UNIVERSIDAD SAN PABLO - CEU

ESCUELA POLITÉCNICA SUPERIOR

GRADO EN INGENIERÍA DE SISTEMAS DE INFORMACIÓN



TRABAJO FIN DE GRADO

**JOBS : diseño e implementación de una aplicación RESTful para búsqueda de empleo para estudiantes/graduados**

Autor: Carlos Soler Martín

Tutores: Sergio Saugar García (U. San Pablo)

Alex Rayón Jerez (U. Deusto)

Junio 2022

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Datos del alumno   |  | | --- | | Nombre: Carlos Soler Martín |   Datos del Trabajo   |  | | --- | | TÍTULO DEL PROYECTO: JOBS - diseño e implementación de una aplicación RESTful para búsqueda de empleo para estudiantes/graduados |   Tribunal calificador   |  |  | | --- | --- | | Presidente: | Fdo.: |  |  |  | | --- | --- | | Secretario: | Fdo.: |  |  |  | | --- | --- | | Vocal: | Fdo.: |  |  | | --- | | Reunido este tribunal el \_\_\_ /\_\_\_ /\_\_\_\_\_\_\_, acuerda otorgar al Trabajo Fin de Grado presentado por D./Dña. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ la calificación de \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | |

Resumen

Texto con, a lo sumo, 200 palabras.

Palabras Clave

Abstract

Resumen anterior, en inglés.

Keywords

Índice de contenidos

[Capítulo 1 Introducción 1](#_Toc106906157)

[Capítulo 2 Estado del arte 5](#_Toc106906158)

[2.1 Servicios Web 5](#_Toc106906159)

[2.2 Algoritmo de recomendación 5](#_Toc106906160)

[2.2.1 Introducción a la ciencia de los datos [1] 5](#_Toc106906161)

[2.2.2 Machine Learning: métodos de aprendizaje [2] 6](#_Toc106906162)

[2.2.3 Técnicas de análisis de datos prescriptivas: sistemas de recomendación [4] 8](#_Toc106906163)

[2.2.4 Medidas de similitud o distancia [6] 10](#_Toc106906164)

[2.2.5 Lenguaje utilizado: R 11](#_Toc106906165)

[Capítulo 3 Gestión del Proyecto 13](#_Toc106906166)

[3.1 Modelo de ciclo de vida 13](#_Toc106906167)

[3.2 Papeles desempeñados en el proyecto 14](#_Toc106906168)

[3.3 Planificación 14](#_Toc106906169)

[3.4 Presupuesto 15](#_Toc106906170)

[3.5 Ejecución 15](#_Toc106906171)

[Capítulo 4 Análisis y arquitectura 17](#_Toc106906172)

[4.1 Análisis de dominio 17](#_Toc106906173)

[4.2 Especificación de requisitos 17](#_Toc106906174)

[4.3 Análisis de los casos de uso y de las clases de análisis 19](#_Toc106906175)

[4.4 Análisis de seguridad 20](#_Toc106906176)

[4.5 Análisis desde la perspectiva del RGPD (si procede) 20](#_Toc106906177)

[4.6 Arquitectura del sistema 20](#_Toc106906178)

[Capítulo 5 Diseño 21](#_Toc106906179)

[5.1 Diseño del Backend 21](#_Toc106906180)

[5.1.1 Diseño del Servicio Web RESTful 21](#_Toc106906181)

[5.1.2 Diseño del algoritmo de recomendación 39](#_Toc106906182)

[5.1.3 Diseño de la base de datos 39](#_Toc106906183)

[5.2 Diseño del Frontend 40](#_Toc106906184)

[5.2.1 Diseño de la interfaz de usuario 40](#_Toc106906185)

[5.3 Modelo de clases de diseño 40](#_Toc106906186)

[5.4 Entorno de construcción 40](#_Toc106906187)

[5.5 Plan de pruebas 40](#_Toc106906188)

[5.6 Diagrama de infraestructuras de nivel 3 41](#_Toc106906189)

[5.7 Diagrama de infraestructuras de nivel 2 (si procede) 41](#_Toc106906190)

[Capítulo 6 Implementación 42](#_Toc106906191)

[6.1 Implementación Backend 42](#_Toc106906192)

[6.1.1 Implementación Servicio REST 42](#_Toc106906193)

[6.1.2 Implementación Algoritmo recomendación 42](#_Toc106906194)

[6.1.3 Implementación Base de datos 42](#_Toc106906195)

[6.2 Implementación Frontend 42](#_Toc106906196)

[6.2.1 XXX 42](#_Toc106906197)

[6.2.2 XXX 42](#_Toc106906198)

[Capítulo 7 Construcción 43](#_Toc106906199)

[7.1 Referencia al repositorio de software 43](#_Toc106906200)

[7.2 Manuales 43](#_Toc106906201)

[Capítulo 8 Algoritmo de Recomendación 44](#_Toc106906202)

[8.1 Sistema de recomendación de JOBS 44](#_Toc106906203)

[8.1.1 Estructura del sistema de recomendación 44](#_Toc106906204)

[8.1.2 API “ofertas\_nuevas.R” 44](#_Toc106906205)

[8.1.3 API “alumno\_nuevo.R” 46](#_Toc106906206)

[8.1.4 Llamada de las APIs 47](#_Toc106906207)

[8.1.5 Intercambio de ficheros Web-APIs 47](#_Toc106906208)

[Capítulo 9 Conclusiones y líneas futuras 49](#_Toc106906209)

[Referencias 51](#_Toc106906210)

[Anexo I 52](#_Toc106906211)

[Otras posibilidades para realizar el análisis y el diseño 52](#_Toc106906212)

[Anexo II 53](#_Toc106906213)

[Fuente de inspiración del presente documento 53](#_Toc106906214)

[Glosario de términos 55](#_Toc106906215)

[Anexos 57](#_Toc106906216)

Índice de ilustraciones

Ilustración 1

Ilustración 2

Índice de tablas

Tabla 1

Tabla 2

# Introducción

Se explica en qué consiste el proyecto y se enumeran los capítulos. Debe incluir sus objetivos.

Desde el inicio de la pandemia, la situación de desempleo en España ha ido a peor y esto perjudica a todos los jóvenes estudiantes a la hora de encontrar su primer empleo o las prácticas que son necesarias para poder conseguir el título universitario. Este proyecto tiene como objetivo contribuir a solventar la dificultad que tienen los graduados universitarios españoles para encontrar un primer empleo. Algunos datos que explican esta problemática:

* Según un estudio realizado por Uniplaces, el 82% de adolescentes españoles se plantean ir a trabajar fuera de España, un porcentaje significativamente mayos comparado con el 63% de los franceses y el 66% de los alemanes.
* “Frustración, miedo e incertidumbre” son las mejores palabras para definir cómo se siente un adolescente que al acabar su grado o máster y no tener perspectiva de encontrar trabajo. Una de las principales razones por la que miles de jóvenes ‘hiperformados’ no optan a un puesto de trabajo es la falta de experiencia. ¿Cómo puedo empezar a trabajar si casi todas las empresas me piden experiencia? ¿Y las empresas que no me piden experiencia como las encuentro? Al final muchos jóvenes se conforman con un trabajo que les permita mantenerse y no consiguen trabajar en algo que realmente les atraiga.
* Cómo consecuencia de la pandemia los universitarios han tenido dificultades para encontrar prácticas curriculares. Miles de estudiantes o no han podido encontrar prácticas o se les han retrasado, por tanto, no han podido terminar el grado hasta que contarán con todos los créditos de las prácticas. El proceso de encontrar trabajo/practicas puede producir una sensación de desesperación/depresión para el estudiante.

El departamento de carreras profesionales de la Universidad CEU San Pablo es el encargado de ayudar al estudiante y graduado a encontrar practicas curriculares/primer trabajo. Cuentan con 5 aplicaciones para la búsqueda de empleo a nivel nacional e internacional. De esas 5 aplicaciones solamente se hace el uso de dos, que son:

* GetHighered: aplicación para buscar trabajo a nivel internacional
* JobTeaser: aplicación para buscar trabajo a nivel nacional.

La funcionalidad en ambas aplicaciones es la misma: el estudiante crea su perfil y adjunta su CV. Una vez realizado lo anterior, el estudiante debe de buscar de manera manual las empresas a las que le gustaría aplicar, una vez encuentra una empresa de trabajo que le interesa debe de enviar a través de la aplicación una pequeña carta de motivación o a veces la empresa te redirige a la sección de recruiting de su web. Por parte de las empresas también deben de registrarse en la aplicación y publicar su oferta de trabajo, o normalmente el equipo de carreras profesionales reciben ofertas de las empresas por correo y las postean en las plataformas de empleo. Cuando son las propias empresas las que publican la oferta de empleo es necesario que alguien de carreras profesionales valide la solicitud.

También cuentan con un portal de prácticas para el alumnado dentro del campus virtual, en el cual se puede: subir el CV del alumno, ver una lista de ofertas de trabajo y poder ponerse en contacto con la empresa, firmar el anexo de las practicas, realizar una memoria sobre las practicas realizadas, realizar un seguimiento del trámite de las prácticas y realizar una encuesta.

La situación descrita es común en el sistema universitario español y presenta una serie de ineficiencias que este proyecto trata de paliar:

* No existe una base de datos integrada de alumnos universitarios que buscan prácticas o un primer empleo.
* Tampoco existe un sistema de información eficiente de las prácticas o primeros empleos ofertados por las empresas. Las relaciones universidad-empresa son bilaterales y no son eficientes.
* Los departamentos de carreras profesionales de las universidades realizan de forma manual el proceso de matching entre alumnos (en base CV) y ofertas de prácticas/primeros empleos.
* Los departamentos de carreras profesionales de las universidades carecen de aplicaciones que les permitan realizar el matching anteriormente mencionado de forma automática, así cómo gestionar de forma eficiente todo el proceso administrativo asociado (contratos, anexos, memorias, etc.).

La plataforma JOBS ayudará a solventar alguna de las deficiencias anteriormente mencionadas ya que tiene como objetivo:

* Constituirse en una base de datos unificada de CVs de alumnos y ofertas de prácticas/primeros empleos.
* Homogeneizar los CVs de los alumnos a través de un modelo de estandarizado de capacidades (“skills”)
* Facilitar un matching automático entre alumnos (en base CV) y ofertas de prácticas/primeros empleos.
* Desarrollar un algoritmo de recomendación que permita dotar a la plataforma de una funcionalidad que recomiende a los alumnos aplicar a puestos de trabajo/prácticas que sean apropiados con su CV. Cuando un nuevo alumno se da de alta en la plataforma e introduce su CV (resumido en la valoración de una serie de cualidades o “skills”) el algoritmo calcula la similitud del CV del nuevo alumno con el CV de alumnos que ya están en la base de datos y los cuales han aplicado con éxito a puestos de trabajo/prácticas. El algoritmo recomienda al nuevo alumno puestos de trabajo/prácticas similares a los que han optado alumnos que tienen un CV similar.

# Estado del arte

## Servicios Web

Los servicios web son una tecnología que facilita un servicio a través de internet. Hacen referencia al conjunto de protocolos, procesos, estándares entre dos máquinas o aplicaciones para comunicarse entre sí. Los servicios web cuentan con dos características fundamentales:

* Multiplataforma: El cliente y servidor no tienen por qué contar con la misma configuración del sistema, lenguaje soportado o plataforma.
* Distribuida: Un servicio web no está dirigido para un único cliente, sino que son varios los que pueden acceder al mismo servicio.

Estos servicios con una arquitectura cliente/servidor, en la que se puede identificar al servidor como las maquinas que almacenan la información y los recursos, y al cliente cualquier dispositivo que se conecte.

Existen dos tipos de servicios web:

* SOAP (Simple Object Access Protocol): Es un protocolo estándar de mensajes basado en XML que permite la comunicación para intercambiar información entre diferentes máquinas. Al ser un protocolo, cuenta con reglas que aumentan la complejidad y sobrecarga, lo cual perjudica el tiempo que tardan las páginas en cargarse.

El envio de la solicitud de datos se basa en HTTP, mientras que los mensajes SOAP que se reciben necesitan ser únicamente documentos XML.

* REST (Representacional State Transfer): Es un estilo de arquitectura software que facilita el intercambio de recursos basados en HTTP. Representa un conjunto de principios que permiten crear, leer, actualizar y eliminar los datos.

Una de las principales características de REST es que es *stateless*, es decir, no almacena los estados, debe de ser el cliente el que introduce el estado en cada llamada realizada.

El modelo de Richardson propone una clasificación de los servicios web mediante el nivel de madurez de cada uno:

* + Nivel 0: The Swamp of POX

Es el más básico. Se utiliza para servicios con un único recurso y método HTTP.

Escala de tiempo

Descripción generada automáticamente con confianza baja

*Ilustración 1: Nivel 0 del Modelo Madurez Richardson*

* + Nivel 1: Resources

Distingue entre diferentes recursos, pero con un único método HTTP.

Gráfico

Descripción generada automáticamente con confianza baja

*Ilustración 2: Nivel 1 del Modelo Madurez Richardson*

* + Nivel 2: HTTP Verbs

Distingue entre diferentes recursos y diferentes métodos HTTP.

Imagen que contiene Escala de tiempo

Descripción generada automáticamente

*Ilustración 3: Nivel 2 del Modelo Madurez Richardson*

* + Nivel 3: Hypermedia Controls

Hace uso de HATEOAS (Hypermedia As The Engine Of Application State). Al hacer un request nos devolverá información adicional (link) sobre otro recurso.

Escala de tiempo

Descripción generada automáticamente

*Ilustración 4: Nivel 3 del Modelo Madurez Richardson*

## Algoritmo de recomendación

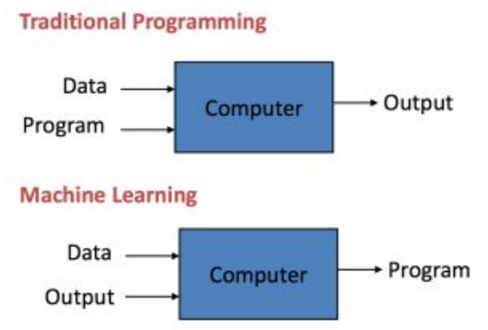
### Introducción a la ciencia de los datos [1]

La automatización de tareas y la capacidad de extraer valor de los datos es el presente y el futuro cercano de cualquier tipo de actividad/función/negocio que implique análisis y toma de decisiones. Otras eras de la historia de la humanidad se han caracterizado por manipular materias primas físicas: cobre, hierro, bronce, etc. Nosotros nos ha tocado vivir la “era de los datos”, y es el dato el nuevo material del que tenemos que extraer valor.

La ciencia de datos es un campo apasionante que combina habilidades estadísticas y cuantitativas avanzadas con programación con el objetivo de sacar valor a la inmensa cantidad de datos con los que nos encontramos en nuestro entorno académico o profesional, que es a lo que nos referimos cuando hablamos de “Big Data”. En general estas técnicas de análisis y explotación de datos es lo que popularmente se ha llamado “Inteligencia Artificial” y, más en concreto, “Machine Learning”, que son algoritmos con la habilidad de aprender sin haber sido explícitamente programados.

### Machine Learning: métodos de aprendizaje [2]

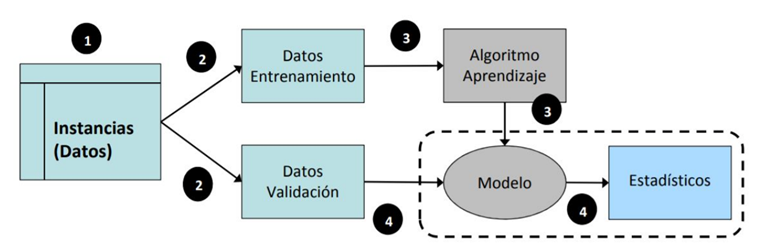
Machine Learning son un conjunto de métodos capaces de detectar automáticamente patrones en los datos con el objetivo de definir modelos que permitan representar de una forma simplificada la realidad. Machine Learning implica definir algoritmos que se encarga de que el modelo aprenda de los datos (se ajuste a los datos), es decir, el algoritmo se encarga de crear una función matemática para representar la realidad a partir de los datos. La diferencia fundamental entre la programación tradicional y la de Machine Learning radica en los elementos de entrada y de salida que tiene un modelo y el otro:



*Ilustración 2: diferencia entre la programación tradicional y el ML*

Los métodos de aprendizaje [3] implican la consecución de una serie de pasos:

* Los datos iniciales son preprocesados, se presentan al sistema de aprendizaje.
* Los datos se dividen en dos grupos. Los datos de entrenamiento servirán para generar el modelo. Los datos de validación servirán para valorar la calidad del modelo.
* El algoritmo de aprendizaje, genera el patrón que mejor se adecua a los datos.
* Se valora la calidad del modelo utilizando los datos validación, para lo cual se utilizan diferentes estadísticos/métricas.



*Ilustración 3: pasos de un modelo de aprendizaje*

Las técnicas y algoritmos de Machine Learning analizan datos con dos objetivos:

* Predecir: intentan responder a la pregunta, ¿Qué podría ocurrir? Implica ser capaz de predecir qué valores de salida se darán para unos valores de entrada determinados.
* Informar: intentan responder a la pregunta, ¿Qué ha ocurrido? Implica extraer información acerca de cómo el modelo, para unas variables de entrada, produce unas variables de salida.

En función del objetivo que persiguen hay dos familias de métodos de aprendizaje:

* Métodos de aprendizaje predictivos o supervisados: el objetivo es estimar una variable a futuro, en función del tipo de variable distinguimos entre:
  + Clasificación: la variable es categórica pudiendo ser binaria (ej. sí/no, paga/no paga…), múltiple (ej. compra a, b o c) o ordenada (ej. riesgo alto, medio, bajo)
  + Regresión: la variable es numérica (ej. precio, cantidad, tiempo…)
* Métodos de aprendizaje descriptivos o no supervisados: el objetivo es explorar los datos
  + Agrupamiento o “Clustering”: el objetivo es agrupar los datos en conjuntos (“clusters”) que se parezcan dentro de un mismo conjunto y que se diferencien de otros conjuntos (Ej. clientes con hábitos de compra similares)
  + Reglas de asociación: buscan reglas (reglas de co-ocurrencia) que describen la mayor parte posible de los datos (ej. productos que se compran juntos)
  + Análisis de Componentes Principales (ACP)

### Técnicas de análisis de datos prescriptivas: sistemas de recomendación [4]

Como acabamos de ver hay dos tipos de técnicas de análisis de datos, las predictivas que nos permiten estimar lo que no conocemos y las descriptivas que nos permiten explorar, analizar y entender mejor los datos. Pero existe un tercer nivel de análisis de datos como extensión natural a los dos anteriores, la analítica prescriptiva que tiene como objetivo realizar recomendaciones sobre las acciones que se han de seguir, por ejemplo, para reducir costes o mejorar los beneficios. Algunos casos de uso típicos de la analítica prescriptiva son: generar planificaciones, establecer estrategias de precios, anticipar la afluencia de público a comercios, etc.

La gran cantidad de sitios especializados en Internet, ofertando millones de productos y/o servicios para su consumo, se ha convertido en un caos de información. Los sistemas de recomendación surgen como solución a este problema. Un sistema de recomendación recibe información del usuario acerca de productos y/o servicios en los que el usuario se encuentra interesado y le recomienda aquéllos cercanos a sus necesidades. Los sistemas de recomendación utilizan métodos matemáticos y estadísticos para explorar los datos y crear recomendaciones adaptadas a cada usuario. Los algoritmos de recomendación son utilizados por Netflix para remendarnos películas, por Spotify para recomendarnos canciones y por Amazon para recomendarnos productos.

La finalidad de un sistema de recomendación es predecir la valoración que un usuario va a hacer de un ítem (libros, películas, páginas, productos, anuncios, etc.) que todavía no ha evaluado. Esta valoración se genera al analizar o las características de cada ítem o las valoraciones de cada usuario a cada ítem. Los sistemas de recomendación se utilizan para recomendar contenido personalizado a los usuarios en base a su perfil, preferencias o intereses.

A la hora de definir y diseñar un sistema de recomendación podemos optar por tres enfoques:

* Filtrado colaborativo [5]: las recomendaciones a un usuario se basan en las preferencias de “usuarios similares”
* Filtrado basado en contenido: las recomendaciones se basan en sus propias preferencias, buscando elementos similares por los que ha mostrado interés
* Filtrado basado en reglas: las recomendaciones se basan en reglas previamente definidas como, por ejemplo, en características sociodemográficas.

Hay dos tipos de filtrados colaborativos:

* Basados en usuario (user-based): para predecir la valoración que un usuario A hará de un ítem X que todavía no ha visto, se buscan usuarios con perfiles similares a A y se utilizan las valoraciones de estos otros usuarios sobre el ítem X como estimación de la valoración de A.
* Basados en ítems (item-based): para predecir la valoración que un usuario A hará de un ítem X que todavía no ha visto, se buscan otros ítems similares (en función del perfil de valoraciones que han recibido) y que el usuario A también haya valorado. Se utilizan las valoraciones que el propio usuario A ha hecho de los ítems similares como predicción de su valoración sobre el ítem X. Este sistema puede parecer similar al basado en contenido, la diferencia se encuentra en que cada ítem no está definido por sus atributos sino por el perfil de valoraciones que ha recibido.

He optado para diseñar el sistema de recomendación de Jobs por un enfoque de filtrado colaborativo. Mi objetivo es recomendar a los usuarios de Jobs ofertas de trabajo o puestos en prácticas a las que son más afines en función de su perfil, es decir, su CV es afín a los requerimientos de la oferta.

### Medidas de similitud o distancia [6]

Todos los sistemas de recomendación tienen una cosa en común, para poder llevar a cabo las predicciones, necesitan definir y cuantificar la similitud entre ítems o usuarios.

Los sistemas de recomendación utilizan el término distancia como cuantificación de la similitud o diferencia entre observaciones. Si se representan las observaciones en un espacio p dimensional, siendo p el número de variables asociadas a cada observación (ítem o usuario), cuando más se asemejen dos observaciones, más próximas estarán, de ahí que se emplee el término distancia. Hay diferentes medidas de similitud:

* Correlación de Pearson: se emplea para calcular la correlación lineal entre dos vectores. Es la medida más apropiada si hay datos nulos.
* Correlación restringida de Pearson: utilizada valores medianos en lugar de medias de rating de dos usuarios.
* Coseno: es una de las medidas de similitud más utilizada. Trata de encontrar cómo dos vectores están relacionados entre ellos midiendo el coseno del ángulo entre los dos vectores. Considera los datos nulos cómo válidos.
* Jaccard: considera el número de preferencias compartidas entre dos usuarios y asume que dos usuarios serán más parecidos cuando tengan valores informados (ratings informados) similares.
* Diferencias de medias cuadradas: considera los ratings absolutos, en lugar de qué ítems han sido valorados. Normalmente se suelen combinar Jaccard.
* PIP: Proximity calcula la diferencia aritmética entre dos ratings. Impact calcula si un ítem le gusta o no al usuario. Popularity calcula la media aritmética del rating de ítems del usuario.
* Correlación de Spearman: Considera el ranking en lugar del rating para calcular la similitud.
* Correlación Tau de Kendall: considera los rankings relativos en lugar del rating para calcular la similitud.

Para el sistema de recomendación de Jobs en optado por utilizar el coseno como medida de similitud. El coseno del ángulo que forman dos vectores puede interpretarse como una medida de similitud de sus orientaciones, independientemente de sus magnitudes. Si dos vectores tienen exactamente la misma orientación (el ángulo que forman es 0º) su coseno toma el valor de 1, si son perpendiculares (forman un ángulo de 90º) su coseno es 0 y si tienen orientaciones opuestas (ángulo de 180º) su coseno es de -1.

### Lenguaje utilizado: R

De entre los diferentes lenguajes de programación que se pueden utilizar para el análisis de datos he decidido utilizar R para programar el algoritmo de selección.

He elegido R ya que he considerado que presenta ventajas significativas: excelente gama de paquetes de código abierto y de alta calidad, dispone funciones y métodos estadísticos integrales muy completos y permite una muy buena visualización de datos con el uso de bibliotecas como ggplot2. También he considerado que R tiene como principal desventaja el rendimiento, que no es el mejor, es decir, R es lento en comparación con otros lenguajes de programación, pero para programar el algoritmo de recomendación es lo suficientemente rápido.

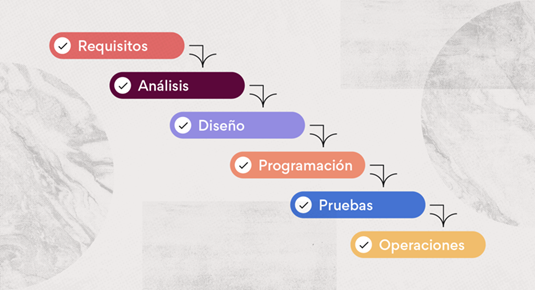
# Gestión del Proyecto

## Modelo de ciclo de vida

Ha de indicarse si se lleva a cabo un modelo de ciclo de vida en cascada, incremental, etc.

El ciclo de vida de un software es una secuencia estructurada y bien definida de las etapas necesarias para desarrollar el software deseado. Para el desarrollo del proyecto se ha decidido utilizar un Modelo de Cascada ya que es una metodología simple y las actividades (requisitos) están divididas en fases secuenciales contando con bucles por si es necesario volver hacia atrás para modificar/mejorar una fase. Dicho modelo se divide en las siguientes fases:

* Requisitos: consiste en la planificación detallada del proyecto donde: hay que estimar el tiempo y los recursos necesarios, evaluar cuidadosamente los riesgos, realizar una lista detallada de tareas, gráficos de dependencia, establecer hitos, ir mejorando los planes a medida que vamos sabiendo más, gráficos de GANTT y PERT
* Análisis: consiste en la especificación del producto, describir el sistema desde la perspectiva del usuario, descripción detallada de los datos de entrada/salida, como serán manejados los errores, todo debe de ser preciso para no tener que volver a fases anteriores.
* Diseño: está dividida en el diseño de la arquitectura (decidir el lenguaje de programación, diseñar las interfaces, muy importante que todas las decisiones de diseño estén especificadas claramente) y diseño detallado (diseñar estructura de datos y los algoritmos).
* Programación: traducir el diseño en código.
* Testing: probar que funciona como se esperaba el código
* Mantenimiento (Operaciones): mantenimiento del programa y refactorización del código.



*Ilustración 1: Fases del modelo en cascada*

## Papeles desempeñados en el proyecto

Además del tutor del trabajo fin de grado (TFG), que será el *dueño de producto*, y del estudiante, en el que colapsarán los papeles de director del proyecto y analista-programador (en caso de que el proyecto esté orientado a desarrollo), se indicará si también participan *stakeholders* y usuarios expertos y quiénes van a ser.

Los dos tutores de este TFG, Sergio Saugar Garcia y Alex Rayón Jerez, son los promotores y propietarios de este proyecto y el alumno, Carlos Soler Martín es el director del proyecto, siendo también el analista-programador que ha diseñado y desarrollado la aplicación JOBS.

## Planificación

A continuación, se listarán las tareas a realizar que se deberán llevar a cabo durante la realización del proyecto:

* Análisis de dominio de la aplicación, donde se definirán los procesos y actores que pertenecen al dominio analizado.
* Especificación de los requisitos obtenidos a través del análisis del dominio.
* Definición de una arquitectura de software que cumpla con los requisitos definidos.
* Diseño de la arquitectura (backend y frontend).
* Implementación de la arquitectura.
* Pruebas y validación del proyecto.
* Documentar el proyecto.

FALTA GANTT

Se identificarán las tareas a realizar, su descripción, sus fechas estimadas de inicio y de fin, sus precedencias, los recursos que requieren, etc. Se podrán utilizar diagramas de Gantt, PERT / CPM, etc. Deberán incluirse también los diferentes hitos a lo largo del proyecto. Para que el estudiante tenga la experiencia real de planificar y controlar su cumplimiento, es **importante** que la planificación se realice con anterioridad a la finalización del proyecto.

Se pueden encontrar explicaciones tanto sobre los diagramas que se indican en esta sección como los que se mencionan en los siguientes apartados en la [documentación sobre técnicas de la metodología Métrica 3](https://administracionelectronica.gob.es/pae_Home/dam/jcr:da7d91fa-d6bd-467c-be32-a72e27c603b3/METRICA_V3_Tecnicas.pdf).

## Presupuesto

En caso de que el desarrollo del TFG en sí mismo implique costes, o tenga sentido un trabajo con proyección a futuro, se indicará qué personal (con sus perfiles, dedicación, etc.), licencias de software, infraestructuras, etc. serán necesarios, así como sus precios y costes correspondientes.

El desarrollo del proyecto de este TFG no ha supuesto incurrir en ningún coste, ya que el software utilizado no ha implicado pagar ninguna licencia y todas las horas de desarrollo han sido asumidas por el alumno. Como el alumno cursa el doble grado de ISI y ADE, su intención es continuar con la elaboración de este proyecto desde un punto de vista empresarial, con objeto de realizar una propuesta de emprendimiento empresarial donde el proyecto software pueda ser la semilla de una futura empresa emergente.

## Ejecución

Se mostrará cómo ha transcurrido realmente el proyecto. Se podrán utilizar los mismos diagramas que en apartado anterior (diagramas de Gantt, PERT / CPM, etc.).

Realizar otro GANTT con timepos mayores

# Análisis y arquitectura

## Análisis de dominio

Xxxx

## Especificación de requisitos

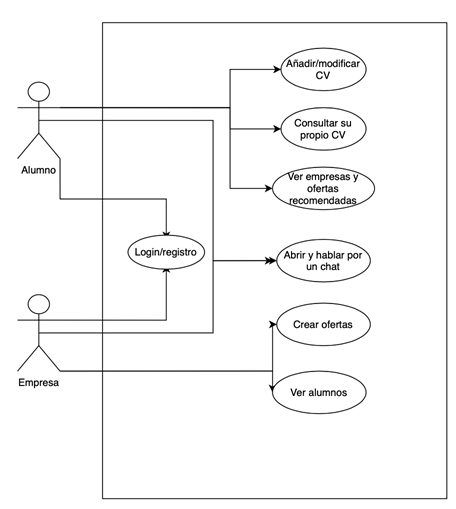
En un escenario habitual, se deberán especificar, al menos, los siguientes tipos de requisitos:

* *Funcionalidades del sistema,* que podrán detallarse, por ejemplo, mediante diagramas de casos de uso con su especificación. Alternativamente se pueden mostrar historias de usuario.
* *Interfaz de usuario* (si procede). Puede incluirse, por ejemplo, el nivel de cumplimiento de la norma de accesibilidad [WCAG](https://www.w3.org/WAI/standards-guidelines/wcag/).
* *Rendimiento*, incluyendo tiempos de respuesta.
* *Capacidad*, incluyendo el número máximo de peticiones que se pueden procesar por unidad de tiempo.
* *Seguridad*. Se indicará el nivel de seguridad que ha de satisfacer el sistema, por ejemplo, según el Esquema Nacional de Seguridad (RD 3/2010) o según alguna norma de la familia ISO 27000. Este nivel de seguridad se establecerá a partir de un análisis de riesgos (véase la sección de análisis de seguridad).
* *Interoperabilidad con otros sistemas* (si procede).
* *Protección de datos* (si procede), en concreto, el cumplimiento del Reglamento General de Protección de Datos ([UE] 2016/679) y de la Ley Orgánica 3/2018 de Protección de Datos Personales y Garantía de los Derechos Digitales (véase la sección de análisis desde la perspectiva del RGPD).
* *Requisitos sobre entorno tecnológico y de comunicaciones*, es decir, requisitos sobre gestores de bases de datos, sistema o los sistemas operativos, hardware, protocolos de red, etc.

Al comienzo del desarrollo del proyecto se definieron una serie de requisitos que para facilitar el desarrollo de la aplicación:

Requisitos funcionales

* Los alumnos deberán de ser capaces de registrarse, hacer login en la aplicación y cerrar sesión.
* Los alumnos deberán añadir su “CV” (skill, capacidades), que posteriormente podrá ser modificado por dicho alumno si lo desea
* Los alumnos podrán consultar su CV
* Los alumnos podrán ver las ofertas de empresas registradas en la aplicación
* Los alumnos deberán de ser capaces de ver las ofertas recomendadas en función del CV introducido
* Las empresas deberán de ser capaces de registrarse y hacer login en la aplicación
* Las empresas serán capaces de crear nuevas ofertas
* Las empresas tendrán acceso a los CV de los alumnos.
* **Chat**

****

*Ilustración 2: Diagrama de casos de uso*

Requisitos de interfaces externas

* Para la gestión del proyecto se utilizará un sistema de control de versiones

Requisitos de rendimiento

Requisitos tecnológicos

* Se utilizará la base de datos PostGreSQL para la gestión de todos los datos de la aplicación

## Análisis de los casos de uso y de las clases de análisis

A partir de los casos de uso, se obtendrá el modelo de clases de análisis, así como la realización de los casos de uso (diagramas de interacción entre objetos). En el diagrama de clases, para cada una de ellas, se identificarán sus atributos, responsabilidades (funcionalidades), asociaciones, agregaciones y generalizaciones. El comportamiento de las clases podrá mostrarse mediante diagramas de transición de estados.

Los diagramas elaborados no estarán condicionados por la tecnología utilizada, sino que estarán centrados en el problema en sí a resolver.

Xxx

## Análisis de seguridad

Se llevará a cabo un análisis de riesgos de acuerdo con las dimensiones de autenticidad, confidencialidad, integridad, disponibilidad y trazabilidad.

Xxx

## Análisis desde la perspectiva del RGPD (si procede)

En caso de que sea necesario, se llevará a cabo una gestión del riesgo y evaluación de impacto en tratamientos de datos personales.

Xxx

## Arquitectura del sistema

Se elaborará un diagrama donde se indiquen las interfaces del sistema, los diferentes servicios que proporciona, los componentes en que se apoyan tales servicios, capa de acceso a datos, etc.

Diagrama de la arquitectura del sistema

# Diseño

## Diseño del Backend

### Diseño del Servicio Web RESTful

#### Definición de un recurso

Tras el análisis de dominio realizado se han obtenido los siguientes recursos:

* Alumno: El recurso Alumno representa a aquellos usuarios, que, desempeñando el rol de Alumno, se conectan a la aplicación. Además, un alumno está asociado a un CV, se le pueden asignar una única oferta y puede ver las ofertas recomendadas. Los atributos del alumno son:
  + Nombre: Nombre completo, incluyendo el segundo nombre en caso de que existiera.
  + Apellido: Primer apellido.
  + Correo electrónico: Correo electrónico del alumno.
  + Número de teléfono: Número de teléfono de contacto del alumno
  + Nombre de usuario: Nombre de usuario para identificar al alumno.
  + Contraseña: Contraseña para verificar al alumno.
* Alumnos: Permite listar todos los alumnos registrados en la aplicación y también permite registrar nuevos alumnos.
* Empresa: El recurso Empresa representa a aquellas empresas, que, desempeñando el rol de Empresa, se conectan a la aplicación. Además, una empresa puede crear ofertas de trabajo. Los atributos de la empresa son:
  + Nombre de la empresa: Nombre completo de la empresa.
  + Correo electrónico: Correo electrónico de la empresa.
  + Número de teléfono: Número de teléfono de contacto de la empresa.
  + Nombre de usuario: Nombre de usuario para identificar a la empresa.
  + Contraseña: Contraseña para verificar a la empresa.
* Empresas: Permite listar todas las empresas registradas en la aplicación y también permite registrar una nueva empresa.
* Administrador: El recurso Administrador representa a aquellos administradores, que, desempeñando el rol de Administrador, se conectan a la aplicación. Además, un administrador puede asignar ofertas de trabajo a alumnos y realizar el cálculo de similitud de ofertas. Los atributos del administrador son:
  + Nombre de usuario: Nombre de usuario para identificar al administrador.
  + Contraseña: Contraseña para verificar al administrador.
* Administradores: Permite listar todos los administradores registrados en la aplicación y también permite registrar un nuevo administrador.
* Oferta: El recurso Oferta representa a aquellas ofertas, que, las crean las empresas y se les asignan a los alumnos. Los atributos de la oferta son:
  + Nombre de la empresa: Nombre de la empresa que realiza la oferta.
  + Título de la oferta: Titulo y descripción de la oferta, ej: Analista Jr. data science.
  + Alumno: alumno asociado a esa oferta de trabajo (en el caso de que la oferta esté asignada a un alumno).
  + Ciudad: Ciudad donde se realizará el trabajo.
  + Grado: Grado requerido para optar por la oferta de trabajo. Los grados disponibles son: Ingeniería Informática, Ingeniería Industrial, Ingeniería de Telecomunicaciones, Data scientist, Matemáticas, Física, ADE, Derecho, Marketing y Sociología.
  + Nota media: Nota media mínima para optar por la oferta de trabajo.
  + Niveles: Las ofertas requieren ciertos niveles de idiomas, de soft-skills, de competencias y lenguajes de programación, cuyo valor asignado es: no aplica, nivel básico, nivel medio y nivel avanzado. Los niveles necesarios por cada oferta son:
    - Idiomas: inglés, alemán y francés.
    - Soft-Skills: trabajo en equipo y comunicación.
    - Competencias: matemáticas, estadística, gestión de proyectos, sostenibilidad y big data.
    - Programación: nivel de programación en general.
  + Asignada: Indica si la oferta está asignada o no.
* Ofertas: Permite listar todas las ofertas y también permite registrar una nueva oferta.
* CV: El recurso CV representa los Curriculum Vitae creados por cada alumno, dicho CV se crea por el propio alumno y también puede ser modificado. Los atributos del CV son:
  + Alumno: Alumno asociado al CV.
  + Grado: Grado cursado/estudiado por el alumno. Los grados disponibles son: Ingeniería Informática, Ingeniería Industrial, Ingeniería de Telecomunicaciones, Data scientist, Matemáticas, Física, ADE, Derecho, Marketing y Sociología.
  + Nota media: Nota media obtenida por el alumno en sus estudios.
  + Niveles: EL CV requiere ciertos niveles de idiomas, de soft-skills, de competencias y lenguajes de programación, cuyo valor asignado es: no aplica, nivel básico, nivel medio y nivel avanzado. Los niveles necesarios por cada CV son:
    - Idiomas: inglés, alemán y francés.
    - Soft-Skills: capacidad analítica, trabajo en equipo, comunicación, pensamiento crítico, innovación, liderazgo, toma de decisiones y resolución a problemas.
    - Competencias: marketing, e-commerce, diseño gráfico, matemáticas, estadística, gestión de proyectos, redes sociales, sostenibilidad, inteligencia artificial, big data, machine learning, análisis de datos, bases de datos, cloud, IoT, redes, sistemas operativos, desarrollo web y diseño web.
    - Lenguajes Programación: R, Java, Pascal y Python.

#### Patrones de las URIs

En esta sección se exponen los patrones de URI asignados a los recursos comentados anteriormente.

* + **/alumnos**: recibe todas las peticiones HTTP relacionadas con los alumnos.
  + **/alumnos/:id**: recibe todas las peticiones HTTP relacionadas un alumno especifico.
  + **/alumnos/:id/CV**: Reúne todas las peticiones HTTP relacionadas con el CV de un alumno especifico.
  + **/empresas**: recibe todas las peticiones HTTP relacionadas con las empresas.
  + **/empresas/:id**: recibe todas las peticiones HTTP relacionadas una empresa especifica.
  + **/empresas/:id/ofertas**: recibe todas las peticiones HTTP relacionadas con todas las ofertas de una empresa especifica.
  + **/empresas/:id/ofertas/:id**: recibe todas las peticiones HTTP relacionadas con una oferta especifica de una empresa especifica.
  + **/ofertas/:id**: recibe todas las peticiones HTTP relacionadas con una oferta especifica.

#### API de cada recurso

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Método | URI | Utilidad | Semántica | Cuerpo Solicitud | Códigos de respuesta |
| POST | /alumnos | Añade un nuevo alumno. | JSON | Atributos de un nuevo alumno | 201 – Created. Location Header  400 – Bad Request  500 – Internal Server Error |
| GET | / alumnos | Se obtienen un listado de todos los alumnos. | JSON |  | 200 – OK  400 – Bad request  500 – Internal Server Error |
| PUT | - | - | - | - | 405 – Method not Allowed |
| DELETE | - | - | - | - | 405 – Method not Allowed |
| PATCH | - | - | - | - | 405 – Method not Allowed |

*Tabla 1: Recurso /alumnos*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Método | URI | Utilidad | Semántica | Cuerpo Solicitud | Códigos de respuesta |
| POST | - | - | - | - | - |
| GET | / alumnos/id | Se obtiene todos los atributos de un alumno por su id | JSON |  | 200 – OK  404 – Not Found  400 – Bad request  500 – Internal Server Error |
| PUT | - | - | - | - | 405 – Method not Allowed |
| DELETE | /alumnos/id | Eliminar un alumno específico | JSON | - | 200 – OK  404 – Not Found  400 – Bad Request  500 – Internal Server Error |
| PATCH | /alumnos/id | Modificar atributos del alumno | JSON | - | 200 – OK  404 – Not Found  400 – Bad request  500 – Internal Server Error |

*Tabla 2: Recurso /alumnos/id*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Método | URI | Utilidad | Semántica | Cuerpo Solicitud | Códigos de respuesta |
| POST | - | - | - | - | 405 – Method not Allowed |
| GET | /alumnos/id/CV | Se obtiene el CV de un alumno por su id | JSON | - | 200 – OK  404 – Not Found  400 – Bad request  500 – Internal Server Error |
| PUT | /alumnos/id/CV | Creación y modificación del CV un alumno especifico mediante su id | JSON | Atributos del CV de un alumno | 200 – OK  404 – Not Found  400 – Bad request  500 – Internal |
| DELETE | - | - | - | - | 405 – Method not Allowed |
| PATCH | - | - | - | - | 405 – Method not Allowed |

*Tabla 3: Recurso /alumnos/id/CV*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Método | URI | Utilidad | Semántica | Cuerpo Solicitud | Códigos de respuesta |
| POST | /empresas | Añade una nueva empresa. | JSON | Atributos de una nueva empresa | 201 – Created. Location Header  400 – Bad Request  500 – Internal Server Error |
| GET | /empresas | Se obtienen un listado de todas las empresas. | JSON |  | 200 – OK  404 – Not Found  400 – Bad request  500 – Internal Server Error |
| PUT | - | - | - | - | 405 – Method not Allowed |
| DELETE | - | - | - | - | 405 – Method not Allowed |
| PATCH | - | - | - | - | 405 – Method not Allowed |

*Tabla 4: Recurso /empresas*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Método | URI | Utilidad | Semántica | Cuerpo Solicitud | Códigos de respuesta |
| POST | - | - | - | - | - |
| GET | /empresas/id | Se obtiene todos los atributos de una empresa por su id | JSON |  | 200 – OK  404 – Not Found  400 – Bad request  500 – Internal Server Error |
| PUT | - | - | - | - | 405 – Method not Allowed |
| DELETE | /empresas/id | Eliminar una empresa especifica | JSON | - | 200 – OK  404 – Not Found  400 – Bad Request  500 – Internal Server Error |
| PATCH | /empresas/id | Modificar un atributo de una empresa | JSON | - | 200 – OK  404 – Not Found  400 – Bad Request  500 – Internal Server Error |

*Tabla 5: Recurso /empresas/id*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Método | URI | Utilidad | Semántica | Cuerpo Solicitud | Códigos de respuesta |
| POST | /empresas/id/ofertas | Una empresa especifica crea una nueva oferta | JSON | Atributos de una oferta | 201 – Created. Location Header  400 – Bad Request  500 – Internal Server Error |
| GET | /empresas/id/ofertas | Se obtiene un listado de todas las ofertas de una empresa especifica | JSON |  | 200 – OK  404 – Not Found  400 – Bad request  500 – Internal Server Error |
| PUT | - | - | - | - | 405 – Method not Allowed |
| DELETE | - |  | - | - | 405 – Method not Allowed |
| PATCH | - | - | - | - | 405 – Method not Allowed |

*Tabla 6: Recurso /empresas/id/ofertas*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Método | URI | Utilidad | Semántica | Cuerpo Solicitud | Códigos de respuesta |
| POST | - | - | - | - | 405 – Method not Allowed |
| GET | /empresas/id/ofertas/id | Se obtiene una oferta especifica de una empresa especifica | JSON |  | 200 – OK  404 – Not Found  400 – Bad request  500 – Internal Server Error |
| PUT | /empresas/id/ofertas/id | Modificación del campo de alumno asignado y del campo para verificar si la oferta esta asignada | JSON | Atributo alumno y asignado de una oferta | 200 – OK  404 – Not Found  400 – Bad request  500 – Internal |
| DELETE | - | - | - | - | 405 – Method not Allowed |
| PATCH | - | - | - | - | 405 – Method not Allowed |

*Tabla 7: Recurso /empresas/id/ofertas/id*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Método | URI | Utilidad | Semántica | Cuerpo Solicitud | Códigos de respuesta |
| POST | - | - | - | - | 405 – Method not Allowed |
| GET | /ofertas/id | Se obtiene una oferta especifica | JSON |  | 200 – OK  404 – Not Found  400 – Bad request  500 – Internal Server Error |
| PUT | /ofertas/id | Modificación del campo de alumno asignado y del campo para verificar si la oferta esta asignada | JSON | Atributo alumno y asignado de una oferta | 200 – OK  404 – Not Found  400 – Bad request  500 – Internal |
| DELETE | - | - | - | - | 405 – Method not Allowed |
| PATCH | - | - | - | - | 405 – Method not Allowed |

*Tabla 8: Recurso /ofertas/id*

### Diseño del algoritmo de recomendación

Xxxx

### Diseño de la base de datos

#### Diseño físico de datos

Se llevará un proceso de normalización y de optimización para obtener las tablas de la base de datos.

#### Migración y carga inicial de los datos (si procede)

En caso de que sea necesario, se elaborará un plan de migración y carga inicial de datos.

## Diseño del Frontend

### Diseño de la interfaz de usuario

Se mostrará cómo serán las ventanas, la navegación, etc. En caso de que sea necesario, podrán utilizarse diagramas de transición de estados.

## Modelo de clases de diseño

El diagrama de clases de análisis se transformará en otro en el que se incluyan aspectos más tecnológicos (accesos a bibliotecas, API de persistencia, etc.).

Xxx

## Entorno de construcción

Se indicará el IDE, *frameworks*, simuladores, etc. que se han utilizado.

Xxx

## Plan de pruebas

Se establecerán los criterios para codificar las pruebas realizadas a través de *test driven development* (TDD). Asimismo, se diseñarán el resto de pruebas que sean necesarias: estrés, carga, evaluación de la accesibilidad, a través de *scripting* (p.ej., con Selenium), con usuarios reales, etc.

Xxx

## Diagrama de infraestructuras de nivel 3

Se mostrarán las diferentes zonas de seguridad, *firewalls*, VLANs, servidores, etc. Los componentes de la infraestructura podrán ser tanto materiales como virtuales.

Xxx

## Diagrama de infraestructuras de nivel 2 (si procede)

Se mostrarán los *routers*, *switches*, conexiones entre ellos, etc. Para cada uno de los dispositivos se indicarán sus características técnicas.

Xxx

# Implementación

## Implementación Backend

### Implementación Servicio REST

Xxxx

### Implementación Algoritmo recomendación

Xxxx

### Implementación Base de datos

Xxxx

## Implementación Frontend

### XXX

XXX

### XXX

Xxx

# Construcción

## Referencia al repositorio de software

Se indicará la URL del repositorio en que está el software para, si procede, que el tribunal pueda probarlo en sus propias máquinas.

Xxx

## Manuales

En caso de que no tenga sentido incluir en el fichero README.md del repositorio alguno de los siguientes manuales: instalación, operación o usuario, se incluirá en la memoria del TFG.

Xxx

# Algoritmo de Recomendación

## Sistema de recomendación de JOBS

### Estructura del sistema de recomendación

El sistema de recomendación se estructura en dos APIs:

* API“ofertas\_nuevas.R”: calcula la similitud, en función de los requerimientos, entre las ofertas nuevas que las empresas han imputado en el sistema y las ofertas que ya han sido asignadas a alumnos.
* API“alumno\_nuevo.R”: calcula la similitud, en función del CV, del alumno nuevo con alumnos a los que ya se les ha asignado una oferta. Recomienda al alumno nuevo la/s oferta/s nueva/s que son similares a las ofertas asignadas a alumnos similares a él. Se considera que si dos alumnos son similares pueden recibir ofertas similares, es decir, tienen afinidad a ofertas similares.

### API “ofertas\_nuevas.R”

Este API hace post a una función que tiene dos parámetros: “a” es el valor inferior del código de oferta nueva (job\_id) y “b” es el valor superior del código de oferta nueva. Este modelo parte de dos imputs: la matriz de datos de ofertas asignadas (70 ofertas x 18 variables) y la matriz de ofertas nuevas (10 ofertas x 18 variables), que contienen información que no es relevante para el análisis (5 variables) e información de los requerimientos de cada oferta (13 variables) que es la base del análisis. Se genera la matriz ofertas (80x18) que es la unión de la matriz de ofertas asignadas y la matriz de ofertas nuevas. Se eliminan las variables (columnas) no relevantes y se obtiene una matriz de ofertas (80x13) que contiene los requerimientos de cada oferta.

Se transpone la matriz de ofertas (13x80) y, utilizando el método del coseno, se calcula la similitud (por columnas) entre las 80 ofertas. Cada oferta es un vector de 13 variables (requerimientos), si dos vectores son iguales el ángulo que forman es 0º y su coseno toma el valor de 1 (máxima similitud), mientras que si son totalmente diferentes forman un ángulo de 90º y su coseno es 0 (similitud nula).

La matriz de similitud entre ofertas (80x80) muestra la similitud de cada oferta con las otras 79, evidentemente cada oferta tiene similitud 1 consigo misma. Por ejemplo, si una oferta tiene una similitud de 0,9348 con otra, implica que los vectores de dichas ofertas forman un ángulo de 20,80º y su coseno es 0,9348. Lo cual implica qué de los 13 requerimientos de dichas ofertas, 10 son iguales y 3 (20,80º/90º x 13) son diferentes.

Se genera la matriz de vecinos (80x6) que contiene las 6 ofertas más similares a cada oferta (6 ofertas). La primera oferta más similar a una determinada oferta es ella misma con una similitud de 1. Una vez eliminada la columna del primer vecino se obtiene una matriz de vecinos (80x5) sobre la cual se genera la matriz vecinos ofertas nuevas (10x5), para lo cual se define un rango de códigos de ofertas (job\_id) que tiene como valor inferior el parámetro “a” (a = 71) y como valor superior el parámetro “b” (b =80). Finalmente se salva el archivo "vecinos\_ofertas\_nuevas.xlsx" que se convierte en un input del API “alumno\_nuevo.R”. El output de este API es un array que muestra para cada una de las 10 ofertas nuevas las 5 ofertas más similares de entre las ofertas asignadas.

### API “alumno\_nuevo.R”

Este API hace post a una función que tiene un parámetro, “alumnonuevo”, que corresponde con el código del alumno (alumno\_id) nuevo. Este modelo parte de dos imputs nuevos: la matriz de alumnos antiguos (70 alumnos x 43 variables) y el vector del alumno nuevo (1 alumno x 43 variables), que contienen información que no es relevante para el análisis (7 variables) e información del CV de los alumnos estructurado en skills (36 variables). Este modelo también tiene cómo input la matriz de ofertas nuevas (también es input del API “ofertas\_nuevas.R”) y la matriz de vecinos ofertas nuevas (output del API “ofertas\_nuevas.R”).

Se genera la matriz de alumnos (71x43) que es la unión de la matriz de alumnos antiguos y el vector del alumno nuevo. Se eliminan las variables (columnas) no relevantes y se obtiene una matriz de alumnos (71x36) que contiene el CV de cada alumno estructurado en 36 skills.

Se transpone la matriz de alumnos (36x71) y, utilizando el método del coseno, se calcula la similitud (por columnas) entre los 71 alumnos. Cada alumno es un vector de 36 variables (skills). La matriz de similitud entre alumnos (71x71) muestra la similitud de cada alumno con los otros 70, evidentemente cada alumno tiene similitud 1 consigo mismo. Por ejemplo, si un alumno tiene una similitud de 0,9059 con otro alumno, implica que los vectores de dichos alumnos forman un ángulo de 25,05º y su coseno es 0,9059. Lo cual implica que de las 36 skills, 26 son iguales y 10 (25,05º/90º x 36) son diferentes.

Se genera la matriz de vecinos (71x6) que contiene los 6 alumnos más similares a cada alumno. El primer alumno más similar a un determinado alumno es él mismo con una similitud de 1. Una vez eliminada la columna del primer vecino se obtiene una matriz de vecinos (71x5) sobre la cual se identifican los 5 vecinos del nuevo alumno (1x5), para lo cual se iguala el código de alumno (alumno\_id) al parámetro “alumnonuevo” (alumnonuevo =71).

Se identifican los vecinos comunes entre el vector de los 5 vecinos del nuevo alumno y la matriz de los 5 vecinos de las 10 ofertas nuevas. Dichos vecinos comunes permiten identificar los códigos (job\_id) de las ofertas nuevas a recomendar al alumno nuevo (rec\_alumno71). Se unen los data frame rec\_alumno71 y ofertas\_nuevas para generar las ofertas nuevas que se recomienda considerar al nuevo alumno. Finalmente se salva el archivo en dos formatos: "REC\_alumno71.xlsx" y “REC\_alumno71.json". El output de este API es un array que muestra al alumno nuevo la/s oferta/s nueva/s que son similares a las ofertas asignadas a alumnos similares a él.

### Llamada de las APIs

Xxx

### Intercambio de ficheros Web-APIs

Xxx

# Conclusiones y líneas futuras

Ideas a las que se llega después del desarrollo del proyecto, así como las líneas posibles de trabajo posterior.

Xxx

# Referencias

1. CIENCIA DE LOS DATOS Kelleher, D. J., Tierney, B. (2021). Ciencia de datos (MIT Press / Conocimientos Esenciales) (Spanish Edition). Ediciones UC.
2. MACHINE LEARNING Roman, V. (2019, 6 de febrero). Introducción al Machine Learning: una guía desde cero. Ciencia y Datos. <https://medium.com/datos-y-ciencia/introduccion-al-machine-learning-una-gu%C3%ADa-desde-cero-b696a2ead359>
3. METODOS DE APRENDIZAJE Cañadas, R. (2021, 7 de septiembre). Algoritmos de machine learning. ABdatum. <https://abdatum.com/machine-learning/algoritmos-machine-learning>.
4. SISTEMAS DE RECOMENDACIÓN González, A. (2020). Sistemas de recomendación de contenido con Machine Learning. Cleverdata. <https://cleverdata.io/sistemas-recomendacion-machine-learning/>
5. FILTRADO COLABORATIVO Bagnato, J.I. (2019, 27 de agosto). Sistemas de Recomendación. Crea en Python un motor de recomendación con Collaborative Filtering. Aprende Machine Learning. <https://www.aprendemachinelearning.com/sistemas-de-recomendacion/>
6. MEDIDAS DE SIMILITUD Mendoza Olguín, G. E., Laureano De Jesús, Y., & Pérez De Celis Herrero, M. D. L. C. (2019). Métricas de similaridad y evaluación para sistemas de recomendación de filtrado colaborativo. Revista de Investigación en Tecnologías de la Información, 7(14), 224–240. <https://doi.org/10.36825/riti.07.14.019>

# Anexo I

# Otras posibilidades para realizar el análisis y el diseño

A lo largo de este documento se ha asumido un desarrollo orientado a objetos. No obstante, se pueden seguir otros enfoques, por ejemplo, un desarrollo estructurado con diagrama de contexto, diagramas de flujos de datos, diagrama entidad-relación, etc.

Asimismo, es posible utilizar el diagrama entidad-relación de forma complementaria a los diagramas de clases.

Por otra parte, el tutor del proyecto puede dar sus propias orientaciones si lo considera oportuno.

# Anexo II

# Fuente de inspiración del presente documento

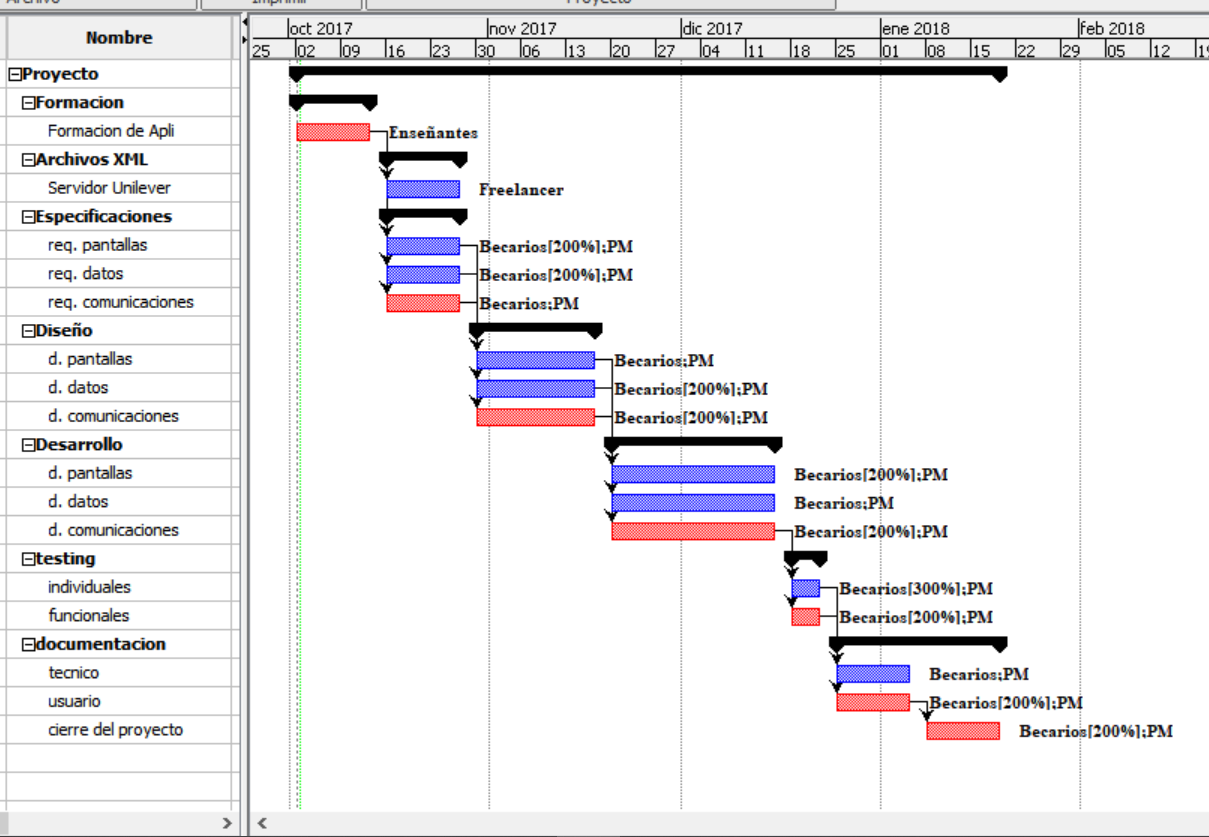
Independiente de que su uso que esté actualmente más o menos extendido, se ha considerado que la metodología [Métrica 3](https://administracionelectronica.gob.es/pae_Home/pae_Documentacion/pae_Metodolog/pae_Metrica_v3.html) es adecuada como marco de referencia en la elaboración de la documentación. En consecuencia, ha sido una fuente de inspiración.

# Glosario de términos

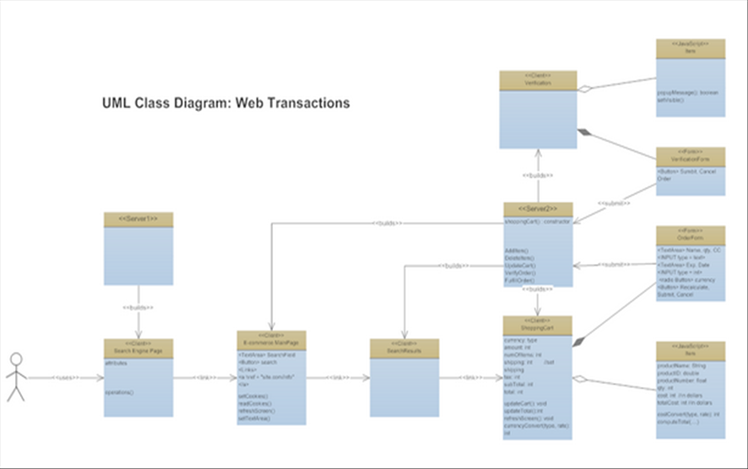
Si es necesario.

# Anexos

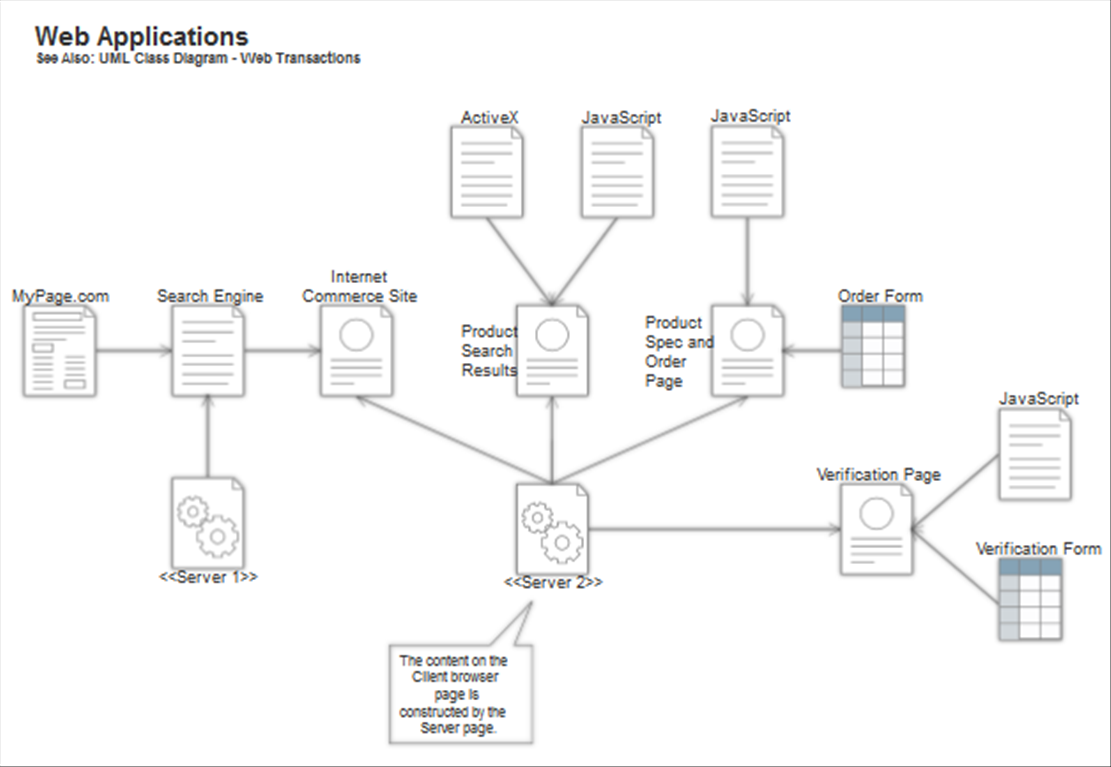
Ejemplo de planificación de un TFG



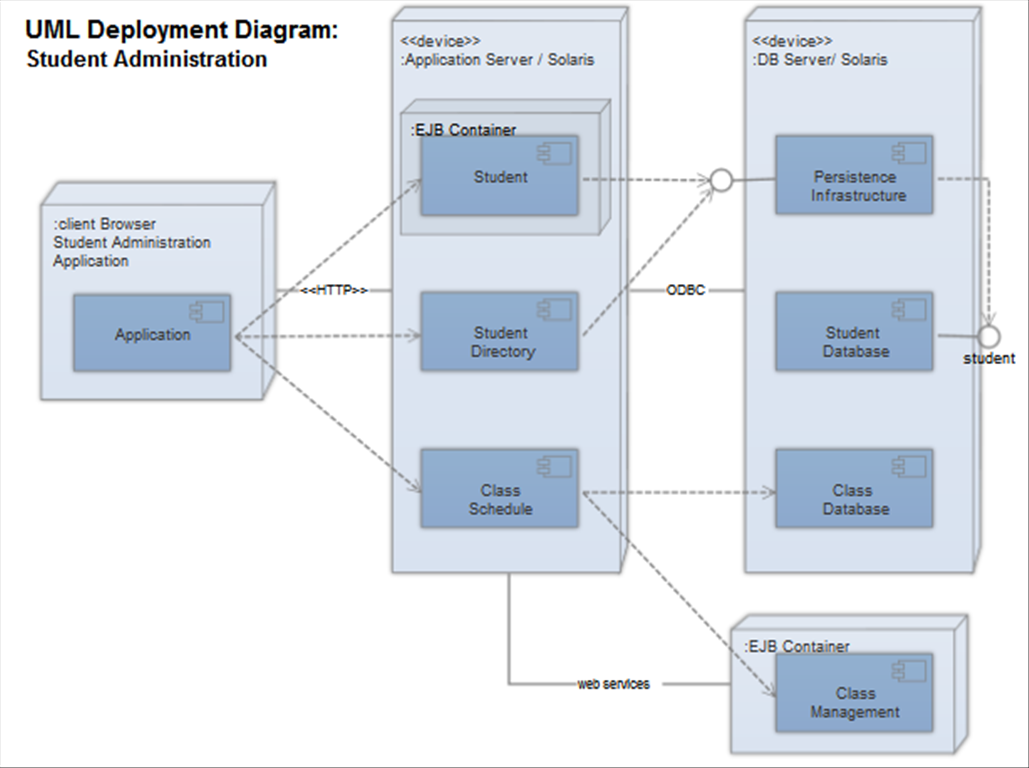
Ejemplo de diagrama de clases



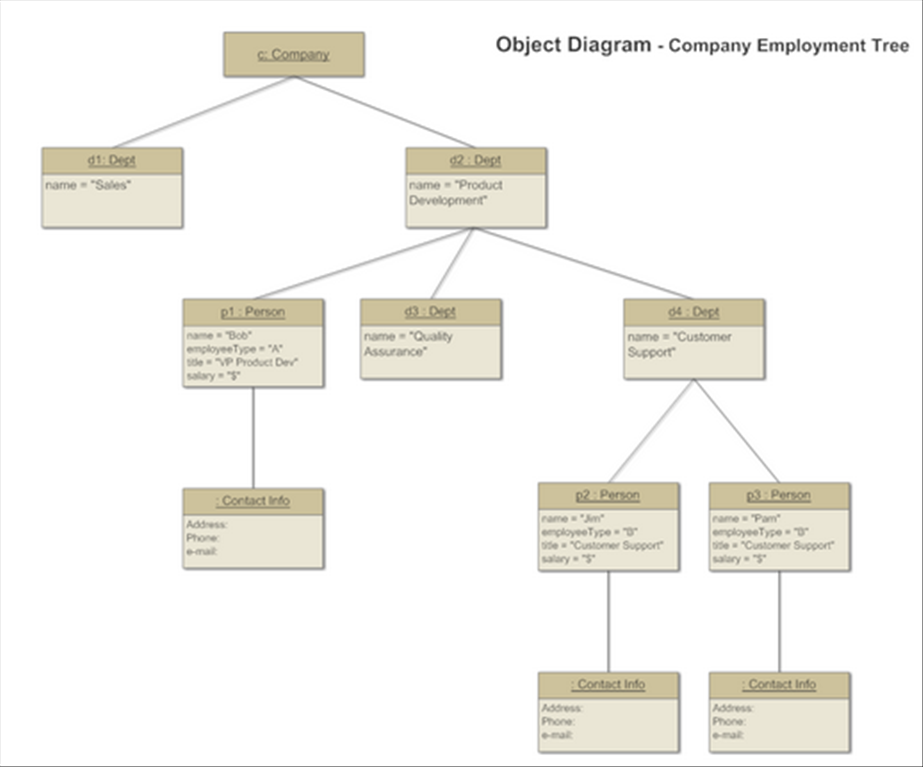
Ejemplo de diagrama de componentes



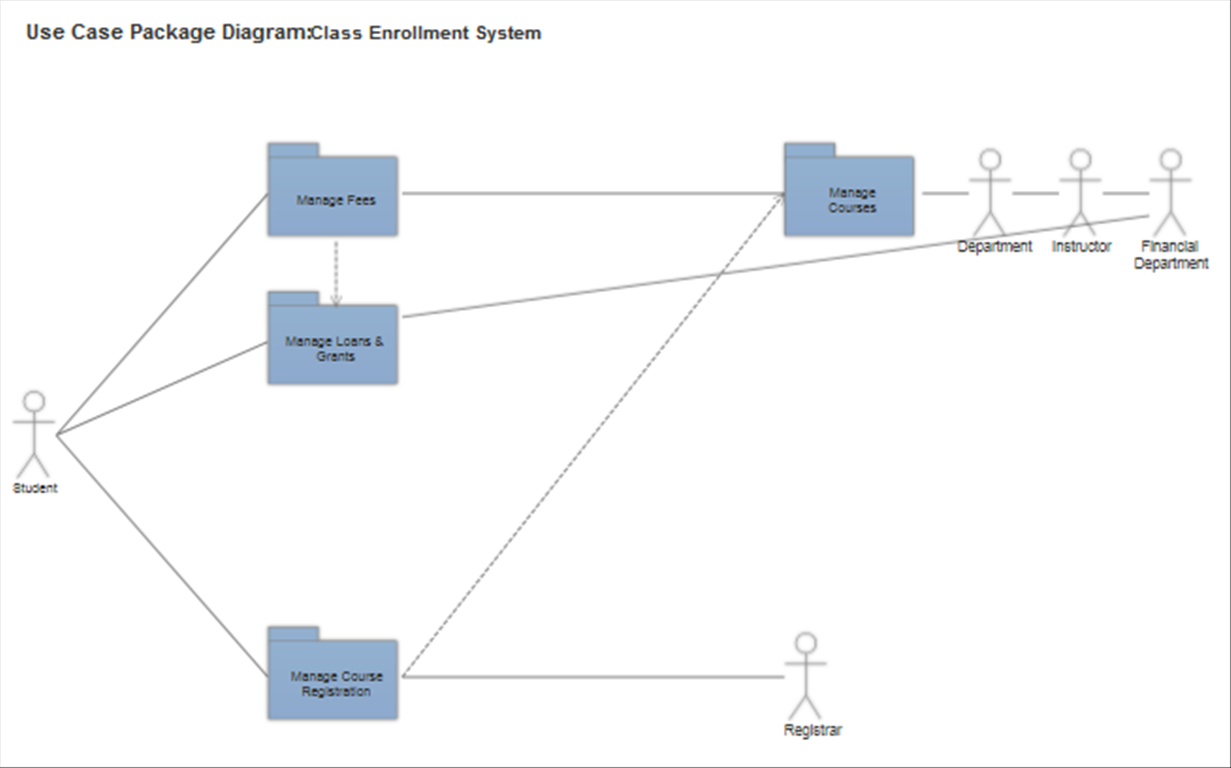
Ejemplo de diagrama de despliegue



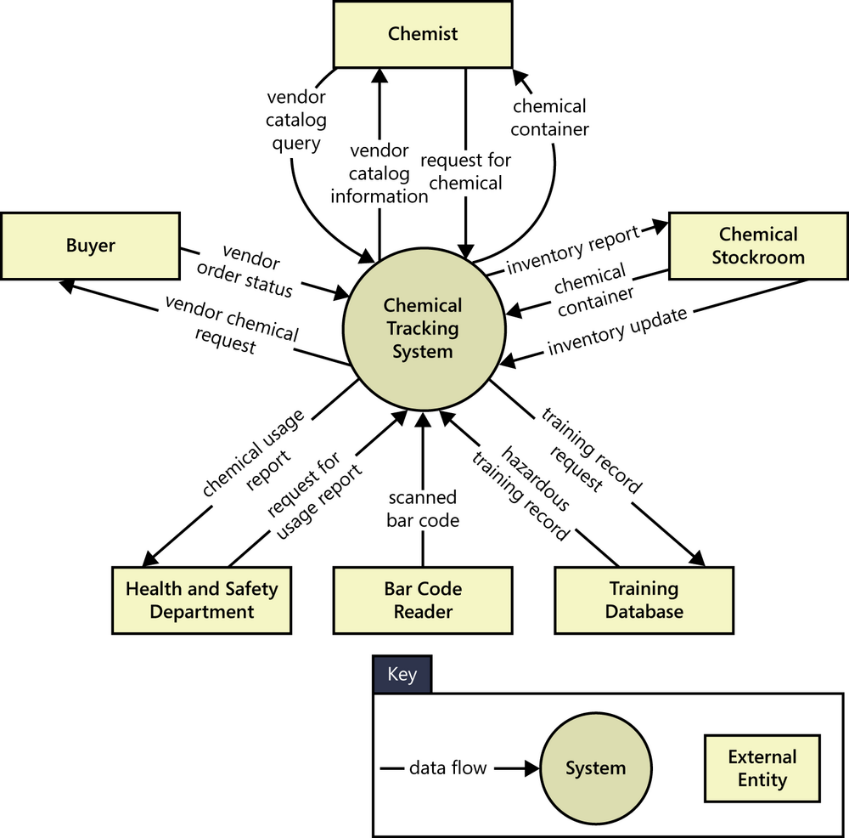
Ejemplo de diagrama de objetos



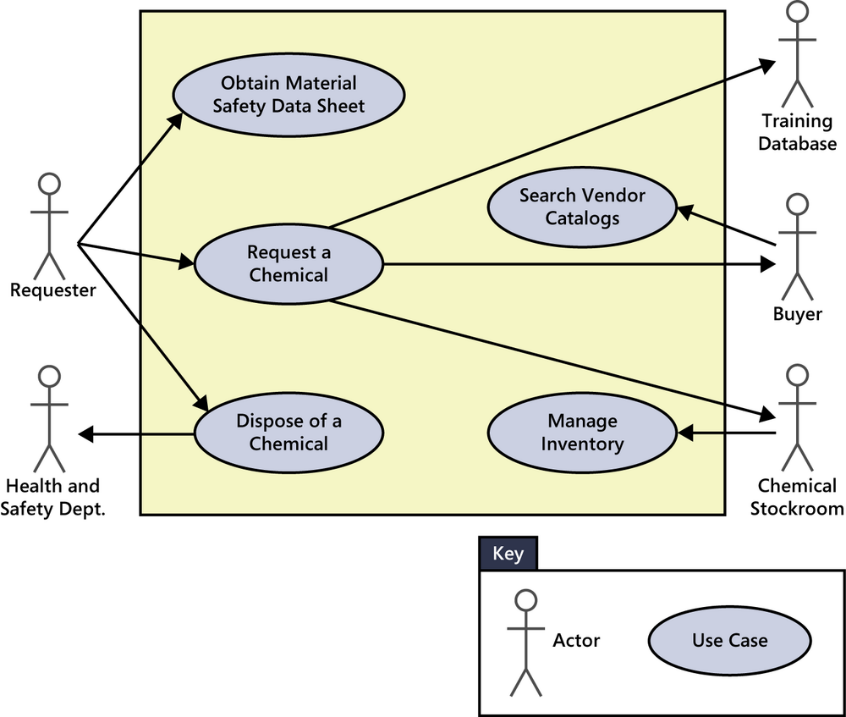
Ejemplo de diagrama de paquetes



Ejemplo de diagrama de contexto



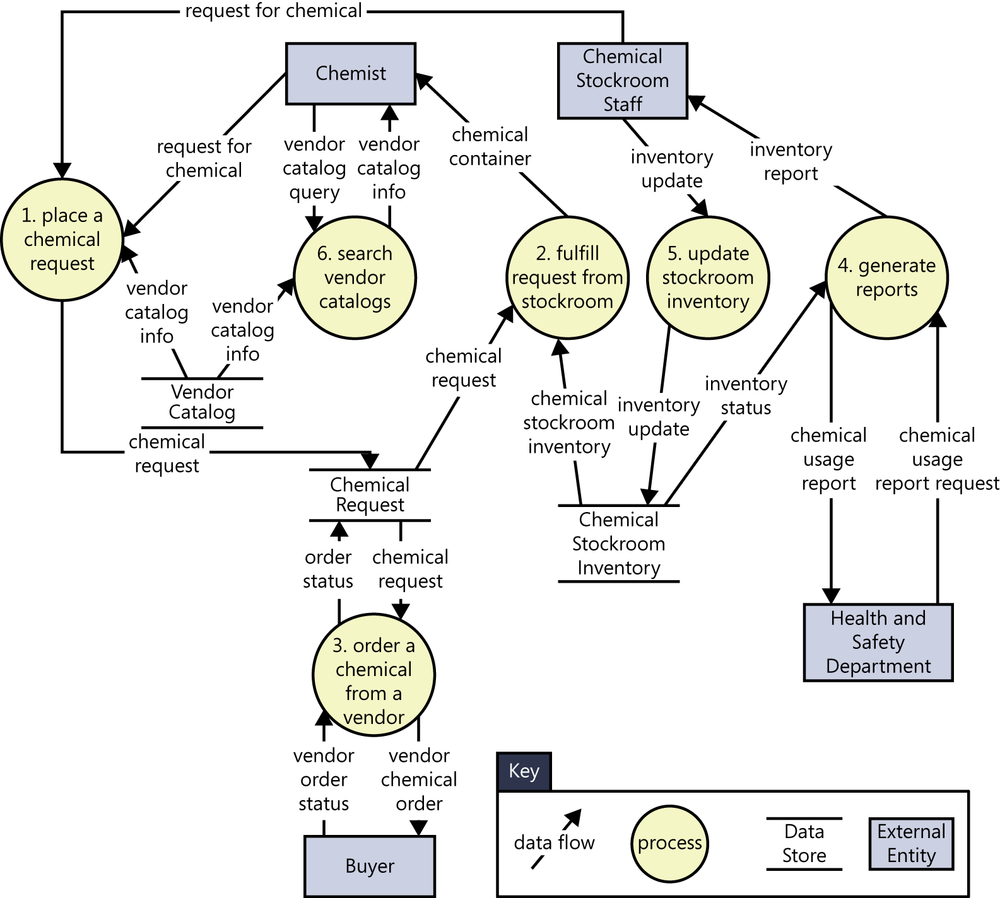
Ejemplo de caso de uso



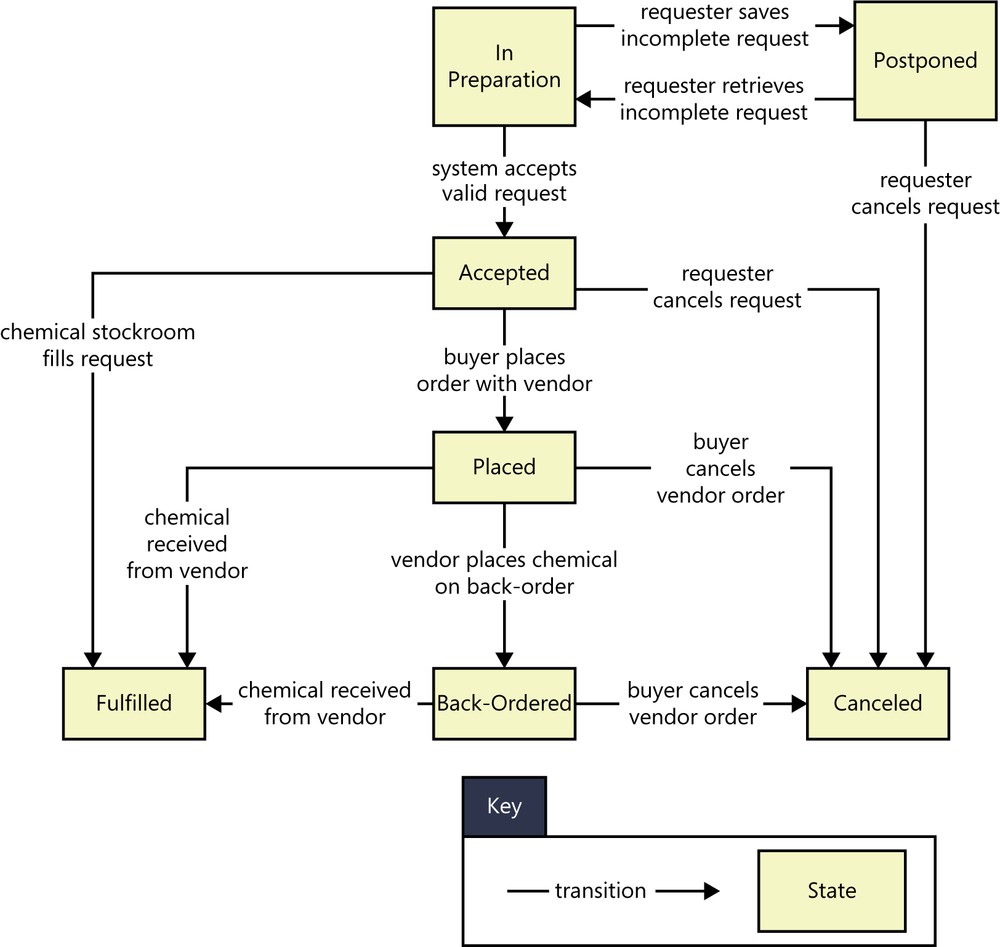
Ejemplo de diagrama de actividades

An illustration showing the normal flow of a use
                  case in the left part of the figure. Starting with a solid
                  dot that represents the use case preconditions, arrows lead
                  downward through a series of steps in rectangles to the use
                  case postconditions. In the middle of the stack of
                  rectangles is a diamond representing a branch point
                  condition. An arrow goes from the diamond to the right,
                  where there is another series of vertical boxes and arrows
                  representing steps in the alternative flow. The last step in
                  the alternative flow has an arrow going back to the left to
                  rejoin the last step in the normal flow.

Ejemplo de DFD Diagrama de Flujo de datos



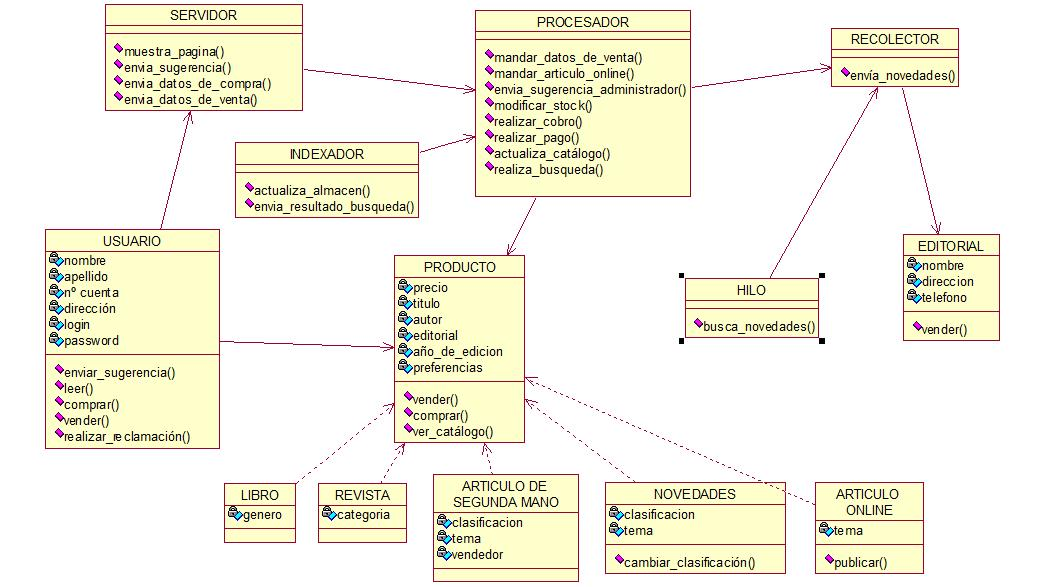
Ejemplo de diagrama de transición



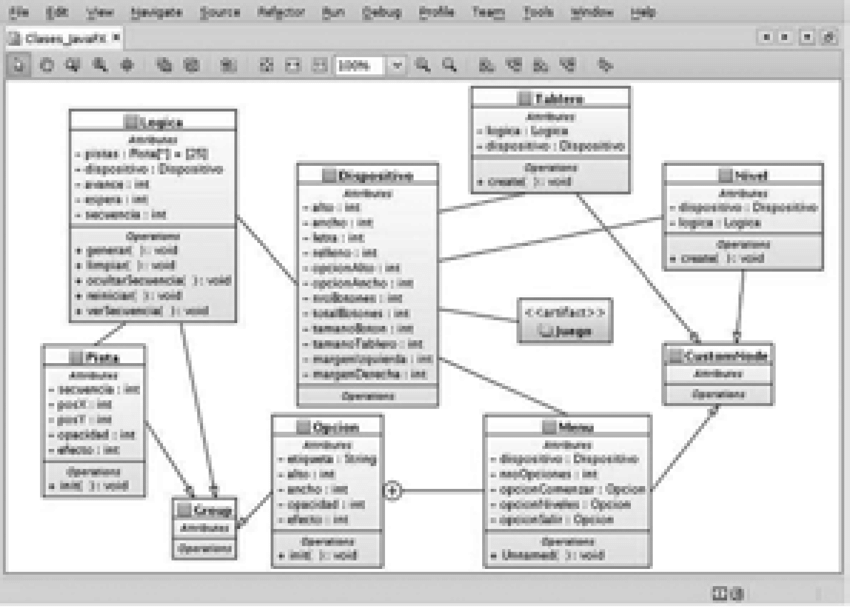
Ejemplo de diagrama lógico de una vpn



Ejemplo de Diagrama E/R



Ejemplo de diagrama de clases de usuarios



Ejemplo de interfaz HMI

