



# MÁSTER EN INGENIERÍA EN SISTEMAS DE DECISIÓN

Curso Académico 2018/2019

Trabajo Fin de Máster

## Marco de trabajo para evaluar la relevancia de los artículos en el dominio científico

Autor : Adrián Alonso Barriuso

Tutor : Dr. Alberto Fernández Isabel



*Dedicado a  
mi familia, pareja, amigos y a todos los que me aguantan, en el buen sentido.*



# **Agradecimientos**



# Resumen





# Summary



# Índice general

<b>1. Introducción</b>	<b>13</b>
1.1. Contexto . . . . .	13
1.1.1. Dominio de aplicación . . . . .	13
1.2. Objetivos . . . . .	13
1.2.1. Objetivo General . . . . .	13
1.2.2. Objetivos específicos . . . . .	13
1.3. Estructura de la memoria . . . . .	13
<b>2. Estado del arte</b>	<b>15</b>
2.1. Algoritmos de reputación . . . . .	15
2.2. Obtención de relevancias . . . . .	15
<b>3. Propuesta</b>	<b>17</b>
3.1. Arquitectura general . . . . .	17
3.2. Recolección y preparación de datos . . . . .	17
3.2.1. Descarga y limpieza de datos . . . . .	17
3.2.2. Cálculo de reputaciones de artículos . . . . .	18
3.3. Gestión del conocimiento . . . . .	19
3.3.1. Obtención de matriz de términos por documentos . . . . .	20
3.3.2. Creación del lexicón . . . . .	20
3.3.3. Visualización . . . . .	22
3.3.4. Evaluación de nuevos artículos . . . . .	22
<b>4. Desarrollo, experimentos y resultados</b>	<b>25</b>
4.1. Planificación temporal . . . . .	25

4.2. Fase de descarga y preparación de datos . . . . .	25
4.3. Fase de procesamiento de lenguaje natural . . . . .	29
4.4. Fase de creación del lexicon . . . . .	31
4.5. Experimentos . . . . .	33
<b>5. Conclusiones</b>	<b>37</b>

# Índice de figuras

3.1. Arquitectura general . . . . .	18
3.2. Nube de palabras del lexicon . . . . .	22
3.3. Histograma de relevancias . . . . .	23
4.1. Diagrama de Gantt del desarrollo . . . . .	25
4.2. Ejemplo de metadatos de un artículo en Semantic Scholar . . . . .	26
4.3. Ejemplo de metadatos de un autor en Semantic Scholar . . . . .	28
4.4. Histograma de reputaciones de artículos . . . . .	29
4.5. Pipeline de Spacy . . . . .	30
4.6. Ejemplo de análisis con Spacy . . . . .	30
4.7. Ejemplo de nube de n-gramas . . . . .	32
4.8. Nube de términos con relevancias entre 0.31 y 0.33 . . . . .	33



# **Capítulo 1**

## **Introducción**

### **1.1. Contexto**

#### **1.1.1. Dominio de aplicación**

### **1.2. Objetivos**

#### **1.2.1. Objetivo General**

#### **1.2.2. Objetivos específicos**

- 
- 

### **1.3. Estructura de la memoria**

1. Estado del arte:
2. Propuesta:
3. Experimentos y resultados:
4. Conclusiones:





## **Capítulo 2**

### **Estado del arte**

#### **2.1. Algoritmos de reputación**

#### **2.2. Obtención de relevancias**



# **Capítulo 3**

## **Propuesta**

### **3.1. Arquitectura general**

### **3.2. Recolección y preparación de datos**

#### **3.2.1. Descarga y limpieza de datos**

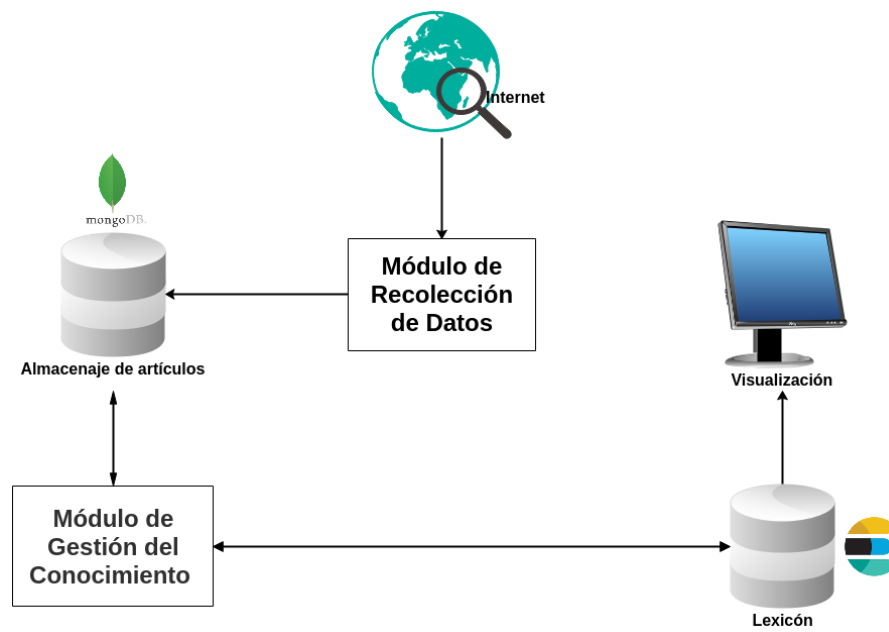


Figura 3.1: Arquitectura general

# Capítulo 4

## Experimentos y resultados

### 4.1. Planificación temporal

En la Figura 4.1 se puede ver un diagrama de Gantt[?] que refleja el tiempo empleado en cada una de las fases del proyecto.

### 4.2. Experimentos

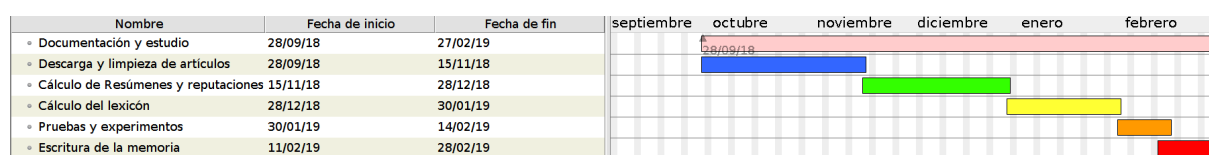


Figura 4.1: Diagrama de Gantt del desarrollo



## **Capítulo 5**

### **Conclusiones**

