



Recuperación de Información

1. Características generales

Nombre:	Recuperación de Información
Sigla:	CI-2414
Créditos:	4
Horas:	4 horas de teoría
Requisitos:	CI-1330 Ingeniería de Software I
Correquisitos:	ninguno
Clasificación:	Curso propio, electivo
Ciclo de carrera:	II ciclo, 4to año
Docente:	Dra. Kryscia Daviana Ramírez Benavides
Datos de contacto:	Oficina 315 – Edificio Anexo ECCI / Casillero 58 – ECCI / Correo electrónico: kryscia.ramirez@ucr.ac.cr
Grupo:	01
Semestre y año:	I ciclo 2020
Aula:	CI303
Horario:	Martes y Viernes de 10:00 a.m. a 11:50 a.m.
Horario de Consulta de Oficina:	Martes 12md-3pm y Viernes 9am-10am
Horario de Atención Opcional:	Vía e-mail (24/7) y vídeo llamada (cita previa)
Asistente:	Por definir.

2. Virtualización – Metodología

La propuesta de virtualización construida es la siguiente:

1. Desarrollar el trabajo en equipo y el aprendizaje autorregulado.
2. Desarrollar nuevas capacidades y habilidades en los estudiantes, quienes construirán juntos su conocimiento y formarán nuevas actitudes. El estudiante será el principal actor en el proceso de enseñanza-aprendizaje.
3. Priorizar contenidos y objetivos específicos esenciales, dar flexibilidad.
4. Reformular las actividades de aprendizaje en relación con los objetivos, el cronograma y el porcentaje de la evaluación.
5. Buscar herramientas que ayuden en la virtualización del curso.
6. Buscar o generar los recursos necesarios para la virtualización.





7. Definir clases sincrónicas y asincrónicas, actividades de aprendizaje y evaluación.
8. Definir un acuerdo de colaboración y compromiso de las actividades que se llevarán a cabo en la virtualidad entre la docente y los estudiantes.
9. Definir en cada clase (sincrónica o asincrónica): 1. Presentación, 2. Objetivos, 3. Actividades a desarrollar, 4. Actividades evaluativas.
10. Dar instrucciones claras y precisas en cada actividad de la virtualidad.
11. Definir una declaración de responsabilidad en cada proceso de autoevaluación y coevaluación de los aprendizajes que se realizarán.

Las clases serán de la siguiente manera:

1. Se realizarán en el horario del curso.
2. Se dejarán recursos para leer y analizar con anterioridad.
3. Se dará una explicación resumida del tema (los estudiantes deberán haber leído los recursos del tema dados por la docente) y se aclararán las consultas respectivas. Esta actividad durará entre 30-40min y será sincrónica, al mismo tiempo se grabará para ponerla a disposición de los estudiantes (pensando en los que no pudieran asistir).
4. Se realizará trabajo en equipo en diferentes objetivos y contenidos del curso. En este momento de la clase virtual, los equipos realizarán una reunión y la docente estará disponible en caso de consultas. Se realizarán actividades de aprendizaje y evaluación.
5. Se realizará el cierre de la clase entre todos, realizando un resumen de los contenidos tratados.

En la virtualización de este curso se utilizarán las siguientes herramientas con su respectivo fin:

1. Zoom y Discord para dar las clases virtuales sincrónicas.
2. Mediación Virtual y Sitio Web propio (<https://www.kramirez.net/>) para el planeamiento de las clases y el almacenamiento de los recursos.
3. Canva, Genially y Power Point para crear el material visual y las presentaciones.
4. Facebook Live y YouTube para generar videos educativos (clases virtuales asincrónicas).

La propuesta de virtualización se realizó con base a lo aprendido en los diferentes cursos que realizó la Escuela de Ciencias de la Comunicación Colectiva de la UCR (en este link se pueden encontrar los videos de los cursos <https://www.youtube.com/user/ECCUCR/videos>):

- ECCC-01 Diseño didáctico de aulas virtuales
- ECCC-02 Videoconferencias: usos y posibilidades
- ECCC-03 Tips de diseño y elaboración de presentaciones con recursos en línea
- ECCC-04 Estrategias de mediación pedagógica en la virtualidad
- ECCC-05 Diseño de audios educativos
- ECCC-06 Creación básica de videos educativos





3. Objetivos

Los objetivos específicos resaltados en negrita serán la prioridad de la virtualización del curso, los demás objetivos se tratarán de ver de una forma reducida.

Objetivo general

Conocer la arquitectura general, la metodología básica y los principales modelos y para el desarrollo de sistemas para la recuperación de información. Incluyendo los sistemas modernos para la recuperación de información en el WWW.

Objetivos específicos

Durante este curso el estudiante adquirirá conocimientos y desarrollará habilidades para:

- 1. Comprender los conceptos generales del área de Recuperación de Información. (conocimiento)**
2. Conocer a fondo ejemplos específicos de sistemas, lenguajes, protocolos, y mecanismos relacionados con los sistemas de recuperación de información en el WWW. (conocimiento)
- 3. Crear las estructuras de datos necesarias para llevar a cabo el proceso de creación de un índice invertido para una colección específica de documentos. (comprensión)**
- 4. Conocer las características básicas de los modelos básicos de recuperación de información. (comprensión)**
- 5. Comparar los modelos básicos de recuperación de información entre sí. (comprensión)**
6. Conocer los diferentes mecanismos para procesamiento de texto y consultas. (comprensión)
- 7. Introducir al estudiante a la investigación en el área de Recuperación de Información mediante el desarrollo de una investigación y una aplicación supervisada en un tema específico. (aplicación)**
- 8. Desarrollar un sistema para recuperación de información sobre una colección específica de documentos. (aplicación)**

Objetivos Transversales

Competencias que se adquirirán sobre trabajo colaborativo:

- 1. Intercambiar información con el equipo de trabajo colaborativo. (comprensión)**
- 2. Identificar adecuadamente las tareas a realizar por el equipo, repartir equitativamente las tareas, establecer fechas de entrega e integrar las partes. (aplicación)**
- 3. Explicar al grupo la tarea realizada, y asegurarse de que todos los demás la han comprendido. (aplicación)**
- 4. Responsabilizarse del trabajo realizado. (aplicación)**





5. Identificar y abordar los conflictos de funcionamiento del equipo. (aplicación)
6. Identificar los aspectos que han ido bien y qué aspectos se pueden mejorar en el funcionamiento del equipo. (aplicación)
7. Aprender de forma autónoma. (aplicación)
8. Desarrollar la capacidad de realizar presentaciones orales del trabajo realizado utilizando los medios adecuados. (aplicación)
9. Autoevaluarse o evaluar a otros a partir de unos criterios dados. (comprensión)
10. Identificar los propios errores. (comprensión)
11. Buscar información relevante para una tarea específica. (aplicación)

4. Contenidos

Los ejes temáticos del curso y los objetivos a los que contribuyen se muestran en la tabla que sigue. Los contenidos resaltados en negrita serán la prioridad de la virtualización del curso, tomando como base los recursos tecnológicos con los que se cuentan; los demás se verán de forma reducida.

Objetivos	Eje temático	Desglose
1	Recuperación de Información	Introducción a la Recuperación de Información Sistema de Recuperación de Información
1, 2, 4, 5	Fundamentos de RI	Modelos clásicos (booleano, vectorial y probabilístico) Modelos alternativos (conjuntos difusos, LSI, redes de inferencia)
1, 2, 6	Evaluación de Resultados	Precisión, factor de recuperación y factor de pérdida Recall Precisión-R, precisión promedio Colección TREC
1, 2, 7	SRI Reales	Arquitectura de Google. SEO y Google. Otros SRI: Yahoo y Bing (Kumo), y comparación con Google





1, 2, 7, 8	Crawlers (<i>Spiders / Arañas</i>)	Funcionamiento de una <i>crawler</i>
1, 2, 7, 8	Procesamiento de Texto	Análisis léxico Eliminación de <i>stopwords</i> Reglas para extracción de términos
1, 2, 6, 7, 8	Operaciones sobre Texto	Algoritmos de similaridad y distancia <ul style="list-style-type: none"> - Distancia de Hamming - Distancia de Levenstein - Bigramas Stemming (Lematización) <ul style="list-style-type: none"> - Porter - Variedad de sucesores - Bigramas Clustering <ul style="list-style-type: none"> - K-Mens - HAC Interfases de visualización <ul style="list-style-type: none"> - Tree Map - Sammon Cluster View - Radial Interactive Visualization Comprensión <ul style="list-style-type: none"> - Huffman orientado a bytes
1, 2, 3, 6, 7, 8	Indexación y Búsqueda	Índices invertidos Búsquedas sobre índices comprimidos Búsqueda secuencial de texto sobre texto





		<ul style="list-style-type: none"> - Shift-Or - Boyer-Moore-Horspool
1, 2, 6, 7, 8	Operaciones sobre Consultas	<p>Retroalimentación por relevancia</p> <p>Expansión de consultas mediante análisis local de contexto</p> <p>Expansión de consultas mediante análisis global de contexto</p>
7, 8	Temas Avanzados	<p>Filtrado de documentos</p> <p>Sumarización</p> <p>Optimización en Buscadores (SEO: <i>Search Engine Optimization</i>)</p> <p>RI Multimedia</p>

5. Evaluación

Se quitó el 20% del desarrollo del juego de mesa y ese valor se pasó a la investigación que realizarán los estudiantes.

A continuación, se presenta la evaluación:

Asignación	Valor
Tareas y Quices	20%
Investigación	40%
Exposición del tema y la aplicación	20%
Material Visual – Vídeo Clip	15%
Evaluaciones de las exposiciones	5%
Proyecto	40%
Etapa I	10%
Etapa II	10%
Etapa III	10%
Etapa IV	10%
TOTAL	100%

6. Cronograma

Se cambiaron las fechas para adecuarlas a las fechas que se tienen disponibles para desarrollar el curso.

Clases





N°	Fecha Inicio	Fecha Fin	Duración (clases)	Tema
1	12/03	12/03	1	Presentación del curso
2	12/03	15/03	2	Recuperación de Información
3	13/04	17/04	2	Fundamentos de RI
4	20/04	24/04	2	Evaluación de Resultados
5	20/04	24/04	2	SRI Reales
6	27/04	30/04	2	Crawlers (Spiders / Arañas)
7	04/05	08/05	2	Procesamiento de Texto
8	11/05	05/06	8	Operaciones sobre Texto
9	08/06	12/06	2	Indexación y Búsqueda
10	15/06	19/06	2	Operaciones sobre Consultas
11	22/06	26/06	2	Temas avanzados
12	29/06	03/07	2	Presentación del proyecto

Tareas

- Tarea #1: Viernes 17 de abril, hora de clase.
- Tarea #2: Viernes 24 de abril, hora de clase.
- Tarea #3: Viernes 8 de mayo, hora de clase.
- Tarea #4: Viernes 15 de mayo, hora de clase.
- Tarea #5: Viernes 22 de mayo, hora de clase.
- Tarea #6: Viernes 12 de junio, hora de clase.

NOTA: Se cambiaron las fechas de algunas tareas.

Investigación

- Entrega de la Presentación: Una semana antes de la fecha de exposición.
- Exposición del Tema y la Aplicación: Del 14 de abril al 19 de junio.
- Entrega del Vídeo Clip: Día de la exposición.

NOTA: Todo sigue igual, los estudiantes en este tiempo de suspensión quedaron en trabajar en la investigación.

Proyecto





- Entrega Etapa I: Viernes 24 de abril.
- Entrega Etapa II: Viernes 15 de mayo.
- Entrega Etapa III: Viernes 5 de junio.
- Entrega Etapa IV: Viernes 3 de julio.
- Entrega Final: Martes 7 de julio, 9am – 11am.

NOTA: Cambiaron las fechas de algunos entregables, los estudiantes en este tiempo de suspensión quedaron en trabajar en la primera etapa del proyecto. La feria quedará suspendida por el momento, su realización dependerá de la situación

NOTA: Las fechas planificadas en este cronograma están sujetas a un transcurso normal del semestre. En caso de cualquier eventualidad, se realizará el cambio de fechas que se acuerden con la mayoría del grupo y la profesora.

