

CARTA AL ESTUDIANTE

Información General

UCR - Universidad de Costa Rica
ECCI - Escuela de Ciencias de la Computación e Informática
CI-2414 Recuperación de Información
Tipo Curso: Teórico-Práctico
4 Créditos
4 Horas clases presenciales a la semana
8 Horas trabajo independiente a la semana
12 Horas totales a la semana
Asistencia Obligatoria¹
Requisito: CI-1330 Ingeniería de Software I
Co-requisito: No tiene
Cupo Mínimo: 10 estudiantes / Cupo Máximo: 22 estudiantes
Curso Electivo del Último Año de la Carrera
Periodo Actual: II Semestre del 2018

Acerca de la Profesora

Dra. Kryscia Daviana Ramírez Benavides
Doctorado Académico en Computación e Informática
Universidad de Costa Rica
Experiencia Docente: 2003-2018 Profesora e investigadora de la UCR
Correos Electrónicos: kryscia.ramirez@eccci.ucr.ac.cr
krysdrb@gmail.com
Oficina: 314 – ECCI (Edificio nuevo – 3er Piso)
Casillero: 58 – ECCI
Teléfono de Oficina: 2511-8019
Teléfono de Secretaria: 2511-8000
Horario de Consulta de Oficina:

- Martes 9:00a.m – 10:00a.m., y 12m.d – 2p.m.

Horario de Atención Opcional: Vía e-mail o Skype (24/7)

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.

¹ Aunque la asistencia a este curso no es obligatoria (de acuerdo al reglamento de la UCR), es necesario que el estudiante asista a todas las clases para poder aprobar el curso con éxito; ya que los objetivos, la metodología, las actividades de enseñanza-aprendizaje y la evaluación están planeadas contando con la participación activa de los estudiantes.

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.

Objetivos

En esta sección se especifican los objetivos formativos que comprende el curso de Estructuras Discretas, desglosándolos en el objetivo general (relacionado con la materia), objetivos específicos (lo que se debe saber al finalizar el curso) y objetivos transversales (capacidades genéricas para un correcto ejercicio de la profesión).

Al mismo tiempo, los objetivos de la asignatura se va a clasificar en tres grupos: conocimiento (información a recordar), comprensión (ser capaz de aplicar un algoritmo previamente conocido), y aplicación (ser capaz de tomar decisiones y decidir entre varias opciones).

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

Al finalizar el curso el estudiante será capaz de:

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 grupo, 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 grupo. (aplicación)
6. Identificar los aspectos que han ido bien y qué aspectos se pueden mejorar en el funcionamiento del grupo. (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)

Contenidos y Cronograma

Nº	Fecha Inicio	Fecha Terminación	Duración (Clases)	Tema
1	14/08	17/08	2	Presentación del curso, lectura de la carta al estudiante. Recuperación de Información: - Introducción a la Recuperación de Información. - Sistema de Recuperación de Información.
2	21/08	31/08	4	Fundamentos de R.I.: - Modelos clásicos (booleano, vectorial y probabilístico). - Modelos alternativos (conjuntos difusos, LSI, redes de inferencia).
3	04/09	07/09	2	Evaluación de Resultados: - Precisión, factor de recuperación y factor de pérdida. - Recall. - Precisión-R, precisión promedio. - Colección TREC.
4	11/09	14/09	2	SRI Reales: - Arquitectura de Google. - SEO y Google. - Otros SRI: Yahoo y Bing (Kumo), y

				comparación con Google (exposición: Grupo #1).
5	18/09	21/09	2	<p><i>Crawlers (Spiders / Arañas):</i></p> <ul style="list-style-type: none"> - Funcionamiento de una <i>crawler</i>. - Aplicación: <i>Crawler</i> (exposición: Grupo #2).
6	25/09	28/09	2	<p>Procesamiento de Texto:</p> <ul style="list-style-type: none"> - Análisis léxico (JLex/JavaCC). - Eliminación de <i>stopwords</i>. - Reglas para extracción de términos. - Aplicación: Analizador (exposición: Grupo #3).
7	02/10	26/10	8	<p>Operaciones sobre Texto:</p> <ul style="list-style-type: none"> - Algoritmos de similaridad y distancia (sobre hileras de texto): <ul style="list-style-type: none"> o Distancia de Hamming. o Distancia de Levenstein. o Bigramas. - Aplicación: Algoritmos de similaridad y distancia (exposición: Grupo #4): <ul style="list-style-type: none"> o Distancia de Hamming. o Distancia de Levenstein. o Bigramas. - <i>Stemming</i> (Lematización): <ul style="list-style-type: none"> o Porter. o Variedad de sucesores. o Bigramas. - Aplicación: Lematizador (exposición: Grupo #5): <ul style="list-style-type: none"> o Porter. o Variedad de sucesores. o Bigramas. - <i>Clustering</i> (Agrupamiento): <ul style="list-style-type: none"> o K-Mens. o HAC. - Aplicación: <i>Clustering</i> (exposición: Grupo #6): <ul style="list-style-type: none"> o K-Mens. o HAC. - Interfaces de visualización: <ul style="list-style-type: none"> o <i>Tree Map</i>. o <i>Sammon Cluster View</i>. o <i>Radial Interactive Visualization</i>. - Aplicación: Interface de visualización (exposición: Grupo #7): <ul style="list-style-type: none"> o <i>Tree Map</i>.

				<ul style="list-style-type: none"> - Compresión: <ul style="list-style-type: none"> o Huffman orientado a bytes. - Aplicación: Compresor (exposición: Grupo #8): <ul style="list-style-type: none"> o Huffman orientado a bytes.
8	30/10	02/11	2	Indexación y Búsqueda: <ul style="list-style-type: none"> - Índices invertidos. - Búsqueda sobre índices comprimidos. - Búsqueda secuencial de texto sobre texto: <ul style="list-style-type: none"> o Shift-Or. o Boyer-Moore Horspool.
9	06/11	09/11	2	Operaciones sobre Consultas: <ul style="list-style-type: none"> - Retroalimentación por relevancia. - Expansión de consultas mediante análisis local de contexto. - Expansión de consultas mediante análisis global de contexto.
10	13/11	16/11	2	Otros Temas: <ul style="list-style-type: none"> - Filtrado de documentos (exposición: Grupo #9). - Sumarización (exposición: Grupo #10). - Optimización en Buscadores (SEO: <i>Search Engine Optimization</i>) (exposición: Grupo #11). - RI Multimedia (exposición: Grupo #12).
11	20/11	30/11	4	Presentación del Proyecto: <ul style="list-style-type: none"> - Sistema de Recuperación de Información.

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).

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.

Partiremos del interés personal de los estudiantes para asignar una serie de temas que serán desarrollados por cada grupo de estudiantes. La investigación, la presentación y la aplicación se harán en grupos de dos personas. Cada grupo desarrollará su tema y una aplicación del mismo, y culminará con una presentación en forma de exposición en clase.

Partiremos de la comprensión de los estudiantes en los temas vistos en clase para asignar un proyecto, el cual está dividido en cuatro etapas, que serán desarrolladas por cada grupo de estudiantes. La realización del proyecto se hará en grupos de dos o tres

personas. Cada grupo desarrollará cada etapa, y al final del semestre culminará con un pequeño Sistema de Recuperación de Información y la presentación del mismo al profesor y al grupo.

Actividades de Enseñanza-Aprendizaje

Carta al Estudiante

El curso Recuperación de Información se planeó como un proceso de enseñanza-aprendizaje conformado por el programa del curso, la profesora y los estudiantes. El programa del curso y la profesora son herramientas cuyas funciones son facilitar el aprendizaje del estudiante. Como cualquier herramienta, sea la más primitiva o la más sofisticada, estos dos elementos no van a producir nada por sí solos y no van a garantizar su éxito en el aprendizaje "por generación espontánea". Se necesita su esfuerzo (es decir su responsabilidad, autocontrol, planificación, capacidad de comunicación efectiva y asertiva) y motivación profesional para lograr los objetivos propuestos.

Para que usted pueda enfrentar con éxito éste y otros cursos de su carrera quisiera darle algunos consejos útiles para el éxito en el aprendizaje:

- Familiarícese con la lógica de la estructura del curso. Comience por leer detalladamente el programa del curso.
- Entienda y aproveche el inmenso potencial que tiene su cerebro. Investigue y descubra las técnicas de estudio que favorezcan su memoria, su creatividad y mejores aprendizajes.
- Desarrolle todo su potencial, todas sus aptitudes, en el sentido de ser un buen escucha y buen comunicador.
- Promueva y aproveche experiencias de aprendizaje que tiendan a una formación integral de su persona.
- Establezca y mantenga el horario de trabajo para aprovechar al máximo este curso. Jerarquice sus tareas, poniendo en primer lugar aquellas que tienen que ver con el estudio y que haga una adecuada distribución y aprovechamiento del tiempo, sin descuidar la importancia del ocio y del tiempo libre.
- No espere que este curso sea fácil, esté preparado para superar las dificultades, ya que lo que todos queremos es la calidad y no la mediocridad. Sea consciente de la responsabilidad por su propio aprendizaje.
- Sea amigable, colaborador y ayude a los demás cuando pueda hacerlo.
- Sea activo, participe en las discusiones en clase y en las actividades dentro y fuera de clase.
- Cuide permanentemente que los profesores y compañeros estudiantes no le disminuyan su nivel de autoestima, ya que una buena autoestima está ligada a un buen aprendizaje.
- Visualícese siempre como un estudiante exitoso y comprométase con las actitudes y acciones para llegar a hacerlo.
- Promueva actitudes, acciones y sentimientos que lo lleven a amar el aprendizaje.
- Si siente que algo anda mal, inmediatamente expréselo para poder resolver el posible problema en forma colaborativa.

Aspectos Generales

En la metodología se describieron las diferentes actividades de enseñanza-aprendizaje que se realizarán en el curso, en dichas actividades el papel del estudiante es muy importante, ya que deberá asumir el rol de un integrante colaborador de un equipo interdisciplinario, en algunos casos será el emisor y en otros el receptor activo de los conocimientos.

A través de actividades de trabajo colaborativo se van a desarrollar diferentes capacidades tanto a nivel de conocimientos propios de la materia como habilidades o competencias transversales del tipo comunicación eficaz con los compañeros, planificación del tiempo, responsabilidad, resolución de conflictos, etc.

La función principal de los equipos de trabajo colaborativo es resolver problemas, se debe hacer énfasis en el consenso, negociación y desarrollo de habilidades de sociales y de equipo. Algunos procedimientos esperados de resolución de problemas son:

- Cada equipo propone su formulación y solución en un papel y se asegura que cada uno de los miembros lo entienda y lo pueda explicar.
- Estudiantes selectos son invitados al azar para presentar su modelo o solución.
- Se espera que todos los miembros de la clase discutan y realicen preguntas de todos los modelos. La discusión se alterna, de toda la clase a un grupo pequeño.
- Cada equipo prepara y entrega un reporte de actividades.
- Los grupos evalúan su efectividad trabajando juntos.

Todos los miembros del equipo deben dar una lista de expectativas de participación y comportamiento en el equipo, los estudiantes:

- Deben generar ideas acerca del comportamiento que pueda interferir en el trabajo en equipo.
- Pueden crear un código de comportamiento para todos los miembros.
- Definir un comportamiento de grupo aceptable.
- Listar los comportamientos que esperan de cada una de las personas, pareja, grupo o de la clase en general.
- Ayudar a que el instructor y los estudiantes muestren comportamientos específicos haciendo que todos se sientan incluidos, expresando, por ejemplo, desacuerdo de manera constructiva, ofreciendo apoyo y soporte, pidiendo aclaraciones, evitando comentarios negativos.

En la sección de contenidos del curso se ha programado los contenidos que se desarrollarán en cada una de las clases, también se presentan las fechas de las diferentes actividades planificadas en la sección [Fechas Importantes](#).

En el curso se utilizarán la estrategia de aprendizaje cooperativo, mediante diferentes técnicas, y el portafolio de evidencias, junto con estrategias de autoevaluación y coevaluación cíclica para la evaluación. Además, se cuenta con el sitio Web, donde se tendrá disponible: calificaciones, presentaciones, trabajos, tareas, etc. En los siguientes links se puede encontrar una descripción más detallada de las actividades planificadas:

- Examen final.
- Tareas y quices.
- Investigación.
- Proyecto.

El número de créditos del curso muestra que se necesitan de 12 horas de trabajo (en clase y fuera de ella) semanales para poder concluir con éxito el aprendizaje de los

contenidos del curso. El estudiante deberá trabajar 8 horas semanales mínimo fuera de la clase para poder cumplir con las actividades planificadas.

Los recursos didácticos con los que se contarán son:

- **La profesora.** El trabajo principal de la profesora 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. Recordar que la profesora estará a disposición en el horario de consultas.
- **Web del curso.** El sitio Web del curso está disponible en <http://www.kramirez.net/ci-2414/>. Toda la información y material de la materia como presentaciones, prácticas, soluciones a algunos ejercicios y actividades estarán disponibles en el mismo. También se puede encontrar las calificaciones.

Evaluación y Medición

La evaluación se usará no tanto como mecanismo de verificación de conocimientos, sino como estímulo para que se hagan las actividades planificadas que conducirán al estudiantado al aprendizaje.

La evaluación de todo el proceso de aprendizaje se va a dividir en:

1. La realización de las actividades organizadas mediante equipos de aprendizaje colaborativo ponderará un 80% de la nota final. Esta parte estará formada por diferentes aspectos de seguimiento de las actividades a través del proyecto, la investigación, las tareas y los quices, controles individuales y de grupo.
2. Un examen final con una ponderación del 20%. En dicha prueba, se va a dar una mayor importancia a la comprensión de un tema de los contenidos del curso por parte del estudiante.

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

Examen Final	20%
Sitio Web del Tema	10%
Evaluación del Tema	10%
Tareas y Quices	20%
Investigación	20%
Exposición del Tema y la Aplicación	15%
Evaluaciones de las Exposiciones	5%
Proyecto	40%
Etapa I	10%
Etapa II	10%
Etapa III	10%
Etapa IV	10%
Total	100%

NOTA: 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 no se reponen. Los quices son quices normales, sin ningún material de apoyo; pueden ser individuales o grupales.
- La investigación consiste en la presentación al grupo, en grupo de dos personas como máximo, de una aplicación sobre temas vistos por el profesor en clases. Se formarán los grupos y se asignarán los temas el primer día de clases.
- El proyecto se puede realizar en grupo de cuatro personas como máximo. Se formarán los grupos 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 grupo, 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 del mismo. 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 y ortografía; por lo que, se rebajará de la nota obtenida un punto por cada falta de ortografía y mala redacción.
- 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 normalizado en el artículo 24 del Reglamento de Régimen Académico Estudiantil.

Fechas Importantes

Examen Final

- Entrega del Sitio Web: Viernes 16 de noviembre.
- Evaluación: Viernes 16 de noviembre.

Tareas

- Tarea #1: Viernes 17 de agosto, hora de clase.
- Tarea #2: Viernes 24 de agosto, hora de clase.
- Tarea #3: Viernes 31 de agosto, hora de clase.
- Tarea #4: Viernes 14 de septiembre, hora de clase.
- Tarea #5: Viernes 28 de septiembre, hora de clase.
- Tarea #6: Viernes 16 de noviembre, hora de clase.

Investigación

- Entrega de la Presentación: Una semana antes de la fecha de exposición.
- Exposición de la Aplicación: Del 14 de septiembre al 16 de noviembre.

Proyecto

- Entrega Etapa I: Viernes 31 de agosto.
- Entrega Etapa II: Viernes 28 de septiembre.
- Entrega Etapa III: Viernes 26 de octubre.
- Entrega Etapa IV: Del 20 al 30 de noviembre.
- Entrega Final: Martes 3 de diciembre, 9 a.m. – 11 a.m.

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.