

Státnicové Bc. okruhy DH KOL UPOL

1. Základní koncepty a teoretické přístupy obecné lingvistiky (lingvistické plány, synchronní a diachronní hledisko, strukturní a generativní modely jazyka ad.)
Přínos a důležitost jazykovědy.
2. Základní koncepty sémiotiky (u F. de Saussura, R. Jakobsona, Ch. S. Peirce, U. Eco, R. Barthesa a dalších). Dyadický a triadický znakový model. Semióza. Typologie znaků. Účel a cíl sémiotiky.
3. Základní reflexe vědy: (1) filosofie vědy, (2) historiografie vědy, (3) sociologie vědy a (4) datová věda o vědě. Pro každou reflexi specifikovat: (a) okolnosti jejího vzniku, (b) hlavní témata, kterým se věnuje, (c) výhody dané reflexe a (d) úskalí jednostranného uplatnění dané reflexe.
4. Charakterizujte pojem „distant reading“ (čtení s odstupem). Vysvětlete vztah „distant reading“ k pojmu „close reading“ z oblasti literárních studií. Na jaké typy textů se distant reading zaměřuje? V jaké škále a s jakou mírou rozlišení je analyzuje? Na příkladu vybrané studie Franka Morettiho přibližte, jakou otázku si autor v studii klade a jakými metodami na ni hledá odpověď. Vysvětlete, co má „distant reading“ společného s pojmem „cultural analytics“ (kulturní analytika), který zavedl Lev Manovich. Popište alespoň jednu oblast kultury a specifického média, na něž se kulturní analytika v Manovichově pojetí zaměřuje. Uveďte příklady technik, pomocí kterých Manovich tuto oblast studuje.
5. Základní přístupy a metody forenzní lingvistiky, měřitelné vlastnosti textů a jejich využití při určování autorství, cíl užití vícerozměrných metod jako hierarchické zhlukování, MDS, PCA a další. Počítání vzdáleností a podobností: Jaccardův koeficient, Diceův koeficient, Manhattanská, Euklidovská, Minkowského a Čebyševova vzdálenost, kosinová nepodobnost, jejich účel a praktické důsledky. Vektorizace textu: indexy a model bag-of-words.
6. Popisná statistika, její cíl a nástroje: průměr, medián, modus, směrodatná odchylka, rozptyl, kvantil, korelace, kovariance, rozdělení. Grafy: bodový graf, histogram, krabicový graf, jejich účel, popis, výhody, nevýhody a klasická řešení

nevýhod. Matice: kovarianční, korelační, matice bodových grafů, jejich účel, výhody, nevýhody a použití. Vztah popisné statistiky se softwarovými nástroji jako jazyk R, Python a Excel.

7. Kognitivní bias: definice, způsoby vypořádání, známé případy a paradoxy. Design experimentů: teorie, hypotéza, interpretace, výběr vzorku, slepý pokus, dvojité a trojitě slepý pokus, kontrolní skupina, důvody využití statistických testů, vztah ke kognitivním biasům, účel a odůvodnění.
8. Práce s vícerozměrnými daty (způsoby detekce redundantních vlastností a jejich odstranění, metody PCA, SVD a jejich účel, vizualizace vícerozměrných dat pomocí PCA, SVD, MDS a tSNE, způsob měření rekonstruované informace). Vytěžování vícerozměrných dat (účel a metody shlukování: hierarchické shlukování, k-průměrů, DBSCAN ad., problematika „prokletí dimenzionality“, shlukování a výpočty vzdáleností/podobností, klasifikační metoda /k/ nejbližších prvků, vztah ke strojovému učení).
9. Analýza sítí a grafů: definice grafu, sítě a aplikace, identifikace klíčových rolí v (sociálních) sítích, k tomuto používané ukazatele, jejich hlavní rozdíly a typy rolí, které dokáží odhalit, detekce komunit, problematika vizualizace sociálních sítí, vztah ke grafovým databázím. Účel, využití a praktické aplikace.
10. Práce se sekvencemi, jejich účel a použití v humanitních a dalších vědách. Vztah sekvencí a textů. Porovnání sekvencí, výpočty vzdálenosti, způsoby modelování, příklady užití a aplikací, Hammingova vzdálenost, Levenstheinova vzdálenost, Damera-Levenstheinova vzdálenost, Needleman-Wunsch, Markovovské procesy a skryté markovovské procesy.
11. Strukturovaná a nestrukturovaná data, rozdíly, důsledky pro jejich vytěžování a příklady jejich zdrojů; role metadat a metainformací, jejich interpretace, příklady. Formáty txt (plain-text), CSV/TSV, XML a JSON. SQL databáze a jejich účel, příklad vložení, vyhledání, aktualizace a vymazání dat. NoSQL databáze a jejich účel. Grafové databáze, jejich účel a příklad vytvoření, nalezení, aktualizace a vymazání dat. Vektorové databáze a jejich účel.
12. Programovací jazyky Python, R, HTML, CSS, JavaScript, jejich účel, využití a specifika při analýze a zpracování jazyka, dat a vytváření interaktivních aplikací.

13. Python frameworky Sanic, Django, Flask, návrh MVC, REST API. Virtualizace, Docker, operační systém Linux a základní práce s ním v příkazové řádce, SSH, GIT. Základní koncepce protokolu HTTP. Obecné zásady programování.