

**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO****Licenciatura en Ciencias de la Computación****Facultad de Ciencias**

Programa de la asignatura

**Denominación de la asignatura:*****Recuperación y Búsqueda de Información en Textos***

| | | | |
|-------------------------------|---|---|----------------------------|
| Clave: | Semestre: 6-8 | Eje temático: Inteligencia Artificial | No. Créditos: 10 |
| Carácter: Optativa | Horas | | Horas por semana |
| Tipo: Teórico-Práctica | Teoría: | Práctica: | Total de Horas |
| | 3 | 4 | |
| Modalidad: Curso | Duración del programa: Semestral | | |

Asignatura con seriación indicativa antecedente: Autómatas y Lenguajes Formales; Modelado y Programación, Probabilidad I

Asignatura con seriación indicativa subsecuente: Ninguna

Objetivo general:

Conocer y aplicar las teorías y algoritmos de los sistemas de recuperación de información en textos. Dar a conocer la metodología del diseño y evaluación de sistemas de recuperación de información.

Índice temático

| Unidad | Temas | Horas | |
|-----------------------------|--|------------|-----------|
| | | Teóricas | Prácticas |
| I | Introducción | 3 | 4 |
| II | Recuperación Booleana | 3 | 4 |
| III | Vocabulario de términos y listas de destinos o declaraciones | 6 | 8 |
| IV | Diccionarios y la recuperación tolerante | 6 | 8 |
| V | Construcción de índices | 6 | 8 |
| VI | Compresión de índices | 3 | 4 |
| VII | Puntaje, peso de los términos y el modelo de espacio vectorial | 6 | 8 |
| VIII | Cálculo de puntajes en un sistema de RI | 6 | 8 |
| IX | Evaluación en RI | 3 | 4 |
| X | Retroalimentación de relevancia y expansión de la pregunta | 6 | 8 |
| Total de horas: | | 48 | 64 |
| Suma total de horas: | | 112 | |

| Contenido temático | |
|--|--|
| Unidad | Tema |
| I Introducción | |
| I.1 | ¿Qué es la recuperación de información? |
| I.2 | Introducción a los problemas en la RI. |
| I.3 | Explosión de la información. |
| II Recuperación Booleana | |
| II.1 | Ejemplo de recuperación de información. |
| II.2 | Índice invertido simple. |
| II.3 | Procesamiento de preguntas Booleanas. |
| II.4 | Comparación contra un modelo de clasificación. |
| III Vocabulario de términos y listas de destinos o declaraciones | |
| III.1 | Conversión de Documentos. |
| III.2 | Determinación del vocabulario de términos. |
| III.3 | Extensión de las listas de destinos o declaraciones. |
| III.4 | Frases como preguntas. |
| IV Diccionarios y la recuperación tolerante | |
| IV.1 | Estructuras de búsqueda. |
| IV.2 | Preguntas con comodines. |
| IV.3 | Corrección de ortografía. |
| IV.4 | Corrección fonética. |
| V Construcción de índices | |
| V.1 | Indexamiento basado en clasificación de términos – documentos por bloques. |
| V.2 | Un solo paso en memoria. |
| V.3 | Indexamiento distribuido. |
| V.4 | Indexamiento dinámico. |
| VI Compresión de índices | |
| VI.1 | Propiedades estadísticas de los términos. |
| VI.2 | Estimación del número de términos. |
| VI.3 | Compresión del diccionario. |
| VI.4 | Compresión de las listas de destinos o declaraciones. |
| VII Puntaje, peso de los términos y el modelo de espacio vectorial | |
| VII.1 | Índices paramétricos y de zona. |
| VII.2 | Frecuencia de los términos y ponderación. |
| VII.3 | Modelo de espacio vectorial para puntaje. |
| VII.4 | Variantes de las funciones tf e idf |
| VIII Cálculo de puntajes en un sistema de RI | |
| VIII.1 | Puntaje y clasificación eficiente. |
| VIII.2 | Componentes de un sistema de RI. |
| VIII.3 | Interacción del modelo de espacio vectorial y operadores en preguntas. |
| IX Evaluación en RI | |

| | |
|--|---|
| IX.1 | Evaluación de un sistema de RI. |
| IX.2 | Colecciones de prueba. |
| IX.3 | Evaluación de conjuntos recuperados sin clasificación. |
| IX.4 | Evaluación de resultados clasificados. |
| IX.5 | Valoración de la relevancia. |
| IX.6 | Snippets de resultado. |
| X Retroalimentación de relevancia y expansión de la pregunta | |
| X.1 | Retroalimentación de relevancia y pseudo-retroalimentación de relevancia. |
| X.2 | Métodos globales para reformulación de la pregunta. |

Bibliografía básica:

1. Baeza-Yates and Ribeiro, *Modern Information Retrieval*, Addison Wesley, 1999 (Z667 B34)
2. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, *Introduction to Information Retrieval*, Cambridge University Press. 2008.
3. Gerard Salton. *Automatic text processing: The transformation, analysis and retrieval of information by computer*. Reading, Massachusetts: Addison-wesley, 1989.

Bibliografía complementaria:

1. Ian H. Witten, Alistair Moffat and Timothy C. Bell. *Managing Gigabytes: Compressing and Indexing Documents and Images*, 2nd Edition (Morgan Kaufmann Series in Multimedia Information and Systems) Morgan Kaufmann Publishers, 1999.
2. Karen Sparck Jones and Peter Willett. *Readings in Information Retrieval* San Francisco: Morgan Kaufmann, 1997.
3. William B. Frakes and Ricardo Baeza-Yates. *Information retrieval: data structures y & algorithms* Englewood Cliffs, N.J.: Prentice Hall, 1992.

| | | | |
|--|------------|---|------------|
| Sugerencias didácticas: | | Métodos de evaluación: | |
| Exposición oral | (X) | Exámenes parciales | (X) |
| Exposición audiovisual | (X) | Examen final escrito | () |
| Ejercicios dentro de clase | (X) | Trabajos y tareas fuera del aula | (X) |
| Ejercicios fuera del aula | () | Prácticas de laboratorio | () |
| Seminarios | (X) | Exposición de seminarios por los alumnos | () |
| Lecturas obligatorias | (X) | Participación en clase | (X) |
| Trabajo de investigación | () | Asistencia | () |
| Prácticas de taller o laboratorio | (X) | Proyectos de programación | (X) |
| Prácticas de campo | () | Proyecto final | () |
| | | Seminario | () |
| Otras: _____ | | Otras: _____ | |

Perfil profesiográfico:

Egresado preferentemente de la Licenciatura en Ciencias de la Computación o matemático con especialidad en computación con amplia experiencia de programación. Es conveniente que posea un posgrado en la disciplina. Con experiencia docente.