



UNIVERSITAT OBERTA DE CATALUNYA (UOC)
MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS (*Data Science*)

TRABAJO FINAL DE MÁSTER

ÁREA: NATURAL LANGUAGE PROCESSING AND VISUAL ANALYTICS
DATA MINING, GRAPHS AND NATURAL LANGUAGE PROCESSING

Desarrollo de un RAG avanzado para la obtención de información sobre convocatorias de ayudas a empresas

Autor: José Luis Rodríguez Andreu

Tutor: Diego Calvo Barreno

Profesor: Josep Anton Mir Tutusaus

Barcelona, 28 de febrero de 2025



Esta obra está sujeta a una licencia de Reconocimiento - NoComercial - SinObraDerivada
3.0 España de Creative Commons.

FICHA DEL TRABAJO FINAL

Título del trabajo:	Desarrollo de un RAG avanzado para la obtención de informaci
Nombre del autor:	José Luis Rodríguez Andreu
Nombre del colaborador/a docente:	Diego Calvo Barreno
Nombre del PRA:	Josep Anton Mir Tutusaus
Fecha de entrega (mm/aaaa):	06/2025
Titulación o programa:	Máster Universitario en Ciencia de Datos
Área del Trabajo Final:	Trabajo Fin de Máster
Idioma del trabajo:	Español
Palabras clave	LLM, RAG, AI

Dedicatoria/Cita

Breves palabras de dedicatoria y/o una cita.

Agradecimientos

Si se considera oportuno, mencionar a las personas, empresas o instituciones que hayan contribuido en la realización de este proyecto.

Abstract

Texto con la síntesis del proyecto, esto es, un texto en el cual se explica de manera concisa la definición del proyecto/problema abordado, sus objetivos/métodos de resolución, y los resultados y conclusiones (no puede ser una lista, sino un texto continuo redactado de manera estructurada). Si es necesario poner una referencia en este texto, ésta será anotada a pie de la misma página. En este apartado se puede usar un lenguaje más literario y coloquial que para el resto del documento.

El Abstract se escribirá por duplicado. Una de las versiones tiene que ser **obligatoriamente en inglés**. La otra versión tiene que estar escrita en catalán o español. En caso de no escribir el resto del documento en inglés, será necesario escribir la segunda versión del Abstract en el idioma utilizado para el resto de la memoria. La palabra Abstract se cambiará por “**Resum**” o “**Resumen**” en la versión catalana y española, respectivamente.

Extensión recomendada: 250 palabras máximo.

Como escribir un buen Abstract (en inglés):

<http://www.ece.cmu.edu/~koopman/essays/abstract.html>

Palabras clave: Keywords del trabajo separadas por comas. Por ejemplo para este documento podrían ser Modelo, Pauta, Plantilla, Memoria, Trabajo de Final de Grado/Máster

Resumen

Texto con la síntesis del proyecto, esto es, un texto en el cual se explica de manera concisa la definición del proyecto/problema abordado, sus objetivos/métodos de resolución, y los resultados y conclusiones (no puede ser una lista, sino un texto continuo redactado de manera estructurada). Si es necesario poner una referencia en este texto, ésta será anotada a pie de la misma página. En este apartado se puede usar un lenguaje más literario y coloquial que para el resto del documento.

El Abstract se escribirá por duplicado. Una de las versiones tiene que ser **obligatoriamente en inglés**. La otra versión tiene que estar escrita en catalán o español. En caso de no escribir el resto del documento en inglés, será necesario escribir la segunda versión del Abstract en el idioma utilizado para el resto de la memoria. La palabra Abstract se cambiará por “**Resum**” o “**Resumen**” en la versión catalana y española, respectivamente.

Extensión recomendada: 250 palabras máximo.

Como escribir un buen Abstract (en inglés):

<http://www.ece.cmu.edu/~koopman/essays/abstract.html>

Palabras clave: Keywords del trabajo separadas por comas. Por ejemplo para este documento podrían ser Modelo, Pauta, Plantilla, Memoria, Trabajo de Final de Grado/Máster

Índice general

Abstract	IX
Resumen	XI
Índice	XIII
Lista de Figuras	XV
Lista de Tablas	1
1. Introducción	3
1. Descripción general del problema	3
1.1. Ejemplo de subsection	3
2. Introducción v2	3
2.1. Contexto y motivación	5
2.2. Objetivos	5
2.3. Sostenibilidad, diversidad y desafíos ético/sociales	5
2.4. Enfoque y metodología	6
2.5. Planificación	6
2.6. Resumen de los productos del proyecto	6
2.7. Breve descripción de los demás capítulos del informe	7
Bibliografía	8

Índice de figuras

1.1. Pie de la imagen.	4
1.2. Ejemplo de figura. Será indexada en la “Lista de Figuras”.	5

Índice de cuadros

1.1. <i>Vertex refinement queries</i>	5
---	---

Capítulo 1

Introducción

1. Descripción general del problema

En la actualidad, los procesos de minería de datos requieren grandes cantidades de datos, que en muchas ocasiones contienen información personal y privada de usuarios o personas. Aunque se realicen procesos básicos de anonimización sobre los datos, es decir, eliminación de los nombres u otros identificadores clave, existen multitud de técnicas de re-identificación que permiten volver a identificar a un usuario dentro de este conjunto de datos. En la Figura [1.1](#) se presenta un mapa donde es posible contextualizar los procesos de anonimización y re-identificación dentro de un proceso de minería de datos.

1.1. Ejemplo de subsection

Aunque se han realizado importantes avances en preservación de la privacidad en publicación de datos, tales como el modelo k -anonymity [\[1\]](#).

Un ejemplo de pseudo-código se puede encontrar en el Código [1](#)

Un ejemplo de tabla se puede ver en la Tabla [1.1](#)

2. Introducción v2

Esta plantilla pretende ser una guía para los estudiantes. Esta plantilla se puede adaptar a las necesidades específicas de cada proyecto si el supervisor del proyecto está de acuerdo con los cambios.



Figura 1.1: Pie de la imagen.

Algorithm 1 Pseudocódigo del algoritmo *Random Switch*

Entrada: El grafo original G y el porcentaje de anonimización p que se desea aplicar.

Salida: El grafo G anonimizado.

```

num = round( $G.num\_edges()$  *  $p$ )
i = 0
while i < num do
     $e_1 = G.random\_edge()$ 
     $e_2 = G.random\_edge()$ 
    new_e1 = ( $e_1.origen, e_2.origen$ )
    new_e2 = ( $e_1.destino, e_2.destino$ )
    if ! $G.exist(new\_e_1)$  and ! $G.exist(new\_e_2)$  then
         $G.add\_edge(new\_e_1)$ 
         $G.add\_edge(new\_e_2)$ 
         $G.delete\_edge(e_1)$ 
         $G.delete\_edge(e_2)$ 
        i = i + 1
    end if
end while
return G

```

Node ID	\mathcal{H}_0	\mathcal{H}_1	\mathcal{H}_2
Alice	ϵ	1	$\{4\}$
Bob	ϵ	4	$\{1, 1, 4, 4\}$
Carol	ϵ	1	$\{4\}$
Dave	ϵ	4	$\{2, 4, 4, 4\}$
Ed	ϵ	4	$\{2, 4, 4, 4\}$
Fred	ϵ	2	$\{4, 4\}$
Greg	ϵ	4	$\{2, 2, 4, 4\}$
Harry	ϵ	2	$\{4, 4\}$

Cuadro 1.1: *Vertex refinement queries.*

Figura 1.2: Ejemplo de figura. Será indexada en la “Lista de Figuras”.

2.1. Contexto y motivación

Punto de partida del proyecto (¿Cuál es el problema que necesita ser resuelto? ¿Por qué es un tema relevante? ¿Cómo se está resolviendo el problema actualmente?) y descripción de la contribución (¿Qué resultado se espera obtener?)

2.2. Objetivos

Listado de los objetivos a alcanzar en este proyecto.

2.3. Sostenibilidad, diversidad y desafíos ético/sociales

Esta sección debe evaluar el impacto positivo/negativo del proyecto en las siguientes dimensiones. No es necesario alcanzar un impacto positivo en todas las dimensiones, pero es necesario considerar y discutir si existe un impacto desde el inicio del proyecto.

Sostenibilidad En el desarrollo del proyecto o durante todo su ciclo de vida (por ejemplo,

despliegue, retiro), ¿tiene el resultado de este proyecto un impacto en la sostenibilidad y/o huella ecológica (consumo/ahorro de energía/recursos, desperdicio, contaminación, agotamiento de materias primas)? ¿Está afectado por leyes o regulaciones sobre este asunto? Considerando otra perspectiva, ¿afecta a alguno de los Objetivos de Desarrollo Sostenible (ODS) relacionados con estas dimensiones? Si no tiene ningún impacto, ya sea positivo o negativo, debe explicar cómo llegó a esta conclusión y justificar su respuesta.

Comportamiento ético y responsabilidad social ¿Es el resultado del proyecto demasiado técnico para tener algún impacto positivo/negativo en aspectos éticos/sociales? ¿Tiene un impacto en leyes/regulaciones (datos, privacidad, trabajo, propiedad intelectual, seguridad personal, ...)? ¿Se adhiere a los principios deontológicos de la profesión? ¿Pone en peligro/mejora/empeora algún puesto de trabajo? Si no tiene ningún impacto, ya sea positivo o negativo, debe explicar cómo llegó a esta conclusión y justificar su respuesta.

Diversidad, género y derechos humanos ¿Es el resultado de este proyecto tan técnico que no tiene impacto positivo/negativo en términos de género, diversidad o derechos humanos? ¿Y en leyes/regulaciones? ¿Y en términos de accesibilidad, discapacidad, ergonomía y/o seguridad de la información? Si no tiene ningún impacto, ya sea positivo o negativo, debe explicar cómo llegó a esta conclusión y justificar su respuesta.

2.4. Enfoque y metodología

Describa las estrategias potenciales para desarrollar este proyecto y explique la estrategia seleccionada. Discuta por qué la estrategia seleccionada es la más adecuada para alcanzar los objetivos del proyecto.

2.5. Planificación

Descripción de los recursos necesarios para desarrollar el proyecto, las tareas a realizar y una planificación temporal de cada tarea utilizando un diagrama de Gantt o equivalente. Esta planificación debe definir los hitos que se completarán en cada Prueba de Evaluación Continua (PEC).

2.6. Resumen de los productos del proyecto

No es necesario describir cada producto en detalle: esto se hará en los capítulos restantes del proyecto.

2.7. Breve descripción de los demás capítulos del informe

Breve descripción de los contenidos de cada capítulo y su relación con el resto del proyecto.

Bibliografía

- [1] Latanya Sweeney. k-anonymity: a model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10:557–570, October 2002.