



Ministerio de Cultura y Educación
Universidad Nacional de San Luis
Facultad de Ciencias Físico Matemáticas y Naturales
Departamento: Informatica
Area: Area I: Datos

(Programa del año 2024)

I - Oferta Académica

Materia	Carrera	Plan	Año	Período
() RECUPERACIÓN DE LA INFORMACIÓN	ING. EN COMPUT.	28/12	2024	2° cuatrimestre
		026/1		
() RECUPERACIÓN DE LA INFORMACIÓN	ING. INFORM.	2-	2024	2° cuatrimestre
		08/15		
() RECUPERACIÓN DE LA INFORMACIÓN	LIC.CS.COMP.	32/12	2024	2° cuatrimestre

II - Equipo Docente

Docente	Función	Cargo	Dedicación
HERRERA, NORMA EDITH	Prof. Responsable	P.Asoc Exc	40 Hs
LUDUEÑA, VERONICA DEL ROSARIO	Prof. Co-Responsable	P.Asoc Exc	40 Hs
REYES, NORA SUSANA	Prof. Co-Responsable	SEC U EX	2 Hs

III - Características del Curso

Credito Horario Semanal				
Teórico/Práctico	Teóricas	Prácticas de Aula	Práct. de lab/ camp/ Resid/ PIP, etc.	Total
5 Hs	Hs	Hs	Hs	5 Hs

Tipificación	Periodo
B - Teoria con prácticas de aula y laboratorio	2° Cuatrimestre

Duración			
Desde	Hasta	Cantidad de Semanas	Cantidad de Horas
05/08/2024	15/11/2024	15	75

IV - Fundamentación

A partir de la expansión y consolidación de Internet se ha puesto a disposición de las personas una importante cantidad de información de todo tipo. En la actualidad, la cantidad de información no estructurada que se genera y distribuye supera ampliamente las posibilidades de los usuarios de realizar un uso eficiente.

Para poder aprovechar al máximo este potencial de información, se necesitan modelos, algoritmos y técnicas que permitan su gestión eficaz y eficiente. En esta materia se abordan los fundamentos teóricos y prácticos de la temática de Recuperación de la Información junto con el estudio de las herramientas que permiten la construcción de un sistema de estas características.

V - Objetivos / Resultados de Aprendizaje

Al completar la asignatura, se espera que los estudiantes sean capaces de:

- Comprender los principales desafíos de recuperación de información y sus diferencias con otras disciplinas relacionadas.
- Conocer conceptos y modelos fundamentales de recuperación de información.
- Comprender los fundamentos teóricos sobre los modelos clásicos de recuperación de información.

- Comprender los métodos de indexación usados en este ámbito.
- Comprender los alcances de la disciplina, junto con criterios que les permitan determinar sus ámbitos de aplicación.
- Hacer un uso eficiente del software existente relacionado a la temática de estudio.

En esta materia se abordan los siguientes ejes transversales:

- Identificación, formulación y resolución de problemas de informática.
- Concepción, diseño y desarrollo de proyectos de informática.
- Gestión, planificación, ejecución y control de proyectos de informática.
- Utilización de técnicas y herramientas de aplicación en la informática.
- Generación de desarrollos tecnológicos y/o innovaciones tecnológicas.
- Fundamentos para el desempeño en equipos de trabajo.
- Fundamentos para la comunicación efectiva.
- Fundamentos para la acción ética y responsable.
- Fundamentos para evaluar y actuar en relación con el impacto social de su actividad en el contexto global y local.
- Fundamentos para el aprendizaje continuo.
- Fundamentos para la acción emprendedora.

VI - Contenidos

1. Introducción

Introducción al problema de recuperación de información. Diferencias con el concepto de recuperación de datos. Arquitectura de un sistema de recuperación de información.

2. Modelos

Modelos de recuperación de la información. Una taxonomía de modelos. Modelos clásicos: booleano, vectorial y probabilístico. Modelos alternativos.

3. Preprocesamiento e Indexación

Análisis y preprocesamiento de textos. Análisis estadístico de las propiedades del texto. Procesamiento de documentos y términos. Eliminación de stopwords y lematización. Indexación: listas invertidas y archivos signatures

4. Consultas.

Consultas basadas en palabras claves. Pattern matching. Expansión de una consulta y retroalimentación. Expansión basada en clustering y expansión basada en thesaurus. Evaluación de la recuperación: exhaustividad (recall) y precisión. Diagramas de Exhaustividad/Precisión.

5. Recuperación de Información en la Web

Características de la recuperación de información en la Web. Arquitectura de los motores de búsquedas. Algoritmos de ranking. Indexación y crawling.

VII - Plan de Trabajos Prácticos

Metodología de Enseñanza

Para cada unidad temática el alumno tendrá disponible el material correspondiente a los contenidos de la unidad: el trabajo práctico a desarrollar, las diapositivas de clase teórica y la recomendación de la bibliografía a leer.

Las actividades prácticas a desarrollar por el estudiante involucran el desarrollo de ejercicios en papel y, cuando la temática así lo requiera, desarrollo de ejercicios en máquina. De cada práctico se brindarán consultas con el objetivo de aclarar dudas que surjan sobre los ejercicios a desarrollar y de incentivar el trabajo con los compañeros.

De cada práctico se requiere que el alumno entregue resuelto algunos ejercicios (a definir por el docente) que serán corregidos y devueltos con el suficiente detalle en las correcciones como para que el alumno pueda medir su avance en la

apropiación de conocimientos. Además de la presentación escrita, se fomentará que los alumnos realicen la presentación oral de la resolución de ejercicios a fin de desarrollar las capacidades de comunicación del alumno y de promover el debate en la clase.

Trabajos Prácticos a Desarrollar

Trabajo Práctico 1: Introducción

Este práctico incluye ejercicios en papel tendientes a lograr que el alumno sea capaz de:

- Identificar la diferencia entre recuperación de datos y recuperación de información.
- Identificar las componentes de un sistema de recuperación de información.

Trabajo Práctico 2: Modelos

Este práctico incluye ejercicios en papel tendientes a lograr que el alumno sea capaz de:

- Entender los conceptos involucrados en los diferentes modelos de recuperación de información.
- Decidir las oportunidades de uso de cada modelo.

Trabajo Práctico 3: Preprocesamiento e Indexación.

Este práctico incluye ejercicios en papel y máquina tendientes a lograr que el alumno sea capaz de:

- Identificar las etapas involucradas en el preprocesamiento de un texto.
- Realizar el análisis estadístico de un texto y su impacto en la etapa de indexación
- Indexar un texto en el entorno de un sistema de recuperación de información.

Trabajo Práctico 4: Consultas

Este práctico incluye ejercicios en papel y máquina tendientes a lograr que el alumno sea capaz de:

- Realizar búsquedas sobre el texto indexado.
- Evaluar la eficiencia de las mismas y, en caso de ser necesario, proponer mejorar la eficiencia.

Trabajo Práctico 5: Recuperación de Información en la Web

Este práctico incluye ejercicios en papel y máquina tendientes a lograr que el alumno sea capaz de:

- Comprender las características de la recuperación de información web.
- Comprender y saber aplicar algoritmos de ranking.
- Entender la problemática de indexación en la web.

Ejes Transversales

A continuación se presenta la descripción del abordaje y evaluación de cada uno de los ejes introducidos previamente en los objetivos.

- Identificación, formulación y resolución de problemas de informática.

Se aborda desde el práctico 2 presentándole al alumno casos reales para el desarrollo de los modelos y la realización de consultas. La evaluación se realiza mediante la entrega de ejercicios.

- Concepción, diseño y desarrollo de proyectos de informática.

Se aborda desde el práctico 2 presentándole al alumno casos reales sobre los cuales se necesita realizar un sistema de recuperación de información. La evaluación se realiza mediante la entrega de ejercicios.

- Gestión, planificación, ejecución y control de proyectos de informática

Se aborda desde el práctico 2 presentándole al alumno casos reales sobre los cuales se necesita realizar un sistema de recuperación de información. La evaluación se realiza mediante la entrega de ejercicios.

- Utilización de técnicas y herramientas de aplicación en la informática

Se aborda desde el práctico 2 presentándole al alumno casos reales sobre los cuales se necesita realizar un sistema de recuperación de información. La evaluación se realiza mediante la entrega de ejercicios.

- Generación de desarrollos tecnológicos y/o innovaciones tecnológicas.

Se aborda desde el práctico 2 presentándole al alumno casos reales sobre los cuales se necesita realizar un sistema de recuperación de información. La evaluación se realiza mediante la entrega de ejercicios.

- Fundamentos para el desempeño en equipos de trabajo.

Se aborda desde el inicio integrando grupos de trabajo para el desarrollo de los prácticos de máquina. La evaluación se realiza mediante la entrega de ejercicios por grupos de trabajo.

- Fundamentos para la comunicación efectiva.

Se aborda desde el inicio mediante la entrega escrita y la presentación oral de ejercicios. La evaluación se realiza teniendo en cuenta la exposición oral y la redacción de las soluciones propuestas.

- Fundamentos para la acción ética y responsable.

Se aborda desde el inicio estableciendo cronogramas de entrega que deben ser cumplidos por el estudiante para regularizar y/o promocionar la materia. La evaluación se realiza teniendo en cuenta que las entregas, para ser aprobadas, tienen que ser realizadas en tiempo y forma por los alumnos.

- Fundamentos para evaluar y actuar en relación con el impacto social de su actividad en el contexto global y local

Se aborda desde el inicio fomentando una actitud crítica que les permita evaluar el impacto social del sistema a desarrollar.

- Fundamentos para el aprendizaje continuo.

Se aborda desde el inicio fomentando el uso de la bibliografía recomendada, apuntes provistos por la cátedra y promoviendo la búsqueda de material por parte del alumno. Se evalúa con la corrección informada de los prácticos.

- Fundamentos para la acción emprendedora.

Se aborda desde el inicio fomentando la búsqueda de casos reales donde puedan aplicar las herramientas y técnicas enseñadas. Se evalúa mediante la entrega y corrección de ejercicios prácticos.

VIII - Regimen de Aprobación

Régimen para estudiantes regulares

Para regularizar la actividad curricular los estudiantes deberán:

1. Asistir al 70% de las clases.
2. Entregar los ejercicios que se le soliciten en las fechas previstas.
3. Realizar y aprobar un trabajo práctico integrador con una nota mínima de 6. Este trabajo consistirá en el desarrollo de una solución completa para una realidad dada y deberá ser entregado de manera escrita y defendido de manera oral.

Régimen para estudiantes promocionales

Para promocionar la actividad curricular los estudiantes deberán:

1. Asistir al 80% de las clases.
2. Entregar los ejercicios que se le soliciten en las fechas previstas.
3. Realizar y aprobar un trabajo práctico integrador con una nota mínima de 7. Este trabajo consistirá en el desarrollo de una solución completa para una realidad dada y deberá ser entregado de manera escrita y defendido de manera oral

Examen final para alumnos regulares

El examen final para alumnos regulares será de carácter teórico práctico y podrá ser escrito y/o oral.

Examen final para alumnos libres

No se admiten alumnos libres

IX - Bibliografía Básica

[1] R. Baeza-Yates, B. Ribeiro-Neto. Modern Information Retrieval. 2nd Ed. Addison-Wesley, 2011.

[2] S. Büttcher, C.L.A. Clarke, G.V. Cormack. Information Retrieval: Implementing and Evaluating Search Engines. The MIT Press, 2016.

[3] M. McCandless, E. Hatcher and O. Gospodnetic. Lucene in Action (2d Edition) Manning 2011.

[4] Stefano Cerri r Alessandro Bozzon, Marco Brambilla , Emanuele Della Valle, Piero Fraternali, Silvia Quarteroni. Web Information Retrieval. Ed. Springer 2013.

X - Bibliografía Complementaria

XI - Resumen de Objetivos

Al completar la asignatura, se espera que los estudiantes sean capaces de:

- Comprender los principales desafíos de recuperación de información y sus diferencias con otras disciplinas relacionadas.
- Comprender los alcances de la disciplina, junto con criterios que les permitan determinar sus ámbitos de aplicación.
- Hacer un uso eficiente del software existente relacionado a la temática de estudio.

XII - Resumen del Programa

Introducción. Modelos. Preprocesamiento e indexación. Consultas. Recuperación de la información en la web.

XIII - Imprevistos

XIV - Otros