

COURSE DESCRIPTION

Department and Course Number: CSCI 517

Course Title: Natural Language Processing

Current Catalog Description: Computer processing of natural language text at orphological, lexical, syntactic, and semantic levels; algorithms and procedures for sentence parsing and analysis; applications of natural language processing techniques.

Total Credits: 3 hours

Coordinator: Steven Schoenly, Associate Professor of Computer and Information Science.

Textbook: Daniel Jurafsky and James H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, 2000; ISBN 0-13-095069-6.

Other required materials: none.

References: <http://sbs.cs.olemiss.edu/>

Terry Winograd, *Language as a Cognitive Process: Syntax*. Addison-Wesley, 1983; ISBN 0-201-08571-2.

Course Goals: The purpose of this course is to give students opportunities for study of topics related to advanced string processing, natural language processing, natural language software interfaces, computational linguistics, and other related issues.

Prerequisites by Topic: There are no specific prerequisite courses. It is assumed that all students enrolled in this class will have advanced expertise in computer science or a discipline closely related to natural language processing.

Major Topics Covered in the Course:

1. string processing techniques (e.g. in Java)
2. other string processing languages (e.g. Icon)
3. philosophy of language, linguistics, computational linguistics
4. morphological processing (e.g. stemming algorithms)
5. lexical processing (dictionary lookup, lexicon design)
6. pattern matching systems and concepts
7. transition network representations of grammar
8. context-free grammars and chart parsers
9. augmented transition networks
10. natural language generation
11. machine translation
12. NLP and information retrieval

Laboratory projects: Course requirements typically include midterm and final semester projects, concerning topics chosen by the individual student.

Estimate of ABET/CAC Category Content:

	CORE	ADVANCED		CORE	ADVANCED
Data Structures	_____	_____ 1 _____	Computer Organization and Architecture	_____	_____
Algorithms	_____	_____ 1 _____	Concepts of Programming Languages	_____	_____ 1 _____
Software Design	_____	_____		_____	_____

Oral and Written Communications:

Every student is required to submit at least 2 written reports (not including exams, tests, quizzes, or commented programs) of typically 5 pages and to make 2 oral presentations of typically 30 minutes duration. This includes only material that is graded for grammar, spelling, style, and so forth, as well as for technical content, completeness, and accuracy.

Social and Ethical Issues:

Social and ethical issues are indirectly relevant in this course, inasmuch as there are such concerns related to voice recognition and security of NLP systems. Social and ethical issues are not separate units or topics.

Theoretical Content (Foundations):

Approximately 10% of the course is devoted to theoretical material concerning philosophy of language and linguistics.

Problem Analysis:

Midterm and semester projects are typically chosen by the individual student, so analysis experiences will vary from student to student.

Solution Design:

Midterm and semester projects are typically chosen by the individual student, so design experiences will vary from student to student.