



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. Descripción

El curso de Recuperación de Información es un curso teórico-práctico dedicado a introducir los fundamentos de representación, almacenamiento, organización, y acceso a la información de manera automática, y desde una perspectiva de las Ciencias de la Computación. El curso se centra en la teoría, algoritmos y técnicas utilizadas para la creación de sistemas de Recuperación de Información.

En cumplimiento de la Ley 7600 de la igualdad de oportunidades se programa la posibilidad de atención de las necesidades educativas especiales de los matriculados. Los interesados favor avisar al profesor por escrito durante la primera semana del curso acerca de las adecuaciones que necesiten. Posteriormente (a más tardar la tercera semana del curso), para hacer efectivas las adecuaciones requeridas se deben presentar los documentos que justifiquen tales necesidades educativas especiales. El estudiante debe estar





dispuesto a trabajar en equipo junto con el docente y la comisión institucional para llevar su necesidad educativa particular.

3. Objetivos

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. Iniciarse en el uso del portafolio de evidencias para clasificar y ordenar el material desarrollado. (aplicación)
8. Aprender de forma autónoma. (aplicación)
9. Desarrollar la capacidad de realizar presentaciones orales del trabajo realizado utilizando los medios adecuados. (aplicación)
10. Autoevaluarse o evaluar a otros a partir de unos criterios dados. (comprensión)
11. Identificar los propios errores. (comprensión)
12. 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.

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 (Spiders / Arañas)	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	<p>Algoritmos de similaridad y distancia</p> <ul style="list-style-type: none"> - Distancia de Hamming - Distancia de Levenstein - Bigramas <p>Stemming (Lematización)</p> <ul style="list-style-type: none"> - Porter - Variedad de sucesores - Bigramas <p>Clustering</p> <ul style="list-style-type: none"> - K-Mens - HAC <p>Interfaces de visualización</p> <ul style="list-style-type: none"> - Tree Map - Sammon Cluster View - Radial Interactive Visualization <p>Comprensión</p> <ul style="list-style-type: none"> - Huffman orientado a bytes
1, 2, 3, 6, 7, 8	Indexación y Búsqueda	<p>Índices invertidos</p> <p>Búsquedas sobre índices comprimidos</p> <p>Búsqueda secuencial de texto sobre texto</p> <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. Metodología

El curso se extenderá a lo largo de un semestre de trabajo, con una equivalencia aproximada de 60 horas presenciales (32 lecciones – 16 semanas).

Este curso promoverá la participación y la colaboración de los estudiantes en el proceso de enseñanza-aprendizaje, y por esta razón requerirá del compromiso constante del estudiante con el curso. Por su parte, el docente proveerá espacios y actividades didácticas que faciliten un aprendizaje colaborativo. Tales actividades se realizarán tanto dentro como fuera del horario de clase y el profesor podrá asignarlas como trabajo individual o en equipos.





Las clases teóricas serán complementadas con la lectura semanal de artículos y documentos que serán comentados en clase y los cuales se relacionarán al tema, o bien a la lectura de material relativo a herramientas, lenguajes o técnicas actualmente usadas para implementar tales sistemas y que serán de utilidad al estudiante para el desarrollo de sus tareas. Además, se partirá del interés de los estudiantes para que realicen una investigación sobre temas avanzados de robótica y un proyecto de un sistema de recuperación de información.

Los estudiantes se organizarán en equipos colaborativos, conformados por cuatro estudiantes, y durante el transcurso de las clases del curso desarrollarán un juego de mesa sobre uno de los temas que se ven en el curso. El primer día de clase, se formarán los equipos y se dará un material de trabajo a partir del cual se debe aprender la dinámica de funcionamiento del equipo. Es necesario que se llegue a un consenso dentro de cada equipo sobre cuáles serán las normas de funcionamiento de dicho equipo. Así en caso de conflicto siempre se podrá intentar solucionarlo en base a dicho reglamento.

La investigación partirá del interés personal de los estudiantes para asignar una serie de temas que serán desarrollados por cada equipo de estudiantes. La investigación, la presentación y la aplicación se harán en equipos de dos personas. Cada equipo desarrollará su tema y una aplicación de este, culminará con una presentación en forma de exposición en clase y la elaboración de un vídeo clip del tema.

El proyecto partirá de la comprensión de los estudiantes en los temas vistos en clase para elegir un problema a resolver. El proyecto está dividido en cuatro etapas, que serán desarrolladas por cada equipo de estudiantes. La realización del proyecto se hará en equipos de cuatro personas. Cada equipo desarrollará cada etapa, y al final del semestre culminará con un Sistema de Recuperación de Información y la presentación de este a la profesora y al grupo.

En todas las actividades del curso se debe adjuntar la autoevaluación y coevaluación de cada miembro del equipo, ya que la calificación final del trabajo o evaluación será el promedio entre la calificación obtenida y las evaluaciones (autoevaluación y coevaluación).

Todo trabajo estará sujeto a la evaluación de su redacción, ortografía, estructura y contenido. Todo material base debe ser referenciado utilizando el estilo APA. Para todo producto realizado durante la ejecución de las actividades del curso se deben generar evidencias sobre el proceso realizado.

Los recursos didácticos con los que se contarán son:

- **El docente.** El trabajo principal del docente es guiar o ayudar al estudiantado a conseguir los objetivos del curso. Su mayor éxito será conseguir que todos los estudiantes aprueben el curso. Durante las actividades estará pendiente del trabajo que esté realizando el estudiantado, y les ofrecerá la ayuda necesaria para que puedan completarlo de forma satisfactoria, aunque procurará que primero





intenten resolverlo por ellos mismos. Hay que recordar que el docente estará a disposición en el horario de consultas.

- **Mediación Virtual.** En Mediación Virtual encontrará toda la información y el material del curso y grupo respectivo, como presentaciones, prácticas, soluciones a algunos ejercicios y actividades estarán disponibles en el mismo. También se puede encontrar las calificaciones y los criterios de evaluación de cada trabajo asignado. El nivel de virtualidad que se trabajará en este curso será “Curso bajo virtual” (75% físico y 25% virtual), por lo que algunas clases serán virtuales utilizando la plataforma.
- **Web del curso.** El sitio Web del curso está disponible en <http://www.kramirez.net/ci-2414/>. Toda la información y el material del curso, como presentaciones, prácticas, soluciones a algunos ejercicios y actividades estarán disponibles en el mismo. También se puede encontrar las calificaciones y los criterios de evaluación de cada trabajo asignado.

6. Evaluación

A continuación, se presenta el resumen de acuerdo con los criterios citados anteriormente:

Asignación	Valor
Juego de Mesa	20%
Tareas y Quices	20%
Investigación	20%
Exposición del tema y la aplicación	7.5%
Material Visual – Vídeo Clip	7.5%
Evaluaciones de las exposiciones	5%
Proyecto	40%
Etapa I	10%
Etapa II	10%
Etapa III	10%
Etapa IV	10%
TOTAL	100%

Para aprobar el curso el estudiante debe tener una nota igual o superior a 6.75. Si la nota final está entre 5.75 y 6.74 tendrá derecho a realizar un examen de ampliación, el cual incluye toda la materia del curso; en dicho examen deberá tener una nota mínima de 7.0 para aprobar el curso, siendo la nota final 7.0. En caso de que el estudiante tenga una nota igual o inferior a 5.74, o bien en caso de presentar el examen de ampliación con una nota inferior a 7.0 reprobará el curso.

Observaciones:

- Las tareas deberán ser entregadas al profesor el día propuesto en el enunciado, por cualquier medio que se indique; son grupales.
- Los quices se harán en todas las lecciones y en cualquier momento durante el transcurso de la lección, y solo se repondrán en los casos que establece





el Reglamento de Régimen Académico Estudiantil en su Artículo 24. En los quices se puede utilizar material de apoyo; pueden ser individuales o grupales.

- La investigación consiste en la presentación al grupo y la realización de un vídeo clip, en equipo de dos personas como máximo, de una aplicación sobre temas vistos por el profesor en clases. Se formarán los equipos y se asignarán los temas el primer día de clases. El material visual y la presentación deberán ser entregadas al profesor el día propuesto en el enunciado, por cualquier medio que se indique; son grupales. Cada estudiante llenará una hoja con comentarios sobre las exposiciones de cada equipo, estos comentarios serán evaluados y pasarán a formar parte de su nota individual; ya que indican el interés, aprovechamiento y asimilación de los temas expuestos.
- El proyecto se puede realizar en equipo de cuatro personas como máximo. Se formarán los equipos el primer día de clases. Cada etapa del proyecto deberá ser entregada al profesor el día propuesto en el enunciado, por cualquier medio que se indique. Se realizará a los integrantes de cada equipo, una comprobación individual del trabajo realizado en cada etapa del proyecto.
- Las comprobaciones del proyecto son quices individuales, sobre el trabajo realizado en cada etapa, que comprueba si el estudiante participó o no en el desarrollo de este. Se realizan el día de entrega de cada etapa. La nota obtenida en cada comprobación se promedia con la nota obtenida en la etapa respectiva del proyecto.
- Los criterios de evaluación de cada trabajo asignado se entregarán oportunamente.
- El uso de lápiz en cualquier evaluación se permite, pero no se acepta reclamos. Por lo que, el uso del lapicero es recomendado.
- Todo trabajo debe ser entregado de forma digital.
- Por cada día natural de retraso en la entrega de cualquier trabajo del curso se rebajará un punto en la escala de 1 a 10.
- Si envía por correo electrónico con uno o más días de retraso, se aplicará la regla de rebajo de puntos expuesta arriba con base en la fecha de envío.
- Cuando el estudiante sepa que tendrá que faltar un día particular en el cual debe entregar algún trabajo, se recomienda que lo comunique a la profesora antes de ese día, para coordinar la entrega de alguna forma.
- En todos los trabajos y las evaluaciones de los estudiantes, se calificará la redacción, ortografía, estructura y contenido. Todo material base debe ser referenciado utilizando el estilo APA. Se rebajará de la nota obtenida un punto por cada falta encontrada en alguno de los aspectos citados.
- En todos los trabajos y las evaluaciones de los estudiantes, se debe entregar una autoevaluación y coevaluación, con el fin de evaluar el aporte de sus compañeros de equipo y el propio. La calificación final del trabajo o evaluación será el promedio entre la calificación obtenida y las evaluaciones (autoevaluación y coevaluación).
- En todos los trabajos y las evaluaciones, se le solicitará al estudiante que firme una lista de entrega, para que el estudiante tenga un documento que compruebe que entregó y realizó lo solicitado por el profesor.
- Los estudiantes pueden discutir los trabajos (excluyendo exámenes y quices) con quien sea. Esto incluye hablar sobre interpretaciones del ejercicio asignado, por donde se podría atacar el problema, inclusive la estrategia completa de solución (si alguno de los que discuten ya lo resolvió); además, se puede sugerir y/o buscar material complementario, etc. Lo que no se puede es





copiar la solución de ninguna fuente, ya sea un(a) compañero(a), un libro, Internet, etc.

- Los trabajos serán revisados por la profesora y/o el(la) asistente, si se encuentra evidencia de cualquier tipo de copia, y es la primera vez, los estudiantes involucrados tendrán un cero de nota. Si ocurre una segunda vez, el caso se remitirá a la comisión disciplinaria de la ECCI para aplicar el reglamento.
- En cuanto a reportes y presentaciones se castigará el plagio, el cual se considera copia y será castigado de la misma manera. Se considera plagio la copia literal de segmentos (texto, figuras, tablas u otros datos no textuales) de otra fuente, sin comillas y sin referencia, aunque sea traducido, así como el parafraseo sin aportes de ningún tipo.
- Cuando un(a) estudiante no pueda asistir a efectuar una evaluación por alguna razón de fuerza mayor: la muerte de un pariente hasta de segundo grado, la enfermedad del estudiante u otra situación de fuerza mayor o caso fortuito; se seguirá con lo normado en el artículo 24 del Reglamento de Régimen Académico Estudiantil.

7. Cronograma

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	19/03	29/03	4	Fundamentos de RI
4	02/04	05/04	2	Evaluación de Resultados
5	09/04	12/04	2	SRI Reales
6	23/04	26/04	2	Crawlers (Spiders / Arañas)
7	30/04	03/05	2	Procesamiento de Texto
8	07/05	31/05	8	Operaciones sobre Texto
9	04/06	07/06	2	Indexación y Búsqueda
10	11/06	14/06	2	Operaciones sobre Consultas
11	18/06	21/06	2	Temas avanzados
12	23/06	26/06	2	Presentación del Juego de Mesa
13	30/06	03/07	2	Presentación del proyecto

Tareas

- Tarea #1: Viernes 20 de marzo, hora de clase.
- Tarea #2: Viernes 27 de marzo, hora de clase.





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

Juego de Mesa

- Entrega del documento del diseño: Viernes 3 de abril, hora de clase.
- Presentación en clase del prototipo: Viernes 29 de mayo, hora de clase.
- Presentación en clase del producto final y plenaria: Del 23 al 26 de junio, hora de clase.
- Entrega del producto final: Martes 7 de julio, 9am – 11am.
- Presentación del producto final en la feria: Martes 7 de julio, 9am – 11am.

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.

Proyecto

- Entrega Etapa I: Viernes 27 de marzo.
- Entrega Etapa II: Viernes 24 de abril.
- Entrega Etapa III: Viernes 29 de mayo.
- Entrega Etapa IV: Del 30 de junio al 3 de julio.
- Entrega Final: Martes 7 de julio, 9am – 11am.

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.

8. Bibliografía

1. Baeza-Yates R. & Ribeiro-Neto B. "Modern Information Retrieval". Addison Wesley-ACM Press. 1999. (LIBRO DE TEXTO)
2. Spark Jones, K. & Willett, P. "Readings in Information Retrieval". Morgan Kaufmann Publishers Inc, San Francisco. USA, 1997.

9. Aspectos relacionados con el sistema de bibliotecas, acceso a los reglamentos estudiantiles, y evaluación por parte de los profesores

El Sistema de Bibliotecas, Documentación e Información (SIBDI) de la Universidad de Costa Rica (<http://sibdi.ucr.ac.cr/>) cuenta con una amplia gama de recursos de información bibliográfica en diferentes formatos como libros, folletos,





publicaciones periódicas, trabajos finales de graduación, entre otros. Algunos de estos recursos se encuentran en Biblioteca Virtual, desde la cual se pueden acceder las publicaciones de conferencias y revistas de ACM, IEEE (<http://sibdi.ucr.ac.cr/dbingenieria.jsp>), o Springer, entre otras. La Biblioteca Luis Demetrio Tinoco ofrece cursos de capacitación a los estudiantes del área de las Ingenierías y Computación.

El sitio web del Consejo Universitario de la UCR contiene las diferentes normativas estudiantiles, que pueden ser consultadas desde el enlace <http://www.cu.ucr.ac.cr/normativa/estudiantil.html>.

Los procedimientos de evaluación y orientación establecidos en el Reglamento de Régimen Académico Estudiantil se encuentran en http://www.cu.ucr.ac.cr/uploads/tx_ucruniversitycouncildatabases/normative/regimen_academico_estudiantil.pdf

Destacamos especialmente los siguientes artículos de dicho Reglamento:

- El Artículo 14 se refiere al contenido que debe tener un programa del curso, incluyendo “las normas de evaluación desglosadas y con las ponderaciones de cada aspecto a evaluar”.
- El Artículo 17 indica que “las normas de evaluación conocidas por los estudiantes pueden ser variadas por el profesor con el consentimiento de la mayoría absoluta (más del 50% de los votos) de los estudiantes matriculados en el curso y grupo respectivo”.
- El Artículo 22 indica que “el profesor debe entregar a los alumnos las evaluaciones calificadas y todo documento o material sujeto a evaluación, a más tardar diez días hábiles después de haberse efectuado las evaluaciones y haber recibido los documentos”.
- El Artículo 24 establece el procedimiento y fechas para realizar la reposición de evaluaciones.

