

Указания за курсова работа по Компютърна лингвистика, триместър А, 2011 г. (спец. Информатика – редовно, IV курс)

Лектор: проф. дмн Г. Тотков

Упражнения: гл. ас. П. Иванова

Всеки студент трябва да разработи курсова работа като избере една от следните 2 възможности:

1. Да напише **текст-реферат** по избрана тема. Рефератът трябва да е около 10 страници добре форматиран и ясно изложен текст, шрифт Times New Roman 12pt, с не повече от 2 см бели полета на страница. Задължително трябва да е изброена използваната литература (поне 5 източника с автор, заглавие, страници, линк и дата на достъп) и да има цитати към нея. Идеи за темите, по които можете да пишете, ще намерите в посочената по-долу литература, както и ще чуете на лекциите и упражненията по тази дисциплина.
2. Да напише **програма** на език за програмиране по избор, която реализира някаква задача, свързана с обработка на естествен език. Необходимо е да представите и кратко описание (около 3 стр.) на задачата, която се решава, какви данни (ресурси) и алгоритми се използват, а също да се посочат и използваните източници.

Можете да работите самостоятелно или по двойки (тройки) като при работа в група обемът и броят на източниците нарастват 2 (3) пъти и трябва да е ясно всеки от авторите какво е правил.

Крайни срокове:

- избор на тема за курсова работа: до **20.10.2011** г. (задължително одобряване от гл.ас. Иванова). Преди да изберете трябва да направите проучване на дадените материали и намерите допълнителни източници, които ще използвате в работата по темата.
- представяне на готовия текст/програма: Предаване в часовете за упражнения (сряда и четвъртък) или изпращане по ел. поща най-късно до **1.12.2011** г. на .doc файл с текста (не се приема .docx, .pdf и др.), или архив (.rar, .zip) на програмата и описанието ѝ, на адрес pavlina@uni-plovdiv.bg. Полето **subject** на е-мейлите да съдържа задължително **CL2011 име и факултетен номер на студента**.
- защита на курсовата работа: **7 и 8.12.2011** г. (последната седмица от триместъра) на упражненията за съответната група. По желание (и по възможност) може да направите презентация за 10-15 мин като го заявите 1 седмица предварително.

Изпит:

Оценка се получава за разработената и **защитена** курсова работа.

Ако някой не е разработил такава се явява на писмен изпит върху въпроси по темите, разгледани на лекциите (има примерен списък с въпроси).

Оценка:

Курсовата работа ще бъде оценявана според приноса на студента (оригиналност, показано разбиране на написаното), свързаност и яснота на изложението, как са използвани източниците и техния брой.

Литература:

1. Ronald Cole, editor. *Survey of the State of the Art in Human Language Technology*. Studies in Natural Language Processing. Cambridge University Press, 1997. <http://www.cslu.ogi.edu/HLTsurvey/>
2. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, *Introduction to Information Retrieval*, Cambridge University Press. 2008. <http://nlp.stanford.edu/IR-book/>
3. Daniel Jurafsky and James H. Martin. *Speech and Language Processing*. Prentice Hall, New Jersey, second edition, 2008. <http://www.cs.colorado.edu/~martin/slp.html> (няма достъп до книгата, освен до глава 1)

Софтуер с отворен код за решаване на задачи, свързани с обработката на естествен език:

NLTK (Python) <http://www.nltk.org>

GATE (Java) <http://gate.ac.uk/>