

Informatika IV

Školska 2004/2005. godina

Literatura:

- G. Salton: Automatic Text Processing, Addison-Wesley Publishing Company, 1989. (glave 8, 9 i 11).
- S. Jones: Text and Context - Document Processing and Storage, Springer-Verlag, 1991 (glave 3 i 10).
- J. Allen: Natural Language Processing, The Benjamin/Cummings Publishing Company, Inc., 1987 (glave 1,2,3,4,7 i 8)
- Barnbrook, G.: Language and computers, Edinburg textbooks in empirical linguistics, 1996
- C. Krstev: Uvod u obeležavanje teksta,
<http://poincare.matf.bg.ac.rs/~cvetana/kurs-markup/index.html>
- C. Krstev: Uvod u XML,
<http://poincare.matf.bg.ac.rs/~cvetana/kurs-xml/index.html>
- C. Krstev: Formati za predstavljanje informacija: otvorena pitanja, Infoteka 1(2), 2000
- C. Krstev: Digitalne biblioteke - razgraničenje pojmova, Infoteka 3(2002)1
- D. Vitas: WordNet baza leksičkih relacija, Svet reči, 13-14, 2001
- Cvetana Krstev: Processing of Serbian, Faculty of Philology, Belgrade 2008, chapter 1
- Cvetana Krstev, Bojana Đorđević, Sanja Antičić, Nevena Ivković-Berček, Zorica Zorica, Vesna Crnogorac, Ljiljana Macura, Kooperativan rad na dogradnji srpskog wordneta, in INFOTEKA : časopis za informatiku i bibliotekarstvo, vol. IX, No.1-2, pp. 57-75, Beograd, 2008.
<http://infoteka.bg.ac.rs/index.php/sr/2014-04-13-15-19-17/2008/136-p-r-iv-n-r-d-n-d-gr-dnji-srps-g-wordneta>
- Biljana Kosanović Pristup naučnim informacijama u Srbiji / iskustvo nakon šest godina
http://infoteka.bg.ac.rs/pdf/Srp/2008/INFOTHECA_IX_1-2_May2008_77-81.pdf

Ispit:

Praktikum, seminarski, pismeni i usmeni

Praktikum se polaže za računarom i podrazumeva poznavanje jezika za rad sa niskama **awk**. Za izradu zadatka na praktikumu na raspolaganju je jedan sat. Položen praktikum nije uslov za potpis ali je obavezan za izlazak na pismeni ispit. Praktikum se polaže u svim redovnim ispitnim rokovima (jun, septembar, oktobar, januar, april). Polaganje praktikuma je nezavisno od polaganja ispita, odnosno ne mora se polagati u istom roku kada i ispit. Jednom položeni praktikum se ne polaže ponovo. Praktikum se ne prijavljuje ispitnom prijavom, ali je obavezno njegovo prijavljivanje u Sekretarijatu Katedre za bibliotekarstvo i informatiku, soba 529, isključivo u terminima oglašanim na oglasnoj tabli Katedre, po pravilu 2-4 nedelje pre ispitnog roka. Ako imate ikakvih dilema u vezi polaganja praktikuma, obratite se Sekretarijatu Katedre.

Seminarski rad studenti dobijaju u drugom semestru i rade ga u okviru obaveznih vežbi iz Informatike IV.

Ispitna pitanja:

Ispitna pitanja za studente koji su Informatiku IV i Primenjenu lingvistiku slušali 2004/05. i 2005/06. školske godine i za stare studente koji se prilagođavaju novom nastavnom planu (nova pitanja su označena crvenom bojom)

1. Organizacija indeksno-sekvencijalnih datoteka.
2. Indeksno-sekvencijalne datoteke i Bulovski upiti. Algoritam spajanja listi dokumenata (upit **or**).
3. Invertovano-indeksne operacije i uslovi udaljenosti.
4. Invertovano-indeksne operacije i težine termina.
5. Invertovano-indeksne operacije i specifikacija sinonima.
6. Invertovano-indeksne operacije i odsecanje termina.
7. Tipična organizacija datoteka kod invertovano-indeksnih sistema.
8. Optimizacija rada sa invertovanim listama smanjivanjem broja indeksnih termina.
9. Optimizacija rada sa invertovanim listama traženjem optimalne formulacije upita.
10. Sistemi za pronalaženje informacija zasnovani na skaniranju teksta. Jednostavni algoritam za skaniranje teksta.
11. Poboljšani algoritam skaniranja teksta.
12. Pojmovi odziv i preciznost kod sistema za pronalaženje informacija.
13. Automatsko indeksiranje --- indeksno okruženje i ciljevi indeksiranja.
14. Indeksiranje pojedinačnim terminima --- frekvencija termina.
15. Indeksiranje pojedinačnim terminima --- diskriminatorna vrednost termina.
16. Veze termina pri indeksiranju --- generisanje kompleksnih identifikatora.
17. Veze termina pri indeksiranju --- formiranje fraznih termina.
18. Veze termina pri indeksiranju --- transformacija tezaursom.
19. Automatsko konstruisanje tezaurusa.
20. Plan za automatsko indeksiranje.
21. Plan za automatsko indeksiranje s korišćenjem povratne informacije o relevantnosti.
22. Baze podataka: Medline, ScienceDirect, Dialog, Questel-Orbit, Lexis-Nexis, Google Scholar, SCI, Cobson.
23. Elektronski časopisi i njihovo korišćenje.
24. Jezici za obeležavanje teksta.
25. Proširiv jezik za obeležavanje --- XML.
26. Definicija tipa dokumenta --- deklaracije elemenata.
27. Definicija tipa dokumenta --- deklaracije atributa.
28. Definicija tipa dokumenta --- deklaracije i korišćenje entiteta.
29. XML komentari, procedure za obradu, označeni odeljci.
30. XML javni identifikatori.
31. XML šeme.
32. Primer korišćenja SGML - Oksfordski rečnik engleskog jezika.
33. Primer korišćenja SGML - TEI (Podsticaj kodiranju teksta) - osnovni moduli.
34. TEI - a-klase elemenata. Primeri.
35. TEI - m-klase elemenata. Primeri.
36. TEI - osnovni modeli sadržaja (phrase.seq, paraContent, specialPara).
37. TEI - strukturiranje dokumenata (elementi text, group, front i bottom i njihov sadržaj).
38. TEI - zaglavlje.
39. Korpusi --- vrste i polja primene.
40. Primena korpusa u leksikografiji --- projekat Cobuild.
41. Kontekstno slobodne gramatike i parsiranje odozgo nadole.
42. Rekurzivne mreže prelaska i parsiranje odozgo nadole.
43. Sintaksička analiza --- tehnike odozgo nadole i odozdo nagore.

44. Znanje o jeziku u sistemima za obradu prirodnih jezika.
45. Elektronski rečnici u analizi prirodnih jezika (morfološki, semantički, Wordnet).