

**UNIVERSIDAD DE LOS ANDES
FACULTAD DE INGENIERÍA
POSTGRADO EN MODELADO Y SIMULACIÓN DE SISTEMAS**

TÓPICOS ESPECIALES: PROCESAMIENTO Y MODELOS DE RECUPERACIÓN DE INFORMACIÓN (RI)

Profesor: Melva J. Márquez Rojas.

Semestre: B-2004.

Nombre del curso: Procesamiento y modelos de Recuperación de información (RI)

JUSTIFICACIÓN

El vertiginoso crecimiento de tecnologías vinculadas con la recuperación y explotación de información general y especializada requiere de la participación de diferentes disciplinas de estudio, entre las cuales se incluyen la estadística, la documentación, la matemática y la lingüística. La recuperación de información (Information Retrieval, en inglés) implica el proceso por el cual un usuario es informado de la existencia o no del documento que ha solicitado. No es un proceso que cambia el conocimiento del usuario con respecto a un tema específico; tampoco constituye un sistema de pregunta-respuesta, ni un proceso de recuperación de datos. **Recuperación de Información** trata con la representación, almacenamiento, organización, y acceso de información.

El curso que ofrecemos se centra en aspectos teóricos y prácticos del proceso de recuperación de información. Sin embargo, nos centramos en el estudio de los corpora documentales, los modelos existentes de recuperación de información y en el lenguaje (natural y documental).

OBJETIVO GENERAL

Iniciarse en el ámbito interdisciplinario de la recuperación de información (RI) a través de la comprensión de su concepto y de su procesamiento.

OBJETIVOS ESPECÍFICOS

1. Comprender conceptos fundamentales en recuperación de información, documentación, terminología y lingüística de corpus.
2. Conocer algunos modelos de recuperación de información y de lenguajes documentales.

3. Diseñar un sistema básico de recuperación de información que extraiga definiciones.

4. A partir de los resultados de (3), redactar un artículo de investigación y someterlo a arbitraje para una revista.

TEMARIO

1. El corpus lingüístico

1.1 Conceptos fundamentales

1.2 Tipología y modelos

1.3 Procesamiento automatizado de corpora lingüísticos

2. La documentación y la terminología

2.1 Conceptos fundamentales

2.2 El discurso especializado

2.3 Las unidades de valor especializado (UVE) en los textos especializados

2.4 La definición: definición y tipos de definición.

3. La recuperación de información

3.1 Conceptos fundamentales

3.2 Recuperación de información vs. Extracción de información

3.3 Modelos de recuperación de información

3.4 Técnicas de procesamiento

3.5 El control de la terminología en la recuperación de la información

3.6 Ontologías, Eurowordnet

3.7 Modelos de recuperación de información a través de Internet: búsqueda, selección y almacenamiento de datos.

ACTIVIDADES

Encuentros semanales de tres horas en los que se discutan las lecturas y se aclaren dudas.

EVALUACIÓN

1. Creación de un corpus documental con textos especializados en el ámbito decidido por común acuerdo entre el profesor y el estudiante.
2. Creación de un sistema básico de recuperación de definiciones.

BIBLIOGRAFÍA RECOMENDADA

- Abadal, E. El control de la terminología en la recuperación de la información. En Cabré, M. Teresa, Lluís Codina y Rosa Estopa (Eds.) (2000). *Terminología i documentació*. Barcelona (España): Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Arntz, R. y Picht H. (1995) *Introducción a la terminología*. [Traducción castellana: A. de Irazzábal et al.] Madrid: Fundación Germán Sánchez Ruipérez (Obra original publicada en 1989).
- 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.T. (200) La terminología y la documentación. En Gonzalo, C. y V. García Yebra, V. (Eds.) *Documentación, terminología y traducción*. Madrid: Síntesis.
- Cabré, M. T. (2003b) Theories of terminology. Their description, prescription and explanation. En *Terminology*, 9 (2) pp. 163-200.
- Cabré, M. T. y Feliu, J. (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
- De Beaugrande, R. y Dressler, W. (1997). *Introducción a la lingüística del texto*. (Versión española: S. Bonilla). Barcelona: Ariel (Obra original

publicada en 1972).

Krzysztof G., y Kulkarni, A. (2001) *FIRE: An Information Retrieval Interface for Intelligent Environments*. Verona. Proceedings of International Workshop on Information Presentation and Natural Multimodal Dialogue.

Clarke, I., Sandberg, O., Wiley, W., y Hong, T. *Freenet: A Distributed Anonymous Information Storage and Retrieval System* [Disponible en: doi.ieeecs.org/10.1109/PTP.2002.1046317]

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

Martí, Ma. A. (Coord.) (2003) *Las tecnologías del lenguaje*. Barcelona: Editorial UOC.

Martín, M. y Portolés, J. (1999) Los marcadores del discurso. En Bosque, I. Y Demonte, V. (Dir.) *Gramática descriptiva de la lengua española*. Madrid: Espasa (3 v; 63).

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