

Statistika za jezikoslovno istraživanje

Proljeće 2010.

4. godina, izborni kolegij za studente svih odjela

Odjel za lingvistiku
Sveučilište u Zadru

Nositelj kolegija: doc. dr. Damir Bavačar

Predavanja: 2 sat/tjedno

Seminar: 1 sat/tjedno

ECTS bodovi: 8

Mogućnost izvođenja na jezicima: hrvatski, engleski, njemački, poljski

Vrijeme i dvorana: Informatički kabinet, vrijeme četvrtak 10-12 (može se promijeniti termin!)

Sati konzultacije: utorak i četvrtak od 12 sati

Pregled

U okviru ovoga kolegija daje se uvod u osnovne metode statističke analize koje se koriste u obradbi teksta i jezika, i to:

- univarijatna analiza (deskriptivna statistika)
- bivarijatna analiza (korelacije)
- multivarijatna analiza (npr. regresijska analiza, klasteraska analiza, "principle component analysis").

Sve će se metode analize primijeniti na obradbu govorenih i tekstovnih jezika i njihovih podataka uporabom računalskih aplikacija poput jezika R, tabličnih kalkulatora kao što je MS Excel ili OpenOffice Spreadsheet.

Studenti će usvojiti temeljne postupke statističke analize i znati izabrati prikladnu metodu analize za određena pitanja i svrhu, kao i vještinu upotrebe uobičajenih računalskih pomagala za obavljanje vlastitih analiza jezika i njihovih podataka.

Literatura i materijali (izbor)

- Abney, Stephen (1996) Statistical methods and linguistics. U: Judith Klavans and Philip Resnik: The Balancing Act. MIT Press Cambridge, MA.
- Goldsmith, John (2001) Probability for linguists. Mscr. University of Chicago.
- Paolillo, John C. (2002) Analyzing Linguistic Variation: Statistical Models and Methods. CSLI Publications. ISBN: 1-57586-276-X
- Manning, Christopher D. and Hinrich Schütze (1999) Foundations of statistical natural language processing. MIT Press. ISBN: 0262133601
- Jurafsky, Daniel and James H. Martin (2000) Speech and language processing: An introduction to natural language processing, computational linguistics and speech recognition. Prentice Hall, Englewood Cliffs, NJ. ISBN: 0-13-095069-6
- Dalgaard, Peter (2002) Introductory Statistics with R. Springer. ISBN: 0-387-95475-9

Obveze

- čitanje literature
- 2 ispita po 5-10 minuta
- aktivno sudjelovanje u nastavi i raspravama
- projekt
- izlaganje projekta
- seminarski rad na osnovi projekta

Dodatno

- Registrirajte si adresu e-pošte. Pošaljite mi dopis e-poštom s oznakom/temom "Statistika za jezikoslovno istraživanje Lista" i vašim imenom u sadržaju dopisa, tako da vas mogu dodati listi za obavijesti e-poštom. Mogu vam slati obavijesti, informacije, poveznice, zadatke za doma e radove i i tanje u elektronskom obliku.
- Najjednostavnije je kontaktirati me e-poštom.
- Doma i radovi ne priznaju se za bodovanje ako se predaju iza roka (osim ako se može dokazati opravdano izostajanje).
- Doma e radove možete predati elektronski putem e-pošte ako imate mogućnost pretvoriti ih u PDF datoteke.
- Doma i radovi se moraju predati u roku i ako niste slušali predavanje.
- Mogućnost za ponavljanje malih ispita ne postoji (to nisu službeni ispiti), samo iznimno u opravdanim slučajevima. Ako ne možete doći na takav ispit, morate me unaprijed obavijestiti.
- Možete preskočiti jedan ispit ili jedan doma i rad bez posljedice.
- Materijali će biti slani e-poštom. Isto tako će biti pristupni na mrežnim stranicama, za preslikavanje kod kopirnice kod referade, te u knjižnici.
- Provjerite redovno vašu e-poštu.

Ocijene

- Konačna ocijena se sastoji od slijedećih elemenata:
- Seminarski rad 35%
- Ispiti 20%
- Sudjelovanje u nastavi 25%
- Izlaganje projekta 20%

Mjera za ocijene

- 5: 90-100%
- 4: 80-89%
- 3: 70-79%
- 2: 65-69%
- 1: 0-64%

Neprihvatljivo akademsko ponašanje

Akademski integritet je poštenje i odgovornost u nastavnom i znanstvenom radu.

Neprihvatljivo akademsko ponašanje predstavljaju aktivnosti i radnje koje narušavaju akademski integritet institucije, a primjeri toga su:

- Korištenje nedozvoljenih materijala kao pomagala na ispitima ili u izradi domaćeg rada.
- Angažiranje druge osobe za polaganje ispita u svoje ime.
- Predaja značajnih dijelova istog akademskog rada za dobivanje bodova više od jedanput, a bez dozvole nastavnika ili programa za koji se rad predaje.
- Falsificiranje ili izmišljanje podataka i informacija u akademskim radovima, uključujući i, ali ne ograničavajući se na, bilješke i izvješća, laboratorijske rezultate i citiranje.
- Plagiranje, odnosno posvajanje radova drugih autora, uključujući i studenata. Svim idejama ili podacima uzetim iz drugih izvora bilo za usmenu ili pismenu uporabu moraju biti jasno naznačeni stvarni autori, ukoliko se ne radi o općem znanju.
- Od studenata i nastavnika se očekuje da se drže općeprihvaćenih standarda akademskog ponašanja. Neprihvatljivo akademsko ponašanje će biti kažnjeno na prikladan način. Kazna može varirati u rasponu od umanjenja ocjene za određeni rad pa sve do negativne ocjene za cijeli kolegij. U nekim slučajevima nastavnik može zatražiti od studenta/ice dodatne radove kako bi zadovoljio/la sve obveze iz kolegija.

- Svi podatci u ovome izvedbenom planu, uključujući i obveze iz kolegija te tjedni plan rada, podložni su promjenama i ne bi trebali biti nadomjestak za pohranje nastave niti za bilo koje informacije dobivene od predmetnog nastavnika.