



UNIVERSIDAD DE BURGOS  
ESCUELA POLITÉCNICA SUPERIOR  
Grado en Ingeniería Informática



**TFG del Grado en Ingeniería  
Informática**

**Data-WareHouse**



Presentado por Mario de la Parte Izquierdo  
en Universidad de Burgos — 15 de mayo  
de 2019

Tutor: Carlos Pardo Aguilar







UNIVERSIDAD DE BURGOS  
ESCUELA POLITÉCNICA SUPERIOR  
Grado en Ingeniería Informática



D. Carlos Pardo Aguilar, profesor del departamento de Ingeniería Civil, área de Lenguajes y Sistemas Informáticos.

Expone:

Que el alumno D. Mario de la Parte Izquierdo, con DNI 71305494C, ha realizado el Trabajo final de Grado en Ingeniería Informática titulado Data-WareHouse.

Y que dicho trabajo ha sido realizado por el alumno bajo la dirección del que suscribe, en virtud de lo cual se autoriza su presentación y defensa.

En Burgos, 15 de mayo de 2019

Vº. Bº. del Tutor:

Vº. Bº. del co-tutor:

D. Carlos Pardo Aguilar

D. nombre co-tutor





## Resumen

En este primer apartado se hace una **breve** presentación del tema que se aborda en el proyecto.

## Descriptores

Palabras separadas por comas que identifiquen el contenido del proyecto Ej: servidor web, buscador de vuelos, android ...

## **Abstract**

A **brief** presentation of the topic addressed in the project.

## **Keywords**

keywords separated by commas.



---

# Índice general

---

---

## Índice de figuras

---

---

# Índice de tablas

---



---

# Introducción

---

En la actualidad, existe una gran cantidad de información o datos, los cuales componen una parte muy importante en las grandes empresas y organizaciones de todo el mundo. Cada día se genera multitud de nueva información y es indispensable almacenarla para posteriormente poder interpretarla correctamente. Es imprescindible por lo tanto, saber extraer e identificar información relevante a partir de ficheros o documentos poco legibles o difíciles de entender a priori.

En este punto es cuando toma especial interés la creación de un Data Warehouse, o lo que es lo mismo, un almacén electrónico. En dichos almacenes se protege y mantiene una gran cantidad de datos e información, de manera fiable, segura y fácil de administrar.

Además de estas funciones de almacenamiento y administración, un Data Warehouse también permite organizar, entender y utilizar los datos para la toma de decisiones. Para esta tarea, es necesario contar con cierta capacidad de análisis, ya que hay que extraer información concreta, destacada y relevante; para posteriormente poder visualizarla con ayuda de elementos visuales como gráficos.

En la realización de este proyecto se propone la creación de un Data Warehouse, para procesar, almacenar y representar visualmente la información sobre la matriculación de alumnos en la Universidad de Burgos.

## 1.1. Estructura de la memoria

La memoria se estructura de la siguiente manera:

- **Introducción:** se describe brevemente el contexto y el proyecto realizado. Posteriormente se realiza una sección donde se expone la estructura de la memoria.
- **Objetivos del proyecto:** se exponen los objetivos del proyecto, divididos en objetivo general y objetivos técnicos.

---

# Objetivos del proyecto

---

A continuación se definen los objetivos del proyecto realizado, divididos en dos apartados:

## 2.1. Objetivo general

- Desarrollar una aplicación para analizar datos relacionados con la matriculación de alumnos en la Universidad de Burgos (UBU).

## 2.2. Objetivos técnicos

- Extraer los datos o información relevante de ficheros Excel (.xls), utilizando librerías concretas de Python.
- Crear las diferentes bases de datos para almacenar la información anteriormente extraída.
- Identificar y crear gráficos o estadísticos que resulten útiles para visualizar y comparar información.
- Desarrollar una aplicación en Python que unifique todo lo anterior, así como realizar una la interfaz gráfica agradable para el usuario.
- Conseguir que la aplicación sea fiable, usable y robusta.





---

# Aspectos relevantes del desarrollo del proyecto

---

En este apartado se van a recoger los aspectos más importantes que han surgido en el desarrollo del proyecto. Se incluirán la toma de decisiones, los posibles cambios, la aparición de problemas y las soluciones establecidas.

## 5.1. Inicio del proyecto

## 5.2. Metodologías

A lo largo del desarrollo del proyecto se ha usado la *metodología Scrum*. Se trata de una metodología ágil basada en *sprints*, en este caso, de desarrollo incremental con revisiones semanales.

Por lo tanto, la duración estimada de cada *sprint* es de una semana. Al finalizar cada *sprint*, se planificaba el siguiente, creando sus *issues* o tareas a realizar en dicho *sprint*. Cuando estas tareas se realizaban, se cambiaba el estado del *issue* correspondiente a *Closed*.

### **5.3. Toma de decisiones**

### **5.4. Librerías para el tratamiento y manipulación de datos**

### **5.5. Interfaz de usuario del proyecto**

### **5.6. Problemas encontrados**

#### **Error al abrir los Excel(.xls) bajados de Sigma con Python**

Los archivos Excel(.xls) suministrados (descargados de plataforma *Sigma*) no cumplen el estándar. Al abrirlos tanto con Excel como con varias librerías de Python, muestran un error de formato y extensión. Por lo tanto la única solución encontrada, ha sido realizar un parseo previo de los Excel suministrados, creando un fichero (.csv) nuevo, con toda la información del fichero original corrupto.

De esta manera, se ha creado un analizador sintáctico capaz de leer los ficheros originales (.xls) en modo texto y obtener un (.xml). Se ha parseado toda la información obteniendo filas, celdas, separaciones entre las mismas, contenidos de cada celda...etc. A la vez que se extrae toda esta información, se crea un fichero (.csv) nuevo y se van introduciendo los datos.