*Imagen que contiene dibujo

Descripción generada automáticamente*

MEMORIA

DEL BACK AL END

Martín Ordoño Caro

Trabajo de Fin de Grado, 2020

Tutora: Inmaculada Hernández Salmerón

Grado en Ingeniería Informática, Ingeniería del Software

Escuela Técnica Superior de Ingeniería Informática

Índice

[1. Resumen 5](#_Toc50099679)

[2. Introducción 5](#_Toc50099680)

[3. Gestión del proyecto 6](#_Toc50099681)

[3.1. Alcance 6](#_Toc50099682)

[3.2. Objetivos 8](#_Toc50099683)

[3.3. Planificación 9](#_Toc50099684)

[3.4. Metodología 14](#_Toc50099685)

[3.5. Herramientas 15](#_Toc50099686)

[3.5.1. Github 15](#_Toc50099687)

[3.5.2. Entornos de desarrollo 15](#_Toc50099688)

[3.5.3. Frameworks 16](#_Toc50099689)

[3.6. Relación con las asignaturas del Grado 16](#_Toc50099690)

[1. Diseño y pruebas 16](#_Toc50099691)

[4. Realización del proyecto 18](#_Toc50099692)

[4.1. Backend 18](#_Toc50099693)

[4.1.1. Análisis 18](#_Toc50099694)

[4.1.2. Diseño 19](#_Toc50099695)

[4.1.3. Implementación 20](#_Toc50099696)

[4.1.4. Pruebas 26](#_Toc50099697)

[4.2. Frontend 26](#_Toc50099698)

[4.2.1. Análisis 26](#_Toc50099699)

[4.2.2. Diseño 27](#_Toc50099700)

[4.2.3. Implementación 29](#_Toc50099701)

[4.2.4. Pruebas 32](#_Toc50099702)

[5. Conclusión 32](#_Toc50099703)

[I. Anexo I – Glosario de términos. 34](#_Toc50099704)

[II. Anexo II – Control de cambios. 35](#_Toc50099705)

# Resumen

Este proyecto se centra en la extracción de datos de varias fuentes para almacenar y mostrar la información de diferentes maneras. Para ello, se realizará una aplicación con arquitectura desacoplada diferenciando entre Frontend y Backend.

Las características más importantes de este proyecto son los distintos Scrapers personalizados que han de extraer los datos de las distintas páginas web junto a él Frontend, que deberá mostrar los datos de forma gráfica.

A lo largo de este documento se detallan las distintas tecnologías usadas para la realización del proyecto, de la cual se destaca Svelte, un framework web muy joven con buenas cualidades que puede llegar a dar mucho rendimiento.

Además, en este documento se encuentra toda la información acerca del alcance y la planificación del proyecto, las metodologías que se usan durante el desarrollo, se detalla la implementación del frontend y backend por separado y se explican todos los problemas encontrados.

# Introducción

Se quiere conseguir comparar series de datos de distintas fuentes, tanto de forma visual como los datos numéricos. No obstante, este objetivo se puede fragmentar en distintos apartados, todos ellos centrados en las series de datos: Extracción y almacenamiento, recuperación y entrega de datos, y, por último, visualización y exportación. Estos tres apartados se corresponderían, respectivamente, a los Scrapers, el backend y el frontend.

Por lo tanto, para llegar a la solución, se ha de conseguir una aplicación que cumpla con los tres apartados anteriores. Además, al optar por una arquitectura desacoplada, el backend puede ser usado como una API con la que aplicaciones de terceros puedan interactuar.

La finalidad es tener un sistema completo que extraiga toda la información relevante y la muestre en el mismo formato independientemente de la fuente. Como se va a extraer información de distintas fuentes, es probable que los datos vengan en distintos formatos y/o estructurados de distinta forma, o que directamente falte información. Por ello, al recolectar la información hay que realizar un tratamiento para almacenar los datos con la misma estructura.

# Gestión del proyecto

## Alcance

Objetivo del proyecto

El objetivo del proyecto es desarrollar una aplicación que permita comparar series de datos.

La aplicación contará con distintos Scrapers que se encarguen de extraer y almacenar las series de datos de las distintas fuentes especificadas por el cliente. Una vez los datos están almacenados, la aplicación deberá mandar las series de datos al frontend. Allí, los usuarios tendrán a su disposición el conjunto de series provenientes de distintas fuentes y serán capaces de elegir cuales comparar tanto gráficamente como numéricamente. Además, los usuarios tendrán la opción de exportar las comparaciones que realicen con solo pulsar un botón.

Restricciones

A continuación, se definen las restricciones del proyecto a desarrollar, estas restricciones limitan:

* Tiempo: es imprescindible que se cumpla la fecha de entrega.
* Calidad: el producto desarrollado tiene que cumplir con un mínimo de calidad que satisfaga tanto al cliente como al tutor de este.
* Riesgos: los riesgos contemplados en el proyecto deben tenerse en cuenta para su correcto desarrollo.
* Costes: al ser un proyecto sin presupuesto inicial, las herramientas y tecnologías usadas han de ser gratuitas. El único gasto que se contempla es el del hosting y dominio si se desplegase la aplicación y no fuese un coste alto.

Riesgos

Lista de riesgos posibles durante la vida del proyecto, cada riesgo ha de tenerse en cuenta para evitar contratiempos:

* **Planificación**: si no se estima y planifica adecuadamente, es probable que no se cumplan los objetivos para la fecha acordada.
* **Tecnología**: al usar un framework desconocido, es posible que el proyecto no se pueda desarrollar acorde a los requisitos establecidos o que las tareas de investigación se alarguen en exceso.
* **COVID-19**: Si se contrae el virus esto provocaría inevitablemente un retraso en el proyecto como mínimo.
* **Pérdida de datos**: Existe la posibilidad de que se pierda información debido al malfuncionamiento de algún software o el fallo en el disco duro. Además, incidentes naturales o de caída de energía pueden provocar la pérdida de datos.

|  |  |  |  |
| --- | --- | --- | --- |
| **Riesgo** | **Impacto** | **Probabilidad** | **Nivel de riesgo** |
| Planificación | 4 | 4 | 8 |
| Tecnología | 3 | 7 | 10 |
| COVID-19 | 10 | 1 | 11 |
| Pérdida de datos | 10 | 2 | 12 |

El nivel de riesgo es el resultado de la suma del ipacto y la probabilidad, y representa lo grave que es el riesgo, por lo que a mayor nivel, más se tendrá que tener en cuenta.

Medidas de control

* **Planificación**: se realizará una planificación pesimista, dejando margen suficiente a las tareas hasta adquirir la capacidad de realizar estimaciones más aproximadas.
* **Tecnología**: Se realizarán tareas de investigación previas al desarrollo con las tecnologías desconocidas.
* **COVID-19**: Se tomarán las medidas de seguridad propuestas por el estado, minimizando así la posibilidad de contraer el virus.
* **Pérdida de datos**: se usará un repositorio privado en GitHub para almacenar todos los documentos y proyectos del producto, actualizándose en cada incremento. Esto reduce el impacto al mínimo, ya que únicamente se podrá perder las modificaciones realizadas al ocurrir la incidencia.

Criterios de aceptación

El proyecto será aceptado únicamente si se cumplen los siguientes criterios de aceptación:

* Cumple los requisitos establecidos.
* Es entregado dentro del plazo correspondiente.

## Objetivos

El Trabajo de fin de grado tiene como objetivo proporcionar al alumno una visión general y completa de la planificación, gestión y metodologías aplicables a un proyecto informático. Debe ser un trabajo con una complejidad considerable, al tener una carga de 12 créditos ECTS. Por lo tanto, uno de los objetivos principales de este proyecto es desarrollar un software completo de principio a fin, resumiendo todo lo aprendido a lo largo del grado.

Al ser un proyecto individual, entre los objetivos también se incluyen las metas personales del alumno, destacando un mayor interés por aprender tecnologías nuevas desconocidas por el mismo. Esto se tendrá en cuenta a la hora de elegir las tecnologías para el producto, a pesar de que usar tecnologías más novedosas puede acarrear inconvenientes.

Por último, pero no menos importante, este es un trabajo desarrollado para un cliente, el cual quiere un producto útil y de una calidad mínima. Por tanto, es importante desarrollar un buen producto con el que impresionar al cliente, ya que puede llegar a ser un factor determinante para el comienzo de la vida laboral del alumno.

Al ser la última asignatura del grado, esta asignatura, junto con las prácticas, son el enlace entre la vida universitaria y la laboral, por lo que elaborar un producto de calidad puede ser un factor determinante a la hora de encontrar un buen puesto y dar a conocer las habilidades del alumno.

## Planificación

A pesar de no seguir una metodología ágil, el desarrollo de este proyecto se divide en tres fases principales. Cada una de las fases tiene un propósito distinto y está bien definida en el tiempo. En la primera fase se realizarán las tareas de investigación, el propósito de esta fase es determinar los frameworks y herramientas que se van a usar (especialmente en el frontend) y adquirir una base para poder comenzar a trabajar con ellos. El objetivo de la segunda fase es construir un prototipo con las funcionalidades clave, este prototipo debe ser completo para poder hacer una demostración al cliente y recibir buen feedback. El objetivo de la tercera y última fase es completar el producto, para ello se tendrá en cuenta el feedback recibido al final de la segunda fase. Además, se implementarán las funcionalidades extras y se mejorará tanto el apartado visual del producto como la API.

Primera fase: Inicio

Como se indica en el párrafo anterior, la finalidad de esta fase es determinar las tecnologías que se van a usar y adquirir cierto conocimiento básico en aquellas que sean desconocidas. Para ello se ha de realizar un análisis de las herramientas y framework propuestos y valorar cuáles son las mejores opciones. En concreto, estas son las alternativas:

**Base de dato**s: Oracle (sql) / MongoDB (nosql).

**Frontend**: Svelte / React.

**Backend**: Spring Boot / Otra.

Segunda fase: Prototipo

El objetivo de esta fase es implementar las funcionalidades principales. Para ello, se creará un scraper para una de las fuentes de datos, INDEA, y se implementará toda la funcionalidad desde el almacenamiento en base de datos hasta la visualización de las series de datos en el frontend.

Tercera fase: Finalización

A lo largo de esta fase, se completará el producto con las tareas detectadas en el feedback obtenido del cliente, además de cualquier funcionalidad extra propuesta. En el caso de que no diese tiempo a implementar alguna funcionalidad en la fase anterior, esta pasaría como tarea de esta fase con una prioridad superior a las demás.

Tareas

A continuación, se encuentra el listado de tareas. Aquí se definen los tiempos de inicio y finalización de las fases, las descripciones de las tareas y sus prioridades. Las tareas que se encuentran en este backlog son tareas generales que se dividirán en subtareas antes de ser desarrolladas.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Primera fase: Inicio** | | | | |
| **Fecha inicio: 01/07/2020** | | | **Fecha fin: 07/07/2020** | |
| **TI.01** | Determinar herramientas y frameworks a usar. | | | A |
| **TI.02** | Obtener conocimientos básicos sobre las herramientas elegidas desconocidas. | | | A |
| * Svelte | | | | 3h |
| * MongoDB | | | | 1h |
| **Segunda fase: Prototipo** | | | | |
| **Fecha Inicio: 08/07/2020** | | | **Fecha fin: 28/07/2020** | |
| **TP.01** | Inicialización del proyecto backend | | | A |
| * Crear proyecto base con Spring initializr. | | | | 2h |
| * Instalar dependencias (POM). | | | | 1h |
| **TP.02** | Tabla de series. | | | A |
| * Componente para mostrar las series en una tabla en Svelte. | | | | 5h |
| * Paginación en la tabla. | | | | 5h |
| * Endpoint con paginación en la API. | | | | 1h |
| **TP.03** | Almacenamiento de datos del sistema sobre series y sus datos en MongoDB. | | | A |
| * Dominio java para Serie con la siguiente información: código, periodo, fuente y colección de datos. | | | | 2h |
| * Dominio java para Datos con la siguiente información: momento y valor. | | | | 2h |
| * Repositorios y servicios para Series y Datos | | | | 2h |
| * BDD MongoDB y conectar con los repositorios anteriores | | | | 2h |
| **TP.04** | Scraper INDEA | | | A |
| * Scraper que recolecte series de INDEA e inserte las series en el sistema. | | | | 8h |
| **TP.05** | API backend | | | A |
| * Endpoints básicos para enviar las series de datos al frontend. | | | | 3h |
| * CORS. | | | | 2h |
| **TP.06** | Visualización de las series en forma de tabla. | | | A |
| * Componente tabla de datos Svelte. | | | | 5h |
| **TP.07** | Implementación de e-chart para la visualización de series. | | | A |
| * Componente e-chart en Svelte. | | | | 5h |
| * Visualización de varias series en el mismo gráfico. | | | | 5h |
| **TP.08** | Calcular la tasa de variación y visualizarla. | | | M |
| * Calcular tasa de variación en el backend. | | | | 2h |
| * Serie DTO para enviar la tasa con las series al frontend | | | | 3h |
| **TP.09** | Botón de descarga del gráfico. | | | M |
| * Botón para descargar el gráfico que se esté visualizando en el momento. | | | | 2h |
| **TP.10** | Mejorar apartado visual. | | | M |
| * Añadir estilo al frontend | | | | 6h |
| **TP.11** | Filtrado de series | | | A |
| * Formulario de búsqueda que debe incluir los siguientes criterios: palabra clave, fuente, origen, tipo de datos y periodicidad | | | | 10h |
| * Endpoint en la API que devuelva un listado de series filtrado según los criterios descritos anteriormente. | | | | 3h |
| **TP.12** | Scraper MINECO | | | A |
| * Scraper que recolecte series del Ministerio de economía e inserte las series en el sistema. | | | | 6h |
| * Datos de las fuentes coherentes y mostrados correctamente. | | | | 3h |
| Tercera fase: Finalización | | | | |
| Fecha Inicio: 29/07/2020 | | Fecha fin: 29/08/2020 | | |
| **TF.01** | Scraper INE, EUROSTAT. | | | A |
| * Scraper para fuente: INE. | | | | 6h |
| * Scraper para fuente: EUROSTAT. | | | | 6h |
| **TF.02** | Mejorar frontend – tareas extra frontend | | | B |
| * Menú principal. | | | | 5h |
| * Mejorar estilo, apartado visual del sistema. | | | | 5h |
| * Control de errores. | | | | 3h |
| **TF.03** | Mejorar backend – tareas extra backend | | | B |
| * Control de excepciones. | | | | 8h |
| * Documentación Swagger. | | | | 3h |
| **TF.04** | Avisos del sistema | | | M |
| * Aviso (pop up) al visualizar series con unidades o periodicidad distinta. | | | | 4h |

Prioridades: A – Alta, M – Media, B - Baja

Costes

El coste más importante y el único coste directo que presenta este proyecto es el salario, como este proyecto se está desarrollando en una asignatura, para el cálculo del sueldo se han usado las 300 horas determinadas por la asignatura. Para este cálculo, también se ha usado [Glassdoor](https://www.glassdoor.es/Sueldos/sevilla-programador-junior-sueldo-SRCH_IL.0,7_IM1014_KO8,26.htm) para determinar el sueldo anual de un programador Junior en Sevilla.

La fórmula usada para calcular el coste es:

->

S: Salario anual. H: Horas de la asignatura. JA: Jornada anual

Además del salario, se tiene en cuenta el coste de amortización del equipo usado durante el desarrollo. Se ha calculado aplicando un porcentaje de pérdida de valor anual del 20% al precio original sin impuestos del equipo, lo que supone la siguiente fórmula:

->

Por lo que el coste total del proyecto seria **2.871,68 €.**

## Metodología

Existen varias metodologías que se pueden usar a la hora de desarrollar un proyecto software. Estas metodologías se suelen separar entre lineales e iterativos. Los modelos lineales, llamados también modelos en cascada, se caracterizan por ser secuenciales, es decir, el proyecto pasa por las siguientes fases de manera sistemática: análisis, diseño, desarrollo, pruebas, mantenimiento. Por el contrario, los modelos iterativos se caracterizan por fragmentar los requisitos en distintos grupos y realizar una iteración para implementar cada uno de dichos grupos. Es decir, un desarrollo iterativo es similar a realizar varios desarrollos en cascada con menos carga de trabajo.

Por otro lado, el backend y el frontend van a ser desarrollados de manera individual, lo que se conoce como arquitectura desacoplada. Posiblemente, la ventaja principal de esta arquitectura es la fluidez que adquiere el sistema, ya que el intercambio de información entre backend y frontend es extremadamente ligero, en cada petición y respuesta entre el cliente y servidor solo se intercambian archivos *Json*. Además, el backend puede ser utilizado por otras interfaces web o aplicaciones, lo que añade valor al producto.

Imagen que contiene tabla

Descripción generada automáticamente

## Herramientas

### Github

Github es una plataforma de desarrollo colaborativo en la que se pueden almacenar proyectos usando Git. En este proyecto, a pesar de que no sea colaborativo al ser desarrollado por una persona, se va a usar Github para el control de versiones y la gestión de las tareas.

Además, al usar Github para almacenar el proyecto y para la gestión de tareas, se van a usar funcionalidades más avanzadas, como asociar cada subida de código o de archivos a una tarea concreta.

### Entornos de desarrollo

Se van a usar dos entornos de desarrollo distintos: IntelliJ para el backend y Visual Studio Code para el frontend.

IntelliJ es un buen IDE que ha sido seleccionado entre otros, como por ejemplo Eclipse, debido a que ofrece muchas funcionalidades útiles para desarrollar en Spring Boot. En concreto, se puede conectar directamente con Spring Initializr, lo cual es muy eficiente a la hora de crear proyectos nuevos. Además, es un IDE usado habitualmente por el desarrollador del proyecto, por lo que no hay alternativa mejor.

Se ha elegido Visual Studio Code para el frontend por ser usado habitualmente por el desarrollador del proyecto, al igual que IntelliJ, además de ser bastante popular para el desarrollo web.

### Frameworks

Spring Boot

Spring Boot es un framework de desarrollo de proyectos en Java muy conocido tanto por la comunidad de desarrolladores como por el desarrollador del proyecto. Al ser tan conocido, se minimiza el riesgo del desarrollo del backend, y permite que se comience a trabajar lo antes posible.

Svelte

Respecto al frontend, la tecnología decidida finalmente para el desarrollo ha sido Svelte. Este es un framework de interfaces de usuario muy reciente que tiene varias características interesantes. Por el contrario, aunque ofrece ventajas sobre otros frameworks gracias a la reactividad o que se necesita escribir menos código, al ser nuevo, hay muchísima menos documentación, dudas resueltas y componentes ya hechos en internet.

MongoDB

Para la base de datos se va a usar un sistema NoSQL, es decir, en vez de guardar los datos en tablas, como se hace tradicionalmente, se guardan en un esquema dinámico sin relaciones. MongoDB es capaz de devolver los datos de manera más fácil y rápida que las bases de datos relacionales, por lo que, al no tener un sistema complejo de relaciones, es el sistema más adecuado para este proyecto.

## Relación con las asignaturas del Grado

### Diseño y pruebas

Diseño y pruebas son dos asignaturas (DP 1 y DP 2) de tercer curso en las cuales se forma al alumno para ser capaz de diseñar e implementar sistemas web a partir de unos requisitos establecidos por el cliente. Estas asignaturas, a pesar de ser muy intensas, son muy importantes para aprender conceptos fundamentales para un Ingeniero del Software que se dedique a sistemas web. A lo largo de esta asignatura no solo se aprenden a diseñar, implementar, probar y desplegar aplicaciones web, si no que el alumno tendrá que pasar por todas las dificultades que supone realizar una aplicación web con muy poca experiencia.

Adicionalmente, el framework usado para esta asignatura es una versión antigua del mismo que se usa para el backend de este proyecto: Spring.

Ingeniería del Software y Práctica Profesional

Esta asignatura, al igual que Diseño y Pruebas está fuertemente ligada a este proyecto. ISPP es una asignatura de último curso en el que los grupos de alumnos tienen que crear un producto desde cero, esto implica que la idea del proyecto ha de ser propia de los alumnos.

Esta asignatura es más completa que la anterior, sin embargo, el alumnado tiene más experiencia por lo que se le puede exigir más. A pesar de no haberse utilizado en este proyecto ninguna de las tecnologías usadas en la asignatura, la arquitectura del sistema desarrollado también es desacoplada. Esto ha supuesto que ya se dominase el concepto de backend y frontend, y lo que conlleva desarrollar un producto de estas características.

Planificación y Gestión de Proyectos Informáticos

Esta asignatura se centra completamente en las fases de planificación y gestión de proyectos software, como su nombre indica, por lo que se complementa con la asignatura ISPP y DP. Esta asignatura, también realizada en grupo, tiene como objetivo proporcionar al alumno una visión clara de cómo se gestiona y dirige un proyecto software, sus contenidos trata sobre la organización, planificación y documentación de un proyecto teniendo en cuenta la normativa y legislación aplicables. Esta asignatura se destaca bastante ya que trata aspectos que no se tratan en otras asignaturas del grado.

Complementos de bases de datos

Esta es una asignatura optativa de cuarto curso cuyo objetivo es consolidar y ampliar los conocimientos sobre bases de datos adquiridos a lo largo del grado. A lo largo de esta asigatura se estudian todos los tipos y variantes de bases de datos, en concreto, las bases de datos no relacionales se estudian por primera vez en CBD. De hecho, en esta asignatura se explica porqué las bases de datos NoSQL son mejores para tratar grandes cantidades de datos y tienen más escalabilidad horizontal, lo cual ha sido el factor determinante para elegir el tipo de base de datos para este proyecto.

Acceso Inteligente a la Información

Acceso inteligente a la información es otra asignatura optativa de último curso, en esta asignatura, además de aprender a implementar interfaces de usuario y a usar Django, se enseña a extraer información desde páginas HTML estáticas en Python.

A pesar de ya haber trabajado con Crawlers en el pasado, en esta asignatura se enseña a localizar y extraer los elementos deseados de una página HTML, estos conocimientos son extremadamente útiles a la hora de construir Scrapers, por lo que esta asignatura también guarda relación con este proyecto.

# Realización del proyecto

## Backend

### Análisis

Requisitos de información

RI.01 - Se quiere guardar la siguiente información acerca de cada serie:

* Nombre.
* Periodo (fecha de inicio y fin de datos disponibles).
* Fuente (IECA/INE/EUROSTAR).
* Periodicidad (Mensual/Trimestral/Anual).

En el caso de encontrar más información útil mediante los Scrapers, se podrán añadir campos adicionales.

RI.02 - Para cada serie, hay que guardar los datos correspondientes a cada momento en el tiempo, los cuales constan de:

* Año.
* Valor.
* Periodo.

Es necesario que se guarde la información del momento en el tiempo en dos campos distintos debido a que la periodicidad de los datos puede cambiar. Por tanto, si los datos fuesen trimestrales, el año puede no variar de un dato a otro, pero el conjunto de año y periodo debe ser único para cada dato de la série.

Requisitos funcionales

RFB.01 – El backend deberá poder conectarse con la base de datos para almacenar, modificar o eliminar los datos.

RFB.02 – El backend deberá disponer de scrapers para las distintas fuentes especificadas para poder extraer los datos e insertarlos en la base de datos.

RFB.03 – El backend deberá de disponer de una API con la que poder comunicarse con el frontend. Esta API debe contener suficientes endpoints para las distintas operaciones que se realizan desde el frontend.

### Diseño

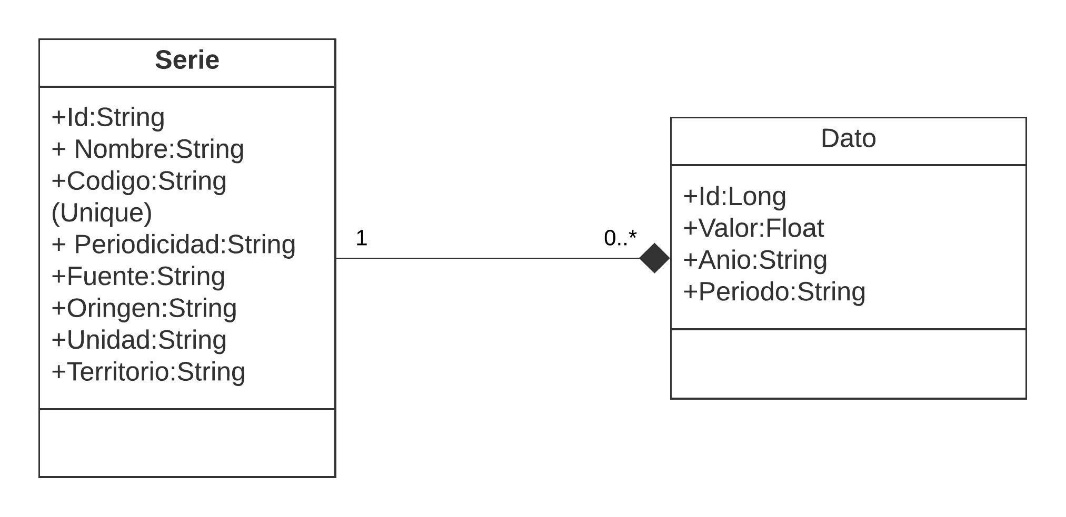
A continuación, se muestra el modelo de dominio, el cual contiene las entidades que han de ser almacenadas por el sistema:

Ilustración : modelo conceptual

Como podemos observar, la relación entre *Serie* y *Dato* no es una relación estándar, sino que es una composición. Esta es una relación especial, indica que las series están compuestas de datos, es decir, no tiene cabida una serie sin datos ni un conjunto de datos sin serie.

### Implementación

Scraper IECA

El Instituto de Estadística y Cartografía de Andalucía (IECA) tiene una web en la que permite seleccionar series para ser descargadas en distintos formatos (Excel, CSV, Json) o guardar la selección para usarla más adelante.

El primer acercamiento para crear el Scraper fue intentar crear un Crawler que recorriese las páginas de las series de datos extrayendo los datos del propio código fuente. Pero esto resultó imposible ya que la página está construida con muchas funciones JavaScript y los datos no aparecen en el código fuente.

Afortunadamente, la página permite guardar una cesta con mucha información sobre las series seleccionadas. Finalmente, el Scraper funciona recorriendo esta cesta y realizando peticiones a la web con los datos de la propia cesta. A continuación, se expone un ejemplo básico:

En la información de la cesta de la Serie X se indica que su código es: 1234, por lo que para acceder a sus datos se construye la URL: www.ieca.es/obtenerSerie?cod=1234

Así, tras recorrer la cesta en la que se incluyen todas las series disponibles, tenemos los datos de todas las series de esta fuente.

Scraper MINECO

Con esta segunda fuente ocurre como con la primera, a pesar de intentar recorrer las páginas de datos con un Crawler estas páginas no existen realmente, sino que son mostradas con funciones JavaScript, lo que hace imposible recorrerlas de manera automática.

La solución encontrada finalmente es parecida a la de IECA. Se descargan los datos completos en un directorio zip. Esta carpeta se recorre y se extraen todos los datos de cada serie.

Se han encontrado distintas dificultades respecto a la fuente anterior. La primera viene dada por el formato de los datos. Al contrario que IECA, que permite la descarga en distintos formatos conocidos, el ministerio de economía tiene su propio formato y solo permite la descarga de todas las series en ese formato, lo que ha provocado un esfuerzo adicional a la hora de extraer los datos y estructurarlos a nuestra manera. Por otra parte, la cantidad de series es muy elevada (22.000 aprox), lo que provoca una gran caída del rendimiento al extraer todos los ficheros de datos. Este último problema se solventó adaptando el algoritmo del Scraper para que extraiga la información de los ficheros individualmente.

MongoDB

La implementación de la base de datos es importante a pesar de su sencillez. El primer paso es incluir las dependencias de MongoDB en el POM de la aplicación. Gracias a Spring Initializr, este paso se realizó automáticamente en la creación de la aplicación base.

El siguiente paso es adaptar las clases del dominio a MongoDB, lo que también es sencillo, únicamente tendremos que usar la anotación @Document(collection = "examples") para especificar un nombre para la colección. Sin embargo, este paso se puede omitir, lo que provocará que el nombre de la colección sea el mismo del de la clase en cuestión. Además, queremos que MongoDB guarde un identificado ID único para cada objeto, por lo que hay que especificar ese atributo en la clase y anotarlo con @Id, mongoDB se encargará de gestionar el valor para este atributo internamente.

Otra característica de MongoDB en Spring Boot que hay que tener en cuenta es que no hace falta la existencia previa de la base de datos, la misma aplicación la generará en el caso de que no la encuentre. Por defecto, el nombre de la base de datos autogenerada es “test”, pero se puede modificar especificando un nombre en las propiedades de la aplicación mediante el siguiente parámetro:

*spring.data.mongodb.database = databasename*

Sin embargo, este no es el único parámetro que se ha tenido que especificar, para que los índices especificados en el dominio se generen en la base de datos hay que añadir este otro parámetro:

spring.data.mongodb.auto-index-creation = true

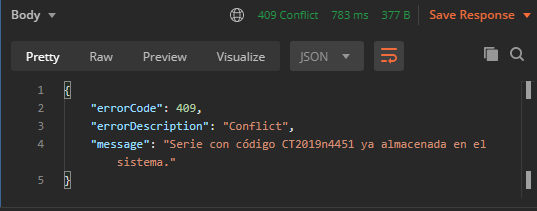
Esta opción permite añadir índices de MongoDB desde el dominio java, en concreto, se ha usado para añadir la restricción de unicidad a los códigos de las series. Esta restricción va a asegurar que no haya series duplicadas en el sistema, si se intentase insertar una serie con un mismo código, MongoDB genera una excepción tipo *“DuplicateKeyException”*. A continuación, se muestra el resultado de dicha excepción junto con el manejo de excepciones explicado más adelante:

Ilustración : Excepción por código duplicado.

API

Al tener una aplicación con arquitectura desacoplada, es necesario que el frontend y el backend se comuniquen mediante algún mecanismo. Para conseguir este objetivo, el backend se ha realizado en formato API, es decir, desde el punto de vista del frontend, va a haber una serie de endpoints disponibles con los que obtener o enviar datos al backend. Esto es muy fácil de realizar en Spring Boot, ya que únicamente hace falta poner las siguientes anotaciones a los controladores que vayan a formar parte de la API:

*@RestController*: Identifica al controlador como parte de la API.

*@RequestMapping*: Indica la ruta hacia el controlador. p.e: (@RequestMapping(“/serie”).

*@GetMapping*: Se colocan sobre los métodos del controlador para especificar el tipo de solicitud HTTP del endpoint y especificar la ruta si es necesario. Existen distintos tipos de solicitudes, los más usados son *Get*, *Post*, *Put* y *Delete*. p.e: (@GetMapping("/{id}")).

Una vez configurada la aplicación con estas anotaciones ya está lista para interactuar con el frontend. Pero para que las dos aplicaciones se puedan comunicar es necesario habilitar el CORS desde el backend hacia el frontend, ya que si no las peticiones serán bloqueadas por este mecanismo.

Hay distintas alternativas para habilitar el CORS, la más simple es especificar mediante la anotación *@CrossOrigin*, la cual se puede colocar en los métodos del controlador o en todo el controlador para aplicar los ajustes a todos los métodos a la vez. No obstante, se ha optado por aplicar una configuración global para toda la aplicación. Esto se ha conseguido modificando la clase de configuración de la aplicación, añadiéndole la siguiente línea de código:

*registry.addMapping("/serie/\*\*").allowedOrigins("http://localhost:5000");*

Esto permite el acceso desde localhost:5000 al recurso serie y todos sus endpoints.

Control de excepciones

Para que los errores que ocurran durante la ejecución de la aplicación estén controlados y los mensajes de error sean comprensibles se ha hecho uso del manejo de excepciones que ofrece Spring Boot, además de crear excepciones propias de la aplicación y mensajes de error personalizados.

Sin embargo, hay que especificar que excepciones queremos controlar a Spring, esto se puede hacer a nivel de controlador, pero como ocurre también con el CORS, es más interesante solucionar este problema a nivel de aplicación, esto se consigue especificando todas las excepciones en una clase aparte marcada con la anotación *@ControllerAdvice.*

En la clase de configuración hay que crear un método para cada excepción que se quiera tratar y marcar el método con la siguiente anotación:

*@ExceptionHandler({ ExampleException.class})*

Esto permite capturar la excepción y tratarla de forma individual, así se puede especificar el código de error HTTP correspondiente a la excepción y crear una respuesta más explicativa.

Captura de pantalla de un celular

Descripción generada automáticamenteA continuación, se pueden observar las diferencias entre controlar una excepción o no, en concreto, este caso es el error provocado por intentar obtener una serie inexistente:

Ilustración : Excepción no controlada

Como se puede observar en la imagen, el error es poco descriptivo y el estatus es “*500 Internal Server Error”*, cuando debería ser “*404 Not Found”* debido a que no es un error del sistema, si no que no se ha encontrado el recurso solicitado.

Captura de pantalla de un celular

Descripción generada automáticamenteA continuación, se muestra el comportamiento de la aplicación con el manejo de excepciones implementado, se puede observar que el mensaje es mucho más explicativo y el status mostrado es el correcto.

Ilustración : Excepción controlada

Swagger

Swagger es una herramienta muy fácil de usar que permite documentar APIs. La finalidad de usar esta herramienta es generar una documentación clara y completa de cómo usar la API que ha sido desarrollada. Swagger da uso de las anotaciones vistas anteriormente, como *@RestController*, para identificar los controladores y métodos que van a ser documentados.

Para configurar Swagger en Spring Boot, únicamente hay que crear un método *Docket* en la clase de configuración. En este punto, podemos definir algunos parámetros para controlar el comportamiento de Swagger, pero en este proyecto se ha dejado la configuración básica ya que se quiere documentar toda la API.

Una vez está configurado, Swagger nos permitirá usar comentarios en el código para documentar los métodos de la API y su funcionamiento, es decir, a través de esta herramienta, las personas que vayan a dar uso de la API, serán capaces de visualizar un listado de todos los endpoints que ofrece, sus definiciones, restricciones y especificaciones, adquiriendo así una idea bastante buena de cómo usar la API. Esto es bastante útil y que normalmente se da el caso de querer usar APIs de terceros pero encontrar documentación mal hecha o incompleta.

Además, Swagger ofrece una interfaz gráfica llamada Swagger UI para poder visualizar toda esta información, a continuación, se muestra una imagen de ejemplo:

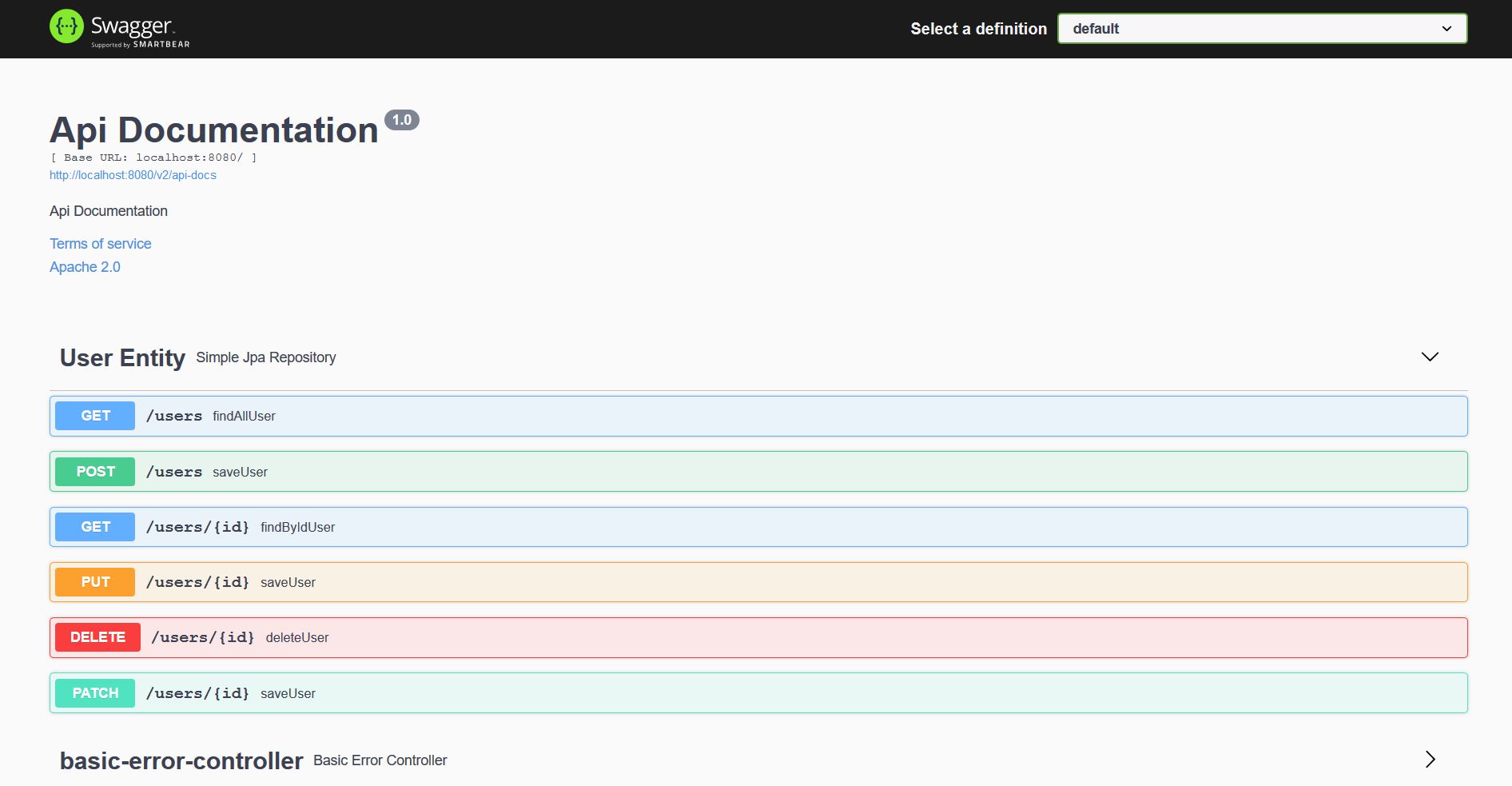


Ilustración : Ejemplo de swagger

Calidad del código

Para controlar la calidad del código se han usado dos plugins para IntelliJ: checkstyle y sonarlint.

Checkstyle es una herramienta que analiza el código y avisa al desarrollador si este no cumple con las reglas de codificación establecidas. Esta herramienta es muy útil ya que es común dejar pasar pequeñas faltas en el código que, al compilar la aplicación, son imperceptibles a la hora de ejecutar la aplicación. Además, esta herramienta ayuda a mantener el código limpio y legible.

Sonarlint es una herramienta que cumple una función similar a Checkstyle, sin embargo, esta herramienta se centra más en los bugs y problemas de seguridad que puedan aparecer.

Al ser plugins de IntelliJ, ambas herramientas pueden analizar el código mientras es escrito, lo cual es una gran ventaja para el desarrollador.

### Pruebas

Para asegurar el correcto funcionamiento de los servicios y controladores se han realizado tests unitarias para cada método que aparece en dichas clases para comprobar que funcionan según se espera.

## Frontend

### Análisis

Requisitos funcionales

RF.01 – El sistema permitirá al usuario listar las series almacenadas.

RF.02 – El sistema permitirá al usuario filtrar la lista de series.

RF.03 – El sistema permitirá al usuario mostrar una o varias series en un gráfico.

RF.04 – El sistema permitirá al usuario mostrar una o varias series en una tabla.

RF.05 – El sistema permitirá al usuario descargar la gráfica que esté visualizando.

RF.06 –El sistema permitirá al usuario descargar la tabla que este visualizando.

Los usuarios del sistema son trabajadores acostumbrados a trabajar con series de datos, por lo que tampoco es necesario realizar una interfaz excesivamente limitada, sin embargo, sí que debe ser intuitiva y fácil de usar.

### Diseño

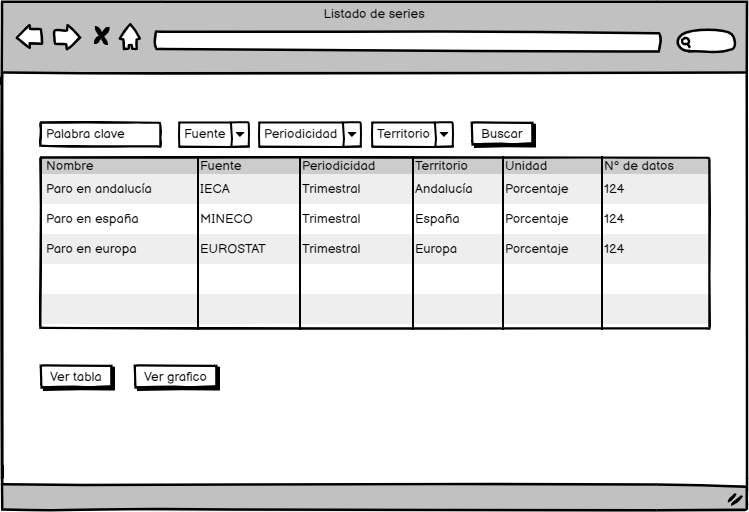
A continuación, se muestran los mockups de las pantallas más importantes de la aplicación, estos representan una imagen aproximada que puede ser usada como guía de la información y estructura que tendrá el frontend una vez se desarrolle.

Ilustración : Mockup listado de series.

Esta es una de las pantallas principales del sistema, en la cual los usuarios deberán poder visualizar las series almacenadas en el sistema. Además, los usuarios deben ser capaces de realizar una selección de series para visualizarlas en una tabla o un gráfico. Las pantallas siguientes son las correspondientes las visualizaciones.

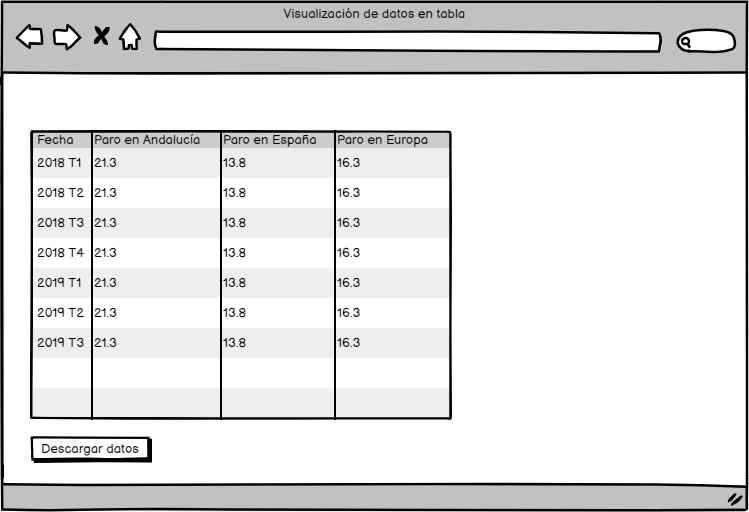


Ilustración : Mockup tabla de datos

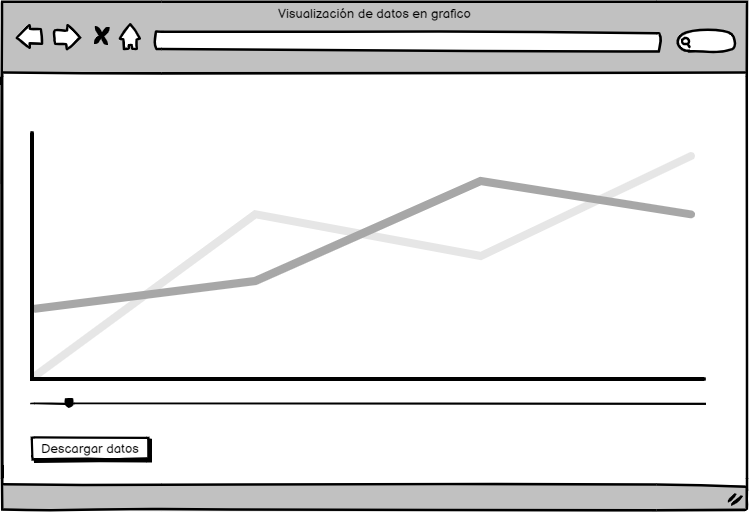


Ilustración : Mockup diagrama de líneas

### Implementación

Svelte

Svelte, como se explica anteriormente, es un framework nuevo sin mucha documentación aún en internet más allá de la oficial. Esto ha supuesto, junto con la falta de experiencia en frontend del desarrollador, que se inviertan muchas horas en investigación. Sin embargo, una vez se aprenden las nociones básicas de Svelte, este resulta ser muy potente, ya que se pueden implementar tareas con menos líneas de código de las que harían falta con otro framework como React. Además, Svelte permite añadir estilo a los componentes desde el mismo fichero, no hace falta crearlo apare, esto es muy útil si se quiere aplicar un estilo concreto solo a los componentes de ese mismo fichero, ya que el resto de la aplicación quedará fuera del alcance.

Por otro lado, Svelte permite realizar declaraciones reactivas. Estas declaraciones provocan una actualización del DOM al cambiar su valor. Esto simplifica mucho el desarrollo de componentes que se actualicen a menudo y deba verse reflejado en la página. Por otro lado, además de variables, también se pueden declarar sentencias como puede ser un *console.log*  o un bloque *if.*

Pero la reactividad de Svelte no acaba ahí, aunque este tipo de declaraciones sean útiles no han sido necesarias, ya que la reactividad de Svelte también es puesta en movimiento a través de las asociaciones. Esto quiere decir que, al asociar un valor a una variable también provocará que se recargue la página, visualizando así los cambios pertinentes. Sin embargo, al ser disparada por las asociaciones, el hecho de añadir un elemento a un *array* o realizar cualquier cambio sobre un objeto que no sea una asignación, la reactividad no se activará. No obstante, hay una alternativa muy simple a este problema, si, por ejemplo, queremos añadir un elemento a un *array* y que el sistema se comporte como si fuese una asociación, podemos asociar el *array* en cuestión a sí mismo, lo que no provocará ningún cambio en los datos, pero solucionará el problema.

Componentes externos

Para mantener un estilo de componentes coherente y de calidad, se ha optado por utilizar los componentes de Svelte Material UI. Material UI es una colección de componentes ya hechos que tiene diferentes versiones para utilizarse en los distintos frameworks (React, Angular, Svelte). Utilizar componentes ya existentes supone un ahorro de tiempo considerable teniendo en cuenta la poca experiencia desarrollando componentes en frontend y en svelte, sin embargo, al ser svelte un framework tan joven, estos componentes apenas han llegado a su primera versión estable, son bastante limitados y apenas están documentados.

El uso de estos componentes es bastante simple, únicamente hay que instalarlo en la aplicación mediante *npm install*, importarlos en el código y ya se podrán usar como cualquier otro componente.

E-Charts

Para la visualización de los datos en gráficos se tuvieron en cuenta varias librerías. Finalmente, se optó por Apache E-Charts, una librería muy potente al ofrecer una gran cantidad de personalización de los gráficos.

Aunque el gráfico que se muestra en la aplicación es relativamente sencillo, solo han de mostrarse los datos de las distintas series en un diagrama de líneas, su implementación se ha llevado a cabo con varias dificultades. Para usar esta librería es necesario importarla como cualquier otro componente, una vez instalada, para mostrar los gráficos, es necesario utilizar un elemento SVG o Canvas como contenedor.

Llegados a este punto, debemos inicializar el echart y especificar sus opciones. La inicialización es sencilla, solamente hace falta una línea de código usando la importación de echarts y usando el método *init*. No obstante, la especificación de opciones es más compleja, se realiza con el método *setOptions*, el cual recibe como parámetro un objeto JavaScript con todas las especificaciones del gráfico, incluyendo los datos. A continuación, se explican todas las propiedades del objeto que se han usado:

* **Legend**: Atributo que habilita el componente de la leyenda del diagrama, la leyenda consta del título de cada serie mostrada y su indicador de color.
* **DataZoom**: Atributo que habilita el componente que permite hacer zoom al gráfico, esta característica es especialmente útil ya que se trata con series con multitud de datos.
* **Toolbox**: Atributo que habilita la barra de herramientas del diagrama, hay varias herramientas disponibles, pero en este caso únicamente se ha habilitado la de descarga del gráfico.
* **Tooltip**: Atributo que habilita el tooltip. Esto es una descripción emergente en la localización del ratón, en este caso se usa para mostrar los datos de las series en el momento correspondiente a la localización del ratón.
* **xAxis**: Atributo en el que se definen las características de eje x del gráfico, según la implementación realizada se especifica que los datos de este eje son de tipo categoría, y se introducen los datos correspondientes, es decir, el listado de momentos que componen las series de datos mostradas.
* **yAxis**: Atributo en el que se especifica el tipo de datos que se va a mostrar en el diagrama, al ser valores numéricos, se utiliza la opción *value*. Al contrario que en el eje x, en este atributo no se introducen datos.
* **Series**: Este es el atributo más importante, en él se introducen todas las opciones de las series y sus datos correspondientes.

Se puede encontrar un listado completo con todas las opciones y sus descripciones completas en la página oficial de [Apache Echarts](https://echarts.apache.org/en/option.html).

Problemas encontrados

El principal problema encontrado a lo largo del desarrollo del frontend ha sido la falta de documentación, lo cual ha supuesto una carga de trabajo de investigación superior a la esperada. Esta falta de documentación no solo se aplica a los conceptos de Svelte, si no a los distintos componentes usados.

Por otro lado, durante la implementación de la visualización de series con e-charts también se han encontrado problemas. En concreto, a pesar de que todas las demás características funcionasen correctamente, el tooltip no se mostraba con la configuración explicada en el apartado anterior. Para solucionar este problema se intentaron varias alternativas, la primera fue utilizar un componente que sirve de envoltorio para e-charts, este componente genera un Canvas con el componente echart listo para pasarle las opciones. Sin embargo, esto no solucionó el problema, ya que es un problema de Svelte en sí mismo, a partir de la versión 3.15 de Svelte, el tooltip deja de ser visible. No se ha encontrado el motivo por el cual no se muestra en las versiones más recientes, por lo que la solución a este problema ha sido alterar a dicha versión de Svelte.

Por último, se han encontrado algunos bugs visuales que han tenido que ser corregidos con CSS.

### Pruebas

Debido a la implementación del frontend, al no tener una navegabilidad mediante URLs, no se han realizado test.

# Conclusión

Respecto a los objetivos establecidos al comienzo de este documento, sería acertado decir que se han alcanzado la totalidad de ellos en mayor o menor medida. Se ha desarrollado una aplicación web desde cero, eligiendo las tecnologías y las metodologías a usar, lo que ha supuesto que el alumno repase todo el conocimiento adquiridos en estos campos durante el grado. Además, se han realizado los requisitos para satisfacer al cliente, esto, junto con una continuación del producto, puede llegar a suponer una oferta laboral, lo que es algo muy importante dado el contexto de ser la última asignatura de la carrera.

A continuación, se encuentra una valoración personal de distintas cuestiones tratadas a lo largo de este proyecto:

¿Es recomendable utilizar un framework nuevo?

La respuesta corta sería no. Sin embargo, hay distintos aspectos que pueden hacer cambiar esa respuesta. Utilizar un framework nuevo, como se explica varias veces a lo largo del proyecto, implica tratar con tecnologías desconocidas y falta de documentación y soporte. No obstante, también es cierto que un framework nuevo es innovador y seguramente aporte muchas ventajas al sistema que se desarrolle. Por lo tanto, la respuesta larga sería que dependiendo de las ventajas que ofrezca el framework nuevo para el producto que se va a desarrollar sí se podría llegar a utilizar. Siempre que se tenga en cuenta que hay que realizar una formación previa al comienzo del desarrollo.

¿Son útiles los Scrapers dada la posible dificultad de implementación?

Es cierto que dependiendo de lo que se quiera extraer la implementación de un scraper puede llegar a ser muy tediosa y compleja, sin embargo, estos programas se realizan al no haber alternativas para obtener los datos deseados. Por lo que sí, los scrapers son útiles siempre que se implementen correctamente.

¿Qué ventajas ofrece una arquitectura desacoplada?

Desde mi punto de vista personal, la arquitectura es el nuevo estándar de desarrollo de aplicaciones web y es recomendable aplicar esta arquitectura o una arquitectura de microservicios en la gran mayoría de productos. Esto es debido a que este tipo de arquitectura aumenta el valor del producto sin aumentar el coste, además de permitir versatilidad a la hora de escoger la tecnología para el frontend y el backend.

Siendo más específico, al usar esta arquitectura, normalmente se crea una API para la comunicación con el frontend, lo que supone un valor añadido al producto ya que dicha API podrá ser utilizada por terceros o por aplicaciones que se desarrollen en un futuro.

Por último, el hecho de realizar un proyecto desde sin ningún compañero ni apoyo ha supuesto supone un aumento de confianza personal y que se ha alcanzado cierta madurez como desarrollador, lo que se puede considerar como un objetivo extra cumplido. Este proyecto demuestra que se han adquirido las capacidades necesarias a lo largo del grado, y que el alumno está preparado para comenzar la vida laboral como Ingeniero de Software.

# Anexo I – Glosario de términos.

**Frontend**: Parte de una aplicación web que contiene la interfaz gráfica con la que interactúa el usuario. Se corresponde con la parte cliente.

**Backend**: Parte de una aplicación web que se encarga de administrar la base de datos y donde se incluye la lógica de negocio. Se corresponde a la parte servidor.

**Serie de datos**: En el contexto de este documento, las series de datos son las series estadísticas que se quieren extraer de las distintas fuentes para ser almacenadas y mostradas por la aplicación. Estas series son conjuntos de datos recogidos por el estado que se encuentran públicos en internet.

**API**: Interfaz de programación de aplicaciones, llamada comúnmente API, es el conjunto de edpoints que ofrece un software para ser usado por otro software como una capa de abstracción. En el contexto de la aplicación, la API supone el conjunto de endpoints expuestos por el backend para que el frontend interactúe con él.

**Endpoint**: Función concreta de una API, cada endpoint debe tener una ruta única para ser diferenciados entre ellos dentro de una misma API.

**Scraper**: Programa que se encarga de extraer información concreta de una web, normalmente recorriendo el código HTML, manteniendo la información deseada y desechando la demás.

**Crawler**: Programa que inspecciona las páginas web de forma sistemática y ordenada.

**CORS**: intercambio de recursos de origen cruzado, es un mecanismo característico de las aplicaciones web que rettringe el acceso a recursos si son solicitados desde un dominio distinto. Esto sucede a menudo en la arquitectura desacoplada, por lo que hay que configurar el cors para permitir que el frontend acceda a los recursos del backend.

**NoSQL**: Tipo de bases de datos que se caracteriza por su flexibilidad y no contemplar relaciones entre tablas de datos. Son reconocidas por su fácil desarrollo y su rendimiento manejando grandes cantidades de datos.

**Reactividad**: Capacidad de un software de mantener un entorno actualizado de manera asíncrona.

# Anexo II – Control de cambios.

|  |  |  |
| --- | --- | --- |
| **Descripción** | **Localización** | **Fecha** |
| Añadida nueva tarea de control de excepciones en el backend | Planificación. | 27-07-2020 |
| Subtarea de filtrado de series adelantada a la fase de prototipo como tarea completa | Planificación. | 28-07-2020 |
| Nueva tarea añadida: Scraper mineco. Movida subtarea “Datos de las fuentes coherentes y mostrados correctamente” a Scraper mineco. | Planificación. | 28/07/202 |
| Nuevo campo incluido en las Series: Origen. | Backend – Diseño. | 19/08/2020 |
| Corrección de costes. | Planificación – costes. | 19/08/2020 |