

Razni zadaci

Zadaci

9.1.1 Neka originalni rečnik indeksiranja nekog dokumenta sadrži termine: **ROSA, KOSA, KORA, BOR**. Sistem za pronalaženje predviđa pronalaženje po upitnim terminima koji dozvoljavaju sve vrste odsecanja.

1. Sastaviti prošireni indeks koji će ovo omogućiti.
2. Ako korisnik postavi navedene upite, kako ove upite treba preformulisati da bi se oni prosledili proširenom indeksu, šta oni pronalaze u proširenom indeksu i kojim to originalnim terminima odgovara:
 - (a) BOR
 - (b) KO*
 - (c) *SA
 - (d) *O*
 - (e) K*A

9.1.2 Neka indeks neke kolekcije dokumenata sadrži sledeće termine:

BAS, DRVO, FAR, AUTO, CENTAR, GOL, ZNAK, STOP

. Kako izgleda sortirani indeks i

1. linearno
2. binarno pretraživanje termina **GOL** i **PENAL**.

Za date indeksne termine formirati

1. binarno pretraživačko stablo
2. balansirano B-stablo (kapaciteta 3)?

9.1.3 Za obrazac **G T A C A G T A** formirati tablicu pomeranja, a zatim demonstrirati rad KMP algoritma pri pretraživanju teksta **G T G T A C G G T A C A G T A**.

9.1.4 Za obrazac **G T A C A G T A** formirati tablicu pomeranja, a zatim demonstrirati rad BM algoritma pri pretraživanju teksta **G T G T A C C G T A C A G T A**.

9.1.5 Za obrazac **B C C B D D B C C** formirati tablicu pomeranja, a zatim demonstrirati rad BM algoritma pri pretraživanju teksta **B C C D M B C C E D B A A N B D D B C C B C**.

9.1.6 U kolekciji od $N = 240,000$ dokumenata, pronalaze se dokumenti koji odgovaraju upitu q . U kolekciji postoji 24,600 dokumenata relevantnih za upit q , ali se u rezultatu pretrage našlo 12,800. U rezultatu pretrage javlja se i 3,600 dokumenata iz ostatka kolekcije. Izračunati odziv, preciznost, tačnost, kao i $F_{0.5}$ meru sistema za pronalaženje u odnosu na taj upit q .