

Sistemas de Información  
3er Año – Ciencia de la Computación – Curso 2021-2022  
Orientaciones para el Proyecto Final

El Proyecto Final de Sistemas de Recuperación de Información en el curso 2021-2022 consiste en el diseño, implementación, evaluación y análisis de un Sistema de Recuperación de Información. El trabajo se desarrollará en equipos de dos estudiantes. **La fecha límite de entrega es el lunes de la semana 16 del curso 2021-2022.**

El sistema a desarrollar debe comprender todas las etapas del proceso de recuperación de información. Es decir, desde el procesamiento de la consulta hecha por un usuario, la representación de los documentos y la consulta, el funcionamiento del motor de búsqueda y la obtención de los resultados. No hay limitaciones respecto al Modelo de Recuperación de Información que deben emplear, puede ser cualquiera de los clásicos o alguno de los alternativos, siempre atendiendo a las características de cada uno y su adecuación al escenario en el que se aplicarán.

El ejercicio académico concluye con la **entrega y discusión de un artículo científico** en formato LNCS, que se adjunta a las orientaciones. Dicho artículo debe describir todo el proceso de diseño, implementación y análisis de los resultados del sistema. Además, **debe adjuntarse el código fuente del sistema implementado**. De manera general, el informe debe incluir:

- Diseño completo del sistema según cada etapa de la recuperación de información. Debe especificar las consideraciones y argumentos para la elección y aplicación de los modelos y enfoques adoptados.
- Presentación de las herramientas empleadas para la programación y aspectos más importantes del código.
- Evaluación del sistema empleando las métricas objetivas y subjetivas estudiadas en clase empleando al menos dos colecciones de prueba distintas (incorporar consultas de ejemplo con los resultados proveídos por la solución, al menos una por cada colección de prueba).
- Análisis crítico de las ventajas y desventajas del sistema desarrollado.
- Recomendaciones para trabajos futuros que mejoren la propuesta.

La discusión del proyecto se hará en función de las condiciones epidemiológicas existentes en el momento de entrega, ya sea que se realice de manera virtual a partir de preguntas y respuestas o en un intercambio presencial en la facultad.

Adicionalmente, exhortamos a la incorporación de funciones adicionales o mejoras a sus soluciones. Esto, sin duda, favorecerá el desempeño de sus sistemas, la consolidación de los contenidos abordados en conferencia y tendrá un impacto positivo en la evaluación final. A continuación, proponemos algunas funcionalidades, aunque son libres de plantear otras.

- Operaciones textuales más avanzadas.
- Incorporación de retroalimentación.
- Expansión de consultas.
- Integración con algoritmos de *Crawling*.

- Uso de bases de conocimiento como ontologías o tesauros.
- Interfaz visual para la realización de consultas.

Adjuntamos a las orientaciones tres colecciones de prueba que pueden ser empleadas para la evaluación de sus sistemas. Son libres de emplear otras a las que tengan acceso. En particular la colección Cranfield cuenta con un conjunto de consultas y la información de los documentos relevantes a las mismas.

- Cranfield
- Reuters-21578
- 20 Newsgroups

Ante cualquier dificultad o impedimento para realizar la actividad, comunicarse inmediatamente con los profesores de la asignatura.