

IB047

Nástroje pro psaní slovníků

Pavel Rychlý

pary@fi.muni.cz

May 13, 2022

Editor pro velké organizace

- správa uživatelů
- řízení přístupu
- předávání/schvalování

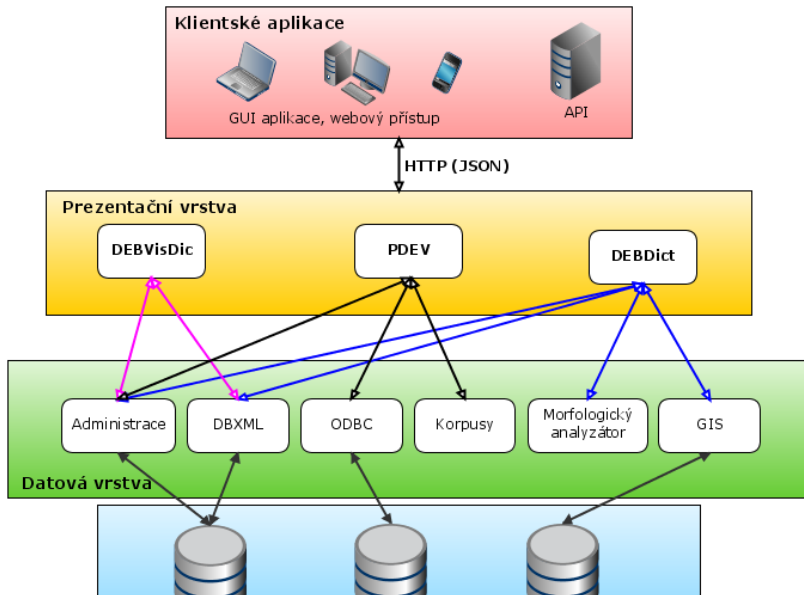
The screenshot displays the 'Entry Editor' application window. The title bar reads 'Entry Editor : (dps2 server)'. The interface is divided into several sections:

- Project Information:** Project: Wednesday, Headword: Wednesday, Senses: 1, Words: 74, Characters: 443.
- Navigation and Tools:** A menu bar with FILE, EDIT, CONFIG, TOOLS, and HELP. A toolbar contains icons for NEW, PRINT, SAVE, TEMPLATE, VERSIONS, and FIND.
- Left Panel:** Contains a search bar, a 'LOCAL STORE' section with 'ENTRY TOC', and a 'REMOTE STORE' section with a tree view showing folders like 'FIRSTLGP' and 'DAYS', and sub-items like 'today', 'Tuesday', 'washday', 'Wednesday', and 'weekday'.
- Main Content Area:** Displays the dictionary entry for 'Wednesday':
 - Wednesday** ■ **noun** the day of the week before Thursday and following Tuesday. ■ **adverb** chiefly N. Amer. on Wednesday. ► (**Wednesdays**) on Wednesdays; each Wednesday.
 - ORIGIN OE *W[onac]dnesdæg*, named after the Germanic god Odin, translation of late L. *Mercurii dies*.
- Right Panel:** Shows a morphological analysis tree for the word 'Wednesday':
 - DF: the day of the week before Thursday and following Tuesday.
 - ES: a report goes before the councillors on Wednesday
 - EX: they finish early on Wednesdays
 - ES: as SY modifier /SY
 - EX: on a Wednesday morning.
 - NLP: DOHCLASS Day
 - SET: POSG: POS
 - HSICT: LG chiefly GE N. Amer. /GE
 - DF: on Wednesday.

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

Platforma DEB



- 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)

DEB Prohlížeč slovníků

všechny slovníky

- SSJ slovník spis. jaz. českého
- SSC slovník cizích slov
- SSČ slovník spis. češtiny
- SSS slovník českých synonym
- slm slov. fráze a idiomy
- slm neslov. fráze a idiomy
- PSJ příruční sl. jaz. českého
- všechny slovníky
- Český WordNet
- Concise Oxford English D.
- Oxford Dictionary of English
- Oxford Thesaurus of English
- Slovníkový slovník
- Komplexní slovník ruštiny
- Diderot
- Google
- Answers.com
- Wikipedia
- Seznam Encyklopedie
- CIA World Factbook
- mapa ČR
- morf. analyzátor ajka

kód Hledat hledat i v definicích Nastavení

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** před. k. název; k-á zkratka

kódovací
před. 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 před. sloužící ke kódování, kódovací systém,

psjc Příruční slovník jazyka českého

Příprava lexikální databáze češtiny v Ústavu pro jazyk český

jazyk
 Zdroj pro heslář fsc+ssjc
 Typ hesla jednosl. podstatné jméno
 STAT. Sg/Pl 12792/3985 FRQ 16777 ARF 6122 Zdroj

SEZNAMY (rozbalit/sbalit)
 SLOVOTVORBA
 Derivovaná slova

Fundace/motivace

Defraz. lexémy

Zpracovatel vvesely |
 Vytvořeno 2008-12-04 10:44 změněno 2009-09-14 12:26

jazyk Homonymie

Zdroj pro heslář fsc+ssjc
 Typ hesla jednosl. Slovní druh/typ podstatné jméno
 U zkratk: Plné znění zkratky
 Pozn. k celému heslu

Uložit a zavřít Uložit Zavřít

V1
 Původ definice: SSJČ *svalnatý velmi pohyblivý orgán v dutině ústní (u zvířat v tlamě, zobáku atd.); orgán chuti, mluvy*

E1 k V1 Adj+SUBST (rozbalit/sbalit)
E2 k V1 SUBST+Adj. (rozbalit/sbalit)
E3 k V1 SUBST+Subst-gen (rozbalit/sbalit)
E4 k V1 Subst+SUBST-gen (rozbalit/sbalit)
 vyplazování jazyka;
 vyfíznutí jazyka;
 špička jazyka, kořen jazyka

Pozn. k E4 zpracováváné substantivum je samo genitivním

E5 k V1 SUBST+Prep+Subst/SUBST+Subst-ji

Pozn. k E5 předložkami přibírajícím ke zpracovávanému substantivu

ZÁHLAVÍ ADMIN. VÝSL. PŮVOD DĚLENÍ ETYMOL. STAT.
 EXPL. ÚJEMNÍ PRÍZNAK. ROBOVÝ PRÍZNAK. STYLOVÝ PRÍZNAK.
 ZKR./EKVIVALENTY SOUSLOVÍ FRAZÉMY JINÉ VÍCESL. VÝRAZY VÍCESL. NEZAŘAZENO KOMP. VÍCESL. SLOVOTVORBA

Zavřít Sbalit seznamy Sbalit exempl. Duplikovat Zpět Pozn. Tisk NOVÉ DOKLADY DEF Náhled 1 2 Náhled 3 jazyk všec... Hledat

jazyk : FRAZÉMY

- + mit jazyk (ostrý) jako meč Dohledat Odkaz
 Poznámka
 - + mit jazyk jako na obrtliku Dohledat Odkaz
 Poznámka
 - + mit jazyk jako poleno Dohledat Odkaz
 Poznámka
 - + mlčí jako by /mu/ přimrzl jazyk Dohledat Odkaz
 Poznámka

Uložit a zavřít Uložit Zavřít

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

The screenshot displays the CPA Entry Manager interface. On the left is a list of verb entries with columns for the verb, frequency, and other attributes. The 'answer' entry is highlighted. On the right, the 'answer: CPA Patterns' window shows a list of six patterns for the verb 'answer'. The first pattern is selected and expanded to show its grammatical details.

CPA Entry Manager

| Entry | Filter | OEC | E |
|------------------|--------|--------|---|
| amuse | 2 | 17444 | 4 |
| anaesthetize | 2 | 60 | 3 |
| analyse | 1 | 12828 | 4 |
| anger | 2 | 7467 | 8 |
| angle | 4 | 3189 | 2 |
| anglicize | 2 | 122 | 1 |
| anchor | 5 | 4693 | 4 |
| animate | 3 | 4929 | 1 |
| anneal | 1 | 221 | 5 |
| annex | 3 | 2277 | 2 |
| annihilate | 1 | 1831 | 1 |
| annotate | 1 | 1015 | 9 |
| announce | 4 | 92547 | 1 |
| annoy | 2 | 28130 | 5 |
| annul | 1 | 852 | 1 |
| anoint | 1 | 1285 | 1 |
| answer | 13 | 129214 | 9 |
| antagonize | 1 | 782 | 8 |
| antedate | 1 | 152 | 2 |
| anthologize | 1 | 93 | 2 |
| anthropomorph... | 1 | 80 | 3 |
| anticipate | 2 | 20741 | 2 |
| ape | 1 | 892 | 7 |
| apologize | 1 | 17983 | 4 |
| apostrophize | 1 | 16 | 7 |
| appal | 1 | 106 | 1 |
| appeal | 4 | 40303 | 4 |

answer: CPA Patterns

Patterns for: **answer** Add Copy Corpora Preview Renumber Delete Close

Save Sample size Semantic class Aspectual class

- 1** **[[Human]] answer [NO OBJ] (that [CLAUSE] | [QUOTE])**
[[Human]] says (that [CLAUSE] | [QUOTE]) in response to a question or statement by someone else
- 2** **[[Human]] answer [[Ask Activity]]**
[SUB][Human] says or writes something intended to provide relevant information in response to someone else's [OBJ][Ask Activity]
- 3** **[[Human]] answer [[Telephone]]**
[SUB][Human] speaks into [OBJ][Telephone] after it rings
- 4** **[[Human 1]] answer [NO OBJ] (to [[Human 2]] | to [[God]])**
[SUB][Human] has an obligation to account for his/her actions to [[Human 2 | God]]
- 5** **[[Human]] answer [[Mail]]**
[SUB][Human] writes a letter in response to [OBJ][Mail] from someone else
- 6** **[[Human]] answer [[Speech Act = Accusation]]**
[[Human]] says or writes something intended to refute [[Speech Act = Accusation]]

answer Pattern 1 show: Save Save & close Close Test

subject Human Role Lexset Attr.

verb form answer

object no object

adverbial add no adverbial

clausal optional | to/INF [V] -ING that [CLAUSE] WH- [CLAUSE] [QUOTE]

primary implicature semantics idiom

[[Human]] says (that [CLAUSE] | [QUOTE]) in response to a question or state: idiom

Count: 13

Patterns: 1383 Verbs: 356

TeDi – Terminologický slovník

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

TeDi, editace hesla adresa (grafická, dedikační) - Seamonkey

File Edit View Go Bookmarks Tools Window Help

https://apollo.fi.muni.cz:8010/teDi?action=edit&id=adresa_g Search

Home Bookmarks

Autor: ID: **adresa (grafická, dedikační)-11804343501**

obor

heslo česky anglicky

německy francouzsky

varianty +

styl. příznak

definice

příklady +

nadpojem +

podpojem

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

The image displays three overlapping windows of the DEBVisDic application, each showing a different wordnet dictionary. The top window is titled "DEBVisDic" and has a menu bar with "User", "Settings", "Windows", and "Help".

The left window, titled "DEBVisDic English WordNet", shows a search for "code". The results list various codes such as "cipher", "flag", "conduct", "codification", "color", "dress", "error correction", "ethic", and "fire". The selected item is "code:1, codification:2". The definition is "a set of rules or principles or laws (especially written ones)". The domain is "factotum". The SUMO/MILO is "Obligation". The hypernym is "written communication:1, written language:1". The eng_derivative is "codify:1". The hyponym is "Bushido:1". The legal code is "building code:1, dress code:1, fire code:1, omerta:1, sanitary code:1, health code:1, Highway Code:1". The status bar shows "Querying a dictionary is complete." and "Item(s): 30".

The middle window, titled "DEBVisDic Greek Wordnet", shows a search for "κωδικός:2". The definition is "σύλλογος κανόνων ή αρχών ή νόμων γραπτών". The hypernym is "[hyponym] γραπτή επικοινωνία:1, λόγος". The status bar shows "Querying a dictionary is complete." and "Item(s): 1".

The right window, titled "DEBVisDic Korean WordNet Top 500", shows a search for "교화". The status bar shows "Querying a dictionary is complete. Found 4 item(s). Item(s): 1".

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

- předgenerování z korpusu
- anotace
- posteditace

- 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 .!?

- *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$
```