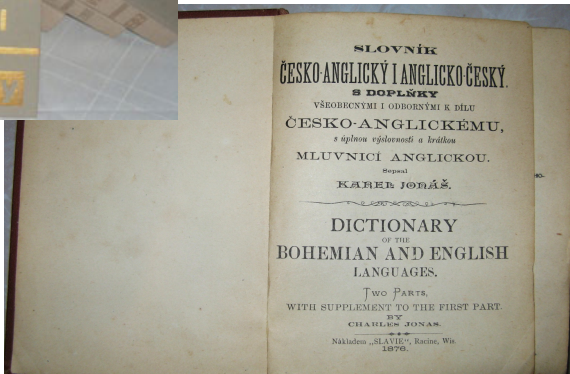


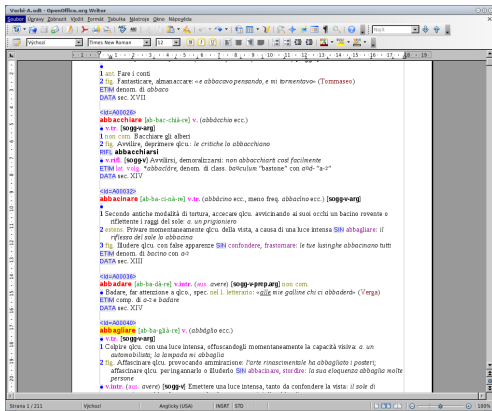
Dictionary Writing Systems

Adam Rambousek

Natural Language Processing Center
Faculty of Informatics, Masaryk University
`xrambous@fi.muni.cz`







Dictionaries

- digital formats
 - typesetting
 - electronic versions (CD, web)
 - further processing in language applications
- no widely used standard
- custom SGML, XML formats

TEI: Text Encoding Initiative

- founded in 1987
- guidelines for the digital encoding of literary and linguistic texts
- publishing *TEI Guidelines* (1st version in 1990)
- current version *P5 Guidelines* published in 2007
 - new support for multimedia, graphics, people and places data
 - interoperability with other XML formats
- modular, customizable
- TEI Lite “designed to meet 90 % of the needs of 90 % of the TEI user community”.

XDXF: XML Dictionary Exchange Format

- founded in 2004 to unify formats of freely available dictionaries
- supported by the most of open-source dictionary viewers
- over 300 bilingual and explanatory dictionaries

LMF: Lexical Markup Framework

- ISO 24613:2008
- developed since 2003, published in 2008
- “representing data in lexical databases used with monolingual and multilingual computer applications”
- main objective is machine processing and data exchange
- modular, extensible, connected with other ISO standards
- extended information: one XML tag with *name* and *value*

IDM DPS

- commercial, developed by IDM (France)
- “suitable for monolingual and bilingual dictionaries of any type and can also be used to produce a thesaurus, a biographical dictionary, a dictionary of quotations or an encyclopaedia”
- used by Oxford University Press and Cambridge University Press
- client-server application, both parts Windows only

Entry Editor : (dps2 server)

User: adminidm

Project: Wednesday
Headword: Wednesday
Senses: 1
Words: 74 Characters: 443

NEW PRINT SAVE TEMPLATE VERSIONS FIND

FILE EDIT CONFIG TOOLS HELP

UNLOCK UNLOAD

LOCAL STORE ENTRY TOC

Wednesday
noun
adverb

SEARCH

quick adminidm

REMOTE STORE

FIRSTLGP

FIRSTLGPES02

DAYS

DAYSES02

today
Tuesday
Wednesday
Wednesday
weekday

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[odnes]dææg*, named after the Germanic god Odin; translation of late L. *Mercurii dies*.

the day of the week before Thursday and following Tuesday.

EG a report goes before the councillors on Wednesday

EX they finish early on Wednesdays

as SV modifier /sv

EX on a Wednesday morning.

NLP

DOMCLASS Day

SET

POSS

POS

HSDICT

LG chiefly GE N. Amer. /at

CF on Wednesday.

EG

EX see you Wednesday.

HSDICT

FG

F Wednesdays

CF on Wednesdays; each Wednesday.

RS

- independent modules:
 - Central Database/Repository
 - documents stored in XML format, optionally with DTD
 - implemented in Oracle, MS SQL, or PostgreSQL
 - Authoring XML Editor
 - client application for authors
 - showing tree structure of XML document and XSLT generated preview
 - download the documents for off-line editing
 - Search Engine
 - custom query language, XML structure, regular expressions, lemmatization...
 - Work Allocation and Workflow Manager
 - managing team work, batch editing
 - Microsoft Project import/export
 - Proofing Tool
 - generate PDF for proofing
 - Adobe InDesign support for automatic corrections import

iLEX

- commercial, developed by Erlandsen Media Publishing (Denmark)
- current version 2.2, January 2010
- price: 400 to 7000 EUR
- developed in Java, supported on Windows, Linux and MacOS
- client-server or standalone application
- used by Royal Swedish Academy of Sciences to compile Svenska Akademiens Ordbok

The screenshot displays the iLEX software interface with three main panes:

- Left Pane (MW-coil):** A list of documents, with 'house' selected at the bottom.
- Middle Pane (Selected MW-coil document):** Shows the dictionary entry for 'chamber'.

1 chamber \ˈtʃæm.bər/ n [ME *chambre*, fr. AF, fr. LL *camera*, fr. L, arched roof, fr. Gk *kamara* vault] (13c) **1:** ROOM; esp: BEDROOM **2:** a natural or artificial enclosed space or cavity **3 a:** a hall for the meetings of a deliberative, legislative, or judicial body «the senate → **b:** a room where a judge transacts business — *una*, used in pl. **c:** the reception room of a person of rank or authority **4 a:** a legislative or judicial body; esp: either of the houses of a bicameral legislature **b:** a rotary board or council **5 a:** the part of the bore of a gun that holds the charge **b:** a compartment in the cartridge cylinder of a revolver — **chambered** \-berd/ adj

2 chamber or chambered, cham-ber-ing \-b(ə)-rɪŋ/ (1575) **1:** to place in or as if in a chamber **2:** to serve as a chamber for; esp: to accommodate in the chamber of a firearm

3 chamber adj (1706) being, relating to, or performing chamber music
- Right Pane (Selected MW-coil document):** Shows the dictionary entry for 'house'.

1 house \ˈhaʊs/ n, pl **houses** \ˈhaʊz-əz/ also -sɪz/ (often attrib) [ME *hous*, fr. OE *hūs*: akin to OHG *hūs* house] (bef. 12c) **1:** a building that serves as living quarters for one or a few families: **HOUSE** **2 a (1):** a shelter or refuge (as a nest or den) of a wild animal **(2):** a natural covering (as a tent or shell) that encloses and protects an animal or a colony of insects **b:** a building in which something is housed «a carriage → **3 a:** one of the 12 equal sectors in which the celestial sphere is divided in astrology **b:** 1 a zodiacal sign that is the seat of a planet's greatest influence **4 a:** HOUSEHOLD **b:** a family including ancestors, descendants, and kindred «the - of Tudor» **5 a:** a residence for a religious community or for students **b:** the community or students in residence **6 a:** a legislative, deliberative, or consultative assembly; esp: one constituting a division of a bicameral body

2 **house** **1-sense**

<dt> the building or chamber where such an assembly meets </dt>

3 **house** **1-sense**

«c: a quorum of such an assembly **7 a:** a place of business or entertainment «a movie → **b (1):** a business organization «a publishing → **(2):** a gambling establishment **c:** the audience in a theater or concert hall «a full — on opening night» **8:** the circular area 12 feet in diameter surrounding the tee and within which a curling stone must rest in order to count **9 fr:** The Warehouse, Chicago dance club that pioneered the style: a type of dance music mixed by a disc jockey that features overlapping with a heavy repetitive drumbeat and repeated electronic melody lines — **house-ful** \ˈhaʊz-ful/ n — **house-less** \ˈhaʊz-ləs/ adj — **house-less-ness** n

<dt> </dt>

<dt> slip on the house </dt>

<dt> 1 without charge; FREE </dt>

</dt>

4 **house**

Path: /home/lex1/jj/dictionaries/9/sense/2/ Model: Text Not Allowed

Project: MW-coil ID: 10001695

Locate: - Character: -

Schema context: default

Last phase: Error Site: Def
- Far Right Pane (Element panel):** Shows the 'sense' element with various options like 'Forward tabulation' and 'Operators'.

- built-in XML database, own query language and XQuery
- setting stored as a XML document in special database
- user accounts and permissions controlled by database
- output defined by XSLT, XSL:FO or own format
- customizable user interface, selectable panels
 - document XML structure, editing form, entry preview, statistics, links, multimedia library, ...
- extensible with many plug-ins
 - spellcheck, lexicography, SmartEdit, workflow, changes tracking,...

TLex (TshwaneLex)

- commercial, developed by TsdwaneDJe (South Africa)
- current version 5, October 2010
- price: 150 to 1900 EUR
- supported on Windows and MacOS
- single user or client-server application
- important clients include Oxford University Press, Pearson/Longman, or Macmillan Publishers

TshwaneLex - [C:\English - Chinese.tldict]

File Edit View Lemma Dictionary Format Tools Window Help

Bilingual Re

yarn
 ye
 yeah
 year
 yell
 yellow
 yep
 yes
 yesterday
 you
 year
 a
 yeah
 ye
 yacht

Lemma: year LemmaSign=year, Frequency=60, Part
 Sense: 1 SenseNumber=1
 TE: TE=年
 TE: TE=一年
 TE: TE=一年間
 Sense: 2 SenseNumber=2
 TE: TE=歲
 Subentry: Term=>s
 Sense: SenseNumber=1
 TE: TE=年齡
 Definition: Definition=age
 Example: Example=He is ten years

Attributes (F1) Attributes (F2) Search

Lemma:	<input type="checkbox"/> Incomplete
LemmaSign	year
Pronunciation	
Frequency	60

Sense: 1

TE:	
TE	年
Pinyin	
TE:	
TF	年

(Default styles)

year noun 年, 一年, 一年間 歲 -s 年齡
age: He is ten years old. 他十歲. 老年, 老人, 長者: a man of years 老人 多年: We have known each other for years. 我們是舊交了.
 年度: a financial year 會計年度 年代: the years of riots 暴動時代
yell verb 吼
yellow adjective 黃色的
yep interjection
yes interjection 是
yesterday adverb 昨天
yet adverb
yield¹ verb
yield² noun
you /ju/ pronoun (personal pronoun) 你 (m) /ni/, 你們 (pl) /nimen/, 妳 (f) /ni/, 妳們 (f pl) /nimen/, 您 (polite) /nin/, 您們 (polite pl) /ninmen/ 當主詞或受詞 一般人, 任何人 (特定土地之) 人們 (通常不譯出): Do you have much rainfall here? 當地雨下的多嗎?
young adjective 年輕
youngster noun 年紀小的人

- document structure set by DTD or graphic interface
- data stored in custom format
- import from XML or CSV, export to XML, HTML, MS Word, RTF, Adobe InDesign
- editing in forms or output preview
- localizable user interface
- multimedia support
- tracking changes and comparing document of several users
- integrated corpus tools

Mātāpuna

- free, open-source, developed by Thinktank Consulting (New Zealand)
- not developed since 2005
- web application, programmed in Perl, data stored in PostgreSQL, export to XML
- designed for Maori dictionary project
- document structure can be changed only by source code modification
- workflow and statistics
- data checks (spellcheck, defining dictionary,...)
- 2004 Computerworld Excellence Award for the Use of IT in Government

Matāpuna - edit Headword

[Home](#)
[Add Headword](#)
[My Editing](#)
[Search](#)
[Search corpus](#)
[Reports](#)
[Logout](#)

naku¹. **tmw**. Ka tapahi haere, ka ketuketu (i te whenua) kia taea ai tētahi mea te tango ake, te huna rānei. *Ki konā te kurī naku ai i te whenua me kore e taea tana wheua te huna.* {nanaku}

Headword

Sense :

Category

Atua domain

Word Class

Secondary word class


base - adjective
 base - adverb
 base - noun - common
 base - verb - intransitive
 base - verb - stative
 base - verb - transitive

Passive forms

Nominalising forms

Glossword

- free, open-source, developed by Dmitry N. Shilnikov
- development canceled in 2009, revived in 2010 by Glossword.biz Team
- web application, programmed in PHP, data stored in MySQL
- import/export XML and CSV
- creating explanatory dictionaries, glossaries, references
- document structure can be changed only by source code modification
- multiple dictionaries with one installation
- built-in editor to customize output templates
- feedback form for each entry
- virtual keyboard with special characters

Dictionary Users Add a term Online Admin User 03:51:43 Logout

- Search
- Terms
- Dictionaries
- Topics
- Custom pages
- Visual themes
- Abbreviations
- Virtual keyboard
- Custom alphabetic order
- Search queries log
- Users
- System settings
- Edit
- Maintenance
- Menu manager

Control Panel Home

Recently updated dictionaries

1	Edit	Add a term	Dictionary Name	15 June 2008
---	----------------------	----------------------------	-----------------	--------------


Recently updated terms

1	Edit	Term example	Dictionary Name
---	----------------------	--------------	-----------------

Tips

Create a new dictionary
Dictionary → Add → Fill the form, then click on "Save".

Add a term
Dictionary → Browse → Select dictionary from topic → Add a term → Fill the form, then click on "Save".

Control Panel Home • [to main page](#)  Powered by Glossword 1.8.10 [↑top of page](#)

DEB II

- free, open-source, developed by NLP Center, FI MU
- data stored in Oracle Berkeley DBXML, programmed in Ruby
- client-server architecture
- libraries and modules for dictionary application development
 - data storage and manipulation
 - user and dictionary management, cooperation
 - import, export, backup
- client: GUI, web application, API
- more than 10 projects – dictionaries, ontologies, lexical databases
 - DEBDict – general dictionary browser
 - DEBVisDic – wordnet editor and browser
 - Praled – Czech lexical database editor
 - FaNUK – dictionary of family names in UK
 - ...

Jazyk

Přidat význam Odebrat význam Heslo Význam 1 Význam 2 Význam 3

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

SEZNAMY (rozbalit/sbalit)
SLOVOTVORBA
Derivovaná slova
Fundace/motivace
Defraz. lexémy

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

Heslo **jazyk** Homonymie

Zdroj pro heslář fsc+ssjc

Typ hesla jednosl. Slovní druh/typ podstatné jméno

U zkratek: Plně znění zkratky

Pozn. k celému heslu

Uložit a zavřít Uložit Zavřít

V1

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

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;
vyříznutí jazyka;
špička jazyka, kofen jazyka
Pozn. k E4 zpracovávané substantivum je samo genitivním
E5 k V1 SUBST+Prep+Subst/SUBST+Subst-ji

Pozn. k E5 následující náboj zastává zpracovávané jméno

ZÁHLAVÍ ADMIN. VÝSL. PŮVOD DĚLENÍ ETYMOL. STAT.
EXPR. ÚZEMNÍ PRÁZNAK DOBOVÝ PRÁZNAK STYLOVÝ PRÁZNAK
ZKR./EKVIVALENTY SOUSLOVÍ FRAZÉMY JINÉ VÍCESL. VÝRAZÝ VÍCESL. NEZARÁZENÉ KOMP. VÍCESL. SLOVOTVORBA

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

Jazyk: ZÁHLAVÍ

Heslo **jazyk** Homonymie

Zdroj pro heslář fsc+ssjc

Typ hesla jednosl. Slovní druh/typ podstatné jméno

U zkratek: Plně znění zkratky

Pozn. k celému heslu

Uložit a zavřít Uložit Zavřít

Jazyk: FRAZÉMY

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

Uložit a zavřít Uložit Zavřít