

Úvod



Počítačové nástroje pro češtinu
Jaro 2020

Helena Geržová



Organizační informace

- docházka (2 absence)
- seminární úkol
- 3 kredity
- kontaktovat cvičící

Co je to počítačová lingvistika

- obor mezi informatikou a lingvistikou
- detailní analýza jazyka a jeho formální popis
- výsledkem jsou denně používané aplikace
 - korektor překlepů a gramatiky
 - vyhledávání na webu
 - prediktivní psaní
 - překladače jazyků



Počítačové zpracování přirozeného jazyka

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

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

- urychlení a zefektivnění práce lingvisty
- ověřování existujících teorií
- objevení nového jazykového jevu, zákonitosti
- co mohu a nemohu od nástroje očekávat
- co a jak mohu 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, počítačová), jazykové inženýrství, počítačové 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

Obsah kurzu

- **počítačová lexikografie** – DEBDict, DEBWrite, lexikální databáze, Vokabulář webový a další
- **korpusová lingvistika** – KonText, Sketch Engine
- **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**
- seminární práce, hry

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, JÚĽŠ SAV Bratislava – <http://korpus.juls.savba.sk/>

Bonus

- Internetová jazyková příručka – <http://prirucka.ujc.cas.cz>
- WebMetaTrans – <http://metatrans.fi.muni.cz>
- sociální robot Karel Pepper –
<https://nlp.fi.muni.cz/trac/pepper/wiki/NlpPepperShows>
- jak fungují počítače –
<https://cs.khanacademy.org/computing/computer-science/how-computers-work2/v/khan-academy-and-codeorg-introducing-how-computers-work>
- chatbot - <https://www.elbot.com/>

Úkol do 29. 2. 2020

- doplnit celkem 15 slov s jejich definicí (1. slovo je příklad, nahradíte je); definice nemusí být dlouhé, dokonale přesné, ani obsáhlé, mohou být naopak (není to podmínkou) kreativní i vtipné; mustr pro slovníkové výrazy je k nalezení ve studijních materiálech:
https://is.muni.cz/el/phil/jaro2020/CJBB85/um/01_uvod/
- zaregistrovat se na http://deb.fi.muni.cz/proj_debwrite.php & poslat uživatelské jméno na mrkyvka@phil.muni.cz.

Děkuji za pozornost

