

Počítačová lexikografie Úvod

Adam Rambousek

Podmínky ukončení

- study (verb) the act of texting, eating, and watching TV with an open textbook nearby
- účast nepovinná
- zápočet
 - písemný test (60 %)
 - týmový projekt vytvořte si vlastní slovník
 - pravděpodobně 11. prosince
- (20. listopadu volno)



Lexikografie

Seznámení

Jak být lexikografem?





Historie

Historie

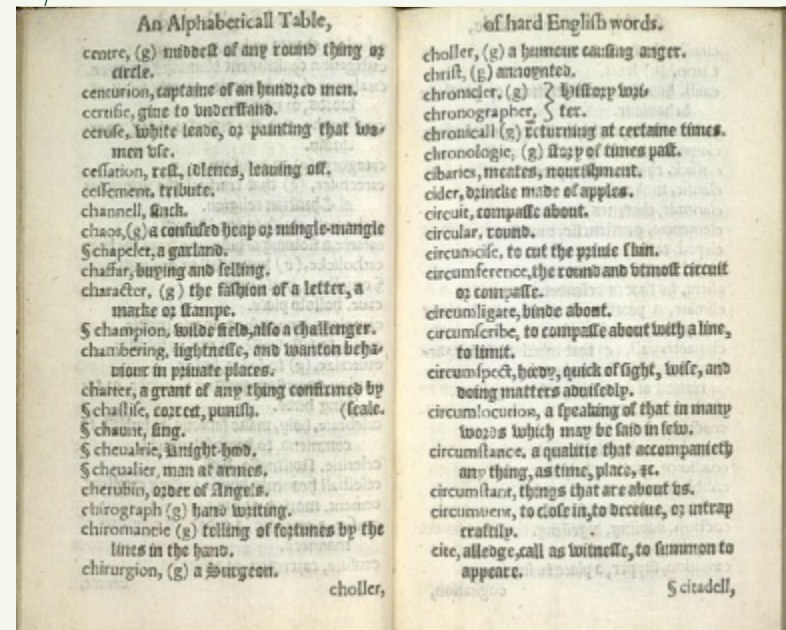
- hliněné tabulky z Ebla (Sýrie)
- sumerština eblaština
- cca 2500 – 2250 př.n.l.



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"



Historie

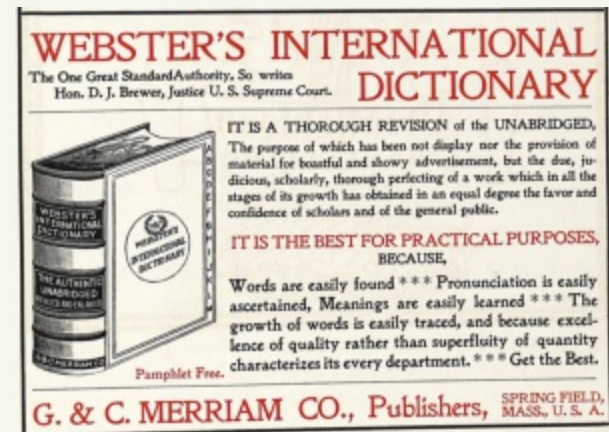
- moderní slovníky, 17. století
- Samuel Johnson
 - 1747, Plan of a Dictionary of the English Language
 - analýza a žádost o peníze
 - 1755, A Dictionary of the English Language
 - 42 773 hesel
 - "to preserve the purity and ascertain the meaning our English idiom" "to fix the English language"



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.*

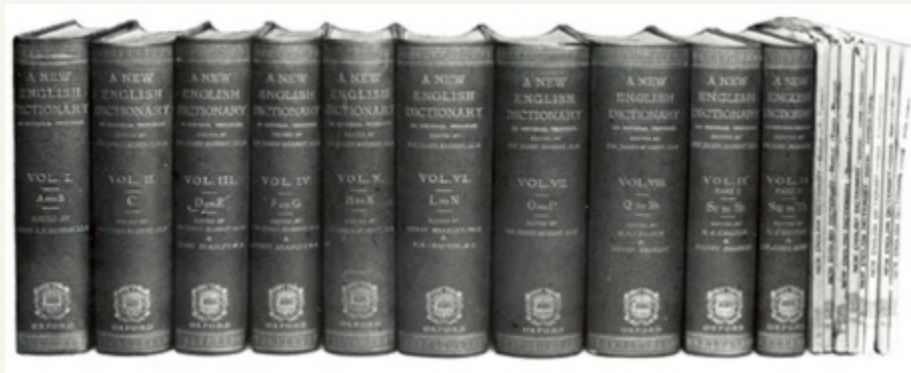
Historie

- 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



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



Historie

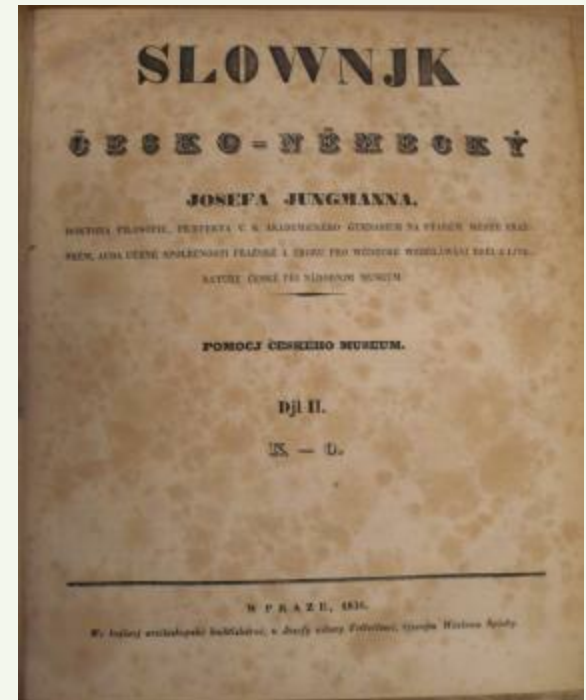
- Bartoloměj z Chlumce, Klaret, 14. století
 - latinsko české slovníky
 - Vokabulář (gramatický) odborná terminologie
 - Bohemář
 - Glosář 7000 dvojic
 - Raro sequens gesta de bestiis cernis honesta.
Lew leo wlkque lupusque le[e]na lwicze, nedvied ursus
Ursaque nedviedicze, lupa wlczicze, dic ovis owcze,
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,
frazologie
 - 20 let příprav... požár Lešna

Historie

- Josef Jungmann, 19. století
 - Slovník česko německý
 - 1815 1833, vyd 1835 1839
 - 5 svazků, 4694 stran
 - 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



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



Význam 1



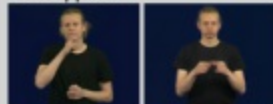
 *Informatika*

synonymum  e-mail

synonymum   predikát

Příklad:
[kapacita](#)

Příklady použití:



Význam 2



 *Matematika*

synonymum   predikát

antonymum   výrok

Příklad:
[komunikace](#)
[výrok](#)

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
 - k dov £n – 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
- 1987, Text Encoding Initiative
- 1990, TEI Guidelines



Slovníky

Rozdělení slovníků

- Jazyk: jednojazyčné, dvoujazyčné (směr), vícejazyčné
- Obsah: obecný jazyk, terminologie, encyklopedie, oblast
- Čas: synchronní, diachronní
- Velikost (počet hesel)
- Médium: tisk, CD, web
- Řazení: abecední, frekvenční, věcné
- Uživatelé: lingvisté, veřejnost, studenti, děti, studenti jazyka