

Počítačové nástroje pro češtinu

Mgr. Dana Hlaváčková, Ph.D.

hlavack@fi.muni.cz

Centrum zpracování přirozeného jazyka FI MU

budova B, 2. patro, prostřední laboratoř (B206)

Centrum počítačové lingvistiky FF MU

budova U1 (Úvoz), 4. patro, pracovna 412

Počítačové zpracování češtiny

- **přirozený jazyk** x počítačové zpracování
- **jak funguje přirozený jazyk?**
- **jak funguje počítač?**
- **algoritmus** – návod, postup při řešení daného problému
- pravidelnost v jazyce (cca 80 %) – algoritmický popis

Počítačové zpracování češtiny – pár zásad

- **proč** to chceme? (cíl, účel, uživatel)
- **jak** toho dosáhneme? (efektivita)
- maximum **automatizace** – minimum ruční práce
- zpracování **velkého objemu** dat
- **univerzálnost** (široká množina vstupů)
- **nezávislost** na jednotlivých lingvistických teoriích
- **PŘESNOST** („ono to nefunguje“ 😞)

Počítačové zpracování češtiny

- urychlení a zefektivnění práce lingvisty
- ověřování existujících teorií
- objevení nového jazykového jevu, zákonitosti
- co a jak mohu použít
- co mohu a nemohu od nástroje očekávat
- autorská práva a přístupy

Mezioborová spolupráce

- **informatika – lingvistika** („společný jazyk“)
- **počítačová lingvistika** (matematická, počítačnická), jazykové inženýrství, počítačové zpracování přirozeného jazyka
- **Natural Language Processing (NLP)**

Hlavní oblasti (uživatelský přístup)

- syntéza a analýza řeči
- počítačová lexikografie
- formální analýza jazyka (morfologická, slovotvorná, syntaktická, sémantická)
- korpusová lingvistika
- dialogové systémy, umělá inteligence

Obsah kurzu

- počítačová lexikografie, prohlížeč a editor slovníků – DebDict a další
- rozpoznávání a syntéza řeči
- jazykové korpusy – Sketch Engine, Word Sketches
- morfologická analýza – ajka, morče
- derivační rozhraní – Deriv, Morfio
- syntaktická analýza – synt, set, PDT
- sémantická analýza – WordNet, ontologie
- valenční databáze – Vallex, VerbaLex
- seminární práce

Příbuzná pracoviště

- Centrum zpracování přirozeného jazyka FI MU Brno – <http://nlp.fi.muni.cz/>
- Ústav formální a aplikované lingvistiky MFF UK Praha – <http://ufal.mff.cuni.cz>
- Ústav teoretické a počítačové lingvistiky FF UK Praha – <http://utkl.ff.cuni.cz>
- Ústav Českého národního korpusu FF UK Praha – <http://www.korpus.cz>
- Ústav pro jazyk český AV ČR – <http://www.ujc.cas.cz>

Příbuzná pracoviště

- Fakulta informačních technologií VUT Brno – <http://www.fit.vutbr.cz>
- Katedra informatiky a výpočetní techniky FAV ZCU Plzeň – <http://www.kiv.zcu.cz>, Katedra kybernetiky <http://www.kky.zcu.cz>
- Ústav informačních technologií a elektroniky FM TU Liberec – <http://www.fm.tul.cz>

- Internetová jazyková příručka
<http://prirucka.ujc.cas.cz>