

Základy využití korpusu v praxi **cjbb75**

Středa: 9.10-10.50 G13

24. 2. Úvodní hodina (definice **korpusu**, korpusová pracoviště a dostupné korpusy, korpus a internet, korpusové **manažery**, **ÚČNK**, webové rozhraní **Kontext**, registrace uživatele).

2. 3. Dú: získat přístupová práva (**uživatelské jméno/heslo**), přečíst si charakteristiky dostupných korpusů ÚČNK, znát odpovědi na otázky v hdt.

9. 3. Kontext – **příkazový řádek** – vyhledávání podle atributů (**word, lemma, tag, lc**), **cql**, **regulární výrazy**, **KWIC**, **konkordanční seznam a frekvenční seznam** jako zdroj poznání o jazykových vlastnostech vyhledaných jednotek (způsob psaní/pravopis: přejatá slova /*mailovat/mejlovat/majlovat*/, aplikace pravidel – varianty /*ismus/izmus*/, varianty pravopisu zdobnělin vlastních jmen – Terinka/Terynka).

16. 3. Dú: Jaká jsou lemmata/tvary adaptovaných slov a jaké je frekvenční rozložení variantních způsobů jejich grafické realizace (slovesa utvořená od anglických základů: *použí(va)t Skype = ????*, *použí(va)t google = ???*, *použí(va)t twitter = ???*).

23. 3. Korpus jako slovník – počítačová lexikografie. **Slovníky založené na korpusech**. Využití korpusu jako on-line slovníku.

30. 3. Dú: cvičení na hledání významů slov – prezentace.

6. 4. Korpus jako zdroj dat pro výzkum slovo tvorby. Jak zadat dotaz pro hledání slov, která mají společné slovo tvorné vlastnosti (lingvistické termíny: morfém, afix, kořen, kmen, slovní druh, substantivum, adjektivum, sloveso, ..., gramatická kategorie, rod, číslo, pád, osoba ...; termíny pro formální zadání: řetězec, token, lemma, morfologická značka/tag, pozice/atribut, hodnota, technické řešení, ...). Adjektiva od sloves typu *lyžující*. Slovní formulace – tvrzení, která musí platit, podmínka nutná/dostačující, využití regulárních výrazů a logických operací, využití filtrů, pozorování dat a úprava dotazu.

13. 4. Jak lze v korpusu vyhledat substantiva tvořená příponou –č, která označují osoby vykonávající činnost označenou základovým slovesem: *řidič, nosič, sběrač, očišřovač, hráč, ...*

20. 4. Korpus jako zdroj dat pro výzkum tvarosloví. Jak zadat dotaz pro hledání slov, která mají společné tvaroslovné vlastnosti (lingvistické termíny: morfém, afix, kořen, kmen, slovní druh, substantivum, adjektivum, sloveso, ..., gramatická kategorie, rod, číslo, pád, osoba ..., vzor, třída, ...; termíny pro formální zadání: řetězec, token, lemma, morfologická značka/tag, pozice/atribut, hodnota, technické řešení, ...). Slovesa III. třídy vzor *kupovat*. Slovní formulace – tvrzení, která musí platit, podmínka nutná/dostačující, využití regulárních výrazů a logických operací, využití filtrů, pozorování dat a úprava dotazu.

27. 4. Jak lze v korpusu vyhledat slovesa I. třídy vzoru *nést*, substantiva skloňovaná podle vzoru *žena*, adjektiva typu *otcův/matčín*.

4. 5. Korpus jako zdroj dat pro výzkum syntaxe. Syntaktické kritérium určení slovnědruhově platnosti slov. (lingvistické termíny: slovní druh, větný člen, primární a sekundární větněčlenské funkce slovních

druhů, slovnědruhový přechod/transpozice, substantivizace, prepozicionalizace, slovosled uvnitř jmenné skupiny, gramatická shoda...; termíny pro formální zadání: řetězec, token, lemma, morfologická značka/tag, pravý a levý kontext, pozice vpravo a vlevo od KWIC, filtr, ...). Jakou slovnědruhovou platnost může mít slovo *hnát*, *popravčí*, *pila*? Které syntaktické vlastnosti můžeme využít, abychom v korpusu našli chyby v disambiguaci?

11. 5. Sleduj v korpusu slovnědruhové značkování jednotek (*kolem*, *místo*, *výchovné*), které lze interpretovat vícero způsoby a pokus se navrhnout, jak postupovat při odhalení chyb v disambiguaci.
„Místo vašich schůzek sis měl lépe vybírat.“

18. 5. Písemka

Podmínky udělení zápočtu:

1. Odevzdání všech domácích úkolů. Úkoly je třeba dodávat průběžně. Na začátku každé hodiny může být kdokoli požádán, aby vysvětlil, oč v úkolu šlo. Je také možné zodpovědět případné nejasnosti.
2. Písemný test 18. 5.