



UNIVERSIDAD NACIONAL DE SAN LUIS
RECTORADO

SAN LUIS, 19 de mayo de 2022

VISTO:

El EXPE-143/2022 mediante el cual se solicita la protocolización del Curso de Posgrado: BÚSQUEDA INTELIGENTE DE INFORMACIÓN; y

CONSIDERANDO:

Que el Curso de Posgrado se propone dictar en el ámbito de la Facultad de Ciencias Físico Matemáticas y Naturales del 10 de junio al 20 de agosto de 2022 con un crédito horario de SESENTA (60) horas presenciales y bajo la coordinación de la Dra. Ana GARIS.

Que la Comisión Asesora de Posgrado de la Facultad de Ciencias Físico Matemáticas y Naturales recomienda aprobar el curso de referencia.

Que el Consejo de Posgrado de la Universidad Nacional de San Luis en su reunión del 12 de abril de 2022, analizó la propuesta y observa que el programa del curso, bibliografía, metodología de evaluación y docentes a cargo, constituyen una propuesta de formación de posgrado de calidad en su campo específico de estudio.

Que la RCS N° 400/2020 contiene las decisiones y propuestas de funcionamiento de las actividades de posgrado en el marco de la situación sanitaria vigente COVID – 19, y que esta actividad se enmarca en las acciones orientadas a continuar y sostener el dictado de las actividades previstas en cronogramas de estudiantes y propuestas.

Que, por lo expuesto, el Consejo de Posgrado aprueba la propuesta como Curso de Posgrado, según lo establecido en Ordenanza CS N° 35/2016.

Que corresponde su protocolización.

Por ello y en uso de sus atribuciones:

EL RECTOR DE LA UNIVERSIDAD NACIONAL DE SAN LUIS

RESUELVE:

ARTÍCULO 1º.- Protocolizar el dictado del Curso de Posgrado: BÚSQUEDA INTELIGENTE DE INFORMACIÓN, en el ámbito de la Facultad de Ciencias Físico Matemáticas y Naturales del 10 de junio al 20 de agosto de 2022 con un crédito horario de SESENTA (60) horas presenciales.



UNIVERSIDAD NACIONAL DE SAN LUIS
RECTORADO

ARTÍCULO 2°.- Protocolizar como docente responsable del curso a la Dra. Graciela Verónica GIL COSTA, DU N° 27135820 y Auxiliar Ing. Fernando LOOR, DU N° 34408504 ambos de la Universidad Nacional de San Luis.

ARTÍCULO 3°.- Aprobar el programa del Curso de referencia, de acuerdo al ANEXO de la presente disposición.

ARTÍCULO 4°.- Comuníquese, Publíquese en el Digesto Administrativo de la Universidad Nacional de San Luis, insértese en el Libro de Resoluciones, y archívese.

MSS

Documento firmado digitalmente según Ordenanza Rectoral N° 15/2021 por: Vicerrector FLORES, Héctor Daniel (a cargo RR-807/2022) – Secretaria de Posgrado REYES, Nora Susana



UNIVERSIDAD NACIONAL DE SAN LUIS
RECTORADO



ANEXO

IDENTIFICACIÓN DEL CURSO

UNIDAD ACADÉMICA RESPONSABLE: Facultad de Ciencias Físico Matemáticas y Naturales

DENOMINACIÓN DEL CURSO: BÚSQUEDA INTELIGENTE DE INFORMACIÓN

CATEGORIZACIÓN: Perfeccionamiento

FECHA DE DICTADO DEL CURSO: 10 de junio al 20 de agosto de 2022

MODALIDAD DE DICTADO: Presencial.

CRÉDITO HORARIO TOTAL: Sesenta (60) horas (20 horas teóricas, 20 horas de prácticas de aula y 20 horas de prácticas de laboratorio)

COORDINADORA: Dra. Ana GARIS DU N° 25700407

EQUIPO DOCENTE

RESPONSABLE: Dra. Graciela Verónica GIL COSTA

AUXILIAR: Ing. Fernando LOOR

- Tareas a desarrollar: Asistir en las actividades docentes a la Prof. Responsable y colaborar en actividades de laboratorio y prácticas.

PROGRAMA ANALÍTICO

FUNDAMENTACIÓN:

Diversos autores y trabajos de investigación confiables destacan que toda la información producida por la humanidad desde la antigüedad hasta fines del siglo xx, es menor al volumen de datos que es accesible en la Web desde la última década y que gestionamos a través de buscadores Webs y Metabuscadore. Los motores de búsqueda comerciales no siempre son eficaces y en la mayoría de las oportunidades son ineficientes cuando las búsquedas abandonan el ámbito de lo trivial. Hay también que considerar el crecimiento de las redes sociales y foros de opinión, que provocan un aumento exponencial de los datos en circulación, ya no sólo en forma de texto sino también en lenguaje multimedia. Este módulo está destinado a analizar fortalezas y debilidades de los sistemas de recuperación de información y está orientado a ampliar las base conceptual y la generación



de habilidades para acceder a los contenidos pretendidos, a proporcionar conceptos para comprender los procesos intervinientes y para la utilización de la Inteligencia Artificial como un área a explorar para la evolución y mejora de las herramientas para consultas cada vez más precisas.

OBJETIVOS

- Estudiar en profundidad, mediante un análisis comparativo, los aspectos relacionados con recuperación de información vs. recuperación de datos.
- Analizar los procesos de recuperación de información y proponer innovaciones.
- Discutir y proponer aspectos concretos respecto de la utilización del comportamiento del usuario como realimentación.
- Examinar las principales características de los buscadores de Internet más conocidos: arquitectura y operadores. Proponer mejoras. Enfoques de tiempo real.
- Estudiar comparativamente diversos buscadores de la Web.
- Identificar los parámetros que tienen mayor impacto en los tiempos de respuesta de las consultas Web mediante análisis de sensibilidad.

CONTENIDOS MÍNIMOS

Recuperación de información vs. recuperación de datos. Procesos de recuperación de información. Herramientas de búsqueda en Internet. Buscadores y Meta-Buscadores. Sistemas de pregunta-respuesta. El usuario y los buscadores. Tipos de consultas. Comportamiento del usuario como realimentación. Examen de las principales características de los buscadores de Internet más conocidos: Arquitectura y operadores. Operadores de búsqueda. Resultados. Métricas de evaluación. Identificación de parámetros relevantes mediante análisis de sensibilidad. Problemas en la recuperación de información. Recuperación de Información y técnicas de representación de información y ranking de resultados. Arquitectura de motores de búsqueda. Web caching y procesamiento de consultas Multithreading en motores de búsqueda Estructuras de datos para índices y búsquedas. Bases de datos multimediales. Estructuras de datos para espacios métricos. Técnicas de paralelización de índices y clusters de alto rendimiento. La Web: caracterización, indexación, algoritmos de crawling, algoritmos de ranking, centros de datos, capacity planning, diversificación de resultados, tópicos avanzados.

PROGRAMA DETALLADO:

Unidad I: Recuperación de información vs. recuperación de datos

Búsqueda sobre datos estructurados. Bases de Datos, SQL. Portal vertical, “búsqueda avanzada”, formularios de búsqueda. Recuperación de información; búsqueda abierta. Búsqueda específica. Búsqueda exhaustiva vs. algún buen ejemplo, búsqueda de los mejores documentos vs. recopilación de fuentes. Búsqueda navegacional. Búsqueda con incertidumbre: En la propia consulta, en el significado de la información disponible, en el grado de cumplimiento de las condiciones de consulta. Búsquedas para satisfacer



necesidades de información: De un usuario, de un sistema, por iniciativa del usuario (consulta vs. browsing) o del sistema. Espacio de búsqueda no estructurado (o no totalmente): Texto, Multimedia, Objetos de un dominio (productos, personas, etc.)

Unidad II: Arquitectura de motores de búsqueda

Componentes: Crawler, Indexador, Interfaz, Motor de búsqueda. Motores de Búsqueda basados en servicios. Particionado y replicas. Tipos de servicios. Consumo de energía. Motores de búsqueda verticales. Redes de comunicación. Fat-Tree. Listas invertidas. Algoritmos de ranking. Particionado de índice. Listas invertidas en paralelo.

Unidad III: Web caching y procesamiento de consultas

Procesamiento Round-Robin de Consultas: Modo Asíncronico, Modo Sincronico. Clustering de tópicos. Técnicas de caching. Cache estática. Refresco de la cache. Políticas de reemplazo. SDC y LandLord. Jerarquía de cachés. Location cache. Flujo de consultas.

Unidad IV: Multithreading en motores de búsqueda

Arquitectura multicore. Planificación de threads. Real time search. Conflicto de transacciones R/W. Transacciones atómicas y seriales. Query Solver. Algoritmo Bulk Processing. Alzas bruscas en el tráfico de consultas.

Unidad V: Índices para Espacios Métricos

Definición. Dimensionalidad. Aplicaciones. Tipo de consultas. Índices. Clasificación de las Técnicas de Búsqueda: basados en pivotes, clustering, híbridos y permutaciones. Construcción de índices y algoritmos de búsqueda en ambientes distribuidos.

Unidad VI: Técnicas de optimización

Algoritmos de enrutamiento. Respuestas aproximadas mediante técnicas de machine learning y PCAP. Scheduling. Mapping se de servicios. Tolerancia a fallos. Balance de carga. Compresión de Índices. Generación de Snnipets. Diversificación de resultados. Diversificación implícita o explícita. Algoritmos greedy y heurísticas. Análisis de Sensibilidad.

Unidad VII: Análisis comparativo de los buscadores más conocidos

Examen de las principales características de los buscadores de Internet más conocidos: Arquitectura y operadores. Operadores de búsqueda. Resultados. Problemas en la recuperación de información. Naturaleza de un metabuscador: Sus ventajas. Estudio de diversos metabuscadores de la Web.

Unidad VIII: Recuperación de Información en la Web con Técnicas de Soft-Computing

Estudio del diseño, pruebas e implementación de modelos principalmente basados en técnicas de Soft-Computing de representación conceptual de documentos. Diferentes métodos y técnicas de inferencia, clasificación, clustering, procesado del lenguaje natural. Mejorar la eficiencia conceptual y de la búsqueda de información en la Web.

SISTEMA DE EVALUACIÓN:

Para aprobar el curso, se requerirá una asistencia no inferior al 80 %. Además, el estudiante deberá desarrollar los casos prácticos correspondientes a cada tópico del curso



y para deberá presentar un proyecto integrador individual, con el objetivo de evaluar si se ha logrado incorporar los conocimientos del curso.

BIBLIOGRAFÍA

- Al-Fares, M; Loukissas, A; Vahdat, A. A scalable, commodity data center network architecture. Proceedings of the ACM SIGCOMM 2008 conference on Data communication. ACM. Pág: 63–74. USA. 2008.
- Arroyuelo, Diego; Gil-Costa, Veronica; Marin, Mauricio; Gonzales, Senen; Silva, Mauricio. Distributed search based on self-indexed compressed text. Information Processing and Management (IPM) Special Issue on Large-Scale Distributed systems for information retrieval. Elsevier. 2011.
- Baeza-Yates, R. A; Gionis, A; Junqueira, F; Murdock, V; Plachouras, V; Silvestri, F. Design trade-offs for search engine caching. ACM Trans. Web. Vol:2. Num: 4. Pág: 1-28. ACM. USA. 2008.
- Baeza-Yates, R; Ribeiro-Neto, B. Modern Information Retrieval - the concepts and technology behind search, 2nd edn. Pearson Education Ltd. 2011.
- Broder, A.Z; Carmel, D; Herscovici, M; Soffer, A; Zien, J.Y. Efficient query evaluation using a two-level retrieval process. CIKM. Pág: 426–434. 2003.
- Buettcher, S; Clarke, C; Cormack, G. Information Retrieval: Implementing and Evaluating Search Engines. The MIT Press. 2010.
- Dover, Danny; Dafforn, Erik. Search Engine Optimization (SEO) Secrets. Wiley, 2011.
- Fagni, T; Perego, R; Silvestri, F; Orlando, S. Boosting the performance of web search engines: Caching and prefetching query results by exploiting historical usage data. ACM Trans. Inf. Syst. Vol: 24. Pág: 51–78. 2006.
- Gan, Q; Suel, T. Improved techniques for result caching in web search engines. Proceedings of the 18th international conference on World Wide Web WWW 09. ACM. Pág: 431–440. USA. 2009.
- Gaioso [Roussian R. A.](#) , Gil-Costa Veronica, [Hélio Guardia](#), [Hermes Senger](#): Performance evaluation of single vs. batch of queries on GPUs. [Concurr. Comput. Pract. Exp. 32\(20\)](#) (2020)
- Gil-Costa, Veronica; Santos, Rodrygo; Macdonald, Craig; Ounis, Iadh. Modelling Efficient Novelty-based Search Result Diversification in Metric Spaces. Journal of Discrete Algorithms. Elsevier. 2012.
- Gil-Costa, Veronica; Lobos, Jair; Inostrosa-Psijas, Alonso; Marin, Mauricio. Capacity Planning for Vertical Search Engines: An approach based on Coloured Petri Nets. 33rd International Conference on Application and Theory of Petri Nets and Concurrency - Petri Nets 2012. Germany. June 25–29, 2012.
- Gil-Costa Veronica, [Inostrosa-Psijas](#) Alonso, Marin Mauricio: Elastic and Real-time Capacity Planning for Web Search Engines. [PDP 2020](#): 331-338
- Hartman, Karen; Ackermann, Ernest. Searching and Researching on the Internet and the World Wide Web 5th Edition. Franklin, Beedle & Associates Inc. 2010.



Loor Fernando, Manriquez [Manuel](#), Gil-Costa Veronica Gil-Costa, Marin [Mauricio](#) Feasibility of P2P-STB based crowdsourcing to speed-up photo classification for natural disasters. [Clust. Comput. 25\(1\)](#): 279-302 (2022)

[Marín](#) Mauricio, Gil-Costa Veronica, [Inostrosa-Psijas](#) Alonso, [Bonacic](#) Carolina:Hybrid capacity planning methodology for web search engines. [Simul. Model. Pract. Theory 93](#): 148-163 (2019)

Moffat, A; Webber, W; Zobel, J; Baeza-Yates, R. A pipelined architecture for distributed text query evaluation. Information Retrieval. Vol: 10. Kluwer Academic Publishers. Pág: 205–231. USA. 2007.

Molina [Romina](#), Loor [Fernando](#), Gil-Costa Veronica, [Franco Maria Nardini](#), [Raffaele Perego](#), [Salvatore Trani](#): Efficient traversal of decision tree ensembles with FPGAs. [J. Parallel Distributed Comput. 155](#): 38-49 (2021)

Siri, Laura. Internet Búsquedas y Buscadores (Spanish Edition). Grupo Editorial Norma. 2005.

CARACTERÍSTICAS DEL CURSO

DESTINATARIOS Y REQUISITOS DE INSCRIPCIÓN: Profesionales de carreras universitarias o carreras de 4 años de duración como mínimo: Profesionales en Informática y títulos de grado similares correspondientes a carreras con planes de estudio que impliquen 4 o más años de duración.

CUPO: Mínimo: 5 personas. Máximo: 20 personas

PROCESO DE ADMISIÓN: En caso de que la demanda exceda el cupo previsto para este curso, se priorizará la admisión de estudiantes de carreras de posgrado de la Facultad de Ciencias Físico Matemáticas y Naturales y se evaluará el CV del estudiante.

CRONOGRAMA DE ACTIVIDADES:

Fecha	Tipo de actividad /temas a desarrollar	Docente/s responsable/s de la actividad	Ámbito/plataforma digital
10 y 11 de Junio	Clases de unidad 1 y 2	Dra. Verónica Gil Costa	Aula Posgrado Dpto. Informática
24 y 25 de Junio	Clases unidad 3, 4	Dra. Verónica Gil Costa Ing. Fernando Loor	Aula Posgrado Dpto. Informática
12 y 13 de Agosto	Clases unidad 5 y 6	Dra. Verónica Gil Costa Ing. Fernando Loor	Aula Posgrado Dpto. Informática
19 y 20 de Agosto	Clases unidad 7 y 8	Dra. Verónica Gil Costa	Aula Posgrado Dpto. Informática



UNIVERSIDAD NACIONAL DE SAN LUIS
RECTORADO



LUGAR DE DICTADO: Aula de posgrado del Dpto. de Informática de la FCFMyN

FECHA PREVISTA PARA ELEVAR LA NÓMINA DE ESTUDIANTES
APROBADOS: Diciembre de 2022

FINANCIAMIENTO DEL CURSO

COSTOS: Materiales e insumos.

FUENTES DE FINANCIAMIENTO: UNSL

ARANCEL GENERAL: Gratuito

Hoja de firmas