**Plataforma para el análisis de datos de Instagram**

Trabajo de Fin de Grado

Grado en Ingeniería Informática

Texto, Logotipo

Descripción generada automáticamente

Septiembre de 2022

**Autor**

Alberto Macías Gutiérrez

**Tutores**

Lucía Martín Gómez

Gabriel Villarrubia González

Juan Francisco de Paz Santana

**FIRMA DE TUTORES**

**RESUMEN**

**ABSTRACT**

# Índice

# Índice de ilustraciones

[Figura 1. Estudio de las redes sociales 2022 2](#_Toc111077184)

[Figura 2. Historia de las redes sociales 7](#_Toc111077185)

[Figura 3. Protocolo HTTP 10](#_Toc111077186)

[Figura 4. Modelo Cliente-Servidor 11](#_Toc111077187)

[Figura 5. Historia del Desarrollo Web 14](#_Toc111077188)

[Figura 6. Funcionamiento Framework Django 22](#_Toc111077189)

[Figura 7. Puntos de casos de uso 29](#_Toc111077190)

[Figura 8. Interfaz EZEstimate 29](#_Toc111077191)

# Introducción

En las últimas décadas, el nacimiento de diferentes redes sociales ha supuesto una revolución en la comunicación entre las personas y la difusión de información. Cada vez son más las personas que utilizan las redes sociales, gracias en parte a la popularización del smartphone con conexión a internet (Portaltic, 2022). Concretamente en España un 93% de la población comprendida entre los 12 y 70 años usan los servicios de internet, de los cuales un 85% son usuarios de redes sociales. Además, los usuarios suelen usar durante el mes un promedio de 4,5 redes sociales (ELOGIA, 2022).

Los usuarios que forman parte de las redes sociales y las usan, postean gran cantidad de datos, pero estos no son fáciles de recuperar y analizar manualmente, por lo que los programas para el análisis de la información de las redes sociales se han convertido en una herramienta esencial hoy en día para saber las tendencias o preferencias que siguen los distintos grupos de usuarios.

Un problema se encuentra en la recopilación de datos, aunque algunas redes sociales proporcionen APIs, muchas otras no ofrecen este tipo de facilidades y medios o en otros casos su coste es demasiado elevado (Lucía Martín Gómez et al., 2021).

Interfaz de usuario gráfica

Descripción generada automáticamente

Figura . Estudio de las redes sociales 2022

Como podemos ver en la Figura 1 (ELOGIA, 2022), el contenido más consumido por los usuarios es en formato de foto, llevándose un 67,6% de las interacciones, seguido bastante por detrás de vídeo. Este hecho justifica a Instagram como la red social que mayor interacción tiene, gracias a que se trata de una red social de fotos, así como una de las más usadas diariamente, en concreto la segunda solamente detrás de WhatsApp, además de ser una de las favoritas, para el público comprendido entre los 16 y 24 años (Galeano, 2022).

Este documento tratará de analizar los datos de la red social Instagram, para ello, se ha planteado una propuesta, que recopila datos, mediante web scraping, de los usuarios, hashtags y publicaciones a modo de buscador. Además, también se recopilarán datos de las historias destacadas de los distintos usuarios. Todo esto se va a resolver mediante el desarrollo de una aplicación web multiusuario, que administre al usuario información, gráficos y tablas, construidos gracias a los datos recogidos anteriormente, de forma interactiva, intuitiva y sencilla gracias al uso de interfaces gráficas visuales, usables y educativas para el usuario (Eniun Diseño Web y Marketing Digital, 2019).

Este documento está organizado en los siguientes apartados o secciones:

* **Introducción:** Sección actual, que busca identificar y justificar el problema a tratar.
* **Objetivos:** Busca redactar los diferentes objetivos o metas que se pretenden lograr con este proyecto.
* **Antecedentes:** Pequeño resumen de las tecnologías que se han ido usando para las redes sociales.
* **Conceptos teóricos:** Se detallan las distintas técnicas y herramientas que se usan para el desarrollo del proyecto.
* **Desarrollo de la propuesta:** Se detallan los aspectos más importantes del desarrollo del proyecto.
* **Resultados:** Se muestra la información obtenida del trabajo.
* **Conclusiones y líneas de trabajo futuras:** Se interpretan los resultados que se han extraído de la realización del trabajo y el planteamiento a abordar si se desea continuar con el proyecto.

Este proyecto, también está formado por otro documento compuesto por los siguientes anexos:

* **Plan del proyecto Software:** Se detalla la planificación temporal del proyecto.
* **Especificación de requisitos y análisis:** Se detallan la especificación de requisitos del proyecto y el análisis de este.
* **Especificación de diseño:** Se detalla la documentación del diseño del proyecto.
* **Documentación técnica de programación:** Se detalla la documentación para la comprensión por parte del programador.
* **Manual del usuario:** Se detalla la documentación para dar asistencia a los usuarios que pretendan hacer uso del proyecto.

# Objetivos

El Trabajo de Fin de Grado tiene como objetivo principal ofrecer a los usuarios una herramienta que permita recopilar, tratar y plasmar información de perfiles, hashtags o publicaciones de Instagram, con el fin de representar y mostrar estos datos para poder guiar y dar a entender a los usuarios lo más importante y relevante de estos.

De este objetivo principal derivan los siguientes objetivos específicos:

* **Obtener datos de Instagram a través de web scraping:** Los datos a utilizar por la herramienta se van a obtener mediante técnicas de web scraping.
* **Crear tablas y gráficos con los datos obtenidos:** Los datos a obtener, se van a presentar en forma de tablas y gráficos con el fin de poder transformar esos datos en información más entendible para los usuarios.
* **Crear interfaces para usar la aplicación:** Para que el usuario pueda hacer uso de toda la funcionalidad de la herramienta, se va a crear una aplicación web con varías interfaces escritas en el lenguaje de marcado HTML, con el fin de que se pueda usar la aplicación desde cualquier buscador.
* **Almacenar información:** Para almacenar información de las cuentas de usuario, así como la información sobre sobre las búsquedas, se va a hacer dentro de una base de datos relacional MySQL.
* **Crear un modelo de usuario:** Para que cualquier cliente se pueda registrar como usuario, modificar la cuenta de Instagram de scraping para realizar las búsquedas y poder borrar la cuenta de usuario.
* **Guardar datos importantes encriptados:** Con el fin de crear una aplicación segura, las contraseñas de los usuarios se van a guardar encriptadas dentro de la base de datos.

# Antecedentes

Para poder meternos en contexto debemos recordar el nacimiento de Internet, en 1969 durante la guerra fría, Estados Unidos crea ARPANET una red exclusivamente militar, con el objetivo de mantener las comunicaciones en caso de guerra ante la situación de incertidumbre y temor del momento. Un año más tarde, Ray Tomlinson establece las bases para la creación del correo electrónico.

En 1983, el departamento de defensa de Estados Unidos decidió usar el protocolo TCP/IP en su red ARPANET creando la red Arpa Internet, que con el paso de los años acabó nombrándose Internet. En 1989, Tim Berners Lee describió por primera vez el protocolo de transferencias de hipertextos que daría lugar a la primera web utilizando los recursos HTML, HTTP y un programa llamado Web Browser. La World Wide Web (www), creció exponencialmente pasando de tener 100 World Wide Web Sites en 1993 a más de 200000 en 1997. Actualmente internet sigue creciendo (Bahillo, 2022) .

Una vez explicada la historia de internet veremos la historia de las redes sociales, en el año 1997 aparece SixDegrees, la que fue considerada como la primera red social del mundo, una red que permitía localizar a otros miembros de la red uniendo conocidos con conocidos de conocidos y crear listas de amigos.

En 2002 se crea Friendster, una red social para los amantes de los videojuegos y en 2003 se crea LinkedIn, una red social mucho más profesional orientada a empresas, que hoy en día cuenta con más de 600 millones de registrados.

En 2004 aparece la red social por excelencia, Facebook, creada por el joven estudiante Mark Zuckerberg. En un principio era un portal llamando Facemash, con la finalidad de que los estudiantes de Harvard pudieran compartir opiniones acerca de quienes eran las personas más atractivas de la Universidad. Hoy en día es la mayor red social con más de 2500 millones de usuarios activos al mes.

En 2005 aparece YouTube, una red social creada por Chad Hurley, Steve Chen y Jawn Karim en California. Esta red social se trata de un sitio web dedicado a compartir vídeos. Hoy en día dispone cerca de 2000 millones de usuarios activos al mes.

En 2006 aparece Twitter, una red social creada por Jack Dorsey, Noah Glass, Biz Stone y Evan Williams. Esta red social permite la comunicación con otros usuarios mediante mensajes con un tamaño máximo de 140 caracteres. Hoy en día dispone cerca de 340 millones de usuarios activos al mes.

En 2009 aparece WhatsApp, la app de mensajería instantánea más famosa, creada por el ucraniano Jan Koum y posteriormente comprada por Mark Zuckerberg. Esta se vincula a la agenda de contactos del dispositivo, permitiendo al usuario ver qué estaba haciendo cada persona en cada momento, con el fin de saber si poder iniciar o no una conversación con dicha persona. Hoy en día supera los 2000 millones de usuarios activos al mes.

En 2010 aparece Instagram, la red social de fotografía por excelencia, creada por Kevin Systrom y Mike Krieger. Además, fue pionera junto con Twitter en el uso de hashtags, facilitando a los usuarios descubrir fotografías que otros usuarios compartían sobre un mismo tema. Hoy en día supera los 1000 millones de usuarios activos al mes.

También ese mismo año aparece Pinterest, una red social que colecciona imágenes, que permite a los usuarios almacenarlas en tableros y dotarlas de pines. Hoy en día supera los 300 millones de usuarios activos al mes.

En 2011 aparece Twitch, la plataforma de streaming más grande del mundo, creada por Justin Kan, Emmett Shear, Michael Seibel y Kyle Vogt y posteriormente comprada por Amazon.

En 2016 aparece Tik Tok también conocido como Douyin en China, es una red social que permite compartir videos cortos, desde 15 segundos de duración hasta un máximo de un minuto (Hera, 2022).

Captura de pantalla de un celular

Descripción generada automáticamente

Figura . Historia de las redes sociales

Una vez conocemos la historia y para que esta concebido el uso de ciertas redes sociales, vamos a ver herramientas similares a la que se pretende realizar en este proyecto.

Una de las herramientas más famosas, y que también realiza análisis de datos sobre las redes sociales, pero en este caso sobre Twitter, es la API de Twitter (Twitter, 2022). Esta permite recuperar información pública de Twitter, como solicitar información sobre una determinada cuenta o usuario, acceder a Tweets buscando palabras clave o una cierta cuenta, o brindar acceso a conversaciones por mensajes directos de los usuarios que otorgan permisos de forma explícita (Pío, 2022). Esta API permite gran cantidad de cosas, pero sin en cambio tiene un problema, que no permite recuperar información de cuentas o tweets privados, ya que seria totalmente ilegal.

Otra herramienta para realizar análisis de datos sobre las redes sociales, también para Twitter, es SocioViz (SocioViz, 2014). Esta herramienta además de obtener datos de Twitter permite visualizarlos. La información obtenida puede ser de cualquier palabra clave, hashtag, emoji o mención de usuario filtrando por fecha e idioma, tweets en tiempo real o durante un periodo concreto, además esta información es representada mediante gráficos y tablas para facilitar su comprensión. El funcionamiento de esta herramienta se basa consultar y recopilar resultados de la API de Twitter (Zonín, 2019), por lo que solo permite recuperar información pública.

Una vez hemos visto algunas herramientas similares, vamos a centrar nuestra atención en Instagram. Esta red social ofrece una API de visualización básica para obtener información básica de una cuenta para la que se dispone de credenciales, sin embargo, la API tiene varias limitaciones, ya que no ofrece información sobre las historias o comentarios del usuario, las publicaciones que contienen medios tampoco están disponibles y la información del usuario solo se puede obtener si las credenciales del usuario están disponibles.

Instagram también ofrece una segunda API para cuentas profesionales, con la que se puede obtener y publicar contenido multimedia, administrar y responder comentarios, identificar contenido multimedia con menciones, buscar contenido multimedia usando hashtags y obtener resultados y metadatos básicos sobre otras empresas y creadores de esta red social. La funcionalidad de esta segunda API es mucho más completa, pero solo es posible de aplicar al usuario profesional desde el que se accede, y el usuario también debe de tener una cuenta profesional. Esto limita mucho la recopilación de datos (Lucía Martín Gómez et al., 2021).

Por este motivo, en este proyecto se van a recopilar los datos mediante web scraping, ya que además de recopilar información sobre usuarios públicos, se va a poder recopilar información sobre usuarios privados, debido a que se van a utilizar cuentas de Instagram en concreto para realizar esta búsqueda y si la cuenta con la que se realiza la búsqueda sigue a la que se está inspeccionando, se podrá recopilar información de dicha cuenta.

# Conceptos Teóricos

## Protocolo HTTP

HTTP (Hypertext Transfer Protocol), es el protocolo que permite realizar peticiones de datos y recursos, tales como documentos HTML (MDN contributors, 2020). Es la base de cualquier intercambio de datos de la Web. Diseñado a principios de la década de 1990, HTTP es un protocolo ampliable, que ha ido evolucionando con el tiempo. En la figura 3 se puede observar que HTTP es lo que se conoce como un protocolo de la capa de aplicación, que se transmite sobre el protocolo TCP, o el protocolo encriptado TLS, aunque se podría usar cualquier otro protocolo fiable. Al tratarse de un protocolo ampliable, además de transmitir documentos de hipertexto (HTML) permite transmitir imágenes, vídeos o enviar datos o contenido a los servidores a través de formularios.

Gráfico, Gráfico de rectángulos

Descripción generada automáticamente

Figura . Protocolo HTTP

## Arquitectura Cliente-Servidor

Gran parte de las aplicaciones web usan la arquitectura Cliente Servidor (Schiaffarino, 2019), que es una de las principales usadas en muchos servicios y protocolos de internet. Esta arquitectura tiene dos partes bien diferenciadas, por un lado, la parte del servidor, que suele ser una máquina bastante potente que actúa como depósito de datos y ejecuta la aplicación web, y por otro la parte del cliente o clientes, que son dispositivos o máquinas que demandan servicios al servidor.

Diagrama

Descripción generada automáticamente

Figura . Modelo Cliente-Servidor

En la figura 4 se observa cómo funciona el Modelo Cliente Servidor dentro de internet, por el que los clientes o usuarios envían peticiones y el servidor responde a esas peticiones.

Un ejemplo de la arquitectura Cliente Servidor es la red de internet donde existen ordenadores de diferentes personas conectados alrededor del mundo, las cuales se conectan a través de los servidores de su proveedor de internet por ISP donde son redirigidos a los servidores de las páginas que desean visualizar.

Ventajas:

* Facilita la integración entre diferentes sistemas y comparte información permitiendo que las máquinas ya existentes puedan ser utilizadas mediante una interfaz más amigable para el usuario
* Al tratarse de un sistema que favorece el uso de las interfaces gráficas interactivas, tiene una mayor interacción con el usuario.
* La estructura modular facilita la integración de nuevas tecnologías y el crecimiento de la infraestructura computacional favoreciendo así la estabilidad de las soluciones.
* Este modelo permite proporcionar a las diferentes áreas de una empresa generar un orden de trabajo en donde cada sector puede trabajar en su área, pero accediendo al mismo servidor e información que los demás sin generar conflictos.

Desventajas:

* Requiere habilidad para que un servidor sea reparado.
* Problema de seguridad, el hecho que se comparte canales de información entre servidores y clientes requieren que estas pasen por procesos de validación, es decir, protocolos de seguridad que pueden tener algún tipo de puerta abierta permitiendo que se generen daños físicos, amenazas o ataques de malware.
* Este modelo representa una importante limitación en cuanto a los costos económicos debido a que los servidores son computadoras de alto nivel con un hardware y software específicos, por lo que reemplazar componentes que estén averiados tiene un costo elevado.

## Desarrollo Web

El desarrollo web es un término (Mercedes, 2017) que define la creación o mantenimiento de sitios web para Internet o una intranet. Para conseguirlo se hace uso de tecnologías de software del lado del servidor y del cliente que involucran una combinación de procesos de base de datos con el uso de un navegador web con el fin de realizar unas determinadas tareas o mostrar cierta información.

Un desarrollador web (Seguro, 2021) se encarga de las siguientes tareas:

* Creación de sitios web y aplicaciones.
* Mantenimiento de estos sistemas, para que sean eficientes y ágiles.
* Asegurar que el sistema desarrollado cumpla con las demandas de los usuarios.
* Proporcionar seguridad al entorno de la aplicación web.

### Desarrollo Frontend

El desarrollo frontend es el desarrollo que se encarga de los aspectos funcionales y de la usabilidad. En este ámbito se incluye la UX (experiencia de usuario) y la UI (relativo a interfaces), por lo que a este desarrollo se lo conoce como desarrollo por parte del cliente. Utiliza frameworks de JavaScript, además del lenguaje de marcado HTML.

### Desarrollo Backend

El desarrollo backend es el desarrollo que se enfoca en las estructuras en las que se apoyan los sitios web y aplicaciones, es decir, de aquello que los usuarios no ven. En este se establecen las conexiones necesarias con la base de datos y el servidor, por lo que se lo conoce como desarrollo del lado del servidor. Utiliza lenguajes de programación como JavaScript, Python, SQL, PHP o Java, entre otros.

### Evolución

El desarrollo web (Gima, 2019) ha evolucionado exponencialmente desde su inicio. En la actualidad es uno de los servicios más solicitados, que nos brinda una gran cantidad de funciones distintas.

Escala de tiempo

Descripción generada automáticamente

Figura . Historia del Desarrollo Web

En la figura 5 podemos observar, como hemos pasado de páginas HTML estáticas en los años 1990 a la creación de páginas web dinámicas, gracias al uso de JavaScript para dar dinamismo a las páginas web, las hojas de estilo CSS para dar formato, apariencia y hacer un diseño responsivo o adaptable con el fin de que estas páginas web se puedan representar en pantallas con tamaños y resoluciones muy diferentes.

## Red social

Las redes sociales (Llonch, 2021) son plataformas digitales que conectan entre sí a personas con intereses, actividades o relaciones en común. Estas plataformas permiten el contacto entre los individuos que las componen y funcionan como un medio para intercambiar información. La información publicada por un usuario puede ser pública o privada.

Tipos de redes sociales:

* Redes sociales personales: Las redes sociales personales están pensadas para conectar individuos entre sí basándose en conexiones personales. Un ejemplo es Instagram.
* Redes sociales de entretenimiento: La función principal de estas redes sociales es compartir contenidos. Un ejemplo es YouTube.
* Redes sociales profesionales: Estas están pensadas para hacer contactos, publicar información sobre el currículum y logros, conseguir empleo y otros fines relacionados con el mundo profesional. Un ejemplo es LinkedIn.
* Redes sociales de nicho: Dirigidas a personas con un interés específico común, ya sea profesional o personal. Un ejemplo es DevianArt.

Esta clasificación se basa en la finalidad original de cada red, en realidad muchas redes sociales ofrecen flexibilidad en cuanto a sus usos. Instagram puede considerarse una red personal, de entretenimiento o incluso personal.

## Web Scraping

Se llama web scraping (Zhao, 2017) a la técnica de extraer datos de una página web y guardarlos en un sistema de archivos o base de datos para su posterior recuperación o análisis. Normalmente, los datos de la web se extraen utilizando el protocolo HTTP o a través de un navegador web. Esto se lleva a cabo de forma manual por un usuario, automáticamente por un bot o rastreador web. Debido a la gran cantidad de datos heterogéneos que se generan constantemente en internet, el web scraping está ampliamente reconocido como una técnica eficaz y poderosa para recopilar big data.

Para adaptarse a gran variedad de escenarios, las técnicas actuales de web scraping se han personalizado a partir de procedimientos más pequeños ad hoc, asistidos por humanos, hasta la utilización de sistemas totalmente automatizados que son capaces de convertir sitios web enteros en un conjunto de datos. Las herramientas de web scraping más modernas son capaces de analizar lenguajes de marcado archivos JSON e integrar el análisis visual y el procesamiento del lenguaje natural para simular cómo los usuarios navegan por el contenido de la web.

El proceso de extracción de datos de Internet puede dividirse en dos pasos secuenciales: la adquisición de recursos web y, a continuación, extraer la información deseada de los datos adquiridos. En concreto, un programa de comienza componiendo una petición HTTP para adquirir recursos de un sitio web, esta solicitud puede tener el formato de una URL que contenga una consulta GET o un trozo de mensaje HTTP que contenga una consulta POST. Una vez que la solicitud es recibida y procesada con éxito por el sitio web de destino, el recurso solicitado se recuperará del sitio web y se envía de nuevo a el programa de web scraping. El recurso puede estar en múltiples formatos, como páginas web construidas a partir de HTML, fuentes de datos en formato XML o JSON, datos multimedia como imágenes, audio o archivos de vídeo.

## GraphQL

GraphQL (Acosta, 2021) es un nuevo estándar de API que está diseñado para ser declarativo, flexible y eficiente. Es un lenguaje de consultas para comunicar clientes y servidores. Es de código abierto y fue desarrollado por Facebook, actualmente lo mantiene una gran comunidad de empresas e individuos de todo el mundo.

Fue inventado por Facebook ya que la empresa necesitaba resolver muchos problemas técnicos con su aplicación móvil nativa, de forma que GraphQL permite mejorar la comunicación entre las diferentes partes de una aplicación de software, por lo que hace que esa aplicación sea más fácil de entender, desarrollar, mantener y escalar. El objetivo principal de este es evitar las múltiples consultas al servidor

GraphQL proporciona los bloques de construcción para que los desarrolladores definan la estructura de sus datos, cómo acceder a ellos y qué datos devolver. Una API basada en GraphQL puede ser consumida por muchos clientes, no solo por los clientes de GraphQL. No está vinculado a una base de datos específica, asique no importa con qué motor de almacenamiento tenga que lidiar.

Es agnóstico de plataforma ya que se puede implementar en más de 20 lenguajes de programación, esto se debe a que GraphQL es una especificación, es decir, especifica cómo debe funcionar, permitiendo a cualquiera implementar GraphQL en cualquier lenguaje de programación. Es un lenguaje tipado, es decir, utiliza tipos para definir recursos, añade tipos a los campos de cada recurso y también utiliza tipos para comprobar estáticamente los errores. Es utilizado tanto por empresas grandes como pequeñas, desde Twitter y Facebook hasta Yelp y Twitch. En 2016, Github cambió su API de REST a GraphQL y en 2017, AWS lanzó una plataforma basada en GraphQL para crear aplicaciones.

Diferencias entre API REST y GraphQL (Ortega, 2019):

API REST

* Es una convención, una manera de comunicarse entre el servidor y el cliente, cada uno tiene sus reglas.
* El servidor expone recursos, los clientes se tienen que adecuar a como están expuestos.
* Hace overfetching, envía más información de la que se necesita.
* Múltiples solicitudes por vista, muy costoso en rendimiento.
* Documentación ajena al desarrollo, no hay estándar por lo que depende mucho del desarrollador para mantenerla.

GraphQL

* Lenguaje tipado y validable, se puede dar forma a lo que recibe y lo que devuelve.
* El cliente define que recibe, los clientes hacen una consulta, de la estructura que se define como respuesta.
* Envía lo necesario, se tiene control total de las respuestas que se esperan del servidor.
* Solo una solicitud por vista, prácticamente en una solicitud se puede mandar lo que se necesite.

## Business Intelligence

También conocido como inteligencia de negocio o inteligencia empresarial, business intelligence (media, 2021) hace referencia a uso de estrategias y herramientas que sirven para transformar datos en conocimiento, con el objetivo de mejorar el proceso de toma de decisiones en una empresa. Este concepto ha ganado popularidad por la relevancia empresarial para cualquier empresa o negocio, y la complejidad en recopilar datos, procesarlos, analizarlos y presentarlos de forma que cualquiera los pueda entender.

## Tecnologías y Herramientas para el desarrollo web

### Python

Fue creado por Guido van Rossum (Visus, 2020) a principios de los 90, su nombre se debe a la afición de este al grupo Monty Python. Python es un lenguaje de programación interpretado cuya filosofía es que sea legible por cualquier persona con conocimientos básicos de programación.

Características:

* Es gratis, ya que es un lenguaje de código abierto y no hay que pagar ninguna licencia para utilizarlo.
* Respaldado por una comunidad enorme, por el carácter gratuito hace que continuamente se estén desarrollando nuevas librerías y aplicaciones. Esto es un factor multiplicativo para los programadores, debido a que casi cualquier duda estará resuelta en los foros.
* Es un lenguaje multiparadigma, es decir, combina propiedades de diferentes paradigmas de programación, lo que permite que sea muy flexible y fácil de aprender.
* Sus aplicaciones no se limitan a un área en concreto, ya que al tratarse de un lenguaje multiparadigma permite utilizarlo en diseño de aplicaciones web o inteligencia artificial, entre muchos otros.
* Es acto para todas las plataformas, es decir, se puede ejecutar en cualquier sistema operativo usando el intérprete correspondiente.

El principal inconveniente del lenguaje de programación Python es que se trata de un lenguaje de programación interpretado, es decir, que no se compila, sino que se interpreta en tiempo de ejecución, dando como resultado a que sea más lento que otros lenguajes de programación como Java, C o C++. Pero este inconveniente no es un gran problema actualmente ya que en la actualidad las máquinas son muy potentes.

### Django

Fue desarrollado inicialmente entre 2003 y 2005 por un equipo que era responsable de crear y mantener sitos webs de periódicos con el fin de reutilizar códigos y patrones de diseño comunes. Django es un framework web gratuito (MDN contributors, 2021) de alto nivel que permite el desarrollo rápido de sitios web seguros y mantenibles. Django se encarga de una gran parte de las complicaciones del desarrollo web, por lo que permite al programador centrarse únicamente en escribir las funcionalidades de la aplicación. Este es de código abierto, además de poseer una comunidad próspera y activa, mucha documentación y soporte gratuito. Django se encuentra escrito en Python.

El software escrito en Django es:

* Completo: Provee casi todo lo que los desarrolladores quisieran que traiga de serie un framework web.
* Versátil: Puede ser usado para construir casi cualquier tipo de sitio web, desde sistemas manejadores de contenidos y wikis, hasta redes sociales y sitios de noticias. Puede funcionar con cualquier framework en el lado del cliente, devolver contenido en una gran variedad de formatos (HTML, RSS feeds, JSON, XML, …) y también usar distintos motores de bases de datos y plantillas.
* Seguro: Proporciona una manera segura de administrar cuentas de usuario y contraseñas, evitando errores comunes. Además, permite protección contra algunas vulnerabilidades de forma predeterminada, incluida la inyección SQL, scripts entre sitios, falsificación de solicitudes y clickjacking.
* Escalable: Django usa un componente en la arquitectura shared-nothing, en la que cada parte de esta arquitectura es independiente de las otras, y por lo tanto puede ser reemplazada o cambiada si es necesario. Habiendo una clara separación entre las diferentes partes significa que puede escalar para aumentar el tráfico al agregar hardware en cualquier nivel, ya sean servidores cache, servidores de bases de datos o servidores de aplicación. Un ejemplo de sitios que han escalado a Django son Instagram y Disqus.
* Mantenible: Utiliza el principio No te repitas “Don’t Repeat Yourself” (DRY) para que no exista una duplicación innecesaria, reduciendo la cantidad de código.
* Portable: Está escrito en Python, el cual se ejecuta en muchas plataformas, este puede ejecutar sus aplicaciones en muchas distribuciones de Linux, Windows y Mac OS X. Además de contar con el respaldo de muchos proveedores de alojamiento web, y que a menudo proporcionan una infraestructura específica y documentación para el alojamiento de sitios Django.

Django funciona con patrón de diseño organizando la arquitectura en Modelo Vista Plantilla “Model View Template (MVT)”, este tiene una estructura muy similar a la arquitectura Model View Controller, con la diferencia que la vista es el controlador intermediario entre el modelo, que se comunica con la base de datos, y las plantillas en formato HTML que reproducen la información y los formularios para pedir o cambiar información en la base de datos.

En un sitio web tradicional, una aplicación web espera peticiones HTTP del cliente. Cuando se recibe una petición la aplicación elabora la respuesta que se necesita basándose en la URL y a veces, en una información incluida en los datos de una petición POST o GET. La aplicación también devolverá una respuesta al explorador web, normalmente creando dinámicamente una página HTML con los datos correspondientes a mostrar.

Sin embargo, las aplicaciones web de Django agrupan el código que gestiona cada uno de estos pasos en ficheros separados.

Diagrama

Descripción generada automáticamente

Figura . Funcionamiento Framework Django

URLs: Como se observa en la figura 6, cuando se recibe una petición por parte de un cliente, esta se procesa en el archivo urls.py, de forma que funciona como un mapeador URL para redirigir las peticiones HTTP a la vista apropiada.

View: El mapeador URL redirige la petición a una vista en concreto, la cual accederá a los datos que se necesitan para la petición por medio del modelo y formateará la respuesta al cliente por medio de las plantillas.

Models: Los modelos definen la estructura de los datos de la aplicación y proporcionan mecanismos para gestionar esta (añadir, modificar y borrar) y consultar registros en la base de datos.

Templates: Las plantillas son ficheros de texto que definen la estructura o diagrama de otro fichero, con marcadores de posición para representar contenido real. La vista creará dinámicamente una página usando una plantilla, rellenando esta con los datos del modelo. Normalmente estos ficheros suelen estar codificados en lenguaje HTML pero no tiene porque ser así, pueden ser ficheros de cualquier tipo.

### HTML (HyperText Markup Language)

Fue creado por el científico de computación británico Timohty John Berners-Lee en 1991. HTML (Reyes, 2016) es un lenguaje de marcado de hipertexto (HyperText Markup Language), el cual se escribe en su totalidad con elementos, que a su vez están construidos por etiquetas, contenido y atributos. HTML es el lenguaje que interpreta el navegador web para mostrar los sitios o aplicaciones web, tal y como están descritos.

Los elementos dentro de este lenguaje ayudan a estructurar y dar significado a las distintas partes de un documento HTML. Las etiquetas sirven para delimitar el inicio y el fin de un elemento. El contenido de un elemento pueden ser caracteres, comentarios u otro elemento delimitado dentro de las etiquetas de inicio y cierre.

### CSS (Cascading Style Sheets)

CSS (Santos, 2022) es un lenguaje que se ocupa del diseño y la presentación de las páginas web, funciona junto con el lenguaje HTML que se encarga del contenido básico de las páginas. Sus siglas significan “hojas de estilo en cascada”, ya que puedes tener varias hojas y una de ellas con propiedades heredadas o en cascada de otras. Estas hojas de estilo permiten crear reglas para indicar a la página web como mostrar la información y el estilo de cada elemento de la página.

### JavaScript

JavaScript (Ramos, 2021) es el lenguaje de programación encargado de dar más interactividad y dinamismo a las páginas web. Este no necesita ser compilado, ya que el navegador lee directamente el código sin necesidad de terceros. Se trata de uno de los tres lenguajes nativos de las páginas web junto a HTML (contenido y estructura) y a CSS (diseño del contenido y estructura).

Este lenguaje de programación se ejecuta en la máquina de los clientes, que permite crear efectos y animaciones o responder a eventos causados por el usuario como pulsar botones o modificar el DOM (document object model). El código JavaScript se ejecuta en todos los navegadores, tanto de escritorio como móviles. JavaScript es capaz de detectar errores en formularios, crear sliders que se adapten a cualquier pantalla, hacer cálculos matemáticos de forma eficiente, modificar elemento de una página web de forma sencilla y muchas más cosas.

También, existe una tecnología llamada AJAX que permite intercambiar información con el servidor sin la necesidad de recargar la página. Esta tecnología ha supuesto uno de los principales avances en el desarrollo web. Como ejemplo, esta tecnología es la encargada de conseguir más mensajes, tweets, emails… con el único hecho de pulsar un botón, sin tener que recargar la página.

### Bootstrap

Bootstrap (Invitado, 2020) es un framework CSS desarrollado por Twitter en 2010, para estandarizar las herramientas de la compañía. Este framework combina CSS y JavaScript para estilizar los elementos de las páginas web HTML, además proporciona interactividad en las páginas incluyendo una serie de componentes que facilitan la comunicación con el usuario, como menús de navegación, controles de página y barras de progreso.

El principal objetivo de Bootstrap es permitir la construcción de sitios web responsive (los elementos de la página escalan en función de la resolución de la pantalla) para dispositivos móviles, ordenadores y tablets, de una manera simple y organizada.

La estructura de este framework se compone en dos directorios:

* css: Contiene los archivos necesarios para la estilización de los elementos y una alternativa al tema original.
* js: Contiene la parte posterior del archivo bootstrap.js, responsable de la ejecución de aplicaciones de estilo que requieren manipulación interactiva.

Para asignar una característica a un elemento, hay que informar la clase correspondiente en la propiedad “class” del elemento que será estilizado.

### Base de datos

Una base de datos (Oracle, 2014) es una recopilación organizada de información o datos estructurados, que normalmente se almacenan de forma electrónica en un sistema informático. Normalmente, una base de datos se controla por un sistema de gestión de bases de datos (DBMS). El conjunto de los datos, el DBMS y las aplicaciones asociadas a ellos, reciben el nombre de sistema de bases de datos.

Los datos de los tipos más comunes de bases de datos en funcionamiento actualmente se suelen utilizar como estructuras de filas y columnas en una serie de tablas para aumentar la eficacia del procesamiento y consulta de datos. La mayoría de las bases de datos utilizan un lenguaje de consulta estructurada denominado SQL para escribir y consultar datos.

#### MySQL

Es el sistema de gestión (Robledado, 2019) de bases de datos relacional más extendido en la actualidad al estar basada en código abierto. Cuenta con una doble licencia, por una parte, es de código abierto y por otra cuenta con una versión comercial gestionada por la compañía Oracle.

Las ventajas más evidentes son:

* Trabaja con bases de datos relacionales, es decir, utiliza tablas múltiples que se interconectan entre sí para almacenar la información y organizarla correctamente.
* Al ser basada en código abierto es fácilmente accesible, además de contar con una gran comunidad que ofrece soporte a otros usuarios.

Algunas de sus características son:

* Arquitectura Cliente-Servidor: Basa su funcionamiento en un modelo cliente-servidor. Cada cliente puede hacer consultas a través del sistema de registro para obtener datos, modificarlos, guardar estos cambios o establecer nuevas tablas de registros.
* Compatibilidad con SQL: SQL es un lenguaje generalizado dentro de la industria y MySQL ofrece plena compatibilidad con este.
* Vistas: Ofrece compatibilidad para poder configurar vistas personalizadas.
* Procedimientos almacenados: Posee la característica de no procesar las tablas directamente, sino que a través de procedimientos almacenados es posible incrementar la eficacia de la implementación.
* Desencadenantes: Permite automatizar ciertas tareas en la base de datos, en el momento que se produce un evento otro es lanzado para actualizar registros u optimizar su funcionalidad.
* Transacciones: Una transacción representa la actuación de diversas operaciones en la base de datos. El sistema de base de datos avala que todas las operaciones de una transacción se establezcan correctamente o ninguna de ellas, preservando así la integridad de la base de datos.

### Visual Studio Code

Es un editor (Flores, 2022) de código fuente desarrollado por Microsoft para Windows, Linux y macOS. Se trata de un software libre y multiplataforma, tiene buena integración con Git, cuenta con soporte para la depuración de código y dispone de una gran cantidad de extensiones que dan la posibilidad de escribir y ejecutar código en cualquier lenguaje de programación.

Algunas de sus características son:

* Multiplataforma: Está disponible para cualquier sistema operativo.
* IntelliSense: Proporciona sugerencias de código, terminaciones inteligentes y resaltado de sintaxis, lo que permite ser más ágil a la hora de escribir código.
* Depuración: Incluye la función de depuración que ayuda a detectar errores en el código.
* Uso del control de versiones: Tiene compatibilidad con Git, por lo que permite revisar diferencias, organizar archivos, realizar commits y hacer push y pull desde el editor.
* Extensiones: Permite personalizar y agregar funcionalidad adicional de forma modular y aislada. Las extensiones no afectan al rendimiento del editor, ya que se ejecutan en procesos diferentes.

## Tecnologías y Herramientas de control de versiones

### Git

Es un software libre (Rubio, 2019) de control de versiones distribuidas diseñado por Linus Torvalds, pensado en la eficiencia, la confiabilidad y compatibilidad del mantenimiento de versiones de aplicaciones cuando estas tienen un gran número de archivos de código fuente. Su propósito es lleva un registro de los cambios en archivos de computadora incluyendo coordinar el trabajo que varias personas realizan sobre archivos compartidos en un repositorio de código.

### GitHub

Es una plataforma gratuita (Camacho, 2021) que ofrece a los desarrolladores la posibilidad de crear repositorios de código y guardarlos en la nube de forma segura, usando el sistema de control de versiones Git. Facilita la organización de proyectos y permite la colaboración de varios desarrolladores en tiempo real, es decir, permite centralizar el contenido del repositorio para poder colaborar con otros miembros. Es propiedad de Microsoft.

Ventajas:

* Permite alojar proyectos de forma gratuita.
* Permite alojar tanto repositorios públicos como privados.
* Facilita compartir proyectos de forma fácil y crear un portafolio.
* Permite colaborar para mejorar proyectos

## Tecnologías y Herramientas de documentación del proyecto

### Visual Paradigm

Es una herramienta CASE (Ingeniería de Software Asistida por Computación), que propicia un conjunto de ayudas para el desarrollo de programas informáticos, desde la planificación, pasando por el análisis y el diseño, hasta la generación del código fuente de los programas y la documentación (EcuRed, 2016). Esta ha sido concebida para soportar el ciclo de vida completo del proceso de desarrollo software a través de la representación de todo tipo de diagramas. Constituye una herramienta privada disponible en varias ediciones, cada una destinada a satisfacer diferentes necesidades.

### Microsoft Project

Es un software (Machuca, 2022) de administración de proyectos y programas de proyectos desarrollado y comercializado por Microsoft para asistir a los administradores de proyectos en el desarrollo de planes, asignar recursos a tareas, dar seguimiento al progreso, administrar presupuesto y analizar cargas de trabajo.

### EZEstimate

Herramienta (García, 2021) que permite estimar el esfuerzo de un proyecto a partir de un valor en cada uno de los factores de complejidad técnica y factores de complejidad del entorno, que en su conjunto forman los puntos de los casos de uso de un proyecto. En la figura 6 se pueden observar cada uno de los diferentes factores y en la figura 7 se puede observar la interfaz del programa.

Tabla

Descripción generada automáticamente con confianza media

Figura . Puntos de casos de uso

Interfaz de usuario gráfica

Descripción generada automáticamente

Figura . Interfaz EZEstimate

### Sphinx

Sphinx (Sphinx, 2022) es un generador de documentación de Python o una herramienta que traduce un conjunto de archivos fuente de texto plano a varios formatos de salida, produciendo automáticamente referencias cruzadas, índices, etc. Es decir, si hay un directorio que contiene bastantes documentos reStructuredText o Markdown, Sphinx puede generar una serie de archivos HTML, un archivo PDF, páginas de manual y mucho más. Sphinx se enfoca, en particular, en la documentación escita a mano.

# Desarrollo de la Propuesta

## Definición de la Propuesta

## Diseño de la Base de Datos

## Proceso de Obtención de Datos de Instagram

## Desarrollo de la Aplicación Web

# Resultados

# Conclusiones y Líneas de Trabajo Futuras

# Bibliografía

Acosta, J. (29 de Junio de 2021). *GraphQL: Qué es y qué ventajas ofrece*. OpenWebinars: https://openwebinars.net/blog/graphql-que-es-y-que-ventajas-ofrece/

Bahillo, L. (16 de Mayo de 2022). *Historia de Internet: cómo nació y cuál fue su evolución*. Marketing4ecommerce: https://marketing4ecommerce.net/historia-de-internet/

Camacho, D. (2021). *Qué es GitHub y cómo usarlo para aprovechar sus beneficios*. Platzi: https://platzi.com/blog/que-es-github-como-funciona/

EcuRed. (14 de Noviembre de 2016). *Visual Paradigm*. https://www.ecured.cu/Visual\_Paradigm

ELOGIA. (18 de Mayo de 2022). *IAB SPAIN PRESENTA EL ‘ESTUDIO DE REDES SOCIALES 2022’.* iabspain: https://iabspain.es/iab-spain-presenta-el-estudio-de-redes-sociales-2022/

Eniun Diseño Web y Marketing Digital. (25 de Agosto de 2019). *Características de una interfaz web: usable, visual, educativa y actualizada*. https://www.eniun.com/caracteristicas-interfaz-usable-visual-educativa-actualizada/

Flores, F. (22 de Julio de 2022). *Qué es Visual Studio Code y qué ventajas ofrece*. OpenWebinars: https://openwebinars.net/blog/que-es-visual-studio-code-y-que-ventajas-ofrece/

Galeano, S. (28 de Enero de 2022). *Cuáles son las redes sociales con más usuarios del mundo (2022)*. Marketing4ecommerce: https://marketing4ecommerce.net/cuales-redes-sociales-con-mas-usuarios-mundo-ranking/

García, M. N. (2021). *Práctica 1 Estimación del esfuerzo.* Universidad de Salamanca.

Gima, N. (30 de Septiembre de 2019). *La evolución del diseño web*. HostGator: https://www.hostgator.mx/blog/evolucion-del-diseno-web/

Hera, C. d. (2 de Junio de 2022). *Historia de las redes sociales: cómo nacieron y cuál fue su evolución*. Marketing4ecommerce: https://marketing4ecommerce.net/historia-de-las-redes-sociales-evolucion/

Invitado, A. (12 de Abril de 2020). *Bootstrap: guía para principiantes de qué es, por qué y cómo usarlo*. Rockcontent: https://rockcontent.com/es/blog/bootstrap/

Llonch, E. (25 de Mayo de 2021). *¿Qué son las redes sociales y cuáles son las más importantes?* Cyberclick: https://www.cyberclick.es/numerical-blog/que-son-las-redes-sociales-y-cuales-son-las-mas-importantes

Lucía Martín Gómez, R. C. (30 de Octubre de 2021). *Business Benefits of Instagram Scraping: Questionable Uses of Data.* Springer: https://link.springer.com/chapter/10.1007/978-3-030-87687-6\_22

Machuca, F. (1 de Junio de 2022). *Descubre cómo funciona Microsoft Project y cómo usarlo para la gestión de proyectos*. crehana: https://www.crehana.com/blog/desarrollo-web/que-es-microsoft-project/

MDN contributors. (8 de Diciembre de 2020). *Generalidades del protocolo HTTP*. MDN web docs: https://developer.mozilla.org/es/docs/Web/HTTP/Overview

MDN contributors. (11 de Febrero de 2021). *Introducción a Django*. MDN Web Docs: https://developer.mozilla.org/es/docs/Learn/Server-side/Django/Introduction

media. (14 de Abril de 2021). *¿Qué es Business Intelligence (BI) y qué herramientas existen?* Signaturit Blog: https://blog.signaturit.com/es/que-es-business-intelligence-bi-y-que-herramientas-existen

Mercedes. (11 de Septiembre de 2017). *¿Qué es el desarrollo web?* Openclassrooms: https://blog.openclassrooms.com/es/2017/09/11/que-es-el-desarrollo-web/

Oracle. (31 de Julio de 2014). *¿Qué es una base de datos?* https://www.oracle.com/es/database/what-is-database/

Ortega, J. M. (4 de Abril de 2019). *Que es GraphQL?* Medium: https://medium.com/@jmz12/que-es-graphql-bf835e55960

Pío. (2022). *Información sobre las API de Twitter*. Centro de Ayuda de Twitter: https://help.twitter.com/es/rules-and-policies/twitter-api#:~:text=Twitter%20permite%20acceder%20a%20partes,opini%C3%B3n%20del%20cliente%20en%20Twitter.

Portaltic. (09 de Febrero de 2022). *Nueve de cada diez personas usan redes sociales en España y dedican casi 2 horas diarias a ellas*. https://www.europapress.es/portaltic/socialmedia/noticia-nueve-cada-diez-personas-usan-redes-sociales-espana-dedican-casi-horas-diarias-ellas-20220209163949.html

Ramos, R. (25 de Septiembre de 2021). *¿Qué es JavaScript y para qué sirve?* Soyrafaramos: https://soyrafaramos.com/que-es-javascript-para-que-sirve/

Reyes, J. J. (6 de Mayo de 2016). *¿Qué es HTML?* DevCode: https://devcode.la/blog/que-es-html/

Robledado, A. (24 de Septiembre de 2019). *Qué es MySQL: Características y ventajas*. OpenWebinars: https://openwebinars.net/blog/que-es-mysql/

Rubio, J. C. (25 de Febrero de 2019). *Qué es GIT y para qué sirve*. OpenWebinars: https://openwebinars.net/blog/que-es-git-y-para-que-sirve/

Santos, D. (8 de Agosto de 2022). *Introducción al CSS: qué es, para qué sirve y otras 10 preguntas frecuentes*. HubSpot: https://blog.hubspot.es/website/que-es-css

Schiaffarino, A. (12 de Marzo de 2019). *Modelo cliente servidor*. Infranetworking: https://blog.infranetworking.com/modelo-cliente-servidor/

Seguro, N. (17 de Septiembre de 2021). *¿Qué es el desarrollo Web y por qué es importante?* Coderhouse: https://www.coderhouse.es/blog/que-es-el-desarrollo-web

SocioViz. (2014). *Twitter analytics made easy.* https://socioviz.net/

Sphinx. (6 de Agosto de 2022). *Getting Started*. https://www.sphinx-doc.org/en/master/usage/quickstart.html

Twitter. (2022). *Twitter API*. Developer Platform: https://developer.twitter.com/en/products/twitter-api

Visus, A. (Octubre de 2020). *¿Para qué sirve Python? Razones para utilizar este lenguaje de programación*. Esic: https://www.esic.edu/rethink/tecnologia/para-que-sirve-python

Zhao, B. (Mayo de 2017). *Web Scraping.* researchgate: https://www.researchgate.net/profile/Bo-Zhao-3/publication/317177787\_Web\_Scraping/links/5c293f85a6fdccfc7073192f/Web-Scraping.pdf

Zonín, A. (6 de Junio de 2019). *SocioViz es completamente nuevo*. Historias de datos digitales: https://digitaldatastories.it/2019/06/06/socioviz-is-completely-new/