

ANTEPROYECTO DEL TRABAJO DE FIN DE GRADO

| | | | | | |
|--|--|--------------------------|-----------|-------------------------------------|--|
| Alumno/a | Enrique López Encinas | | | | |
| Titulación: | Grado en Ingeniería Informática | | | | |
| Tutor/es: | Eduardo Guzmán De los Riscos | | | | |
| Título | Herramienta de crawling de publicaciones científicas | | | | |
| Subtítulo <i>(solo si en grupo)</i> | | | | | |
| Título en inglés | Scientific publications crawling tool | | | | |
| Subtítulo en inglés <i>(solo si en grupo)</i> | | | | | |
| Trabajo en grupo: | Sí | <input type="checkbox"/> | No | <input checked="" type="checkbox"/> | |
| Otros integrantes del grupo: | | | | | |

Contextualización del problema a resolver. Describir claramente de dónde surge la necesidad de este TFG y el dominio de aplicación. En caso de que el TFG se base en trabajos previos, debe aclararse cuáles son las aportaciones del TFG.

El objetivo es analizar la evolución de la investigación en Informática y en especial, en Inteligencia Artificial, en las últimas décadas. Para ello se montará una arquitectura basada en microservicios que contenga servicios de extracción en Python y una pequeña aplicación web encargada de integrar los datos obtenidos en un modelo que limpie los datos de entrada y los estructure como paso previo al análisis.

Descripción detallada de en qué consistirá el TFG. En caso de que el objeto principal del TFG sea el desarrollo de software, además de los objetivos generales deben describirse sus funcionalidades a alto nivel.

Con este TFG se busca desarrollar una aplicación web que recopile información de varios sitios web dedicadas a la divulgación de artículos académicos (por ejemplo, Google Scholar) así como de sus autores para ofrecer al usuario una herramienta que contenga toda la información que pueda necesitar sin la necesidad de consultar diferentes fuentes externas, ahorrando de esta forma tiempo y recursos.

Constará de una base de datos relacional donde se guardará toda la información que se obtenga de las búsquedas de los usuarios, que se realizarán a través de una aplicación web basada en Angular.

Esta información se obtendrá usando el método de “web scraping”, a través de scripts capaces de recoger y organizar todos los datos de una página web.

Listado de resultados que generará el TFG (aplicaciones, estudios, manuales, etc.)

Aplicación web

Memoria del TFG

METODOLOGÍA:

Descripción de la metodología empleada en el desarrollo del TFG. Especificar cómo se va a desarrollar. Concretar si se trata de alguna metodología existente y, en caso contrario, describir y justificar adecuadamente los métodos que se aplicarán.

Se usará una metodología ágil, que permite gran flexibilidad a la hora de realizar las diferentes tareas del proyecto según se vayan necesitando.

A pesar de existir diferentes fases de trabajo en el proyecto, esta metodología facilita saltar de una a otra si se encuentran errores que necesitan ser arreglados con urgencia o se necesita rehacer o cambiar alguna parte anterior.

FASES DE TRABAJO:

Enumeración y breve descripción de las fases de trabajo en las que consistirá el TFG.

Contextualización del proyecto: Esta fase está enfocada en entender el alcance del proyecto. Tienen que quedar claras las fuentes de información (páginas web) de la que obtener los datos, así como las tecnologías necesarias para realizar el proyecto según los requisitos que se definan.

Fase de aprendizaje: En esta fase el alumno investigará las tecnologías que requiera para la realización del TFG, así como aprender lo básico para manejarse con ellas.

Fase de desarrollo del crawler: Se desarrollará el script en Python que se encarga de realizar el crawling a Google Scholar y de la limpieza de los datos para su uso posterior.

Fase de desarrollo del Backend y Frontend: En estas dos etapas se montará la aplicación web donde se guardarán y mostrarán los datos obtenidos al usuario final.

Documentación y pruebas: Durante todo el proyecto se llevarán a cabo periodos de pruebas del código para verificar que todo funcione como debe, y se documentará conforme se vaya trabajando en el proyecto.

Memoria final del proyecto: Finalmente se elaborará la memoria del proyecto.

TEMPORIZACIÓN:

La siguiente tabla deberá contener una fila por cada una de las fases enumeradas en la sección anterior. En caso de tratarse de un trabajo en grupo, se añadirá una columna HORAS por cada miembro del equipo. Debe especificarse claramente el número de horas dedicado por cada alumno/a y la suma de horas individual deberá ser también de 296.

| FASE | HORAS |
|---|-----------------------|
| | Enrique López Encinas |
| Contextualización del proyecto | 24 |
| Investigar páginas web que se usarán para extraer datos | 8 |
| Definir tecnologías y lenguajes que se usarán | 4 |



| | |
|--|------------|
| Estructurar trabajo a realizar durante el proyecto | 4 |
| Rellenar plantilla anteproyecto | 5 |
| Crear repositorio git para guardar el proyecto | 3 |
| Aprendizaje de los lenguajes y tecnologías a usar. | 43 |
| Aprendizaje Python | 4 |
| Aprendizaje HTML y CSS | 4 |
| Aprendizaje Typescript | 8 |
| Aprendizaje Git | 3 |
| Aprendizaje Angular | 16 |
| Aprendizaje de técnicas de scraping de páginas web | 8 |
| Desarrollo de los Crawlers | 64 |
| Realizar crawler para Google Scholar | 24 |
| Realizar crawler para otra fuente de datos | 24 |
| Script para estructurar y exportar los datos obtenidos a la BBDD | 16 |
| Desarrollo del Backend | 36 |
| Creación de la base de datos relacional | 8 |
| Crear modelo lógico y tablas a usar | 24 |
| Introducir los datos obtenidos en la base de datos | 4 |
| Desarrollo del Frontend | 72 |
| Crear proyecto Angular | 8 |
| Creación página Inicio | 8 |
| Creación de página para introducir filtros de búsqueda | 24 |
| Creación de página para visualización de los datos obtenidos | 16 |
| Dar estilo a las páginas anteriores | 16 |
| Documentación y pruebas | 36 |
| Pruebas y Test sobre el crawler | 8 |
| Pruebas y Test sobre Backend | 8 |
| Pruebas y Test sobre el Frontend | 8 |
| Documentación de los scripts del crawler | 4 |
| Documentación del Backend | 4 |
| Documentación del Frontend | 4 |
| Realización de la memoria final del proyecto | 21 |
| | 296 |



TECNOLOGÍAS EMPLEADAS:

Enumeración de las tecnologías utilizadas (lenguajes de programación, frameworks, sistemas gestores de bases de datos, etc.) en el desarrollo del TFG.

Python, Angular, Typescript, HTML, CSS, Git.

Gestor de base de datos relacional (MySQL, MariaDB, PostgreSQL, ...)

RECURSOS SOFTWARE Y HARDWARE:

Listado de dispositivos (placas de desarrollo, microcontroladores, procesadores, sensores, robots, etc.) o software (IDE, editores, etc.) empleados en el desarrollo del TFG.

Visual Studio Code.

Listado de referencias (libros, páginas web, etc.)

[Google Scholar](#)

[Web of Science](#)

[Exaly](#)

[ORCID](#)

[DBPL](#)

[ResearchGate](#)

Málaga, _____ de _____ de _____

Firma tutor/tutora:

Firma cotutor/a:

Firma tutor/a coordinador/a: