

Tratamiento informático de la lengua española

Carlos Subirats Rüggeberg

Objetivos

El objetivo central de este curso es ofrecer una introducción a la construcción de gramáticas contextuales en forma de expresiones regulares o autómatas; dichas gramáticas, junto con léxicos electrónicos, serán utilizadas por un sistema de aplicaciones, que permiten reconocer automáticamente construcciones sintácticas.

Las clases se desarrollarán en el Aula de Informática de la Facultad de Letras y todos los estudiantes dispondrán durante el curso de una cuenta individual en los ordenadores de los Servicios Informáticos de la UAB, destinados a la docencia; en dichas cuentas, los estudiantes del curso deberán realizar ejercicios y prácticas, que determinarán

Temario

1. El sistema operativo OpenVMS-AXP versión 7.2. Comandos básicos del sistema operativo y del editor EDIT.
2. Autómatas finitos deterministas y no deterministas. Equivalencia entre autómatas y expresiones regulares: teorema de Kleene. Creación de gramáticas de restricciones contextuales con expresiones regulares o autómatas. Utilización de una base de conocimiento léxico en soporte electrónico: información léxica, lemas, códigos categoriales y morfológicos.
3. Conversión automática de expresiones regulares en autómatas deterministas mínimos. Reconocimiento automático de construcciones sintácticas con autómatas y léxicos electrónicos. Creación automática de concordancias.
4. Reconocimiento automático del grupo nominal en español. Gramáticas de restricciones contextuales y propiedades transformacionales de los nombres predicativos: formalización de características léxicamente condicionadas. Los grupos nominales coordinados.
5. Gramática contextual de las formas verbales simples y compuestas, activas y pasivas con ser y estar. Gramática contextual de las combinaciones de pronombres clíticos en español. Gramática contextual de las formas verbales simples y compuestas con pronombres clíticos.

Evaluación

La asistencia a las clases teóricas y a las prácticas es imprescindible para aprobar el curso. La nota final será el resultado de la media de las calificaciones obtenidas en los ejercicios que se realizarán con ordenador y que se presentarán a lo largo del curso, en fechas que se fijarán previamente.

Bibliografía

Alarcos Llorach, E. 1994. Gramática de la lengua española. Madrid: Espasa Calpe.

Borges, Quim, Joan Serra & Josep M. Arqués. 1996. Teoria d'autòmats. Bellaterra: Universidad Autónoma de Barcelona.

Casas, Rafel & Lluís Màrquez. 1997. Llenguatges, gramàtiques i autòmats. Barcelona: Edicions UPC.

Fernández Ramírez, S. 1986. Gramática española. El verbo y la oración, ed. de I. Bosque. Madrid: Arco/Libros.

Harris, Zellig et al. 1989. The Form of Information in Science. Analysis of an immunology sublanguage. Dordrecht: Kluwer Academic Press.

John E. Hopcroft y Joffrey D. Ullman. 1979. Introduction to Automata Theory, Languages, and Computation. Reading, MA: Addison-Wesley.

Johnson, Stephen B. 1987. An Analyzer for the Information Content of Sentences. Tesis doctoral, Computer Science Dept., New York University, New York.

Kelley, Dean. 1995. Teoría de autómatas y lenguajes formales. Madrid: Prentice Hall.

Marcos Marín, F.A. 1994. Informática y humanidades. Madrid: Gredos.

Martín, Aurora. 1994. Una propuesta de codificación morfosintáctica para corpus de referencia en lengua española. Tesis doctoral, Universidad Autónoma de Madrid.

Partee, Barbara H, Alice ter Meulen y Robert E. Wall. 1993. Mathematical Methods in Linguistics. Dordrecht: Kluwer Academic Publishers.

Roche, Emmanuel y Yves Shabes. 1995. Deterministic part-of-speech tagging with finite-state transducers. Computational Linguistics 21.2:227-253.

Ryckman, Thomas A. 1986. Grammar and Information. An investigation in linguistic metatheory. Tesis doctoral, Columbia University, New York.

Silberztein, Max. 1993. Dictionnaires électroniques et analyse automatique de textes. Paris: Masson.

