



Universidad Politécnica
de Madrid

**Escuela Técnica Superior de
Ingenieros Informáticos**



Grado en Ingeniería del Software

Trabajo Fin de Grado

**Sistema de Extracción y Análisis de Información en
el Boletín Oficial del Estado**

Autor: Alberto Lara García
Tutor: Carlos Badenes Olmedo

CURSO 2023/2024

RESUMEN

Para encontrar información concreta en el Boletín Oficial del Estado (BOE), ya sea de carácter legal, sobre subastas o convocatorias de oposiciones, se puede hacer a través de la herramienta oficial del BOE. Sin embargo, la extracción de conocimiento y el análisis sobre las grandes cantidades de datos e información que publica el BOE diariamente representa un desafío significativo, ya que no existen herramientas específicas que asistan en esta tarea. Este trabajo de fin de grado se motiva por la necesidad de facilitar el acceso y la interpretación de la información que se publica en el BOE, tanto para expertos en el análisis de datos como para personas sin conocimientos en el tema.

El principal problema que aborda este proyecto es la necesidad de disponer de herramientas que ayuden a interpretar y extraer conocimiento de la información publicada diariamente en el Boletín Oficial del Estado. Para poder lograr los objetivos definidos en el proyecto ha habido que enfrentarse a una serie de retos como el manejo de grandes cantidades de datos y la actualización continua de información publicada en el BOE. Además, gracias a la revisión del estado del arte se ha visto que, aunque existen herramientas oficiales en el BOE para realizar búsquedas concretas, no hay ninguna forma de realizar análisis profundos y personalizados sobre los datos, problema que también se ha abordado en el proyecto.

La propuesta de este TFG consiste en desarrollar una herramienta que facilite la búsqueda y el análisis de la información publicada en el BOE. Propuesta que ha sido llevada a cabo siguiendo una metodología tradicional en la que primero se realizó un planteamiento y análisis del problema, seguido del diseño de un diagrama de la estructura de la herramienta. Luego se continuó con la implementación y codificación de cada módulo definido en el diagrama. Finalmente se evaluó el resultado del proyecto mediante pruebas de los módulos desarrollados, asegurando que cumplieran con los objetivos definidos.

Tras la ejecución del proyecto, los resultados mostraron que la herramienta es capaz de obtener, procesar y transformar la información del BOE, facilitando su evaluación e interpretación. Esto permite a los usuarios extraer conocimiento de manera eficiente y personalizada.

En conclusión, la definición del prototipo resultante de este TFG ofrece una solución inicial que permitirá el desarrollo de un sistema que, además de obtener, procesar y transformar la información del BOE, incluya una interfaz gráfica y más funcionalidades en base al feedback de los usuarios. Sistema que pretende ser de gran ayuda para las personas que busquen obtener conocimiento a partir del BOE, como juristas y otros profesionales.

ABSTRACT

To find specific information in the Boletín Oficial del Estado (BOE), whether legal, about auctions, or calls for competitive examinations, you can use the official BOE tool. However, extracting knowledge and analysing the vast amounts of data and information published by the BOE daily presents a significant challenge, as there are no specific tools to assist in this task. This final degree project is motivated by the need to facilitate access and interpretation of the information published in the BOE, both for data analysis experts and people without knowledge in the field.

The main problem addressed by this project is the need of tools to help understand and extract knowledge from the information published daily in the Boletín Oficial del Estado. To achieve the objectives defined in the project, several challenges had to be faced, such as managing large amounts of data and the continuous updating of information published in the BOE. Moreover, a review of the state of the art showed that although there are official tools in the BOE for specific searches, there is no way to perform deep and personalized analysis of the data, a problem also addressed in this project.

The proposal of this final degree project consists of developing a tool that facilitates the search and analysis of the information published in the BOE. This proposal was carried out following a traditional methodology where first, the problem was defined and analysed, followed by designing a diagram of the tool's structure. Next, each of the modules defined in the diagram was implemented and coded. Finally, the results of the project were evaluated by testing the developed modules, ensuring that they met the defined objectives.

After the implementation of the project, the results showed that the tool is capable of obtaining, processing and transforming information from the BOE, facilitating its evaluation and interpretation. This allows users to extract knowledge in an efficient and personalised way.

In conclusion, the definition of the prototype resulting from this final degree project offers an initial solution that will allow the development of a system that, in addition to obtaining, processing and transforming information from the BOE, includes a graphical interface and more functionalities based on user feedback, which will be of great help to people seeking to obtain knowledge from the BOE, such as legal professionals and other experts.

ÍNDICE GENERAL

Introducción	1
1.1. Contexto del proyecto	1
1.2. Objetivos	3
1.3. Retos del proyecto	3
1.4. Estructura del Documento	4
Trabajos Relacionados	4
2.1 BOE.es	4
2.2 Mi BOE.....	9
2.3 Herramientas no oficiales:	9
2.3.1 Aplicación Móvil Boletín Oficial de España.....	10
2.3.2 JurisTracker - App BOE	10
2.3.3 SuperBOE.....	11
Metodología	12
3.1. Análisis	12
3.2. Diseño	14
3.3. Implementación.....	15
3.3.1. Obtención del BOE:	15
3.3.2. Almacenamiento del BOE en MongoDB:.....	17
3.3.3. Integración con Neo4j:	19
3.4. Pruebas	21
3.5 Recursos Generados	24
3.6. Despliegue	34
Conclusiones Personales	40
Trabajo Futuro	41
Responsabilidad Social y Ambiental	41
Bibliografía	43

ÍNDICE DE FIGURAS

Figura 1. Ejemplo disposición general	2
Figura 2. Menú inicial BOE.....	5
Figura 3. Último BOE 8 Mayo 2024.....	6
Figura 4. Elegir BOE por fecha.....	7
Figura 5. Búsqueda sencilla BOE.....	7
Figura 6. Búsqueda filtrada simple.....	8
Figura 7. Búsqueda filtrada avanzada.....	8
Figura 8. MiBOE seleccionar temas de interés	9
Figura 9. MiBOE alerta.....	9
Figura 10. App móvil BOE.....	10
Figura 11. JurisTracker App BOE.....	10
Figura 12. Super BOE	11
Figura 13. Metodología seguida.....	12
Figura 14. Workflow.....	14
Figura 15. Ejemplo de sumario.....	16
Figura 16. Almacenamiento XML.....	17
Figura 17. Número XML descargados.....	17
Figura 18. N° publicaciones del sumario.....	17
Figura 19. XML vs JSON.....	18
Figura 20. XML de ejemplo	19
Figura 21. JSON de ejemplo.....	19
Figura 22. Boe sencillo.....	20
Figura 23. Almacenamiento XML.....	21
Figura 24. Número XML descargados.....	21
Figura 25. N° publicaciones del sumario.....	21
Figura 26. Conexión a MongoDB.....	22
Figura 27. Find en MongoDB.....	22
Figura 28. Log Neo4j.....	23
Figura 29. Neo4j resultados	23
Figura 30. BOE en Neo4j.....	24
Figura 31. Resultados desdeHasta.....	24
Figura 32. Cuenta bbdd vacia MongoDB	25
Figura 33. Cuenta bbdd llena	25
Figura 34. Resultado mostrarBOE().....	26
Figura 35. Resultados mostraBOE() en MongoDB.....	26
Figura 36. Resultados filtrados por departamento	27
Figura 37. Resultados cuenta en Neo4j	27
Figura 38. Resultad cuenta en MongoDB.....	27
Figura 39. Resultado java getboedepartamento	28
Figura 40. Resultado java getBoeDepartamentoRango().....	28
Figura 41. Ejemplo referencias anteriores	29
Figura 42. Referencias.....	30
Figura 43. Árbol de evolucionBOE.....	31
Figura 44. Referencias 13261.....	31
Figura 45. Comprobación 1er hijo	32
Figura 46. Referencias 23537.....	32
Figura 47. Referencias 23537 en Neo4j.....	32
Figura 48. Referencias 24842.....	33
Figura 49. Referencias 24842 en Neo4j.....	33
Figura 50. Aplicación algoritmo.....	33
Figura 51. Imagen Neo4j usada.....	35
Figura 52. Imagen MongoDB usada.....	35
Figura 53. Repositorio GitHub	35
Figura 54. Contenedores creados en Docker.....	36
Figura 55. Conexión a MongoDB.....	37
Figura 56. Conexión a Neo4j.....	37
Figura 57. descargarBOE original.....	37
Figura 58. descargarBOE editado.....	37
Figura 59. Ejemplo simple del Main para 1era ejecución.....	38
Figura 60. Inicio sesión Neo4j.....	38
Figura 60. Comprobación n° publicaciones en MongoDB	39
Figura 61. Publicación en MongoDB.....	39
Figura 62. Publicación en Neo4j	39

Introducción

1.1. Contexto del proyecto

El Boletín Oficial del Estado (BOE) representa una fuente de información esencial para toda la sociedad española, ya que se trata del medio oficial y principal en el que se publican las leyes, disposiciones y actos que deben ser conocidos por todos los españoles. Este proyecto abarca la necesidad de acceder a la gran cantidad de información que publica el BOE de una forma sencilla, rápida y que permita análisis complejos.

El BOE proporciona a cualquier individuo, empresa u organización la capacidad de conocer las normativas que les afectan directamente, permitiéndoles planificar y ajustar sus actividades en consecuencia. También contribuye al buen funcionamiento de la sociedad y su información es fundamental para investigadores, docentes, abogados, funcionarios públicos...

Siendo conscientes de la relevancia y la cantidad de personas y entidades que utilizan la información del BOE, este proyecto propone simplificar el acceso a estos datos para su procesamiento y análisis de manera eficiente y efectiva.

Los datos sobre los que esta herramienta va a trabajar son las publicaciones del BOE. El BOE se publica diariamente y contiene una gran variedad de documentos, incluyendo leyes, reales decretos, resoluciones, órdenes ministeriales, convocatorias de oposiciones, subastas, anuncios oficiales y otros actos administrativos. Estos documentos están estructurados en diferentes secciones, como:

- Disposiciones Generales: Leyes, reales decretos y disposiciones de carácter general.
- Autoridades y personal: Nombramientos, situaciones e incidencias.
- Otras disposiciones: Convenios colectivos, becas, premios y otros anuncios.
- Anuncios: Subastas, procedimientos judiciales, notificaciones, y otros anuncios de interés público.

Las publicaciones del BOE se realizan con una frecuencia diaria (excepto los domingos, salvo casos extraordinarios), lo que implica una actualización constante de la información, más aun teniendo en cuenta que no todo lo que se publica es nuevo, si no que se pueden derogar o actualizar documentos previamente publicados. Estas publicaciones cuentan con una estructura detallada y técnica, con un lenguaje jurídico que puede ser complejo para quienes no están familiarizados con él. Además, los documentos del BOE se publican en la web oficial (boe.es) en formato PDF, XML y EPUB. Siendo el único formato reconocido como oficial el PDF, ya que estos cuentan con un código CVE que identifica de forma única a cada documento publicado en el BOE. El resto de los formatos son considerados como meramente informativos, pero tienen el mismo contenido.

A continuación, se puede ver el ejemplo de una disposición general publicada en formato PDF, firmado electrónicamente con el código CVE situado en vertical en la zona inferior derecha del documento:

I. DISPOSICIONES GENERALES

PRESIDENCIA DEL GOBIERNO

7479 *Real Decreto 363/2024, de 9 de abril, por el que se convocan elecciones de diputados y diputadas al Parlamento Europeo.*

El Acta de 20 de septiembre de 1976 relativa a la elección de los diputados al Parlamento Europeo por sufragio universal directo, aneja a la Decisión 76/787 del Consejo, de 20 de septiembre de 1976, modificada por las Decisiones del Consejo 93/81, de 1 de febrero de 1993, 95/1, de 1 de enero de 1995, 2002/772, de 25 de junio de 2002 y de 23 de septiembre de 2002, y 218/994, de 13 de julio de 2018, regula las elecciones al Parlamento Europeo.

De conformidad con su artículo 8, y hasta la aprobación de un procedimiento uniforme, el procedimiento electoral se regirá en cada Estado miembro por sus disposiciones nacionales.

En España, el procedimiento se recoge en la Ley Orgánica 5/1985, de 19 de junio, del Régimen Electoral General, así como por las normas reglamentarias reguladoras de los procesos electorales.

Por Declaración del Consejo, de 22 de mayo de 2023, el período contemplado en el artículo 10, apartado 1, del Acta de 20 de septiembre de 1976 relativa a la elección de los diputados al Parlamento Europeo por sufragio universal directo, queda fijado, para la décima elección, del 6 al 9 de junio de 2024.

De conformidad con el apartado 2 del artículo 10 del Acta de 20 de septiembre de 1976 relativo a la elección de los diputados al Parlamento Europeo por sufragio universal directo, no se publicará ningún resultado oficial en ningún Estado miembro hasta después del cierre de los locales electorales del Estado miembro cuyos electores y electoras voten en último lugar.

Conforme al artículo 14 del Tratado de la Unión Europea, el Parlamento Europeo estará compuesto por un número de representantes que no excederá de setecientos cincuenta (750), más la persona que ostente la presidencia (+1). Dispone que la representación será decrecientemente proporcional, con un mínimo de seis (6) representantes por Estado miembro, y que no se asignará a ningún Estado miembro más de noventa y seis (96) escaños.

La Decisión (UE) 2023/2061 del Consejo Europeo, de 22 de septiembre de 2023, por la que se fija la composición del Parlamento Europeo, concreta el número total de escaños del Parlamento Europeo en 720, habida cuenta de los cambios demográficos acontecidos en los Estados miembros de la UE. Determina también el número de escaños asignados a cada Estado miembro para la legislatura 2024-2029, correspondiéndole a España la elección de 61 representantes.

En su virtud, a propuesta del Presidente del Gobierno, y previa deliberación del Consejo de Ministros en su reunión del día 9 de abril de 2024,

DISPONGO:

Artículo 1. Convocatoria de elecciones.

Se convocan elecciones al Parlamento Europeo, que se celebrarán el domingo día 9 de junio de 2024.

Artículo 2. Número de diputados y diputadas que se eligen.

El número de diputados y diputadas al Parlamento Europeo será de 61, en aplicación de lo dispuesto en el artículo 215 de la Ley Orgánica 5/1985, de 19 de junio, del

one: BOE-A-2024-7479
Verificable en <https://www.boe.es>

Figura 1. Ejemplo disposición general

Como hemos visto, la explotación y análisis de la información del BOE presenta varias dificultades: la gran cantidad de documentos publicados diariamente hace que sea difícil mantenerse al día con todas las actualizaciones, los documentos están en formatos que, aunque estandarizados, no son fácilmente manipulables para el análisis sin un procesamiento previo. Además, la estructura y el lenguaje legal de los documentos pueden resultar compleja y se requiere de cierto conocimiento del lenguaje jurídico para su interpretación, lo que puede ser un obstáculo.

No obstante, este proyecto se encarga del procesamiento de los datos para simplificar su acceso y análisis, se asegura de tener los datos disponibles y actualizados y facilita una futura incorporación de una interfaz gráfica con la que sea más accesible y simple la interpretación de los datos.

1.2. Objetivos

Los objetivos definidos en esta sección establecen las metas y propósitos específicos que han guiado el desarrollo e implementación del *Sistema de Extracción y Análisis de Información en el Boletín Oficial del Estado*

El presente proyecto desarrolla una solución software que permite a cualquier tipo de usuario realizar consultas y obtener información sobre el Boletín Oficial del Estado (BOE).

Aquí se detallan los objetivos clave que se abordarán en el marco de este proyecto:

- **OB.1-Procesar documentación legal:** Se pretende incorporar herramientas de procesamiento de datos que permitan estructurar y organizar la información del BOE de manera sistemática. Esto incluye técnicas de extracción, transformación y carga de datos (ETL) para mantener actualizada la base de datos.
- **OB.2-Facilitar la búsqueda y filtrado de información sobre textos legales:** La herramienta permitirá realizar búsquedas complejas y detalladas en el Boletín Oficial del Estado, las cuales no son posibles o resultan muy complicadas de ejecutar mediante la web oficial del BOE.
- **OB.3-Analizar la Calidad de los Resultados:** Se garantizará la precisión y la actualización constante de los datos del BOE en las bases de datos de la herramienta. Esto implica implementar mecanismos de verificación y validación para evitar errores, duplicados o información desactualizada.
- **OB.4-Facilitar la Navegación de Información en el BOE:** Se posibilitará la utilización y aplicación de algoritmos de grafos sobre los documentos del BOE creando relaciones entre sus publicaciones.
-

1.3. Retos del proyecto

Los retos que se abordan en este proyecto son tales como: el gran volumen de datos, la diversidad de la información del BOE, la actualización continua de la información, las necesidades de los usuarios finales...

Cada uno de estos retos representa un aspecto crítico que ha requerido de soluciones y estrategias para afrontarlos y lograr el éxito del proyecto:

- **R.1-Actualización continua:** La información del BOE se actualiza constantemente con nuevas publicaciones. Todos los días de la semana se publica el BOE, por lo que mantener la base de datos actualizada y en tiempo real implica desafíos en términos de velocidad y carga de almacenamiento.
- **R.2-Control de datos duplicados:** Se puede dar el caso de que se la herramienta descargue varias veces el mismo documento del BOE, dando lugar a duplicados y problemas de almacenamiento.
- **R.3-Manejo de grandes volúmenes de datos:** Entre los años 2018 y 2022 se realizaron anualmente un total de 101.061 publicaciones de disposiciones en el BOE, las cuales suelen tener una extensión en PDF de entre 1 y 60 páginas. Además, que los datos del BOE se obtengan en XML en vez de en PDF implica que, además del texto, también se tengan que almacenar las etiquetas XML, incrementando en gran medida la carga de almacenamiento.

1.4. Estructura del Documento

Este documento de memoria del TFG se organiza del siguiente modo:

Se ha comenzado con la sección de Introducción, la cual proporciona una descripción del contexto en el cual se enmarca el proyecto, destacando la relevancia y necesidad de abordar el tema investigado. Posteriormente se exponen los objetivos a cumplir del TFG, indicando lo que se pretende lograr como resultado del proyecto y se finaliza mencionando los desafíos y dificultades que se enfrentaron durante el desarrollo del proyecto.

En la siguiente sección se habla sobre los Trabajos Relacionados para conocer el funcionamiento y estructura del BOE y se realiza una revisión exhaustiva de la literatura existente, analizando trabajos previos, desarrollos tecnológicos actuales en el área y sus limitaciones.

Luego se continúa con la sección de Metodología, en la que se detallan los métodos, enfoques y procedimientos utilizados para llevar a cabo el proyecto, incluyendo la explicación detallada de lo que se ha hecho en cada etapa del desarrollo; análisis, diseño, implementación... Además, se incluye una explicación de los recursos generados, y de las herramientas y tecnologías utilizadas durante la ejecución del proyecto.

Finalmente se exponen las Conclusiones y reflexiones personales sobre el trabajo desarrollado, se continúa con una sección sobre el Trabajo Futuro, en la que se comentan los siguientes pasos que tomará el desarrollo realizado en este TFG. También en una última sección, se habla sobre el impacto y la Responsabilidad Social y Ambiental en la que se enmarca el proyecto.

Para concluir la memoria se incluye la sección de la bibliografía, en la que se define la lista de referencias bibliográficas que han sido utilizadas en la elaboración de la memoria y que han sido importantes para el desarrollo y éxito del proyecto.

Trabajos Relacionados

Para entender mejor el contexto del proyecto y conocer el estado actual del conocimiento en este campo, esta sección se va a utilizar para analizar y explicar el trabajo y proyectos previos en el campo de estudio de este TFG.

2.1 BOE.es

Como se ha comentado anteriormente, el BOE es el diario oficial de España en el que el estado publica todas las disposiciones de carácter general, como leyes, decretos, reglamentos disposiciones administrativas y actos que requieran ser publicados oficialmente para el conocimiento de la ciudadanía. El BOE es una herramienta fundamental para la transparencia y la legalidad, ya que garantiza que todas las normas y disposiciones que afectan a la ciudadanía sean de conocimiento público.

En la actualidad las herramientas que existen para consultar el BOE y poder realizar análisis sobre su contenido son escasas e incompletas.

La herramienta principal y oficial para consultar el Boletín Oficial del Estado (BOE) es la propia página del BOE (<https://www.boe.es/>), la cual pertenece a la Agencia Estatal del Boletín Oficial del Estado (perteneciente al Ministerio de la Presidencia, Justicia y Relaciones con las Cortes).

Diarios oficiales

 <p>Boletín Oficial del Estado</p> <ul style="list-style-type: none">Último BOEBuscar en BOECalendario	 <p>Boletín Oficial del Registro Mercantil</p> <ul style="list-style-type: none">Último BORMEBuscar en BORMECalendario	 <p>Otros diarios oficiales</p> <ul style="list-style-type: none">Unión Europea (DOUE)Boletines autonómicosBoletines provinciales
--	--	--

Información Jurídica

 <p>Todo el Derecho</p> <ul style="list-style-type: none">LegislaciónTribunal ConstitucionalJurisprudencia (CENDOJ)Doctrina de la FiscalíaDerechos fundamentalesDictámenes del Consejo de EstadoAbogacía del EstadoDerecho de la Unión EuropeaColección histórica: Gazeta (1661-1959) <p>Más búsquedas 🔍</p>	 <p>Biblioteca Jurídica Digital</p> <p>Publicaciones jurídicas para descargar de forma libre y gratuita:</p> <ul style="list-style-type: none">Libros jurídicos electrónicosCódigos electrónicosCódigos electrónicos universitariosMaterial de oposicionesAnuarios y revistasLa Editorial del BOELa Librería del BOE
--	---

Figura 2. Menú inicial BOE

El BOE se publica todos los días de la semana salvo los domingos. En caso de querer consultar la publicación del día se puede hacer accediendo al apartado de *Último BOE* que se muestra en la Figura 2.

Resultado a día 8 de Mayo de 2024:

Agencia Estatal Boletín Oficial del Estado

Castellano | Buscar | MI BOE | Menú

Está Ud. en Inicio > BOE > Calendario > 08/05/2024 - Sumario < anterior

Boletín Oficial del Estado: miércoles 8 de mayo de 2024, Núm. 112

Disposiciones y Anuncios | Notificaciones | Edictos Judiciales

Sumario BOE-S-2024-112: PDF XML

Secciones | Departamentos

I. Disposiciones generales

MINISTERIO DE LA PRESIDENCIA, JUSTICIA Y RELACIONES CON LAS CORTES

Terrorismo

Orden PJC/406/2024, de 7 de mayo, por la que se publica la Estrategia Nacional contra el Terrorismo 2023, aprobada por el Consejo de Seguridad Nacional.

PDF (BOE-A-2024-9149 - 22 págs. - 332 KB) | Otros formatos

Registro Mercantil. Modelos de cuentas anuales

Resolución de 23 de abril de 2024, de la Dirección General de Seguridad Jurídica y Fe Pública, referida a los modelos para la presentación en el Registro Mercantil de las cuentas anuales consolidadas de los sujetos obligados a su publicación.

PDF (BOE-A-2024-9150 - 2 págs. - 196 KB) | Otros formatos

Resolución de 23 de abril de 2024, de la Dirección General de Seguridad Jurídica y Fe Pública, referida a los modelos para la presentación en el Registro Mercantil de las cuentas anuales de los sujetos obligados a su publicación.

PDF (BOE-A-2024-9151 - 2 págs. - 195 KB) | Otros formatos

MINISTERIO DE DEFENSA

Control económico y financiero

Orden DEF/407/2024, de 25 de abril, por la que se aprueban las normas para la elaboración y seguimiento de los planes de acción del Departamento derivados del ejercicio del control financiero permanente y de la auditoría pública de la gestión económica-financiera del Ministerio de Defensa.

PDF (BOE-A-2024-9152 - 6 págs. - 220 KB) | Otros formatos

Figura 3. Último BOE 8 Mayo 2024

En la *Figura 3* se puede ver parte del contenido publicado por el BOE el día 8 de Mayo de 2024. Para un mejor entendimiento de este proyecto y del funcionamiento del BOE en general, es importante saber que el *Sumario BOE-S-2024-112* que se ve en la imagen, contiene el nombre de todos los documentos del BOE que se han publicado en ese día, en este caso 8/05/2024. Los nombres del BOE siguen este estilo: *BOE-A-2024-9149*, pero esto se explicará con mayor detalle en secciones posteriores.

Si se quiere consultar una publicación concreta, habría que hacer clic en el enlace de la publicación en la que estemos interesados. Si quisiésemos consultar la disposición del Ministerio de Defensa sobre Control económico y financiero, siguiendo la *Figura 3*, iríamos al enlace <https://www.boe.es/boe/dias/2024/05/08/pdfs/BOE-A-2024-9152.pdf>.

Otra forma para consultar las publicaciones del BOE es acceder al apartado de *Calendario* que se puede ver en la *Figura 2*, esto nos permitirá elegir cualquier fecha en la que se ha publicado el BOE y acceder a todo el contenido publicado ese día.

Pantalla del calendario a día 8 de Mayo de 2024:

Agencia Estatal Boletín Oficial del Estado

Castellano | Buscar | MI BOE | Menú

Está Vd. en > Inicio > BOE (Diario oficial Boletín Oficial del Estado)

Último BOE publicado (8 de mayo de 2024)

mayo 2024						
lun	mar	mié	jue	vie	sáb	dom
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

Calendario anual

Datos abiertos

Ayuda

Ir al BOE de fecha
dd/mm/aaaa

Buscar

Verificar documentos mediante CVE
BOE-

Verificar

Suplementos en [catalán](#) [gallego](#) [valenciano](#) [euskera](#)

Sección del [Tribunal Constitucional](#)

Figura 4. Elegir BOE por fecha

En cuanto a herramientas para realizar búsquedas filtradas sobre el BOE, la propia página oficial ofrece una serie de ellas. Estas son: accediendo al apartado de *buscar* que se muestra en la parte superior de la página (de la *Figura 4* por ejemplo) o accediendo al apartado de *Buscar en BOE* que aparece en la pantalla inicial (*Figura 2*).

En caso de elegir la primera opción nos aparecerá esta pantalla:

Buscar

Búsqueda rápida:

Todo el BOE | agua

Legislación

Todo el BOE

Notificaciones

Edictos

BOE

- Legislación
- Personal
- Otras Disposiciones
- Todo el BOE
- Últimas becas, ayudas y premios

Anuncios

- Administración de justicia en BOE
- Anuncios en BOE
- Anuncios en BORME

Suplementos

- Notificaciones
- Tablón Edictal Judicial Único

Otras bases de datos

- Jurisprudencia Constitucional
- Jurisprudencia (CENDOJ)
- Doctrina de la Fiscalía
- Derechos fundamentales
- Dictámenes del Consejo de Estado
- Abogacía del Estado

Suplementos cooficiales

- Catalán
- Gallego
- Valenciano
- Euskera

Figura 5. Búsqueda sencilla BOE

Esta opción nos permite buscar documentos publicados en el BOE filtrados por *Todo el BOE*, *Legislación*, *Notificaciones* o *Edictos* relacionados con la palabra que se ponga en el buscador.

Si se pone la palabra *agua* con el filtro de *Todo el BOE* daría este resultado:

Buscar en todo el BOE

Búsqueda **Resultados** Guardar búsqueda

Resultados 1 a 50 de 17.247

1 2 3 ... 345 [Pág. siguiente](#)

Ministerio para la Transición Ecológica y el Reto Demográfico
BOE 115 de 11/05/2024 - V. Anuncios

Anuncio de la Confederación Hidrográfica del Miño-Sil, O.A. de información pública del expediente de concesión para aprovechamiento de 0,67 l/s de agua procedente del manantial "Vilavella" sito en el lugar de Vilavella, parroquia de Vilavella (Santa María), término municipal de Triacastela (Lugo) para abastecimiento del núcleo Vilavella (doméstico y ganadero).A/27/36676.

[Más... \(Referencia BOE-B-2024-17130\)](#)

Otros Entes
BOE 114 de 10/05/2024 - V. Anuncios

Anuncio de la Comunidad de usuarios de agua subterráneas Goda sobre convocatoria de asamblea general de constitución.

[Más... \(Referencia BOE-B-2024-16829\)](#)

Ministerio de Ciencia, Innovación y Universidades
BOE 114 de 10/05/2024 - V. Anuncios

Anuncio de formalización de contratos de: Secretaría General del Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT). Objeto: Suministro de agua potable mediante fuentes dispensadoras para los edificios del CIEMAT en el Centro de Moncloa-Madrid. Expediente: 298774.

[Más... \(Referencia BOE-B-2024-16769\)](#)

Figura 6. Búsqueda filtrada simple

Pese a que este filtro pueda resultar muy útil, solo funciona con sustantivos que aparezcan de forma general en las publicaciones del BOE, por lo que es bastante limitada. En el caso de que queramos realizar una búsqueda en el apartado de *Buscar en BOE* (Figura 2), llegaremos a esta pantalla:

Buscar en todo el BOE

Disposiciones publicadas en las secciones I, II, y III desde 1960. Anuncios de la sección IV, desde 2009 y de la V desde 1960. Resoluciones del Tribunal Constitucional desde 1981. Para buscar los anuncios de notificación y los edictos judiciales consulte el apartado de la web correspondiente a [Notificaciones](#) y [Tablón Edictal Judicial Unico](#).

[Ayuda y contenido](#)

Sección

Sección I Sección II Sección III Sección IV Sección V Sección TC

Título Y

Departamento Y

Texto Y

Núm. de boletín Y

Núm. oficial Y

Fecha de publicación de a

Documentos por página Ordenados por fecha de publicación

Figura 7. Búsqueda filtrada avanzada

Dentro de las herramientas de búsqueda que proporciona la página oficial del BOE está es de las más completas y precisas, ya que te permite utilizar y combinar una varios filtros.

2.2 Mi BOE

La última herramienta oficial que ofrece la Agencia Estatal del Boletín Oficial del Estado es *Mi BOE* (https://www.boe.es/mi_boe/). Consiste en un servicio de alertas por suscripción gratuito que envía una notificación al correo electrónico del usuario cuando se publica un documento relacionado con una materia de interés para este, o cuando se actualiza alguna disposición, normativa... a la que nos hemos suscrito previamente.

Alertas legislativas

Legislación | Nombramientos, oposiciones y concursos | Anuncios de contratación | Temáticas

Ofrece las novedades legislativas publicadas en el Boletín Oficial del Estado y en el Diario Oficial de la Unión Europea

Seleccione las materias de su interés.

<input type="checkbox"/> Administración de Justicia	<input type="checkbox"/> Administración electrónica	<input type="checkbox"/> Agricultura
<input type="checkbox"/> Alimentación	<input type="checkbox"/> Asociaciones profesionales	<input type="checkbox"/> Asuntos sociales
<input type="checkbox"/> Comercio	<input type="checkbox"/> Consumidores y usuarios	<input type="checkbox"/> Cultura y ocio
<input type="checkbox"/> Deporte	<input type="checkbox"/> Derecho Administrativo	<input type="checkbox"/> Derecho Civil
<input type="checkbox"/> Derecho Constitucional	<input type="checkbox"/> Derecho Mercantil	<input type="checkbox"/> Derecho Penal
<input type="checkbox"/> Discapacidad	<input type="checkbox"/> Educación y enseñanza	<input type="checkbox"/> Energía
<input type="checkbox"/> Extranjería	<input type="checkbox"/> Función Pública	<input type="checkbox"/> Ganadería y animales
<input type="checkbox"/> Industria	<input type="checkbox"/> Medio ambiente	<input type="checkbox"/> Obras y construcciones
<input type="checkbox"/> Organización de la Administración	<input type="checkbox"/> Pesca	<input type="checkbox"/> Relaciones internacionales
<input type="checkbox"/> Sanidad	<input type="checkbox"/> Seguridad Social	<input type="checkbox"/> Seguridad y Defensa
<input type="checkbox"/> Sistema financiero	<input type="checkbox"/> Sistema tributario	<input type="checkbox"/> Tecnología e investigación
<input type="checkbox"/> Telecomunicaciones	<input type="checkbox"/> Trabajo y empleo	<input type="checkbox"/> Transportes y tráfico
<input type="checkbox"/> Turismo	<input type="checkbox"/> Unión Europea	<input type="checkbox"/> Vivienda y urbanismo

Figura 8. MiBOE seleccionar temas de interés

Publicado en: «BOE» núm. 282, de 25/11/1994.
Entrada en vigor: 01/01/1995
Departamento: Jefatura del Estado
Referencia: BOE-A-1994-26003
Permalink ELI: <https://www.boe.es/eli/es/l/1994/11/24/29/con>

PDF ePUB

Diccionario Jurídico RAE

Completo Solo Texto Índice Análisis

Alertas Jurisprudencia

Mi BOE

Recibirá una alerta por correo electrónico cada vez que se actualice el texto consolidado de esta disposición.

Recibir correo
 No seguir

Figura 9. MiBOE alerta

En conclusión, las herramientas oficiales del BOE proponen una serie de opciones para realizar búsquedas u operaciones sobre las publicaciones del BOE, pero estas se encuentran bastante limitadas. A pesar de ofrecer distintas herramientas de búsqueda, son todas muy similares, y a la hora de realizar búsquedas concretas la tarea se puede demorar más de lo esperado. Además, tanto las búsquedas como el sistema de alertas simplemente retornan el acceso a una publicación del BOE, lo que limita las posibilidades de realizar operaciones más complejas como propone este proyecto que, incluyendo la posibilidad de realizar búsquedas filtradas, también permite aplicar algoritmos de Data Science sobre el BOE gracias a la formación de grafos.

2.3 Herramientas no oficiales:

Además de las herramientas oficiales que ofrece la Agencia Estatal del Boletín Oficial del Estado para realizar búsquedas y operaciones sobre el BOE, contamos con los siguientes desarrollos que abordan este tema:

2.3.1 Aplicación Móvil Boletín Oficial de España

Se trata de una aplicación que permite la consulta del BOE desde dispositivos móviles. Permite consultar el último diario disponible, buscar un diario en una fecha seleccionada, filtrar los artículos del diario según su sección, buscar artículos según su rango, así como añadir y eliminar artículos en una lista de favoritos.

Es una aplicación que no aporta ninguna funcionalidad nueva respecto a la página oficial del BOE. Además, como se puede ver en los comentarios y opiniones de la aplicación, esta no funciona correctamente, no se actualiza diariamente y no ofrece una buena usabilidad al usuario.



Figura 10. App móvil BOE

2.3.2 JurisTracker - App BOE

Consiste en una aplicación móvil que permite la consulta del BOE. Esta no cuenta con un sistema de búsquedas filtradas, pero permite a los usuarios consultar un BOE en específico, visualizar los cambios y realizar comparaciones sobre distintas versiones de una misma publicación. También incluye un sistema de alertas por suscripción similar al de *Mi BOE*.

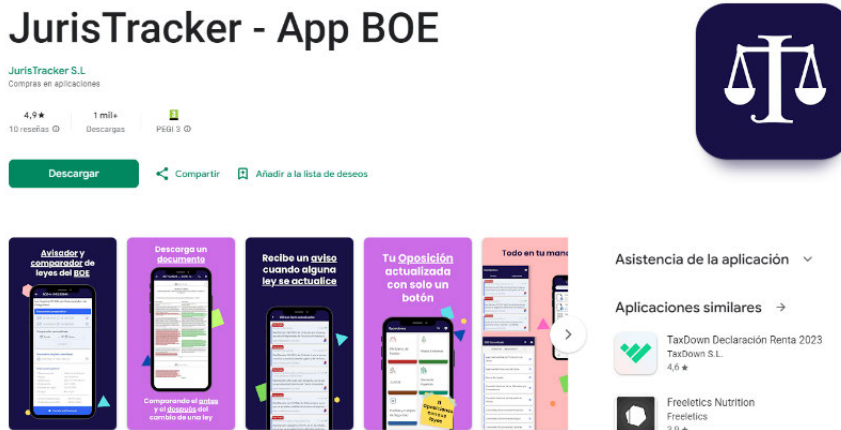


Figura 11. JurisTracker App BOE

2.3.3 SuperBOE

SuperBOE es una página web (<https://superboe.es/>) que permite al usuario seleccionar una publicación del BOE para que este luego elija un formato de visualización para dicha publicación. La página web cambia el formato original del BOE modificando su tipografía, márgenes, interlineado... para una lectura más cómoda, permitir realizar apuntes o añadir datos personalizados.

Esta herramienta surge de querer facilitar y agilizar la lectura del BOE a juristas u otros profesionales que tengan que consistentemente leerse las publicaciones del BOE para sus propósitos o trabajos. Según indica el autor de esta herramienta, gracias a SuperBOE se podría reducir el número de páginas de lectura de cada publicación en un 5%.



Figura 12. Super BOE

Metodología

El desarrollo de la herramienta se estructuró en varias fases, cada una con objetivos específicos y un enfoque metodológico claro. Se utilizó una metodología tradicional, en cascada, en la que se realizaron las etapas de análisis, diseño, implementación, pruebas y despliegue.

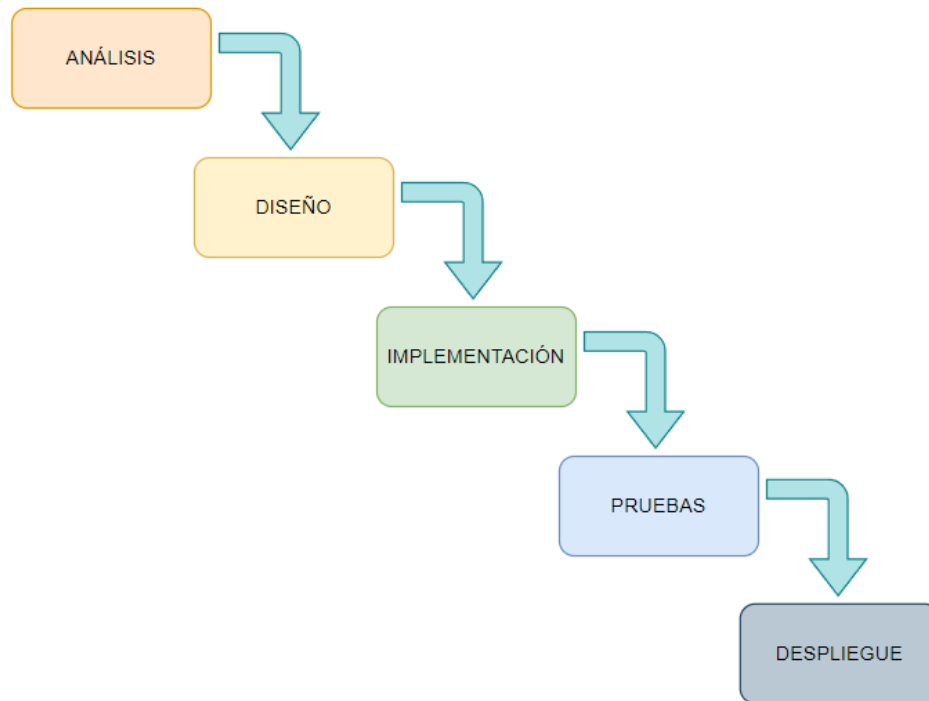


Figura 13. Metodología seguida

3.1. Análisis

Se comenzó por una fase de análisis, en la que se estudió la problemática en detalle para definir los objetivos del proyecto, incluyendo un estudio detallado del estado del arte. Esta fase fue crucial para definir los objetivos del proyecto, los posibles desafíos técnicos y establecer una base sólida sobre la cual diseñar una solución.

Una vez definidos los objetivos y los retos, se realizó una evaluación sobre estos para poder organizar el desarrollo del proyecto, estudiar las tecnologías a utilizar y poder comenzar con el diseño funcional de la herramienta.

Los objetivos y retos del proyecto están definidos al principio del documento.

Estudio de tecnologías a utilizar:

El primer problema al que había que enfrentarse era la obtención de las publicaciones del BOE para poder explotar sus datos. La manera más eficiente de obtener el BOE es a través de la página web oficial (<https://www.boe.es/>), ya que esta permite descargar el BOE en diferentes formatos: PDF, EPUB y XML. De estos formatos el que nos interesa es el XML, ya que este formato es

estructurado y legible por máquina, lo que facilita la extracción y el procesamiento automatizado de la información contenida en los documentos.

El formato XML permite acceder directamente a los diferentes campos de metadatos y contenido del BOE, tales como títulos, fechas, departamentos, y el texto de las disposiciones legales, de manera más eficiente y precisa en comparación con los formatos PDF y EPUB, que están más orientados a la presentación visual del documento.

Una vez sabemos de dónde y en que formato obtener el BOE necesitamos almacenarlo para poder operar sobre sus datos. Entre las opciones de sistemas de bases de datos existentes, hay que considerar si es mejor para nuestro caso una base de datos relacional o una NoSQL.

Existen bases de datos relacionales como MySQL o PostgreSQL que ofrecen soporte nativo para datos XML, permiten realizar consultas SQL sobre los datos que se almacenen utilizando funciones y métodos específicos para XML, además las bases de datos relacionales son muy adecuadas a la hora de realizar consultas complejas y tienen un ecosistema de soporte más amplio al llevar más tiempo en uso y desarrollo.

Por otro lado, tenemos las bases de datos NoSQL que ofrecen un modelo de datos más flexible que las BBDDs relacionales y no se necesita ningún tipo de conversión para almacenar archivos XML. Además, cuentan con lenguajes de consultas específicos que son muy potentes y fáciles de usar, y existen BBDDs NoSQL como MongoDB que tienen una gran escalabilidad horizontal permitiendo añadir nuevos servidores y manejar así el crecimiento de datos.

Una vez exploradas las distintas opciones de bases de datos, se eligió para este proyecto una base de datos NoSQL, MongoDB, ya que esta ofrece un lenguaje de consultas potente y fácil de usar, es altamente escalable horizontalmente (ayudando así con el reto de los posibles problemas de almacenamiento), y es fácil de utilizar con el lenguaje de programación elegido para este proyecto (Java).

Con los datos almacenados en MongoDB ya podemos explotarlos y realizar consultas sobre las publicaciones del BOE, pero aún nos falta una herramienta para poder añadir la posibilidad de aplicar algoritmos e IA sobre el BOE de manera eficiente. Para ello utilizaremos otra base de datos NoSQL, Neo4j, en la que podremos almacenar ciertos datos concretos de las publicaciones del BOE, para así agilizar consultas rápidas y permitir la realización de consultas más complejas que en MongoDB.

La utilización de Neo4j nos va a abrir también una gran serie de posibilidades y aplicaciones. Al tratarse de una base de datos orientada a grafos nos permite relacionar publicaciones del BOE entre sí y posibilitar la aplicación de algoritmos de grafos. Además, Neo4j está compuesta por un conjunto de aplicaciones y herramientas como *Neo4j Bloom*¹ que da al usuario una visualización amigable de los datos, pero también muy completa para usuarios más expertos en el análisis de datos. Neo4j también cuenta con *Neo4j Graph Data Science*², que es la herramienta que nos va a permitir aplicar algoritmos, técnicas de datos conectados y la posibilidad de utilizar modelos de machine learning gracias a su biblioteca integrada.

¹ <https://neo4j.com/product/bloom/?ref=product>

² <https://neo4j.com/product/graph-data-science/?ref=product>

Finalmente, con el conocimiento de las tecnologías a utilizar y con una idea preliminar de los pasos a seguir para la construcción de la herramienta, se terminó la etapa de análisis y dio comienzo la fase de diseño.

3.2. Diseño

En esta etapa se diseñó un diagrama que ilustra los módulos que componen la herramienta y el orden en el que se pretendía desarrollarlos. La realización de este diagrama de módulos se utilizó también como un workflow, es decir, se descompuso el proyecto en fases y módulos específicos, proporcionando una visión clara y organizada del desarrollo. De esta manera se facilitó la gestión del proyecto, permitiendo una asignación más precisa de recursos y un seguimiento eficaz del progreso.

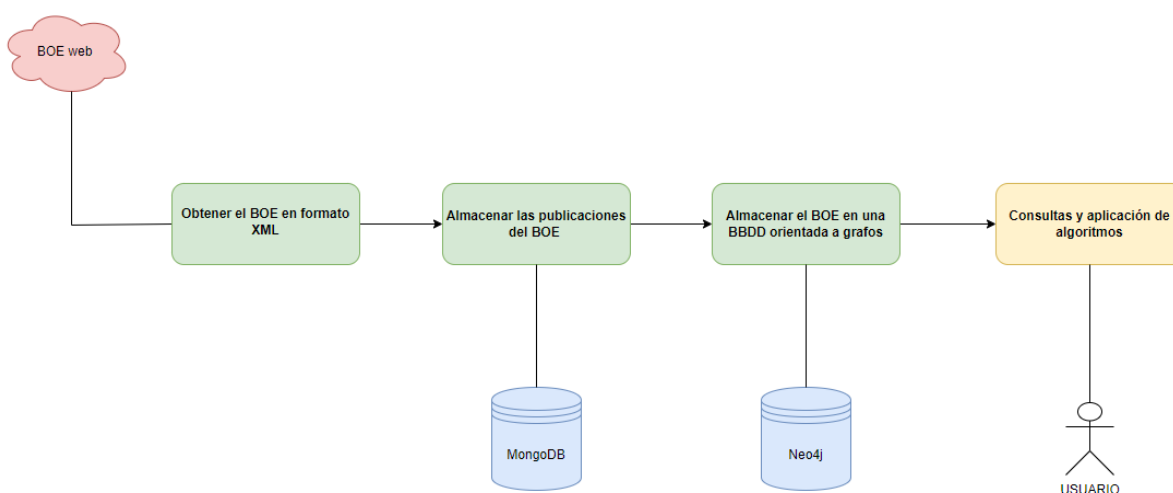


Figura 14. Workflow

El diagrama destaca los siguientes módulos principales:

- **Obtención del BOE:** Este módulo se encarga de la extracción del Boletín Oficial del Estado (BOE) desde su web en formato XML.
- **Almacenamiento en NoSQL:** Los datos XML se transformaron en JSON y se almacenaron en una base de datos NoSQL, específicamente MongoDB. Este módulo se encargó de la organización y gestión del almacenamiento de datos, asegurando un acceso rápido y eficiente.
- **Integración con Neo4j:** Se desarrolló un módulo para integrar los datos almacenados en MongoDB con la base de datos de grafos Neo4j. Este paso fue crucial para permitir análisis complejos y consultas avanzadas sobre los datos del BOE.
- **Generación de Informes y Análisis:** El último módulo es la herramienta resultante del desarrollo del proyecto, con la que el usuario interactúa y puede ejecutar consultas, aplicar algoritmos y realizar análisis basados en los datos almacenados. También se pueden visualizar ciertos resultados gracias a la interfaz de Neo4j, y en un futuro todos los datos gracias a una interfaz gráfica.

Para terminar con esta sección, se concluye que la aplicación de una metodología en cascada permitió un desarrollo estructurado y controlado del proyecto que, facilitó no solo alcanzar los objetivos del proyecto, sino también enfrentar y resolver los retos técnicos y operativos que surgieron durante el desarrollo. El buen planteamiento y análisis del problema permitió establecer una base sólida sobre la cual se pudo diseñar una solución efectiva, eficiente y que permitió trazar el progreso.

3.3. Implementación

Este apartado abarca la implementación práctica del proyecto, donde se detallan los procesos, técnicas y herramientas empleadas para convertir el diseño de la solución en un sistema funcional. En esta sección se describe de manera exhaustiva cómo se llevaron a cabo las tareas de desarrollo de cada uno de los módulos que componen la herramienta final.

Como se ha explicado en el punto anterior, el desarrollo del proyecto se estructuró en varias fases, siguiendo el orden del diagrama de la *figura 14*. En la fase de implementación también se abordaron aspectos críticos como el control de duplicados, la actualización continua y demás retos.

Es importante saber que la implementación de la herramienta a utilizado como lenguaje de programación Java, lenguaje elegido por contar con una amplia variedad de librerías y frameworks, que han sido de gran ayuda.

3.3.1. Obtención del BOE:

El primer módulo para implementar es el de la obtención del BOE en formato XML. Como para poder aprovechar la herramienta es necesario tener publicaciones del BOE almacenadas, y descargar todas las publicaciones existentes del BOE supondría un coste inicial en tiempo y recursos enorme, se decidió que la herramienta descargaría de forma automática todas las publicaciones del BOE que se publicasen diariamente, es decir, cada día de la semana que se publica el BOE (menos los domingos salvo casos extraordinarios) la herramienta descargaría todo el contenido publicado relativo a ese día. Así ya existe un punto de partida sobre el que poder empezar el desarrollo de dicho módulo.

Para la obtención del BOE será necesario descargar el XML a través de la página web oficial del BOE. Todos los enlaces de las publicaciones del BOE cuentan con la misma estructura, así que lo aprovecharemos para que nuestro módulo en Java realice una conexión http por cada publicación del BOE, siguiendo la estructura fija que propone la página oficial.

La estructura de las URL es la siguiente:

[https://boe.es/diario_boe/xml.php?id=\[PUB\]-\[I\]-\[FECHA\]](https://boe.es/diario_boe/xml.php?id=[PUB]-[I]-[FECHA])

- [PUB]. Tipo de boletín. En nuestro caso será BOE.
- [I]. Tipo de ítem. Siempre que es S se trata de un sumario.
- [FECHA]. Fecha de la publicación en formato AAAAMMDD

Ejemplo de enlace del sumario publicado el 6 de octubre de 2014:
https://boe.es/diario_boe/xml.php?id=BOE-S-20141006

Contenido de dicho sumario:

```
<sumario>
  <meta>
    <pub>BOE</pub>
    <anno>2014</anno>
    <fecha>06/10/2014</fecha>
    <fechaInv>2014/10/06</fechaInv>
    <fechaAnt>04/10/2014</fechaAnt>
    <fechaAntAnt>03/10/2014</fechaAntAnt>
    <fechaSig>07/10/2014</fechaSig>
    <fechaPub>lunes 6 de octubre de 2014</fechaPub>
    <pubDate>Mon, 06 Oct 2014 00:00:00 +0200</pubDate>
  </meta>
  <diario nbo="242">
    <sumario_nbo id="BOE-S-2014-242">
      <urlPdf szBytes="313838" szKBytes="306"/>/boe/dias/2014/10/06/pdfs/BOE-S-2014-242.pdf</urlPdf>
    </sumario_nbo>
    <seccion num="1" nombre="I. Disposiciones generales">
      <departamento nombre="MINISTERIO DE ASUNTOS EXTERIORES Y DE COOPERACIÓN" etq="7120">
        <epigrafe nombre="Organización">
          <item id="BOE-A-2014-10113" control="2014/9990">
            <titulo>Orden AEC/1806/2014, de 15 de septiembre, por la que se crea una Oficina Consular Honoraria de España en el Estado de Nevada (Estados Unidos).</titulo>
            <urlPdf szBytes="137464" szKBytes="134" numPag="1"/>/boe/dias/2014/10/06/pdfs/BOE-A-2014-10113.pdf</urlPdf>
            <urlHtm>/diario_boe/txt.php?id=BOE-A-2014-10113</urlHtm>
            <urlXml>/diario_boe/xml.php?id=BOE-A-2014-10113</urlXml>
          </item>
        </epigrafe>
      </departamento>
    </seccion>
    <seccion num="2A" nombre="II. Autoridades y personal. - A. Nombramientos, situaciones e incidencias">
      <departamento nombre="CONSEJO GENERAL DEL PODER JUDICIAL" etq="1820">
        <epigrafe nombre="Adscripciones">
          <item id="BOE-A-2014-10114" control="2014/9994">
            <titulo>Acuerdo de 23 de septiembre de 2014, de la Comisión Permanente del Consejo General del Poder Judicial, por el que se adscribe al Magistrado don Juan Ma
            Navarra.</titulo>
            <urlPdf szBytes="134375" szKBytes="131" numPag="1"/>/boe/dias/2014/10/06/pdfs/BOE-A-2014-10114.pdf</urlPdf>
            <urlHtm>/diario_boe/txt.php?id=BOE-A-2014-10114</urlHtm>
            <urlXml>/diario_boe/xml.php?id=BOE-A-2014-10114</urlXml>
          </item>
        </epigrafe>
        <epigrafe nombre="Situaciones">
          <item id="BOE-A-2014-10115" control="2014/9995">
            <titulo>Acuerdo de 23 de septiembre de 2014, de la Comisión Permanente del Consejo General del Poder Judicial, por el que queda a disposición del Presidente de
            </titulo>
            <urlPdf szBytes="134631" szKBytes="131" numPag="1"/>/boe/dias/2014/10/06/pdfs/BOE-A-2014-10115.pdf</urlPdf>
            <urlHtm>/diario_boe/txt.php?id=BOE-A-2014-10115</urlHtm>
            <urlXml>/diario_boe/xml.php?id=BOE-A-2014-10115</urlXml>
          </item>
          <item id="BOE-A-2014-10116" control="2014/9996">
            <titulo>Acuerdo de 23 de septiembre de 2014, de la Comisión Permanente del Consejo General del Poder Judicial, por el que se declara en la situación administra
            Gómez Duarros.</titulo>
            <urlPdf szBytes="134542" szKBytes="131" numPag="1"/>/boe/dias/2014/10/06/pdfs/BOE-A-2014-10116.pdf</urlPdf>
            <urlHtm>/diario_boe/txt.php?id=BOE-A-2014-10116</urlHtm>
            <urlXml>/diario_boe/xml.php?id=BOE-A-2014-10116</urlXml>
          </item>
        </epigrafe>
      </departamento>
    </seccion>
  </diario nbo="242">
  <departamento nombre="MINISTERIO DEL INTERIOR" etq="7320">
```

Figura 15. Ejemplo de sumario

El sumario es una de las publicaciones diarias que realiza el BOE y la que más nos interesa de todas para obtener los documentos del BOE, pues se trata de un documento en el que se incluyen todas las disposiciones que se han publicado en esa fecha, por lo que recorriendo dicho documento podremos descargar todas las disposiciones del BOE publicadas en ese día.

Como descargamos el sumario en formato XML, nos interesa la etiqueta `<urlXml>`, que es en donde se encuentran las URLs de todas las publicaciones del día. Por lo que haciendo uso de XPATH³, un lenguaje que permite construir expresiones para recorrer y procesar documentos XML, lo usaremos para localizar la etiqueta que nos interesa (`<urlXml>`), obtener el contenido de esta y almacenarlo en una lista de URLs.

El módulo una vez tenga la lista con las URLs simplemente realizará la conexión web y descargará cada publicación individualmente para ser almacenada.

³ <https://www.mclibre.org/consultar/xml/lecciones/xml-xpath.html>

La implementación del método y su ejecución a día 29 de Mayo de 2024 daría el siguiente resultado:









 BOE-A-2024-10776	29/05/2024 13:07	Microsoft Edge H...	3 KB
 BOE-A-2024-10761	29/05/2024 13:07	Microsoft Edge H...	40 KB
 BOE-A-2024-10762	29/05/2024 13:07	Microsoft Edge H...	110 KB
 BOE-A-2024-10763	29/05/2024 13:07	Microsoft Edge H...	97 KB
 BOE-A-2024-10764	29/05/2024 13:07	Microsoft Edge H...	28 KB
 BOE-A-2024-10765	29/05/2024 13:07	Microsoft Edge H...	636 KB
 BOE-A-2024-10766	29/05/2024 13:07	Microsoft Edge H...	136 KB
 BOE-A-2024-10767	29/05/2024 13:07	Microsoft Edge H...	59 KB

Figura 16. Almacenamiento XML

En total se han descargado 216 archivos:



Figura 17. Número XML descargados

Si buscamos cuántas apariciones de la etiqueta `<urlXml>` existen en el sumario haciendo uso de Ctrl + F para saber cuántos documentos del BOE se han publicado ese día, nos sale 216, el mismo número de publicaciones descargadas. La / de la búsqueda se pone porque las etiquetas de XML tienen inicio y cierre similar, y aparecería el número de resultados duplicado.

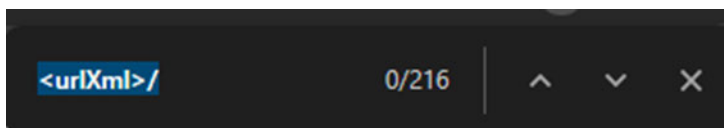


Figura 18. N° publicaciones del sumario

3.3.2. Almacenamiento del BOE en MongoDB:

El siguiente módulo consiste en el almacenamiento del BOE XML descargado en la base de datos NoSQL MongoDB. Para la implementación de este módulo se decidió convertir el BOE XML a formato JSON. A pesar de que se puede almacenar un archivo XML directamente en MongoDB, la conversión a JSON no añade una complejidad significativa al módulo y permite almacenar los datos de una manera más “limpia” en MongoDB ya que JSON es el formato de datos nativo de esta base de datos, lo que optimiza las tareas de almacenamiento y consulta de datos.

También gracias a JSON se reduce ligeramente la carga de almacenamiento por la eliminación de las etiquetas XML y se facilita el intercambio de datos con otras aplicaciones, pues JSON es mucho más compatible con otros lenguajes y herramientas.

Los campos que va a tener cada documento almacenado en MongoDB son los mismos que existen en el documento original del XML del BOE, ya que cualquiera de estos puede ser de utilidad más adelante para realizar búsquedas filtradas o relacionar publicaciones entre sí, por lo que no se va a despreciar ningún campo.

Ya que este módulo se encarga del almacenamiento de las publicaciones en MongoDB, tenemos que enfrentarnos a la existencia de duplicados:

Todas las publicaciones del BOE cuentan con un identificador único, que también se va a utilizar como su nombre, y sigue esta estructura fija: [PUB]-[I]-[AAAA]-[NNN].

- [PUB]. Tipo de boletín. En nuestro caso siempre será BOE.
- [I]. Tipo de ítem.
- [AAAA]. Año de publicación del boletín.
- [NNN]. Número de boletín junto con el que se publica.

Como todas las publicaciones cuentan con un identificador único y además la estructura de este es fija, se creó un índice sobre el campo \$Identificador en MongoDB de la siguiente forma:
`db.publicaciones.createIndex({ identificador: 1 })`

Con el índice creado, el módulo comprobará mediante una consulta de mongo (`db.publicaciones.find({ identificador: "$id"}`) si en la colección en la que estamos almacenando el BOE existe algún documento con el mismo identificador ya almacenado. En caso de que la publicación aún no exista en la bbdd, se inserta en la colección, pero si ya existe no se inserta nada.

Tanto para el proceso de búsqueda del identificador como para el de insertar el documento en MongoDB, el módulo realiza una conexión a una base de datos creada previamente. Por el funcionamiento específico de MongoDB se ha creado primero una base de datos con el nombre de *boe* y dentro de ella una colección llamada *publicaciones* en la que almacenamos todas las publicaciones que descargamos del BOE.

En este proyecto para la implementación de las bases de datos se ha utilizado Docker⁴: Docker es una herramienta para la virtualización de aplicaciones basada en contenedores, en la que implementas imágenes de aplicaciones aisladas en cualquier sistema que ejecute Docker, es decir, puedes empaquetar una aplicación y sus dependencias en un contenedor que al ejecutarse es como si ejecutaras la aplicación en tu sistema. El objetivo de usar Docker en este proyecto se debe a las facilidades que aporta esta herramienta al proceso de desarrollo, implementación y pruebas, y a la reproducibilidad que ofrece para ejecutar esta aplicación en distintos entornos.

Análisis de carga de almacenamiento XML vs JSON:

Como se ha mencionado anteriormente, la eliminación de las etiquetas XML supone un alivio en cuanto a carga de almacenamiento, pero hay que saber si la transformación del formato de datos merece la pena. Para ello vamos a convertir aleatoriamente algunas de las publicaciones del BOE que hemos descargado con anterioridad a JSON para comparar su tamaño.









Nombre	Tipo	Tamaño
 BOE-A-2024-10416	Archivo JSON	140 KB
 BOE-A-2024-10416	Microsoft Edge H...	172 KB
 BOE-A-2024-10426	Archivo JSON	470 KB
 BOE-A-2024-10426	Microsoft Edge H...	539 KB
 BOE-A-2024-10761	Archivo JSON	36 KB
 BOE-A-2024-10761	Microsoft Edge H...	40 KB
 BOE-A-2024-10763	Archivo JSON	93 KB
 BOE-A-2024-10763	Microsoft Edge H...	97 KB

Figura 19. XML vs JSON

⁴ <https://www.docker.com/>

Como se ve en la imagen, para una misma publicación el tamaño en formato XML es superior al de JSON. La diferencia media de tamaño para estas publicaciones del BOE es:

$$\frac{\left(1 - \frac{140}{172}\right) \cdot 100 + \left(1 - \frac{470}{539}\right) \cdot 100 + \left(1 - \frac{36}{40}\right) \cdot 100 + \left(1 - \frac{93}{97}\right) \cdot 100}{4} = 11,38\%$$

Es decir, gracias a la conversión de XML a JSON de estas publicaciones ahorramos un 11,38% de carga de almacenamiento (para estas publicaciones).

La diferencia de tamaño entre ambos formatos reside en la extensión de cada publicación, cuanto más larga sea, más etiquetas XML tendrá y mayor será el beneficio de la conversión a JSON.

Comparación del XML y el JSON del BOE-A-2024-10761:

```

▼<documento fecha_actualizacion="20240529115601">
  ▼<metadatos>
    <identificador>BOE-A-2024-10761</identificador>
    <origen_legislativo codigo="1">Estatal</origen_legislativo>
    <departamento codigo="9562">Ministerio de Asuntos Exteriores, Unión Europea y Cooperación</departamento>
    <rango codigo="1180">Acuerdo Internacional</rango>
    <fecha_disposicion>20211210</fecha_disposicion>
    <numero_oficial/>
    <titulo>Modificaciones de los anexos I, II y IV del Protocolo para la protección del Mar Mediterráneo con
    <diario codigo="BOE">Boletín Oficial del Estado</diario>
    <fecha_publicacion>20240529</fecha_publicacion>
    <diario_numero>130</diario_numero>
    <seccion>1</seccion>
    <subseccion/>
    <pagina_inicial>61912</pagina_inicial>
    <pagina_final>61920</pagina_final>
    <suplemento_pagina_inicial/>
    <suplemento_pagina_final/>
    <url_pdf>/boe/dias/2024/05/29/pdfs/BOE-A-2024-10761.pdf</url_pdf>
    <url_epub>/diario_boe/epub.php?id=BOE-A-2024-10761</url_epub>
  
```

Figura 20. XML de ejemplo

```

{"fecha_actualizacion":"20240529115601","metadatos":{"identificador":"BOE-A-2024-10761","origen_legislativo":{"codigo":"1","":"Estatal"},"departamento":{"codigo":"9562","":"Ministerio de Asuntos Exteriores, Unión Europea y Cooperación"},"rango":{"codigo":"1180","":"Acuerdo Internacional"},"fecha_disposicion":"20211210","numero_oficial":"","titulo":"Modificaciones de los anexos I, II y IV del Protocolo para la protección del Mar Mediterráneo contra la contaminación causada por fuentes y actividades situadas en tierra, adoptadas en Antalya el 10 de diciembre de 2021.","diario":{"codigo":"BOE","":"Boletín Oficial del Estado"},"fecha_publicacion":"20240529","diario_numero":"130","seccion":"1","subseccion":"","pagina_inicial":"61912","pagina_final":"61920","suplemento_pagina_inicial":"","suplemento_pagina_final":"","url_pdf":"/boe/dias/2024/05/29/pdfs/BOE-A-2024-10761.pdf","url_epub":"/diario_boe/epub.php?id=BOE-A-2024-10761","url_pdf_catalan":"","url_pdf_euskera":"","url_pdf_gallego":"","url_pdf_valenciano":"","estatus_legislativo":"L","fecha_vigencia":"20230121","est
  
```

Figura 21. JSON de ejemplo

3.3.3. Integración con Neo4j:

Este módulo sirve para transferir el contenido de las publicaciones del BOE almacenadas en MongoDB a Neo4j. En este caso, no se almacenará el contenido al completo de las publicaciones del BOE, pues este ya se encuentra en MongoDB y sería duplicar información y almacenamiento de forma innecesaria. Por ello solo se almacenará el identificador y la metainformación de las publicaciones del BOE en Neo4j, para así facilitar búsquedas filtradas y el resto de las funcionalidades.

La metainformación del BOE son datos tales como el título del BOE, el origen legislativo, el rango y el departamento.

Aquí un ejemplo sencillo de una publicación del BOE para ver el contenido de la metainformación:

```

<documento fecha_actualizacion="20181023201505">
  <metadatos>
    <identificador>BOE-A-2014-10115</identificador>
    <origen_legislativo codigo="1">Estatal</origen_legislativo>
    <departamento codigo="1020">Consejo General del Poder Judicial</departamento>
    <rango codigo="1020">Acuerdo</rango>
    <fecha_disposicion>20140923</fecha_disposicion>
    <numero_oficial/>
    <título>Acuerdo de 23 de septiembre de 2014, de la Comisión Permanente del Consejo General del Poder Judicial, por el que queda a disposición del Presidente del Tribunal Superior de Justicia de Andalucía el Magistrado don Álvaro Salvador Prieto.</título>
    <diario codigo="BOE">Boletín Oficial del Estado</diario>
    <fecha_publicacion>20141006</fecha_publicacion>
    <diario_numero>242</diario_numero>
    <seccion/>
    <subseccionA</subseccion>
    <pagina_inicial>79683</pagina_inicial>
    <pagina_final>79683</pagina_final>
    <suplemento_pagina_inicial/>
    <suplemento_pagina_final/>
    <url_pdf>/boe/dias/2014/10/06/pdf/BOE-A-2014-10115.pdf</url_pdf>
    <url_epub/>
    <url_pdf_catalan/>
    <url_pdf_euskera/>
    <url_pdf_gallego/>
    <url_pdf_valenciano/>
    <estatus_legislativo/>
    <fecha_vigencia/>
    <estatus_derogacionN</estatus_derogacion>
    <fecha_derogacion/>
    <judicialmente_anuladaN</judicialmente_anulada>
    <fecha_anulacion/>
    <vigencia_agotadaN</vigencia_agotada>
    <estado_consolidacion codigo="0"/>
    <letra_imagenA</letra_imagen>
    <suplemento_letra_imagen/>
  </metadatos>
  <analisis>
    <materias/>
    <notas/>
    <referencias>
      <anteriores/>
      <posteriores/>
    </referencias>
    <letras/>
  </analisis>
  <texto>
    <p class="parrafo">De conformidad con lo dispuesto en el artículo 355.bis.2 de la Ley Orgánica 6/1995, de 1 de julio, del Poder Judicial, la Comisión Permanente del Consejo General del Poder Judicial, en su reunión del día 23 de septiembre de 2014, ha acordado que el Magistrado don Álvaro Salvador Prieto, quede a disposición del Presidente del Tribunal Superior de Justicia de Andalucía, a partir del día en que haga efectiva la reincorporación el Magistrado don Mario Vicente Alonso Alonso a la plaza reservada, determinando la Sala de Gobierno los puestos en que ha de prestar servicios mientras permanezca en dicha situación y ocupando vacante que se produzca en los Juzgados del mismo orden jurisdiccional del lugar de la plaza reservada.</p>
    <p class="parrafo_2">Madrid, 23 de septiembre de 2014.-El Presidente del Consejo General del Poder Judicial, Carlos Lesmes Serrano.</p>
  </texto>
</documento>

```

Figura 22. Boe sencillo

El módulo de integración con Neo4j trabaja junto con el de almacenamiento en MongoDB, por lo que cada vez que se almacene una publicación del BOE en MongoDB también se almacenará (su metainformación) en Neo4j. De la misma forma que cuando se quiere almacenar una publicación que ya existe en la base de datos de MongoDB, esta no deja, tampoco se almacenará en Neo4j. Siguiendo el procedimiento de almacenar primero en MongoDB, para que después de verificar que no existe ningún duplicado, almacenar también en Neo4j, se controlan los duplicados en ambas bases de datos sin necesidad de añadir ningún nuevo procedimiento ni una mayor complejidad.

En resumen, la herramienta cuando requiere almacenar una publicación del BOE en Neo4j también lo debe hacer en MongoDB para tener el contenido completo de la publicación, así siempre se pasa por el control de duplicados y no es necesario controlarlos adicionalmente en Neo4j.

Uno de los retos que este módulo ha tenido que enfrentar es la existencia de nulos, pues solo los campos de título e identificador son campos obligatorios para todas las publicaciones del BOE, el resto de los campos que almacenamos no siempre aparecen en las publicaciones, por lo que hay que controlar los nulos para mantener la integridad de los datos y evitar problemas en la inserción.

Para dar solución a este problema los campos que sean nulos se detectan previamente a ser insertados en las bases de datos mediante la búsqueda con XPATH del contenido de las etiquetas XML de dichos campos; `<rango>`, `<departamento>`... Y se sustituyen por un mensaje del tipo “No existe un *rango* para esta publicación”.

Para todo el proceso de almacenamiento en Neo4j se han utilizado sentencias en Cypher⁵, el lenguaje de consultas nativo de Neo4j.

⁵ <https://neo4j.com/docs/cypher-manual/current/introduction/>

3.4. Pruebas

En esta sección se presenta un resumen de los resultados obtenidos tras la realización de pruebas funcionales sobre la implementación del proyecto. Además de evaluar si se han cumplido los objetivos.

Lo primero es saber dónde se van a mostrar los resultados, es decir, dónde vamos a comprobar que los módulos funcionan correctamente. Para la obtención del BOE, se mostrarán los resultados en los archivos locales del ordenador en el que se ha realizado el proyecto y así comprender mejor el proceso de obtención y transformación de los datos. Finalmente, para el resto de los resultados se utilizará la consola de MongoDB y la interfaz gráfica de Neo4j.

3.4.1 Obtención del BOE:

La obtención del BOE se implementó mediante un método en java, *DescargarBOE()*, y se siguió un proceso en el que primero se obtenía el BOE en formato XML, luego se transformaba a JSON para almacenarlo en MongoDB y se integraba parte de ese documento en Neo4j (la metainformación). Además, este método funciona descargando el BOE diario.

Para mostrar la obtención del BOE vamos a partir desde el punto en el que se explicó y mostró la obtención del BOE en el apartado de su [implementación](#).

El resultado de la ejecución a día 29 de Mayo de 2024 fue el siguiente:









 BOE-A-2024-10776	29/05/2024 13:07	Microsoft Edge H...	3 KB
 BOE-A-2024-10761	29/05/2024 13:07	Microsoft Edge H...	40 KB
 BOE-A-2024-10762	29/05/2024 13:07	Microsoft Edge H...	110 KB
 BOE-A-2024-10763	29/05/2024 13:07	Microsoft Edge H...	97 KB
 BOE-A-2024-10764	29/05/2024 13:07	Microsoft Edge H...	28 KB
 BOE-A-2024-10765	29/05/2024 13:07	Microsoft Edge H...	636 KB
 BOE-A-2024-10766	29/05/2024 13:07	Microsoft Edge H...	136 KB
 BOE-A-2024-10767	29/05/2024 13:07	Microsoft Edge H...	59 KB

Figura 23. Almacenamiento XML

En total se han descargado 216 archivos:



Figura 24. Número XML descargados

Buscando cuantas apariciones de la etiqueta `<urlXml>` existen para saber cuántos documentos del BOE se publicaron ese día, y por lo tanto tenemos que descargar, nos sale 216, el mismo número de publicaciones descargadas.

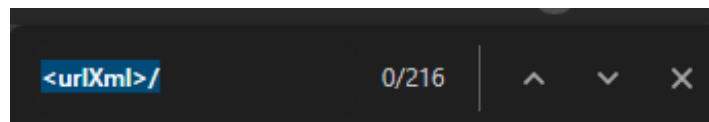


Figura 25. N° publicaciones del sumario

3.4.2 Almacenamiento en MongoDB:

Una vez ya hemos visto los resultados y el funcionamiento de la obtención del BOE, queda su almacenamiento. Empezaremos por el almacenamiento del BOE en la base de datos de MongoDB. Para poder utilizar MongoDB tenemos que conectarnos primero a la base de datos desde la consola.

```
Current Mongosh Log ID: 66576b4f84ccea4c4a3c28
Connecting to:      mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000&appName=mongosh+2.1.5
Using MongoDB:     7.0.6
Using Mongosh:     2.1.5

For mongosh info see: https://docs.mongodb.com/mongodb-shell/

-----
The server generated these startup warnings when booting
2024-05-29T17:52:04.209+00:00: Using the XFS filesystem is strongly recommended with the WiredTiger storage engine. See http://dochub.mongodb.org/core/prodnotes-filesystem
2024-05-29T17:52:04.813+00:00: Access control is not enabled for the database. Read and write access to data and configuration is unrestricted
2024-05-29T17:52:04.813+00:00: /sys/kernel/mm/transparent_hugepage/enabled is 'always'. We suggest setting it to 'never'
2024-05-29T17:52:04.813+00:00: vm.max_map_count is too low
-----

test> use boe
switched to db boe
boe>
```

Figura 26. Conexión a MongoDB

En este proyecto se creó una base de datos con el nombre de *BOE* y una colección con el nombre de *publicaciones* en la que almacenamos todos los documentos del BOE descargados.

Ejecutamos el método *descargarBOE()* y vemos los resultados en MongoDB. Al tener 216 publicaciones almacenadas tras ejecutar el método, se mostrará únicamente la publicación del ejemplo anterior ejecutando una búsqueda en Mongo de la siguiente manera:

```
boe> db.publicaciones.find({"metadatos.identificador" : "BOE-A-2024-10761"})
[
  {
    _id: ObjectId('66576cd51a77146854e33828'),
    fecha_actualizacion: '20240529115601',
    metadatos: {
      identificador: 'BOE-A-2024-10761',
      origen_legislativo: { codigo: '1', ':': 'Estatal' },
      departamento: {
        codigo: '9562',
        ':': 'Ministerio de Asuntos Exteriores, Unión Europea y Cooperación'
      },
      rango: { codigo: '1180', ':': 'Acuerdo Internacional' },
      fecha_disposicion: '20211210',
      numero_oficial: '',
      titulo: 'Modificaciones de los anexos I, II y IV del Protocolo para la protección del Mar Mediterráneo contra la contaminación causada por fuentes y actividades situadas en tierra, adoptadas en Antalya el 10 de diciembre de 2021.',
      diario: { codigo: 'BOE', ':': 'Boletín Oficial del Estado' },
      fecha_publicacion: '20240529',
      diario_numero: '130',
      seccion: '1',
      subseccion: '',
      pagina_inicial: '61912',
      pagina_final: '61920',
      suplemento_pagina_inicial: '',
      suplemento_pagina_final: ''
    }
  }
]
```

Figura 27. Find en MongoDB

3.4.3 Integración en Neo4j:

Ahora falta ver los resultados de la integración con Neo4j. Para visualizar los resultados en esta base de datos, tenemos que realizar una conexión a la base de datos e iniciar sesión para ver la interfaz gráfica de Neo4j.

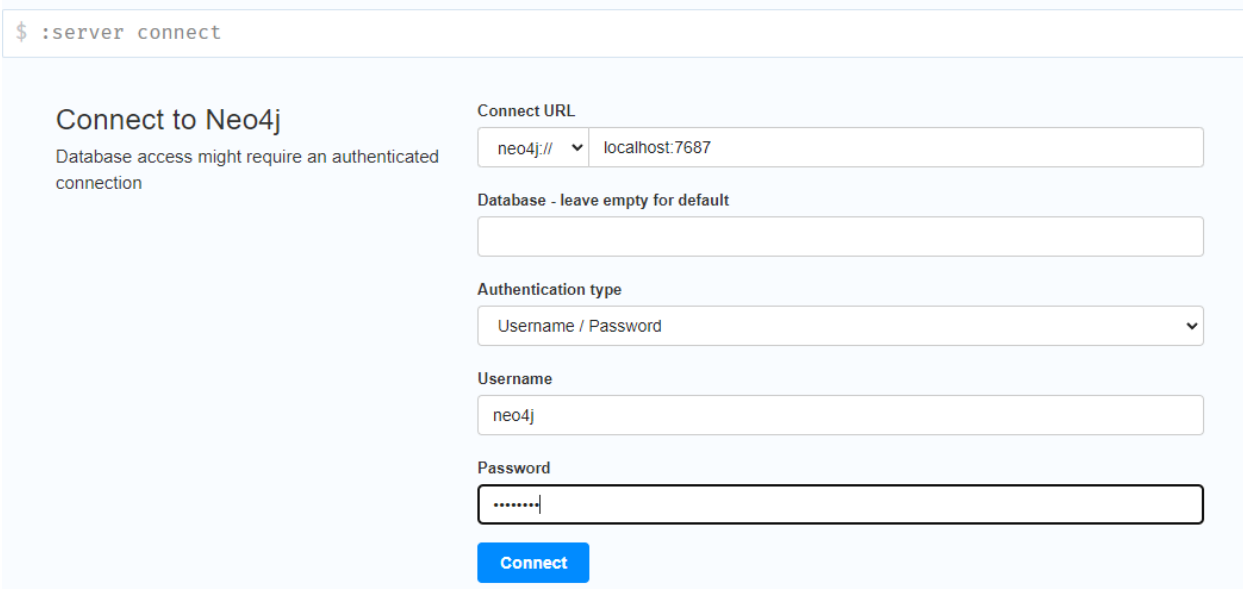


Figura 28. Log Neo4j

Como se puede ver en la imagen de abajo, se han creado 216 nodos, 1 por cada publicación.

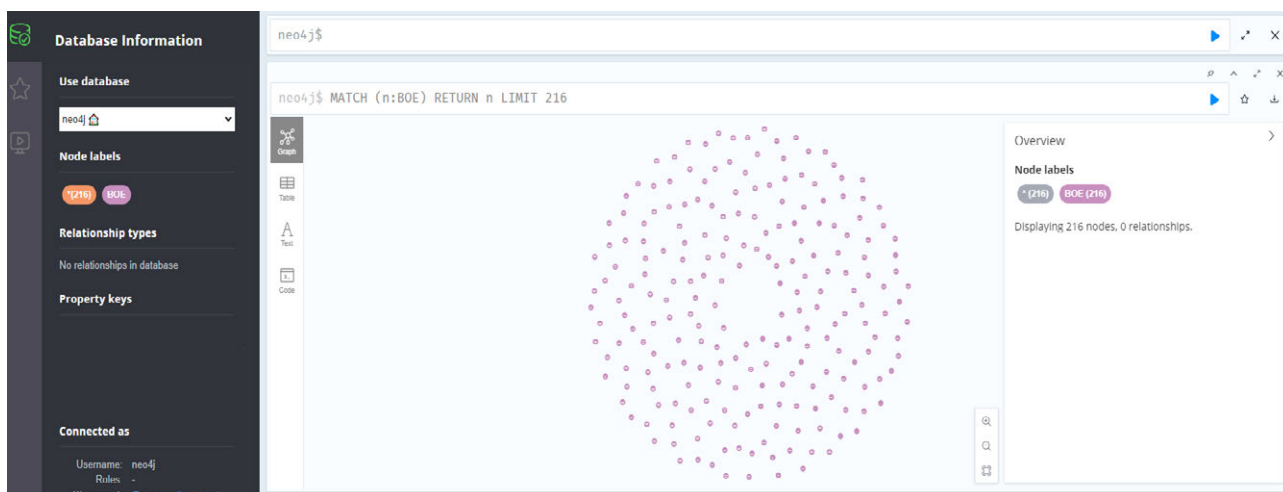


Figura 29. Neo4j resultados

Además, según la lógica de la herramienta, se debería crear un nodo en Neo4j con la metainformación que hemos elegido, en este caso serían: Identificador, rango, orden legislativo, departamento y título por cada publicación almacenada en MongoDB. Vamos a ver si es así utilizando la misma publicación de los ejemplos previos; BOE-A-2024-10761.



Figura 30. BOE en Neo4j

Finalmente comprobamos que tanto la metainformación de los nodos, el funcionamiento del método DescargarBOE() y el conjunto de los módulos trabajando para la extracción del BOE desde su página oficial hasta el almacenamiento en dos bases de datos distintas es correcto.

3.5 Recursos Generados

Gracias a los módulos anteriores estamos más cerca de cumplir con los objetivos del proyecto y dar los últimos pasos para conseguir una herramienta funcional. Como este documento pretende servir también como guía para desarrollos similares y para continuar con el trabajo desarrollado en este proyecto, se han propuesto distintas funcionalidades para mostrar el valor de la herramienta, la usabilidad con la que cuenta tras finalizar el proyecto y probar de manera más exhaustiva el funcionamiento de esta.

3.5.1 DescargarBoeDesdeHasta():

Se ha incluido otra forma de extraer el BOE con el método *DescargarBoeDesdeHasta()*, que descarga todas las publicaciones del BOE entre dos fechas dadas. El interés de este método reside en que el usuario final pueda descargarse una cantidad de publicaciones acotadas en una fecha y que explote los datos de dichas publicaciones aplicando el resto de funcionalidades de la herramienta.

Para probar el funcionamiento de este método se utilizarán las fechas 3-05-2024 hasta 6-05-2024, solo 3 días para que no tener una cantidad elevada de datos a la hora de enseñar los resultados. PD: se descargan 3 días porque el día 5-05-2024 es domingo y los domingos no se publica el BOE salvo en casos extraordinarios (este no lo fue), así también se prueba si la herramienta tiene algún problema al intentar descargar en fechas en las que no se han realizado ninguna publicación.

Se ejecuta el método y estos son los resultados:

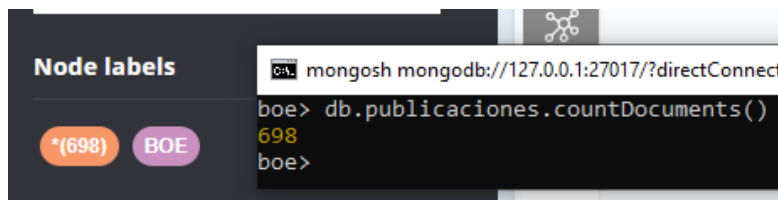
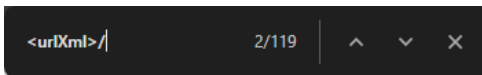


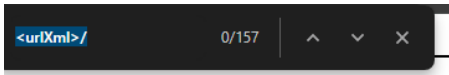
Figura 31. Resultados desdeHasta

Se han almacenado en total 698 publicaciones del BOE, así que hay que comprobar si el número de publicaciones entre esas fechas es correcto. Para ello iremos a los tres sumarios en XML publicados y contaremos cuantos documentos se subieron cada día, contando las apariciones de la etiqueta `<urlXml>` en el buscador de palabras Ctrl + F.

Para el BOE del día 3-05-2024: 119 publicaciones.



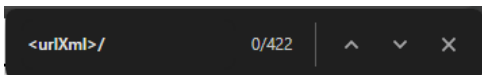
Para el BOE del día 4-05-2024: 157 publicaciones.



Para el BOE del día 5-05-2024: 0 publicaciones ya que al ser domingo no se publica el BOE, ni hubo publicaciones extraordinarias.

mayo 2024						
lun	mar	mié	jue	vie	sáb	dom
		<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>	<u>12</u>

Para el BOE del día 6-05-2024: 422 publicaciones.



Total de publicaciones: $119 + 157 + 422 = 698$.

3.5.2 Almacenar en Neo4j:

AlmacenarEnNeo4j() se trata de otro método para almacenar el BOE, pero en este caso sirve para, si por motivo de una funcionalidad futura u otra razón, se quiere transferir todo el contenido de la base de datos de MongoDB a la de Neo4j. En este caso la se recorrerá la colección completa de MongoDB (publicaciones) y se insertará publicación por publicación en Neo4j.

Para probar este método borraremos con el comando Cypher *MATCH (n) OPTIONAL MATCH (n)-[r]-() DELETE n, r;* todo el contenido de la base de datos de Neo4j y ejecutaremos el método con las 698 publicaciones que tenemos almacenadas en MongoDB.

Base de datos de Neo4j vacía:



Figura 32. Cuenta bbdd vacía MongoDB

Resultado de ejecutar el método:



Figura 33. Cuenta bbdd llena

3.5.3 Búsquedas filtradas:

Ahora vamos a ver los resultados de la implementación de unos métodos que realizan búsquedas filtradas. Algunas de las búsquedas filtradas son consultas que se pueden realizar actualmente en la web oficial del BOE y otras son consultas que actualmente no se pueden realizar con ninguna herramienta. Los resultados de los métodos se van a mostrar en la consola de Java ya que la herramienta carece de una interfaz gráfica propia.

MostrarBOE():

Este método es un ejemplo de consulta, en la que se busca una publicación del BOE por su nombre o identificador. El método buscará el BOE en la bbdd de Neo4j mediante una consulta Cypher, y únicamente mostrará la metainformación de la publicación almacenada en el nodo.

Este método devuelve el BOE que quieres buscar por su identificador/nombre. Vamos a comprobar si el método funciona con la publicación BOE-A-2024-10846.

```
13:57:21.474 [main] DEBUG io.netty.util.internal.PlatformDependent -- org.jctools-core.MpscChunkedArrayQueue: available
departamento: Consejo General del Poder Judicial
titulo: Acuerdo de 24 de abril de 2024, de la Comisión Permanente del Consejo General del Poder Judicial, por el que se declara en
rango: Acuerdo
identificador: BOE-A-2024-9027
origen_legislativo: Estatal
```

Figura 34. Resultado mostrarBOE()

Como se ve en la imagen el método funciona correctamente. Para demostrar que también se puede realizar la búsqueda en MongoDB se va a mostrar el resultado a continuación:

```
metadatos=Document{{identificador=BOE-A-2024-10846, origen_legislativo=Document{{codigo=1, =Estatal}},
departamento=Document{{codigo=9575, =Ministerio para la Transición Ecológica y el Reto Demográfico}}, rango=Document{{codigo=1370, =Resolución}}, fecha_disposicion=20240513,
texto=Document{{p=[Document{{class=centro_negrita, =Antecedentes de hecho}}, Document{{class=parrafo_2, =Con fecha 27 de diciembre de 2021, tiene entrada
```

Figura 35. Resultados muestraBOE() en MongoDB

GetBoeDepartamento():

Este filtro sirve para buscar las publicaciones del BOE realizadas por un departamento en concreto, dentro de nuestra base de datos. Actualmente tenemos almacenadas las 698 publicaciones del ejemplo anterior así que vamos a probar a filtrar por departamento, es decir, la herramienta debería devolver todas las publicaciones que hayan sido publicadas por un departamento en concreto (que se pasa por parámetro). Vamos a probar con el departamento “Consejo General del Poder Judicial”.

En este caso la búsqueda filtrada se ha realizado sobre la bbdd de Neo4j para que se vieran los resultados en la consola de java, ya que en MongoDB se recupera la publicación al completo y es muy difícil mostrarla mediante imágenes.

El resultado en la consola de Java sería:

```

departamento: Consejo General del Poder Judicial
titulo: Acuerdo de 8 de febrero de 2024, de la Comisión Permanente del Consejo General del Poder Judicial
rango: Acuerdo
identificador: BOE-A-2024-8901
origen_legislativo: Estatal
departamento: Consejo General del Poder Judicial
titulo: Acuerdo de 29 de abril de 2024, de la Comisión de Selección prevista en el artículo 305 de la Ley
rango: Acuerdo
identificador: BOE-A-2024-8978
origen_legislativo: Estatal
departamento: Consejo General del Poder Judicial
titulo: Acuerdo de 24 de abril de 2024, de la Comisión Permanente del Consejo General del Poder Judicial,
rango: Acuerdo
identificador: BOE-A-2024-9027
origen_legislativo: Estatal

```

Figura 36. Resultados filtrados por departamento

En este caso solo hay 3 documentos de los 698 almacenados que han sido publicados por el Consejo General del Poder Judicial en el BOE. Se comprobará si son correctas buscando las publicaciones en Neo4j y contando en MongoDB cuantos documentos con dicho departamento tenemos almacenados.

```
neo4j$ match (n:BOE) where n.departamento = 'Consejo General del Poder Judicial' return n
```



Figura 37. Resultados cuenta en Neo4j

```

boe> db.publicaciones.countDocuments({"metadatos.departamento.": "Consejo General del Poder Judicial"})
3
boe>

```

Figura 38. Resultad cuenta en MongoDB

El filtro funciona correctamente, pues en ambas fotos podemos comprobar que solo existen en 3 publicaciones dadas de alta por el departamento” Consejo General del Poder Judicial”.

getBoeDepartamentoRango():

Esta consulta utiliza un “doble filtro”, uno de los que ya hemos visto antes (departamento) en combinación con el filtrado por rango. En este caso esta consulta no se puede realizar con la web del BOE actualmente, no porque no se puedan combinar filtros, sino porque no existe ningún tipo de búsqueda filtrada por rango. En vez de mostrar simplemente el filtro de rango, que es muy simple, se aprovechará también para mostrar que se pueden combinar filtros y que la herramienta cuenta con consultas que no existen en los sistemas actuales.

Esta consulta filtra las publicaciones que pertenezcan a un departamento concreto, y que a su vez pertenezcan a un rango determinado. La búsqueda filtrada se realizará igual que la del filtro de

departamento, en Neo4j, pero también se puede hacer con MongoDB para obtener el contenido completo de las publicaciones del BOE.

Para realizar las comprobaciones de que el filtro funciona vamos primero a utilizar el filtro único de departamento (GetBoeDepartamento()) para que nos devuelva todas las publicaciones del Ministerio de Hacienda:

```

departamento: Ministerio de Hacienda
titulo: Anuncio de formalización de contratos de: Presidencia de la Autoridad Independiente de Responsabilidad Fiscal. Objeto: Servicio de adecuación a la normativa sobre accesibilidad de los documentos de la Autoridad Inde
rango: No tiene rango
identificador: BOE-B-2024-15810
origen_legislativo: No tiene origen legislativo
departamento: Ministerio de Hacienda
titulo: Resolución de 25 de abril de 2024, de la Intervención General de la Administración del Estado, por la que se publica el Convenio con la Comunidad Autónoma de Cantabria, para la adhesión a los sistemas departamentales
rango: Resolución
identificador: BOE-A-2024-9016
origen_legislativo: Estatal
departamento: Ministerio de Hacienda
titulo: Extracto de la Resolución de 25 de abril de 2024 del Consejo de Administración del Consorcio de la Ciudad de Toledo por la que se aprueban las bases reguladoras de la subvención directa en especie del Consorcio de la
rango: No tiene rango
identificador: BOE-B-2024-15925
origen_legislativo: No tiene origen legislativo
departamento: Ministerio de Hacienda
titulo: Extracto de la Resolución de 25 de abril de 2024 del Consejo de Administración del Consorcio de la Ciudad de Toledo por la que se aprueban las bases reguladoras de la subvención directa en especie del Consorcio de la
rango: No tiene rango
identificador: BOE-B-2024-15926
origen_legislativo: No tiene origen legislativo
departamento: Ministerio de Hacienda
titulo: Resolución de 26 de abril de 2024, de la Dirección General del Catastro, por la que se publica el Convenio con el Ayuntamiento de Atarfe.
rango: Resolución
identificador: BOE-A-2024-9094

```

Figura 39. Resultado java getboedepartamento

Para este departamento hay una gran cantidad de publicaciones, y muchas de ellas no tienen rango (pues no es un campo obligatorio), pero se puede observar en la imagen que los BOE-A-2024-9016 y BOE-A-2024-9094 si tienen rango, “Resolución”. Ahora ejecutaremos el “filtrado doble” para el departamento que hemos visto antes y el rango “Resolución” para ver si existen más publicaciones con departamento: Ministerio de Hacienda y Rango: Resolución:

```

14:24:24.917 [main] DEBUG io.netty.util.internal.PlatformDependent -- org.jctools-core.MpscChunkedArrayQueue: available
departamento: Ministerio de Hacienda
titulo: Resolución de 25 de abril de 2024, de la Intervención General de la Administración del Estado, por la que se publica el Convenio con la C
rango: Resolución
identificador: BOE-A-2024-9016
origen_legislativo: Estatal
departamento: Ministerio de Hacienda
titulo: Resolución de 26 de abril de 2024, de la Dirección General del Catastro, por la que se publica el Convenio con el Ayuntamiento de Atarfe.
rango: Resolución
identificador: BOE-A-2024-9094
origen_legislativo: Estatal
Process finished with exit code 0

```

Figura 40. Resultado java getBoeDepartamentoRango()

En este caso solo existían esas dos publicaciones con esas características particulares de rango y departamento. Ahora comprobaremos en ambas bases de datos si el método ha funcionado correctamente.

Contamos las publicaciones con un método en MongoDB:

```

boe> db.publicaciones.countDocuments({'$and':{'metadatos.rango':"Resolución"},{'metadatos.departamento':"Ministerio de Hacienda"}})
2
boe>

```

Comprobación en Neo4j:

```

1 MATCH (n:BOE)
2 WHERE n.departamento = 'Ministerio de Hacienda' AND n.rango = 'Resolución'
3 RETURN n
4

```



The image shows the Neo4j interface with a Cypher query and its results. The query is: `1 MATCH (n:BOE) 2 WHERE n.departamento = 'Ministerio de Hacienda' AND n.rango = 'Resolución' 3 RETURN n 4`. The results pane shows two nodes, both labeled 'Resoluci...'. The 'Overview' panel on the right indicates 'Node labels: * (2) BOE (2)' and 'Displaying 2 nodes, 0 relationships.'

Comprobamos que los datos son correctos y se finaliza la prueba.

Finalmente vemos que todos los métodos de filtrado que se han propuesto funcionan correctamente, y que tener los datos almacenados en distintas bases de datos permite realizar búsquedas muy rápidas y sencillas, y se da flexibilidad a la hora de elegir si se quieren realizar búsquedas sobre MongoDB para obtener el contenido completo de las publicaciones del BOE o si se requiere de más rapidez y sencillez en la presentación de los datos y se devuelven desde Neo4j.

La idea es que con estos filtrados puedas seleccionar el BOE que quieras según tus preferencias, y luego haciendo uso de una futura interfaz gráfica puedas tanto ver el contenido completo de una de las publicaciones del BOE que selecciones o que puedas operar sobre dichas publicaciones haciendo uso del resto de funcionalidades de la herramienta.

3.5.4 EvoluciónBOE():

Hasta ahora las funcionalidades vistas servían para almacenar las publicaciones en las distintas bases de datos del proyecto y búsquedas filtradas sobre las mismas, pero parte de lo anterior ya se puede hacer con la página oficial del BOE y las herramientas mostradas en el apartado del Estado del Arte. Este método, a diferencia de los demás, crea relaciones entre las publicaciones del BOE, relaciones que se ven representadas en la base de datos de Neo4j como relaciones entre nodos. Esto permite que las publicaciones del BOE ya no sean independientes entre sí, y que al estar relacionadas formen grafos dentro de Neo4j permitiendo aplicar algoritmos de grafos sobre ellos.

El método que se va a presentar a continuación relaciona las publicaciones del BOE según sus *referencias anteriores* para mostrar la evolución que ha tenido un artículo en el tiempo. Estas referencias se encuentran en las publicaciones del BOE en formato XML de la siguiente manera:

```
<documento fecha_actualizacion="20240426092601">
  <metadatos>
    ...
  </metadatos>
  <metadata-eli>
    ...
  </metadata-eli>
  <analisis>
    <materias>
      ...
    </materias>
    <notas/>
    <referencias>
      <anteriores>
        <anterior referencia="BOE-A-2021-13261" orden="1010">
          <palabra codigo="210">DEROGA</palabra>
          <texto>el Real Decreto 683/2021, de 3 de agosto</texto>
        </anterior>
        <anterior referencia="BOE-A-2023-24842" orden="3060">
          <palabra codigo="440">DE CONFORMIDAD con</palabra>
          <texto>el Real Decreto 1009/2023, de 5 de diciembre</texto>
        </anterior>
        <anterior referencia="BOE-A-2023-23537" orden="5020">
          <palabra codigo="330">CITA</palabra>
          <texto>Real Decreto 829/2023, de 20 de noviembre</texto>
        </anterior>
      </anteriores>
      <posteriores>
        ...
      </posteriores>
    </referencias>
  </analisis>
  <alertas>
    ...
  </alertas>
  <texto>
    ...
  </texto>
</documento>
```

Figura 41. Ejemplo referencias anteriores

Estas referencias vienen nombradas por el identificador del BOE y no todas las publicaciones del BOE tienen referencias.

Como el objetivo del método *evolucionBOE()* es mostrar la evolución de una publicación en el tiempo, hay que indicarle un nivel de profundidad en el que se van a buscar las referencias, es decir, si queremos ver la evolución de la publicación del BOE en sus últimas 3 referencias o en sus últimas 20 referencias si las hubiere. Por otro lado, la herramienta actual solo tiene almacenadas las publicaciones que se hayan descargado ejecutando la descarga del BOE diario que se ha visto anteriormente y la descarga del BOE desde una fecha inicio a una fecha fin. Luego las referencias de una publicación no se buscan en las bases de datos, ya que limitaría la funcionalidad de esta, si no que se buscan en la web oficial del BOE.

Para probar este método necesitaremos una publicación que tenga referencias anteriores, es decir, que tenga contenido en la etiqueta XML `<referencias> <anteriores> <anterior referencia = ““>`. Lo que haremos entonces será borrar todas las publicaciones de ambas bases de datos para poder visualizar (mejor) en Neo4j los resultados de la ejecución del método, que debería ser la construcción de un grafo en forma de árbol.

Con ambas bases de datos vacías almacenaremos el BOE-A-2024-5481 (figura 31). En la imagen de abajo se muestran las referencias que queremos.

```
▼<referencias>
  ▼<anteriores>
    ▼<anterior referencia="BOE-A-2021-13261" orden="1010">
      <palabra codigo="210">DEROGA</palabra>
      <texto>el Real Decreto 683/2021, de 3 de agosto</texto>
    </anterior>
    ▼<anterior referencia="BOE-A-2023-24842" orden="3060">
      <palabra codigo="440">DE CONFORMIDAD con</palabra>
      <texto>el Real Decreto 1009/2023, de 5 de diciembre</texto>
    </anterior>
    ▼<anterior referencia="BOE-A-2023-23537" orden="5020">
      <palabra codigo="330">CITA</palabra>
      <texto>Real Decreto 829/2023, de 20 de noviembre</texto>
    </anterior>
  </anteriores>
```

Figura 42. Referencias

Para nosotros un grafo de altura 4, será un grafo con 4 niveles, siendo el nivel 0 el nodo del BOE del que queremos obtener las referencias, el nivel 1 serían los nodos hijos correspondientes a las referencias, en el caso del BOE de la imagen anterior serían las referencias que se ven, el nivel 2 estaría conformado por los hijos de las referencias del nivel 1 y así sucesivamente. El último nivel sería el 3, y solo habrá niveles si los nodos de cada nivel del grafo tienen a su vez hijos (referencias), y los niveles o altura máximos que puede tener un grafo tras la ejecución de este método vendrá marcada por la profundidad elegida por el usuario.

Ahora ejecutaremos el método *EvolucionBOE()* para el BOE-A-2024-5481 con profundidad 2, es decir, como resultado de la ejecución deberíamos obtener como máximo un árbol de 3 alturas.

Ejecutamos el método y obtenemos:

MATCH (n:BOE) RETURN n LIMIT 25

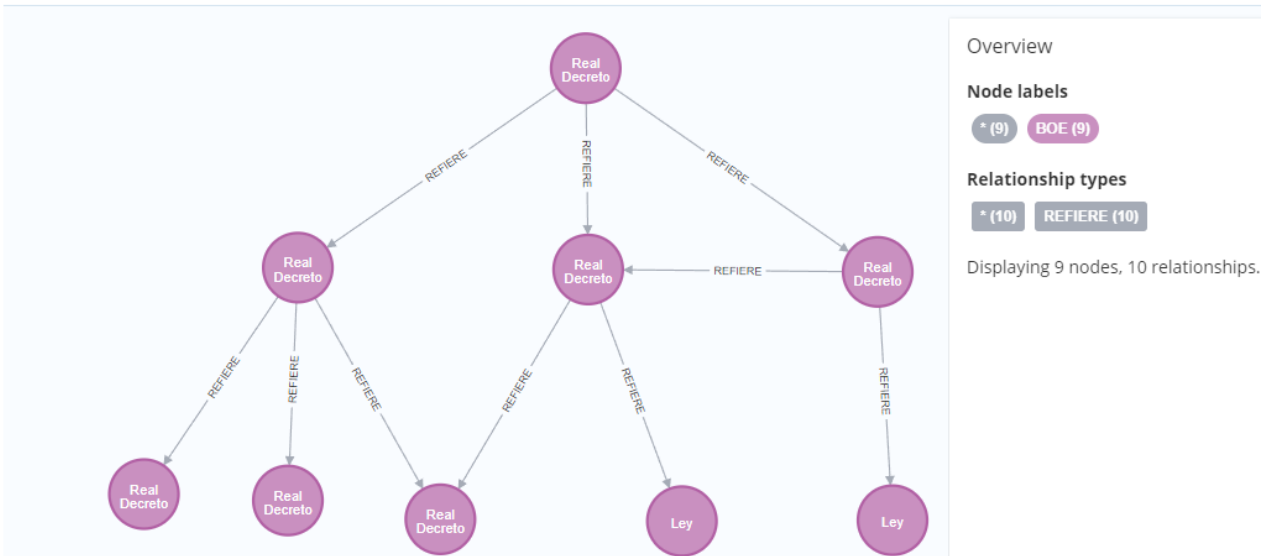


Figura 43. Árbol de evolucionBOE

En total tenemos 9 nodos, el inicial que ya teníamos, los 3 nodos de las referencias que se vieron en la imagen de la publicación inicial, y otros 5 nodos de las referencias de las referencias.

También tendremos las nuevas publicaciones almacenadas en MongoDB:

```
boe> db.publicaciones.countDocuments()  
9  
boe>
```

Vamos a comprobar entonces si los nodos del grafo se tratan de las referencias, y si estas son correctas:

El nodo de la izquierda del todo en el 1er nivel se corresponde con el BOE-A-2021-13261, y estas son sus referencias:

```
▼<referencias>  
  ▼<anteriores>  
    ▼<anterior referencia="BOE-A-2020-1996" orden="1010">  
      <palabra codigo="210">DEROGA</palabra>  
      <texto>el Real Decreto 307/2020, de 11 de febrero</texto>  
    </anterior>  
    ▼<anterior referencia="BOE-A-2020-1246" orden="3060">  
      <palabra codigo="440">DE CONFORMIDAD con</palabra>  
      <texto>el Real Decreto 139/2020, de 28 de enero</texto>  
    </anterior>  
    ▼<anterior referencia="BOE-A-2020-410" orden="3060">  
      <palabra codigo="440">DE CONFORMIDAD con</palabra>  
      <texto>el Real Decreto 2/2020, de 12 de enero</texto>  
    </anterior>  
  </anteriores>
```

Figura 44. Referencias 13261

Comprobamos ahora si sus tres nodos hijo son dichas referencias:

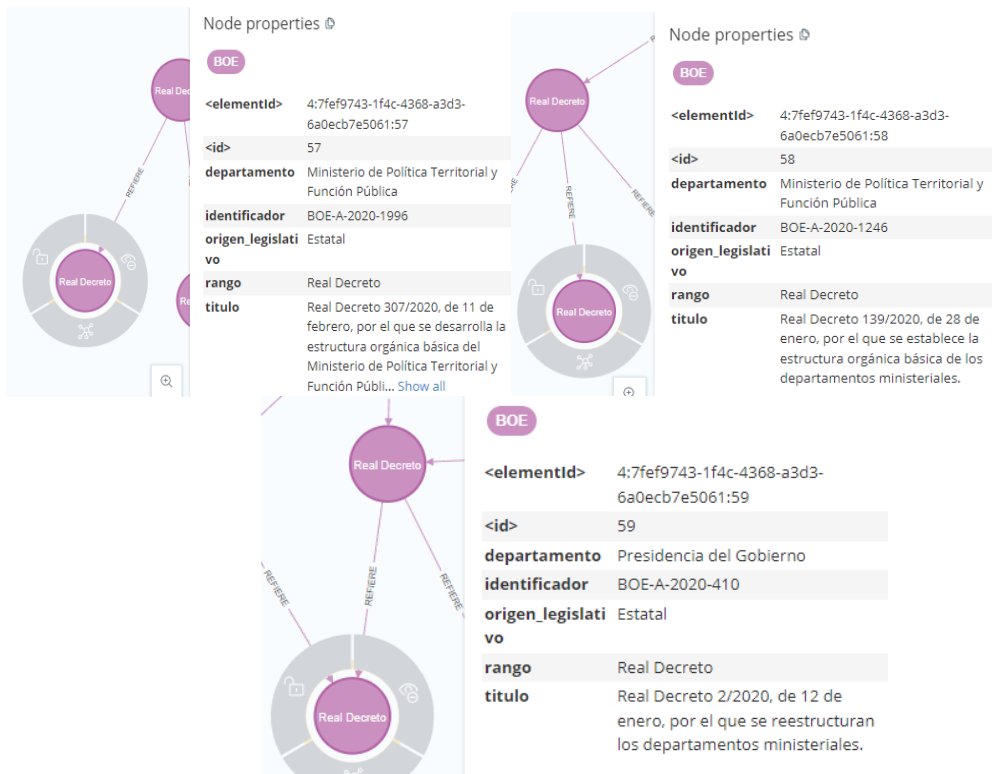


Figura 45. Comprobación 1er hijo

Vemos que sí se corresponden con las referencias de la figura 56. Ahora vamos a ver si las referencias del nodo del medio (BOE-A-2023-23537) del 1ºer nivel son correctas:

```

▼ <referencias>
  ▼ <anteriores>
    ▼ <anterior referencia="BOE-A-2020-410" orden="1010">
      <palabra codigo="210">DEROGA</palabra>
      <texto>el Real Decreto 2/2020, de 12 de enero</texto>
    </anterior>
    ▼ <anterior referencia="BOE-A-1997-25336" orden="3060">
      <palabra codigo="440">DE CONFORMIDAD con</palabra>
      <texto>el art. 2.2.j) de la Ley 50/1997, de 27 de noviembre</texto>
    </anterior>
  </anteriores>

```

Figura 46. Referencias 23537

Nodos hijo:

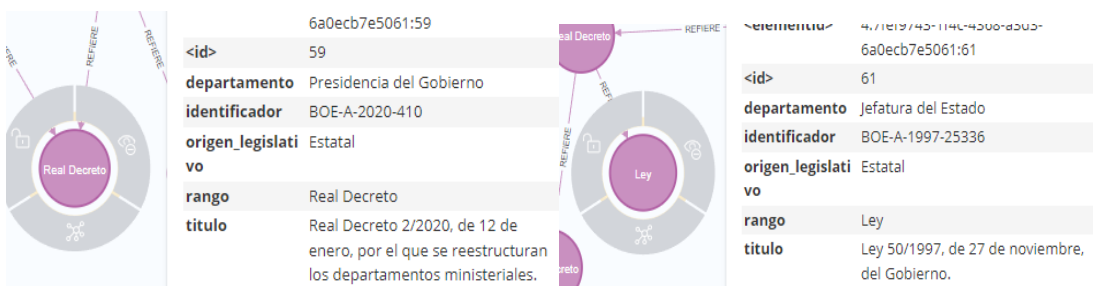


Figura 47. Referencias 23537 en Neo4j

Hasta ahora todos los nodos y referencias son correctas. Hacemos entonces la última comprobación de las referencias del nodo restante, BOE-A-2023-24842:

```

▼ <anteriores>
  ▼ <anterior referencia="BOE-A-2023-23537" orden="3060">
    <palabra codigo="440">DE CONFORMIDAD con</palabra>
    <texto>el Real Decreto 829/2023, de 20 de noviembre</texto>
  </anterior>
  ▼ <anterior referencia="BOE-A-2015-10566" orden="5020">
    <palabra codigo="330">CITA</palabra>
    <texto>Ley 40/2015, de 1 de octubre</texto>
  </anterior>
</anteriores>

```

Figura 48. Referencias 24842

Comprobamos ahora si sus referencias son correctas:

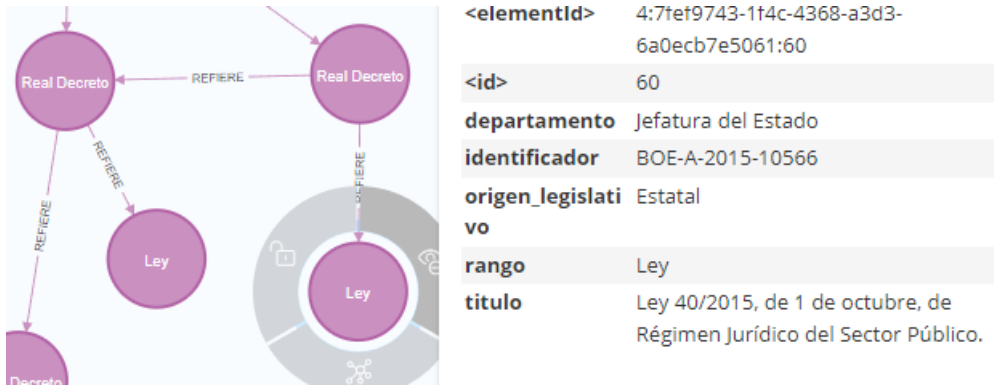


Figura 49. Referencias 24842 en Neo4j

Vemos que las referencias son correctas, pues se ha creado el nodo BOE-A-2023-10566 correctamente y también se ha creado la relación entre BOE-A-2023-24842 y BOE-A-2023-23537, que ya existía.

Con todas estas comprobaciones vemos que el método EvoluciónBOE() funciona correctamente y crea los grafos que pueden ser utilizados en un futuro para aplicar algoritmos sobre ellos.

Para demostrar que es posible aplicar algoritmos en la herramienta creada, vamos a realizar un pequeño ejemplo con el grafo que hemos creado gracias a EvoluciónBOE(). Se va a aplicar un algoritmo para encontrar el camino más corto entre dos nodos:

```

1 MATCH (start:BOE {identificador:'BOE-A-2024-5481'}), (end:BOE {identificador:'BOE-A-2020-410'})
2
3 MATCH path = shortestPath((start)-[:REFIERE*]- (end))
4 RETURN path;

```

Overview

- Node labels: (3) BOE (3)
- Relationship types: (2) REFIERE (2)

Displaying 3 nodes, 2 relationships.

Figura 50. Aplicación algoritmo

En este caso se ha buscado el camino más corto entre el nodo inicial BOE-A-2024-5481 y el nodo BOE-A-2020-410. A pesar de que el grafo es pequeño, se puede ver como se llega al nodo BOE-

A-2020-410 siguiendo el camino más corto, pues se podría llegar visitando el nodo BOE-A-2023-24842 (después del inicial) para llegar al nodo BOE-A-2023-23537, y finalmente acabar en el nodo objetivo.

Para finalizar con este apartado, vamos a comprobar si hemos logrado cumplir con los objetivos del proyecto y abordar los retos que se han planteado a lo largo del desarrollo de la herramienta:

En primer lugar, se puede decir que los retos del proyecto han sido abordados con éxito, pues desde el análisis y el diseño se afrontaron los retos sobre el manejo de grandes volúmenes de datos gracias a la elección de la base de datos de MongoDB y el paso previo a almacenar el BOE tras convertirlo a JSON para disminuir la carga de almacenamiento alrededor de un 10%. En la implementación se afrontó el reto de los duplicados mediante comprobaciones de los identificadores del BOE dentro de MongoDB. También se combatió la actualización continua de datos del BOE con el método DescargarBOE(), que descarga todo el contenido publicado del BOE cada día.

En segundo lugar, se han logrado cumplir con todos los objetivos del proyecto, el OB-1 se ha completado gracias al proceso de transformación de los datos para el almacenamiento en las distintas bases de datos de la herramienta. El OB-2 se ha afrontado con la codificación de métodos que permiten búsquedas filtradas, algunas de ellas imposibles de realizar con las herramientas actuales, pero que este proyecto posibilita, por ejemplo, con el filtrado por rango. El OB-3 también se puede dar por cumplido gracias a la implementación de medidas de control de duplicados, estructuras condicionales en la codificación de la herramienta y la actualización continua de la información.

Finalmente, el OB-4 se puede dar por logrado ya que como se vio en el apartado del estado del arte, no existe ninguna herramienta en la actualidad que procese los datos del BOE permitiendo no solo realizar búsquedas filtradas inviables con los sistemas actuales, sino que también permite estudiar el BOE relacionando sus publicaciones y posibilitando la aplicación de algoritmos de grafos sobre estas. El método EvoluciónBOE() es un ejemplo de la consecución de este objetivo, pues relaciona publicaciones entre sí creando un grafo en Neo4j, y posibilitando así la aplicación de algoritmos DAG⁶ con las herramientas de algoritmos y data science de Neo4j.

3.6. Despliegue

En esta sección se detalla el proceso de despliegue de la herramienta desarrollada con el objetivo de que cualquier usuario pueda descargar, instalar y comenzar a utilizar la herramienta o seguir desarrollándola. El despliegue del sistema implica varios pasos, que incluyen la descarga de los archivos necesarios, la configuración del entorno de ejecución, y la ejecución de la aplicación.

3.6.1 Requisitos Previos:

Para poder hacer uso del proyecto será necesario tener descargado Java, Maven y las bases de datos NoSQL; MongoDB y Neo4j. Para la elaboración de este proyecto se ha utilizado el entorno de desarrollo IntelliJ IDEA⁷ ya que funciona perfectamente con Java y está integrado con Maven. Por ello se recomienda su uso para quien quiera descargar el proyecto.

En cuanto a las bases de datos, para la elaboración del proyecto se utilizó Docker, una herramienta para la virtualización de aplicaciones basada en contenedores. En esta se tendrían que descargar

⁶ <https://es.cointelegraph.com/explained/what-is-a-dag-and-how-do-it-work>

⁷ <https://www.jetbrains.com/es-es/idea/>

las imágenes de MongoDB y Neo4j para poder hacer uso de ellas.

Las imágenes en Docker serían las siguientes:

Neo4j:

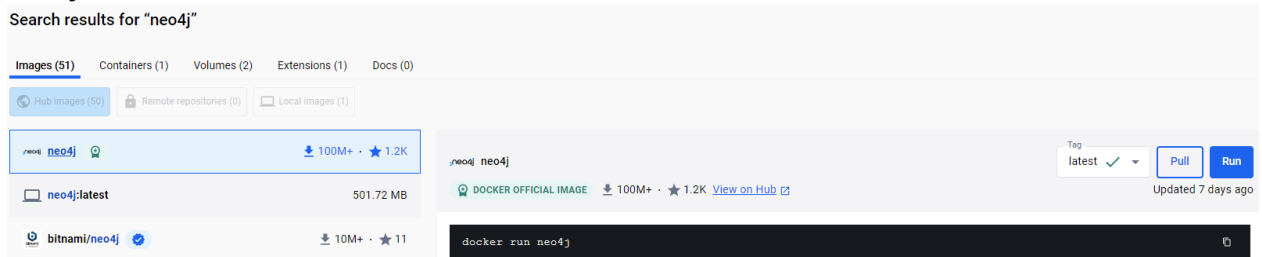


Figura 51. Imagen Neo4j usada

MongoDB:



Figura 52. Imagen MongoDB usada

3.6.2 Obtención del Proyecto:

Una vez cumplimos con los requisitos previos, se puede obtener el proyecto del siguiente repositorio de GitHub: <https://github.com/albertoupm/tfg-br0290.git>.

Para tenerlo en local se puede clonar el proyecto haciendo uso de Git⁸ y el siguiente comando: `git clone https://github.com/albertoupm/tfg-br0290.git`. También se puede descargar el zip que se ve en la imagen de abajo.

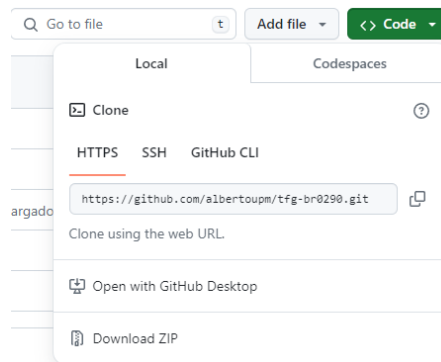


Figura 53. Repositorio GitHub

⁸ <https://git-scm.com/>

3.6.3 Configuración del entorno:

Una vez hemos descargado el proyecto de GitHub debemos abrir las clases de Java para comprobar que no salta ningún error en el código y que se han importado todas las librerías y dependencias del proyecto correctamente.

Lo siguiente para empezar a usar la aplicación es inicializar las dos bases de datos. En caso de no hacerlo con Docker, para Neo4j habría que crear una nueva BBDD y apuntar la contraseña que utilizemos, ya que será importante más adelante, e iniciarla dando al botón de *Start*. En el caso de MongoDB habría que crear una nueva base de datos y una colección dentro de esta base de datos, la cual usaremos para almacenar los documentos del BOE. Para MongoDB deberemos también apuntar los nombres de la base de datos y la colección.

En el caso de utilizar Docker, tendremos que crear un contenedor de Neo4j con el siguiente comando en la CMD:

```
docker run \
  --name neo4j-container \
  -p 7474:7474 -p 7687:7687 \
  -d \
  -e NEO4J_AUTH=neo4j/password \
  neo4j
```

En el cual estamos indicando el nombre del contenedor, que también es el de la nueva base de datos (“neo4j-container”), el puerto para la interfaz gráfica de Neo4j (7474:7474 -p 7687:7687) y la contraseña de la base de datos (“password”).

Para el uso de MongoDB en Docker tendremos que ejecutar el siguiente comando en la CMD: `docker run --name my-mongo-db -p 27017:27017`

Siendo “my-mongo-db” el nombre de la bbdd y 27017:27017 el puerto para la conexión. Como también necesitamos crear una colección, tendremos que meternos dentro de la base de datos que hemos creado para crear allí nuestra colección.

Comando para entrar en nuestra bbdd de MongoDB: `docker exec -it my-mongo-db sh`

Comando para crear la colección: 1º hacemos; `use my-mongo-db` 2º; `db.createCollection(“nombre de la colección”)`.

Una vez tenemos los contenedores creados con las bases de datos, tendremos que inicializarlas dando al botón de *Start* antes de utilizar el proyecto de Java.

Image	Status	CPU (%)	Port(s)	Last started	Actions
neo4j	Exited	N/A	7474:7474 Show all ports (2)	21 days ago	
mongo:latest	Exited	N/A	27017:27017	21 days ago	

Figura 54. Contenedores creados en Docker

Cuando tengamos tanto MongoDB como Neo4j corriendo en nuestro ordenador, para poder utilizar la herramienta solo faltaría configurar las conexiones desde el proyecto Java a las bases de datos.

En el caso de MongoDB habría que actualizar esta parte del código con el puerto, nombre de la bbdd y nombre de la colección elegidos, respectivamente.

```
private static final String MONGO_DB_URL = "mongodb://localhost:27017";
4 usages
private static final String MONGO_DB_NAME = "boe";
4 usages
private static final String MONGO_COLLECTION_NAME = "publicaciones";
```

Figura 55. Conexión a MongoDB

Para Neo4j habría que actualizar esta parte del código con el puerto y contraseña elegidos, respectivamente. En este caso el nombre de la base de datos no se utiliza y se usa *neo4j* como usuario predeterminado.

```
private static final String NE04J_URL = "bolt://localhost:7687";
6 usages
private static final String NE04J_USER = "neo4j";
6 usages
private static final String NE04J_PASSWORD = "password";
```

Figura 56. Conexión a Neo4j

Finalmente, con estos pasos realizados ya deberíamos poder utilizar la herramienta, así que haremos una pequeña comprobación para ver si todo funciona bien. Comprobaremos que se hace correctamente la conexión a la página oficial del BOE y que se almacenan las publicaciones descargadas en ambas bases de datos. Para que sea más rápida y simple esta comprobación descargaremos únicamente una publicación del BOE, para la cual solo necesitaremos su enlace con el contenido en XML.

En este ejemplo se utilizará la siguiente publicación https://www.boe.es/diario_boe/xml.php?id=BOE-A-2024-11379, pero se puede utilizar cualquiera.

En el método `DescargarBOE()` tendremos que sustituir `SacarUrl(fechaActual)` por `new ArrayList<>()` y agregar a `listaUrls` el enlace de nuestra publicación empezando por `diario_boe/`

El método antes del cambio:

```
public void descargarBOE() {
    Calendar calendario = GregorianCalendar.getInstance();
    Date fechaActual = calendario.getTime();
    List<String> listaUrls = SacarUrl(fechaActual);
    for (String lista : listaUrls) {
        // URL del BOE
        String urlBOE = "https://www.boe.es/" + lista;
        // Establecer conexión HTTP
        URL url;
        HttpURLConnection conexion;
```

Figura 57. descargarBOE original

Después:

```
public void descargarBOE() {
    Calendar calendario = GregorianCalendar.getInstance();
    Date fechaActual = calendario.getTime();
    List<String> listaUrls = new ArrayList<>();
    listaUrls.add("diario_boe/xml.php?id=BOE-A-2024-11379");
    for (String lista : listaUrls) {
        // URL del BOE
        String urlBOE = "https://www.boe.es/" + lista;
        // Establecer conexión HTTP
        URL url;
        HttpURLConnection conexion;
```

Figura 58. descargarBOE editado

Y ejecutamos la clase Main de esta forma:

```
public class Main {
    public static void main(String[] args){
        DescargadorBOE tfg = new DescargadorBOE();
        tfg.descargarBOE();
    }
}
```

Figura 59. Ejemplo simple del Main para 1era ejecución

Finalmente comprobamos ambas bases de datos para ver si existe la publicación, y si esta tiene los datos correspondientes.

Para la conexión a las bases de datos tendremos que seguir una serie de pasos. En caso de que no estemos usando Docker para MongoDB, solo tendremos que abrir una interfaz gráfica para mongo como puede ser Mongo Compass⁹. Si estamos usando Docker tendremos que abrir la CMD y utilizar el siguiente comando: `docker exec -it nombre_Contenedor mongosh`. Este comando hará que entremos a la shell de Mongo de nuestro contenedor de Docker, desde ahí podremos ver las bases de datos que tenemos con el comando `show dbs` y escribiendo `use nombre_BBDD` accederemos a nuestra base de datos.

Si queremos acceder a Neo4j tendremos que escribir en el buscador web que queramos la siguiente dirección: localhost:7474. El 7474 lo sustituiremos por el puerto que hayamos elegido.

Accediendo a esa dirección nos encontraremos la siguiente pantalla en la que tendremos que poner las credenciales que hayamos elegido:

\$:server connect

Connect to Neo4j
Database access might require an authenticated connection

Connect URL
neo4j:// localhost:7687

Database - leave empty for default

Authentication type
Username / Password

Username
neo4j

Password

Connect

Figura 60. Inicio sesión Neo4j

En MongoDB ejecutamos (dentro de nuestra bdd) el comando `db.nombreColección.find()` que saca todas las publicaciones, en este caso solo 1:

```
boe> db.publicaciones.countDocuments()
1
```

⁹ <https://www.mongodb.com/es/products/tools/compass>

Figura 60. Comprobación nº publicaciones en MongoDB

```
boe> db.publicaciones.find()
[
  {
    _id: ObjectId('6675ca38f1123470d7019135'),
    fecha_actualizacion: '20240610085601',
    metadatos: {
      identificador: 'BOE-A-2024-11379',
      origen_legislativo: { codigo: '1', ': 'Estatal' },
      departamento: { codigo: '9591', ': 'Ministerio de Industria y Turismo' },
      rango: { codigo: '1340', ': 'Real Decreto' },
      fecha_disposicion: '20240604',
      numero_oficial: '519/2024',
      titulo: 'Real Decreto 519/2024, de 4 de junio, por el que se modifican el R
22, de 5 de diciembre, por los que se regula la concesión directa de subvenciones
diario: { codigo: 'BOE', ': 'Boletín Oficial del Estado' },
      fecha_publicacion: '20240605',
      diario_numero: '136',
      seccion: '1',
      subseccion: '',
      pagina_inicial: '65442',
      pagina_final: '65448',
      suplemento_pagina_inicial: '',
      suplemento_pagina_final: '',
      url_pdf: '/boe/dias/2024/06/05/pdfs/BOE-A-2024-11379.pdf',
      url_epub: '/diario_boe/epub.php?id=BOE-A-2024-11379',
      url_pdf_catalan: '',
      url_pdf_euskera: '',
      url_pdf_gallego: ''
    },
    estatus_legislativo: 'L',
    fecha_vigencia: '20240606',
    estatus_derogacion: 'N',
    fecha_derogacion: '',
    judicialmente_anulada: 'N',
    fecha_anulacion: '',
    vigencia_agotada: 'N',
    estado_consolidacion: { codigo: '0' },
    letra_imagen: 'A',
    suplemento_letra_imagen: '',
    url_eli: 'https://www.boe.es/eli/es/rd/2024/06/04/519'
  },
  'metadata-eli': {
    RDF: {

```

Figura 61. Publicación en MongoDB

Vemos que se trata de nuestra publicación comprobando el identificador.

En Neo4j no es necesario ejecutar ningún comando para ver y contar las publicaciones, ya que se ve en la interfaz gráfica:

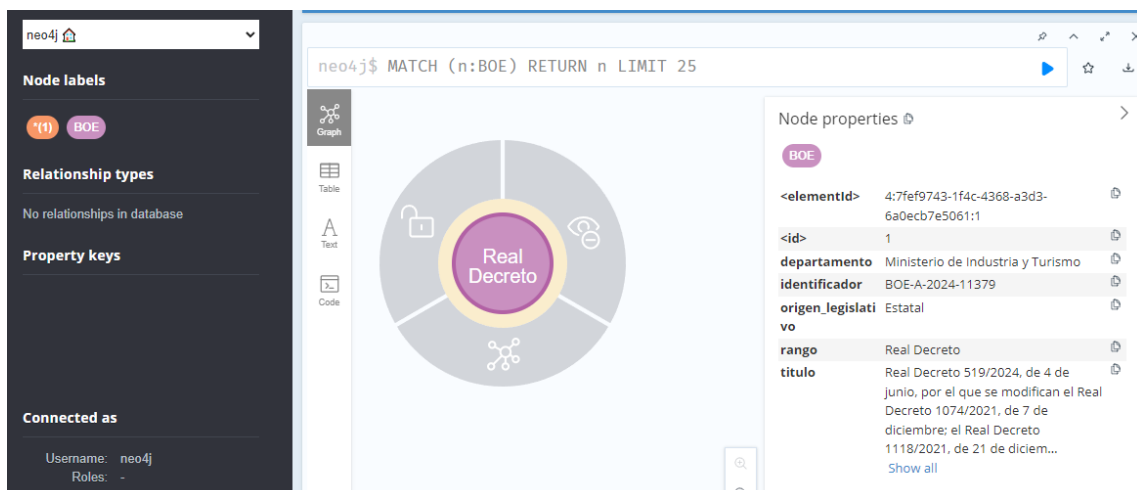


Figura 62. Publicación en Neo4j

Vemos también que el identificador de la publicación coincide y damos por terminada la verificación del sistema, y comprobamos que tenemos la herramienta correctamente instalada y lista para ser usada.

Enlaces de utilidad:

Java: https://www.java.com/es/download/ie_manual.jsp

Maven: <https://maven.apache.org/install.html>

IntelliJ Idea: <https://www.jetbrains.com/idea/download/?section=windows>

MongoDB: <https://www.mongodb.com/try/download/community>

Neo4j: <https://neo4j.com/download/>

Docker: <https://www.docker.com/products/docker-desktop/>

Compass: <https://www.mongodb.com/try/download/compass>

Conclusiones Personales

En esta esta sección se comentan las conclusiones personales sobre el trabajo realizado y se realiza una reflexión sobre la experiencia de la realización del TFG, comentando los conocimientos adquiridos. También se incluye la perspectiva sobre el impacto y relevancia del trabajo realizado:

La realización de este TFG ha sido una experiencia divertida y formativa, que me ha permitido afianzar y expandir mis conocimientos sobre la ingeniería del software. A lo largo del proyecto, he tenido la oportunidad de aplicar conceptos teóricos aprendidos durante el grado y de enfrentarme a desafíos prácticos que me han ayudado a desarrollar y comprobar mis habilidades técnicas y de gestión de proyectos.

Uno de los aspectos más destacados de esta experiencia ha sido el trabajo con bases de datos NoSQL como MongoDB y Neo4j. Aunque estos temas ya los había tocado superficialmente en las asignaturas de Bases de Datos, este proyecto me ha permitido profundizar en su funcionamiento y en mejores prácticas para su implementación y uso. He aprendido a manejar grandes volúmenes de datos y a realizar consultas complejas con Cypher y Mongo, lo cual ha sido fundamental para el éxito del proyecto.

Otro aspecto significativo ha sido la sensibilización respecto a la problemática que enfrentan juristas y otros profesionales que estudian el Boletín Oficial del Estado. Antes de comenzar este TFG, no era consciente de las dificultades que estas personas enfrentan debido a la falta de herramientas adecuadas para manejar y analizar la información publicada en el BOE y con este proyecto he podido comprender mejor estas necesidades y trabajar en una solución que facilite su trabajo.

Finalmente, aunque se alcanzaron los objetivos planteados al comienzo del proyecto, soy consciente de que el proyecto aún tiene mucho potencial para seguir desarrollándose, pues a pesar de contar con funcionalidades originales y de valor, hay numerosas mejoras que se pueden implementar para hacer la herramienta más completa y útil. Por ello me gustaría que este proyecto siguiera refinándose y desarrollándose para ofrecer un mayor valor a sus usuarios en el futuro.

En resumen, este TFG ha sido una experiencia valiosa que ha fortalecido mis conocimientos técnicos, ha aumentado mi conciencia sobre problemas reales en el ámbito jurídico y ha reafirmado mi compromiso con el desarrollo de soluciones tecnológicas útiles y relevantes.

Trabajo Futuro

En esta sección se abordan las posibles direcciones y pasos siguientes que podrían tomarse a partir del desarrollo realizado en el TFG. Se identifican áreas de mejora, oportunidades para investigaciones adicionales, y se sugieren proyectos que no se pudieron realizar en el este trabajo pero que serían valiosos para explorar en el futuro:

El desarrollo de este proyecto ha sentado las bases para una herramienta valiosa, pero aún queda un camino significativo por recorrer para alcanzar todo su potencial. Como posibles direcciones y pasos siguientes que podrían tomarse a partir del desarrollo realizado, se debería seguir desarrollando y añadiendo funcionalidades conforme los usuarios den su feedback.

El proyecto actual presenta únicamente un backend con funcionalidades limitadas por el tiempo en el que se ha desarrollado el proyecto, por lo que sería crítico el desarrollo de una interfaz gráfica. La interfaz gráfica debería ser intuitiva y accesible, permitiendo tanto a expertos como no expertos ejecutar y visualizar los resultados de las funcionalidades de manera sencilla y eficiente, implementando visualizaciones interactivas que permitan a los usuarios entender mejor los datos y resultados generados por la herramienta. También habría que trabajar en la portabilidad de la herramienta para que funcione correctamente en una variedad de dispositivos, incluyendo ordenadores, tablets y teléfonos inteligentes.

En resumen, aunque los objetivos planteados para este TFG han sido alcanzados, el proyecto tiene un inmenso potencial de crecimiento y mejora. La continuación del desarrollo basado en el feedback de los usuarios, la creación de una interfaz gráfica intuitiva, y la exploración de nuevas áreas de investigación asegurarán que esta herramienta evolucione y se convierta en un recurso indispensable para juristas y otros profesionales del sector legal.

Responsabilidad Social y Ambiental

Para finalizar la memoria tenemos que hablar sobre el impacto social y medioambiental que supone este proyecto. Al tratarse de un trabajo con el fin de mejorar la accesibilidad y el análisis de la información del BOE, facilitando no solo el trabajo de juristas y otros profesionales, sino también facilitando el acceso a esta información a cualquier tipo de persona. Se podría decir que este proyecto promueve una sociedad más informada y participativa, donde la información pública está al alcance de todos, independientemente del nivel de especialización o recursos.

Para tratar la responsabilidad social se ha pensado en la relación del proyecto con algunos de los ODS, como el 9 sobre Industria, Innovación e Infraestructura, que se centra en construir infraestructuras resilientes, promover la industrialización inclusiva y sostenible, y fomentar la innovación. Este proyecto facilita la innovación al proporcionar una herramienta para el análisis y acceso a la información publicada en el BOE mediante una búsqueda más eficiente y funcionalidades que previas al desarrollo del proyecto no existían sobre el BOE. El proyecto contribuye a la creación de nuevas oportunidades para la investigación y el desarrollo en el ámbito legal, académico y empresarial.

Al mejorar el acceso y la interpretación de la información legal y administrativa, esta herramienta apoya a investigadores, legisladores, abogados y otros profesionales, fomentando un entorno más informado y digitalizado.

Otro de los ODS con los que tiene relación el proyecto es el ODS 16, el cual tiene como objetivo

promover sociedades justas, pacíficas e inclusivas, y fortalecer las instituciones a todos los niveles. Este proyecto facilita el acceso a información jurídica y administrativa, la cual es fundamental para el buen funcionamiento de las instituciones y la participación ciudadana.

Al proporcionar una herramienta que simplifica la búsqueda y el análisis de la información del BOE, el proyecto promueve la transparencia y la accesibilidad a la información pública. Esto es crucial para que todos los ciudadanos españoles puedan conocer las leyes, disposiciones y actos oficiales que les afectan, como las normativas laborales, convocatorias de oposiciones y subastas públicas. Además, de ayudar a empresas y sus trabajadores a estar informados sobre sus derechos y obligaciones.

Sobre la responsabilidad ambiental, este proyecto podría tener un impacto negativo debido a la contaminación digital que provoca cualquier proyecto que implique el almacenamiento y procesamiento de grandes conjuntos de datos. Al final, la contaminación digital se refiere al impacto sobre el medio ambiente que se genera por el consumo de energía y los recursos necesarios para mantener la infraestructura tecnológica de servidores y centros de datos.

A pesar de lo anterior siempre se pueden mitigar los daños causados por el almacenamiento de grandes cantidades de datos. Una de las formas de mitigar estos daños es la eficiencia en el almacenamiento, como puede ser el control de duplicados, ya que al reducir el espacio necesario para almacenar información disminuye la cantidad de energía asociada al almacenamiento. En este proyecto contamos con varias medidas que ayudan en este aspecto, como el control de duplicados y el haber reducido en un 10% o más la carga de almacenamiento gracias a la transformación de las publicaciones de formato XML a formato JSON.

Otra forma en la que se ha mitigado este impacto negativo es mediante la optimización de consultas, ya que, si se minimiza la cantidad de procesamiento para obtener cualquier información, también se minimiza la carga de trabajo en los servidores y su consumo energético.

Debido al alcance de este proyecto, no se han podido implementar medidas adicionales contra la contaminación ambiental, además de las anteriormente explicadas. Sin embargo, sería interesante que, en el futuro, si el proyecto continúa desarrollándose como se explicó en la sección anterior, los servidores y centros de datos que se utilicen funcionen con energías renovables o con proveedores de servicios en la nube comprometidos con la sostenibilidad ambiental.

En resumen, este proyecto contribuye al desarrollo sostenible, apoyando la innovación y el fortaleciendo las instituciones mediante una mayor transparencia y acceso a la información, y, a pesar de que el almacenamiento y procesamiento de grandes conjuntos de datos presenten un impacto ambiental negativo, este proyecto intenta mitigar su impacto mediante la implementación de estrategias sostenibles y eficientes que contribuyan a la reducción de la contaminación digital.

Bibliografía

- [1] Oposiciones Flou . (2024, marzo 26). Cómo buscar Oposiciones en el BOE: una guía imprescindible. Oposiciones Flou. <https://oposicionesflou.com/blog/buscar-oposiciones-boe/#:~:text=Una%20vez%20dentro%20del%20BOE,del%20Bolet%C3%ADn%20Oficial%20del%20Estado>.
- [2] Silva Tools. (n.d.). BOE a la carta. Google Play Store. https://play.google.com/store/apps/details?id=com.silvatools.boe&hl=es_EC
- [3] JurisTracker. (n.d.). JurisTracker. <https://www.juristracker.com/>
- [4] Gómez-Obregón, J. (2022, octubre 29). SuperBOE. <https://superboe.es/>
- [5] Uppers. (2022, noviembre 8). Cuando el BOE dejó de ser un 'ladrillo': la nueva herramienta que te pone fácil consultarlo. Uppers. https://www.uppers.es/ciencia-y-tecnologia/nuevas-tecnologias/20221108/herramienta-boe-permitira-consultar-ajustar-diseno_18_07918390.html
- [6] EPUB. (2024, febrero 20). En Wikipedia. <https://es.wikipedia.org/wiki/EPUB>
- [7] Neo4j. (n.d.). Neo4j. <https://neo4j.com/es/producto/>
- [8] Emilio. (2020, marzo 30). Introducción, análisis y uso de Neo4j. <https://todobi.com/introduccion-analisis-y-uso-de-neo4j/>
- [9] Neo4j. (n.d.). Docker. <https://neo4j.com/docs/operations-manual/current/docker/>
- [10] Agencia Estatal Boletín Oficial del Estado. (n.d.). Sumarios del BOE. https://www.boe.es/datosabiertos/documentos/SumariosBOE_v_1_0.pdf
- [11] Maldonado José. (2020, abril 2). ¿Qué es un DAG y cómo funciona? Cointelegraph. <https://es.cointelegraph.com/explained/what-is-a-dag-and-how-do-it-work>
- [12] Agencia Estatal Boletín Oficial del Estado. (n.d.). Ayuda del BOE. https://www.boe.es/diario_boe/ayuda.php
- [13] Agencia Estatal Boletín Oficial del Estado. (n.d.). Información sobre el organismo. <https://www.boe.es/organismo/>
- [14] Universitat Oberta de Catalunya. (2018, junio 11). Introducción a Cypher. <https://blogs.uoc.edu/informatica/es/introduccion-a-cypher/>
- [15] Naciones Unidas. (n.d.). Objetivos de Desarrollo Sostenible. <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>
- [16] Delfino.cr. (2023, enero 7). La utilidad y aplicación de los grafos y sistemas de información geográfica. <https://delfino.cr/2023/01/la-utilidad-y-aplicacion-de-los-grafos-y-sistemas-de-informacion-geografica>
- [17] Immedia IT. (2024, febrero 2). Conectando los puntos: el papel de los grafos de conocimiento en la transformación digital. <https://immediait.com/es/conectando-los-puntos-el-papel-de-los-grafos-de-conocimiento-en-la-transformacion-digital#:~:=%C2%BFQu%C3%A9%20es%20un%20Grafo%20de,entre%20diferentes%20entidades%20o%20conceptos>

[18] Iberdrola. (n.d). La sostenibilidad digital como solución para reducir nuestro impacto ambiental. <https://www.iberdrola.com/sostenibilidad/contaminacion-digital>

[19] Garret Caroline .(2022, noviembre 14). Contaminación digital: ¿cómo se puede reducir? Selectra. <https://climate.selectra.com/es/que-es/contaminacion-internet>