

CC72J OLAP Y MINERIA DE DATOS (10 UD)

Prof. Carlos Hurtado Larraín - Semestre 2002/1

Requisitos: CC42A/CC55A y Autorización del Departamento

Objetivos

Por "procesamiento analítico en línea" (OLAP) y "minería de datos" (Data Mining) nos referimos a categorías de técnicas de procesamiento de información que han irrumpido fuertemente en la academia e industria desde mediados de la década de los noventa. Ambas Tecnologías están enfocadas en la computación eficiente de resúmenes y patrones que describen conocimiento extraído de grandes volúmenes de información. El objetivo de este curso es introducir al alumno en los conceptos y técnicas fundamentales de estas tecnologías.

Metodología

Dos clases a la semana. Para la primera parte del curso (OLAP) el alumno deberá rendir una interrogación escrita (20% de la calificación final). Para la segunda parte del curso, el alumno deberá elegir y presentar un artículo técnico (30% de la calificación final), algunos de los tópicos del curso serán cubiertos en estas presentaciones. El resto del trabajo del curso se concentra en desarrollar un proyecto (50% de la calificación final) a especificar.

PROGRAMA

I. OLAP

- 1. OLAP vs. procesamiento transaccional en línea (OLTP).**
- 2. Modelación lógica de datos:** El modelo multidimensional, tablas de hechos (facts) y dimensiones.
- 3. Lenguajes de consultas:** Extensiones de SQL, cubos de datos, consultas envasadas vs. consultas ad-hoc, agregaciones a múltiples granularidades.
- 4. Organización física de datos:** Servidores OLAP relacionales (ROLAP) vs. servidores OLAP multidimensionales (MOLAP), selección de agregados a materializar, índices.

5. **Procesamiento de consultas:** Computación eficiente de cubos de datos, navegación en agregados.

II. Minería de Datos

1. **Búsqueda de reglas de asociación:** Algoritmos, búsqueda de asociaciones a múltiples niveles de abstracción y dimensiones.
2. **Clasificación:** Inducción de árboles de decisión, clasificación Bayesiana, clasificación basada en vecindarios, estimación de precisión en clasificación.
3. **Búsqueda de agrupaciones (Clustering):** Tipos de datos en búsqueda de agrupaciones, algoritmos basados en particiones, algoritmos jerárquicos, búsqueda de agrupaciones en espacios de numerosas dimensiones y en grandes volúmenes de datos.
4. **Técnicas de reducción de datos:** Análisis de relevancia de atributos, análisis de componentes, técnicas de muestreo.
5. **Minería de Datos en la Web:** Minería de uso y estructura de la Web.

Bibliografía

- Surajit Chaudhuri and Umesh Dayal, An Overview of Data Warehousing and OLAP Technology, ACM SIGMOD Record 26(1), March 1997.
<http://www.acm.org/sigmod/record/issues/9703/chaudhuri.ps>
- Jarke M., Lenzerini M., Vassiliou Y. and Vassiliadis P. (Eds.) Fundamentals of Data Warehouses¹. Springer-Verlag, 1999.
- J. Han and M. Kamber. Data Mining Concepts and Techniques. Morgan Kaufman Publishers, 2001
- Themistoklis Palpanas. Knowledge Discovery in Data Warehouses. Sigmod Record 29 (3), September 2000.
<http://www.cs.toronto.edu/themis/publications/sigrec00.ps>
- R. Kosala and H. Blockeel. Web Mining Research: A Survey. In SIGKDD Explorations, 2(1):1-15. July 2000.
<http://www.acm.org/sigs/sigkdd/explorations/issue2-1/kosala.pdf>.

¹ Sólo algunos capítulos de este libro están dedicados a OLAP

- Fuentes de artículos en la Web:

- Data Warehousing and OLAP, A Research-Oriented Bibliography.
<http://www.cs.toronto.edu/mendel/dwbib.html>.
- Knuggets, Data Mining, Web Mining & Knowledge Discovery.
<http://www.kdnuggets.com/>
- NECI Scientific Literature Digital Library.
<http://www.rearchindex.com/cs>.