

# Počítačová lexikografie

PLIN059

Mgr. Dana Hlaváčková, Ph.D.

# Počítačová lexikografie

- od klasické lexikografie k počítačové
- webové formuláře a aplikace
- značkovací jazyk XML
- ukázky

# Klasická lexikografie – tvorba slovníků

- 1891 – Česká akademie věd a umění
- 1911 – **Kancelář slovníku jazyka českého**
- započal sběr slovníkového materiálu
  - excerpční lístky (próza, poezie, odborná literatura, publicistika)
  - z textů od r. 1770
  - **Lístkový lexikální archiv**
  - 1911–1991 **8 696 850** excerpt
  - v 2007 dokončeno skenování





**Ústav pro jazyk český AV ČR, Letenská 4, Praha**

# Klasická lexikografie – tvorba slovníků

- 1935–1957 *Příruční slovník jazyka českého (PSJČ)*
  - v té době je k dispozici již 5 500 000 lístků
  - 9 svazků, 250 000 hesel
- 1946 založen **Ústav pro jazyk český**
- 1958–1970 *Slovník spisovného jazyka českého (SSJČ)*
  - 4 svazky, přes 190 000 hesel
  - v r. 1989 v 8 svazcích
- 1978 *Slovník spisovné češtiny pro školu a veřejnost (SSČ)* – kodifikační
  - téměř 50 000 slov

# Klasická lexikografie – tvorba slovníků

- **excerpta** – výběr heslových slov, uvádění příkladů použití (často beletrie)
- psaní slovníku podle abecedy různými autory (nevyváženost)
- slovníky pouze v tištěné podobě v několika vydáních
- až SSČ vzniká v elektronické podobě (diskety)
- struktura slovníkového hesla je často nekonzistentní

# Slovníkové heslo SSČ (1978)

- **zámek<sup>1</sup>**, -mku m **1.** zařízení (ovládané klíčem) umožňující něco zamknout, zavřít: z. na dveřích; patentní, visací, dozický z.; zavřít na z. ♦ (zavřít, být zavřen) na tři z-y, na pět, devět, desatero z-ů důkladně, nedobytně **2.** jiné zařízení sloužící k uzavření, upevnění, spojení ap.: z. lovecké pušky; **zámkový** příd.: z-é pero; **zámeček<sup>1</sup>**, -čku m zdrob.
- **zámek<sup>2</sup>**, -mku m rozlehlá výstavná budova, někdejší šlechtické sídlo: renesanční z.; z. Hluboká ♦ (stavět si) vzdušné, větrné z-y (dělat si) plané naděje; **zámecký** příd.: z-á věž; z. park; bývalí z-čtí páni; **zámeček<sup>2</sup>**, -čku m zdrob.

# Počítačová lexikografie

- využití počítačových nástrojů
- využití elektronických zdrojů jazykových dat
- **tvorba a editace slovníků**
  - webové formuláře
  - lexikografické stanice
- **jazyková data**
  - korpusy, výběr heslových slov, příklady užití, frazeologie, sémantické informace
- **prohlížení slovníků**
  - webová rozhraní



# Počítačová lexikografie – typické rysy

- možnost zpracovat velké **množství** dat
- **rychlost** zpracování
- orientace na užívání slov v **reálném kontextu**
- **multimediální** slovníky
- zachycení **aktuálního stavu** jazyka
- je možná proměna struktury slovníkového hesla
- vysoká konzistence
- uživatelské rozhraní – **GUI** (Graphical User Interface)
- **XML** – e**X**tensible **M**arkup **L**anguage (rozšiřitelný značkovací jazyk) – zakódování slovníkového hesla
- možné převody do dalších formátů (PDF, HTML)
- **multiplatformnost**

## terrorismus

[-iz-] (dř. též -ism), -mu m. (z lat.) *způsob vlády vymáhající terorem poslušnost; hrůzovláda, krutovláda, despotismus*: vojenský t.; nesnesitelný t.; demagogie a t.; přen. expr. to je t., nedejte si to líbit

<entry>

<hw>

<orth>terrorismus</orth>

</hw>

<senses>

<sense>

<def>způsob vlády vymáhající terorem poslušnost</def>

<def>hrůzovláda</def>

<def>krutovláda</def>

<def>despotismus</def>

<eg>vojenský terorismus</eg>

<eg>nesnesitelný terorismus</eg>

<eg>demagogie a terorismus</eg>

<usg type=style>přen.expr.</usg>

to je terorismus, nedejte si to líbit

</eg>

</sense>

</senses>

</entry>

# Počítačová lexikografie

- **XML**
  - počáteční a ukončovací značky
  - možnost zanořené struktury
- **DTD** (Document Type Definition) – popis použitých značek
- heslové slovo – headword – lemma
- **lemma** = základní tvar slova
  - substantiva – nominativ singuláru (nom. sg.)
  - adjektiva – nominativ singuláru maskulina (nom. sg. masc.)
  - slovesa – infinitiv (inf.)

# Současné slovníkové platformy

- **Elektronické slovníky a zdroje ÚJČ AV ČR**
  - <http://ujc.cas.cz/>
- **LEXIKO** – webové hnízdo ÚJČ AV ČR
  - [www.lexiko.ujc.cas.cz](http://www.lexiko.ujc.cas.cz)
- **DEB II (Dictionary Editor and Browser)**
  - <http://deb.fi.muni.cz>
  - CZPJ FI MU Brno
  - Martin Povolný, Aleš Horák, **Adam Rambousek**
  - klient-server, servlety, webové rozhraní, XML
  - **klienti** – DEBDict, DEBWrite, NESČ
- **LEXONOMY**
  - <https://www.lexonomy.eu/>

# Současné slovníkové platformy

- **Slovník jedním kliknutím**
  - aplikace ve Sketch Engine
  - automatické vytvoření slovníku na základě korpusových dat
- Akademický slovník současné češtiny
  - <https://www.slovníkcestiny.cz/uvod.php>