

Počítačová lexikografie

Úvod

Adam Rambousek

Lexikografie

Seminární

- počítačová lexikografie
- historie a současnost lexikografie
- historie a současnost počítačové lexikografie
- počítačová lexikografie
- historie a současnost počítačové lexikografie
- počítačová lexikografie
- historie a současnost počítačové lexikografie
- počítačová lexikografie

Jak byt lexikografem?

Historie

Historie

- historie a současnost
- historie a současnost
- historie a současnost

Historie

- historie a současnost
- historie a současnost
- historie a současnost

Historie

- historie a současnost
- historie a současnost
- historie a současnost

Historie

- historie a současnost
- historie a současnost
- historie a současnost

Historie

- historie a současnost
- historie a současnost
- historie a současnost

Budoucnost

- historie a současnost
- historie a současnost
- historie a současnost

Historie česká

Historie

- historie a současnost
- historie a současnost
- historie a současnost

Historie

- historie a současnost
- historie a současnost
- historie a současnost

Historie

- historie a současnost
- historie a současnost
- historie a současnost

Historie

- historie a současnost
- historie a současnost
- historie a současnost

Historie?

- historie a současnost
- historie a současnost
- historie a současnost

Budoucnost?

- historie a současnost
- historie a současnost
- historie a současnost

Námě

- historie a současnost
- historie a současnost
- historie a současnost

Počítačová lexikografie

Slovníky a počítače

- historie a současnost
- historie a současnost
- historie a současnost

Slovníky a počítače

- historie a současnost
- historie a současnost
- historie a současnost

Podmínky ukončení

- historie a současnost
- historie a současnost
- historie a současnost



Počítačová lexikografie

Úvod

Adam Rambousek

Počítačová lexikografie

Úvod

Adam Rambousek

Lexikografie

Seznámení

- podoblast **lexikologie**
- **lexicography, lexikografie**
 - the activity or occupation of compiling dictionaries (Oxford d.)
 - the editing or making of a dictionary (Merriam-Webster d.)
 - the job of writing a dictionary (Macmillan d.)
- praktická lexikografie
- teoretická lexikografie – analýza a popis slovní zásoby, teorie o prvcích slovníku, skupinách uživatelů, hodnocení
- Slovník národního jazyka nřelí mezi prvními požadovanými vzdělávacími činnostmi.

Jak být lexikografem?




Historie


Historie

- hliněná tabulka z Uru (Sýrie)
- sumerština - obilovina
- cca 2500-2250 př.n.l.





Historie

- Sir Thomas Elyot: Dictionary (Bibliotheca Elyotae), 1538
- John Withals: A Shorte Dictionarie for Yang Yongeiners, 1553
- Robert Cawdrey: A Table Alphabeticall, 1604
- první výkladový slovník angličtiny
- "hard words, borrowed from... for the benefit & helpe of ladies, gentlewomen, or any other unskilfull persons"



Historie

- moderní slovníky, 17. století
- Samuel Johnson
 - 1747, Plan of a Dictionary of the English Language
 - analýza a důkladná práce
 - 1755, A Dictionary of the English Language
 - 42 773 hesel
 - "to preserve the purity and ascertain the meaning of our English Idioms" "to fix the English language"

Historie

- 19. století
- Noah Webster
 - An American Dictionary
 - 70 000 hesel, seriózní
 - odmělné zaplacení do
- Joseph Worcester
 - A Dictionary of the English Language
 - 1830-1864 "náhla slovníků"
 - připravovaný Webster
 - Merriam-Webster

Historie česká

Historie

- Bartoloměj z Chlumece, Klavet, 14. století
 - latinsko-češské slovníky
 - Volukář (pravaričský) - odborná terminologie
 - Bohemif
 - Glagol - 7000 hesel
 - Raro sapientis glosa de bestis carnis honesta, Lexico siliquo lupusque lefeljus hieca, nevidet ursus Ursaque nevidalica, lupu vicaica, die ovis orcaae, koas capra, vulpes lylaque carnicula rasta.
 - Půdny st. castus, gnosnoascez nominativus, Ast vocativus uslarok danolque dative, Accusativus ysnovavet primaque prvo.

Historie

- Daniel Adam z Veleslavína, 16. století
 - Dictionarium linguae latinae - latina-čeština
 - Nomenclator trilinguus - latina-čeština-němčina
 - Nomenclator quadrilinguus - švédsko-čeština-němčina-latina-řeština-němčina, 958-300 stran, řazeno český, záložek hrdolavání
- Jan Amos Komenský, 17. století
 - Thesaurus linguae Bohemicae - latinsko-češský, česko-latinský, synchronní, duchovní, levičání, gramatické informace, frazeologie
 - 20 let příprav... požár Lesna

Historie

- Josef Jungmann, 19. století
 - Slovník česko-německý
 - 1815-1831, vydáno 1835-1839
 - 5 svazků, 4694 stran, 120000 hesel
 - popisný výkladový slovník



Historie

- Kancelář Slovníku jazyka
 - 18r slovník velkého m. dobrovolníci
 - výpisky z próby, básně, literární, publicist.
 - Přínutí slovník jazyka
 - 10 821 stran, 250 000 hesel
 - časová "náhla slovníků"
 - Ústav pro jazyk český.

Podmínky ukončení

- **study (verb)** the act of looking, eating, and watching TV with an open textbook nearby
- ústně nepovinná
- zápočet
 - písemný test (60 %)
 - projekt - vytvořte si vlastní slovník

Počítačová lexikografie

Slovníky a počítače

- 60. léta - používají se počítače, lexikografové přili na papír, specialně přepisují do databáze, Brown Corpus
- 1978, Longman Dictionary of Contemporary English
 - první s omezeným slovníkem detmici, kontrolovací stroje
 - kódování pro NLP výzkum

Slovníky a počítače

- 1980, COBUILD, University of Birmingham - Collins
 - korpus současných textů (Bank of English)
 - 1987, Collins-COBUILD English Language Dictionary
 - první slovník založený na korpusových datech
 - nový styl definice - celé věty
 - if a person, animal, or other living thing is **killed**, something or someone causes them to die.
- 90. léta - vývoj specializovaných systémů pro tvorbu slovníků (dictionary writing systems)
 - 1987, Text Encoding Initiative
 - 1990, TEI Guidelines

Podmínky ukončení

- **study (verb)** *the act of texting, eating, and watching TV with an open textbook nearby*
- účast nepovinná
- zápočet
 - písemný test (60 %)
 - projekt - vytvořte si vlastní slovník

Lexikografie



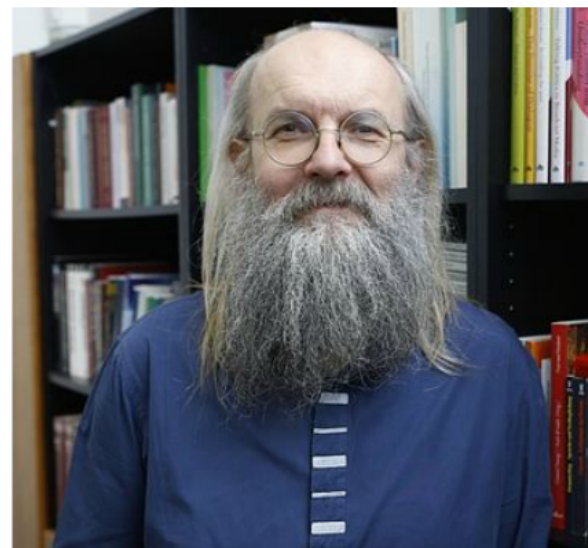
Seznámení

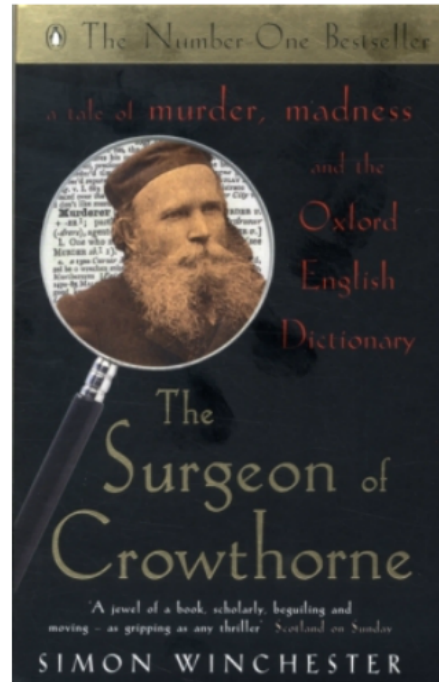
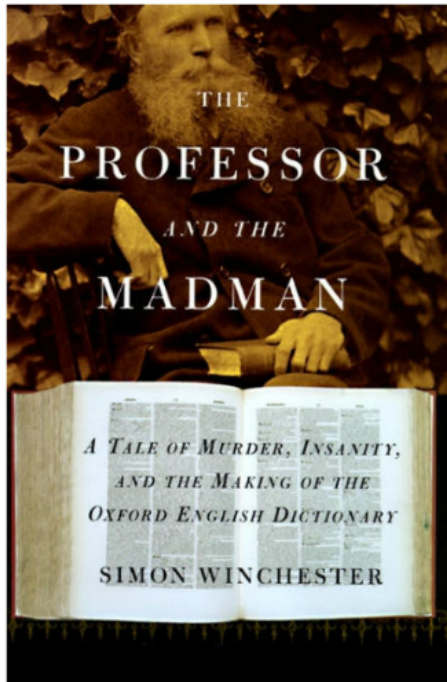
- podoblast **lexikologie**
- lexicography, **lexikografie**

Seznámení

- podoblast **lexikologie**
- lexicography, **lexikografie**
 - *the activity or occupation of compiling dictionaries (Oxford d.)*
 - *the editing or making of a dictionary (Merriam-Webster d.)*
 - *the job of writing a dictionary (Macmillan d.)*
- praktická lexikografie
- teoretická lexikografie - analýza a popis slovní zásoby, teorie o prvcích slovníku, skupinách uživatelů, hodnocení
- *Slovník národního jazyka náleží mezi první potřeby vzdělaného člověka.*

Jak být lexikografem?







Historie



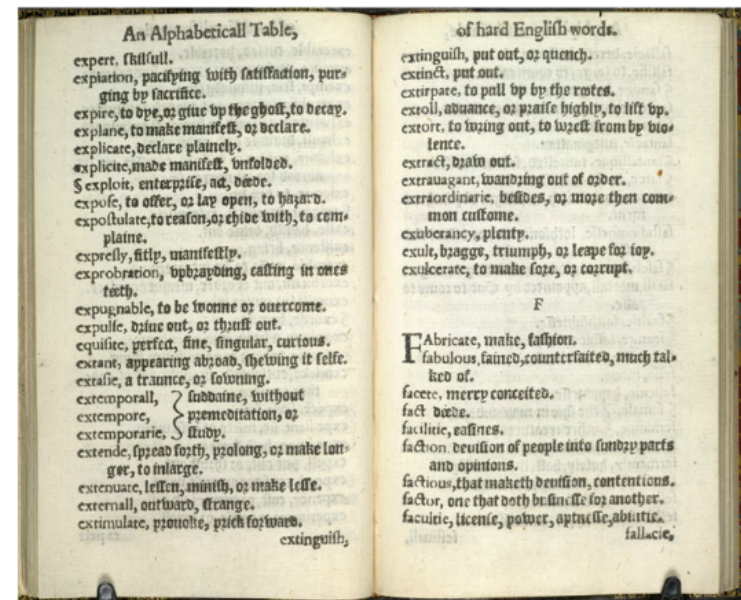
Historie

- hliněné tabulky z Ebla (Sýrie)
- sumerština - eblaština
- cca 2500-2250 př.n.l.



Historie

- Sir Thomas Elyot: Dictionary (Bibliotheca Eliotae), 1538
- John Withals: A Shorte Dictionarie for Yong Begynners, 1553
- Robert Cawdrey: A Table Alphabeticall, 1604
 - první výkladový slovník angličtiny
 - "hard wordes, borrowed from... for the benefit & helpe of Ladies, Gentlewomen, or any other unskilfull persons"



An Alphabetical Table,

expert, skillfull.
 expiation, pacifying with satisfaction, purging by sacrifice.
 expire, to dye, or giue vp the ghost, to decay.
 explane, to make manifest, or declare.
 explicate, declare plainly.
 explicite, made manifest, unfolded.
 exploit, enterprise, act, deede.
 expose, to offer, or lay open, to hazard.
 expostulate, to reason, or chide with, to complain.
 expressly, fitly, manifestly.
 exprobration, vpbzayding, casting in ones teeth.
 expugnable, to be swonne or ouercome.
 expulse, drine out, or thrust out.
 equisite, perfect, fine, singular, curious.
 extant, appearing abroad, shewing it selfe.
 extasie, a traunce, or swooning.
 extemporall, } suddaine, without
 extempore, } premeditation, or
 extemporarie, } study.
 extende, spread forth, prolong, or make longer, to enlarge.
 extenuate, lessen, minish, or make lesse.
 externall, outward, strange.
 extrimulate, prouoke, prick forward.
 extinguish,

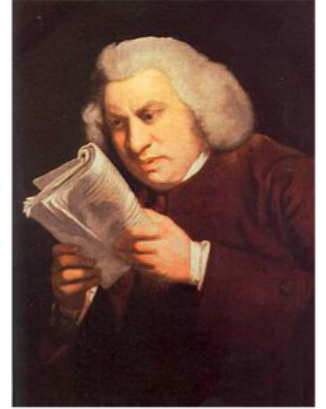
of hard English words.

extinguish, put out, or quench.
 extinct, put out.
 extirpate, to pull by by the rootes.
 extoll, aduance, or praise highly, to lift vp.
 extort, to wring out, to wrest from by violence.
 extract, drato out.
 extrauagant, wandring out of order.
 extraordinarie, besides, or more then common custome.
 exuberancy, plenty.
 exult, bragge, triumph, or leape for toy.
 exulcerate, to make soze, or corrupt.

F

Fabricate, make, fashion.
 Fabulous, fained, counterfeitd, much talked of.
 facete, merry conceited.
 fact deede.
 facilitie, easines.
 faction. deuision of people into sundry parts and opinions.
 factious, that maketh deuision, contentious.
 factor, one that doth businesse for another.
 facultie, license, power, aptnesse, abilitie.
 fallacie,

Historie




- moderní slovníky, 17. století
- **Samuel Johnson**
 - 1747, Plan of a Dictionary of the English Language
 - analýza a žádost o peníze
 - 1755, *A Dictionary of the English Language*
- 42 773 hesel
- *"to preserve the purity and ascertain the meaning our English idiom" "to fix the English language"*

VULCĀNO. *n. s.* [Italian.] A burning mountain; volcano.
Earth calcin'd, flies off into the air; the ashes of burning
mountains, in *vulcano's*, will be carried to great distances. *Art.*

Historie

- 19. století
- **Noah Webster**
 - *An American Dictionary of the English Language*, 1828
 - 70 000 hesel, srovnání britské a americké angličtiny
 - odmítal zařazovat do slovníků neslušná slova
- Joseph Worcester
 - *A Dictionary of the English Language*
- 1830-1864 "válka slovníků"
 - přepracovaný Websterův slovník
 - Merriam-Webster

WEBSTER'S INTERNATIONAL DICTIONARY
The One Great Standard Authority, So writes
Hon. D. J. Brewer, Justice U. S. Supreme Court.



IT IS A THOROUGH REVISION of the UNABRIDGED, The purpose of which has been not display nor the provision of material for boastful and showy advertisement, but the due, judicious, scholarly, thorough perfecting of a work which in all the stages of its growth has obtained in an equal degree the favor and confidence of scholars and of the general public.

IT IS THE BEST FOR PRACTICAL PURPOSES, BECAUSE,

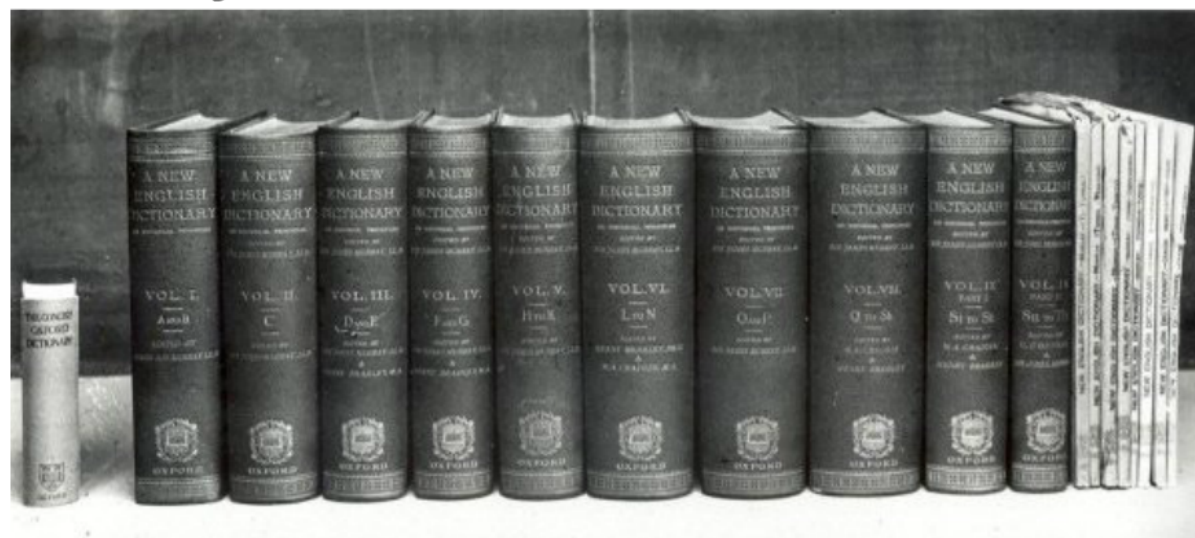
Words are easily found *** Pronunciation is easily ascertained, Meanings are easily learned *** The growth of words is easily traced, and because excellence of quality rather than superfluity of quantity characterizes its every department. *** Get the Best.

Pamphlet Free.

G. & C. MERRIAM CO., Publishers, SPRING FIELD, MASS., U. S. A.

Historie

- *The Oxford English Dictionary (A New English Dictionary)*
 - 1857, Philological Society, R. C. Trench, kritika slovníků
 - 1859, začátek sběru dat, dobrovolní čtenáři
 - 1879, *James A. H. Murray* jmenován hlavním editorem
 - 1882-1928, vychází 12 svazků 15 487 stran, 240 000 hesel



Budoucnost

- *The Oxford English Dictionary 3rd Edition*
 - 2000-2037?
 - rozpočet 34M £
 - 70 lexikografů Oxford + New York
 - "Every word in the Dictionary is being reviewed"
 - aktualizace 4x ročně
 - úpravy v dávkách (frekvence, etymologie, aktuálnost...)

The screenshot shows the Oxford English Dictionary (OED) entry for the word "worstest". The page header includes the OED logo and the text "Oxford English Dictionary the definitive record of the English language". A search bar is visible in the top right corner. The main heading is "worstest, *adj.* and *n.*". Below this, there are sections for "Pronunciation" (with British and American options), "Formed from" (from "worst" and "est"), "Origin" (from "worst" and "est"), and "Compare with" (with "worst" and "worstest"). The entry also includes a "Chiefly humorous and in representations of non-standard speech" note and an "In emphatic use" note. A list of citations is provided at the bottom, including references to "The Spectator" and "The Spectator" from 1760, 1890, 1974, and 2012.

worstest, *adj.* and *n.*

View as: [Outline](#) | [Full entry](#)Quotations: [Show all](#) | [Hide all](#) | [K](#)**Pronunciation:** Brit.  /'wɜːstɪst/, U.S.  /'wɜrstɪst/**Forms:** see *worst* *adj.* and *-est* *suffix*.**Origin:** Formed within English, by derivation. **Etymons:** *worst* *adj.*, *-est* *suffix*.**Etymology:** < *worst* *adj.* + *-est* *suffix*.Compare earlier *worsest* *adj.* With the double superlative formation, compare earlier *bestest* *adj.* ([Show Less](#))Chiefly *humorous* and in representations of nonstandard speech.
A. adj.In emphatic use: = *worst* *adj.* (in various senses).

Categor

1768 *Mod. Courtship* i. iv. 14 I think I have not the worstest voice in the world.1856 H. J. CONWAY *Dred* i. iv. 9 Dis yer is de wustest shirt I ebber see.1892 *News & Observer* (Raleigh, N. Carolina) 23 Sept. Worst—oh, worstest thing of all! I have to stay a girl.1953 J. M. BREWER *Word on Brazos* 11 You know dey hab a sayin' in de Bottoms in de ole days dat de preachuh hab de wustes' chillun in de wo1974 W. LEEDS *Herefordshire Speech* 10 The worstest weather.2012 S. KIRKBY *Capt. Awesome to Rescue* iii. 30 Then he realized the worstest, awfulest truth: *She can read minds!*[\(Hide qu](#)



Historie česká



Historie

- **Bartoloměj z Chlumce**, Klaret, 14. století
 - latinsko-české slovníky
 - *Vokabulář (gramatický)* - odborná terminologie

Historie

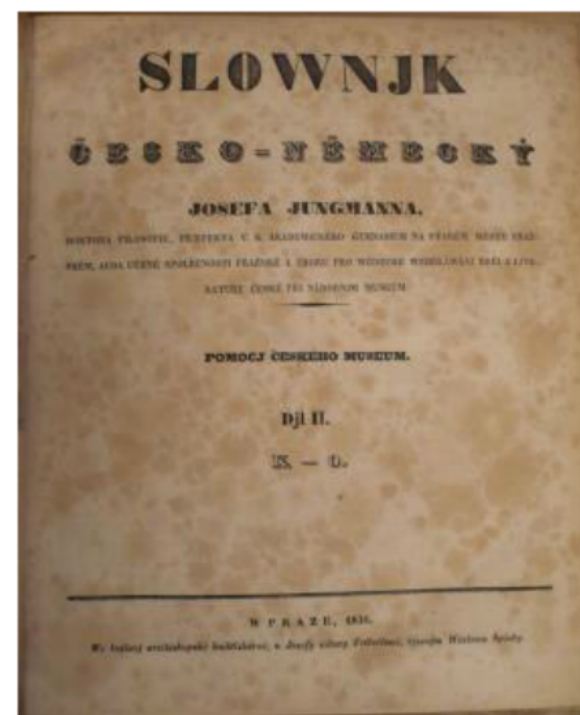
- **Bartoloměj z Chlumce**, Klaret, 14. století
 - latinsko-české slovníky
 - *Vokabulář (gramatický)* - odborná terminologie
 - *Bohemář*
 - *Glosář* - 7000 dvojic
 - *Raro sequens gesta de bestiis cernis honesta. Lew leo wlkque lupusque le[e]na lwicze, nedvied ursus Ursaque nedviedicze, lupa wlczicze, dic ovis owcze, Koza capra, vulpes lyskaque canicula tysta.*
 - *Padenye sit casus, gmenowanecz nominativus, Ast vocativus wolanek danekque dativus, Accusativus wynowanek primaque prwa.*

Historie

- **Daniel Adam z Veleslavína**, 16. století
 - *Dictionarium linguae latinae* - latina-čeština
 - *Nomenclator tribus linguis* - latina-čeština-němčina
 - *Nomenclator quadrilinguis + Silva quadrilinguis* čeština-latina-řečtina-němčina, 958+300 stran, řazeno česky, zárodek hnízdování
- **Jan Amos Komenský**, 17. století
 - *Thesaurus linguae Bohemicae* - latinsko-český, česko-latinský, synchronní, diachronní, lexikální, gramatické informace, frazeologie
 - 20 let příprav... požár Lešna

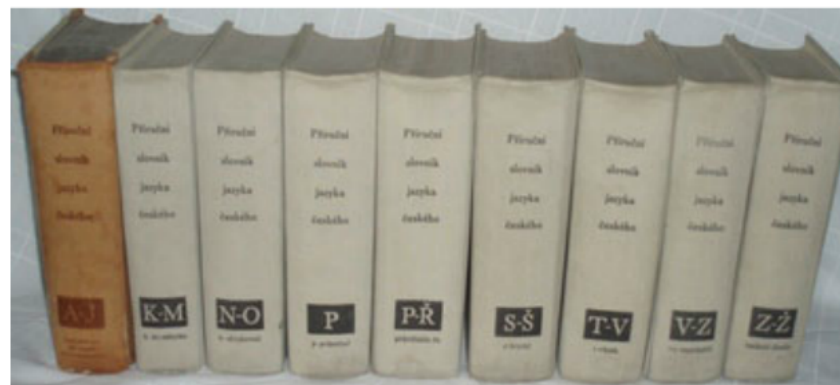
Historie

- **Josef Jungmann, 19. století**
 - *Slovník česko-německý*
 - 1815-1833, vydáno 1835-1839
 - 5 svazků, 4694 stran, 120000 hesel
 - popisný výkladový slovník



Historie

- **Kancelář Slovníku jazyka českého, 1911**
 - sběr slovníkového materiálu, dobrovolníci
 - výpisky z prózy, básní, odborné literatury, publicistických článků
 - *Příruční slovník jazyka českého, 1935-1957*
 - 10 824 stran, 250 000 hesel
 - cenzura "nežádoucích spisovatelů"
- **Ústav pro jazyk český, 1946**



Historie?

- *Slovník spisovného jazyka českého*, 1960-1971, 1989
 - plán na menší slovník -> 4 500 stran, 192 908 hesel
 - první slovník s kodifikačním statutem
 - více terminologie (učebnice)
 - novinky: hnízdování, původ přejatých slov



Budoucnost?

- *Akademický slovník současné češtiny*, 2005, 2012-2016-?
 - předpokládaný rozsah 120-150 tisíc hesel
 - současná čeština oficiální i neoficiální
 - specializované, nářeční omezeně
 - zveřejněno A
 - www.slovníkcestiny.cz

The screenshot shows the website 'Akademický slovník současné češtiny'. The navigation bar includes 'Úvod', 'Autoři', 'Předmluva', 'Zásady zpracování', 'Zkratky a značky', and 'Rychlé v'. The search bar is set to 'hesláři' and shows 'Vyhledat slovo:'. The search results for 'akuvrtačka' are displayed, including the definition: 'akuvrtačka -čky (2. mn. -ček) ž. vrtačka napájená z akumulátoru, akumulátorová vrtačka příklepová akuvrtačka akuvrtačka s příslušenstvím díry vyvrtal akuvrtačkou'. The source is cited as 'Slovník spisovného jazyka českého (1960–1971)'. There is a link to 'Exportovat heslo do PDF'. A list of words starting with 'A' is shown on the left, including 'akutnost', 'akuvrtačka', 'akuzativ', 'akuzativní', 'akva-', 'akva-', 'akvabela', 'akvacentrum', 'akvacentrum', and 'akvadukt'.

Akademický slovník současné češtiny

[Úvod](#) [Autoři](#) [Předmluva](#) [Zásady zpracování](#) [Zkratky a značky](#) [Rychlé v](#)

Hledej ve slovníku

Hledat v

hesláři celém odstavci

Vyhledat slovo:

A

|< < 1 2 3 ... 31 32 33 **34** 35 36 37 ... 134 135 136

> >|

akutnost

akuvrtačka

akuzativ

akuzativní

akva-

aqua-

akvabela

akvacentrum

aquacentrum

akvadukt

akuvrtačka -čky (2. mn. -ček) ž.

vrtačka napájená z akumulátoru, akumulátorová vrtačka

příklepová akuvrtačka

akuvrtačka s příslušenstvím

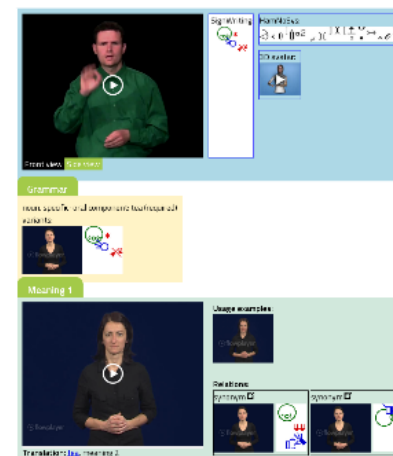
díry vyvrtal akuvrtačkou

Slovník spisovného jazyka českého (1960–1971)

[Exportovat heslo do PDF](#)

Různé

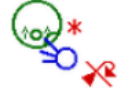
- diachronní: *Elektronický slovník staré češtiny*
- nářeční: *Český jazykový atlas*
- slang: *Slovník nespisovné češtiny*
- onomastický: *Slovník pomístních jmen na Moravě a ve Slezsku, Jak se bude vaše dítě jmenovat?*
- frazeologie: *Slovník české frazeologie a idiomatiky*
- frekvenční řazení: *Frekvenční slovník češtiny*
- gramatický: *Slovesa pro praxi*
- neologizmy: *Nová slova v češtině 1+2*



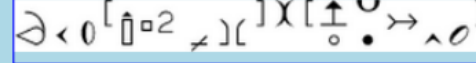


Front view Side view

SignWriting:



HamNoSys:



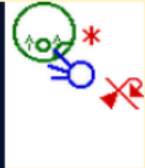
3D avatar:



Grammar

noun, specific; oral component: tea (required)

variants:



Meaning 1



Translation: [tea](#), meaning 2

Usage examples:



Relations:

synonym ↗



synonym ↗



- *vokabular (gramaticky) - odborná terminologie*
- *Bohemář*
- *Glosář - 7000 dvojic*
- *Raro sequens gesta de bestiis cernis honesta. Lew leo wlkque lupusque le[e]na lwicze, nedvied ursus Ursaque nedviedicze, lupa wlczicze, dic ovis owcze, Koza capra, vulpes lyskaque canicula tysta.*
- *Padenye sit casus, gmenowanecz nominativus, Ast vocativus wolanek danekque dativus, Accusativus wynowanek primaque prwa.*

- *Nomenclator t*
- *Nomenclator d*
latina-řečtin
zárodek hníz
- **Jan Amos Kome**
- *Thesaurus ling*
latinský, syn
- informace, fr
- 20 let příprav

Počítačová lexikografie

Slovníky a počítače

- 60. léta - používají se počítače, lexikografové píší na papír, specialisté přepisují do databáze, Brown Corpus
- 1978, *Longman Dictionary of Contemporary English*
 - první s omezeným slovníkem definicí, kontrolováno strojově
 - kódování pro NLP výzkum

Slovníky a p

- 1980, **COBUILD**
 - korpus soudč
 - 1987, *Collins*
 - první slovní
 - nový styl de
 - *If a person, a*
or someone c
- 90. léta - vývoj s
slovníků (dictio
- 1987 *Text Enc*

Slovníky a počítače

- 60. léta - používají se počítače, lexikografové píší na papír, specialisté přepisují do databáze, Brown Corpus
- 1978, *Longman Dictionary of Contemporary English*
 - první s omezeným slovníkem definicí, kontrolováno strojově
 - kódování pro NLP výzkum

Slovníky a počítače

- 1980, **COBUILD**, University of Birmingham + Collins
 - korpus současných textů (*Bank of English*)
 - 1987, *Collins COBUILD English Language Dictionary*
 - první slovník založený na korpusových datech
 - nový styl definice - celé věty
 - *If a person, animal, or other living thing **is killed**, something or someone causes them to die.*
- 90. léta - vývoj specializovaných systémů pro tvorbu slovníků (dictionary writing systems)
- 1987, Text Encoding Initiative
- 1990, TEI Guidelines