

TÓPICOS ESPECIALES: PROCESAMIENTO LINGÜÍSTICO DE INFORMACIÓN ESPECIALIZADA.

Profesor: Melva J. Márquez Rojas.

Semestre: A2002.

Nombre del curso: Documentación, terminología y diseño de corpora lingüísticos

JUSTIFICACIÓN

El desarrollo y vertiginoso crecimiento de tecnologías vinculadas con la recuperación y explotación de información que vehicula el conocimiento especializado ha originado un creciente interés por el procesamiento del lenguaje natural en muchos aspectos tales como el diseño de sistemas de traducción automática, el desarrollo de sistemas de análisis morfosintácticos y semánticos y el diseño de corpus lingüísticos que conducen a diversas aplicaciones entre las cuales podemos mencionar algunas:

- Extracción de terminologías para desarrollo de glosarios, vocabularios y diccionarios.
- Explotación de recursos para desarrollo de sistemas ontológicos que describan la circulación del conocimiento a través de unidades lingüísticas
- Recuperación de información para fines didácticos
- Elaboración automática de resúmenes
- Extracción de descriptores para indización de documentos en bibliotecas y centros de referencia.

El curso que ofrecemos se centra en aspectos teóricos y prácticos del proceso de recuperación de información (documentación) de las unidades lingüísticas de significación especializada (terminología) en la conformación de corpus lingüísticos conformados por textos especializados.

Para llevar a cabo la tarea nada fácil de intentar procesar el lenguaje humano en cualquier forma, es necesario proveer al estudiante formado en ingeniería de ciertas herramientas teóricas y prácticas que le faciliten la comprensión de su actividad y lo vinculen con la ciencia de la lingüística. Por ello, en este curso se pretende iniciar al estudiante en el estudio de la documentación, la terminología como interdisciplina y la lingüística de corpus, teniendo como base únicamente el lenguaje especializado de textos escritos.

OBJETIVO GENERAL

Iniciar al estudiante en el estudio de la documentación, la terminología y la lingüística de corpus para el procesamiento de textos especializados.

OBJETIVOS ESPECÍFICOS

(1) Entender conceptos fundamentales en documentación, terminología y lingüística de corpus, haciendo hincapié en las corrientes teóricas del estudio de la Terminología como disciplina de la Lingüística Aplicada.

(2) Conocer herramientas computacionales y criticar los modelos lingüísticos que se emplean para el diseño y manejo de corpora lingüísticos.

(3) Integrar los conocimientos adquiridos y el manejo de las herramientas en un proyecto específico vinculado con necesidades específicas de los especialistas.

TEMARIO

1. La Terminología

- 1.1. Corrientes teóricas del estudio de la terminología: ¿ciencia?, ¿disciplina?, ¿método?
- 1.2. Las Unidades Léxicas Especializadas: de conocimiento, de información, de valor y de significación.
- 1.3. La comunicación especializada y los colectivos beneficiarios de la Terminología
- 1.4. Necesidad de sistematización en el reconocimiento de las unidades lingüísticas portadoras de información especializada para la comprensión de los textos.
- 1.5. El discurso especializado y la tipología textual.

2. La recuperación automatizada de información

- 2.1. La documentación: disciplina auxiliar de la terminología
- 2.2. Métodos de documentación
- 2.3. Modelos de recuperación de información a través de Internet: búsqueda, selección y almacenamiento de datos.

3. Los corpora lingüísticos en la tecnología: una nueva aproximación hacia el estudio del léxico.

- 3.1. Naturaleza, constitución y uso de corpora lingüísticos.
- 3.2. Análisis crítico de modelos existentes de corpora lingüísticos y su utilidad en la investigación lingüística.
- 3.3. Análisis y procesamiento de modelo de corpus: caso del corpus tècnic del Institut Universitari de Lingüística Aplicada (IULA) de la Universidad Pompeu Fabra.
- 3.4. Programas informáticos de tratamiento textual con base estocástica: TACT (Text Análisis Computing Tools) y WordSmith Tools ®.

ACTIVIDADES

Las actividades se centrarán en dos grandes bloques:

PRIMER BLOQUE (60% DEL CURSO):

Asistencia regular al curso de Documentación y Terminología dictado a los estudiantes de la opción de Traducción de la Licenciatura en Idiomas Modernos.

Lugar y hora: Lunes y miércoles de 3:45 a 5:30 p.m. en el aula C-17 Edificio C Facultad de Humanidades y Educación.

SEGUNDO BLOQUE (40% DEL CURSO):

Participación en el Seminario-Taller "El léxico especializado en las nuevas tecnologías de la información", que se dictará en la sede del INESCO (Instituto de Estudios del Conocimiento) de la Universidad Simón Bolívar los días 22, 23 y 24 de mayo de 2002 con una duración de 16 horas.

EVALUACIÓN

PRIMER BLOQUE:	SEGUNDO BLOQUE
Dos exámenes parciales Presentación oral de un tema del curso Proyecto de terminología	Diseño y explotación de corpus

BIBLIOGRAFÍA RECOMENDADA

Almela P., Ramón (1999) *Procedimientos de formación de palabras en español*. Barcelona: Ariel.

Biber, Douglas, Susan Conrad y Randi Reppen (1998). *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.

Cabré, M.T. (1999) *La terminología: Representación y comunicación*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Cabré, M. Teresa y Judit Feliu (Eds.) (2001) *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica*. Barcelona (España): Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra

Cabré, M. Teresa, Lluís Codina y Rosa Estopa (Eds.) (2000). *Terminología i documentació*. Barcelona (España): Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Gonzalo G., Consuelo y V. García Yebra (Eds.) (2000) *Documentación, terminología y traducción*. Madrid: Síntesis.

Lerat, P. (1997) *Las lenguas especializadas* (Trad. por Albert Ribas). Barcelona: Ariel.

McEnery, Tony y Andrew Wilson (2001) *Corpus Linguistics, an introduction*. 2nd ed.- Edinburgh: Edinburgh University Press. (Edinburgh Textbooks in Empirical Linguistics).

Wichmann, Anne, Steven Figelstone, Tony McEnery y Gerny Knowles (1997) *Teaching and Language Corpora*. Essex: Addison Wesley Longman Limited.

SOFTWARE

- TACT (Programa de libre distribución)

Se puede obtener en <http://www.chass.utoronto.ca/cch/tact.html>

- WordSmith Tools (Programa comercial)

Demo en <http://www.liv.ac.uk/~ms2928/index.htm>

- Buscadores y megabuscaores en línea.