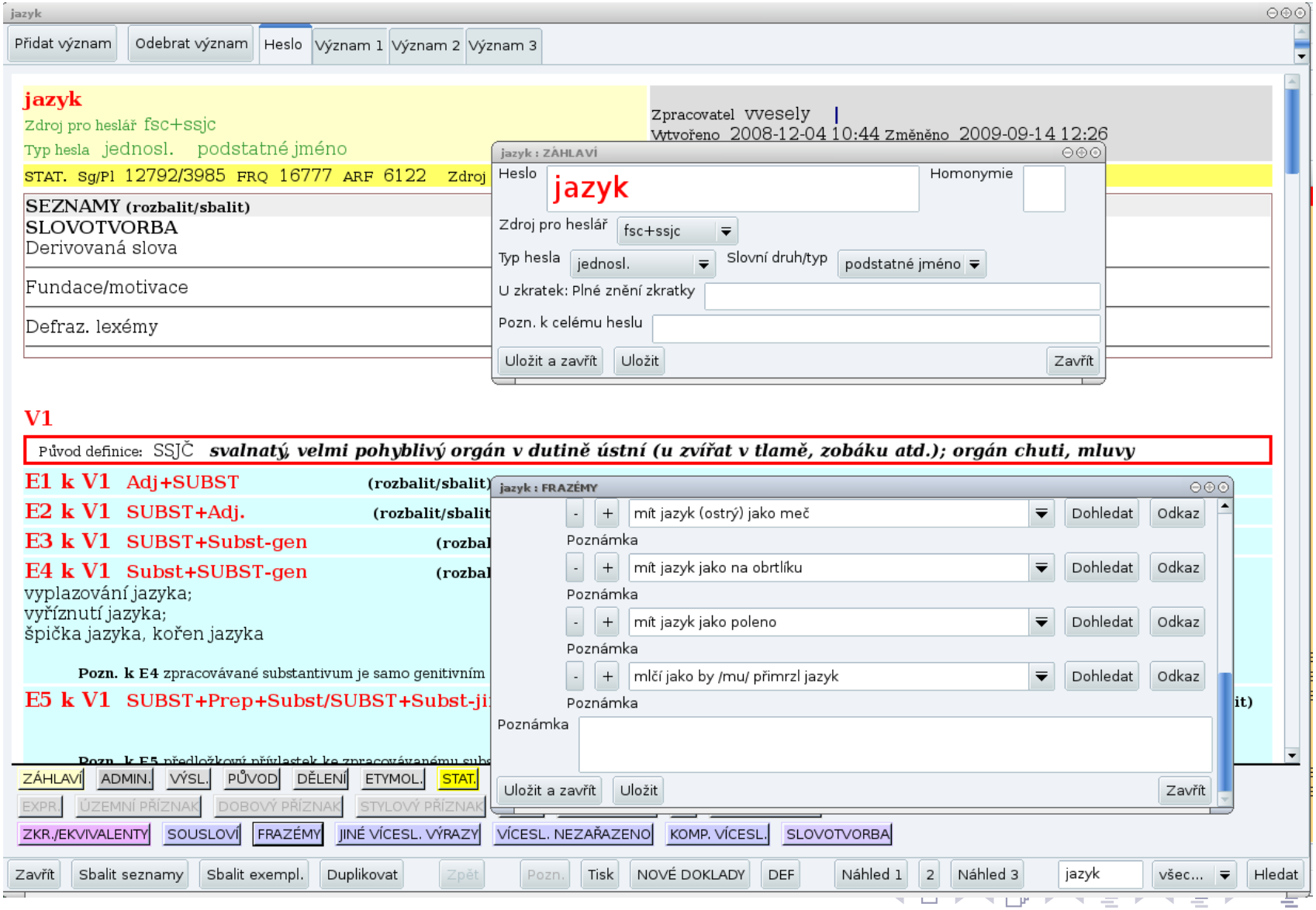
- prohlížeč slovníků
- 7 slovníků češtiny pro veřejnost, další přístupné jen pro část uživatelů
- napojení na morfologický analyzátor
- data z externích zdrojů
- přes 900 uživatelů (podepsané prohlášení)

DEBDict


DEB Prohlížeč slovníků
všechny slovníky
kód
Hledat
hledat i v definicích
Nastavení
slovník spis. jaz. českého
slovník cizích slov
slovník spis. češtiny
slovník českých synonym
slov. fráze a idiomy
neslov. fráze a idiomy
příruční sl. jaz. českého
všechny slovníky
Český WordNet
Concise Oxford English D.
Oxford Dictionary of English
Oxford Thesaurus of English
Slovinský slovník
Komplexní slovník ruštiny
Diderot
Google
Answers.com
Wikipedia
Seznam Encyklopedie
CIA World Factbook
mapa ČR
morf. analyzátor ajka
ssjc Slovník spisovného jazyka českého
kód
-u m. (6. j. -u) (z fr.) odb. předpis pro převod jedné soustavy znaků pro sdělování zpráv do jiné takové soustavy: používat kódu; tajný k. známý jen určitému okruhu osob; telegrafní k. předpis pro převod elektrických, optických n. jiných signálů do soustavy jazykových znaků; kódový příd.: k. název; k-á zkratka
kódovací
příd. odb. týkající se kódování, sloužící ke kódování: k. soustava
kódovati
ned. i dok. odb. (co) sestavovat, sestavit podle kódu: k. telegram
kódový
v. kód
ssc Slovník spisovné češtiny
kód
-u m <f>
1. systém znaků pro přenos informace telegrafní, dálnopisný kód, jazykový kód,
2. výp. tech. pravidla pro jednoznačné přiřazení významu k znakům n. signálům
kódovat
ned. i dok. sestavovat, sestavit podle kódu kódovat zprávu,
kódovací příd. sloužící ke kódování, kódovací systém,
psjc Příruční slovník jazyka českého

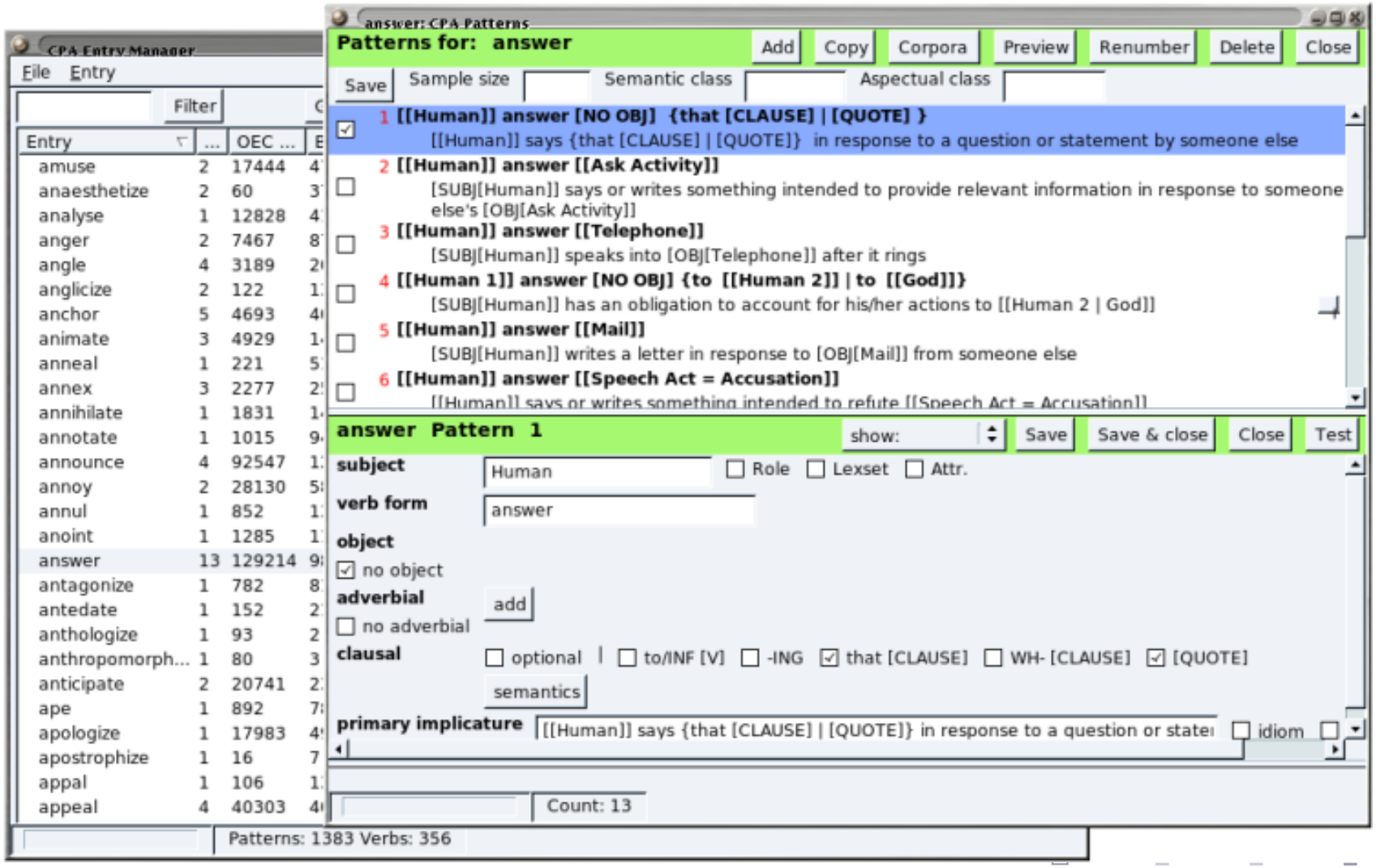
Praled

Příprava lexikální databáze češtiny v Ústavu pro jazyk český



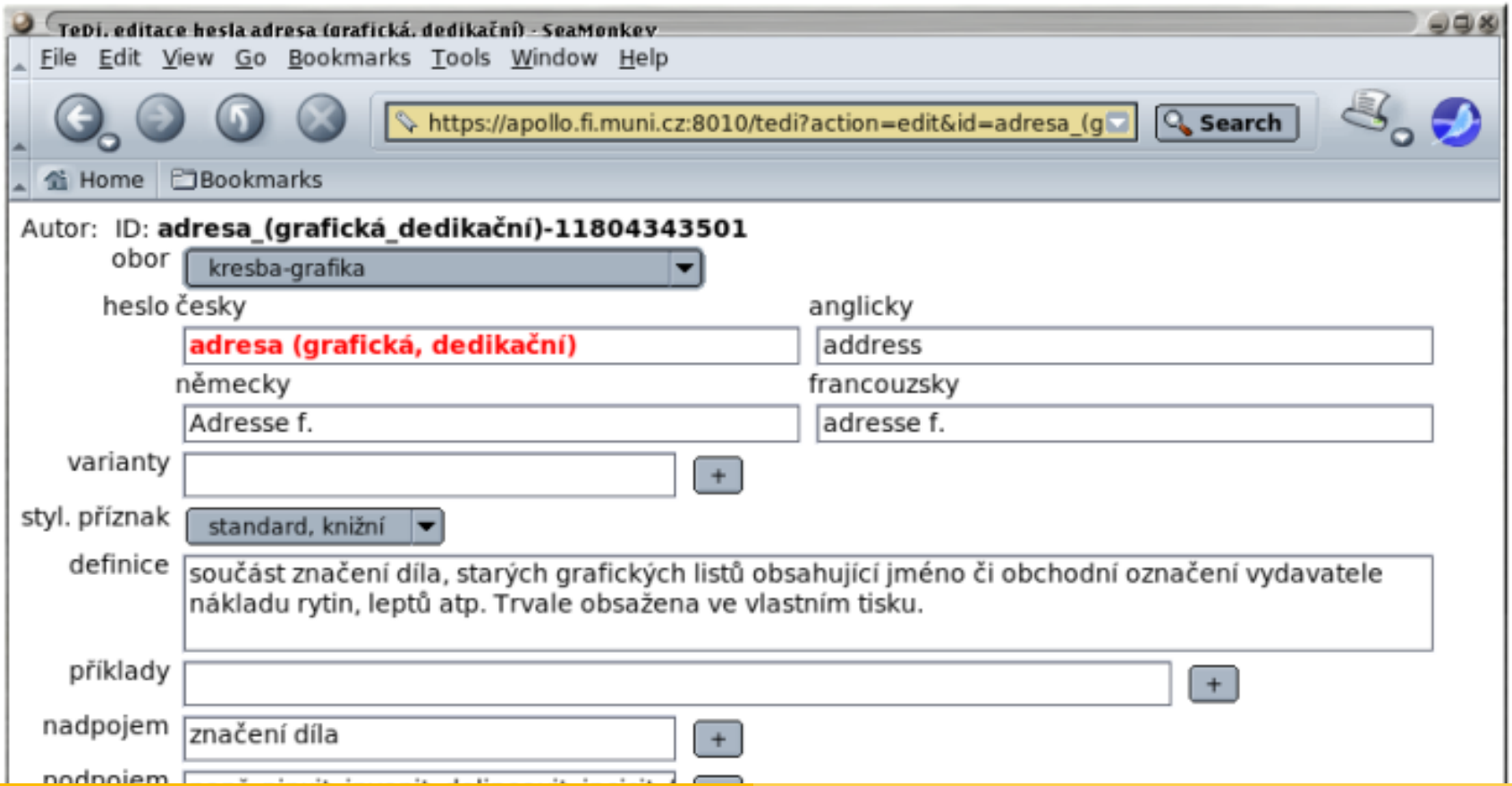
PDEV (CPA)

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



TeDi – Terminologický slovník

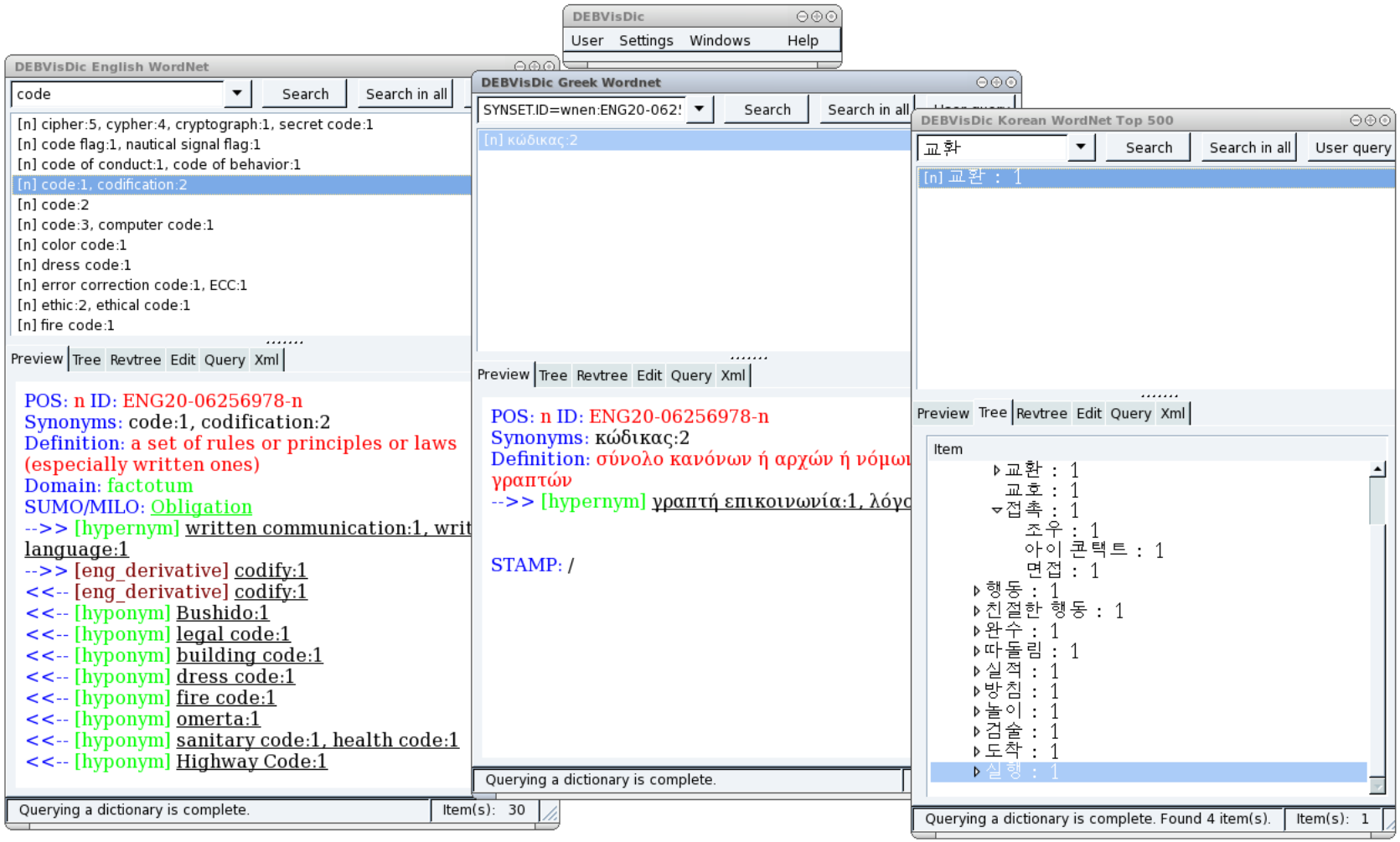
- společný projekt s Fakultou výtvarných umění VUT
- glosář výtvarných pojmů
- multimediální prvky
- nově také Divadelní fakulta JAMU, Agronomická fakulta MU



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

DEBVisDic


DEBVisDic
User Settings Windows Help
DEBVisDic English WordNet
code
Search
Search in all
[n] cipher:5, cypher:4, cryptograph:1, secret code:1
[n] code flag:1, nautical signal flag:1
[n] code of conduct:1, code of behavior:1
[n] code:1, codification:2
[n] code:2
[n] code:3, computer code:1
[n] color code:1
[n] dress code:1
[n] error correction code:1, ECC:1
[n] ethic:2, ethical code:1
[n] fire code:1
Preview Tree Revtree Edit Query Xml
POS: n ID: ENG20-06256978-n
Synonyms: code:1, codification:2
Definition: a set of rules or principles or laws (especially written ones)
Domain: factotum
SUMO/MILO: Obligation
-->> [hypernym] written communication:1, writ
language:1
-->> [eng_derivative] codify:1
<<-- [eng_derivative] codify:1
<<-- [hyponym] Bushido:1
<<-- [hyponym] legal code:1
<<-- [hyponym] building code:1
<<-- [hyponym] dress code:1
<<-- [hyponym] fire code:1
<<-- [hyponym] omerta:1
<<-- [hyponym] sanitary code:1, health code:1
<<-- [hyponym] Highway Code:1
Querying a dictionary is complete.
Item(s): 30
DEBVisDic Greek Wordnet
SYNSET.ID=wnen:ENG20-062!
Search
Search in all
[n] κώδικας:2
Preview Tree Revtree Edit Query Xml
POS: n ID: ENG20-06256978-n
Synonyms: κώδικας:2
Definition: σύνολο κανόνων ή αρχών ή νόμω
γραπτών
-->> [hypernym] γραπτή επικοινωνία:1, λόγ
STAMP: /
Querying a dictionary is complete.
DEBVisDic Korean WordNet Top 500
교환
Search
Search in all
User query
[n] 교환 : 1
Preview Tree Revtree Edit Query Xml
Item
교환 : 1
교호 : 1
접촉 : 1
조우 : 1
아이 콘택트 : 1
면접 : 1
행동 : 1
친절한 행동 : 1
완수 : 1
따돌림 : 1
실적 : 1
방침 : 1
놀이 : 1
검술 : 1
도착 : 1
실행 : 1
Querying a dictionary is complete. Found 4 item(s).
Item(s): 1

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

Budoucnost tvorby slovníků

- předgenerování z korpusu
- anotace
- posteditace

Informations from Corpora

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

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.

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 .!?

Domains/usage

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

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