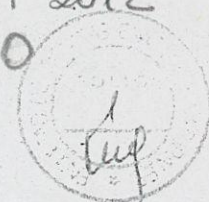


com 2012  
20



*Numero*

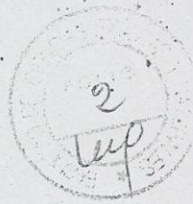


Universidad de Buenos Aires  
Facultad de Ciencias Exactas y Naturales

Planilla a completar para presentación de Cursos de Posgrado

- 1.- DEPARTAMENTO de COMPUTACIÓN
- 2.- NOMBRE DEL CURSO: **Recuperación de Información y Web Mining**
- 3.- DOCENTES:
  - RESPONSABLE/S: Jose M. Castaño.....
  - COLABORADORES: .....
  - AUXILIARES: ..Viviana Cotik...
- 4.- CARRERA de DOCTORADO
- 5.- AÑO: 2012..... CUATRIMESTRE/S: SEGUNDO
- 6.- PUNTAJE PROPUESTO PARA CARRERA DE DOCTORADO: 2 puntos
- 7.- DURACIÓN (anual, cuatrimestral, bimestral u otra): Cuatrimestral
- 8.- CARGA HORARIA SEMANAL:
  - Teóricas: .....
  - Problemas: ...1.....
  - Laboratorio: .....
  - Seminarios: .....
  - Teórico – Práctico: 2.....
  - Salida a Campo: .....
- 9.- CARGA HORARIA TOTAL: 48.....
- 10.- FORMA DE EVALUACIÓN: Trabajo Practico Final
- 11.- PROGRAMA ANALÍTICO:
  - Introducción a Recuperación de Información. Indices invertidos y consultas booleanas. Optimización de las consultas. La naturaleza del texto no-estructurado y del semi-estructurado.





- Codificación del Texto: Segmentación en 'tokens', extracción de lemas, 'stop words', y frases. Optimización de índices para el procesamiento de las consultas. Proximidad de frases y de consultas. Índices posicionales.
- Recuperación tolerante. Corrección ortográfica, sinónimos. Consultas con símbolos comodines ('wild cards'), permutación de índices. Índices de n-gramas. Distancia de Edición, "Soundex", detección del lenguaje.
- Construcción de Índices. Estimación del tamaño de los 'postings'. Indexación de los n-gramas. Cuestiones prácticas.
- Compresión de Índices. Compresión del léxico y de las listas de 'postings'. Codificación de 'gaps', códigos gamma, Ley de Zipf, Bloqueo, compresión extrema.
- Búsqueda parametrizada o por dominios. Zonas de documentos. El modelo del Espacio Vectorial ('Vector Space Model'). Esquemas de asignación de pesos. El esquema 'tf-idf'. Asignación de un valor ('score') a los documentos.
- Valoración en el modelo del Espacio Vectorial. La medida del Coseno. Consideraciones de eficiencia. Técnicas del vecino próximo. Aproximaciones de dimensionalidad reducida, proyecciones al azar.
- Presentación de resultados. Resúmenes estáticos y dinámicos. Evaluación de resultados. Satisfacción del usuario. Medidas de Precisión, Recuperación y "F". Creación de colecciones. Medida kappa, concordancia de anotación. Relevancia, aproximación a la recuperación vectorial.
- Feedback de la relevancia. Pseudo-feedback. Expansión de las queries. Generación automática de un thesaurus. Recuperación basada en los sentidos de los términos.
- Agrupamiento ("Clustering") y sus Métodos.
- Clasificación de textos y Métodos.
- Recuperación de información de la Web. Protocolos. Robots de navegación de la "web" ("Arañas"). Ranking basado en conducta y en 'links'. Análisis de 'links'.
- Recuperación y Extracción de Información. Sistemas de Respuesta a preguntas. .

12.- BIBLIOGRAFÍA:

Texto principal:

Introduction to Information Retrieval,(IIR) by C. Manning, P. Raghavan, and H. Schutze

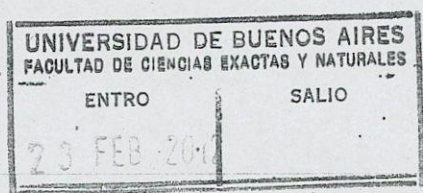
• Textos adicionales:

Managing Gigabytes. I. Witten, A. Moffat, and T. Bell.

Modern Information Retrieval. Baeza-Yates and B. Ribeiro-Neto.

Mining the Web, by S. Chakrabarti.

Foundations of Statistical Natural Language Processing, by C. Manning and H. Schutze.



*Isabel Méndez Díaz*  
ISABEL MÉNDEZ DÍAZ  
Departamento de Computación  
FCEYN - U.B.A.





Universidad de Buenos Aires  
Facultad de Ciencias Exactas y Naturales

Referencia Expte. N° 500.673/2012

Buenos Aires, 6 AGO 2012

**VISTO:**

la nota de la Dra. Isabel Mendez Diaz del Departamento de Computación, mediante la cual eleva la información del curso de posgrado **Recuperación de información y web mining**, que será dictada durante el segundo cuatrimestre de 2012 por José M. Castaño y como docente auxiliar Viviana Cotik,

**CONSIDERANDO:**

lo actuado por la Comisión de Doctorado de esta Facultad el 03/07/2012

lo actuado por la Comisión de Enseñanza, Programas, Planes de Estudio y Posgrado,

lo actuado por este Cuerpo en la sesión realizada en el día de la fecha,

en uso de las atribuciones que le confiere el Artículo 113° del Estatuto Universitario,

**EL CONSEJO DIRECTIVO DE LA FACULTAD DE  
CIENCIAS EXACTAS Y NATURALES  
RESUELVE:**

**Artículo 1°:** Autorizar el dictado del curso de posgrado **Recuperación de información y web mining** de 48 hs. de duración.

**Artículo 2°:** Aprobar el programa del curso de posgrado **Recuperación de información y web mining** obrante a fojas 1 y 2 del expediente de la referencia.

**Artículo 3°:** Aprobar un puntaje máximo de dos (2) puntos para la Carrera del Doctorado.

**Artículo 4°:** Aprobar un arancel de 20 módulos. Disponer que los montos recaudados serán utilizados conforme a lo dispuesto por Resolución CD N° 072/03.

**Artículo 5°:** Comuníquese a la Dirección del Departamento de Computación, a la Biblioteca de la FCEN y a la Subsecretaría de Postgrado (con fotocopia de programa fs 1 y 2). Cumplido Archívese.

Resolución CD N° 1720

SPimed/ 11/07/2012

Dr. JAVIER LÓPEZ DE CASENAVE  
SECRETARIO ACADEMICO

Dr. JORGE ALIAGA  
DECANO