



UNIVERSIDAD DE BURGOS  
ESCUELA POLITÉCNICA SUPERIOR  
Grado en Ingeniería Informática



TFG del Grado en Ingeniería  
Informática

Sistema de Información sobre  
Matriculación



Presentado por Mario de la Parte Izquierdo  
en Universidad de Burgos — 26 de junio  
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 *Sistema de Información sobre Matriculación*.

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, 26 de junio de 2019

Vº. Bº. del Tutor:

D. Carlos Pardo Aguilar





## Resumen

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 adecuadamente.

El desarrollo de este proyecto viene motivado por realizar un Sistema de Información, o lo que es lo mismo, un almacén electrónico sobre la matriculación de alumnos en la Universidad de Burgos.

Con la creación del **Sistema de Información sobre Matriculación** se pretende crear una aplicación que sea capaz de procesar, almacenar, administrar, organizar y visualizar correctamente información relevante a la matriculación.

De esta manera, la Universidad de Burgos contará con un sistema cuya información se podrá utilizar para la toma de decisiones.

## Descriptores

Preprocesado Sigma, Sistema de Información, Matrícula

### **Abstract**

Nowadays, there is a large amount of data and information, which is a very great value for big companies and organizations all around the world. Every single day a multitude of new information is created, and it is indispensable to keep it saved, in order to be used properly later on.

The research and development of this project is to create a new Information System to keep an electronic store, based on Burgos University students's enrolment data.

With this new **Enrolment Information System**, I pretend to create an application that is capable of processing, storing, managing, organising and correctly visualising relevant enrolment information.

In this way, Burgos University will have an Enrolment System whose information can be used to find out things or to make decisions.

### **Keywords**

Preprocessed Sigma, Information System, Enrollment



---

# Índice general

---

Índice general	III
Índice de figuras	V
Índice de tablas	VI
Introducción	1
1.1. Estructura de la memoria . . . . .	2
Objetivos del proyecto	3
2.1. Objetivo general . . . . .	3
2.2. Objetivos técnicos . . . . .	3
Conceptos teóricos	5
3.1. Sistema de Información . . . . .	5
3.2. Parseado de Datos . . . . .	7
3.3. Base de datos . . . . .	7
3.4. Aplicación de escritorio . . . . .	7
3.5. Gráficos Representados . . . . .	7
Técnicas y herramientas	9
4.1. Metodologías . . . . .	9
4.2. Lenguaje de Programación . . . . .	10
4.3. Entorno de Desarrollo . . . . .	10
4.4. Control de Versiones . . . . .	11
4.5. Documentación . . . . .	11
4.6. Otras Herramientas . . . . .	12

<b>Aspectos relevantes del desarrollo del proyecto</b>	<b>15</b>
5.1. Inicio del proyecto . . . . .	15
5.2. Metodologías . . . . .	16
5.3. Toma de decisiones . . . . .	17
5.4. Librerías para el tratamiento y manipulación de datos . . . . .	17
5.5. Interfaz de usuario del proyecto . . . . .	19
5.6. Problemas encontrados . . . . .	19
<b>Trabajos relacionados</b>	<b>27</b>
<b>Conclusiones y Líneas de trabajo futuras</b>	<b>29</b>
<b>Bibliografía</b>	<b>31</b>

---

## Índice de figuras

---

3.1. Componentes de un Sistema de Información . . . . .	6
3.2. Información de Caja y Bigotes . . . . .	8
5.3. Logotipo del Sistema de Información sobre Matriculación . . . . .	16

---

## 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 Sistema de Información, 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 Sistema de Información 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 Sistema de Información, para procesar, almacenar y representar visualmente la información sobre la matriculación de alumnos en la Universidad de Burgos.

De esta forma, en el proyecto se podrán diferenciar varias funcionalidades:

- Preprocesamiento de los ficheros Excel(.xls) descargados de Sigma.
- Creación de la Base de Datos(BBDD).

- Carga de datos en la Base de Datos (BBDD) a partir de los ficheros (.csv) generados.
- Visualización de diferentes tipos de gráficos en función de los datos de la BBDD y lo que el usuario seleccione.

## 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.
- **Conceptos teóricos:** se exponen los conceptos teóricos y básicos para comprender tanto el proyecto como el desarrollo del mismo.
- **Técnicas y herramientas:** se explican las metodologías y herramientas utilizadas durante el desarrollo del proyecto.
- **Conclusiones y líneas de trabajo futuras:** se explican las conclusiones finales que se obtienen después de la realización del proyecto, así como futuras mejoras del mismo.

---

# 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

Se han propuesto cinco objetivos técnicos a realizar:

- Extraer los datos o información relevante de ficheros Excel (.xls), utilizando librerías concretas de Python.
- Crear la Base de Datos(BBDD) para almacenar la información anteriormente extraída, con una estructura de tablas, campos y claves adecuada.
- 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 interfaz gráfica agradable para el usuario.
- Conseguir que la aplicación sea fiable, usable y robusta.





---

# Conceptos teóricos

---

En este apartado se van a explicar aquellos conceptos teóricos básicos que son necesarios para comprender el proyecto.

## 3.1. Sistema de Información

### Definición

Para comenzar, hay que explicar que no existe una definición de consenso en la propia definición de Sistema de Información. De hecho, existen multitud de definiciones diferentes sobre cómo se define un Sistema de Información.

Los autores Laudon y Laudon definen un Sistema de Información como un conjunto de módulos relacionados entre sí que son capaces de obtener (o reutilizar), procesar, almacenar y distribuir cierta información para que sirva de apoyo para la toma de decisiones [1]. A parte de suministrar apoyo en decisiones importantes, también pueden ayudar a detectar problemas o carencias difíciles de ver sin la ayuda de estos sistemas.

### Componentes de un Sistema de Información

Aunque existen numerosas definiciones y no existe una definición general o global, la mayoría de Sistemas de Información pueden representarse a través del diagrama de la figura 3.1.

Se pueden apreciar 5 elementos principales. En primer lugar los elementos de entrada, que en nuestro proyecto serían los ficheros (.xls) originales descargados de *Sigma*.

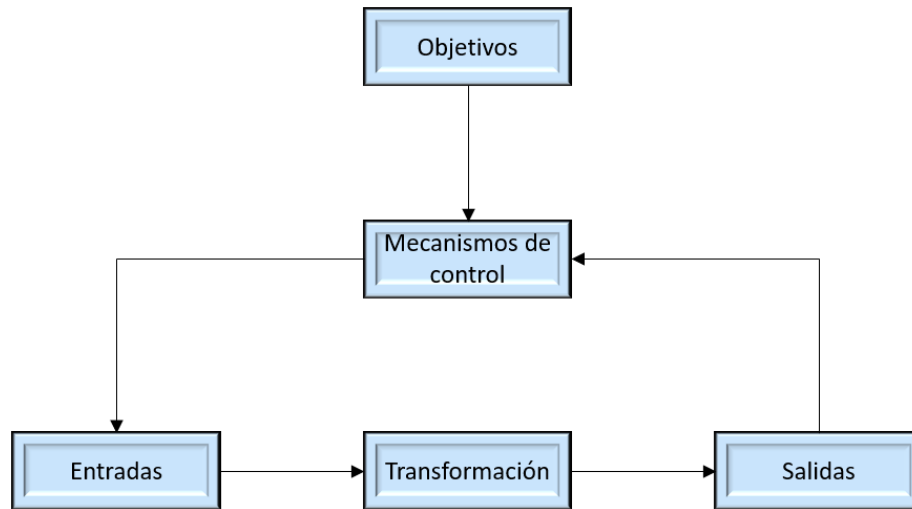


Figura 3.1: Componentes de un Sistema de Información

A continuación estaría un elemento de modificación o transformación, que en nuestro caso sería el preprocesado de los ficheros originales anteriores en ficheros (.csv) reordenados, modificados y sin ningún tipo de error. También se podría incluir la carga de datos a la Base de Datos creada con anterioridad.

Seguidamente estaría el sistema de salida, donde se visualizan los resultados obtenidos. En este proyecto, el sistema de salida serían los diferentes tipos de gráficos que se pueden obtener a partir de la información que seleccione el usuario y los datos existentes o disponibles en la BBDD.

Además de estas 3 secciones, se aprecian otras dos secciones más. Una de ellas es el mecanismo de control, que es el proceso encargado de lograr los objetivos, que sería el quinto y último elemento. En nuestro proyecto se podrían identificar numerosos mecanismos de control, como por ejemplo que los ficheros que se puedan seleccionar en los botones de *Preprocesar* y *Cargar Archivos* sean únicamente (.xls) y (.csv) respectivamente. Otros mecanismos de control serían la no introducción de datos repetidos en la BBDD o la selección de opciones de datos que realmente se encuentran en la BBDD, entre otros.

En cuanto a los objetivos de nuestro sistema de información, hay que destacar que se definen en el apartado anterior denominado *Objetivos del proyecto*.

## **Características de un Sistema de Información**

## **Tipos de Sistemas de Información**

## **Ventajas de un Sistema de Información**

### **3.2. Parseado de Datos**

El Preprocesado / Parseado de ficheros es un proceso mediante el cual...

### **3.3. Base de datos**

Una base de datos es ...

### **3.4. Aplicación de escritorio**

### **3.5. Gráficos Representados**

#### **Diagrama de Caja y Bigotes**

Los diagramas de cajas y bigotes (o diagramas de cuartiles) son un tipo de gráficas que representan una gran cantidad de información de manera muy visual y esquematizada. A su vez, se pueden apreciar características estadísticas relevantes como la simetría y la dispersión de un conjunto de datos.

Toda esta valiosa información se representa mediante unas pequeñas cajas muy intuitivas, como se aprecia a continuación:

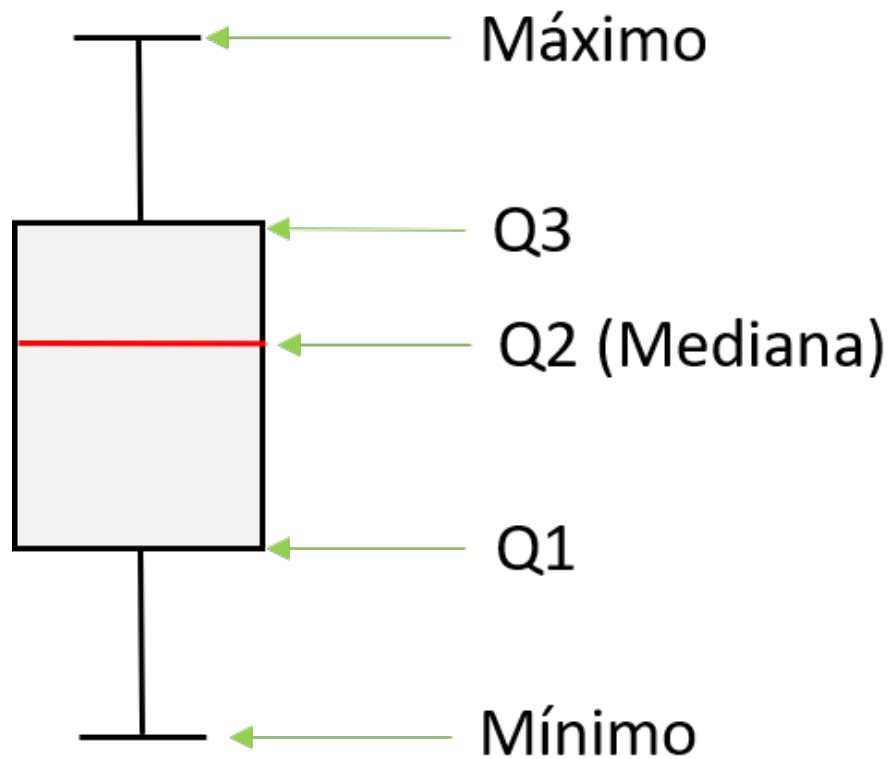


Figura 3.2: Información de Caja y Bigotes

Se diferencian cinco partes fundamentales:

- **Tres Cuartiles (Q1, Q2 y Q3).** Hay que destacar que el segundo cuartil(Q2) coincide con la mediana y representa la relación entre el primer y tercer cuartil. El primer cuartil identifica el valor por debajo del cual queda un 25 % de todos los datos de la muestra ordenada. Del mismo modo, el tercer cuartil es el valor por debajo del cual quedan el 75 % de los datos de la muestra.
- **Máximo.** Representa el valor máximo de los datos.
- **Mínimo.** Representa el valor mínimo de los datos.

## Gráfico de Barras

---

# Técnicas y herramientas

---

En este apartado se van a exponer las técnicas metodológicas y herramientas de desarrollo que se han utilizado para la realización del proyecto. Se detallarán las razones principales por los que se ha usado esa herramienta y no otra.

## 4.1. Metodologías

### Scrum

Para realizar la planificación correcta del proyecto, se ha utilizado *Scrum*, que es una metodología ágil de desarrollo.

- Se ha utilizado una estrategia orientada a un desarrollo incremental y basada en *sprints*.
- La duración media de cada *sprint* era aproximadamente de una semana.
- Al inicio de cada *sprint* se definían las tareas o *issues* a realizar, las cuales tenían que ser realizadas en un cierto intervalo de tiempo.
- Cada *sprint* se planificaba cuando se finalizaban las tareas o *issues* del anterior *sprint*.
- Al final de cada *sprint* se revisan todas las tareas realizadas, así como ver si se han logrado los objetivos fijados y solucionado los problemas encontrados.

Con la utilización de esta metodología se ha logrado evitar la realización de una planificación y ejecución completa desde el inicio del proyecto.

## 4.2. Lenguaje de Programación

### Python

El lenguaje de programación utilizado ha sido Python<sup>1</sup> en la versión 3.7.1.

Las razones por las que se ha decidido utilizar Python son las siguientes:

- Es uno de los lenguajes de programación más sencillos de aprender ya que su sintaxis es muy entendible.
- Es un lenguaje gratuito, multiplataforma y de código abierto.
- Gracias a las dos anteriores razones, se ha convertido en un lenguaje tan popular y utilizado, que ha dado lugar a que se desarrollen multitud de librerías, módulos y programas de software libre. Gran parte de estas librerías destacan en el ámbito de manejo de ficheros, tratamiento y visualización de datos.
- Del mismo modo, al ser un lenguaje utilizado por tantas personas, hace que existan numerosos foros, blogs y páginas en las que apoyarse cuando surgen dudas o se necesita ayuda.

## 4.3. Entorno de Desarrollo

### Jupyter Notebook

Como entorno de desarrollo principal se ha utilizado Jupyter Notebook<sup>2</sup> en la versión 5.7.4.

Se trata de una aplicación web de código abierto que permite tanto el desarrollo como la ejecución del código. Esta aplicación se puede lanzar directamente desde un navegador(sin instalar nada) o se puede instalar con *Anaconda Navigator*. Para el desarrollo del proyecto se utilizó la segunda opción.

Otra de las grandes ventajas de esta aplicación es la agilidad en el desarrollo, ya que al tratarse de una aplicación cuya ejecución es en vivo, se pueden realizar pruebas de manera rápida e intuitiva.

---

<sup>1</sup>[www.python.org](http://www.python.org)

<sup>2</sup>[www.jupyter.org](http://www.jupyter.org)

## SQLite

SQLite<sup>3</sup> se trata de un sistema de gestión de bases de datos(BBDD) relacionales de pequeño tamaño. Una de las características de este sistema de gestión es que no necesita un servidor para poder utilizarse, ya que los datos se almacenan en un único fichero en el sistema host. Hay que destacar que Python incluye soporte para SQLite desde la versión 2.5 incorporado en la Biblioteca Estándar como el módulo *sqlite3*, que es el módulo que se ha utilizado en el desarrollo del proyecto [4].

## 4.4. Control de Versiones

### GitHub

GitHub<sup>4</sup> se trata de una plataforma cuya función principal es la de hospedar repositorios y permitir el desarrollo colaborativo. Es una plataforma de las más usadas y por esta razón es la que se ha utilizado a lo largo del grado y en particular en la realización de este proyecto. Hay que destacar que gracias a formar parte de la Universidad de Burgos y ser estudiante, se ha obtenido la versión *PRO*(licencia de estudiantes). Aun así hay que destacar que se trata de una herramienta gratuita. Por último comentar que se ha utilizado tanto *GitHub Desktop*(aplicación de escritorio) como la plataforma web.

## 4.5. Documentación

### Texmaker

Para la realización de la documentación con L<sup>A</sup>T<sub>E</sub>X<sup>5</sup> se ha utilizado el editor Texmaker<sup>6</sup>. Se trata de un editor gratuito, el cual contiene las herramientas y características necesarias para desarrollar y editar documentos con L<sup>A</sup>T<sub>E</sub>X.

Hay que señalar que también incluye corrección ortográfica, auto-completado, plegado de código y un visor incorporado en pdf con soporte de synctex y un modo de visualización continua [3].

---

<sup>3</sup>[www.sqlite.org](http://www.sqlite.org)

<sup>4</sup>[www.github.com](http://www.github.com)

<sup>5</sup>[www.latex-project.org](http://www.latex-project.org)

<sup>6</sup>[www.xmlmath.net/texmaker](http://www.xmlmath.net/texmaker)

## 4.6. Otras Herramientas

En este apartado se van a explicar otras herramientas destacadas que se han utilizado a lo largo del proyecto.

### DB Browser

DB Browser<sup>7</sup> es una herramienta gratuita y de código abierto cuyo principal objetivo es la administración de Bases de Datos que utilizan SQLite como motor de las mismas. Esta herramienta cuenta con numerosas funcionalidades, entre las que se encuentran la creación de BBDD, tablas, índices, entradas, importar y exportar archivos, entre otras.

Hay que destacar que en este proyecto se ha utilizado la aplicación de escritorio para tareas de visualización de datos de la BBDD y comprobación de los mismos.

### Sublime Text 3 y Notepad++

Tanto Sublime Text 3<sup>8</sup> como Notepad++<sup>9</sup> son editores de código que pueden ser utilizados como entornos de desarrollo, ya que pueden interpretar numerosos lenguajes de programación.

En un primer lugar ambos editores se utilizaron para el desarrollo del proyecto, pero finalmente el uso de estas herramientas fue la de edición y visualización de ficheros. Hay que destacar funcionalidades como la mostrar caracteres ocultos de *Notepad++*, la visualización del texto en función de la sintaxis o lenguaje de programación que se elija en *Sublime Text 3* y las herramientas de búsqueda de ambos editores.

### Nitro Pro

Nitro Pro<sup>10</sup> es una herramienta gráfica cuya funcionalidad reside en la creación y edición de ficheros (.pdf). Hay que destacar que al tratarse de una herramienta de pago, se ha utilizado la versión de prueba de la misma, ya que contaba con las funcionalidades necesarias para el proyecto.

Esta herramienta se ha utilizado para la edición de ficheros (.pdf) como *about.pdf* de la interfaz gráfica.

---

<sup>7</sup>[www.sqlitebrowser.org](http://www.sqlitebrowser.org)

<sup>8</sup>[www.sublimetext.com](http://www.sublimetext.com)

<sup>9</sup>[www.notepad-plus-plus.org](http://www.notepad-plus-plus.org)

<sup>10</sup>[www.gonitro.com](http://www.gonitro.com)



## Excel

Excel<sup>11</sup> pertenece a la categoría de programas conocidos como hojas de cálculo [2].

De hecho es una de las herramientas más utilizadas y potentes en el análisis de datos, ya que cuenta con una gran cantidad de funcionalidades relacionadas con este área.

En la realización del proyecto, se ha utilizado principalmente para modificar(pruebas) y visualizar tanto los ficheros originales (.xls), como los generados (.csv).

## Photoshop

Photoshop<sup>12</sup> es un editor de imágenes dedicado principalmente para el retoque de fotografías y creación de gráficos.

Esta herramienta se ha utilizado para pequeñas tareas de diseño gráfico, como la realización de los botones de tipos de gráfico, el logotipo de la aplicación...etc.

---

<sup>11</sup>[www.support.office.com/es-es/excel](http://www.support.office.com/es-es/excel)

<sup>12</sup>[www.photoshop.com](http://www.photoshop.com)



---

# 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

Al principio se propuso la idea de la creación de una variante de *Data Ware-House* o almacén de datos, para almacenar información relevante para la matriculación de alumnos en la Universidad de Burgos(UBU) y su futura utilización para la toma de decisiones.

Por esta razón, la idea transmitida por el tutor fue la creación de un sistema para poder almacenar y visualizar cualquier grado o máster de la UBU.

El tutor comentó que los archivos que se iban a utilizar de partida, eran un tipo de ficheros descargado desde una aplicación denominada *Sigma* y tenían un error de formato.

Con este conjunto de ideas y proposiciones, se realizaron los *Objetivos del proyecto* y se enviaron al tutor para la supervisión de los mismos.

Tras la aprobación o visto bueno, se empieza a realizar el proyecto, con la creación del repositorio en *GitHub* el día 26 de Febrero de 2019.



Figura 5.3: Logotipo del Sistema de Información sobre Matriculación

## 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, si bien ha habido varios *sprints* que han tenido una duración superior. 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* o cerrado.

## 5.3. Toma de decisiones

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

En este apartado se van a explicar las diferentes librerías o bibliotecas que se han utilizado en el desarrollo del proyecto y su función principal. Todas las librerías explicadas a continuación son de *Python*.

### re

La primera librería que se ha utilizado en el proyecto ha sido [re](https://docs.python.org/3/library/re)<sup>13</sup>, ya que dicha librería contiene las funciones necesarias para trabajar con expresiones regulares.

Las expresiones regulares se han utilizado sobretodo en la parte inicial de parsear los ficheros descargados de Sigma. De esta manera se ha podido separar y extraer información por filas, celdas y contenido de las mismas.

### pandas

Pandas<sup>14</sup> es una librería que ofrece numerosas estructuras de datos de gran rendimiento y herramientas de análisis de datos.

Esta librería se ha utilizado principalmente para abrir y crear archivos con *Python* y para crear *dataframes* o estructuras auxiliares donde guardar datos.

### sqlite3

Como ya se ha comentado en un apartado anterior, la librería [sqlite3](https://docs.python.org/2/library/sqlite3.html)<sup>15</sup> proporciona una base de datos relacional de pequeño tamaño, ya que no necesita un servidor para poder utilizarse, ya que los datos se almacenan en un único fichero en el sistema host.

Esta librería se ha utilizado para la creación de la base de datos(BBDD), así como para la carga de datos y los procesos de consultas a la misma.

---

<sup>13</sup>[www.docs.python.org/3/library/re](https://docs.python.org/3/library/re)

<sup>14</sup>[www.pandas.pydata.org](https://www.pandas.pydata.org)

<sup>15</sup>[www.docs.python.org/2/library/sqlite3.html](https://www.docs.python.org/2/library/sqlite3.html)

## Tkinter

Tkinter<sup>16</sup> es una librería con numerosas funciones para hacer posible la creación de una interfaz gráfica en *Python*.

Se trata de una librería orientada a objetos y gracias a su facilidad de uso y rapidez para realizar una GUI, es una de las librerías más destacadas de este lenguaje de programación.

Por lo tanto, se ha utilizado principalmente en la creación de la interfaz gráfica(botones, despleables, ventanas nuevas...etc).

## math

La librería o componente math<sup>17</sup> incluye principalmente funciones matemáticas, para realizar operaciones aritméticas.

## matplotlib

La librería matplotlib<sup>18</sup> cuenta con multitud de funciones y características para generar gráficos. Hay que destacar que se pueden generar una gran variedad de gráficos(de sector, de barras...), así como personalizar los mismos(ejes, etiquetas, fuente, leyenda...)

Esta librería se ha utilizado para la realización y personalización de los gráficos de la interfaz gráfica.

## os

La librería os<sup>19</sup> de Python permite poder usar funcionalidades relacionadas con el Sistema Operativo.

Las funciones más destacadas de esta librería y las que se han utilizado son las que informan sobre el entorno del Sistema Operativo y las que permiten navegar por la estructura de directorios, ya sea para leer o modificar archivos.

---

<sup>16</sup>[www.docs.python.org/2/library/tkinter.html](http://www.docs.python.org/2/library/tkinter.html)

<sup>17</sup>[www.docs.python.org/3/library/math.html](http://www.docs.python.org/3/library/math.html)

<sup>18</sup>[www.matplotlib.org](http://www.matplotlib.org)

<sup>19</sup>[www.docs.python.org/3/library/os.html](http://www.docs.python.org/3/library/os.html)

## 5.5. Interfaz de usuario del proyecto

Para la realización de este proyecto se ha realizado una GUI (Graphical User Interface) o interfaz de usuario para hacer más sencillo la comunicación entre el usuario y el sistema.

Se ha optado por la realización de una aplicación de escritorio, ya que era la mejor opción para el uso que se espera que se le dé en un futuro a la aplicación.

## 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 (.xml) y finalmente obtener un (.csv). 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.

También se ha decidido modificar el fichero (.csv) resultante, añadiendo una nueva columna al final del encabezado principal de los datos. Esta nueva columna se ha llamado *Plan* y de esta manera, se evita tener que estar volviendo a introducir cada encabezado de datos(repetido) por cada diferente Plan o Titulación que se incluya en el fichero (.xls).

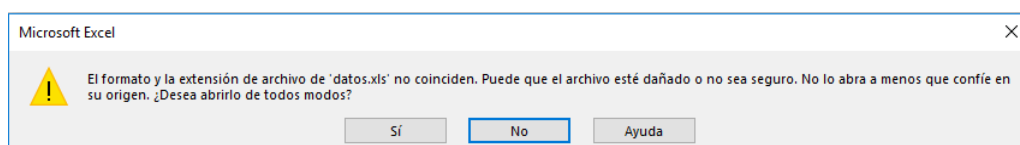
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG																										
1	Control de Grupos de Matrícula																																																										
2	Año ac 2018/19-0    Fech 29/03 Centro (c 13-) Ccerf 13 Plan 263 Plan 126+ Típc Todas																																																										
3	Plan 263 - MÁSTER UNIVERSITARIO EN INTELIGENCIA DE NEGOCIO Y BIG DATA EN ENTORNOS SEGUROS																																																										
4	Plan 264 - MÁSTER UNIVERSITARIO EN INGENIERIA INFORMÁTICA (PLAN 2018)																																																										
5																																																											
6	Asignatura																																																										
7	Código	Descripción	Curso	Grupo	bología académica	Activ.	Tp	Vp	Turno	Exámenes programaticidad	Rep./lo Rep/lejo	Repl/lo Repl/lejo	Repl/lo Repl/lejo	Repl/lo Repl/lejo	Propios	Externos	gr. in Total	disg Descripción	Asig. ar/b desde hasta	PF	NIU	Nombre y Apellidos	CDS	Actas	Docencia	Responsable																																	
8	8093	ATURA PARA	1	90	Teoría	S	S	1	Mixto	0	27	0	0	8	19	0	0	0	8	1	Manual	4594	1014645	3 MUÑOZ, FRANCIS	S	N	S	S																															
9	8094	ROGRAMA/O	1	90	Teoría	S	S	1	Mixto	0	27	0	0	8	19	0	0	0	8	1	Manual	1125	53453	maz Gonzalez, Alvar	S	S	S	N																															
10	8095	ECTURAS BK	1	90	Teoría	S	S	1	Mixto	0	27	0	0	8	19	0	0	0	8	1	Manual	1335	55897	Pastor, Jose Franc	S	S	S	S																															
11	8107	DE NEGOCIO	1	90	Teoría	S	S	2	Mixto	0	27	0	0	7	20	0	0	0	7	1	Manual	4745	1016217	rñez Prieto, Miguel	S	S	S	N																															
12	8110	JO FN DE M	1	90	Teoría	S	S	2	Mixto	0	27	0	0	4	23	0	0	0	4	1	Manual	4582	1016218	J CASTILLO, JOSE	S	N	S	N																															
13	8110	JO FN DE M	1	90	Teoría	S	S	2	Mixto	0	27	0	0	4	23	0	0	0	4	1	Manual	1335	55897	Pastor, Jose Franc	S	N	S	N																															
14	8110	JO FN DE M	1	91	Teoría	S	S	2	Mixto	0	27	0	0	0	27	0	0	0	0	1	Manual	1799	1000466	lguerez Diaz, Juan J	S	N	S	N																															
15	Plan 264 - MÁSTER UNIVERSITARIO EN INGENIERIA INFORMÁTICA (PLAN 2018)																																																										
16	Asignatura																																																										
17	Código	Descripción	Curso	Grupo	bología académica	Activ.	Tp	Vp	Turno	Exámenes programaticidad	Rep./lo Rep/lejo	Repl/lo Repl/lejo	Repl/lo Repl/lejo	Repl/lo Repl/lejo	Propios	Externos	gr. in Total	disg Descripción	Asig. ar/b desde hasta	PF	NIU	Nombre y Apellidos	CDS	Actas	Docencia	Responsable																																	
18	8111	ENICA Y ESTR	1	90	Teoría	S	S	1	Mixto	0	15	0	0	15	0	12	3	0	0	0	12	1	Manual	2912	63707	rcos Arcos, Sanjaiz	S	S	S	S																													
19	8112	AVANZA/OA	1	90	Teoría	S	S	1	Mixto	0	10	0	0	10	0	8	2	0	0	0	8	1	Manual	1403	1000235	ara Nebreda, José f	S	S	S	S																													
20	8113	GENCIA ARTI	1	90	Teoría	S	S	1	Mixto	0	15	0	0	15	0	11	4	0	0	0	11	1	Manual	1125	53453	maz Gonzalez, Alvar	S	S	S	S																													
21	8113	GENCIA ARTI	1	90	Teoría	S	S	1	Mixto	0	15	0	0	15	0	11	4	0	0	0	11	1	Manual	1642	39809	lerrero Cosío, Alvar S	S	S	S	S																													
22	8114	AY SERVIC	1	90	Teoría	S	S	1	Mixto	0	10	0	0	10	0	7	3	0	0	0	7	1	Manual	3137	35796	fondo Guevra, Bac	S	N	N	N																													



De esta manera, obtenemos:

[illegible]

En un primer lugar no se esperaba que estos ficheros fueran a generar muchos problemas, ya que al tratarse de una hoja de cálculo de Excel con extensión (.xls), al abrir los ficheros con dicho programa, se mostraba una pantalla de error, indicando el siguiente tipo de error, el cual el propio programa era capaz de solucionar, pudiendo visualizar todos los datos contenidos en el archivo:



Como se conseguía ver el contenido de los ficheros, así como su formato (color de celdas, celdas unificadas...); se pensó que podría ser un pequeño problema. Realmente se apreció la dimensión del problema cuando se intentó importar y abrir los datos con librerías específicas de *Python* como *Pandas*.

Al comprobar que no se podía cargar u obtener la información de de ninguna manera, se optó por crear un analizador sintáctico o parser para realizar un parseo de los datos. De esta manera se abría el fichero (.xls) como si fuera un fichero de texto, con lo cual obteníamos un fichero (.xml) para analizar. Tras analizar y parsear este último fichero, se obtenía toda la información del (.xls) original (ROW, CELL, MergeDown, MergeAcross, DATA...) y con esta información se creaba un fichero resultante (.csv).

Más adelante se optó por eliminar las cabeceras repetidas en el caso de tener más de una Titulación o Plan en un mismo fichero (recuadro en rojo de la imagen x).

Como en las cabeceras, la información del Plan (recuadro verde de la imagen x) no se debía perder, se añadió una nueva columna (imagen y) para almacenar esa información de las cabeceras repetidas que se iban a prescindir.

## 23

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG																											
1	Control de Grupos de Matrícula																																																											
2	Año ac 2018/19-0    Fech 29/03 Centro (c 13)- I Cerf 13 Plan 263 Plan 126+ Típc Todas																																																											
3	Plan 263 - MÁSTER UNIVERSITARIO EN INTELIGENCIA DE NEGOCIO Y BIG DATA EN ENTORNOS SEGUROS																																																											
4	Plan 263 - MÁSTER UNIVERSITARIO EN INTELIGENCIA DE NEGOCIO Y BIG DATA EN ENTORNOS SEGUROS																																																											
5	Asignatura																																																											
6	Código	Descripción	Curso	Grupo	bología académica	Activ.	Tp	Vp	Turno	Exámenes propiamente Rep. (lo Rep)	Replazamiento	Propios	Max.	Exce.	Asig.	Disp.	Externos	Max.	Asig.	Disp.	gr. in Total	Logro descripción	Asig. arb.	desde hasta	PF	NIU	Nombre y Apellidos	CDS	Actas	Docencia	Responsable																													
7	8093	ATURA PARA	1	90	Teoría	S	S	1	Mixto	0	27	0	0	27	0	8	19	0	0	0	0	8	1	Manual	4564	1014643	MUÑOZ, FRANCIS	S	N	S	S																													
8	8094	ROGRAMACIÓN	1	90	Teoría	S	S	1	Mixto	0	27	0	0	27	0	8	19	0	0	0	0	8	1	Manual	1125	53453	maz Gonzalez, Alvar	S	S	S	S																													
9	8094	ROGRAMACIÓN	1	90	Teoría	S	S	1	Mixto	0	27	0	0	27	0	8	19	0	0	0	0	8	1	Manual	1335	55897	Pastor, Jose Franc	S	S	S	S																													
10	8095	ECTURAS BK	1	90	Teoría	S	S	1	Mixto	0	27	0	0	27	0	8	19	0	0	0	0	8	1	Manual	4745	1016827	rñez Prieto, Miguel	S	S	S	S																													
11	8107	DE NEGOCI	1	90	Teoría	S	S	2	Mixto	0	27	0	0	27	0	7	20	0	0	0	0	7	1	Manual	4582	1015078	CASTILLO, JOSE	S	N	S	N																													
12	8110	JO FIN DE M	1	90	Teoría	S	S	2	Mixto	0	27	0	0	27	0	4	23	0	0	0	0	4	1	Manual	1335	55897	Pastor, Jose Franc	S	N	S	N																													
13	8110	JO FIN DE M	1	90	Teoría	S	S	2	Mixto	0	27	0	0	27	0	4	23	0	0	0	0	4	1	Manual	1799	1000466	Iriguez Diez, Juan J	S	N	S	N																													
14	8110	JO FIN DE M	1	91	Teoría	S	S	2	Mixto	0	27	0	0	27	0	0	27	0	0	0	0	0	1	Manual																																				
15	Plan 264 - MÁSTER UNIVERSITARIO EN INGENIERIA INFORMÁTICA (PLAN 2018)																																																											
16	Asignatura																																																											
17	Código	Descripción	Curso	Grupo	bología académica	Activ.	Tp	Vp	Turno	Exámenes propiamente Rep. (lo Rep)	Replazamiento	Propios	Max.	Exce.	Asig.	Disp.	Externos	Max.	Asig.	Disp.	gr. in Total	Logro descripción	Asig. arb.	desde hasta	PF	NIU	Nombre y Apellidos	CDS	Actas	Docencia	Responsable																													
18	8111	ENICA Y ESTR	1	90	Teoría	S	S	1	Mixto	0	15	0	0	15	0	12	3	0	0	0	0	12	1	Manual	2912	63707	rcos Arcos, Sanluis	S	S	S	S																													
19	8112	AVANZADA	1	90	Teoría	S	S	1	Mixto	0	10	0	0	10	0	8	2	0	0	0	0	8	1	Manual	1403	1000235	ara Nebreda, José F	S	S	S	S																													
20	8113	GENCIA ARTI	1	90	Teoría	S	S	1	Mixto	0	15	0	0	15	0	11	4	0	0	0	0	11	1	Manual	1125	53453	maz Gonzalez, Alvar	S	S	S	S																													
21	8113	GENCIA ARTI	1	90	Teoría	S	S	1	Mixto	0	15	0	0	15	0	11	4	0	0	0	0	11	1	Manual	1642	39809	lerra Cosío, Alvar	S	S	S	S																													
22	8114	GENCIA ARTI	1	90	Teoría	S	S	1	Mixto	0	15	0	0	15	0	11	4	0	0	0	0	11	1	Manual	3137	35796	fondo Guevara, Rak	S	N	N	N																													
23	8114	AY SERVIC	1	90	Teoría	S	S	1	Mixto	0	10	0	0	10	0	7	3	0	0	0	0	7	1	Manual	3137	35796	fondo Guevara, Rak	S	N	N	N																													

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
1	Control de Grupos de Matrícula																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								</

De esta manera se obtuvo un fichero (.csv) final mejor estructurado y preparado para hacer más eficiente la carga de datos a la BBDD.



---

## Trabajos relacionados

---

Este apartado sería parecido a un estado del arte de una tesis o tesina. En un trabajo final grado no parece obligada su presencia, aunque se puede dejar a juicio del tutor el incluir un pequeño resumen comentado de los trabajos y proyectos ya realizados en el campo del proyecto en curso.





---

## **Conclusiones y Líneas de trabajo futuras**

---

Todo proyecto debe incluir las conclusiones que se derivan de su desarrollo. Éstas pueden ser de diferente índole, dependiendo de la tipología del proyecto, pero normalmente van a estar presentes un conjunto de conclusiones relacionadas con los resultados del proyecto y un conjunto de conclusiones técnicas. Además, resulta muy útil realizar un informe crítico indicando cómo se puede mejorar el proyecto, o cómo se puede continuar trabajando en la línea del proyecto realizado.



---

## Bibliografía

---

- [1] Vicenç Fernández Alarcón. *Desarrollo de Sistemas de Información. Una Metodología Basada En El Modelado*. Univ. Politèc. de Catalunya, 2006.
- [2] Moisés Ortíz. Excel total, 2019.
- [3] Wikipedia. Texmaker — wikipedia, la enciclopedia libre, 2018.
- [4] Wikipedia. Sqlite — wikipedia, la enciclopedia libre, 2019.