



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería en Informática



TFG del Grado en Ingeniería Informática

Sistema de Recomendación de
Asignaturas Optativas



Presentado por Clara Palacios Rodrigo
en Universidad de Burgos — 24 de mayo de 2018
Tutor: Dr. José Ignacio Santos Martín, Virginia
Ahedo García



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería en Informática



Dr. José Ignacio Santos Martín y Dña. Virginia Ahedo García, profesores del departamento de Ingeniería Civil , área de Organización de Empresas.

Expone:

Que el alumno D. Clara Palacios Rodrigo, con DNI 71307844-R, ha realizado el Trabajo final de Grado en Ingeniería Informática titulado Sistema de Recomendación de Asignaturas Optativas .

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, 24 de mayo de 2018

Vº. Bº. del Tutor:

Vº. Bº. del Tutor:

Dr. José Ignacio Santos Martín

Dña. Virginia Ahedo García

Resumen

Los sistemas de recomendación son una herramienta ampliamente utilizada en el área informática, tanto en el comercio electrónico como en diferentes estudios del mercado en las empresas. Su funcionalidad se basa en la utilización de los algoritmos para la selección de los items. Sin embargo, a pesar del buen funcionamiento de los sistemas de recomendación con gran número de usuarios, resulta problemático para aquellos que comienzan a utilizarlo, por la falta de datos significativos de las preferencias del usuario, de forma que se puede recomendar únicamente aquellos productos preferentes entre los clientes.

La base de un buen sistema de recomendación se centra en los datos recabados así como el algoritmo utilizado, ya que diferentes algoritmos pueden ofrecer diferentes ponderaciones a un mismo producto. Al ser un método de cálculos de preferencias tan utilizado en la actualidad, se ha decidido orientar el Trabajo Final de Grado a la utilización de los sistemas de recomendación para la selección de asignaturas de 4º curso.

Descriptores

Python, Recomendación, Predicción, Formalización, Filtro Colaborativo (FC), usuarios, items, Red Neuronal, Algoritmos.

Abstract

Recommendation systems are a widely used tool in the computer area, both in electronic commerce and in different market studies in companies. Its functionality is based on the use of algorithms for the selection of items. However, despite the good functioning of the recommendation systems with a large number of users, it is problematic for those who start using it, due to the lack of significant data of the user's preferences, so that only those preferred products can be recommended among the customers

The basis of a good recommendation system focuses on the data collected as well as the algorithm used, since different algorithms can offer different weightings to the same product. As it is a method of preference calculations so widely used today, it has been decided to orient the Final Degree Project to the use of recommendation systems for the selection of 4th year subjects.

Keywords

Python, Recommendation, Prediction, Formalization, Collaborative Filter (CF) .

Índice general

Índice general	III
Índice de figuras	V
Índice de cuadros	VI
Introducción	1
Objetivos del proyecto	2
Conceptos teóricos	3
3.1. Sistemas de Recomendación	3
Técnicas y herramientas	6
4.1. Metodología Ágil	6
4.2. Repositorio de código y control de versiones	7
4.3. Gestores de Tareas	9
4.4. Editores de Texto	10
4.5. Gestores de Referencias Bibliográficas	11
4.6. Editores de código	13
4.7. Recogida de datos	13
4.8. Integración de funcionalidades del cuestionario	15
4.9. Herramienta para el desarrollo	15
Aspectos relevantes del desarrollo del proyecto	16
Trabajos relacionados	17
6.1. Amazon	17
6.2. Spotify	17
6.3. Youtube	18

Conclusiones y Líneas de trabajo futuras	19
Asignaturas de Ingeniería Informática	20
8.1. Asignaturas Obligatorias	20
8.2. Asignaturas Optativas	25
Bibliografía	28

Índice de figuras

Índice de cuadros

3.1.	Tabla de ejemplo de filtro colaborativo basado en modelo	4
8.2.	Tabla de asignaturas del Primer Curso Académico	21
8.3.	Tabla de asignaturas del Segundo Curso Académico	23
8.4.	Tabla de asignaturas del Tercer Curso Académico	24
8.5.	Tabla de asignaturas del Cuarto Curso Académico	27

Introducción

Este proyecto consiste en resolver el problema de la elección de las asignaturas optativas del último curso entre los estudiantes del grado de Ingeniería Informática.

Dado el desconocimiento de la materia impartida en 4º curso y la falta de visión de la relación entre la misma, un gran porcentaje de estudiantes universitarios consideran que su elección en alguna de las asignaturas no ha sido la correcta.

Por ese motivo, para evitar confusiones por parte de los alumnos, se creará una aplicación mediante la cual, el usuario rellenará un cuestionario anónimo, introduciendo una calificación por cada una de las asignaturas cursadas, de forma que el sistema sea capaz de indicarle cuál de las optativas se ajusta más a su criterio.

Sin embargo, a pesar de poner en práctica diferentes filtros colaborativos, los resultados no son exactos, ya que el número de datos recopilados hasta el momento son mínimos con respecto a los sistemas de recomendación utilizados en grandes empresas, de forma que la recomendación ofrecida al usuario es simplemente orientativa.

Objetivos del proyecto

Los objetivos del proyecto se basan en la resolución de la incógnita de qué asignaturas optativas escoger en el último curso académico de Ingeniería Informática basándose en las preferencias de los cursos anteriores. Para ello se debe:

- Analizar el problema propuesto y escoger el algoritmo del sistema de recomendación preferente:
 - Documentarse en los diferentes Sistemas de Recomendación existentes y la decisión de qué sistema se ajusta más a este problema.
 - Documentarse en diferentes métodos de creación de Cuestionarios Anónimos así como la recogida de los datos.
- Creación de un cuestionario y la obtención de los datos.
 - Difusión entre los alumnos que estén cursando -o hayan finalizado la carrera- del cuestionario para la obtención de los datos de sus preferencias en las asignaturas.
 - Tratamiento de dichos datos con el sistema de recomendación escogido.
- Creación de una base de datos.
 - Creación de una base de datos que aloje la información de los usuarios registrados.
 - Sincronización de la base de datos con la aplicación de forma que un usuario pueda iniciar sesión con su correo y contraseña.

Conceptos teóricos

3.1. Sistemas de Recomendación

Un sistema de recomendación es aquella herramienta encargada de recopilar, filtrar y devolver un conjunto de datos destinados para ofrecer al consumidor una orientación en base a sus gustos o en base a los gustos de otros usuarios cuyos aspectos introducidos coincidan.

Los sistemas de recomendación se subdividen en dos grupos:

Filtro colaborativo

Técnica utilizada en los sistemas de recomendación basada en predicción de las preferencias basado en las selecciones de los usuarios. Para ello, el sistema funciona de la siguiente forma:

1. El usuario rellena el formulario con respecto a las preferencias de aquellos elementos conocidos.
2. El sistema recopila los datos y realiza las comparaciones con otros usuarios para determinar las similitudes.
3. El sistema informa de aquellos items en los que el cliente podría coincidir con el resto de usuarios pero no ha rellenado. [35]

Una explicación más sencilla sería en que el sistema ofrece a un usuario lo más popular entre aquellos con sus mismos gustos. Existen dos clases de Sistemas de Recomendación con filtro colaborativo:

Basado en modelos

La estimación de los items se basa en el rating de los items de los diferentes usuarios. Por ello, para su utilización, se crea inicialmente una tabla, en la que

Cuadro 3.1: Tabla de ejemplo de filtro colaborativo basado en modelo

Example	$item_1$	$item_2$...	$item_n$
$user_1$	5	2	...	4
$user_2$	5	1	...	3
...	2	5	...	5
$user_n$	4	?	...	3

las filas son los diferentes usuarios, mientras que las columnas hacen referencia a los diferentes elementos. Dicha estimación utiliza la minería de datos con algoritmos de aprendizaje automático. Entre los ejemplos de los algoritmos existentes, podemos indicar los clasificadores bayesianos, redes neuronales y algoritmos genéticos.

Posteriormente se ajusta el modelo para proponer al usuario un valor recomendado para el ítem desconocido (?).

Basado en memoria

La estimación de los ítems se basa en los rating del cliente con respecto a los ítems de los diferentes usuarios. El sistema, tras rellenar la tabla de usuarios/ítems busca las similitudes entre las preferencias de los usuarios para ofrecer una recomendación. A su vez, el sistema de recomendación con filtro colaborativo basado en memoria se puede subdividir en:

1. Basado en Usuarios

La estimación del $item_i$ de un $usuario_u$ concreto se basa en los rating del $item_i$ de los usuarios con mayor proximidad al $usuario_u$.

Se han utilizado dos fórmulas diferentes para el cálculo del sistema de recomendación basado en usuarios:

- Similitud de coseno:

$$sim_{u,v} = \frac{\sum_{i \in I} r_{u,i} r_{v,i}}{\sqrt{\sum_{i \in I} r_{u,i}^2} \sqrt{\sum_{i \in I} r_{v,i}^2}} \quad (3.1)$$

- Coeficiente de Pearson:

$$sim_{u,v} = \frac{\sum_{i \in I} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I} (r_{v,i} - \bar{r}_v)^2}} \quad (3.2)$$

La fórmula para estimar un ranking de un usuario es la siguiente :

$$\hat{r}_{u,i} = \bar{r}_u + \frac{\sum_{v \in \mathcal{U}} (r_{v,i} - \bar{r}_v) sim_{u,v}}{\sum_{v \in \mathcal{U}} |sim_{u,v}|} \quad (3.3)$$

2. Basado en Productos

La estimación del $item_i$ de un $usuario_u$ concreto se basa en los rating del $usuario_u$ a los items similares a $item_i$.

- Similitud de coseno :

$$sim_{i,j} = \frac{\sum_{u \in \mathcal{U}} r_{u,i} r_{u,j}}{\sqrt{\sum_{u \in \mathcal{U}} r_{u,i}^2} \sqrt{\sum_{u \in \mathcal{U}} r_{u,j}^2}} \quad (3.4)$$

- Coeficiente de Pearson:

$$sim_{i,j} = \frac{\sum_{u \in \mathcal{U}} (r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in \mathcal{U}} (r_{u,i} - \bar{r}_i)^2} \sqrt{\sum_{u \in \mathcal{U}} (r_{u,j} - \bar{r}_j)^2}} \quad (3.5)$$

La fórmula para estimar un ranking en un sistema basado en productos es la siguiente:

$$\hat{r}_{u,i} = \frac{\sum_{j \in \mathcal{I}} r_{u,j} sim_{i,j}}{\sum_{j \in \mathcal{I}} |sim_{i,j}|} \quad (3.6)$$

Filtrado basado en contenido

Técnica utilizada en los sistemas de recomendación, basada en asignar unos pesos a cada uno de los items conocidos por el usuario, de forma que se pueda utilizar el algoritmo para extraer y recomendar aquellos items que tengan mayor similitud con los gustos del usuario. Su explicación se basaría en ofrecer al usuario aquello que se relaciona más con lo que le ha gustado.

Técnicas y herramientas

4.1. Metodología Ágil

La metodología ágil es aquella forma de toma de decisiones en los proyectos software basado en un desarrollo iterativo, evaluando las necesidades y tareas necesarias a la par de la realización, para añadir funcionalidad en el proyecto.[34]

KANBAN

Sistema de información basado en la metodología ágil. Utiliza tarjetas para representar la información de forma visual, mejorando la distribución de trabajo y la organización del mismo. [30] La representación Kanban se desarrolla en un tablero, asignando tareas a los diferentes miembros, utilizando tarjetas para indicar las diferentes etapas en las que se encuentran las subtarefas del proyecto. [41]

SCRUM

Modelo de referencia basado en metodología ágil basado en un desarrollo incremental que permite modificar tareas e ideas una vez comenzado el proyecto. Hemos seleccionado esta opción, ya que entre sus ventajas podemos indicar:

- Flexibilidad ante las diferentes necesidades y posibles cambios.
- Entrega progresiva del proyecto.
- Control en todas las etapas del proyecto.

4.2. Repositorio de código y control de versiones

VersionOne

Plataforma compacta basada en gestión ágil, está estructurada para admitir diferentes metodologías ágiles, tales como Scrum, Kanban, Lean o XP, permitiendo la rápida escalabilidad de los proyectos. [21]

Como ventajas podemos indicar que:

- Fácil de utilizar.
- Gran cantidad de métricas ágiles
- Fácilmente escalable.

Sin embargo, es un software de pago, ya que las utilidades necesarias para el Trabajo Final de Carrera no están incluidas en la versión gratuita por lo que ha sido descartada esta opción.

Bitbucket

Plataforma de almacenamiento de código que utiliza el sistema de control de versiones basado en gestión ágil.

- Ventajas
 - Número ilimitado de creación de repositorios privados gratuitos.
 - Control de versiones Git y Mercurial
 - Software de comunicación entre los diferentes colaboradores de un proyecto.
- Inconvenientes
 - Número limitado de colaboradores en un proyecto gratuito.
 - Exclusión de un soporte para el sistema operativo Linux.
 - El coste por tener un mayor número de colaboradores en un proyecto es muy superior a Github-cuyo coste se basa en el número de repositorios privados existentes.

Gitlab

Sistema de control de versiones basado en gestión ágil y gestor de repositorio de código con licencia MIT-de software libre permisivo-

- Ventajas

- Facilidad en adjuntar archivos en los issues.
- Protección de ramificaciones frente a cambios utilizando el sistema de niveles de autorización. [1]
- Posibilidad de crear repositorios privados de forma gratuita.
- No hay limitaciones ni en el número de repositorios privados ni en el número de colaboradores.

■ Inconvenientes

- Menor cantidad de miembros en la comunidad.
- Gestor de repositorio totalmente nuevo, por lo que se debería estudiar su funcionamiento.

SourceForge

Sitio web de software libre para almacenar y compartir proyectos de código abierto.

■ Ventajas

- Gran número de proyectos alojados.

■ Inconvenientes

- Monetización de la descarga de los proyectos basándose en la inclusión de anuncios publicitarios en los instaladores.[43]
- Necesidad de aprendizaje

Github

GitHub es una plataforma de desarrollo colaborativo de código para alojar proyectos utilizando el sistema de control de versiones Git. Permite gestionar los cambios de código indicando qué se ha modificado, añadido o eliminado en cualquier commit realizado.

Entre sus ventajas e inconvenientes, podemos indicar:

■ Ventajas

1. Servicio gratuito en los repositorios públicos.
2. No es necesario realizar copias de seguridad del código, ya que utiliza un sistema de control de versiones.
3. Compatible con diferentes sistemas operativos (Windows, Linux, OS...).

- Inconvenientes

1. Los repositorios privados no son gratuitos.

Sin embargo, a pesar del inconveniente del pago en caso de desear un repositorio privado, las ventajas de la utilización de GitHub tienen un mayor peso, además de haber sido la plataforma utilizada en los cursos académicos del grado de Ingeniería Informática.

4.3. Gestores de Tareas

Trello

Herramienta para la gestión y seguimiento de las tareas en tiempo real. Para acceder a dicha herramienta, se debe crear una cuenta en Trello.com. Sin embargo, para evitar el uso masivo de herramientas y aplicaciones innecesarias en nuestro proyecto, descartamos esta opción.

Hojas de Cálculo

Una hoja de cálculo es una herramienta para la manipulación de datos y gráficos en hojas divididas en celdas. Se podría Sin embargo, ante la falta de orden y de un correcto seguimiento, se ha descartado dicha herramienta desde un comienzo.

Zenhub

ZenHub es una extensión integrada en Chrome orientada a la gestión y el control de proyectos utilizando pizarras o columnas en las que se ordenan las issues creadas para mejorar la clasificación de las mismas. Utilizaremos esta herramienta para la gestión y el seguimiento de las diferentes tareas, creando nuevas Pipelines y utilizando las existentes. Entre ellas, podemos encontrar:

- ICEBOX

Issues que se hayan comenzado, y sin embargo, por cualquier motivo, se deben dejar paradas.

- Backlog

Issues que puede que se vayan a desarrollar más adelante.

- To do

Issues pendientes por desarrollar.

- In progress

Issues que están en desarrollo en ese mismo momento

- Closed
Issues que se han terminado y han sido cerradas.

4.4. Editores de Texto

Microsoft Word

Microsoft Word es un programa de la empresa Microsoft creado para la edición de texto[42]

- Ventajas
 - Interfaz gráfica que permite una mejor comprensión de la herramienta para los usuarios nóveles.
 - Sencillez ante la edición de texto así como la utilización de los diferentes formatos.
- Inconvenientes
 - Fallo de seguridad en los complementos la herramienta que permite el robo de archivos mediante la introducción de un documento con código oculto.[24]
 - Limitación de la capacidad para la inclusión y el tratamiento de las imágenes.
 - Problemática en el mantenimiento del mismo formato en todo el documento.

OpenOffice

Paquete de software de código abierto destinado para el procesamiento y la edición de texto.

- Ventajas
 - Al igual que Microsoft Word, consta de una interfaz gráfica fácil de utilizar en la edición y tratamiento de texto, gráficos y tablas.
 - No tiene coste de licencia-al contrario que Microsoft Word.
 - Multiplataforma.
- Inconvenientes
 - Lentitud en el procesamiento, edición y guardado de archivos

- Incompatibilidad con Microsoft Word, por lo que varias de las acciones disponibles en OpenOffice Writer no se reconocen en Microsoft Word; acciones tales como es el pegado de una tabla.
- Al igual que ocurre en Microsoft Word, puede ser dificultoso mantener el mismo formato en todo el documento.

L^AT_EX

Sistema de edición y procesamiento de texto basado en comandos de TeX¹. Los archivos de L^AT_EX tienen la extensión .tex, cuyo contenido debe ser compilado de forma previa a su visualización. En este proyecto, la compilación será PDFLaTeX, para obtener el formato PDF del fichero escrito.

- Ventajas
 - Sencillez en la utilización del mismo formato durante el proyecto, referencias cruzadas y numeración (Documento estructurado)
 - Calidad en la edición de texto y funciones matemáticas.
 - Multiplataforma.
 - Es una herramienta portable y gratuita.
- Inconvenientes
 - Es una herramienta compleja, ya que requiere un periodo previo a su utilización de aprendizaje de comandos básicos.
 - Dificultad en la introducción de imágenes y bibliografías.
 - Necesidad de compilación por cada cambio realizado para observar los resultados finales.

L^AT_EX será el sistema de edición de texto escogido, ya que, a pesar de la dificultad de su aprendizaje, hemos considerado que tiene mayor peso las ventajas que los inconvenientes propuestos.

4.5. Gestores de Referencias Bibliográficas

Mendeley

Mendeley es un software gratuito destinado a la gestión de referencias bibliográficas. Compatible con diferentes versiones de Microsoft Word, OpenOffice y BibTex tiene como característica principal la sincronización de las referencias

¹Sistema de tipografía desarrollado en 1985 por Donald E. Knuth, de licencia libre, es utilizado en entornos académicos por el alto número de funcionalidades que ofrece[44].

tanto en el equipo propio como en Web. Entre sus ventajas e inconvenientes, podemos destacar:

- Multiplataforma
- Organización de las referencias.
- Compartir referencias bibliográficas entre diferentes usuarios.

Zotero

Software libre de código abierto destinado a la gestión de referencias bibliográficas. [10] Utilizado en Firefox, Chrome, Safari y Opera, sigue cinco principios básicos:

1. Recopilar información.
2. Organizar los recursos en la biblioteca.
3. Citar las referencias bibliográficas de manera automática en la edición de texto.
4. Sincronización de la biblioteca en un servidor.
5. Colaboración al permitir compartir las bibliotecas creadas con el resto de usuarios.

■ Ventajas

- Multiplataforma
- Extracción automática de las citas en la edición de texto.
- Sincronización con diferentes editores de texto (Microsoft Word, OpenOffice y LibreOffice)[38]
- Importación de metadatos de un amplio número de soportes.
- Reconoce bibliotecas en formato BibTex.

■ Inconvenientes

- No dispone de lector de PDF.
- Necesidad de utilizar aplicaciones de terceros para obtener todas las funcionalidades disponibles.

BibTex

La herramienta elegida para la bibliografía es BibTex, diseñada como una utilidad de apoyo bibliográfico para LaTeX.

Para utilizarlo, se emplean los ficheros.bib en los que se encuentre la bibliografía necesaria (librerías), de forma que BibTex añadirá a la bibliografía las librerías que hayan sido citadas en el documento.

Además, para obtener los datos de la bibliografía, basta con abrir una pestaña de "school.google.es, marcar en Configuración la opción "Mostrar enlaces para importar citas a BibTex". De esta forma, una vez encontrada la página deseada, tan sólo es necesario copiar la bibliografía que se muestre y guardarla en nuestra librería.[25]

4.6. Editores de código

4.7. Recogida de datos

La recogida de datos se ha realizado mediante un cuestionario anónimo difundido entre los Alumnos que hayan cursado la carrera del Grado de Ingeniería Informática en Burgos. Entre las diferentes opciones que existían destacamos:

SurveyMonkey

Herramienta de creación de cuestionarios online con versión gratuita para realizar:

- 15 tipos de preguntas diferentes.
- 10 preguntas máximo por cuestionario.
- Un máximo de 100 respuestas.

Hemos descartado dicha herramienta, ya que en la versión gratuita no permite la exportación de datos en ningún soporte, además de la limitación del número de preguntas por cuestionario.

Zoho Survey

Herramienta de creación de cuestionarios online con una versión gratuita con las siguientes características:

- Número ilimitado de cuestionarios.
- 15 preguntas posibles por cuestionario.

- 150 respuestas posibles por cuestionario.

Sin embargo, hemos descartado dicha herramienta por la limitación del número de preguntas por cuestionario.

Eval & Go

Herramienta de creación de cuestionarios online con posibilidad de creación de una cuenta gratuita con las siguientes características:

- Número de encuestas ilimitadas por cuenta.
- Posibilidad de exportación de los datos recogidos en formato Excel o CSV.

En un principio consideramos utilizar dicha herramienta, sin embargo, tiene un máximo de 150 respuestas al mes, y al necesitar el mayor número de respuestas de alumnos, tendríamos esa limitación, por lo que no podríamos recoger el máximo número de respuestas.

TypeForm

Es una herramienta de creación de cuestionarios online anónimos, con una versión gratuita con las siguientes características:

- Número ilimitado de preguntas, respuestas y cuestionarios por cuenta.
- Posibilidad de utilizar alguna plantilla predefinida o desarrollar una propia.
- Formularios responsive (adaptados a los diferentes dispositivos con los que se interaccione)
- Exportación de los datos en formato Excel en Drive o Google+, de forma que se puedan ver los resultados en tiempo real.

El cuestionario realizado se encuentra en la siguiente dirección <https://clarapalacios.typeform.com/to/RQRRfY>. Es un cuestionario sencillo, en donde se da una explicación breve del funcionamiento del mismo, se solicitan las ponderaciones por cada asignatura (obligatorias para los 3 primeros cursos académicos y optativas para el 4º año). No se permite que mismo usuario responda más de una vez al cuestionario, ya que, a pesar de ser anónimo, se almacena un identificador para evitar que un mismo alumno pueda rellenar varias veces el cuestionario.

4.8. Integración de funcionalidades del cuestionario

Los datos se guardarán en un documento Excel-sincronizado en Drive- llamado **DatosTFG_SistemasRecomendacion**, almacenado en la siguiente dirección: https://docs.google.com/spreadsheets/d/1Dtu6HPKu_d0zToz2b16dhfz4k70luqAASYbve/edit?usp=sharing

En un principio, teníamos pensado utilizar la librería de Python OpenPyxl de forma que se nos permitirá leer y almacenar en una matriz los datos recogidos del Excel. Sin embargo, nos encontramos problemas con la lectura, por lo que hemos decidido utilizar el API de GoogleDrive, descargando un fichero .json para la obtener la clave privada y almacenarla en el ordenador, en el directorio en donde se encuentre el código en Python de la lectura de datos de cuestionario anónimo.

Además, serán necesarias dos librerías, **oauth2client** y **gspread**, pudiéndose instalar con un único código en cmd (Símbolo de Sistema de Windows) mediante el comando “**pip install gspread oauth2client**”. con la versión de pip 8.1.2, a pesar de no ser la más actual existente.

Funcionamiento del Google Drive API

La activación e integración del fichero situado en Drive para poder utilizarlo desde Jupyter se ha realizado de la siguiente manera:

- Iniciar sesión en Google Drive
- Activar en la consola de APIs de Google la API.
- Crear (o en caso de tener uno creado, abrir) un nuevo proyecto. Hay un máximo de 11 proyectos.
- Activar las credenciales de Google Drive.
- Crear credenciales para Servidor Web para acceder a la aplicación.
- Dar un nombre al proyecto y asignarle un rol (Project-Editor)
- Descargar el fichero .JSON correspondiente.
- Copiar el fichero (renombrado a client_secret.json) en el directorio en donde obtengamos los datos en python del Excel.
- Copiar el client_email situado en el fichero, y compartirlo con el fichero en Drive.
- Instalar ambas librerías, e importarlas en el Notebook de Python.

4.9. Herramienta para el desarrollo

Aspectos relevantes del desarrollo del proyecto

[4]

Trabajos relacionados

Los sistemas de recomendación son una herramienta muy utilizada actualmente, marcando un punto de inflexión tanto en el comercio electrónico como en las diferentes webs y apps, tales como Youtube, Spotify... Sin embargo, en relación a diferentes proyectos de Trabajo Final de Grado, éste es el primer proyecto orientado a dicho campo.

Entre las diferentes plataformas, podemos especificar:

6.1. Amazon

Amazon utiliza el sistema de recomendación "ítem to ítem collaborative filtering" inventado por la propia empresa en 1998 ante el problema de los filtros colaborativos basados en usuarios:

- Fallo en la recomendación al disponer de una gran cantidad de elementos y pocas votaciones
- Gran coste en el cálculo de las similitudes de usuarios.
- Cambio de perfil de usuarios de forma rápida.

Por otro lado, utiliza un software dirigido por un equipo de trabajadores encargado de enviar ofertas específicas a los diferentes usuarios según las últimas compras realizadas.

6.2. Spotify

Spotify utiliza 3 sistemas de recomendación diferentes para mejorar las recomendaciones de sus canciones en Discovery Weekly:

- Filtro colaborativo basado en usuarios.
- Modelo basado en el audio en bruto: Sistema de recomendación que utiliza una red neuronal (convolutional neural network) para obtener las características musicales de las canciones escogidas por el usuario.
- NLP Sistema de procesamiento de lenguaje natural para encontrar nuevas tendencias musicales con las que poder trabajar. [7]

6.3. Youtube

El sistema de recomendación de Youtube es bastante simple, ya que utiliza un filtro colaborativo basado en memoria (basado en usuarios). Sin embargo, lo filtra mediante calificaciones de forma que se mejora la ponderación del vídeo en los siguientes aspectos:

- Regionalización: Filtrando el FC en el área geográfica en donde se encuentre el usuario o en el lenguaje de las búsquedas previas realizadas.
- Temporización: El sistema de recomendación de Youtube favorece los vídeos con una visualización de mayor duración, sin importar el porcentaje del vídeo reproducido.
- Número de vídeos vistos por los usuarios a partir de uno determinado en una misma sesión.
- Determina si un vídeo gusta o no a los clientes utilizando inteligencia artificial para observar los comentarios inscritos en el vídeo.

Conclusiones y Líneas de trabajo futuras

Asignaturas de Ingeniería Informática

Las asignaturas del Grado de Ingeniería Informática en Burgos se subdividen en dos categorías dependiendo de su posibilidad de elección o no.

- Asignaturas Obligatorias.
- Asignaturas Optativas.

8.1. Asignaturas Obligatorias

Las asignaturas obligatorias son aquellas que se cursan en los tres primeros años de la carrera. En su conjunto forman 180 créditos, a seis créditos cada asignatura.

- Primer Semestre
 1. Fundamentos Deontológicos y Jurídicos de las TIC
Conocimiento y estudio de la legislación española y su aplicación a las Tecnologías de la Información y la Comunicación, así como la responsabilidad ética y moral de las mismas.
 2. Álgebra Lineal
Conocimiento y aprendizaje de la resolución de diferentes problemas matemáticos de Álgebra Lineal.
 3. Informática Básica
Aprendizaje de los conocimientos básicos del funcionamiento lógico de las computadoras, así como la aplicación de la Matemática en los mismos.

Cuadro 8.2: Tabla de asignaturas del Primer Curso Académico

Example	Orientación	Programación?	Lenguaje de Programación
Fundamentos Deontológicos y Jurídicos de las TIC	Derecho	No	-
Álgebra Lineal	Matemática	No	-
Informática Básica	Ofimática	Sí	JavaScript
Fundamentos físicos de la Informática	Física	No	-
Matemática discreta	Matemática	No	-
Inglés Aplicado a la Informática	Inglés	No	-
Cálculo	Matemática	No	-
Programación	Programación estructurada	Sí	C
Fundamentos de los Computadores	Programación en ensamblador	Sí	Ensamblador
Sistemas Operativos	Uso avanzado de equipos	Sí	Bash

4. Fundamentos físicos de la Informática

Conocimiento básico de resolución de problemas físicos (electromagnetismo, circuitos eléctricos, principio físico de semiconductores y dispositivos electrónicos)

5. Matemática discreta

Conocimiento de conceptos básicos de la lógica matemática y algoritmia, así como la resolución de los problemas aplicados a la Informática.

■ Segundo Semestre

6. Inglés Aplicado a la Informática

Aprendizaje del vocabulario, expresiones y síntesis de la lengua Inglesa orientada a la informática.

7. Cálculo

Aprendizaje y resolución de problemas orientados a los conceptos básicos de la matemática, lógica, algorítmica.

8. Programación

Aprendizaje del lenguaje de programación C. así como la introducción a la programación estructurada.

9. Fundamentos de los Computadores

Aprendizaje del funcionamiento básico de los computadores y la programación de los mismos en lenguaje ensamblador.

10. Sistemas Operativos

Aprendizaje de la arquitectura informática centralizada o distribuida, gestión, desarrollo y mantenimiento de sistemas informáticos.

La siguiente tabla corresponde a las asignaturas del Primer Curso 8.2

■ Tercer Semestre

1. Metodología de la Programación
Aprendizaje de diferentes conceptos tales como modularidad, lenguaje orientado a objetos, conceptos estáticos y dinámicos tales como clases, objetos, herencia, genericidad y resolución de problemas acerca de la robustez de un programa.
2. Estadística
Aprendizaje de medición y la resolución de problemas matemáticos orientados al álgebra lineal, cálculo diferencial e integral estadística y optimización.
3. Ingeniería del Software
Aprendizaje de la metodología del desarrollo software, resolución del modelaje de funciones, utilización de la herramienta CASE y comprensión de la teoría de validación software, aplicándolo al desarrollo de aplicaciones.
4. Bases de Datos
Aprendizaje de la gestión y procesamiento del almacenamiento y acceso a los Sistemas de Información, incluyendo los sistemas basados en Web.
5. Arquitectura de Computadores
Aprendizaje del funcionamiento de una computadora, comprendiendo las unidades funcionales básicas de los computadores, así como el rendimiento de dichas unidades.

■ Cuarto Semestre

6. Estructuras de Datos
Aprendizaje del diseño, desarrollo y evaluación de los sistemas informáticos, calculando la eficiencia algorítmica del acceso a los datos, diferenciando los tipos abstractos de datos y diferenciando el tipo de datos con su abstracción.
7. Redes
Aprendizaje de la configuración de redes locales, VLANs, enrutamiento IPv4 y IPv6, los diferentes protocolos de transporte de la información y la configuración de servicios de red (DHCP, DNS, SMTP, POP Y HTTP)
8. Interacción Hombre-Máquina
Aprendizaje de las técnicas, bases y soluciones de desarrollo de interfaces teniendo en cuenta la usabilidad ofrecida.
9. Fundamentos de Organización y Gestión de Empresas
Aprendizaje de la economía de la empresa, la naturaleza, el propósito, la problemática y el desarrollo de la misma, centrándose en las finanzas y el marketing, función directiva y de recursos humanos.

Cuadro 8.3: Tabla de asignaturas del Segundo Curso Académico

Example	Orientación	Programación?	Lenguaje de Programación
Metodología de la Programación	Informática	Si	Java
Estadística	Matemática	No	-
Ingeniería del Software	Informática	No	-
Bases de Datos	Informática	Sí	SQL
Arquitectura de Computadores	Informática	Sí	C++, Cuda
Estructuras de Datos	Informática	Sí	Java
Redes	Informática	Sí	Matlab
Interacción Hombre-Máquina	Informática	Sí	C#, PHP
Fundamentos de Organización y Gestión de Empresas	Empresa	No	-
Análisis y Diseño de Sistemas	Informática	No	-

10. Análisis y Diseño de Sistemas

Aprendizaje del desarrollo de los sistemas software, aportando calidad en los diferentes requisitos, subdividiendo el aprendizaje de la gestión de proyectos Software en Dirección, organización, planificación y gestión de proyectos informáticos de forma que el alumno sea capaz de seleccionar y aplicar los patrones de diseño en la construcción de la aplicación.

La siguiente tabla corresponde a las asignaturas del Segundo Curso 8.3

■ Quinto Semestre

1. Arquitecturas Paralelas

Aprendizaje del diseño y construcción de técnicas de programación paralela orientadas a aplicaciones de programación paralela, arquitecturas multinúcleo, memoria compartida y distribuida.

2. Sistemas Inteligentes

Aprendizaje del lenguaje lógico de proposiciones y diferentes modelos y algoritmos de búsqueda para la resolución de problemas basados en espacios de estados.

3. Gestión de Proyectos

Aprendizaje de la dirección, organización, gestión y control de proyectos, riesgos y recursos humanos.

4. Diseño y Administración de Sistemas y Redes

Aprendizaje de la administración de servicios y recursos en sistemas operativos y redes de computadores, optimizando los sistemas, analizando la calidad de los mismos.

5. Procesadores del Lenguaje

Aprendizaje de atributos (heredados y sintetizados) orientados a la sintaxis XML y diseño de compiladores (Analizadores sintácticos ascendentes y descendentes).

Cuadro 8.4: Tabla de asignaturas del Tercer Curso Académico

Example	Orientación	Programación?	Lenguaje de Programación
Arquitecturas Paralelas	Informática	Si	C++
Sistemas Inteligentes	Informática	Sí	Python
Gestión de Proyectos	Informática	No	-
Diseño y Administración de Sistemas y Redes	Informática	Sí	Bash
Procesadores del Lenguaje	Informática	Sí	Flex, Bison
Programación Concurrente	Informática	Sí	Java, Ada
Seguridad Informática	Informática	No	-
Aplicaciones de Bases de Datos	Informática	Sí	Java, SQL
Algoritmia	Informática	Sí	Python
Métodos Numéricos y Optimización	Informática	Sí	Matlab

■ Sexto Semestre

6. Programación Concurrente

Aprendizaje y diseño de aplicaciones con diferentes técnicas de sincronización, tiempo real y aplicación de los problemas de programación concurrente a dichas aplicaciones.

7. Seguridad Informática

Aprendizaje de la problemática de la seguridad en los diferentes sistemas, su detección y resolución. Además, se estudiarán los planes de contingencia, su estructura y el ámbito legal en el que se relacionan.

8. Aplicaciones de Bases de Datos

Aprendizaje de las arquitecturas de las bases de datos, las API para integrar comandos de SQL, transacciones y diferentes niveles de aislamiento y aplicar dichos conocimientos en un lenguaje de programación conocido (Java).

9. Algoritmia

Aprendizaje y resolución de problemas de complejidad de diferentes algoritmos (incluyendo algoritmos recursivos), diferenciar y seleccionar el algoritmo que mejor convenga para un problema determinado e implementar el esquema junto con el propio algoritmo.

10. Métodos Numéricos y Optimización

Aprendizaje y resolución de problemas de optimización numérica (discreta y continua) así como la resolución de ecuaciones, sistemas lineales, interpolación global y segmentaria.

La siguiente tabla corresponde a las asignaturas del Tercer Curso 8.4

8.2. Asignaturas Optativas

Las asignaturas optativas son todas aquellas que el alumno tiene posibilidad de elección en el Cuarto curso académico. Constan de un máximo de 48 créditos, pudiendo conmutar dos asignaturas por Prácticas curriculares.

■ Séptimo Semestre

1. Diseño e Implementación de Sistemas Digitales
Aprendizaje y diseño de la programación de circuitos electrónicos digitales aplicando el diseño a un microprocesador sencillo.
2. Gestión de la Información
Aprendizaje y desarrollo de los sistemas de información, aplicar pequeñas aplicaciones de Big Data (sistemas de recomendación) y diferentes marcos de problemas de gestión de información.
3. Diseño y Mantenimiento del Software
Aprendizaje y desarrollo de diferentes patrones de diseño orientados a la aplicación software evaluando en qué momento es necesaria su aplicación.
4. Organización y Gestión de Empresas
Aprendizaje de los principios de los mercados centrándose en el área cinanciera y los subsistemas productivos.
5. Mantenimiento de Equipos Informáticos
Aprendizaje de los diferentes elementos de una computadora, su misión y su funcionamiento, así como el cálculo de su rendimiento, calidad, disponibilidad, fiabilidad y escalabilidad.
6. Hardware de Aplicación Específica
Aprendizaje del tratamiento de imágenes y vídeos utilizando Matlab.
7. Control por Computador
Aprendizaje, diseño e implementación de los diferentes sistemas de control y la comprensión de los sistemas dinámicos.
8. Validación y Pruebas
Aprendizaje del desarrollo, mantenimiento, evaluación y solución de problemas de servicios y sistemas software, evaluando la usabilidad, ergonomía, accesibilidad y seguridad de los sistemas.
9. Computación Neuronal y Evolutiva
Aprendizaje de las diferentes redes neuronales, su evolución y aplicación, además de la programación y el entrenamietno de las mismas. Aprendizaje de los algoritmos evolutivos, sus variantes y modelos empleados en la actualidad.

10. Programación de Sistemas Operativos

Aprendizaje de la estructura de un sistema operativo, su funcionamiento y el análisis para mejorar su eficiencia lo máximo posible orientado a la multitarea, multiproceso y el tiempo de implantación del mismo.

■ Octavo Semestre

1. Sistemas Distribuidos

Aprendizaje de la estructura y funcionalidades de los sistemas distribuidos e implementar soluciones a diferentes problemas de los sistemas distribuidos en el ámbito de Internet.

2. Sistemas Empotrados y de Tiempo Real

Aprendizaje de la arquitectura hardware, su funcionamiento, la programación de los mismos de forma que se permita programar algoritmos de control digital y la programación de microcontroladores.

3. Métodos Formales

Aprendizaje de los fundamentos de métodos formales, la estandarización de los lenguajes y la obtención de código ejecutable a partir de las especificaciones.

4. Nuevas Tecnologías y Empresa

Aprendizaje del cambio de modelo en el negocio actual en base a las nuevas tecnologías, y la relación de las mismas con el intercambio de información, productos y servicios con las empresas y personas.

5. Minería de Datos

Aprendizaje de los conocimientos de la minería de datos e implementar métodos para el análisis de datos.

6. Desarrollo Avanzado de Sistemas Software

aprendizaje y de los diferentes estándares para la gestión de calidad del software, definiendo métricas para su evaluación, y reestructurar programas mediante la refactorización.

La siguiente tabla corresponde a las asignaturas del Cuarto Curso 8.5

Cuadro 8.5: Tabla de asignaturas del Cuarto Curso Académico

Example	Orientación	Programación?	Lenguaje de Programación
Diseño e Implementación de Sistemas Digitales	Informática	Si	HDL
Gestión de la Información	Informática	Sí	Python
Diseño y Mantenimiento del Software	Informática	Sí	Java
Organización y Gestión de Empresas	Empresa	Sí	Matlab
Mantenimiento de Equipos Informáticos	Informática	No	-
Hardware de Aplicación Específica	Informática	Sí	Matlab
Control por Computador	Informática	No	
Validación y Pruebas	Informática	Sí	
Computación Neuronal y Evolutiva	Informática	Sí	Matlab, Java
Programación de Sistemas Operativos	Informática	Sí	C++, C
Sistemas Distribuidos	Informática	Sí	Java
Sistemas Embebidos y de Tiempo Real	Informática	Sí	
Métodos Formales	Informática	Sí	
Nuevas Tecnologías y Empresa	Informática	Sí	Python
Minería de Datos	Informática	Sí	Java, Python
Desarrollo Avanzado de Sistemas Software	Informática	Sí	Java

Bibliografía

- [1] Esaú A. Sourceforge vs github vs bitbucket vs gitlab—open webinars, 2016. [Internet; descargado 02-febrero-2018].
- [2] Proyectos Agiles. Beneficios de scrum — proyectosagiles.org, 2018. [Internet; descargado 28-enero-2018].
- [3] Greg Baugues. Google spreadsheets and python, 2017. [Internet; descargado 8-marzo-2018].
- [4] Zachary J Bortolot and Randolph H Wynne. Estimating forest biomass using small footprint lidar data: An individual tree-based approach that incorporates training data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(6):342–360, 2005.
- [5] Lucho CAstillo. Conociendo-github, 2012. [Internet; descargado 4-abril-2018].
- [6] Nick Davis. Las desventajas de openoffice, 2018. [Internet; descargado 05-febrero-2018].
- [7] Isaac Gonzalez Díaz. Caso spotify: cómo usa advanced analytics para recomendar canciones, 2018. [Internet; descargado 02-mayo-2018].
- [8] Universidad de Burgos. Guías docentes, 2018. [Internet; descargado 26-febrero-2018].
- [9] Universidad de Chile. Gestor de citas bibliográficas mendeley, 2018. [Internet; descargado 10-febrero-2018].
- [10] Universidad de Salamanca. Zotero — gestores bibliográficos, 2018. [Internet; descargado 07-febrero-2018].

- [11] Marketing Directo. Analizamos el sistema de recomendaciones de youtube, ¿es una herramienta demasiado simplona?, 2013. [Internet; descargado 02-mayo-2018].
- [12] Lorenzo Hidalgo (Universidad de Granada) Enrique Herrera-Viedma (Universidad de Granada), Carlos Porcel (Universidad de Jaén). Sistemas de recomendaciones: herramientas para el filtrado de información en internet. Technical report, Universitat Pompeu Fabra, 2004. [Internet; descargado 12-febrero-2018].
- [13] etsamadrid. ¿qué ventajas e inconvenientes tiene tex/latex?, 2018. [Internet; descargado 05-febrero-2018].
- [14] Andrés González. Sistemas de recomendación de contenido con Machine Learning, September 2014.
- [15] Almudena Ruiz Iniesta. Sistemas de recomendación, presente y futuro de la web. Technical report, Universidad Complutense de Madrid, 2018. [Internet; descargado 10-febrero-2018].
- [16] José Manuel Galán José Ignacio Santos. Sistemas de recomendación: filtro colaborativo basado en memoria. Technical report, Universidad de Burgos, 2018. [Internet; descargado 15-abril-2018].
- [17] Manu Mateos. Así funcionan las recomendaciones de amazon, 2014. [Internet; descargado 05-mayo-2018].
- [18] Ric Mazereeuw. ¿cómo funciona el algoritmo de youtube? guía para conseguir más vistas, 2017. [Internet; descargado 02-mayo-2018].
- [19] Martin McNulty. Cambio en el sistema de recomendación de vídeos de youtube, 2012. [Internet; descargado 02-mayo-2018].
- [20] Mikenieva. Gitlab: el github para tus propios servidores, 2015. [Internet; descargado 03-febrero-2018].
- [21] Finances Online. Versionone revisiÓn, software de gestión de proyectos — financesonline.org, 2018. [Internet; descargado 29-enero-2018].
- [22] Apache OpenOffice. Por qué apache openoffice — apache openoffice, 2018. [Internet; descargado 05-febrero-2018].
- [23] Jose Juan Perez. Cómo ayuda microsoft word en la vida — elblogdel1, 2015. [Internet; descargado 05-febrero-2018].
- [24] Monografias Plus. Ventajas y deventajas de microsoft word y excel — monografias plus, 2018. [Internet; descargado 05-febrero-2018].

- [25] Miguel Mata Pérez. Bibliografía en latex. *Una guía concisa de BibTEX*, 1968.
- [26] Elena Queralt. ¿qué es el método scrum y que beneficios y ventajas va a aportar a mi negocio? — iebs school, 2014. [Internet; descargado 28-enero-2018].
- [27] Erlinis Quintana. 10 razones para usar github, 2015. [Internet; descargado 7-abril-2018].
- [28] RocReguant. Github vs bitbucket: Que control de versiones git usar?, 2014. [Internet; descargado 01-febrero-2018].
- [29] Carlos Guerra Terol. 10 herramientas gratuitas para hacer encuestas online, 2016. [Internet; descargado 26-febrero-2018].
- [30] Kanban Tool. ¿por qué utilizar la metodología kanban? — método básico y principios kanban para ayudarle a empezar, 2009. [Internet; descargado 26-enero-2018].
- [31] Jacob Rodríguez Torres. Guía de trello – gestor de tareas con kanban, 2012. [Internet; descargado 4-abril-2018].
- [32] Rubén Velasco. Bitbucket, una alternativa al conocido controlador de versiones github, 2014. [Internet; descargado 01-febrero-2018].
- [33] Wikipedia. Bitbucket — wikipedia, la enciclopedia libre, 2017. [Internet; descargado 29-enero-2018].
- [34] Wikipedia. Desarrollo ágil de software — wikipedia, la enciclopedia libre, 2017. [Internet; descargado 25-enero-2018].
- [35] Wikipedia. Filtrado colaborativo — wikipedia, la enciclopedia libre, Nov 2017. [Internet; descargado 09-febrero-2018].
- [36] Wikipedia. Licencia mit — wikipedia, la enciclopedia libre, 2017. [Internet; descargado 01-febrero-2018].
- [37] Wikipedia. Sistema de recomendación — wikipedia, la enciclopedia libre, Apr 2017. [Internet; descargado 10-febrero-2018].
- [38] Wikipedia. Zotero — wikipedia, la enciclopedia libre, Oct 2017. [Internet; descargado 07-febrero-2018].
- [39] Wikipedia. Gitlab — wikipedia, la enciclopedia libre, 2018. [Internet; descargado 03-febrero-2018].
- [40] Wikipedia. Item-item collaborative filtering — wikipedia, la enciclopedia libre, 2018. [Internet; descargado 05-mayo-2018].

- [41] Wikipedia. Kanban — wikipedia, la enciclopedia libre, 2018. [Internet; descargado 26-enero-2018].
- [42] Wikipedia. Microsoft word — wikipedia, la enciclopedia libre, 2018. [Internet; descargado 04-febrero-2018].
- [43] Wikipedia. Sourceforge — wikipedia, la enciclopedia libre, feb 2018. [Internet; descargado 05-febrero-2018].
- [44] Wikipedia. Tex — wikipedia, la enciclopedia libre, 2018. [Internet; descargado 05-febrero-2018].