

Úvodní hodina



Počítačové nástroje pro češtinu
jaro 2021

Markéta Audy Masopustová

Organizační informace

- celkem 10 seminářů, ke konci semestru „volno“ na zpracování úkolu
- docházka
- zakončení: seminární úkol
 - podrobné informace se dozvíte v průběhu výuky
- obecné informace: kontaktovat M. Audy Masopustovou (audy.masopustova@phil.muni.cz)
- tematické informace: kontaktovat vyučující daného semináře/tématu

Co je počítačová lingvistika

- obor na pomezí informatiky a lingvistiky
- analyzuje jazyk, jeho využití a formalizuje jeho popis
- výsledkem jsou různé aplikace:
 - korektor překlepů a gramatiky
 - vyhledávání na webu
 - prediktivní psaní
 - překladače

Počítačové zpracování přirozeného jazyka

- počítačové zpracování × přirozený jazyk
- Jak funguje přirozený jazyk? × Jak funguje počítač?
- algoritmus – návod, postup při řešení problému
- formální popis jazyka
- v jazyce jsou pravidelnosti (cca 80 %) – je možné ho *nějak* popsat

Počítačové zpracování češtiny: zásady

- proč to chceme? (cíl, účel, uživatel)
- jak toho dosáhneme? (efektivita)
- maximum automatizace – minimum ruční práce (při vytváření i používání)
- zpracování velkého objemu dat
- univerzálnost (široká množina vstupů, spojování více nástrojů do jednoho)
- nezávislost na jednotlivých lingvistických teoriích
- při zpracování i používání je nutná PŘESNOST

Počítačové zpracování češtiny: důsledky

- urychlení a zefektivnění práce lingvisty
- ověřování existujících teorií
- objevení nového jazykového jevu, zákonitosti

Počítačové zpracování češtiny: v semináři

- co lze a nelze od nástroje očekávat
- co a jak je možné použít
- autorská práva a přístupy k nástrojům
 - veřejně dostupné (dostupné na MU)
 - hromadný přístup (společné heslo)
 - vlastní přístup (registrace)

Mezioborová spolupráce

- informatika – lingvistika („společný jazyk“)
- počítačová lingvistika (matematická, počítačnická), jazykové inženýrství, počítačové zpracování přirozeného jazyka
- obor zpracování přirozeného jazyka (Natural Language Processing, NLP)
- hlavní oblasti (uživatelský přístup)
 - syntéza a analýza řeči
 - počítačová lexikografie
 - formální analýza jazyka (morfologická, slovo tvorná, syntaktická, sémantická, textová)
 - korpusová lingvistika
 - dialogové systémy, umělá inteligence

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

Obsah kurzu

- **počítačová lexikografie** – DEBDict, DEBWrite, lexikální databáze, Vokabulář webový a další
- **korpusová lingvistika** – KonText, Sketch Engine (2 semináře)
- **morfologická analýza** – Ajka, Majka, Morče, MorphoDiTa, atributivní a poziční systém
- **derivační rozhraní** – Deriv, Morfio a další
- **syntaktická analýza** – Synt, Set, PDT (stromové banky)
- **sémantická analýza** – WordNet, FrameNet, VerbNet
- **valenční databáze** – Vallex, VerbaLex
- **slovotvorba** – Deriv, Morfio, DeriNet
- **rozpoznávání a syntéza řeči**

Aplikace k využití

- Internetová jazyková příručka – <http://prirucka.ujc.cas.cz>
- Nový encyklopedický slovník češtiny <https://www.czechency.org/>
- překlady:
 - WebMetaTrans – <http://metatrans.fi.muni.cz>
 - databáze překladových ekvivalentů <https://treq.korpus.cz/>
- sociální robot Karel Pepper – <https://nlp.fi.muni.cz/trac/pepper/wiki/NlpPepperShows>
- chatbot - <https://www.elbot.com/>
- Uhádni slovo! https://nlp.fi.muni.cz/projekty/uhadni_to_slovo/

Dotazy?

Děkuji za pozornost.

