

## **Ispitna pitanja iz predmeta Pretraživanje informacija**

### I grupa: IR sistemi i Bulov model pretraživanja

1. Pojam pretraživanja informacija i IR sistema
2. Nacrtati strukturu IR sistema i objanisti ulogu svake komponente u njemu.
3. Šta definiše model pretraživanja, kako se modeli pretraživanja dele?
4. Objasniti Bulov model pretraživanja.
5. Objasniti implementaciju Bulovog modela pretraživanja korišćenjem matrice termina i dokumenata
6. Objasniti implementaciju Bulovog modela pretraživanja korišćenjem invertovanog indeksa
7. Rangiranje dokumenata u proširenom Bulovom modelu pretraživanja
8. Proširenje Bulovog modela pretraživanja dodavanjem parametara blizine.

### II grupa: Invertovani indeks

1. Šta je invertovani indeks? Objasniti osnovnu strukturu invertovanog indeksa.
2. Koji su osnovni zadaci sistema za automatsko indeksiranje dokumenata? Kratko objasniti svaki od njih.
3. Šta su stemeri? Objasniti 2 osnovne metode u implementaciji stemera.
4. Navesti načine za predstavljanje rečnika u IR sistemima i prednosti i nedostatke svakog od njih.
5. Navesti algoritme za kreiranje invertovanog indeksa i kratko objasniti svaki od njih.
6. Objasniti postupak distribuiranog indeksiranja.
7. Navesti metode za kompresiju rečnika termina i kratko objasniti svaku od njih.
8. Navesti metode za kompresiju postig listi i kratko objasniti svaku od njih.

### III grupa: Vektorski model pretraživanja i evaluacija IR sistema

1. Objasniti vektorski model pretraživanja.
2. Navesti mere koje se u vektorskom modelu pretraživanja koriste kao težina termina u dokumentu ili upitu i objasniti logaritamsku frekvencu pojavljivanja termina
3. Navesti mere koje se u vektorskom modelu pretraživanja koriste kao težina termina u dokumentu ili upitu i objasniti *tf-idf* meru
4. Objasniti algoritam traženja u vektorskom modelu pretraživanja
5. Objasniti neheurističke metode za optimizaciju algoritma traženja u vektorskom modelu pretraživanja
6. Objasniti heurističke metode za optimizaciju algoritma traženja u vektorskom modelu pretraživanja
7. Navesti mere koje se koriste za evaluaciju IR sistema
8. Objanisti postupak testiranja IR sistema. Šta standardna kolekcija za testiranje treba da sadrži?

#### IV grupa: Pretraživanje weba

1. Šta čini pretraživanje weba komplikovanijim u odnosu na pretraživanje poznatog korpusa dokumenata?
2. Šta je SEO? Navesti i kratko objasniti tehnike koje SEO koristi.
3. Kako se vrši procena kvaliteta indeksiranja weba?
4. Navesti i kratko objasniti metode za prepoznavanje duplikata na Webu.
5. Šta je crawler?
6. Šta definiše fajl robot.txt?
7. Objasniti algoritam crawl-ovanja?
8. Objasniti arhitekturu web crawlera.
9. Objasniti PageRank algoritam za rangiranje stranica na osnovu linkova.
10. Objasniti HITS algoritam za rangiranje stranica na osnovu linkova.

---

**NAPOMENA:** Ovo su primeri kako ispitna pitanja mogu da izgledaju, što ne znači da će pitanja izgledati doslovno tako.