



# TRABAJO FIN DE GRADO

---

*Extracción de información a partir  
de descripciones de cursos*

---

## PLAN DE TRABAJO

**AUTOR:** RUTH VERÓNICA OCAMPO PRADO  
**TUTORA:** MARÍA NAVAS LORO  
**COTUTORA:** PATRICIA MARTÍN  
**GRADO:** INGENIERÍA INFORMÁTICA

# **CONTENIDO**

<b>1. Descripción general del trabajo .....</b>	<b>1</b>
<b>2. Lista de objetivos .....</b>	<b>2</b>
<b>3. Lista de tareas .....</b>	<b>2</b>
<b>4. Diagrama de Gantt .....</b>	<b>3</b>
<b>5. Copia de la propuesta de trabajo del tutor .....</b>	<b>4</b>

# **1. Descripción general del trabajo**

Este trabajo de fin de grado tiene como fin colaborar con el proyecto europeo AI4LABOUR, el cual tiene como objetivo predecir qué puestos de trabajos serán en un futuro reemplazados por máquinas y, además, conseguir que las personas que ocupan estos puestos puedan ser formadas o llevadas a otro puesto, en el que realmente se pueda aprovechar al máximo las capacidades que posee dicha persona.

Este proceso se llevará a cabo de la siguiente manera, si una persona cumple con la mayoría de los requisitos que se pide para un determinado puesto, será formada para que cumpla todos los requerimientos que le falta por cumplir y luego, será trasladada a ese puesto.

El trabajo solicitado por los tutores es ayudar a desarrollar e implementar un algoritmo que compare las distintas habilidades que puede poseer un trabajador, y determinar si tienen el mismo significado, puesto que, en muchas ocasiones, las descripciones de las habilidades que se adquieren en los cursos o que vienen descritas en los currículums pueden significar lo mismo, pero estar escritas de manera diferente.

Con este proyecto se espera reducir el número de personas afectadas por la implantación de máquinas en su área de trabajo, lo que, en la mayoría de los casos, termina siendo un despido o un cambio a un puesto que no se corresponde con las habilidades reales del trabajador.

## **2. Lista de objetivos**

- Diseño e implementación de algoritmo de identificación de habilidades y tareas similares de distintas bases de datos.
- Diseño e implementación de extracción de dichas tareas y habilidades a partir de descripciones textuales de cursos.
- Evaluación de los algoritmos.

## **3. Lista de tareas**

- Análisis de scripts previos, datos disponibles y estado del arte (50h).
- Diseño de algoritmo de identificación de habilidades/tareas (entidades) similares (10h).
- Implementación del algoritmo (40h).
- Evaluación del algoritmo (20h).
- Diseño del algoritmo de extracción de entidades a partir de descripciones de cursos (10h).
- Implementación del algoritmo (87h).
- Evaluación del algoritmo (20h).
- Documentación y escritura del TFG (40h).
- Tutorías/reuniones progreso (20h).

## 4. Diagrama de Gantt

				SEMANAS																
TAREAS / Semanas	Inicio	Duración horas	Fin	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Reunión inicial	18-sep	1	24-sep	1,0																
Análisis de scripts previos	18-sep	10	24-sep	8,5	1,5															
Análisis de datos disponibles	25-sep	11	02-oct		10,0	1,0														
Reuniones/progresos 1	02-oct	1,5	02-oct			1,5														
Análisis de estado del arte	02-oct	29	15-oct			13,0	16,0													
Reuniones/progresos 2	09-oct	2	09-oct				2,0													
Diseño de algoritmo de identificación de habilidades	16-oct	10	22-oct					10,0												
Reuniones/progresos 3	16-oct	2	16-oct					2,0												
Implementación del algoritmo	16-oct	40	12-nov					4,0	15,0	16,0	5,0									
Reuniones/progresos 4	23-oct	2	23-oct						2,0											
Reuniones/progresos 5	30-oct	1,5	30-oct							1,5										
Evaluación del algoritmo	06-nov	20	18-nov								10,0	10,0								
Diseño del algoritmo de extracción de entidades	13-nov	10	23-nov									2,0	8,0							
Reuniones/progresos 7	20-nov	2	20-nov										2,0							
Implementación del algoritmo	20-nov	87	24-dic										7,0	20,0	20,0	20,0	20,0			
Reuniones/progresos 8	04-dic	2	04-dic												2,0					
Reuniones/progresos 9	18-dic	2	18-dic														2,0			
Evaluación del algoritmo	26-dic	20	09-ene															5,0	15,0	
Documentación y escritura del TFG 1ª parte	30-oct	8	19-nov							2,0	3,0	3,0								
Reuniones/progresos 6	19-nov	2	19-nov									2,0								
Documentación y escritura del TFG 2ª parte	20-nov	32	15-ene										3,0	2,0	2,0	2,0	2,0	6,0	4,0	11,0
Reuniones/progresos 10	15-ene	2	15-ene																	2,0

## **5. Copia de la propuesta de trabajo del tutor**

En el contexto del proyecto europeo AI4LABOUR, necesitamos extraer las habilidades y tareas que cubren distintos cursos de plataformas MOOC y guías docentes universitarias. Partiendo de distintas bases de datos de habilidades y tareas, el alumno deberá identificar las habilidades y tareas similares (por ejemplo, en una base de datos una habilidad puede ser “Programación en C” y en otra “Lenguaje C”, pero realmente son la misma) y localizar las apariciones de estas en descripciones de texto usando técnicas de similaridad semántica, sistemas de reglas u otras técnicas de procesamiento del lenguaje.

Se proporcionarán al alumno scripts previos de similaridad semántica como punto de partida (en Python).