ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA
INGENIERÍA DEL SOFTWARE

HERRAMIENTA PARA LA EXTRACCIÓN DE INFORMACIÓN SOBRE PUBLICACIONES CIENTÍFICAS

A TOOL FOR SCIENTIFIC PUBLICATION RETRIEVAL

Realizado por
NAHUEL VERDUGO REVIGLIONO
Tutorizado por
EDUARDO GUZMÁN DE LOS RISCOS
Departamento
LENGUAJES Y CIENCIAS DE LA COMPUTACIÓN

UNIVERSIDAD DE MÁLAGA
MÁLAGA, JULIO 2015

Fecha defensa:
El Secretario del Tribunal

Resumen: En este trabajo de fin de grado se presenta una herramienta que permite la extracción sobre las publicaciones científicas de diversos investigadores a partir de diferentes fuentes. La obtención de las mismas se realizara utilizando *crawlers* y extrayendo publicaciones de revistas indexadas en el JCR y conferencias del ranking CORE. Cada *crawler* puede ser ejecutado de manera particular o a través de un proceso principal. Las fuentes de información son: Web of Knowledge, Researchgate, Scopus y Google Scholar, pero dejando la posibilidad de añadir más implementando la interfaz apropiada. Toda la información recolectada se puede consultar a través de una interfaz web que proporcionara al usuario diferentes herramientas para visualizar estadísticas, la posibilidad de realizar comparaciones entre uno o varios investigadores, ver la relación que tiene con otros investigadores y centros, y ver las palabras claves más utilizadas en sus publicaciones. Este TFG ha sido financiado por el Vicerrectorado de Investigación y Transferencia de la Universidad de Málaga.

Palabras claves: web crawling, ranking CORE, index JCR, recuperación de la información, extracción de datos, web of knowledge, researchgate, scopus, google scholar

Abstract: In this final project, a tool that allows the extraction of scientific publications of several researchers from different sources is shown. The extraction of journal's publications from the JCR index and conferences from the CORE ranking is performed using *crawlers*. Each *crawler* can be executed in a particular way or through a parent process. The sources of information are: Web of Knowledge, ResearchGate, Scopus and Google Scholar, but leaving the possibility of adding a new one that implements an appropriate interface. All information collected is available through a web interface that provides the user several tools to visualize statistics, the possibility of making comparisons between one or more researchers, see the relationship that a researcher have with other research or centers and see the keywords most used in their publications. This final project has been funded by the Vice-rectorade for research and knowledge transfer of the University of Málaga.

Keywords: web crawling, data scraping, CORE ranking, JCR index, information retrieval, data extraction, web of knowledge, researchgate, scopus, google scholar

ÍNDICE

1 INTRODUCCIÓN

En esta sección se informará, de forma resumida, del contenido de esta memoria de proyecto titulada “Herramienta para la extracción de información sobre publicaciones científicas”.

1.1 OBJETIVOS

El objetivo principal de este proyecto es desarrollar una herramienta que permita extraer y visualizar información sobre las publicaciones científicas de diversos investigadores a través de sus perfiles y firmas en diversas plataformas, a partir de un listado de nombres de esos investigadores. La herramienta obtendrá sus publicaciones en revistas indexadas en el *Journal Citations Reports* (JCR) y publicaciones en conferencias indexadas en *CORE Conference Ranking*. A partir de ellas se podrá obtener diferentes estadísticas tales como:

- factor de impacto máximo conseguido por un investigador en un año,
- publicaciones organizadas por cuartiles o,
- firmas asociadas a un investigador,
- relación con otros investigadores o entidades,
- comparaciones entre investigadores, centros, departamentos y grupos de investigación,
- y palabras claves obtenidas de sus perfiles y publicaciones.

Esta tarea se realizaría utilizando diferentes *crawlers* programados para realizar búsquedas en sitios webs que contienen información sobre publicaciones e investigadores. Los sitios a buscar serán:

- **ResearchGate**: red social dirigida a personas que investigan en cualquier disciplina, donde pueden cargar sus propias publicaciones, además de información relacionada con su área de conocimiento y competencias. A principios de este año ya contaba con más de 6 millones de usuarios.
- **Scopus**: es una base de datos bibliográfica de resúmenes y citas de artículos de revistas científicas, editada por Elsevier. En su base de datos cubre un total de 18.000 títulos de 5.000 editores diferentes, incluyendo 16.500 revistas.
- **Web of Knowledge**: es un servicio en línea de información científica, suministrado por Institute for Scientific Information (ISI), grupo integrado en Thomson Reuters. El conjunto de bases de datos combinados incluye 23.000 revistas científicas y académicas, 110.000 publicaciones en conferencias y cubre más de 40 millones de documentos.
- **Google Scholar**: es un buscador de Google especializado en artículos de revistas científicas, enfocado en el mundo académico. Se estima que contiene en total 160 millones de documentos a fecha de mayo de 2014.
- Para facilitar la representación y acceso a los datos obtenidos se desarrollará una interfaz web y las tareas periódicas necesarias para que los mismos sean correctos y estén actualizados.

1.1.1 CRAWLERS, WEB CRAWLING Y DATA SCRAPING

Antes de comenzar, es necesario explicar en qué se basa este proyecto y por eso el lector tiene que estar familiarizado con tres conceptos fundamentales.

Primero, *data scraping* es la técnica de obtención de información de un recurso como puede ser ordenador, una base de datos, una API o una web. En este sentido su uso se centra en convertir datos sin estructura (como puede ser el código fuente de una web) en datos estructurados para su posterior almacenamiento y análisis.

Entre las técnicas más destacadas a la hora de realizar esta acción se encuentran: el uso de expresiones regulares, algoritmos de minería de datos, procesamiento (parseo) de documentos HTML, entre otras.

Por otra parte se encuentra el termino *web crawling*, que es un procedimiento cuyo objetivo es inspeccionar páginas webs de forma automática realizando una copia local de las mismas para un procesado posterior. Su principal cometido es buscar hiperenlaces para poder localizar nuevos sitios, con el objetivo de descargarlos a partir de una dirección URL base dada.

Para finalizar se encuentra el *crawler*, también llamado extractor a lo largo de este TFG, que es el programa que unifica estos dos conceptos. En nuestro caso, el cada *crawler* está diseñado para realizar la obtención y extracción de información de una forma concreta. En el siguiente diagrama se puede observar a groso modo como se realiza este proceso.

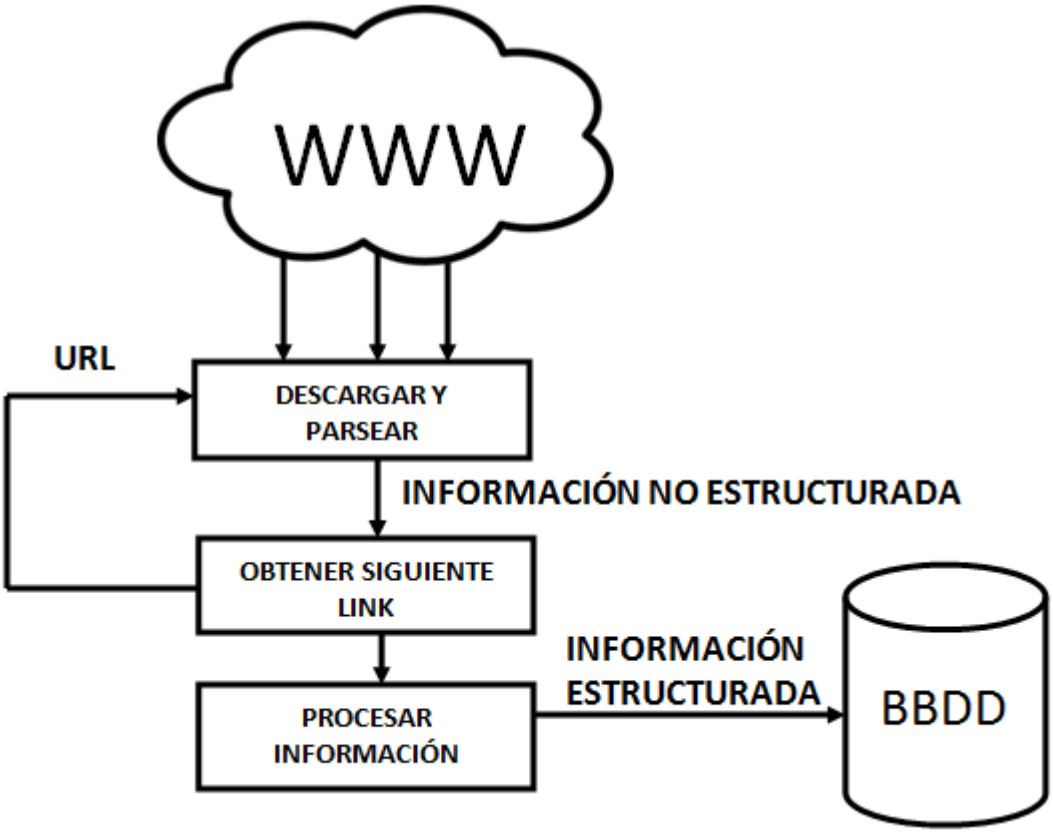


Ilustración 1.1 Diagrama del proceso de un crawler

1.1.2 NECESIDAD A CUBRIR

El desarrollo de este TFG forma parte del proyecto OGMIOS financiado por el Vicerrectorado de Investigación y Transferencias, órgano de la universidad encargado de la gestión de la investigación y de las infraestructuras que la sustentan, que conlleva a la calidad y excelencia de la producción de resultados de investigación de los distintos grupos y la trasferencia de los mismos a la sociedad.

A través de los resultados del proyecto OGMIOS se intentan conseguir los siguientes objetivos:

1. Facilitar la evaluación de la actividad investigadora del profesorado y de los investigadores contratados.
2. Formalizar y homogeneizar a través del sistema, la evaluación y reconocimiento de la actividad investigadora para todos los investigadores a través de criterios uniformes.
3. Ofrecer una herramienta para que los investigadores puedan tener al día su currículo de forma semiautomática.

Las herramientas de extracción, junto a la interfaz web y las tareas periódicas se encargarán de dar con la mayor exactitud los medios necesarios para poder llevar a cabo estas tareas.

1.1.3 TECNOLOGÍAS Y HERRAMIENTAS UTILIZADAS

En la siguiente tabla se describen las principales herramientas, plataformas de programación y librerías empleadas en este proyecto. A pesar de ser una amplia lista de tecnologías y herramientas, el objetivo ha sido intentar evitar reinventar la rueda y hacer uso de los proyectos con licencia de software libre en la mayoría de los casos.

PLATAFORMAS DE DESARROLLO

	**JAVA EE** Plataforma de desarrollo de Oracle basada en Java orientado a las arquitecturas web. Ha sido el lenguaje de programación más utilizado a lo largo de nuestra formación.

SOFTWARE

	**ECLIPSE IDE** IDE de desarrollo de código abierto, utilizado a lo largo de nuestra formación académica, y con una extensión especializada para desarrollar aplicaciones web en Java EE. Además de plugins que ayudan a agilizar el trabajo.

MAGICDRAW UML

Otra herramienta muy utilizada a lo largo de la carrera. Herramienta CASE utilizada para realizar los diagramas de uso, de clase o de secuencia usando la notación UML que aparecen a lo largo de esta memoria.

MAVEN

Sirve para la gestión y construcción de proyectos en Java, con Licencia Apache 2.0. Facilita el trabajo mediante el uso de un archivo XML de configuración para describir dependencias, tanto de otros módulos como de componentes externos, y como construir el proyecto a base de objetivos predefinidos.

BASE DE DATOS

MONGO DB

El moderno sistema de base de datos NoSQL de código abierto, que guarda los documentos de forma dinámica en formato JSON. Con características que lo hacen apropiado para el sistema como escalabilidad horizontal, flexibilidad en los modelos de datos a la hora de modificar los esquemas, posibilidad de índices secundarios y una comunidad de desarrolladores que crece cada día. Licencia GNU AGPL v3.0.

FRAMEWORK DE SEGURIDAD

SPRING SECURITY SAML

Es un framework para Java EE que proporciona autenticación y autorización, además de otras características de seguridad. Otro proyecto bajo la Licencia Apache. Simple de configurar e implementar sin necesidad de modificar nada en el proyecto.

LIBRERÍAS JAVA

QUARTZ

Librería abierta con Licencia Apache 2.0, integra fácilmente la posibilidad de crear tareas programadas sin importar el número y simplificando el desarrollo a la utilización de una interfaz para cada tarea.

JASPERREPORTS

Especializada en la creación de informes, permite entregar contenido enriquecido en múltiples formatos, con licencia libre GNU.

JSOUP

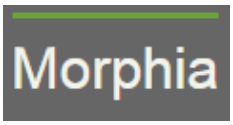
Librería perfecta en cuanto a extracción y manipulación de documentos HTML se refiere. Licencia MIT. Licencia Apache 2.0.

LOG4J

Herramienta muy extendida para escribir mensajes de registro. Permite múltiples configuraciones y organizar los mensajes en función de su importancia.

CRAWLER4J

Ofrece una interfaz simple para realizar *crawling* sobre la web. Se pueden configurar en pocos minutos crawlers multihebra. Con Licencia Apache 2.0.



MORPHIA

Librería que encarga de crear un envoltorio sobre el driver oficial de MongoDB para Java. Abstrayendo la complejidad de su uso y prestando herramientas tan interesantes como implementar persistencia en nuestra BBDD proporcionándonos las interfaces necesarias. Licencia Apache 2.0.

DESARROLLO WEB

JQUERY

Es una librería de JavaScript que proporciona numerosas herramientas, desde manipular el árbol DOM[1] hasta la interacción con AJAX. Con Licencia GPL y MIT.

BOOTSTRAP

Framework que contiene diferentes componentes webs y otros elementos de diseños combinando HTML y CSS, además de extensiones de JavaScript opcionales. Licencia Apache 2.0

[1] Document Object Model – API para representar documentos HTML y XML.

1.2 CONTENIDO DE LA MEMORIA EN CAPÍTULOS

La organización de la memoria se ha estructurado de tal forma que se asemeje a las etapas seguidas en cualquier desarrollo de software, permitiendo al lector observar la evolución de manera iterativa del planteamiento expuesto en la introducción.

1.2.1 TECNOLOGÍAS UTILIZADAS

En esta sección de la documentación se explicará el por qué de la selección y uso que se le da a algunas de las tecnologías ya nombradas.

Entre las explicadas se encontrarán Maven, Mongo DB y Morphia, Spring Security SAML, JSOUP, Quartz y Bootstrap por ser a las que más tiempo se le han dedicado para su integración en el proyecto y por ser las bases del desarrollo.

1.2.2 ANÁLISIS DE REQUISITOS

Durante esta primera fase se describirán con más detalles los requisitos que se obtienen a partir de las diferentes reuniones mantenidas con el tutor del proyecto. Se agruparán según su función dentro del proyecto que puede ser la de EXTRACTOR, APLICACIÓN WEB o TAREAS PERIÓDICAS.

1.2.3 DISEÑO

De cada requisito funcional obtenido en la primera fase se documentará los casos de uso más significativos. En ellos se detallará los actores, escenarios y las post/pre condiciones que deben cumplirse. A partir de ellos será posible la definición del sistema y del diseño del diagrama de clases. Serán una pieza clave para la implementación.

1.2.4 IMPLEMENTACIÓN Y PRUEBAS

En la última fase del proceso, se hará un resumen de lo más destacado en cuanto a problemas encontrados, solución propuesta y herramientas utilizadas.

1.2.5 CONCLUSIONES

En esta sección habrá un resumen de lo aprendido durante todo el proyecto, haciendo una comparación entre los objetivos y los resultados conseguidos. Se darán las conclusiones obtenidas durante el proceso de desarrollo y también de las personales.

2 TECNOLOGÍAS UTILIZADAS

En esta sección se explicará mejor la elección y el uso de las tecnologías utilizadas que suponen un mayor beneficio para el proyecto.

2.1 MAVEN

Desarrollada por Apache Software Foundation, es una herramienta destinada a gestionar y construir proyectos de Java. Contiene una serie de ventajas a la hora de comenzar cualquier proyecto y simplemente realizando la configuración del mismo a través de un archivo XML, llamado POM (Project Object Model).

En él se especifica el software a construir, dependencias, componentes, orden de construcción entre otras. Sin contar que tiene objetivos definidos para realizar tareas como es el caso de compilación, empaquetado, pruebas, generación de documentación. Entre las más destacadas están:

- Gestión de dependencias: a través del repositorio central se pueden importar gran cantidad de librerías y utilidades, sin necesidad de guardar cada una de ellas y despreocupando al programador de las mismas a la hora de compilar.
- Estandariza la estructura de los directorios, todos los proyectos comparten una estructura similar que ayuda a mantener el orden de los mismo, también permite configurarlos de forma personalizada.
- Integrado perfectamente con la IDE con la que se ha trabajado.
- Aporta además una implementación de ciclo de vida del proyecto, siendo las principales *compile, test, package, install, deploy*.

2.2 SPRING SECURITY SAML

Framework enmarcado dentro de las soluciones ofrecidas por Spring, por lo que las ventajas que conlleva es su uso dentro de proyectos de referencia del sector Java. Solución madura, facilidad de configuración y parametrización e integración con los sistemas de autenticación como es SAML 2.0 y proveedores de identidad.

Era la solución óptima para poder realizar la autenticación de los usuarios a través de la iDUMA (sistema de autenticación de la UMA) y que era posible de utilizar a través del repositorio de Maven.

Para entender el funcionamiento de SAML 2.0 primero es necesario aclarar los tres roles que participan:

- Principal: es el usuario que solicita la petición.
- Proveedor de servicio (SP): nuestra aplicación.
- Proveedor de identidad (idP): donde se identifica al usuario y el mismo realiza la autenticación.

El caso típico es aquel en el que el rol principal solicita acceso un recurso del SP, este comprueba si ya está autentificado y, si no es así, solicita al idP una confirmación de identidad. El SP teniendo esa confirmación recibida toma la decisión oportuna acerca del acceso autorizado del usuario al recurso. Se puede observar en el siguiente diagrama.

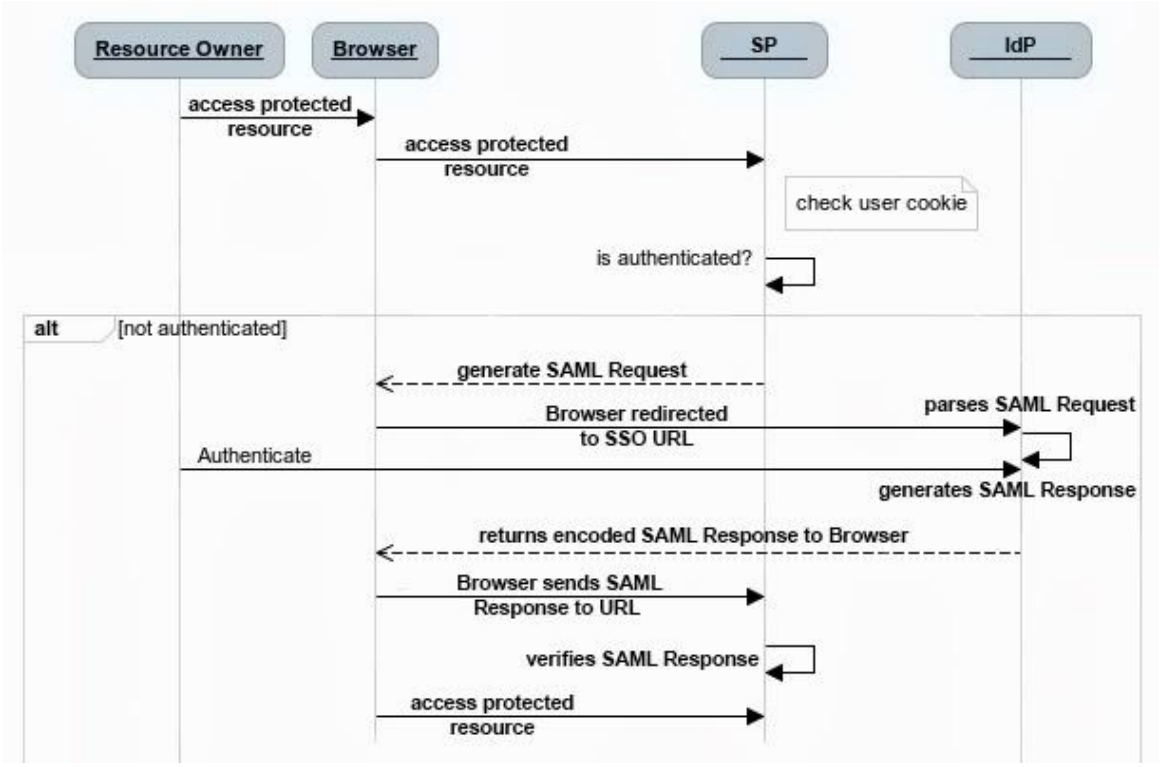


Ilustración 2.1: codeproject.com - How to integrate spring-oauth2 with spring-saml.

Además SAML en este caso requiere seguridad a nivel transporte mediante SSL y firma y encriptación XML para la seguridad del mensaje. Los certificados y llaves utilizados fueron provistos por el Servicio Central de Informática de la Universidad de Málaga.

2.3 MongoDB y Morphia

Los extractores, a medida que van recolectando datos, necesitan guardar esa información en algún sitio; los requisitos del proyecto, a pesar de estar bien definidos, no se sabe cómo podrían variar en un futuro, ya sea, la cantidad o el tipo de información recolectada. El tratamiento de la información y los informes que se quieren obtener de ellas podrían ser más complejos también. Por ese motivo se escogió un sistema base de datos NoSQL orientado a documentos, en este caso MongoDB.

En este sistema las tablas son suplantadas por colecciones y las filas por documentos. Tienen las ventajas de poder manejar enormes cantidades de datos, no generan cuellos de botella, escalamiento sencillo, rendimiento aceptable en máquinas con hardware simple.

Para el mapeo de los objetos de Java en MongoDB y la persistencia se ha utilizado la librería Morphia, que ofrece una API para realizar consultas a la BBDD, validación de campos, anotaciones y clases para poder implementar DAO (Data Access Object) sin demasiada complicación. Durante el desarrollo del proyecto Morphia llego a la versión 1.0 totalmente compatible con la versión 3 de MongoDB.

2.4 JSOUP

Esta librería, por su uso dentro del proyecto, se puede decir que es la base para la mayoría de los extractores. Se encarga perfectamente de obtener el código HTML (no lo que

se visualiza) de la página que se desea, pero aparte de esto cabe destacar las siguientes funcionalidades:

- Seleccionar DOM de forma similar a la utilizada en jQuery, incluyendo por CSS.
- Manipular los elementos HTML, atributos y textos.
- Posibilidad de obtener y utilizar cookies en las solicitudes.
- Añadir un agente de navegación a las solicitudes.

Estas ventajas junto a la facilidad de su utilización, convierten JSOUP en un elemento imprescindible en cualquier proyecto en el que se tengan que manipular archivos con código HTML.

2.5 QUARTZ

La librería encargada de la planificación y gestión de tareas, mejora y nos aparta de complicaciones de tener que utilizar la clase java.util.TimerTask. Las características que lo hacen destacado dentro del proyecto, tanto para el presente o como un futuro si se necesitara más funcionalidades, son:

- Flexibilidad para definir los instantes de ejecución, con una nomenclatura similar a la de cron de Linux.
- API completa.
- Valido para aplicaciones Java EE.
- Almacenamiento persistente.
- Si se requiere, balanceo de carga al trabajar en modo clúster.
- Historial de ejecuciones de tareas.

Su estructura se basa en tres componentes:

1. La **tarea** que implantará la interface *Job*.
2. El **trigger**, que se puede implementar o utilizar las implementaciones que vienen por defecto, se utilizará *org.cuartz.CronTrigger* que permite especificar instantes más concretos y utilizando expresiones de cron.
3. Y por último el **scheduler**, que es el que almacena y planifica las tareas en base a los trigger, realiza reintentos tras operaciones fallidas y gestiona el estado del sistema de planificación. Su configuración se realiza mediante un archivo llamado *quartz.properties*.

2.6 BOOTSTRAP

Por último, y por nombrar también algunas de las tecnologías utilizadas en el diseño de la interfaz, hay que destacar el uso de conjunto de herramientas que contiene plantillas de diseños, tipografías, botones, menús desplegables y de navegación y otra gran cantidad de elementos.

Hay que hacer especial mención en las siguientes utilidades frecuentemente recurridas a lo largo del diseño:

- Grid System de 12 columnas que puede ser responsivo si se requiere.
- Galería de iconos que representan multitud de acciones y son gratuitos.

- Utilización de *Modal* para desplegar diálogos y ventanas emergentes dentro de la pantalla.
- *Tooltips* para mostrar información de los botones y sus acciones.

Bootstrap es fundamental para los desarrolladores que poseemos pocos conocimientos en diseño, ya que es un conjunto de buenas prácticas, nos ayuda a adaptar una forma de trabajo, en lo que se refiere a tratar CSS + HTML5, y existen multitud de webs donde se pueden visualizar ejemplos realizados con esta tecnología y adaptar a nuestros proyectos.

3 ANÁLISIS DE REQUISITOS

3.1 GLOSARIO

JCR Es la publicación anual realizada por la empresa *Thomson Scientific* donde se evalúa el impacto y relevancia de las revistas científicas de ciencias aplicadas y ciencias sociales.

CORE Conference Ranking Es un ranking de conferencias realizado por *Computing Research and Education Association of Australasia*, una asociación de departamentos de universidades australianas y neozelandesas de informática.

iDUMA Servicio de identidad de la Universidad de Málaga, donde se realiza la autenticación centralizada.

Entidad Hace referencia a cualquier institución que tenga investigadores a su cargo, como puede ser la UMA.

Publicación Un texto científico hecho público a través de una revista científica o una conferencia.

Fuente Nombre de la revista o la conferencia donde se realiza una publicación.

3.2 ANÁLISIS

Antes de comenzar a listar los requisitos que se nos piden, los actores que tenemos en el sistema y los diversos casos de uso que vamos a tener que desarrollar, es conveniente introducir al lector en la situación que nos encontramos al comenzar el proyecto.

Debido a la diversidad de fuentes (públicas o privadas) donde un investigador puede dar a conocer o compartir sus publicaciones, estar al tanto de todo lo que ha publicado se hace una tarea ardua que quita tiempo debido a los registros en webs, buscadores, comparación de títulos, comprobación de propiedad y otras tareas. A esto hay que añadir que no todas las fuentes muestran la misma información.

Continuando esta tarea, se necesita poder cuantificar la cantidad y calidad del trabajo realizado por un investigador a lo largo de su carrera, no existe un modelo claro a seguir a la hora de hacerla. Una primera aproximación sería utilizar un factor que sea común a las publicaciones y que permitan hacer esta calificación. Existen dos listas a nivel mundial, índice ***JCR*** para las revistas científicas y ***CORE Conference Ranking*** para las conferencias internacionales que permiten conseguirla.

A partir de este escenario surgen los *EXTRACTORES,* procesos que se encargarán de realizar esta tarea de asociar publicaciones a investigadores de forma automática.

El proceso anterior tiene una gran complejidad que reside en la desambiguación de los nombres utilizados por los investigadores ya sea por su nombre en sí o por la firma utilizada por el mismo en las publicaciones. Para poder llevar un mejor control sobre el problema que se plantea, es necesaria una interfaz donde un investigador o responsables de la aplicación realicen tareas de mantenimiento para comprar que los datos en ella son correctos, tales como, comprobar firmas, perfiles y publicaciones obtenidas.

Esta interfaz a su vez servirá de apoyo para poder realizar otras funciones que puedan ser de utilidad a la hora de calificar al investigador dando las herramientas necesarias para tener una visión global del mismo o comparándolo con otro o un grupo de ellos. Esta aplicación tiene que poder ser accedida por múltiples plataformas, por lo que se opta que sea una *APLICACIÓN WEB*.

Para facilitar las tareas de mantenimiento y poder obtener mejores resultados a medida que se avanza en el tiempo, se proponen *PROCESOS PERIÓDICOS* que, a partir de las diferentes acciones realizadas por un usuario de la aplicación, ejecutará un proceso.

Este breve análisis se irá detallando y entendiendo mejor a medida que se avance dentro de esta sección.

3.3 REQUISITOS

3.3.1 REQUISITOS FUNCIONALES

Aquí se van a listar los requisitos funcionales, que son los que determinan la funcionalidad de la aplicación y definen las acciones que pueden realizar los diferentes roles de la misma. Las listas están divididas según la categoría, un identificador, un nombre y una descripción para poder llevar una trazabilidad. Vamos a comenzar listando los requisitos de los extractores:

Categoría	ID	Nombre	Descripción
EXTRACTOR	EX1	Extraer perfiles	El usuario podrá extraer todos los perfiles asociados a una institución.
	EX2	Extraer publicaciones de perfiles	El usuario podrá extraer de un perfil en un rango de años dado.
	EX3	Extraer publicaciones de firma	El usuario podrá extraer a partir de una firma en un rango de años dado.
	EX4	Extraer información investigador	El usuario podrá extraer información complementaria de un investigador.

A parte de los listados, también serán necesarios dos extractores auxiliares para poder generar la base de datos de revistas y conferencias con la que trabajaremos.

ID	Nombre	Descripción
EXR1	Extraer JCR	El usuario podrá extraer toda la información sobre revistas del JCR.
EXC1	Extraer CORE	El usuario podrá extraer toda la información sobre conferencias del CORE.

Listado de requisitos correspondientes a la aplicación web:

Categoría	ID	Nombre	Descripción
APLICACIÓN WEB	AW1	Realizar login	Los usuarios deberán autentificarse en la aplicación.
	AW2	Realizar logout	Los usuarios podrán salir de la aplicación.
	AW3	RU Perfil	Los usuarios podrán ver y editar la información de su perfil, como es el caso de firmas y perfiles, siendo investigadores o de todos los investigadores siendo administradores.

AW4	Recibir firma recomendada	La aplicación recomendará firmas a los usuarios, siempre que sea posible, en caso de no tener ninguna asociada o pocas.
AW5	Ver estadística	Los usuarios verán las estadísticas para un perfil dando la posibilidad de filtrar por años.
AW6	Listar investigadores	Los administradores podrán listar y buscar investigadores por centros/departamentos/grupos.
AW7	Realizar búsqueda	Los administradores podrán realizar búsquedas por grupos/departamentos/centros.
AW8	Comparar búsquedas	Los administradores en las estadísticas podrán comparar una búsqueda con investigador/grupos/departamentos/centros /entidad.
AW9	Imprimir resumen	Los administradores podrán extraer un archivo con los datos de una búsqueda.
AW10	Ver relaciones	Los usuarios en las estadísticas podrán ver relaciones con otras entidades e investigadores.
AW11	Listar publicaciones	Los usuarios en las estadísticas podrán listar todas las publicaciones en revistas o en conferencias asociadas.
AW12	Ver publicaciones	Los usuarios al listar las publicaciones de unas estadísticas podrán ver las mismas, las fuentes de obtención, filtrar, buscar, ordenar y ver un índice de fiabilidad.
AW13	Validar publicación	Los usuarios podrán validar publicaciones.
AW14	Rechazar publicación	Los usuarios podrán rechazar publicaciones.
AW15	Cambiar fuente	Los usuarios podrán cambiar la fuente de sus publicaciones.
AW16	Imprimir publicaciones	Los usuarios podrán imprimir un resumen de las publicaciones.

Y estos serían los últimos requisitos funcionales que corresponden con los procesos periódicos:

Categoría	ID	Nombre	Descripción
PROCESOS PERIÓDICOS	PP1	Programar validación	El sistema asociará las publicaciones validadas por los investigadores con ellos.
	PP2	Programar descarte	El sistema descartará las publicaciones rechazadas por los investigadores.
	PP3	Programar extracción	Por cada extractor, el sistema realizará una extracción de todas las publicaciones de todos los investigadores que tengan perfiles para ese extractor y pertenezcan a la entidad.
	PP4	Programar actualización perfil	Por cada extractor, el sistema actualizará los perfiles nuevos añadidos.
	PP5	Programar actualización búsquedas	El sistema actualizará todas las búsquedas de investigadores.
	PP6	Programar actualización fuente	El sistema actualizará los cambios de revistas o conferencias de las publicaciones.

3.3.2 REQUISITOS NO FUNCIONALES

Estos requisitos determinan restricciones o condiciones sobre las que se ejecuta o desarrolla el sistema. La lista de los mismos se hará en la tabla con una categoría del requisito, identificador y descripción.

Categoría	ID	Descripción
Seguridad	RNF1	La autenticación se deberá llevar a cabo utilizando la iDUMA.
Seguridad	RNF2	No se permitirá que usuarios ajenos al sistema tengan acceso.
Usabilidad	RNF3	El sistema informará del resultado de realizar cualquier acción a través de mensajes por pantalla.
Usabilidad	RNF4	El sistema controlará los datos introducidos en los campos para evitar datos erróneos.
Interfaz	RNF5	Los extractores se podrán utilizar a través de una consola.
Accesibilidad	RNF6	El sistema podrá utilizarse en las versiones más recientes de los navegadores web Chrome, Firefox, Internet Explorer y Safari.
Concurrencia	RNF7	El sistema tendrá que manejar y controlar los accesos concurrentes a la base de datos.
Hardware	RNF8	El sistema tiene que ser capaz de utilizarse en cualquier plataforma Java 1.7 o superior con acceso a internet y una BBDD MongoDB.

3.4 ACTORES

Describimos a continuación brevemente en qué consiste cada actor/rol del sistema.

- **Usuario:** Este actor representa cualquier persona que tenga acceso a la utilización de los extractores.
- **No autentificado:** Dentro de la aplicación web es todo aquel usuario que intenta acceder a la misma sin estar autentificado.
- **Anónimo:** Es el rol que tiene un usuario que no se le ha asignado ningún rol y que por lo tanto sus permisos no le dejan hacer prácticamente nada dentro de la aplicación, como es el caso del inicio de sesión por parte de un alumno.
- **Investigador:** Es un rol dentro de la aplicación web, y puede ser compatible con otros roles. Es cualquier investigador que su DNI coincide con los datos de un investigador perteneciente a la los que están guardados en la BBDD.
- **Administrador:** Es otro rol de la aplicación web, representa a las personas que pueden realizar tareas de mantenimiento o que trabajen directamente sobre los datos extraídos de la aplicación.
- **Sistema:** Representa la máquina donde estará corriendo el planificador de tareas.

3.5 DIAGRAMA DE CASOS DE USO

Aquí se representan los diferentes diagramas de cada uso para cada categoría.

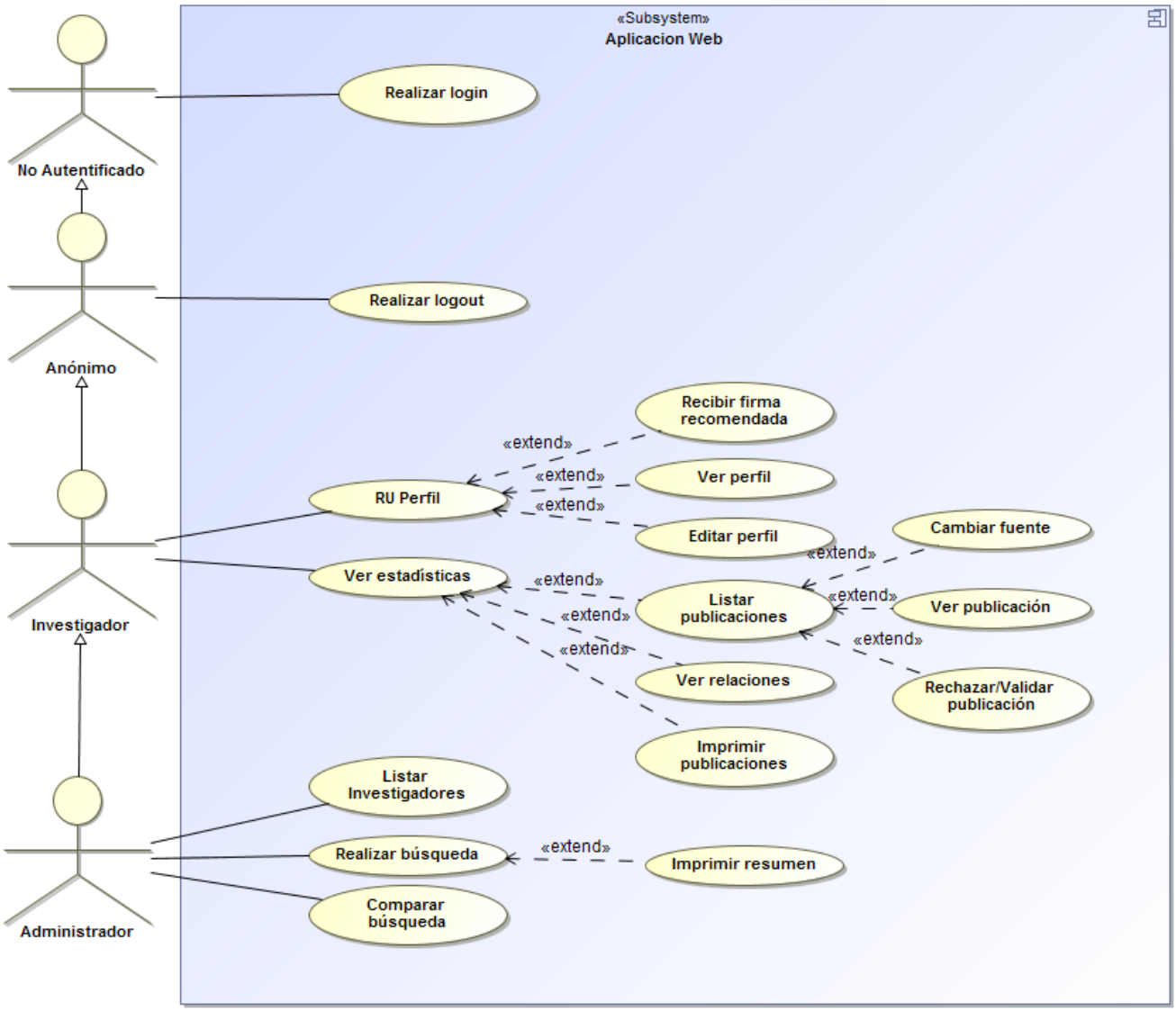


Ilustración 3.2: Diagrama caso de uso de la aplicación web.

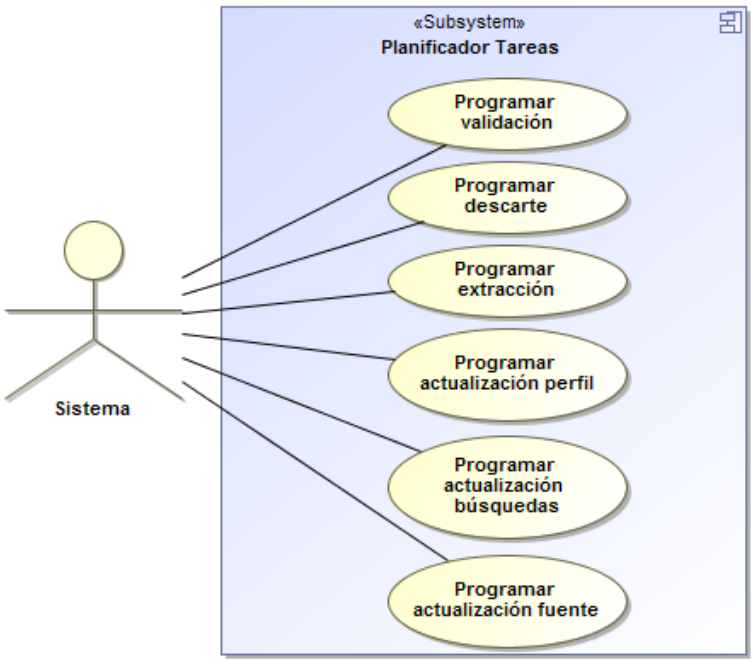


Ilustración 3.1: Diagrama caso de uso del planificador de tareas.

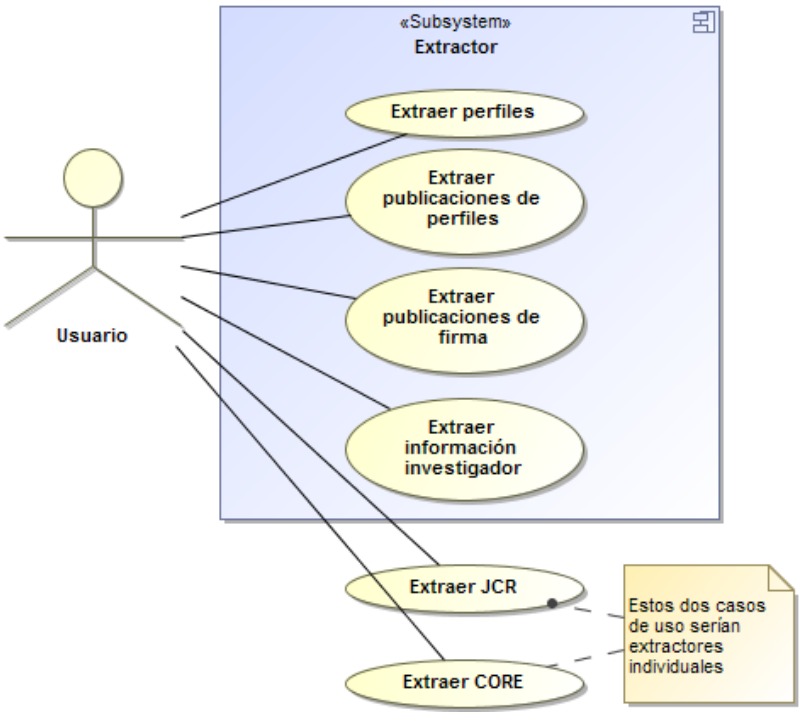


Ilustración 3.3: Diagrama caso de uso de los extractores.

4 DISEÑO

4.1 DESCRIPCIÓN DE LOS CASOS DE USO

En las siguientes tablas se recoge de forma detallada los casos de uso, no están expuestos todos los escenarios pero si los más destacados.

ID	U1
Caso de uso	Realizar login
Actores	No autentificado
Descripción	Acceso típico a la aplicación.
Precondiciones	No haber iniciado sesión en la aplicación.
Escenario principal	1. El usuario accede a cualquier url perteneciente a la aplicación. 2. Es usuario es redirigido a la web de autenticación de iDUMA. 3. El usuario introduce sus datos correctos en el formulario. 4. El usuario es redirigido a la url que solicitaba.
Escenario alternativo	3. El usuario introduce sus datos de forma incorrecta. 4. El usuario es redirigido otra vez a la web de autenticación de iDUMA y esta muestra un mensaje de error en los datos introducidos.
Requisito asociado	AW1

ID	U2
Caso de uso	Realizar logout
Actores	Anónimo
Descripción	Cierre de sesión de la aplicación
Precondiciones	Estar autenticado en la aplicación.
Escenario principal	1. El usuario accede al menú superior derecho de la aplicación y pulsa cerrar sesión. 2. El sistema cierra sesión en iDUMA y dentro de la aplicación. 3. El usuario es redirigido a una pantalla que muestra que ha cerrado sesión.
Escenario alternativo	2. Ocurre un error en la comunicación con la iDUMA y no se cierra sesión correctamente. 3. El usuario es redirigido a una ventana que muestra que ha ocurrido un error.
Requisito asociado	AW2

ID	U3
Caso de uso	Ver perfil
Actores	Investigador
Descripción	Ver información de su perfil de investigador.
Precondiciones	Estar autenticado en la aplicación con rol investigador.
Escenario principal	1. El usuario accede a la pestaña del menú “Mi perfil”. 2. Visualiza la información de su perfil.
Escenario alternativo	-
Requisito asociado	AW3

ID	U4
Caso de uso	Editar perfil
Actores	Investigador
Descripción	El investigador desea borrar una firma que tiene asociada a su perfil o el administrador visualizando el perfil de un investigador.
Precondiciones	Estar autenticado en la aplicación. Rol investigador haber realizado U3 y rol administrador U13.
Escenario principal	1. El usuario pulsa sobre la “x” que aparece al lado de su firma para borrarla. 2. El sistema informa que la firma ha sido borrada correctamente.
Escenario alternativo	2. El sistema informa que ha ocurrido un error al borrar su firma.
Requisito asociado	AW3

ID	U5
Caso de uso	Recibir firma recomendada
Actores	Investigador
Descripción	El investigador quiere conocer si hay alguna firma que esté relacionada con su nombre dentro de su departamento.
Precondiciones	Estar autenticado en la aplicación. Rol investigador haber realizado U3 y rol administrador U13.
Escenario principal	1. El usuario realiza clic sobre “Firmas recomendadas”. 2. La aplicación despliega un diálogo con posibles firmas dispuestas para seleccionar. 3. El usuario selecciona las deseadas y pulsa “Añadir”. 4. El sistema informa a través de una notificación por pantalla que se han añadido correctamente.
Escenario alternativo	2. La aplicación despliega un diálogo sin ninguna firma posible. 3. El usuario pulsa “Cancelar” y vuelve a su perfil.
Requisito asociado	AW4

ID	U6
Caso de uso	Ver estadísticas
Actores	Investigador
Descripción	Ver las estadísticas asociadas.
Precondiciones	Estar autenticado en la aplicación. Rol investigador haber realizado U3 y rol administrador U13 o U14.
Escenario principal	1. Pulsar el botón de “Ver estadísticas” al final del perfil. 2. El usuario es dirigido a una vista de estadísticas donde puede ver un resumen de sus publicaciones en revista y conferencias.
Escenario alternativo	-
Requisito asociado	AW5

ID	U7
Caso de uso	Listar publicaciones
Actores	Investigador
Descripción	Listar las publicaciones que han encontrado los extractores.
Precondiciones	Estar autenticado en la aplicación. Haber realizado U6.
Escenario principal	1. El usuario accede a la pestaña del menú “Ver mis publicaciones en revista”. 2. El sistema carga al final de la página la lista de publicaciones.
Escenario alternativo	2. El sistema no carga nada, porque no tiene ninguna publicación.
Requisito asociado	AW11

ID	U8
Caso de uso	Ver relaciones
Actores	Investigador
Descripción	Ver información de su perfil de investigador.
Precondiciones	Estar autenticado en la aplicación. Haber realizado U6.
Escenario principal	1. El usuario accede a la pestaña del menú “Relaciones”. 2. El sistema carga al final de la página un grafo con las relaciones del investigador con otros investigadores y otras entidades. También 2 tablas por si no es lo suficientemente visible en el grafo.
Escenario alternativo	-
Requisito asociado	AW10

ID	U9
Caso de uso	Imprimir publicaciones
Actores	Investigador

Descripción	Descargar un documento con formato RTF o PDF con las publicaciones que aparecen al listar publicaciones.
Precondiciones	Estar autenticado en la aplicación. Haber realizado U6.
Escenario principal	1. El usuario pulsa el botón “Versión imprimible” 2. El sistema despliega un cuadro de diálogo de configuración. 3. El usuario selecciona el tipo de archivo que desea y pulsa imprimir. 4. El sistema genera un archivo del tipo seleccionado por el usuario para que este lo descargue.
Escenario alternativo	3. El usuario se arrepiente imprimir. 4. Cierra el cuadro de diálogo y vuelve a la vista de estadísticas.
Requisito asociado	AW16

ID	**U10**
Caso de uso	Cambiar fuente
Actores	Investigador
Descripción	Cambiar el nombre de la fuente de una revista o conferencia que no son correctos.
Precondiciones	Estar autenticado en la aplicación. Haber realizado U7.
Escenario principal	1. El usuario pulsa sobre el nombre de la revista. 2. El sistema despliega un cuadro de diálogo. 3. El usuario introduce el nombre apropiado para la revista. 4. Se despliega una lista de posibles revistas. 5. El usuario selecciona la que está buscando. 6. Se cierra el cuadro de diálogo y se notifica al usuario que se realizará el cambio.
Escenario alternativo	3. El usuario se arrepiente de cambiar el nombre. 4. Cierra el cuadro de diálogo y vuelve a la vista de estadísticas.
Requisito asociado	AW15

ID	**U11**
Caso de uso	Ver publicaciones
Actores	Investigador
Descripción	Ver información que no se muestra por defecto en la lista desplegada al ver las publicaciones.
Precondiciones	Estar autenticado en la aplicación. Haber realizado U7.
Escenario principal	1. El usuario pulsa sobre un icono con forma de lupa. 2. El sistema despliega un cuadro de diálogo con información adicional de la publicación.
Escenario alternativo	-
Requisito asociado	AW12

ID	U12
Caso de uso	Rechazar/Validar publicación
Actores	Investigador
Descripción	Rechazar o validar una publicación listada.
Precondiciones	Estar autenticado en la aplicación. Haber realizado U7.
Escenario principal	1. El usuario pulsa sobre el botón con un pulgar hacia arriba para validar la publicación. 2. El sistema informa de que se ha realizado correctamente.
Escenario alternativo	2. El sistema informa de que ha ocurrido un error al validar.
Requisito asociado	AW13,AW14

ID	U13
Caso de uso	Listar investigadores
Actores	Administrador
Descripción	Ver información de su perfil de investigador.
Precondiciones	Estar autenticado en la aplicación.
Escenario principal	1. El administrador accede a la pestaña del menú “Mi perfil”. 2. El sistema despliega una lista paginada de investigadores. 3. El administrador introduce un nombre para buscar. 4. El sistema despliega una lista paginada de investigadores que coinciden con el valor buscado.
Escenario alternativo	4. El sistema informa al usuario que no hay ningún nombre en el sistema que coincida con ese valor.
Requisito asociado	AW6

ID	U14
Caso de uso	Realizar búsqueda
Actores	Administrador
Descripción	Realizar búsqueda para ver estadísticas de un centro/grupo/departamento/entidad.
Precondiciones	Estar autenticado en la aplicación.
Escenario principal	1. El administrador accede a la pestaña del menú “Otras búsquedas”. 2. Selecciona un centro. 3. El sistema muestra el nombre del centro para indicar que es lo que se va a buscar. 4. El administrador pulsa el botón “Buscar”. 5. Es redirigido a las estadísticas del centro.
Escenario alternativo	-
Requisito asociado	AW7

ID	U15
Caso de uso	Comparar búsqueda
Actores	Administrador
Descripción	Realizar una comparación de un investigador/centro/grupo/departamento/entidad.
Precondiciones	Estar autenticado en la aplicación. Haber realizado U13 o U1
Escenario principal	1. El administrador accede a la opción del menú lateral “Comparar”. 2. El sistema abre un cuadro de diálogo donde el usuario puede seleccionar contra que comparar. 3. El administrador escribe el nombre de un investigador y pulsa comparar. 4. El sistema carga al final de la página una gráfica comparativa y dos tablas con los datos de ambas estadísticas.
Escenario alternativo	-
Requisito asociado	AW8

ID	U16
Caso de uso	Imprimir resumen
Actores	Administrador
Descripción	Exporta a CSV la lista de investigadores que se ha filtrado .
Precondiciones	Estar autenticado en la aplicación. Haber realizado U13.
Escenario principal	1. El administrador pulsa el botón “Exportar a CSV”. 2. El sistema abre un cuadro de diálogo donde el usuario puede seleccionar las obras a exportar y el rango de año. 3. El administrador selecciona el tipo de obra y el rango de años, solo si lo desea. 4. El sistema genera un archivo CSV con los datos que el usuario puede descargar.
Escenario alternativo	-
Requisito asociado	AW9

ID	U17
Caso de uso	Extraer perfiles
Actores	Usuario
Descripción	Extraer todos los perfiles asociados a una entidad y mostrarlos por pantalla.
Precondiciones	Tener acceso al ordenador donde se encuentra el extractor.
Escenario principal	1. El usuario ejecuta el programa con la opción “perfiles”. 2. El extractor imprime por pantalla todos los perfiles encontrados.

Escenario alternativo	2. Se produce un error durante la extracción de perfiles. 3. El sistema notifica por email al responsable de la aplicación de que ha ocurrido un error y crea una entrada con este en el log.
Requisito asociado	EX1

ID	U18
Caso de uso	Extraer publicaciones de perfiles
Actores	Usuario
Descripción	Extraer las publicaciones asociadas a un perfil a través del id.
Precondiciones	Tener acceso al ordenador donde se encuentra el extractor.
Escenario principal	1. El usuario ejecuta el programa con la opción “id” y un id válido. 2. El sistema lista todas las publicaciones asociadas al id.
Escenario alternativo	2. Se produce un error durante la extracción de publicaciones. 3. El sistema notifica por email al responsable de la aplicación de que ha ocurrido un error y crea una entrada con este en el log.
Requisito asociado	EX2

El caso U19 sería similar al 18, haría referencia a Extraer publicaciones de firma (EX3), pero utilizando la firma en lugar del id del perfil.

ID	U20
Caso de uso	Extraer información investigador
Actores	Usuario
Descripción	Ver información adicional de un investigador.
Precondiciones	Tener acceso al ordenador donde se encuentra el extractor.
Escenario principal	1. El usuario ejecuta el programa con la opción “investigador” y un DNI válido. 2. El sistema devuelve el investigador con la información encontrada.
Escenario alternativo	2. Se produce un error durante la extracción de información. 3. El sistema notifica por email al responsable de la aplicación de que ha ocurrido un error y crea una entrada con este en el log.
Requisito asociado	EX4

ID	U21
Caso de uso	Extraer JCR
Actores	Usuario

Descripción	Guardar en BBDD todas las revistas del índice JCR.
Precondiciones	Tener acceso al ordenador donde se encuentra el extractor.
Escenario principal	1. El usuario lanza ejecuta la aplicación. 2. El extractor guarda en BBDD todas las revistas del índice JCR.
Escenario alternativo	2. Se produce un error durante la extracción de perfiles. 3. El sistema notifica por email al responsable de la aplicación de que ha ocurrido un error y crea una entrada con este en el log.
Requisito asociado	EXR1

El caso U22 sería similar al 23, haría referencia a Extraer CORE (EXC1)

ID	U23
Caso de uso	Programar validación
Actores	Sistema
Descripción	Activar la tarea para validar obras.
Precondiciones	Realizarla cada 1 minuto, de lunes a viernes de 0 a 20 horas.
Escenario principal	1. El sistema creará una entrada en el log de que ha comenzado. 2. El sistema activará el proceso. 3. El sistema creará una entrada en el log de que ha terminado.
Escenario alternativo	2. El proceso falla durante su ejecución. 3. El sistema creará una entrada en el log de que ha fallado. 4. El sistema enviará un email al responsable de la aplicación con el fallo.
Requisito asociado	PP1

Por la similitud de los siguientes casos de usos al este último, estos serán descritos de manera abreviada indicando cuando serán programados.

- U24 - Programar descarte (PP2): Realizarla cada 1 minuto, de lunes a viernes de 0 a 20 horas.
- U25 - Programar extracción (PP3): Realizarla cada 1 vez cada dos meses dependiendo del extractor se hará en un mes par o impar y en la primera o tercera semana. Comenzando un sábado a las 0 horas.
- U26 - Programar actualización perfil (PP4): Realizarla cada X minutos desentendiendo del extractor y evitando que se ejecuten 2 al mismo tiempo para no cargar demasiado al servidor, de lunes a viernes de 0 a 20 horas.
- U27 - Programar actualización búsquedas (PP5): Realizarla de lunes a viernes a las 22 horas.
- U28 - Programar actualización fuente (PP6): Realizarla cada 1 minuto, de lunes a viernes entre las 20 y 22 horas.

Están organizadas de tal manera que se ejecuten de forma simultánea las que menos tardan para dosificar la carga en el servidor.

4.2 MODELO DE CLASES

Se muestran los atributos más significativos.

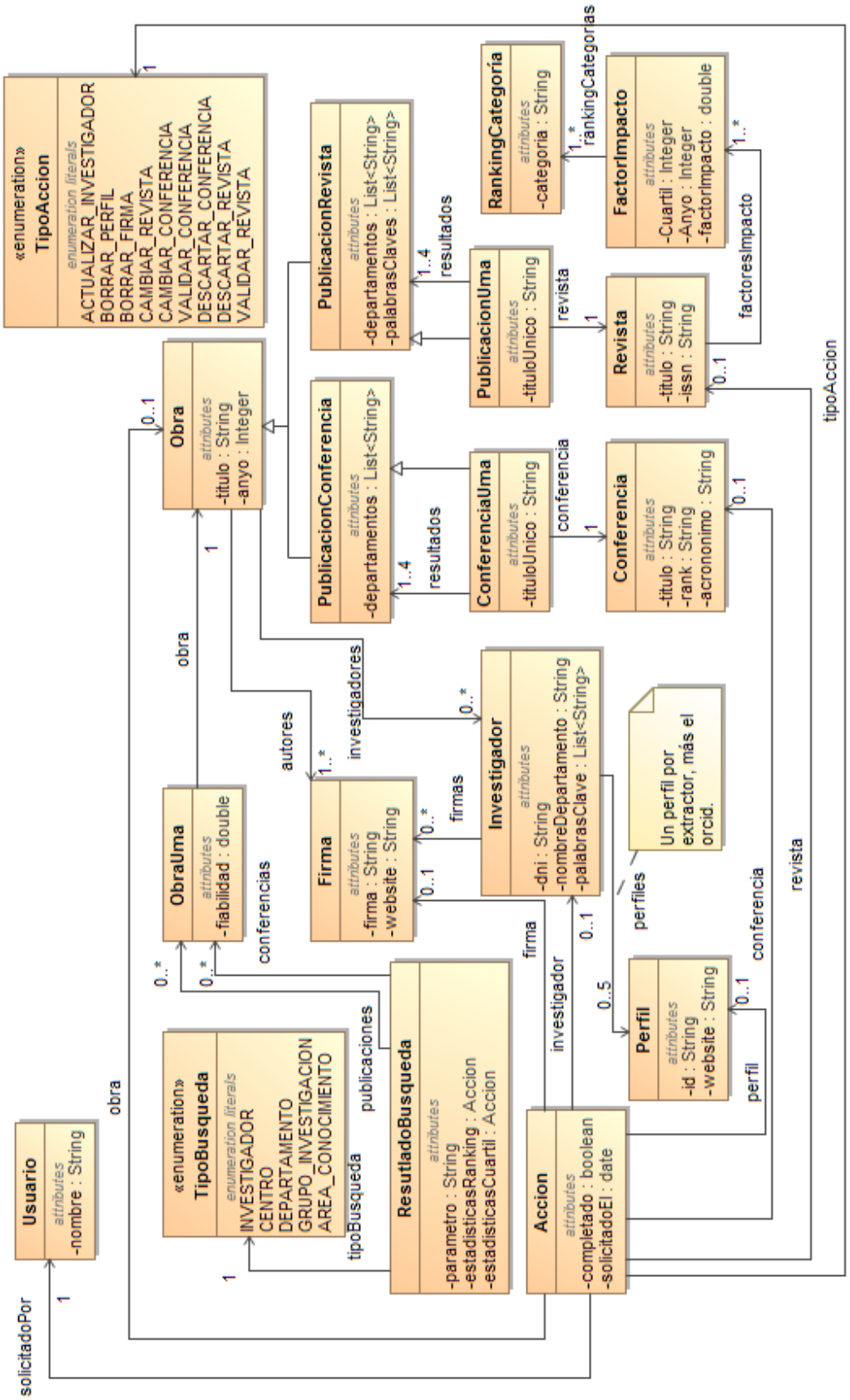


Ilustración 4.1 Diagrama de clases

4.3 MATRIZ DE TRAZABILIDAD

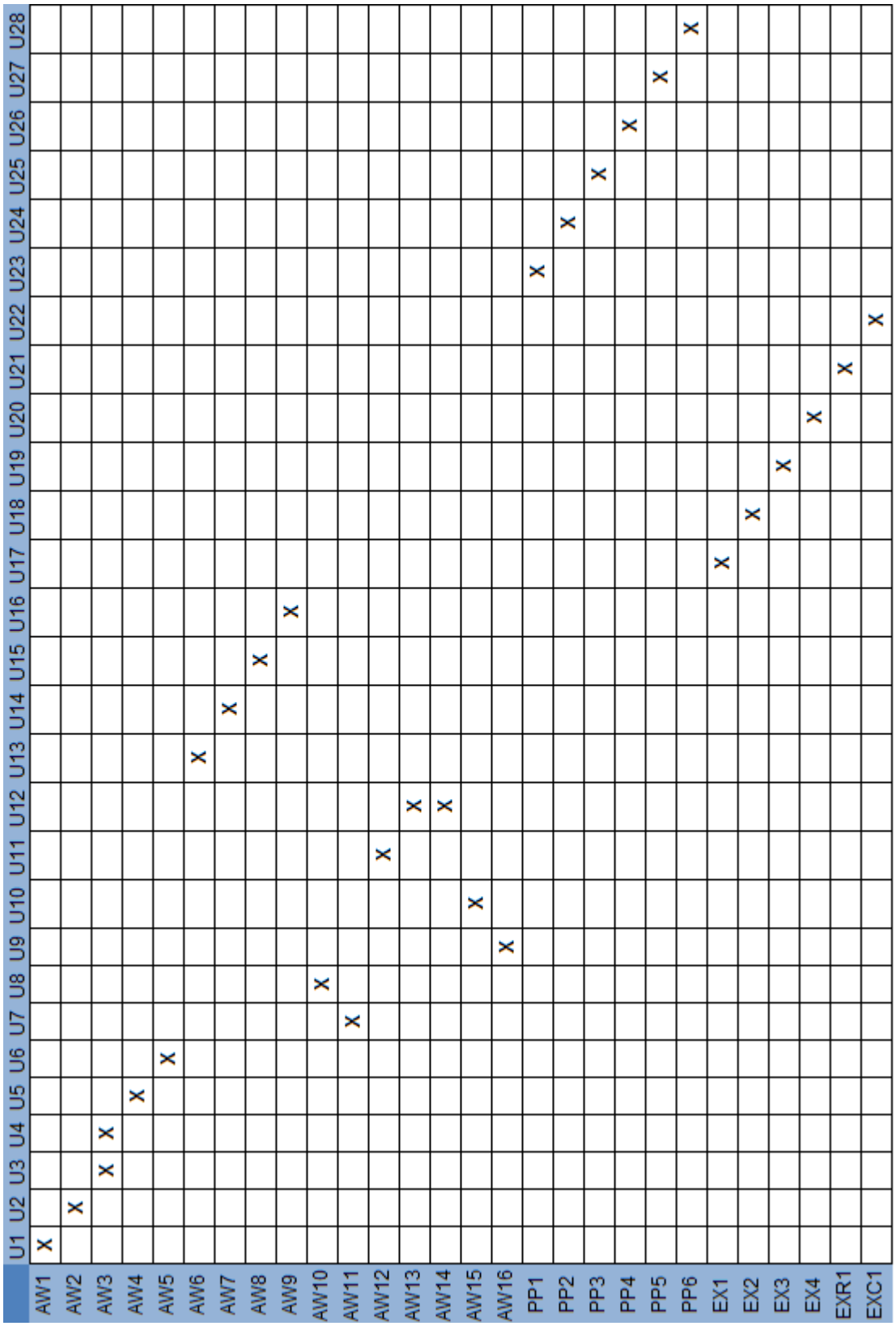


Ilustración 4.2 Matriz de trazabilidad

5 IMPLEMENTACIÓN

En esta sección se comentarán los problemas surgidos durante el desarrollo y las soluciones planteadas, tanto las que afectan al proyecto en general como a los extractores.

5.1 PROYECTO

5.1.1 ASOCIAR FIRMAS A NOMBRES

Uno de los grandes problemas encontrados al comienzo del proyecto fue la asociación de nombres de perfiles o firmas encontrados en publicaciones y perfiles a investigadores. Este reside básicamente en:

1. Los nombres completos no se suelen utilizar por su longitud. Por ejemplo "Juan Ignacio Lopez Perez" puede utilizar el nombre "Juan Lopez Perez".
2. En las firmas se tiende a abreviar el nombre, por ejemplo "Juan Lopez Perez" podría firmar como "J. Perez".
3. Se utiliza el formato "Apellidos, Nombre" en muchas ocasiones.

Es cierto que sin tener ninguna forma más para relacionar nombres, simplemente con dos cadenas de texto, es necesaria alguna herramienta o algoritmo que nos pueda dar un valor para poder calcular la similitud entre dos cadenas.

Para conseguir esto se optó por no reinventar la rueda y utilizar una solución ya probada *fuzzywuzzy-java* [2], una implementación que apoyándose en la *distancia de Levenshtein*[3] calcula la similitud de dos cadenas de texto sin tener en cuenta el orden, las mayúsculas y devolviendo un valor entre 0 y 1. A esta función se le realizaron cambios para que a la hora de calcular este valor tuviese en cuenta las tildes o caracteres como la "ñ", las comas y las abreviaturas en el primer nombre.

```
Similitud entre Juan Rodriguez Lopez y Juan Lopez: 1.0
Similitud entre Juan Rodriguez Lopez y Rodriguez Lopez, Juan : 1.0
Similitud entre Juan Rodriguez Lopez y J. Rodriguez Lopez : 0.84
Similitud entre Juan Rodriguez Lopez y Rodriguez Lopez, J. : 0.84
Similitud entre Juan Rodriguez Lopez y Rodriguez, J. : 0.75
Similitud entre Juan Rodriguez Lopez y Joan Lupez : 0.4
```

Ilustración 5.1 Resultados de ejecución de fuzzywuzzy-java

A partir de 0.7 se puede decir, a través de la experiencia y los test realizados, que es un resultado aceptable. Sin duda alguna fue uno de los problemas que más se tardó en resolver.

5.1.2 CONCURRENCIA EN MONGODB

Al comenzar las primeras pruebas de con los extractores trabajando de forma simultánea, mientras se actualizaban o modificaban investigadores y publicaciones Morphia comenzaba a dar errores de concurrencia al tratar de modificar dos objetos al mismo tiempo.

Esto se debe a que la librería como forma de protección utiliza un control de concurrencia optimista: cada entidad mapeada en MongoDB lleva una variable de tipo Long con la anotación @Version que aumenta cada vez que es modificada. Al guardar un objeto

[2] https://github.com/msubhash/fuzzywuzzy-java
[3] Algoritmo que mide el número de operaciones de inserción, borrado y sustitución para convertir una cadena de texto en otra.

comprueba que ninguna otra transacción haya modificado la entidad leída comparando esta variable. Si esto ocurre lanza una excepción del tipo: `ConcurrentModificationException.`

Para evitar que esto ocurra, a la hora de guardar objetos que puedan provocar esta excepción se utiliza un try/catch dentro de un do/while.

```
boolean failUpdate = true;
do {
    obj = dao.get(id);
    // modificaciones necesarias
    try {
        dao.save(obj);
        failUpdate = false;
    } catch (ConcurrentModificationException ex) {
        LOG.warn("Error ConcurrentModificationException controlado");
    }
} while (failUpdate);
```

Ilustración 5.2 Ejemplo de código para control de concurrencia

5.1.3 Compartir una conexión a MongoDB entre DAOs

Otro problema que se encontró durante el desarrollo del proyecto, fue el coste en tiempo y memoria de crear una conexión con MongoDB cada vez que un DAO quería acceder, crear, modificar o borrar una entidad.

```
public class MongoDB {
    private static final Logger LOG = Logger.getLogger(MongoDB.class.getName());
    private static final MongoDB INSTANCE = new MongoDB();
    private final Datastore datastore;
    public static final String DB = "ogmios";
    private MongoDB() {
        MongoClientOptions mongoOptions = MongoClientOptions.builder()
                .socketTimeout(60000)
                .connectTimeout(1200000)
                .maxWaitTime(60000)
                .build();
        MongoClient mongoClient;
        mongoClient = new MongoClient(new ServerAddress("127.0.0.1", 27017),
                mongoOptions);

        mongoClient.setWriteConcern(WriteConcern.SAFE);
        datastore = new Morphia().createDatastore(mongoClient, DB);
        datastore.ensureIndexes();
        LOG.info("Conexion a la base de datos '" + DB + "' inicializada.");
    }

    public static MongoDB instance() {
        return INSTANCE;
    }
    public Datastore getDatabase() {
        return datastore;
    }
}
```

Ilustración 5.3 Patrón singleton utilizado

Al principio cada DAO tenía su objeto `Datastore` que es así como Morphia se encarga de realizar la comunicación con el servidor de base de datos, entonces cada llamada a algún método implicaba la creación de un nuevo objeto.

La solución que se utilizó fue utilizar un patrón Singleton para que solo exista una instancia de la conexión y que solo exista un punto de acceso global a ella.

5.1.4 Obtener posibles firmas

En dos partes del proyecto era importante tener una función que permitiera conocer las posibles firmas que podrían pertenecer a un investigador.

La forma más óptima que se encontró fue realizar una consulta sobre mongo que devolviera todas las firmas de las publicaciones del departamento del investigador y que esta firma al menos contuviese el primer apellido del investigador.

Sobre las firmas encontradas se aplica el algoritmo de similitud de nombres, anteriormente comentado, y las que tengas un valor de similitud superior a 0.7 se consideran posibles firmas.

5.1.5 Fiabilidad de las obras encontradas

Al listar las publicaciones era necesario tener un índice de fiabilidad, el administrador podría no conocer el investigador al que está consultando sus estadísticas y no saber si se puede confiar de los datos que está viendo.

Por eso al guardar una búsqueda, cada lista de publicaciones o revista, se envuelve el objeto `Obra` en otro objeto llamado `ObraUma` que este tiene una variable que indica la fiabilidad de la misma.

El cálculo de la fiabilidad se realiza teniendo en cuenta si los coautores de la misma pertenecen al departamento del investigador, comparando las firmas de estos con las firmas de la obra. También si la obra fue obtenida utilizando algún perfil del investigador. Y por último cada extractor tiene un peso asignado según su oficialidad (Web of Knowledge y Scopus > Researchgate > GoogleScholar) y que dependiendo de este tendrá más o menos fiabilidad.

5.2 Extractores

5.2.1 Procesos de extracción

Para poder añadir un nuevo extractor es necesario que el sitio web donde se desea realizar *crawling* cumpla ciertos requisitos:

1. Disponer de perfiles de usuarios y que estos tengan publicaciones que al menos contengan año de publicación, fuente y título.
2. Se puedan listar los perfiles de una entidad, p.e. Universidad de Málaga.
3. Se puedan hacer búsquedas por autor.

Una vez que se cumpla esto, el desarrollo del *crawler* se basa en la implementación de una interfaz llamada *IExtractor*. Que tiene los métodos utilizados para los procesos de extracción de todos los investigadores y de los procesos de actualización de los perfiles.

Siguiendo este esquema, la tarea de añadir un nuevo extractor es trivial. Por lo tanto, para añadir un proceso periódico para extraer todo se extiende de la clase abstracta

`ExtraerTodoJob` y para un proceso de actualizar perfil la clase abstracta `ActualizarExtractorJob`.

```
public class ScopusActualizarExtractorJob extends ActualizarExtractorJob {

    @Override
    public IExtractor getExtractor() {
        return new Scopus();
    }
}
```

Ilustración 5.4 Implementación de Scopus para actualizar un perfil

```
public class ScopusExtraerTodoJob extends ExtraerTodoJob {

    @Override
    public IExtractor getExtractor() {
        return new Scopus();
    }

}
```

Ilustración 5.5 Implementación de Scopus para extraer todo

Los siguientes diagramas ilustrarán mejor los pasos seguidos por un extractor durante la extracción de todas las publicaciones de todos los investigadores del sistema.

1. Extraer todos los perfiles y realizar emparejamiento entre los perfiles y los investigadores.

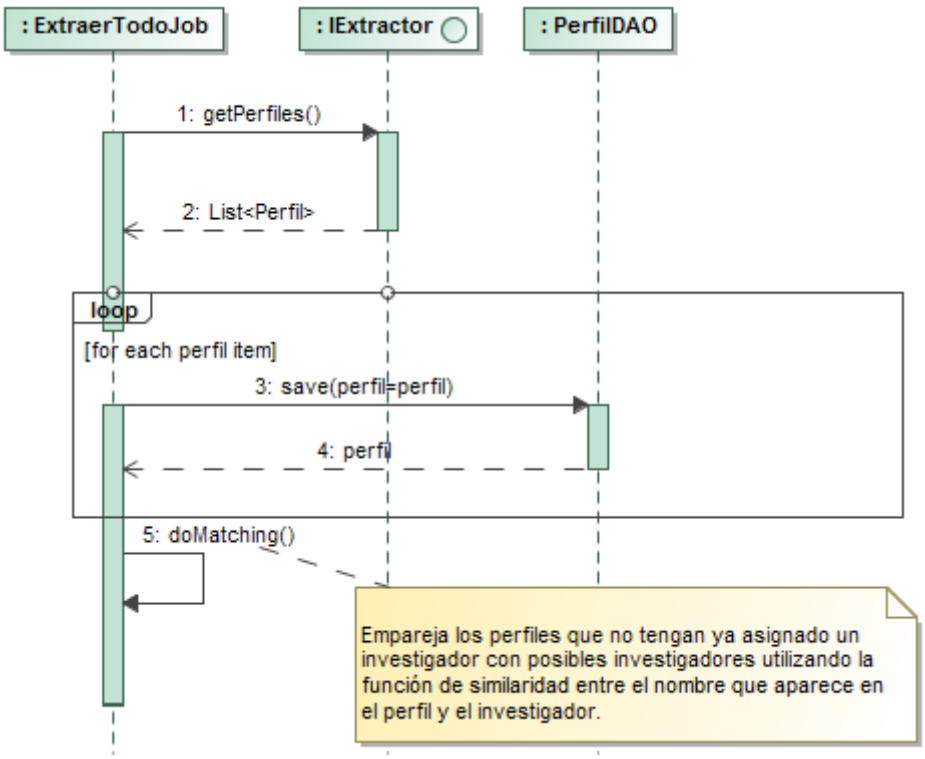


Ilustración 5.6 Diagrama de secuencia del primer paso

2. Completar información de los investigadores que tengan perfiles para este extractor.

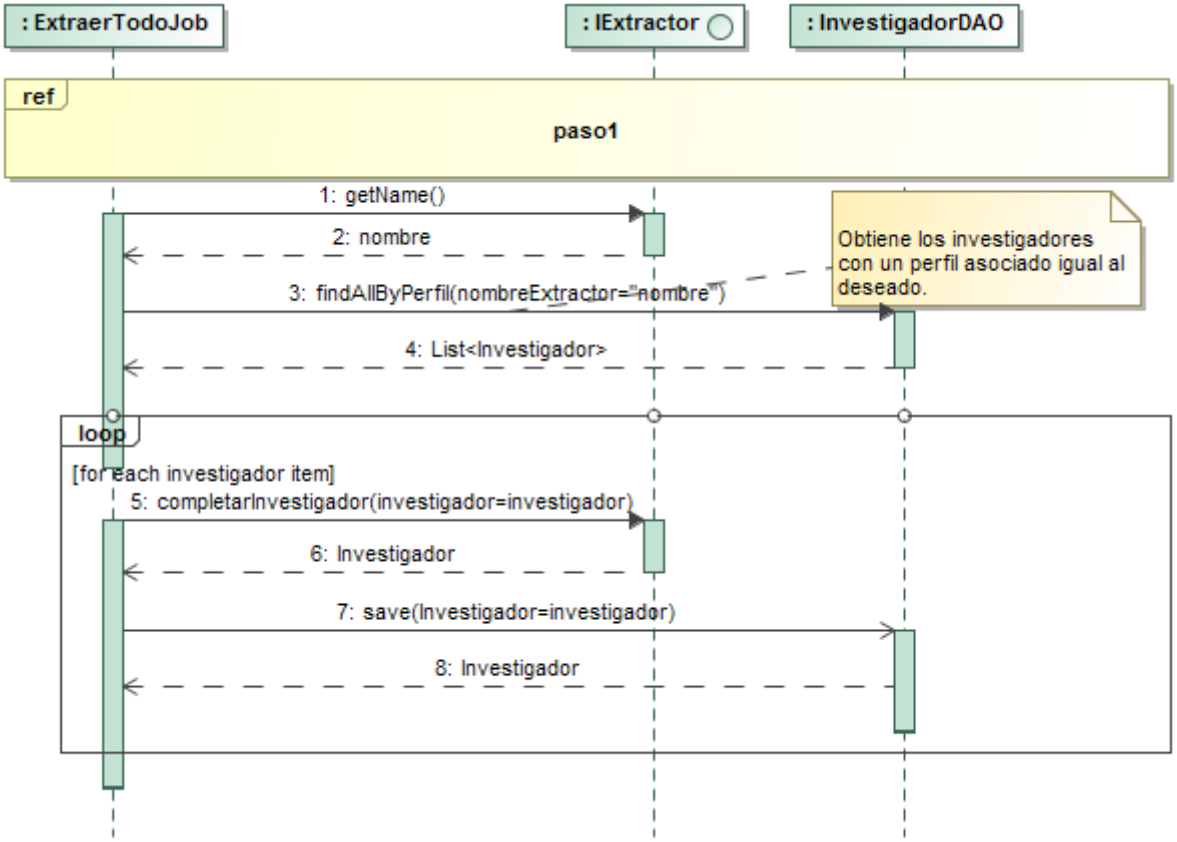


Ilustración 5.7 Diagrama de secuencia del segundo paso

3. Extraer todas las publicaciones para cada investigador que tengan perfiles para este extractor.
4. Obtener posibles firmas para los investigadores que no tienen perfil en para este extractor.
5. Este último paso es similar al 3, se recorren los investigadores a los que se le ha encontrado una posible firma en el paso 4 y de la firma con mayor similitud se utiliza el método `busquedaPorAutor(String autor)` de la interfaz `IExtractor`.

Los diagramas de `ActualizarExtractorJob` no se incluyen al ser similares a los de `ExtractorTodoJob` y variando en que su ejecución se realiza solo para un investigador, en lugar de todos.

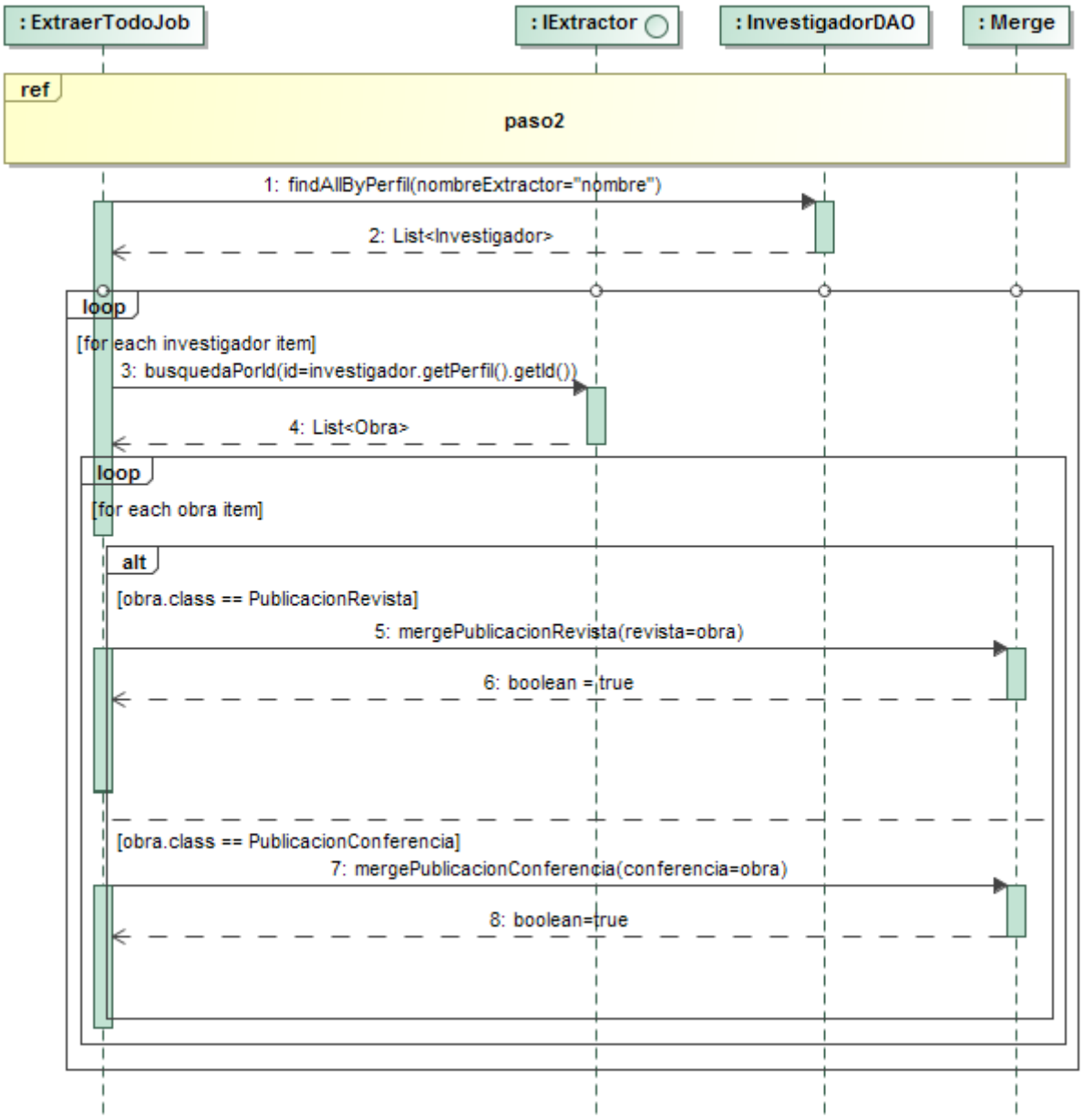


Ilustración 5.8 ilustración 5.9 Diagrama de secuencia del tercer paso

Ilustración 5.10 Diagrama de secuencia del cuarto paso

5.2.2 FUNCIONAMIENTO DE LOS EXTRACTORES WEB

Para comprender cómo funcionan los *crawlers* con las webs, se utilizará un ejemplo de un caso concreto, ya que el funcionamiento es común en la mayoría de los casos.

En la imagen que se muestra a continuación se puede ver lo que nosotros vemos al acceder, lo que ve el extractor y lo que le interesa extraer.

La configuración de dónde se encuentra cada uno de estos datos podría variar, por lo que existe un archivo de configuración general que permite cambiar en donde se busca la información de ser necesario. Además este archivo ofrece la posibilidad de adaptar los extractores a otras entidades.

El ejemplo mostrado hace referencia a la web donde se muestra una publicación de revista en Google Scholar.

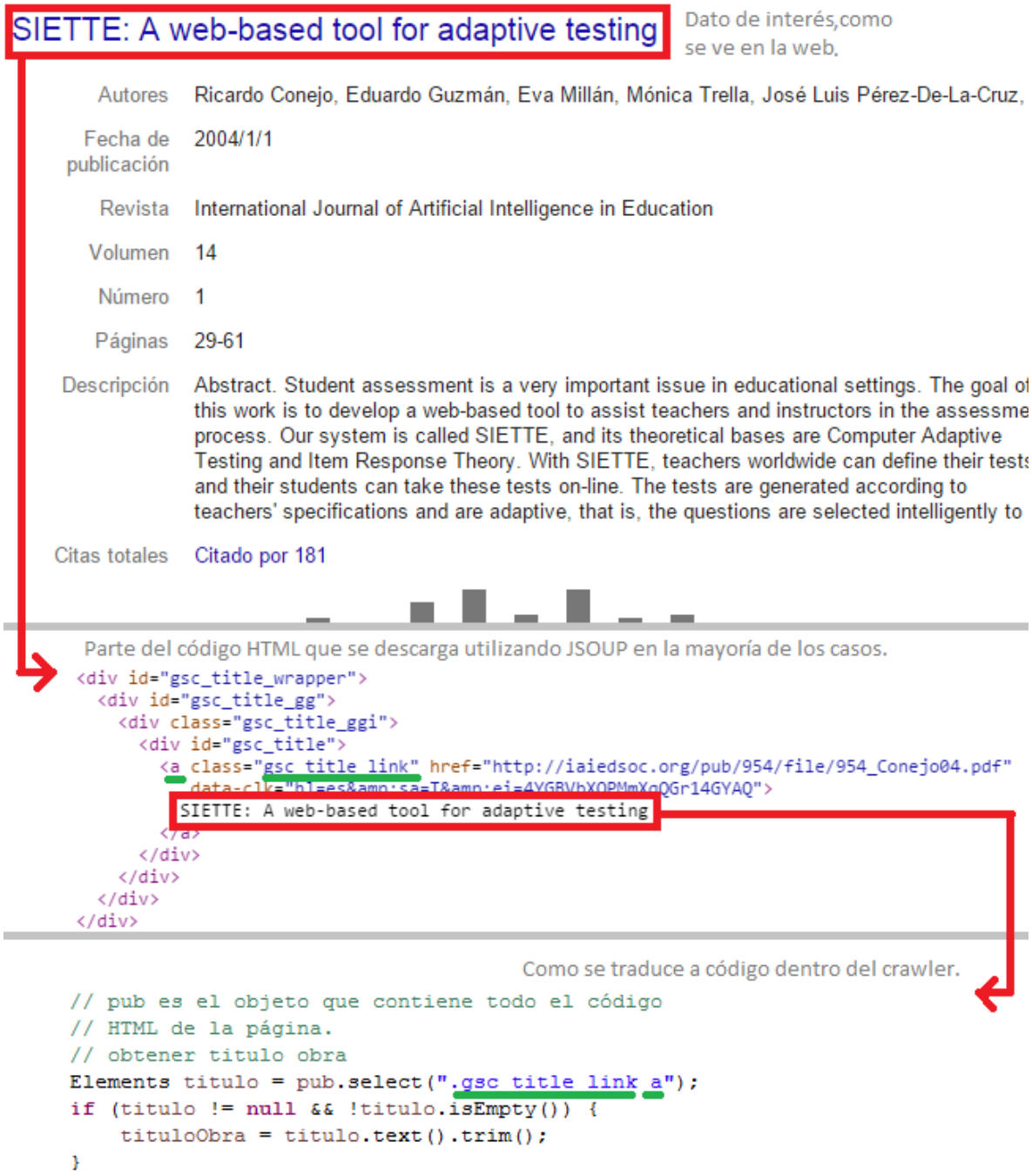


Ilustración 5.11 Ejemplo de funcionamiento del crawler de GoogleScholar.

5.2.3 GOOGLE SCHOLAR

BLOQUEOS

Por defecto Google suele bloquear a cualquier *crawler* que trabaje sobre alguno de sus dominios sin respetar los tiempos entre cada petición o realizando descargas. Cuando sospechan que una IP está realizando esta actividad, realizan un bloqueo por el cual el usuario, para poder continuar navegando en cada página, tiene que reconocer la palabra

escrita dentro de una imagen con ruido; lo que tradicionalmente se conoce como un CAPTCHA[4].

Para que el *crawler* de Google Scholar no sea bloqueado por este sistema y su comportamiento fuese similar al de una persona que navega, las soluciones propuestas fueron:

1. Utilizar diferentes agentes de navegación web escogidos aleatoriamente de una lista, es decir, al realizar una petición a una web simular ser un navegador.
2. Utilizar las cookies de navegación que se enviaban en la respuesta para futuras peticiones.
3. Utilizar intervalos de tiempo aleatorios a la hora de realizar una consulta, que siempre fuese mayor a 4 segundos como mínimo.

Estas soluciones evitan el bloqueo a cambio de asumir costes de tiempo para la ejecución de las tareas. Por otra parte, para la búsqueda por firmas era necesario realizar una descarga del documento en formato RIS[5] , en este caso el proceso de extracción suele fallar, al realizarse de forma continua, a pesar de utilizar los pasos descritos anteriormente.

5.2.4 RESEARCHGATE

BUSCAR INVESTIGADORES

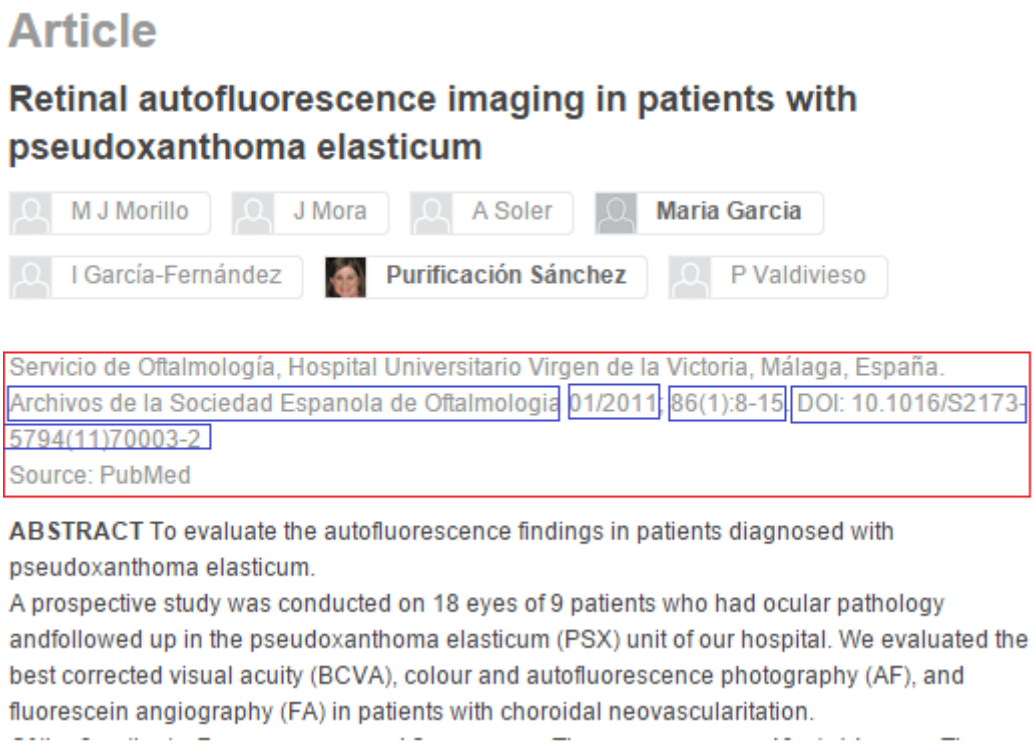
El principal inconveniente que tiene esta red social es al acceder de forma anónima, ya que es imposible realizar una búsqueda por nombre de un investigador. Para poder realizar esta acción aunque no la ofrezca la plataforma se utilizó un buscador externo para realizar búsquedas en la web, como es el caso de DuckDuckGo.

La idea consiste en realizar una búsqueda del tipo: “Nombre Apellido site: www.researchgate.net”, con lo que hace una búsqueda en todas las webs indexadas con ese dominio. Del resultado mostrado, entra en los 10 primeros resultados, busca todos los nombres de investigador que encuentra que lleven a un link de perfil (registrado o no) y sobre los encontrados obtiene el perfil con el nombre con mayor similitud y si este pertenece a la entidad que se busca (p.e. Universidad de Málaga) le da preferencia sobre el resto de encontrados en caso de empate.

BÚSQUEDA DE INFORMACIÓN DENTRO DE LAS DESCRIPCIONES

A pesar de ser una red social muy utilizada, a simple vista cuesta ver el año de publicación de las obras o en la fuente que fue publicada. Estos datos se encuentran dentro de cada publicación en la descripción de la siguiente forma:

[4] Completely Automated Public Turing test to tell Computers and Humans Apart
[5] Research Information Systems


Article

Retinal autofluorescence imaging in patients with pseudoxanthoma elasticum

M J Morillo · J Mora · A Soler · Maria Garcia · I García-Fernández · Purificación Sánchez · P Valdivieso

Servicio de Oftalmología, Hospital Universitario Virgen de la Victoria, Málaga, España.

Archivos de la Sociedad Espanola de Oftalmologia 01/2011; 86(1):8-15. DOI: 10.1016/S2173-5794(11)70003-2

Source: PubMed

ABSTRACT To evaluate the autofluorescence findings in patients diagnosed with pseudoxanthoma elasticum.

A prospective study was conducted on 18 eyes of 9 patients who had ocular pathology andfollowed up in the pseudoxanthoma elasticum (PSX) unit of our hospital. We evaluated the best corrected visual acuity (BCVA), colour and autofluorescence photography (AF), and fluorescein angiography (FA) in patients with choroidal neovascularitation.

Ilustración 5.12 En rojo descripción y en azul los datos de interés

Para poder obtener esta información fue necesario el uso de expresiones regulares para obtención de fechas, páginas de la publicación, volumen, títulos de fuente, DOI[6], ISSN, entre otros.

Puede que por estos dos motivos el *crawler* de ResearchGate junto al de Google Scholar sean los más lentos.

5.2.5 Web Of Knowledge

Búsqueda con API

De las bases de datos existentes con información sobre publicaciones científica, *Web Of Knowledge*, es una de la pocas que ofrece una conexión a la misma a través de una API. Ofrecen a las entidades públicas un acceso gratuito a una versión Lite a través de IPs previamente validadas.

El servicio web que ofrecen se comunica a través de mensajes SOAP y es necesario crear un cliente, que fue generado utilizando los documentos WSDL de sus servicios y la librería JAXB. Es interesante tener en cuenta que este servicio web tienen restricciones sobre el número de solicitudes que se pueden hacer en un minuto. Además no da acceso a todas las bases de datos y la información que devuelve sobre las publicaciones es limitada.

A pesar de todo esto, ofrecen la información necesaria para poder realizar *crawling* y los datos de perfiles de los investigadores se pueden obtener de la web researcherid.com. Es el sitio donde hay el menor número de investigadores registrados.

[6] Identificador Digital de Objetos

5.2.6 SCOPUS

DESCARGA DE PERFILES

Es el extractor que menos problemas ha dado a lo largo de su desarrollo, solo tiene una dificultad consistente en que, para simplificar el proceso de extracción, necesita descargar todas las publicaciones en un único fichero en formato RIS y luego recorrerlas.

La facilidad de leer los archivos RIS reside en que los datos vienen organizados de la siguiente manera TAG – INFORMACIÓN y cada uno por línea, siendo obligatorios al principio y fin los tags TY (tipo publicación) y ER (fin de referencia).

Para poder realizar la descarga es necesario que el usuario introduzca a mano las cookies utilizadas para realizar la misma acción en un navegador. Estas se pueden ver a través de la consola de desarrollador de cualquier navegador moderno, una vez localizadas se copian en el archivo de configuración.

6 CONCLUSIONES

Esta última sección está dedicada a las conclusiones obtenidas tanto de forma personal, del desarrollo y académicas en el resultado final. También se comentará futuras modificaciones pensadas para continuar el desarrollo de la aplicación.

6.1 RESULTADO FINAL Y CONCLUSIONES

Antes de comenzar a exponer mis conclusiones sobre el desarrollo del proyecto en la siguiente tabla se muestra los datos obtenidos por cada extractor y el número total de publicaciones obtenidas y el de perfiles. También se indica el número de perfiles que se han asignado.

Extractor	Perfiles	Perfiles asignados	Pub. en revistas	Pub. en conferencias
ResearchGate	1.694	1.091	10.634	575
GoogleScholar	507	416	7.118	1.588
WebOfKnowledge	269	216	25.776	1.477
Scopus	4.356	1.474	9.680	1.193

También se han obtenido datos para **15595** revistas pertenecientes al ISI JCR entre los años 1997 – 2013 y para las **2134** conferencias listadas en el CORE Conference Ranking.

De estos resultados se puede decir que han sido satisfactorios y cumpliéndose las expectativas, tal es el caso, que el proyecto a día de hoy sigue en desarrollo e implementando nuevas funcionalidades.

Las conclusiones obtenidas del desarrollo las podría agrupar dependiendo de la fase del proyecto.

- Análisis: es lo más importante del proyecto si nuestra idea es intentar reescribir código o realizar cambios, asegurarse que es exactamente lo que quiere el cliente antes de escribir una línea de código.
- Maquetación web: es la asignatura pendiente de cualquier persona que solo se dedique a desarrollar, es la parte que más me costó pensar en ocasiones y la que ha día de hoy no me termina de convencer.
- Diseño: esta parte ha sido tanto satisfactoria como frustrante. Lo que he aprendido es a no “reinventar la rueda”, no intentar desarrollar funciones complejas si ya están desarrolladas o probadas en otras librerías. También a organizar el mismo en pequeñas tareas y tener un flujo de trabajo a la hora de afrontar un problema.
- Documentación: nunca dejarla para el final. No hay nada peor que tener un producto y escribir de él cuando ya lo tienes desarrollado. Lo mismo con los comentarios dentro del código.

A nivel académico o a lo que se refiere en lo aprendido a lo largo de este trabajo de fin de grado, podría decir que he puesto en práctica el cómo llevar lo estudiado a lo largo de la carrera en gran parte de las asignaturas, a un proyecto que soluciona un problema real.

Me ha ayudado a darme cuenta que soy autosuficiente y organizado, capaz de buscar una solución a los problemas que se me presentan. También autodidacta, al pasar varias horas leyendo, aprendiendo y documentándome para buscar las mejores opciones para el proyecto.

En general, la experiencia obtenida en este TFG ha sido muy enriquecedora y cierra una etapa en mi vida al terminarlo, pero abre otra en la que quiero aspirar a más.

6.2 PLANES FUTUROS

El desarrollo de la aplicación tanto web y de nuevos *crawlers* continúa para llegar más lejos en los resultados obtenidos. Las mejoras que se están haciendo o se plantean desarrollar son:

- Añadir *crawlers* específicos para otras áreas dentro de la entidad, como pueden ser algunos para repositorios de publicaciones más específicos de otras áreas tales como Humanidades y Arte, donde las publicaciones no son el único factor a tener en cuenta.
- Poder recomendar a los investigadores compañeros afines con los que poder publicar y hacer sugerencias de publicaciones necesarias para conseguir los requisitos de los sexenios.
- Utilizar algoritmos de aprendizaje para evitar cometer errores como los que ocurren a la hora de asociar un investigador con un perfil.
- Implementar un servicio web donde los investigadores interesados puedan compartir la lista de sus publicaciones.

7 BIBLIOGRAFÍA

http://wokinfo.com/citationconnection/ - The Citation Connection - Real Facts - IP & Science - Thomson Reuters - [Accedido el 1 de Marzo de 2015]

https://www.deutschland.de/en/topic/knowledge/networks-partnerships/the-research-network-researchgate - The research network ResearchGate - Deutschland.de - Your link to Germany - [Accedido el 1 de Marzo de 2015]

http://www.core.edu.au/index.php/conference-rankings - Computing Research & Education - Conference Rankings - [Accedido el 1 de Marzo de 2015]

https://www.mongodb.org/ - MongoDB - [Accedido el 1 de Marzo de 2015]

https://eclipse.org/ - Eclipse - The Eclipse Foundation open source community website. - [Accedido el 1 de Marzo de 2015]

http://es.wikipedia.org/wiki/Java_EE - Java EE - Wikipedia, la enciclopedia libre - [Accedido el 2 de Mayo de 2015]

http://jama.jamanetwork.com/article.aspx-articleid=184519 - JAMA Network | JAMA | Comparisons of Citations in Web of Science, Scopus, and Google Scholar for Articles Published in General Medical Journals - [Accedido el 3 de Mayo de 2015]

http://quartz-scheduler.org/ - Quartz Scheduler | - [Accedido el 15 de Mayo de 2015]

https://jquery.com/ - jQuery - [Accedido el 12 de Mayo de 2015]

https://maven.apache.org/ - Maven - Welcome to Apache Maven - [Accedido el 20 de Mayo de 2015]

http://es.wikipedia.org/wiki/MagicDraw_UML - MagicDraw UML - Wikipedia, la enciclopedia libre - [Accedido el 13 de Abril de 2015]

http://projects.spring.io/spring-security-saml/ - Spring Security SAML - [Accedido el 15 de Abril de 2015]

https://github.com/mongodb/morphia - mongodb/morphia - GitHub - [Accedido el 2 de Marzo de 2015]

http://logging.apache.org/log4j/2.x/ - Log4j - Log4j 2 Guide - Apache Log4j 2 - [Accedido el 5 de Mayo de 2015]

https://github.com/yasserg/crawler4j - yasserg/crawler4j - GitHub - [Accedido el 7 de Abril de 2015]

http://community.jaspersoft.com/project/jasperreports-library - JasperReports- Library | Jaspersoft Community - [Accedido el 23 de Mayo de 2015]

http://www.stackoverflow.com/ - Stack Overflow - [Accedido a lo largo del desarrollo]

http://www.movired.com/blog/quartz/ - Introducción a Quartz Framework. Planificador de Tareas - Movired Blog - [Accedido el 8 de Mayo de 2015]

http://www.uma.es/personal-docente-e-investigador/cms/menu/gestion-pdi/sexenios-pdi-contratado/ - Sexenios PDI contratado - Universidad de Málaga - [Accedido el 12 de Mayo de 2015]

http://apps.webofknowledge.com/ - Web of Science - Please Sign In to Access Web of Science - [Accedido el 17 de Mayo de 2015]

http://www.codeproject.com/Articles/598581/How-to-integrate-Spring-oAuth-with-Spring-SAML - How to integrate Spring-oAuth2 with Spring-SAML – CodeProject - [Accedido el 28 de Mayo de 2015]

http://xlinux.nist.gov/dads//HTML/Levenshtein.html - Levenshtein distance - [Accedido el 21 de Abril]

https://en.wikipedia.org/wiki/RIS_(file_format) - RIS (file format) - [Accedido el 14 de Abril]

https://es.wikipedia.org/wiki/Web_scraping - Web scraping - [Accedido el 28 de Junio]

https://es.wikipedia.org/wiki/Ara%C3%B1a_web - Araña web - [Accedido el 28 de Junio]

https://www.promptcloud.com/blog/data-scraping-vs-data-crawling/ - Data Scraping Vs Data Crawling | Web Crawling | PromptCloud - [Accedido el 28 de Junio]

ANEXO I: MANUAL DE USUARIO

1 MENÚ

1.1 OPCIONES

MI PERFIL
Muestra la información de su perfil de investigador dentro de la aplicación. Tenga en cuenta que esta información ha sido obtenida de forma automatizada y debe ser comprobada.

NOTIFICACIONES
Área donde se muestra la actividad de los coautores de las publicaciones. Estas podrían ser erróneas si no tiene su perfil actualizado.

MENÚ PERSONAL
Opciones adicionales del usuario.

INVESTIGADORES
Buscar investigadores y descargar resúmenes.

GRUPOS PERSONALIZADOS
Administración de los grupos personalizados.

OTRAS BÚSQUEDAS
Realizar búsquedas avanzadas de departamentos, centros, grupos personalizados y áreas de investigación.

2 MI PERFIL

2.1 PERFILES

En esta apartado se encuentras los perfiles que el sistema ha encontrado basándose en su nombre.

BORRAR PERFIL
Si cree que alguno de los perfiles que se muestran no le pertenece, puede borrarlo pulsando el botón con una X.

AÑADIR PERFIL

Puede añadir un perfil pulsando “Añadir perfil”. En el siguiente cuadro de diálogo podrá rellenar:

- **Nombre:** el nombre que aparece en el perfil.
- **Id y Sitio Web:** dependiendo del sitio web que desea utilizar el id lo puede obtener de la barra de direcciones del navegador, en negrita está marcado lo que debería utilizar.
    - **ResearchGate**: http://www.researchgate.net/profile/**Ricardo_Conejo**
    - **ResearcherID**: http://www.researcherid.com/rid/**J-7272-2013**
    - **GooleScholar**: https://scholar.google.es/citations?user=**4eo8Zj8AAAAJ**
    - **Scopus**: http://www.scopus.com/authid/detail.url?authorId=**6602861828**
    - **ORCID**: http://orcid.org/**0000-0003-0810-4608**

2.2 FIRMAS

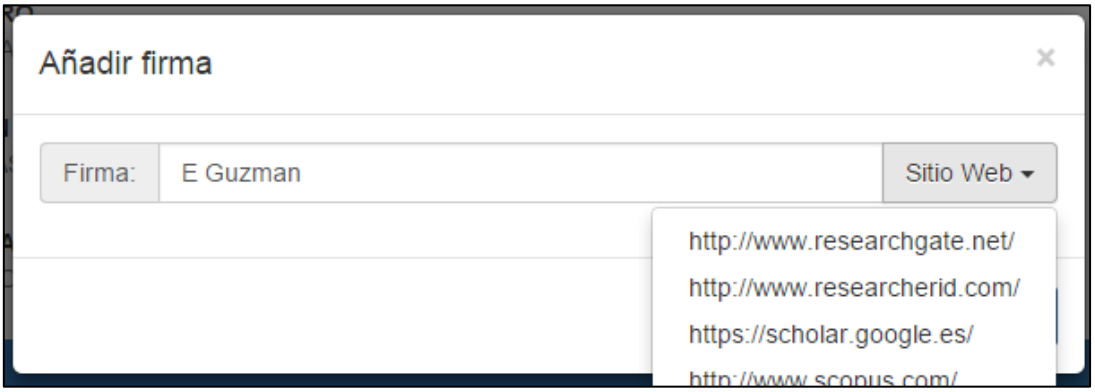
En esta apartado se encuentras las firmas que el sistema ha encontrado en relación a las publicaciones dentro de su departamento y a los perfiles asociados.

BORRAR FIRMA

Si cree que alguna de las firmas que se muestran no se corresponde con su nombre o firma habitual, puede borrarla pulsando el botón con una X.

AÑADIR FIRMA

Puede añadir una firma pulsando en “Añadir firma”. En el siguiente cuadro de diálogo podrá rellenar con su firma y el portal correspondiente.



RECOMENDAR FIRMAS

En base a las firmas de las publicaciones de su departamento, el sistema le dará un listado de posibles firmas a utilizar.

Para utilizar esta función pulse en "Recomendar firmas" , seleccione las que desee y pulse el botón “Añadir seleccionadas”.

2.3 ACTUALIZAR

Una vez que haya actualizado su información referente a perfiles y firmas, puede solicitar una actualización de su perfil pulsando en el botón que se muestra a continuación. El tiempo que tarda en actualizar puede variar dependiendo del volumen de las actualizaciones pendientes.

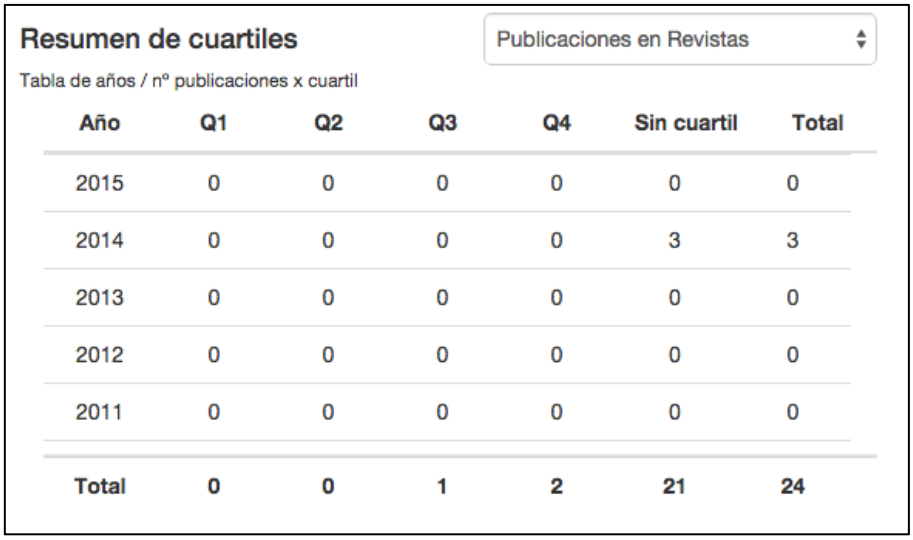
2.4 VER ESTADÍSTICAS

Para visualizar toda las publicaciones obtenidas a partir de sus datos y sus estadísticas, pulse el siguiente botón.

3 ESTADÍSTICAS

3.1 RESUMEN

Visión global de las publicaciones en revistas/conferencias/etc. obtenidas.



Resumen de cuartiles — Publicaciones en Revistas

Tabla de años / nº publicaciones x cuartil

Año	Q1	Q2	Q3	Q4	Sin cuartil	Total
2015	0	0	0	0	0	0
2014	0	0	0	0	3	3
2013	0	0	0	0	0	0
2012	0	0	0	0	0	0
2011	0	0	0	0	0	0
Total	**0**	**0**	**1**	**2**	**21**	**24**

3.2 MENÚ ACCIONES

Listado de acciones disponibles a realizar desde las estadísticas.

3.3 VERSION IMPRIMIBLE

Imprime un resumen de las publicaciones obtenidas (de momento solo imprime publicaciones en revistas).

Para utilizar esta función presionar el siguiente botón:

Se desplegará un cuadro de dialogo como el siguiente, donde podrá elegir el formato y otras opciones.

3.4 VER LÍNEAS

Lista todas las palabras claves asociadas a sus publicaciones, cada una tiene una puntuación basada en un heurístico el cual a su vez se basa en los índices de calidad. Puede ordenar la tabla pulsando sobre las cabeceras de las columnas.

#	Publicacion	Revista	Año	Cuartil	Factor de impacto	Link	Fiabilidad	Opciones
	Algorithm	Artificial	2012	>Q1	<=2		>20	
1	Mining Web-based Educational Systems to Predict Student Learning Achievements	International Journal of Artificial Intelligence and Interactive Multimedia	2015	NO	0.00	Link	95.00%	

Para ver las líneas, pulse la siguiente opción del menú de acciones:

Al final de la página se cargará una tabla como la que se puede ver a continuación:

Keywords

Obtenidas de las publicaciones en revista y con una puntuación mayor a 4. Puntuacion calculada de la forma : (5 * nº en Q1)+(4 * nº en Q2)+(3 * nº en Q3)+(2 * nº en Q4)+(1 * nº sin cuartil)

#	Keyword	Q1	Q2	Q3	Q4	Sin cuartil	Puntuación
1	information systems (is)	0	0	2	2	0	10
2	fuzzy sql	0	0	2	2	0	10

3.5 VER PUBLICACIONES

Lista todas sus publicaciones, con una puntuación basada en los índices de calidad. La fiabilidad es calculada de forma automática basada en los coautores, perfiles, fuentes de obtención. Puede ordenar la tabla pulsando sobre las cabeceras de la misma, para filtrar las columnas utilice expresiones regulares si lo desea.

Para ver esta lista, pulse en la siguiente opción del menú de acciones:

Al final de la página se cargará una tabla como la que se puede ver a continuación:

VALIDAR PUBLICACIÓN

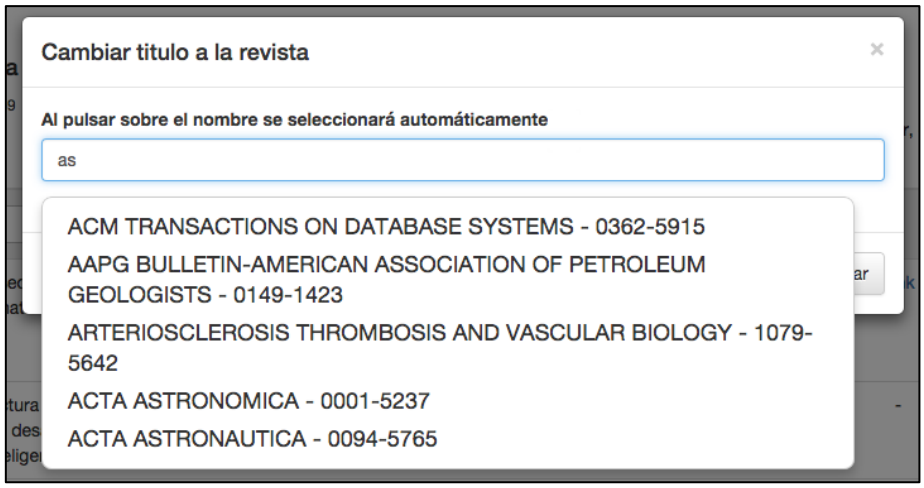
Pulse en el icono con la mano y el dedo pulgar apuntando hacia arriba.

RECHAZAR PUBLICACIÓN

Pulse en el icono con la mano y el dedo pulgar apuntando hacia abajo.

CAMBIAR FUENTE

Pulse sobre el nombre la fuente y se desplegará un cuadro de dialogo como el siguiente:

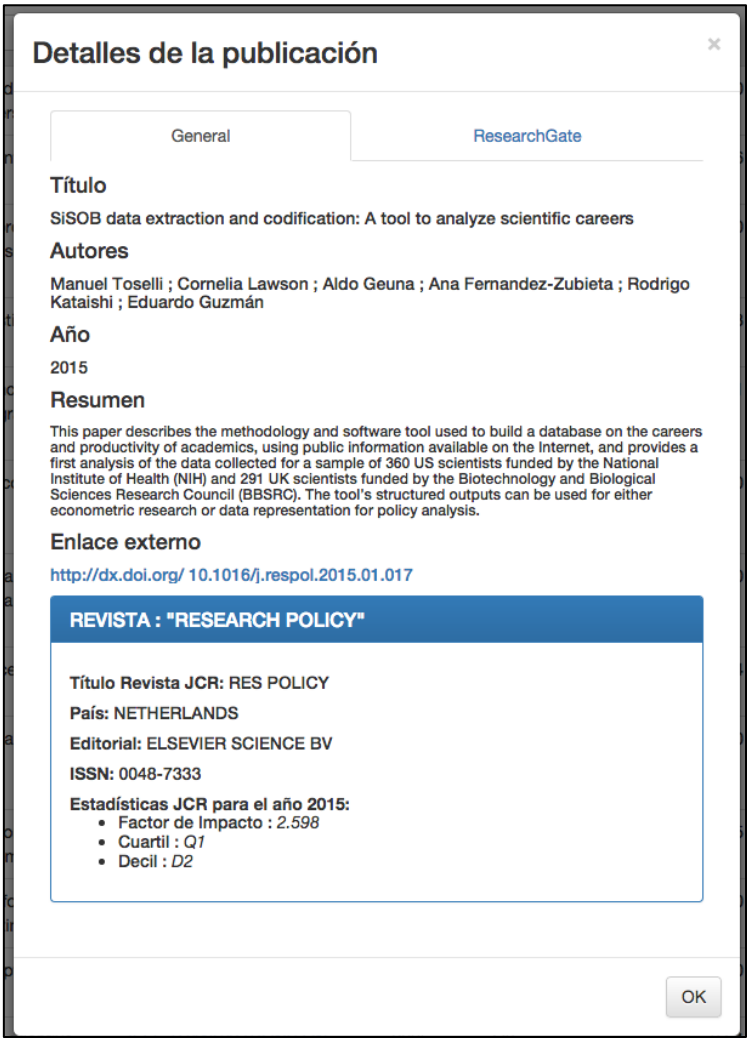


Al desplegarse el cuadro de diálogo escriba el nombre o ISBN de la fuente y pulse sobre el correcto. Recuerde que deberá esperar a que sea aprobada su solicitud.

VER MÁS INFORMACIÓN

Pulse sobre el icono con forma de lupa para ver más información de la publicación .

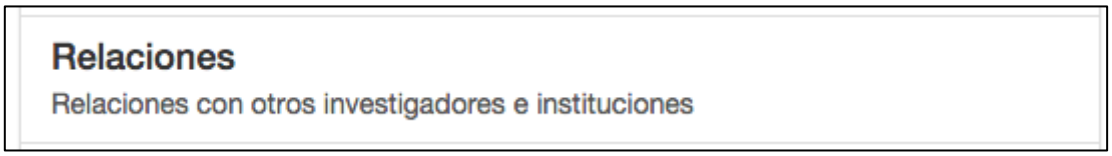
Se desplegará un cuadro de dialogo como el siguiente con la información de la publicación:



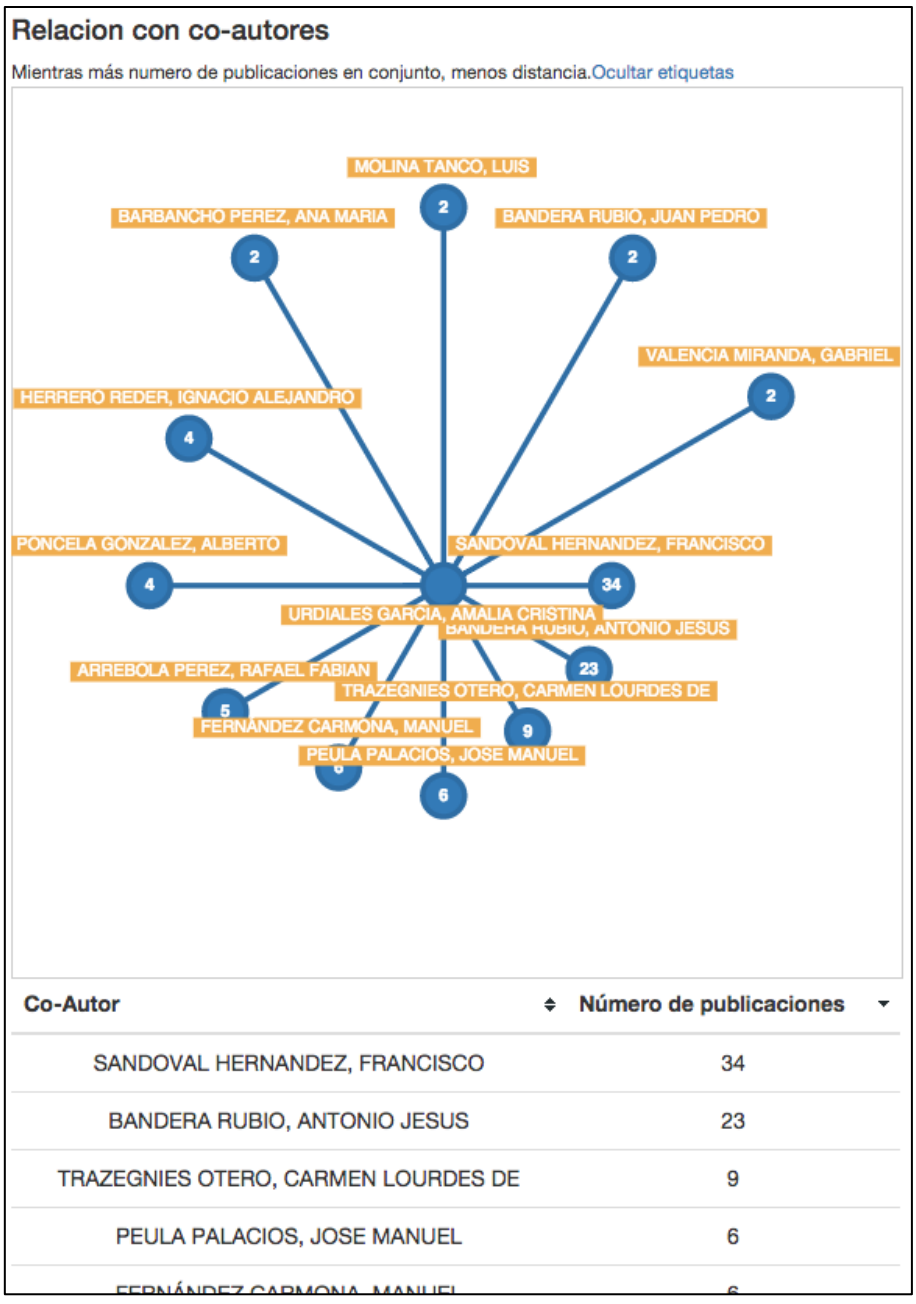
3.6 RELACIONES

Listado de las relaciones con otros investigadores y entidades. El grafo muestra hasta las 16 entidades/investigadores con los que más ha trabajado, a mayor proximidad más publicaciones en conjunto. La tabla, en cambio, contiene toda la información.

Para verlas, pulse la siguiente opción del menú de acciones:



Al final de la página se cargará un grafo y una tabla como la que se puede ver a continuación:



3.7 FILTRAR

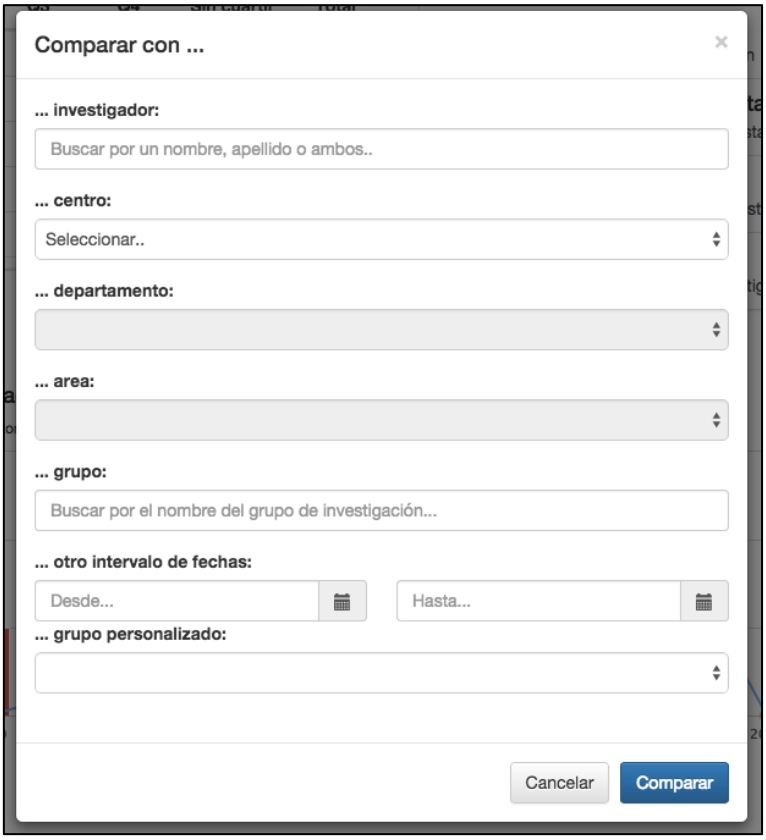
Si lo desea también podrá filtrar por años dentro de las estadísticas.

3.8 COMPARAR

Realiza una comparación a nivel de publicaciones de la estadística actual contra cualquier otra estadística de investigadores/departamentos/áreas de conocimiento/grupos personalizados/centros o incluso en intervalos de año.

Para verlas, pulse la siguiente opción del menú de acciones:

Se desplegará un cuadro de diálogo como el siguiente donde puede elegir frente a qué desea comparar las estadísticas en las que se encuentra:



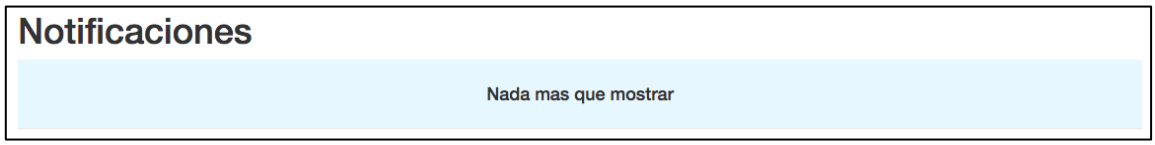
Una vez seleccionado lo que desea comparar, pulse "Comparar"

4 NOTIFICACIONES

4.1 VER NOTIFICACIONES

Podrá ver toda la actividad que se realizan los coautores de sus publicaciones.

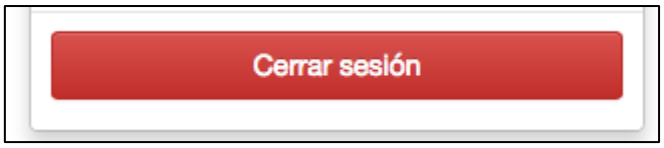
Por ejemplo, si un coautor valida o rechaza una publicación que le pertenece, será informado en esta sección.



5 MENÚ PERSONAL

5.1 CERRAR SESIÓN

Cerrará sesión tanto en iDUMA como en la aplicación.



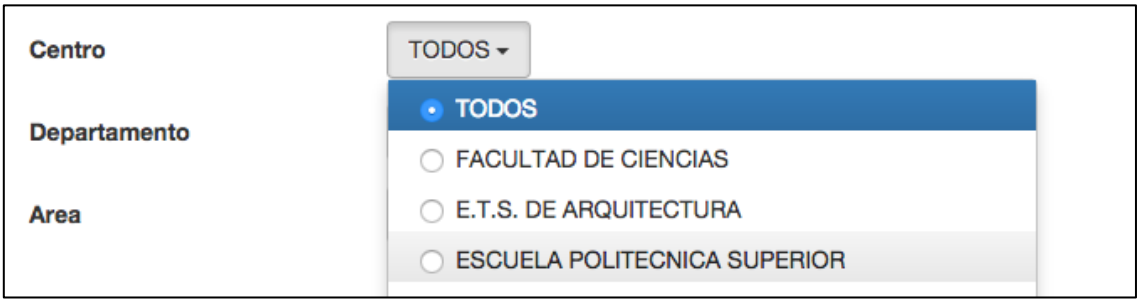
6 INVESTIGADORES

6.1 LISTAR INVESTIGADORES

Al acceder aparecerá el listado de investigadores que puede consultar.

6.2 FILTRAR INVESTIGADORES

Si lo necesita, utilice y aplique los filtros que sean necesarios, una vez que lo haya realizado, pulse en la lupa de buscar para actualizarlos.



6.3 VER INVESTIGADOR

Si desea ver información de un investigador pulse la flecha hacia abajo que se encuentra al final del nombre  .

Una vez pulsada, se desplegará la información del investigador como se ve en la siguiente imagen:

Si desea volver a ocultar la información del investigador simplemente la flecha con dirección hacia arriba al final del nombre  .

6.4 EXPORTAR A CSV

Si desea exportar los datos de la lista que se esta mostrando en cualquier momento, puede hacerlo pulsando el botón "Exportar a CSV".

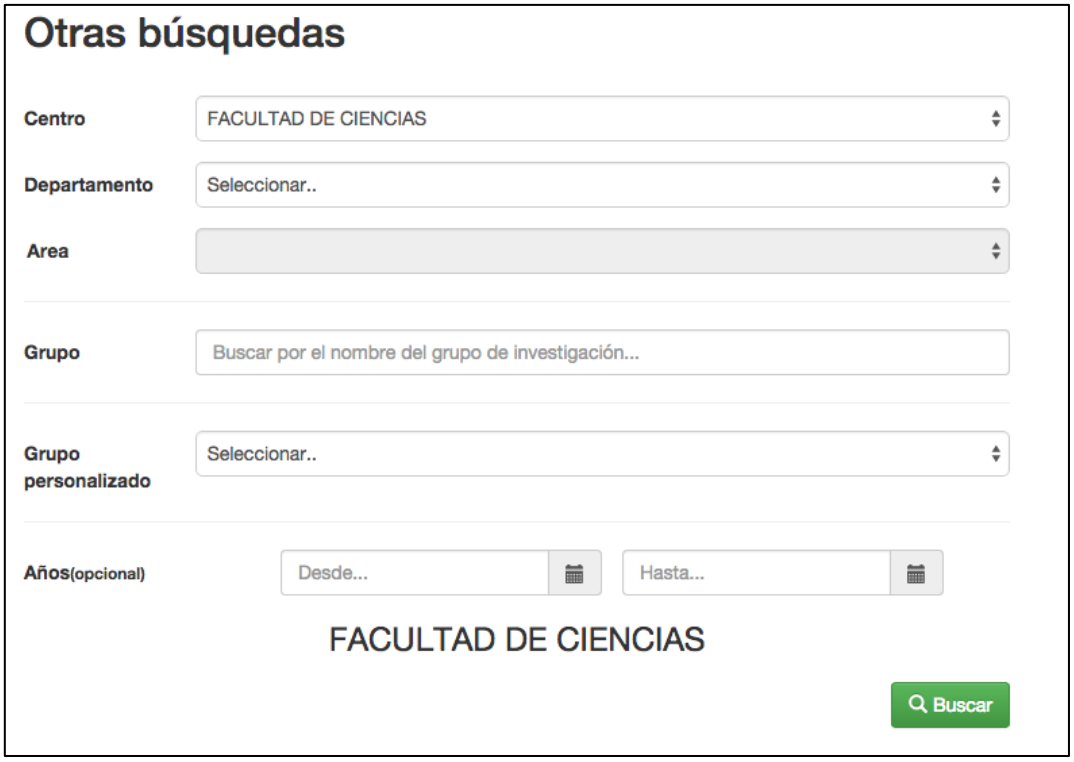
Una vez pulsado, se desplegará un cuadro de diálogo donde puede seleccionar los datos que desea exportar a CSV:

7 OTRAS BÚSQUEDAS

7.1 REALIZAR BÚSQUEDA

Permite realizar búsquedas de departamentos/áreas de conocimiento/grupos personalizados/centros.

Para llevar a cabo una búsqueda, simplemente tiene que seleccionar lo que desea buscar y seguidamente pulsar en "Buscar". Opcionalmente puede utilizar un rango de años deseado.



Las búsquedas que agrupan a varios investigadores pueden llegar a tardar incluso un par de minutos.