

PROGRAMA ANALÍTICO

UNIDAD 1: Recuperación de información

Introducción. Documentos electrónicos. Algoritmos y estructuras básicas. Modelos de recuperación de información. La recuperación multilingüe. La recuperación de documentos multimedia o no textuales. La recuperación de información en Internet. La recuperación basada en la citación. Sistemas hipertextuales y recuperación documental. Sistemas de filtrado y recomendación.

UNIDAD 2: Diseño de Crawlers y Motores de Búsqueda

Diseño de un spider/crawler. Estrategias. Web scrapping. Minería de textos. Minería de datos. Recolección y representación estructural de sitios web/dominios. Recolección de repositorios. Protocolos de recuperación. Tratamiento de grafos. Diseño de un motor de búsqueda. Diseño de un indexador de documentos. Optimización en motores de búsqueda web (SEO/SEM).

UNIDAD 3: Evaluación de la recuperación

Diseño de experimentos y evaluación de la recuperación. Clasificación automática. Métricas para evaluar un clasificador. Introducción a métodos de aprendizaje automático. La búsqueda de respuestas. Métricas para evaluar la recuperación. Aplicación de librerías para evaluar un sistema de recuperación.

UNIDAD 4: Técnicas de recuperación de información avanzadas

Procesamiento del Lenguaje Natural (PLN). Técnicas de representación de documentos. TF-IDF, LSI/LSA, WDM, Bag of Words. Técnicas de extracción de información. Realimentación y expansión de consultas. Resumen automático de documentos.