

Tratamiento informático de la lengua española (2000-2001)

http://seneca.uab.es/ebobes/tratamiento_informatico.html

Eulàlia de Bobes Soler

Objetivos

El objetivo de este curso es presentar las distintas herramientas informáticas de ayuda de que dispone el lingüista, y ofrecer una introducción al desarrollo de herramientas para el tratamiento automático del español.

Metodología

Se combinarán clases teóricas con ejercicios y prácticas. Las clases se desarrollarán en el Aula de Informática de la Facultad de Letras. Todos los estudiantes dispondrán de una dirección de correo electrónico (CESAR), que se utilizará para la docencia y para la entrega de las prácticas.

No es necesario tener conocimientos previos de informática, aunque es recomendable estar familiarizado con el uso de ordenadores personales.

Evaluación

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.

Temario

Herramientas informáticas para el lingüista

- Internet. Mail: recibir y enviar mensajes; transmisión de documentos adjuntos. Listas de distribución: qué son; suscribirse/desuscribirse, otros comandos; listas de distribución dedicadas a la lengua española (Infoling, Hispania, Traducción...). Web: navegación por Internet: direcciones interesantes sobre la lengua española (Centro Virtual Cervantes, Real Academia Española (CREA), Biblioteca Virtual Miguel de Cervantes...); recuperación de documentos de la red. Consulta remota de catálogos de bibliotecas y otras herramientas documentales.
- Correctores y traductores automáticos. Diccionarios en soporte electrónico. Programas para la traducción asistida por ordenador. Uso de escáneres y OCR. Programas de concordancias.

Desrrollo de herramientas para el tratamiento automático de la lengua española

- Etiquetación de textos. Diccionarios electrónicos. Detección de unidades léxicas simples y compuestas. Expresiones ambiguas. Uso de transductores para la resolución automática de ambigüedades y para la detección de predicados compuestos.
- Elaboración de gramáticas electrónicas.
- Detección de predicados compuestos mediante transductores.
- Propiedades de los predicados: propiedades generales y propiedades léxicamente condicionadas.
- Formalización de las propiedades de dependencia. Clases de dependencia.
- Formalización de las propiedades de selección y transformacionales de los predicados.

- Detección de oraciones derivadas y recuperación de oraciones de la base mediante transductores.

Bibliografía

Durante el curso se facilitará bibliografía y direcciones web específicas para cada uno de los temas.

Blecua, José Manuel [et al.]. 1999. Filología e informática nuevas tecnologías en los estudios filológicos. Bellaterra : Seminario de Filología e Informática, Departamento de Filología Española. Universidad Autónoma de Barcelona.

Bobes, Eulàlia de. 2000. Clases de dependencia de las locuciones verbales del español. Trabajo de investigación. Laboratorio de Lingüística Informática, Universidad Autónoma de Barcelona.

Bobes, Eulàlia de. 2000. Gramática electrónica de las locuciones verbales. Documento electrónico, Laboratorio de Lingüística Informática, Universidad Autónoma de Barcelona.

EuroWordnet Viewer. Periscope 1.3.1. Wordnet Samples: Español, Inglés, Italiano y holandés. 1998. CD-ROM; Lernout & Hauspie.

Fernández Prat, H. 1995. Los adjetivos predicativos en español: un estudio transformacional en el léxico. Tesis doctoral: Universidad Autónoma de Barcelona.

Garrido Íñigo, Paloma. 2000. Gramática electrónica de las locuciones predicativas comparativas. Documento electrónico, Laboratorio de Lingüística Informática, Universidad Autónoma de Barcelona.

Gómez, J. y Lorenzo, A.M. [eds.].1996. Lingüística e informática. Santiago de Compostela, Tórculo.

Harris, Zellig S. 1991. Language and Information. A Mathematical Approach. Oxford: Clarendon.

Marcos Marín, Francisco. 1994. Informática y humanidades / Francisco A. Marcos Marín. Madrid, Gredos cop. 1994

Miras, Neus. 1999. Generación de las formas flexivas del español utilizando autómatas. Proyecto de fin de carrera. Departamento de Informática y Laboratorio de Lingüística Informática, Universidad Autónoma de Barcelona.

Moya, Lidia. 2000. Gramática electrónica de los verbos predicativos Pnn. Documento electrónico, Laboratorio de Lingüística Informática, Universidad Autónoma de Barcelona. Observatorio Español de Industrias de la Lengua. 1996. Informe sobre recursos lingüísticos para el español. Alcalá de Henares, Instituto Cervantes.

Ortega, Marc. 2000. Transductores en el análisis léxico y sintáctico de un texto. Laboratorio de Lingüística Informática, Universidad Autónoma de Barcelona.

Palacios, Roser. 1996. Los operadores verbales de primer nivel. Un estudio transformacional en el léxico. Tesis de licenciatura, Laboratorio de Lingüística Informática, Universidad Autónoma de Barcelona.

Pérez Castro, Lois C. 1991. Informática y filología (ordenadores PC) guía del investigador novel II. Madrid, Ediciones Clásicas DL

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

Silberstein, Max. 1998. Finite State Transducers and the processing of natural languages. <http://www.ladl.jussieu.fr/INTEX/Tutorial.htm>

Subirats Rüggeberg, C. 1994. La flexión nominal en el Diccionario Electrónico de Formas Compuestas del Español. *Lingua Franca* 1: 63-69.

Subirats Rüggeberg, C. 1997. Bases de conocimiento lingüístico y análisis automático del discurso. *La Corónica* 26.1:159-166.

Subirats Rüggeberg, C. 2000. *Introducción a la sintaxis léxica*. Frankfurt: Vervuert.

Subirats Rüggeberg, C. y Ortega Gil, M. 1999. Extracción automática de información de grandes corpus. En J. De Kock y C. Gómez, eds. *La lingüística de corpus: aplicaciones*. Salamanca: Ediciones Universidad de Salamanca; también en

<http://seneca.uab.es/csubirats/Extraccion.doc>

Valls, Jordi. 1998. Representación de diccionarios electrónicos con autómatas. Proyecto de fin de carrera. Departamento de Informática y Laboratorio de Lingüística Informática, Universidad Autónoma de Barcelona.