

Creación de un dashboard para usuarios del ticket digital de Mercadona: que permita consultara evolución de productos habitualmente adquiridos, el coste de sus compras por períodos temporales y el monto de gastos por áreas de producto

Santiago Sánchez Sans

Ciclo formativo en desarrollo de aplicaciones web

Memoria del Proyecto de DAW

IES Abastos. Curso 2024/25. Grupo 7X. XX de Junio de 2025

Tutor Individual: Carlos Furones

Índice general

1. Introducción	2
1.1. ¿Qué es el ticket digital de Mercadona?	2
1.2. Identificación de necesidades	2
1.3. Objetivos del proyecto	2
2. Diseño del proyecto	4
2.1. Requisitos Funcionales	4
2.1.1. Requisitos de la aplicación	4
2.1.2. Requisitos de los usuarios	5
2.2. Stack tecnológico	5
2.2.1. Front-End: HTML, CSS y Javascript	5
2.2.2. back-end: Java (SpringBoot) y Python	5
2.2.3. BBDD: MySQL y MongoDB	5
2.3. Diagramas de la aplicación	6
3. Desarrollo del proyecto	7
3.1. GitHub del proyecto	7
3.2. Entornos de desarrollo	7
3.3. desarrollo del front-end	8
3.4. desarrollo back-end	8
3.4.1. Estructura de la aplicacion	8
3.4.2. Autenticacion	8
4. Evaluación y Conclusiones Finales	13
5. ANEXO	14
5.1. Flujo de trabajo habitual en git	14
5.2. Diferencias de seguridad: JWT vs SESSID en cookies seguras	15
5.3. Clases para crear y verificar JWTs	16
5.3.1. Clase JWT	16
5.3.2. Clase Refresh Token	16
5.3.3. Clase Access Token	16
5.4. CODIGO 5	17
7. Bibliografía	17

Introducción

1.1. ¿Qué es el ticket digital de Mercadona?

Mercadona implementa un sistema de tickets digitales que vinculan la tarjeta de débito a un correo electrónico. Cualquier usuario del supermercado que quiera utilizar el ticket digital solamente deberá facilitar estos dos datos y el supermercado le enviará por correo electrónico los tickets de las posteriores compras hechas en cualquier establecimiento de Mercadona.

Las ventajas para el usuario son evidentes: no se pierden los tickets de cara a devoluciones y el cliente del supermercado no debe esperar a la impresión del ticket.

1.2. Identificación de necesidades

Los tickets de cada usuario se acumulan de forma recurrente en el correo electrónico y con un formato estructurado (los asuntos son predecibles e incluyen las fechas) y dentro de cada correo de un ticket digital se encuentra un PDF con el desglose de la compra (producto, unidades vendidas, establecimiento, etc).

Esta información se acumula en el correo del usuario pero a pesar de ser una información estructurada su acceso para el usuario no es simple: no puede visualizar lo que ha gastado, ni el precio de los productos y de su evolución, ni los supermercados en los que ha comprado, ni las veces que lo ha hecho, etc.

1.3. Objetivos del proyecto

Este proyecto quiere responder a estas necesidades. Para ello se plantea la Creación de un dashboard o “cuadro de mando” front-end para que un usuario del ticket digital de Mercadona pueda visualizar la evolución de precios de los productos adquiridos, el coste promedio de sus compras por períodos temporales y sus distribuciones de gastos a partir de los tickets digitales guardados en una base de datos.

Los **Objetivos principales** del proyecto son mostrar al usuario:

- **Evolución de precios** (inflación) a lo largo del tiempo en los productos

habitualmente comprados en el mismo establecimiento¹.

- **Evolución del gasto** total del usuario a lo largo del tiempo.

¹La evolución de precios se mostrará solamente para un mismo centro de Mercadona, dado que distintos centros pueden cambiar los nombres de los productos (por ejemplo, en Cataluña. . .).

Diseño del proyecto

Para implementar los objetivos principales de los que hemos hablado en la sección 1.3 hemos proyectado una serie de requisitos funcionales de la aplicación.

2.1. Requisitos Funcionales

2.1.1. Requisitos de la aplicación

REQUISITO A: Mostrar evolución de los precios de los productos unitarios adquiridos con más frecuencia (visualizable en un gráfico donde en X tendremos el tiempo y en Y el precio en euros). Para los productos de precios muy variables (productos a granel, como frutas, etc.), se mostrará la evolución del precio por kg a lo largo del tiempo.

REQUISITO B: Coste total de la cesta de la compra del usuario a lo largo de distintas ventanas temporales (por meses, períodos de 3, 6 meses y un año), independientemente del centro de Mercadona en el que se compre (todos juntos).

REQUISITO C: Al lado de este mismo coste total mostrado en B), se incluirá un diagrama de queso (o de sectores) desglosando qué porcentaje del dinero se ha destinado a cada una de las siguientes categorías: verdura y hortalizas, frutas, huevos y lácteos, agua y bebidas, aceite y especias, carnes, pescado, hogar e higiene personal. Para ello, dado que no tenemos categorizados todos los productos de Mercadona ni podríamos hacerlo por falta de una lista exhaustiva y de tiempo, se usará un modelo predictivo con word embeddings (módulo Spacy) y cosine similarity (sklearn) para encontrar distancias pequeñas entre las descripciones de los tickets y las categorías, facilitando así la clasificación.

REQUISITO D: Un botón “Actualizar” permitirá al usuario refrescar los datos desde el servidor cuando haya añadido nuevas compras. También podríamos permitir que los PDFs descargados en el servidor se almacenen en una carpeta local del usuario para que pueda verificar la extracción de los datos.

REQUISITO E: Hacer un sistema front-end y back-end que permitan redirigir a los usuarios rápidamente a un registro de forma inteligente. Nos inspiraremos en el sistema de registro e iniciar sesión de Netflix (**POSAR DIAGRAMA A L'ANEX SOBRE EL SISTEMA DEL NETFLIX I EL NOSTRE**)

2.1.2. Requisitos de los usuarios

El correo electrónico y la contraseña de la cuenta de Gmail de alguien que sea usuario del ticket digital de Mercadona y tenga decenas de tickets digitales por analizar, con compras estables y productos recurrentes.

Nota: En la demo se proporcionarán ya muchos tickets digitales (tickets míos, que cederé para mostrar la utilidad de la aplicación). No será necesario recurrir a la extracción de datos de otro usuario de ticket digital. Se mostrarán un mínimo de tickets digitales en un mismo centro de Mercadona para poder evidenciar la evolución de precios y gastos.

2.2. Stack tecnológico

Hemos escogido un stack tecnológico que permite que seamos fieles a los requisitos funcionales que nos hemos marcado para la aplicación:

2.2.1. Front-End: HTML, CSS y Javascript

Se ha usado HTML, CSS y JavaScript. - Para la visualización de gráficos se usará una librería javascript: <https://www.chartjs.org/>. Entre otras cosas se utilizará para hacer los gráficos de la evolución de precios por producto.

2.2.2. back-end: Java (SpringBoot) y Python

- Back-end con Java (springboot para el login y la autenticación de usuarios: con este framework guardaremos datos en la BBDD mySQL).

- Python dentro de un contenedor docker (o python a secas, para descargar los pdfs del correo) y parsear el contenido de los mismos: con sklearn, numpy y spacy que luego se podrán pasarlos a la BBDD mongoDB.

2.2.3. BBDD: MySQL y MongoDB

Para guardar los datos de los usuarios se debe usar un sistema de gestión de base de datos relacional. Hemos escogido MySQL dado que es el que hemos visto en el grado superior y estamos bien versados en ello.

Sin embargo, los productos de Mercadona no los conocemos de antemano ni tenemos una lista exhaustiva de los mismos. Además, el número de productos que

se pueden encontrar en un ticket varía en cada compra, por lo que no podemos usar una base de datos relacional tradicional como MySQL o PostgreSQL por que se trata de información no estructurada. En su lugar, usaremos MongoDB, una BBDD NoSQL que almacena datos en formato JSON y permite, además, búsquedas eficientes.

Para optimizar el backend, intentaremos que un usuario pueda consultar repetidamente sus compras sin sobrecargar el servidor. La primera vez que consulte sus datos, estos se descargarán y almacenarán en localStorage del cliente. En consultas posteriores, los datos se obtendrán directamente de localStorage sin necesidad de hacer peticiones al servidor. Evaluaremos la viabilidad de este sistema durante el desarrollo; Esto es la fase de diseño y como tal, **PUEDE QUE EN LA FASE DE DESARROLLO CAMBIE**, en caso de no ser factible, las consultas se harán directamente en MongoDB.

2.3. Diagramas de la aplicación

Fer un diagrama guapo de tots els components de l'aplicació.

Desarrollo del proyecto

3.1. GitHub del proyecto

Para desarrollar este proyecto se ha trabajado con GitHub y git. Dado que no ha habido trabajo en equipo no se han utilizado pull requests a la rama main sino simplemente se ha seguido la estrategia de crear ramas de característica y, una vez son satisfactorias, hacer un merge en la rama main en local.

Un flujo de trabajo habitual es mediante ramas de característica (puede verse anexo 5.1). También puede verse el GitHub del proyecto a continuación. Dentro del readme del proyecto encontraréis instrucciones para su descarga, clonado y “despliegue” de sus componentes en vuestro ordenador personal si así lo deseáis.

Link al repositorio	https://github.com/blackcub3s/mercApp
Página desplegada	TO DO

Cuadro 3.1: Enlaces importantes del proyecto.

3.2. Entornos de desarrollo

Para el back-end de Java con SpringBoot se ha utilizado el editor Java **IntelliJ Idea community edition** que expone el backend en el puerto **8080**: se han utilizado extensiones necesarias para correr el proyecto que permiten sacar provecho de Lombok sin las cuales correr el proyecto en IntelliJ fallará.

Para el frontend se ha utilizado **VScode**, con la extensión live server para poder hacer llamadas al back-end directamente desde el puerto **5500**. Esto podría hacer las veces de una CDN donde podrían estar alojados los archivos estáticos (HTML, CSS y JavaScript).

3.3. desarrollo del front-end

3.4. desarrollo back-end

3.4.1. Estructura de la aplicacion

TO DO pom.xml application.properties controller service etc.

3.4.2. Autenticacion

3.4.2.1. método utilizado: JWT

Para autenticar y autorizar a los usuarios no utilizaremos sesiones. Las sesiones, tal y como vimos en la asignatura de desarrollo web entorno servidor, requieren guardar un estado en el servidor (si tenemos 100 usuarios conectados necesitamos rastrear 100 personas en el servidor) y un identificador de sesión en una cookie segura con HttpOnly puesto a True guardada en el navegador de cada uno de los usuario conectados que lo identifica en relación al servidor.

Sin embargo, existe un método de acceso por token más escalable que no requiere guardar sesiones en el servidor (es decir, es un método “stateless” o sin estado) con el que nos basta tener solamente la Cookie Segura para guardarlo y ya está. Es un token que está autocontenido: es decir, puede contener ya el ID de usuario, nombre de usuario, roles que luego permitirán dar permisos o no en el servidor para acceder a determinadas APIs o recursos, etc. En definitiva, por todo ello, nos facilita tener una autenticación más eficiente y un control del acceso sin necesidad de almacenar sesiones en el servidor.

Este método se llama JSON Web Token (JWT) y tiene su información encriptada o **firmada digitalmente** con SHA256 mediante una clave privada que tenemos nosotros en el servidor y que es igual para todos los tokens que generemos: la firma digital que emana de esta clave para generar los tokens **será inválida** si un atacante ha modificado el token y nos lo devuelve al servidor; con ello, el servidor rechazará la integridad del token y evitará que pueda acceder a recursos de usuario.

JWT no es perfecto, por supuesto. Una desventaja del JWT es que una vez puesta una fecha de expiración el desarrollador ya no la puede cambiar. En las sesiones del servidor se pueden extender las sesiones si se detecta actividad del usuario, acortarlas si pasa justo lo contrario o incluso cerrar la sesión de un usuario en remoto; pero con JWT no es posible: una vez creado el Token la fecha de actividad del mismo no se puede modificar (porque no puedes invalidar un token ya existente!). La solución

que se propone JWT es tener dos tokens almacenados en el cliente: el “access token” que es el que permite autenticar y autorizar, del que hemos hablado hasta ahora; y el “refresh token”, que lo que hace es validar si el token de refresco sigue siendo válido con el servidor, y si lo es, entonces permite que el servidor emita un nuevo “access token” para el usuario. De este modo el usuario no deberá estar generando nuevos tokens constantemente. En este trabajo solo utilizaremos “access tokens” y ya está. Cuando un usuario se des loguee, lo que haremos será borrar el access token y ya está.

En resumen, **las ventajas** que tiene JWT vs uso de sesiones (assumiendo que tanto JWT como SESSID se guarden en una cookie segura) son las siguientes:

- **No depende del almacenamiento en el servidor**
- **Firmado digitalmente**
- **Mayor control sobre el acceso**
- **y descentralización**
- **Menos carga para el servidor**

Y la **desventaja** más evidente que tiene JWT, es, en nuestra opinión, de cara a su complejidad¹, pues para tener un buen equilibrio entre facilidad de uso y seguridad es necesario almacenar dos tokens en el cliente para conseguir que uno se renueve (el token de acceso y el de refresco):

- **Caducidad y renovación de tokens:**

3.4.2.2. ¿Qué compone un JWT?

El JWT se compone de tres partes. **los headers, el payload y la signatura**. En la página <https://jwt.io/>, como veremos después, se pueden validar tokens [1] y ver su estructura interna, a saber:

- **Los headers:** Aportan información sobre el algoritmo que lo encriptó.
- **El payload:** Es donde está la información que nos interesa del token: el sujeto que lo generó (“sub”), el momento en que se generó el token (“iat”, o “issued at”) y la fecha de expiración (“exp” o “expiration time”). También podemos tener ahí otros pares clave valor que podremos querer definir, por ejemplo, que contengan el id del usuario y sus roles o permisos que son los que nos permitirán dejar que un determinado usuario pueda consultar o no ciertos recursos.
- **La signatura:** Es la parte que garantiza la integridad del token y evita que sea alterado por terceros. Se genera aplicando un algoritmo de hash (en nuestro caso el SHA256) a la combinación del header y el payload, junto con la clave

¹Se puede ver una tabla de diferencias más en profundidad, especialmente en materia de seguridad en el anexo 5.2)

secreta que solo conoce el servidor (es lo que permitirá al servidor rechazar el token si no es válido -i.e. ha sido manipulado).

3.4.2.3. Implementación de JWT en java SpringBoot

Para poder implementarlo añadimos la dependencia **jjwt** en **pom.xml** que es la que nos permite definirlo.

```
<dependency>
    <groupId>io.jsonwebtoken</groupId>
    <artifactId>jjwt</artifactId>
    <version>0.12.6</version>
</dependency>
```

En el proyecto se han creado tres clases dentro de la ruta **src/main/java/miApp.app/seguretat/jwtseguretat** denominadas **JwtUtil**, **AccessToken** y **RefreshToken**.

En la primera clase hemos creado un método que obtiene las claims (pares clave valor, la carga útil) de un JWT y en el constructor hemos creado la definición de una clave privada con la que derivar todas las instancias que hagamos de esa clase -es decir, todos los tokens que hagamos con esa contraseña-. De esta hemos heretado las otras dos: La subclase que nos genera el **token de acceso**, **AccessToken**; y la que nos genera el **token de refresco**, **RefreshToken**. A continuacion podéis ver la clase **AccessToken**:

```
@Component
public class AccessToken extends JwtUtil {

    private static int tExpM; //minuts per a expirar el token

    public AccessToken() {this.tExpM = 10;}

    //FINALITAT: Generar un JWT d'accés.
    public String genera(String correu, int idUsuari, byte permisos) {
        Map<String, Object> dadesExtraApayload = new HashMap<>();
        dadesExtraApayload.put("permisos", permisos);
        dadesExtraApayload.put("idUsuari", idUsuari);

        return Jwts.builder()
            .setClaims(dadesExtraApayload) //dades customitzades
            .setSubject(correu)             //guardo nom subjecte (clau "sub")
            .setIssuedAt(new Date())         //data creacio (clau "iat" payload)
            .setExpiration(new Date(System.currentTimeMillis() + (tExpM*60*1000)))
            .signWith(SignatureAlgorithm.HS256, clauSecreta.getBytes())
            .compact();
    }
}
```

```
}
```

Y con la función **genera()** de la clase `AccessToken`, con los parámetros necesarios, podemos generar un token de acceso en “`accesJWT`”:

```
AccessToken accessToken = new AccessToken();
String accesJWT = accessToken.genera(
    "santo@gmail.com", //campoSub
    2, //idUsuari
    1 //permisos
);
```

El token de `AccessJWT`, que es el que visualizamos a la izquierda, podremos ver que está estructurado en sus tres partes a la derecha de la figura 3.1) siendo la más importante la que está en color rosa, la payload:

Encoded

PASTE A TOKEN HERE

```
eyJhbGciOiJIUzI1NiJ9.eyJwZXJtaXNvcyI6MSwlaWRVc3VhcmkiOiJIsInN1YiI6InNhbnRvQGdtYWlsLmNvbSIsImIhdCI6MTc0MzE1MjM1MCwiZXhwIjoxNzQzMTUyOTUwFQ.ixC6NKkuG99zrxLcrgxfjRRKi8NgYbUrirFgPMEcUC0
```

Decoded

EDIT THE PAYLOAD AND SECRET


HEADER: ALGORITHM & TOKEN TYPE
<pre>{ "alg": "HS256" }</pre>
PAYLOAD: DATA
<pre>{ "permisos": 1, "idUsuari": 2, "sub": "santo@gmail.com", "iat": 1743152350, "exp": 1743152950 }</pre> 
VERIFY SIGNATURE
<pre>HMACSHA256(base64UrlEncode(header) + "." + base64UrlEncode(payload), your-256-a8f7d9g0b6c3r) <input type="checkbox"/> secret base64 encoded</pre>

Figura 3.1: Decodificación de un token de acceso generado con la función `genera` de la clase `AccessToken` mediante jwt.io

```
{
  "permisos": 1,
  "idUsuari": 2,
  "sub": "santo@gmail.com",
  "iat": 1743152350,
  "exp": 1743152950
}
```

Hemos utilizado herencia porque la clave privada es la misma para ambos tipos de token, mientras que los métodos para generar cada uno de los dos tipos de JWT

cambian. En StackOverflow existe un debate para ver si hay que tener una clave privada distinta para cada tipo de token, por si el lector está interesado [2]. Se ha optado por compartir claves para ambos tipos.

En la clase JwtUtils tenemos una función denominada `getClaims()` que es la que utilizaremos en el Service de nuestra aplicación para poder autenticar usuarios. Las tres clases pueden ser consultadas en el anexo 5.3 o en el GitHub del proyecto ([link](#)). Se recomienda encarecidamente al lector optar por esta última opción².

²Al poner las tres clases en anexo se omitieron las funciones main donde se testeaban las funciones de creación de tokens con control de excepciones, comentarios e imports por falta de espacio en el DIN A4.

Evaluación y Conclusiones Finales

ANEXO

5.1. Flujo de trabajo habitual en git

```
# trabajamos con el proyecto y se introduce
# en el staging area
git add -A

# creamos rama para aglutinar los cambios
git branch backEnd

# cambiamos a la rama que acabamos de crear
git checkout backEnd

# guardamos los cambios como nodos dentro de
# la rama con la que desarrollamos.
git commit -m "commit 1"
git commit -m "commit 2"
# [...]
git commit -m "commit n"

#cambiamos a rama main local y luego integramos cambios
git checkout main
git merge backEnd

#Subimos los cambios al repo remoto
git push origin main
```

5.2. Diferencias de seguridad: JWT vs SESSION en cookies seguras

Característica	JWT en cookies seguras	Session ID en cookies seguras
Seguridad contra XSS	Más seguro si la cookie tiene <code>HttpOnly</code> y <code>Secure</code> , ya que no es accesible desde JavaScript.	Más seguro si la cookie tiene <code>HttpOnly</code> y <code>Secure</code> , ya que no es accesible desde JavaScript.
Seguridad contra CSRF	Puede ser vulnerable si la cookie no tiene <code>SameSite=Strict</code> .	Menos vulnerable si la cookie tiene <code>SameSite=Strict</code> .
Estado en el servidor	Stateless (no hay estado en el servidor, el JWT contiene toda la información).	Stateful (el servidor mantiene una sesión activa asociada con el Session ID).
Escalabilidad	Mejor escalabilidad porque no requiere almacenamiento de sesiones en el servidor.	Menos escalable, ya que el servidor debe manejar las sesiones activas.
Expiración y revocación	Difícil de revocar antes de que expire, a menos que se implemente una lista negra en el servidor.	Fácil de invalidar eliminando la sesión en el servidor.
Uso con JavaScript	No accesible desde JavaScript si la cookie tiene <code>HttpOnly</code> .	No accesible desde JavaScript si la cookie tiene <code>HttpOnly</code> .

Cuadro 5.1: Comparación de seguridad entre JWT y Session ID almacenados en cookies seguras con `HttpOnly=True`.

5.3. Clases para crear y verificar JWTs

5.3.1. Clase JWT

```
//NO INSTANCIEM AQUESTA CLASSE MAI. LA FEM ABSTRACTA
@Component
public abstract class JwtUtil {

    //es la clau privada de 256 bits com a minim per encriptar el token (tant el d'accés com el de refresh)
    //veure debat http://bit.ly/3RmBGIK
    protected static String clauSecreta;

    public JwtUtil() {
        this.clauSecreta = "a8f7d9g0b6c3e5h2i4j7k1l0m9n8p6q3r5s2t1u4v0w9x8y7z";
    }

    //METODE QUE PARSEJA EL TOKEN JWT COMPLET. VERIFICA LA FIRMA I EXTRAU LES CLAIMS (parells clau valor en el payload).
    protected Claims getClaims(String token) {
        return Jwts.parser()
            .setSigningKey(clauSecreta.getBytes())
            .build()
            .parseClaimsJws(token)
            .getBody();
    }
}
```

5.3.2. Clase Refresh Token

```
@Component
public class RefreshToken extends JwtUtil {

    private static int tExpDies;

    public RefreshToken() {this.tExpDies = 7;}

    // FINALITAT DEL METODE: Refrescar el token d'accés que genera generaAccesToken().
    public String generaRefreshToken(String correu, int idUsuari) {
        Map<String, Object> dadesExtraApayload = new HashMap<>();
        dadesExtraApayload.put("idUsuari", idUsuari);
        //posar mes dades al payload si es necessari

        return Jwts.builder()
            .setClaims(dadesExtraApayload) //dades customitzades
            .setId(String.valueOf(UUID.randomUUID().toString())) //id unic per a token. Per traSSSabilitat
            .setSubject(correu) //guardo nom subjecte (dins "sub")
            .setIssuedAt(new Date()) //data creacio
            .setExpiration(new Date(System.currentTimeMillis() + tExpDies*86400*1000)) //expiracio
            .signWith(SignatureAlgorithm.HS256, clauSecreta.getBytes())
            .compact();
    }
}
```

5.3.3. Clase Access Token

```
@Component
public class AccessToken extends JwtUtil {

    private static int tExpM; //minuts per a exprirar el token

    public AccessToken() {this.tExpM = 10;}

    //FINALITAT: Generar un JWT d'accés.
    public String genera(String correu, int idUsuari, byte permisos) {
        Map<String, Object> dadesExtraApayload = new HashMap<>();
        dadesExtraApayload.put("permisos", permisos);
        dadesExtraApayload.put("idUsuari", idUsuari);
    }
}
```

```

        return Jwts.builder()
            .setClaims(dadesExtraApayload) //dades customitzades
            .setSubject(correu)             //guardo nom subjecte (clau "sub")
            .setIssuedAt(new Date())         //data creacio (clau "iat" payload)
            .setExpiration(new Date(System.currentTimeMillis() + (tExpM*60*1000)))
            .signWith(SignatureAlgorithm.HS256, clauSecreta.getBytes())
            .compact();
    }
}

```

5.4. CODIGO 5

Bibliografía

- [1] jwt.io. Json web tokens - jwt.io. <https://jwt.io/>. Accedido el 27 marzo de 2025.
- [2] Stack Overflow Community. Should refresh tokens in jwt authentication schemes be signed with a different secret than the access token? <https://stackoverflow.com/questions/63092165/should-refresh-tokens-in-jwt-authentication-schemes-be-signed-with-a-different-secret>. Accedido el 28 marzo de 2025.