

Štátne skúšky – okruhy tém k štátnicovému predmetu

Akademický rok: 2022/2023

Garantujúce pracovisko: Katedra kybernetiky a umelej inteligencie FEI TU – KKUI

Študijný program: **Hospodárska informatika**

Študijný odbor: **Informatika**

Stupeň štúdia: 2. – inžiniersky

Štátnicový predmet: **Hlavné poznatky odboru hospodárska informatika a ich využitie**

Okruhy otázok z predmetu:

1. Vzťahy a rozdiely medzi dátami, informáciami a znalosťami, význam znalostí pre organizácie. Rôzne uhly pohľadu na manažment znalostí (MZ), ich stručná charakteristika a náplň. Podmieneny pohľad na MZ: vplyv charakteristík riešených úloh, typu používaných znalostí, charakteristík organizácie a vonkajšieho prostredia na MZ.
2. Vyhľadávanie informácií, základné modely pre vyhľadávanie informácií, spôsoby vyhodnotenia kvality vyhľadávania. Vyhľadávanie na webe, využitie štruktúry liniek, algoritmy PageRank a HITS.
3. Proces objavovania znalostí z databáz, popis jeho základných krokov. Základné typy znalostí, ktoré je možné získavať z dát. Metodológia CRISP-DM, popis jej jednotlivých krokov. Typy dát, metódy na pochopenie a prípravu dát, príklady aplikácie jednotlivých metód. Metriky a spôsoby vyhodnotenia vytvorených modelov. Príklady aplikácií OZ v praxi.
4. Prediktívne dolovanie v dátach: definícia, klasifikácia a predikcia (regresia), základný postup pri prediktívnom dolovaní v dátach. Výber príznakov pre klasifikáciu. Rozhodovacie stromy, pravidlové klasifikátory. Pravdepodobnostná klasifikácia: Bayesovská klasifikácia a logistická regresia. Klasifikátory na princípe k-najbližších susedov. Semi-kontrolované učenie. Aktívne učenie. Zložená klasifikácia (učenie súborom metód). Vyhodnotenie kvality klasifikátorov.
5. Zhlukovanie, príklady aplikácií. Výber príznakov pre zhlukovanie: filtrovacie, obálkové a hybridné prístupy. Metódy zhlukovania založené na výbere reprezentantov zhlukov (k-stredové metódy). Metódy hierarchického zhlukovania: aglomeratívne a divízivne. Na mriežke založené zhlukovanie, na hustote založené zhlukovanie, na grafoch založené zhlukovanie. Validácia zhlukov: interné a externé kritériá validity zhlukov.
6. Detekcia anomálií, príklady aplikácií, vybrané typy metód na detekciu anomálií (napr. analýza extrémnych hodnôt, metódy založené na vzdialenosti, metódy založené na hustote). Validita anomálií: ROC krivka. Asociačné pravidlá (AP), základné pojmy, mierky zaujímavosti: podpora, spoľahlivosť, lift. Princíp monotónnosti podpory, algoritmus Apriori. Generovanie asociačných pravidiel. Meta-algoritmy AP: vzorkovanie, hierarchické AP, kvantitatívne AP.
7. Objavovanie znalostí v textoch. Predspracovanie textov do vhodnej reprezentácie. Typy úloh objavovania znalostí v textoch: klasifikácia, analýza sentimentu, zhlukovanie, extrahovanie tém a extrahovanie informácií.
8. Základné spôsoby prehľadávania. Lokálne prehľadávanie a jeho komponenty. Iteračné zlepšovanie a únik z lokálneho extrému. Prírodne inšpirované algoritmy a algoritmy pre odhad distribúcie. Výpočtová zložitosť, základné zložitosťné triedy

9. Systémy a manažérske informačné systémy (MIS), charakteristika systémov MIS, metódy a prostriedky realizácie manažérskej úrovne riadenia. Trojvrstvová architektúra systémov MIS (relačná vrstva, multidimenzionálna vrstva, prezentačná vrstva - prostriedky OLAP). Príklady aplikácií MIS v praxi.
10. Relačná vrstva, systémy OLTP ako produkčné systémy a zdroje dát pre MIS, relačný databázový model, charakteristika relačných databáz, podmienky minimálnej relačnosti, pravidlá pre relačné databázové systémy, databázové tabuľky-entity, relačné vzťahy medzi entitami, primárny kľúč a cudzí kľúč, relácie, tri normálne formy.
11. Multidimenzionálna vrstva a multidimenzionálny dátový model, definícia a charakteristika dátových skladov, prvky dátových skladov a ich charakteristika (dimenzie, fakty, metadáta, repozitáre). Metódy, postupy a prostriedky budovania a napĺňania dátových skladov. Prezentačná vrstva, OLAP analýza, definícia a pravidlá OLAP, operácie a spôsoby prezentácie výstupov OLAP analýzy.
12. Definícia veľkých dát, charakteristika veľkých dát – 3V/5V model, princípy distribuovaných súborových systémov a databáz, typy NoSQL databáz a ich charakteristika – stĺpcovo orientované databázy, úložiska typu kľúč : hodnota, dokumentové databázy, grafové databázy.
13. Architektúry systémov pre spracovanie veľkých dát. Distribuované prostredie pre dávkové spracovanie dát – základný model MapReduce. Distribuované prostredie pre spracovanie prúdových dát – dátové toky, fronty správ. Kombinovaná architektúra. Prehľad technológií pre spracovanie veľkých dát.

Košice, 09.02.2023

doc. Ing. Peter Papcun, PhD.
vedúci katedry

