

Compu f. 2001
40

UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES

1. DEPARTAMENTO: Computación
2. CUATRIMESTRE: Segundo de 2001
3. ASIGNATURA: Measurement, modelling and data analysis on the WWW
4. CARRERA: Licenciatura en Ciencias de la Computación
5. CARACTER DE LA MATERIA: optativa
6. NUMERO DE CODIGO DE CARRERA: 18
7. NUMERO DE CODIGO DE MATERIA: C
8. PUNTAJE: 1
9. PLAN DE ESTUDIOS AÑO: (1993)
10. DURACION DE LA MATERIA: semanal intensivo
11. HORAS DE CLASE SEMANAL:
 - a) TEORICAS/PRACTICAS: 15hs
 - b) LABORATORIO: ---
 - c) PRACTICAS: ---
 - d) SEMINARIOS: ---
12. CARGA HORARIA TOTAL: 15hs semanales.
13. ASIGNATURAS CORRELATIVAS: nociones algoritmos elementales de probabilidad y estad
14. FORMA DE EVALUACION: trabajo práctico final
15. PROGRAMA Y BIBLIOGRAFIA: la bibliografía no fue especificada por el docente

Fecha: 29 de agosto 2001


Prof. Responsable
Dr. Andrew Tomkins

Directora Dra. Patricia Borensztein


Dra. PATRICIA BORENSZTEJN
DEPTO. DE CIENCIAS EXACTAS Y NATURALES

N1. "Measurement, modelling and data analysis on the WWW"*Medidas, modelos y análisis de datos sobre la WWW***Dr. Andrew Tomkins***IBM Almaden Research Center, USA*

1. Caracterización de la WWW: se verá un conjunto de resultados concernientes al tamaño, crecimiento y estructura de la WWW que van desde pequeños grupos de páginas, a través de websites individuales, al grafo entero.
2. Modelos sobre la WWW: se describirán algunos modelos teóricos que explican o explotan el comportamiento del usuario, ya sea individualmente (browsing) o en conjunto (avisos, evolución del grafo)
3. Bases de datos para la WWW: se considerarán distintos enfoques de estructuras impuestas a la web para soportar formas más poderosas de consultas. Estos enfoques incluyen extracción de datos, creación de tablas relacionales (a menudo sin realmente materializar la tabla) y alguna forma de lenguaje estructurado de consulta (structured query language)
4. Análisis de datos: por último, se verán algunos problemas algorítmicos en el análisis de datos en la web. Se cubrirán algoritmos de crawling, búsqueda (incluyendo técnicas de álgebra lineal como Google y Clever), taxonomy creation, descubrimiento de comunidades, detección de duplicados, sitios espejo (mirror sites), descubrimiento de páginas relacionadas (related page discovery), y algoritmos de streaming.

Prerequisitos: Nociones de algoritmos, conocimientos elementales de probabilidad y estadística.

Idioma: Inglés.

Andrew Tomkins hizo su doctorado en Ciencias de la Computación de la Universidad de Carnegie Mellon en 1997. Anteriormente se ha desempeñado como investigador y consultor en el Laboratorio Matsushita de Tecnología de la información en Princeton, NJ, trabajando en reconocimiento de escritura manuscrita y OCR, entre otras aplicaciones. Es poseedor de varias patentes en el área de reconocimiento de patrones y pen-based?. Sus principales intereses incluyen a los algoritmos probabilísticos y de tipo on-line como así también algoritmos, análisis y modelado sobre la www y computadoras pen-based. Actualmente trabaja en IBM en el Almaden Research Center, en San José, CA donde es miembro del grupo de principios de metodologías.

Notas del Examen Final

Volver al Cronograma de la ECI 2001


Dra. PATRICIA BORENSZTEJN
DIRECTORA
DEPTO. DE COM. C.
F. C. E. y N. UBA