

Sveučilište u Rijeci
ODJEL ZA INFORMATIKU
Radmile Matejčić 2, Rijeka
Akademska 2013./2014. godina

RAČUNALNA ANALIZA PRIRODNOG JEZIKA

Studiji: Diplomski studij informatike IKS OBAVEZNI + PI IZBORNI
ECTS bodovi: 5
Nastavno opterećenje: 2 + 2
Web stranica predmeta: <http://mudri.uniri.hr>
http://www.inf.uniri.hr/index.php?option=com_classes&Itemid=40&task=displayclass&id=2268

Studiji: Diplomski studij informatike (dvopredmetni) IZBORNI
ECTS bodovi: 4
Nastavno opterećenje: 2 + 2
Web stranica predmeta: <http://mudri.uniri.hr>
http://www.inf.uniri.hr/index.php?option=com_classes&Itemid=40&task=displayclass&id=2268

Nositeljica predmeta:

doc. dr. sc. Sanda Martinčić-Ipšić
e-mail: smarti@inf.uniri.hr
Ured: Radmile Matejčić 2, Rijeka, 409
Vrijeme konzultacija: utorkom po prethodnom
dogovoru e-mailom

Asistent:

Luka Vretenar
e-mail: lvretenar@inf.uniri.hr
Ured: Radmile Matejčić 2, Rijeka, 419
Vrijeme konzultacija: srijedom po prethodnom
dogovoru e-mailom

RAČUNALNA ANALIZA PRIRODNOG JEZIKA

Razvijanje općih i specifičnih kompetencija (znanja i vještina)

Studenti trebaju steći temeljna znanja o postupcima računalne analize prirodnog jezika i računalne lingvistike.

Korespondentnost i korelativnost programa

Program kolegija povezan je sa kolegijima Formalni jezici i jezični procesori 1.

Okvirni sadržaj predmeta

Uvod u područje i pregled pojmova računalne analiza prirodnog jezika. Računalna lingvistika i jezične tehnologije. Jezični resursi, korpusi, rječnici, leksikoni.

Prepoznavanje sintakse. Vjerojatnostni modeli. N-gramski modeli. Perpleksnost. Jezični alati za provjeru gramatike i pravopisa. Morfološki analizatori. Obilježivači vrsta riječi. Parsiranje pomoću kontekstno neovisnih gramatika.

Prepoznavanje semantike. Prepoznavanje leksičkoga i rečeničkoga značenja.

Prepoznavanje pragmatike. Diskurs. Dijalog. Generiranje jezika.

Identifikacija jezika. Strojno prevođenje i strojno potpomognuto prevođenje.

Pretraživanje dokumenata. Strukture indeksiranja dokumenata.

Oblici provođenja nastave i način provjere znanja

Predavanja, vježbe, samostalni zadaci, konzultacije, seminarski rad, E-učenje

Popis literature potrebne za studij i polaganje ispita

1. D. Jurafsky, J. H. Martin: Speech and Language Processing, An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, Prentice Hall, 2000.
2. C. Manning, H. Schütze: Foundations of Statistical NLP, MIT Press, Cambridge, Massachusetts, 1999.

Popis literature koja se preporučuje kao dopunska

1. J. Allen: Natural Language Understanding
2. Minker, W., Bennacef, S. Speech and human-Machine dialog. Kluwer Academic Publishers. Boston. 2004.
3. P. Nugues An Introduction to Language Processing with Perl and Prolog, Springer, 2006
4. http://www.cs.lth.se/home/Pierre_Nugues/ilppp/slides.shtml
5. M. Covington, Natural Language Processing for Prolog Programmers, Prentice Hall 1994.
6. S. Bird, E. Klein, E. Loper: Natural Language Processing with Python, O'Riley 2009

Način praćenja kvalitete i uspješnosti izvedbe predmeta

Kroz ustrojeni sustav osiguranja kvalitete Odjela za informatiku.

Preduvjeti za upis predmeta

R. BR.	OČEKIVANI ISHODI
1.	Definirati i objasniti postupke i pojmove računalne analize prirodnog jezika.
2.	Razumjeti postupke i standarde izgradnje jezičnih resursa.
3.	Izraditi postupke za izgradnju i analizu jezičnih resursa.
4.	Objasniti postupke morfološke analize teksta te obilježavanja vrste riječi.
5.	Izgraditi statistički jezični model te definirati i razumjeti postupke izgradnje.
6.	Razumjeti postupke Early-evog i CYK parsiranja.
7.	Objasniti i definirati formalizme za zapis značenja u jezičnim aplikacijama.
8.	Razumjeti postupke semantičke analize i ekstrakcije podataka iz nestrukturiranih izvora.
9.	Objasnite postupke automatske klasifikacije teksta i naprednog pretraživanja dokumenta.
10.	Nabrojati i definirati postupke analize diskursa.
11.	Objasniti modele i strukturu sustava za automatsko vođenje dijaloga s računalom te ih primijeniti u samostalnom radu.
12.	Objasniti principe strojno potpomognutog prevođenja i/ili određivanja polariteta teksta.

AKTIVNOSTI I OCJENJIVANJE STUDENATA

VRSTA AKTIVNOSTI	ECTS	ISHODI UČENJA	SPECIFIČNA AKTIVNOST	METODA PROCJENJIVANJA	BODOVI MAX.
Prisutnost na nastavi	1	1-12	Prisutnost studenata	Popisivanje (evidencija)	0
Samostalne vježbe na računalu	0,5	3-8	Samostalni zadaci na vježbama	0-5 bodova po zadatku	15
Kontinuirana provjera znanja	1	1-12	Jedna kontrolna zadaća (kolokviji)	0-25 bodova za zadaću, ovisno o stupnju točnosti i potpunosti	25
Samostalni zadaci	1,5	2-12.	Izrada praktičnog seminarskog rada	0-30 bodova za seminarski rad, ovisno o pravilnosti realiziranog postupka ali i samog postupka izrade	30
Završni ispit	1	1-12	Pisani test	0-30 bodova, ovisno o stupnju točnosti i potpunosti	30
UKUPNO	5				100

Na završnom ispitu studenti moraju skupiti barem 50% bodova za prolaz.
Nepoštivanje postavljenih rokova smatra se neizvršavanjem obveza.

RASPORED NASTAVE U AKADEMSKOJ GODINI 2013./2014.

Nastava će se odvijati u ljetnom semestru prema sljedećem rasporedu:

Predavanja	utorkom	10-12	028
Vježbe	srijedom	12-14	366

Izvođenje nastave prikazano je u sljedećoj tablici:

KALENDAR PREDAVANJA I VJEŽBI – ljetni (IV) semestar ak. godine 2013/2014.

Datum	Tema	Nastava	Izvođač
4.3.2014.	Uvod , opće informacije, pregled pojmova	P	Dr.sc. Sanda Martinčić-Ipšić
5.3.2014.	Jezični resursi, korpusi, rječnici, leksikoni.	P	Dr.sc. Sanda Martinčić-Ipšić
11.3.2014.	word-net, korpusi, alati za korpuse	V	Luka Vretenar
12.3.2014.	Python-regularni izrazi	V	Luka Vretenar
18.3.2014.	Morfološki analizatori. Obilježivači vrsta riječi. Lematizatori	P	Dr.sc. Sanda Martinčić-Ipšić
19.3.2014.	Python, NLTK 1	V	Luka Vretenar
25.3.2014.	NLTK 2	V	Luka Vretenar
26.3.2014.	NLTK 3	V	Luka Vretenar
1.4.2014.	Statistički jezični modeli. N-gramski modeli. Perpleksnost.	P	Dr.sc. Sanda Martinčić-Ipšić
2.4.2014.	NLTK 4	V	Luka Vretenar
8.4.2014.	Parsiranje. Prepoznavanje značenja.	P	Dr.sc. Sanda Martinčić-Ipšić
9.4.2014.	POS Tagger	V	Luka Vretenar
15.4.2014.	Prepoznavanje pragmatike. Diskurs. Dijalog.	P	Dr.sc. Sanda Martinčić-Ipšić
16.4.2014.	Stemmer/Lematizator	V	Luka Vretenar
22.4.2014.	Pretraživanje dokumenata. Strukture indeksiranja dokumenata.	P	Dr.sc. Sanda Martinčić-Ipšić
23.4.2014.	Statistički modeli jezika	V	Luka Vretenar
29.4.2014.	Kolokvij	P	Dr.sc. Sanda Martinčić-Ipšić
30.4.2014.	Orange- text mining 1	V	Luka Vretenar
6.5.2014.	Vektorski prostor. Latentno indeksiranje. Klasifikacija dokumenata	P	Dr.sc. Sanda Martinčić-Ipšić
7.5.2014.	Orange- text mining 2	V	Luka Vretenar
13.5.2014.	Page Rank HITS	P	Dr.sc. Sanda Martinčić-Ipšić
14.5.2014.	Orange- text mining 3	V	Luka Vretenar
20.5.2014.	Otkrivanje stava i mišljenja iz teksta	P	Dr.sc. Sanda Martinčić-Ipšić
21.5.2014.	Klasifikacija tekstova	V	Luka Vretenar
27.5.2014.	Dubinska analiza teksta. Sustavi preporuka.	P	Dr.sc. Sanda Martinčić-Ipšić
28.5.2014.	Polarnost tekstova.	V	Luka Vretenar
3.6.2014.	Identifikacija jezika. Strojno prevođenje i strojno potpomognuto prevođenje.	P	Dr.sc. Sanda Martinčić-Ipšić
4.6.2014.	Seminari	V	Luka Vretenar
10.6.2014.	Seminari	P	Dr.sc. Sanda Martinčić-Ipšić
11.6.2014.	Seminari	V	Luka Vretenar

P – predavanja V – vježbe

Ispitni rokovi:

26.06.2014. 10:00 - četvrtak
 03.07.2014. 10:00 - četvrtak
 04.09.2014. 10:00 - četvrtak
 11.09.2014. 10:00 - četvrtak