

CJBB75 Základy využití korpusu v praxi G13 po 10.00-11.30

Harmonogram

Výuka v tomto předmětu bude zaměřena na využití jazykových korpusů v lingvistickém výzkumu. K tomuto účelu bude studentům představeny především korpusy a korpusové nástroje ÚČNK. Dovednosti s užíváním korpusových nástrojů jsou ovšem transformovatelné při práci s dalšími jazykovými korpusy a nástroji.

Výuka bude probíhat jednak formou kontaktní výuky, kdy pedagog demonstruje, jak lze z jazykového korpusu získat data, která vytvoří oporu pro odpověď na předem stanovenou lingvistickou hypotézu. V následující nekontaktní hodině se studenti budou samostatně věnovat vyřešení obdobných úkolů. Úkoly odevzdají do **Odevzdávárny**, a to nejpozději **tři pracovní dny** před konáním následující kontaktní výuky. Pedagog na začátku následující hodiny ohodnotí úkol a rozebere možnosti řešení (včetně doplnění toho, nač studenti nepřišli).

20. 2. Úvodní hodina (definice **korpusu**, korpusová pracoviště a dostupné korpusy, korpus a internet, korpusové **manažery**, **ÚČNK**, webové rozhraní **Kontext**). Kontext – **příkazový řádek** – vyhledávání podle atributů (**word, lemma, tag, lc**), **cql, regulární výrazy, KWIC, konkordanční seznam a frekvenční seznam, Menu Zobrazení**. Cvičení je zaměřeno na tvorbu dotazu v jazyce cql se zřetelem k vyhledávání slov, lemmat, pozorování jejich kontextu a úrovně jejich lemmatizace a pos_taggingu.

27. 2. Úkol

6. 3. Korpus jako slovník – počítačová lexikografie. **Slovníky založené na korpusech**. Využití korpusu jako on-line slovníku. Práce s kolokacemi, nástroj Treq, nástroj Sketch engine - slovní profily.

13.3. Úkol

20. 3. Korpus jako zdroj dat pro výzkum slovo tvorby. Jak zadat dotaz pro hledání slov, která mají společné slovo tvorné vlastnosti. Slovní formulace – tvrzení, která musí platit, podmínka nutná/dostačující, využití regulárních výrazů a logických operací, využití filtrů, pozorování dat a úprava dotazu.

27. 3. Úkol

3. 4. Korpus jako zdroj dat pro výzkum tvarosloví. Tvary substantiv. Slovní formulace – tvrzení, která musí platit, podmínka nutná/dostačující, využití regulárních výrazů a logických operací, využití filtrů, pozorování dat a úprava dotazu.

17. 4. Úkol

24. 4. Korpus jako zdroj dat pro výzkum syntaxe. Syntaktické kritérium určení slovnědruhové platnosti slovního tvaru. Které syntaktické vlastnosti můžeme využít, abychom v korpusu našli chyby v disambiguaci?

1.5. úkol

8.5. On-line cvičný test

15. 5. Písemka : praktické zvládnutí formulace dotazu v cql