

TRABAJO DE FIN DE MÁSTER

DESARROLLO DE UN PROTOTIPO DE MOTOR DE BÚSQUEDA QUE INCORPORE TÉCNICAS BIBLIOMÉTRICAS PARA MEJORAR LA RECUPERACIÓN

Máster Universitario en Ingeniería Informática

Autor: Aythami Estévez Olivas

Tutor: Juan Manuel Fernández Luna

ÍNDICE

- Objetivos
- Planificación
- Contexto
- Análisis
- Diseño
- Desarrollo
- Demostración
- Conclusiones y trabajos futuros

OBJETIVOS

OBJETIVO PRINCIPAL

Desarrollar un sistema de recuperación de información que incorpore medidas bibliométricas para mejorar la recuperación clásica

OTROS OBJETIVOS

- Descomponer el sistema RI: búsqueda autores y artículos
- Desarrollar sistema usable

PLANIFICACIÓN

TEMPORAL

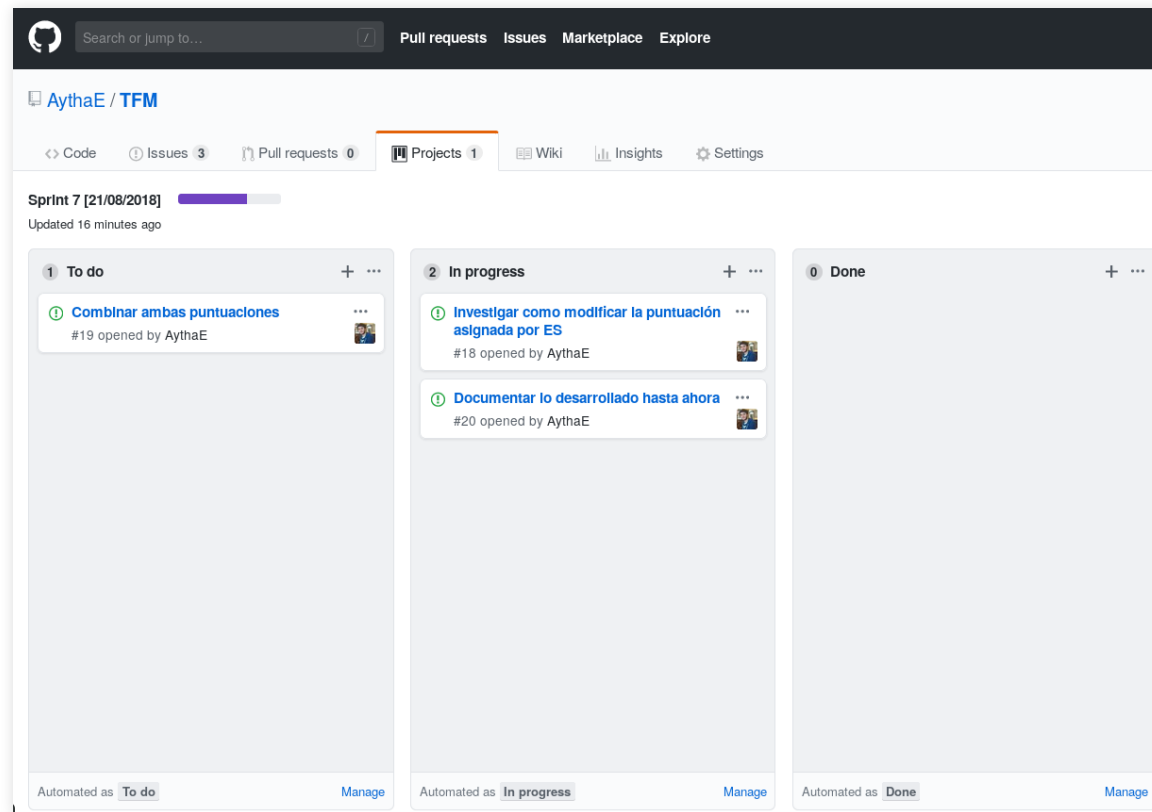
Tarea	Duración
Investigación	8 semanas
Obtención de datos	2 semanas
Procesado de datos	2 semanas
Búsqueda básica	2 semanas
Búsqueda con bibliometría	4 semanas
Refinamiento	1 semana

COSTE

Elemento	Coste
Recursos humanos	4800€
Hardware para el desarrollo	55,55€
Servidor	440€
Software	0€
Gastos editoriales	0€
TOTAL	5295,55€

METODOLOGÍA

Metodología ágil basada en SCRUM → *Sprints*

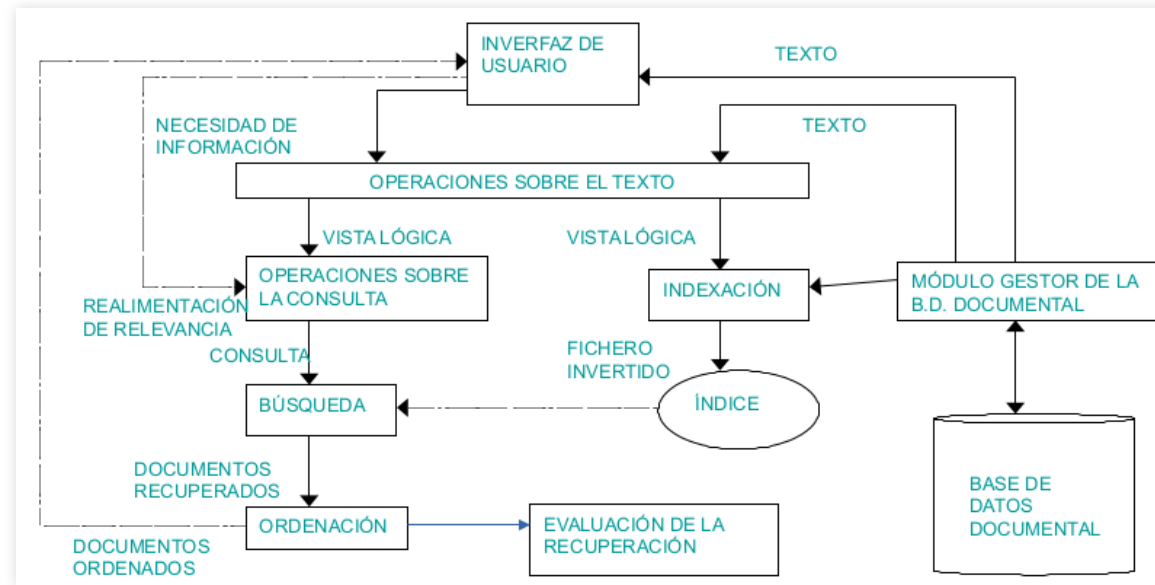


CONTEXTO

RECUPERACIÓN DE INFORMACIÓN: CONCEPTOS

- Disciplina que trata de modelar, diseñar e implementar sistemas capaces de proporcionar acceso basado en contenidos
- Relevancia y similitud
- Modelos: Booleano, Vectorial y Probabilístico

RECUPERACIÓN DE INFORMACIÓN: COMPONENTES SISTEMA



BIBLIOMETRÍA

- Análisis estadístico de publicaciones escritas. Se emplea para ofrecer análisis cuantitativo de la literatura científica.
- Número de citas, Índice H...

ANÁLISIS

ENFOQUE

Modelo clásico de RI con reordenación *a priori* de resultados en función a medidas directas y ordenación *a posteriori* utilizando un grafo de citación

HISTORIAS DE USUARIO

- Seleccionar el método de ordenación a priori de los resultados de búsqueda
- Seleccionar el método de ordenación a posteriori de los resultados
- Realizar búsquedas de autores
- Realizar búsquedas de artículos
- Desplegar una vista detallada de un artículo
- Desplegar una vista detallada de un autor

DISEÑO

MODELO DE DATOS

Author
+ugr_id: integer +scopus_id: string +num_docs: integer +ugr_cites: integer +ugr_hindex: integer +ugr_cites5: integer +ugr_hindex5: integer +scopus_cites: integer +scopus_hindex: integer +scopus_url: string +gscholar_url: string +ugr_url: string +full_name +nick_name: string +speciality: string +investigation_group: string
+save(): void +load(scopus_id: string): Author

Abstract
+scopus_id: string +authors: dictionary +cites: integer +title: string +date: date +doi: string +keywords: list +publication_name: string +publisher: string +abstract: string +references: list +subject_areas: list
+save(): void +load(scopus_id: string): Abstract

ARQUITECTURA INICIAL



DESARROLLO

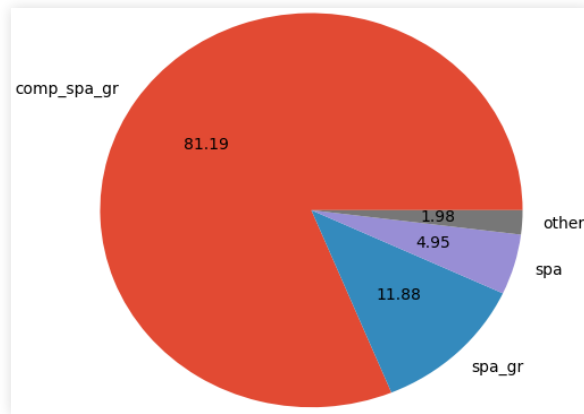
INVESTIGACIÓN

- Primeros 4 *Sprints* de investigación:
 - Recuperación de información
 - Bibliometría
 - Estado del arte
 - Pruebas con motores de búsqueda y conexiones APIs

OBTENCIÓN DE DATOS

- Ranking UGRinvestiga: 214 autores
- Scopus: 202 autores

PROCESAMIENTO DE DATOS



- Limpieza de autores
- 164 Autores
- 891 Artículos
- 742 Referencias

INDEXACIÓN

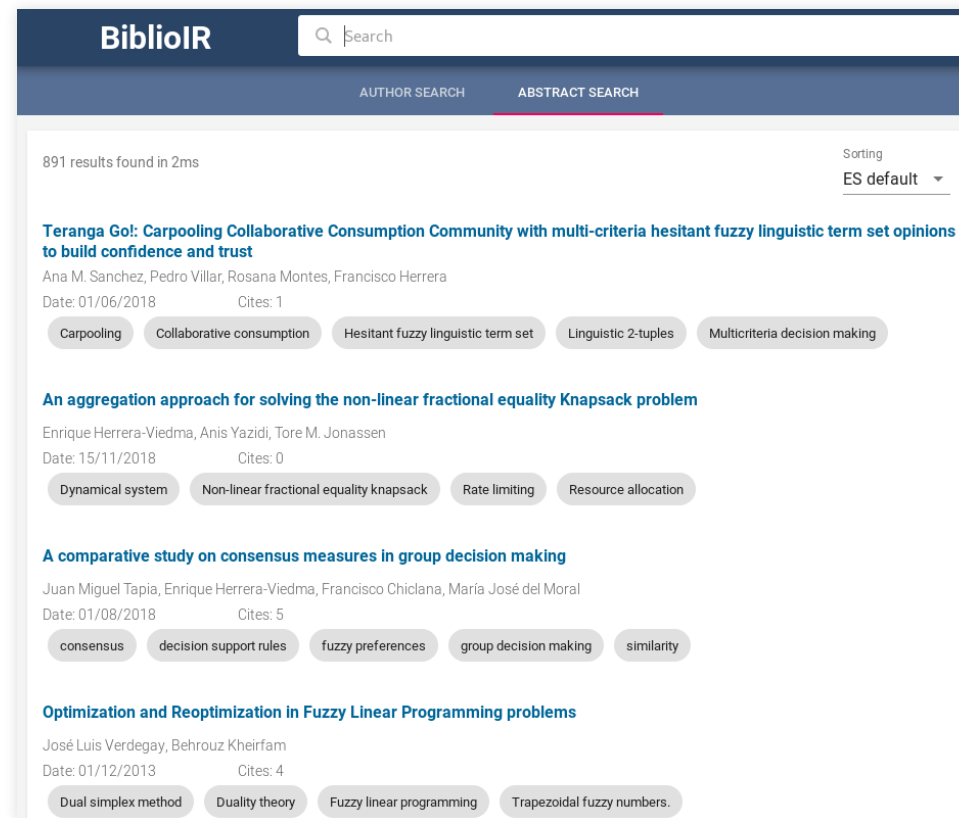
Subconjunto de
campos útiles para la
recuperación

Campo	Tipo
<i>full_name</i>	<i>text</i>
<i>nick_name</i>	<i>text</i>
<i>scopus_cites</i>	<i>integer</i>
<i>scopus_hindex</i>	<i>integer</i>
<i>speciality</i>	<i>text</i>
<i>num_docs</i>	<i>integer</i>
<i>investigation_group</i>	<i>text</i>
<i>ugr_hindex</i>	<i>integer</i>
<i>ugr_cites</i>	<i>integer</i>
<i>ugr_hindex5</i>	<i>integer</i>
<i>ugr_cites5</i>	<i>integer</i>

Campo	Tipo
<i>title</i>	<i>text</i>
<i>abstract</i>	<i>text</i>
<i>authors</i>	<i>text</i>
<i>subject_areas</i>	<i>text</i>
<i>keywords</i>	<i>text</i>
<i>publisher</i>	<i>text</i>
<i>publication_name</i>	<i>text</i>
<i>cites</i>	<i>integer</i>
<i>date</i>	<i>date</i>

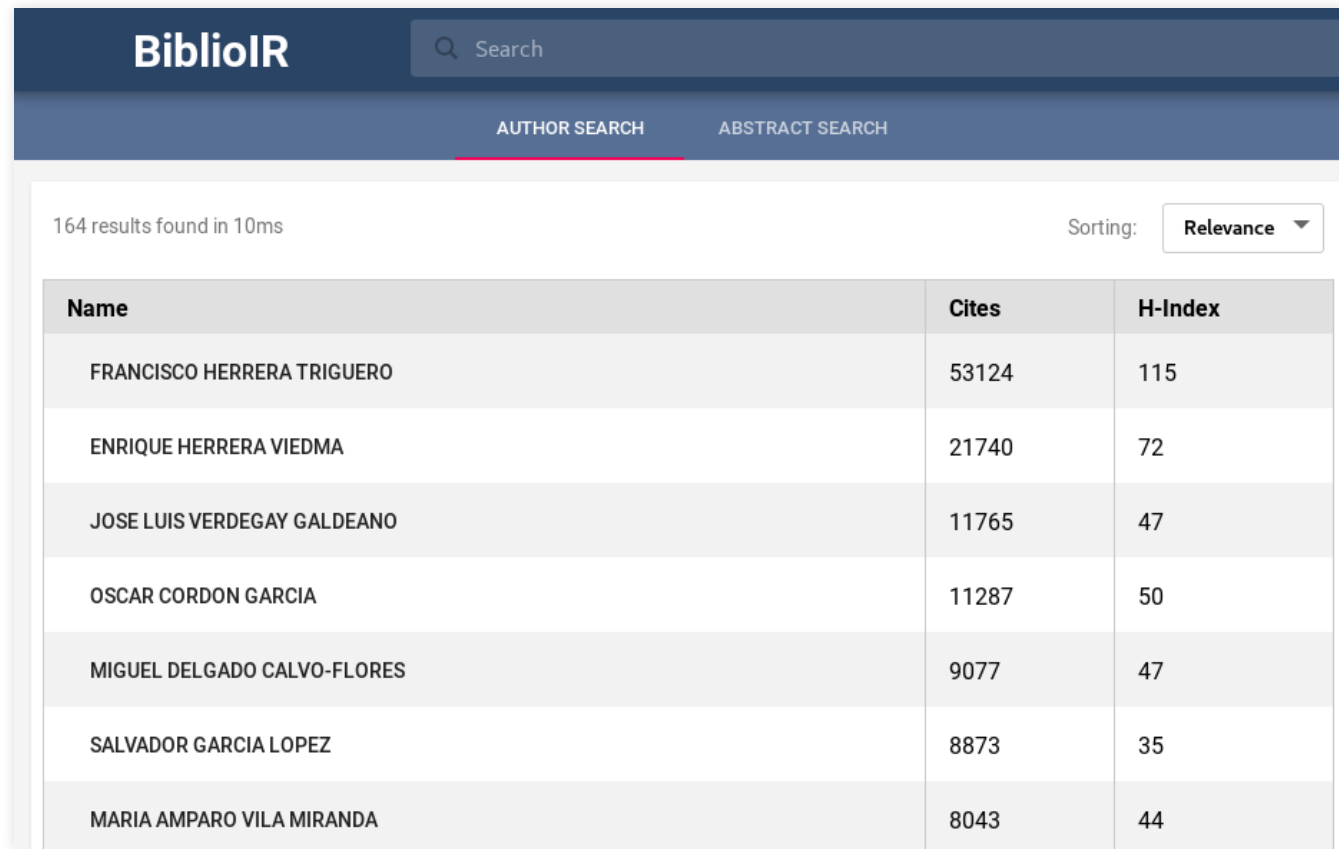
BÚSQUEDA CLÁSICA

Cliente web: *Searchkit, MaterialUI, ReactJS*



BÚSQUEDA CLÁSICA

Cliente web: *Searchkit, MaterialUI, ReactJS*



The screenshot displays the BiblioIR web application interface. At the top, there is a dark blue header with the 'BiblioIR' logo on the left and a search bar on the right. Below the header, a navigation bar contains two tabs: 'AUTHOR SEARCH' (which is active and underlined in red) and 'ABSTRACT SEARCH'. The main content area shows the results of an author search. It includes a status message '164 results found in 10ms' and a sorting dropdown menu set to 'Relevance'. Below this is a table with three columns: 'Name', 'Cites', and 'H-Index'. The table lists seven authors with their respective citation counts and H-index values.

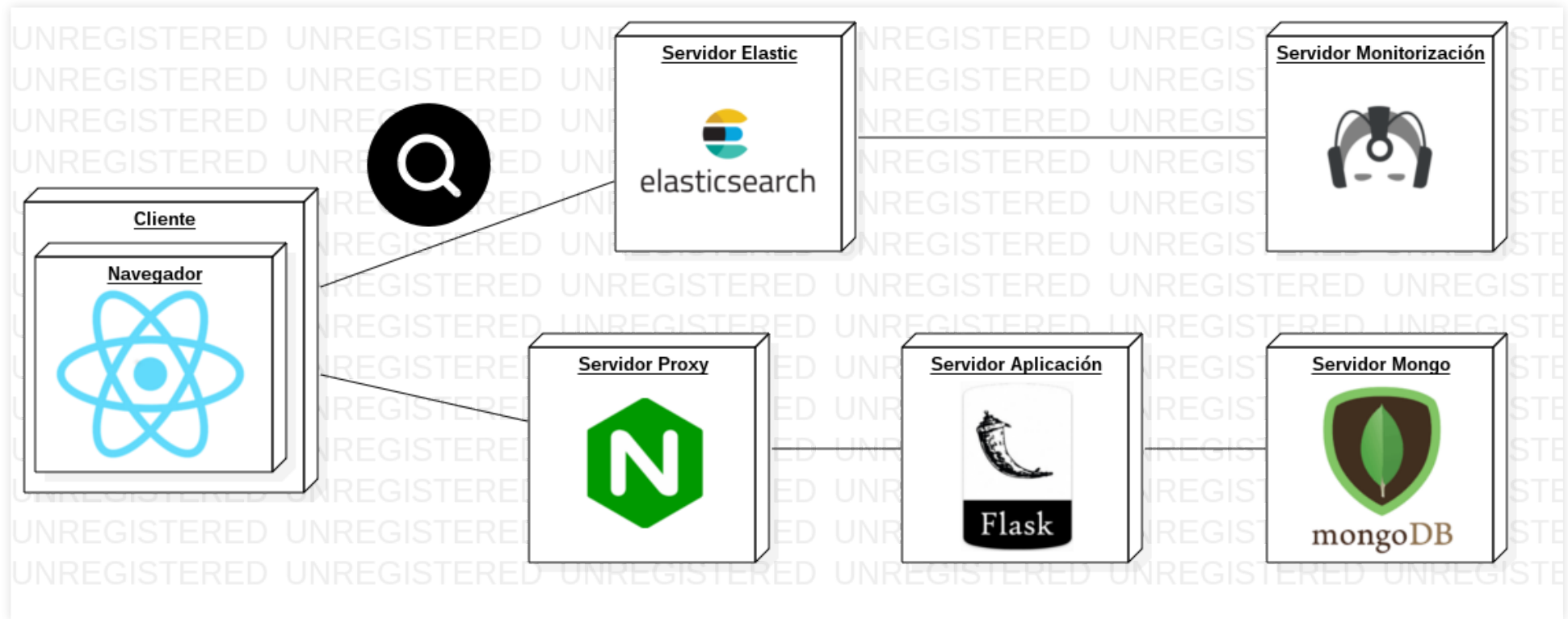
Name	Cites	H-Index
FRANCISCO HERRERA TRIGUERO	53124	115
ENRIQUE HERRERA VIEDMA	21740	72
JOSE LUIS VERDEGAY GALDEANO	11765	47
OSCAR CORDON GARCIA	11287	50
MIGUEL DELGADO CALVO-FLORES	9077	47
SALVADOR GARCIA LOPEZ	8873	35
MARIA AMPARO VILA MIRANDA	8043	44

APLICACIÓN MEDIDAS BIBLIOMÉTRICAS

Autores	Artículos
ES Default	
Nombre	Título
Citas	
Índice H	
CombMAX Citas	
CombSUM Citas	
CombMAX Índice H	
CombSUM Índice H	

- Combinación con score ES → **Normalización**
- Medidas → Citas, Índice h
- Algoritmos → **CombMAX, CombSUM**

ARQUITECTURA FINAL



DEMOSTRACIÓN

CONCLUSIONES Y TRABAJOS FUTUROS

CONCLUSIONES

- Proyecto interesante, he aprendido bastante y satisfecho curiosidad
- CombSUM varía más que CombMAX
- Ordenación *a posteriori* con citas: 742 citas pero solo 41,53% de los artículos con alguna cita

TRABAJOS FUTUROS

- Evaluación: utilizar encuestas a expertos
- Redes de citación para *relevance feedback*: necesarias más citas.

FIN

Gracias por su atención