



Escuela  
Politécnica  
Superior

# Sistema de recomendación de titulaciones universitarias basado en datos abiertos



Máster Universitario en Ingeniería Informática

## Trabajo Fin de Máster

Autor:

Mohamad Yousef HatHat

Tutor/es:

David Tomás Díaz

José Norberto Mazón López

Septiembre 2016



Universitat d'Alacant  
Universidad de Alicante





Universidad de Alicante  
Escuela Politécnica Superior

**Sistema de recomendación de  
titulaciones universitarias basado en  
datos abiertos**

Autor

**Mohamad Yousef HatHat**

Tutor

**David Tomás Díaz**

Cotutor

**José Norberto Mazón López**

Septiembre 2016



# Agradecimientos

En primer lugar agradecer a mi familia, en especial mi madre, por todo el tiempo que ha dedicado en levantarme los ánimos, actuar como guía profesional a parte de madre, aguantarme todos los días y saber escucharme cuando más lo he necesitado.

A todos mis compañeros del master en general, sin olvidar los buenos amigos que he tenido el placer y honor de haber conocido, en especial a Pedro Amorós y Emilio Correas por hacer amenas todas las clases con gracias, fueras de tono, trabajo en equipo y sobre todo compañerismo.

A mis dos compañeros de batalla Juan Antonio Brotons y Jose Manuel Bru. Por todos esos momentos de recaídas, desmotivación y frustración con los obstáculos que se presentan en la vida estudiantil, laboral y personal. En gran parte también, a mi colega Héctor Sempere por estar ahí y aunque este año no ha estado presente en la carrera hacia el master y la posterior colegiación futura, si lo ha estado en el ámbito laboral.

Especial agradecimiento a mis tutores David Tomás y Jose Norberto, que con esmero y arduo trabajo han conseguido levantar este proyecto conmigo y darle una solución lo más óptima posible. Muchas gracias por su paciencia conmigo y por contestar a mis correos por muy tarde que fuese.

Este año, el año que espero que cierre mi etapa como master en ingeniería informática, ha sido un año y medio de experiencias sorprendentes, satisfactorias y en gran medida alegres, en el que he aprendido a solventar errores más enfocados a un ámbito laboral que solventarlos de cara a una y atiborrado de presentaciones.

Contento por haber puesto mi grano de arena en la Escuela Politécnica Superior y haber aprendido en mi estancia como estudiante, si Dios quiere, volver para ser docente o volver a estar involucrado en la universidad que me vio crecer como persona y mejorar en todos los aspectos de la vida.



# Índice general

1. Introducción y justificación del proyecto.....	9
2. Estado de la cuestión.....	13
3. Objetivos .....	17
4. Metodología.....	18
4.1. Modelo de negocio.....	19
4.1.1. Segmento de clientes .....	21
4.1.2. Problema .....	21
4.1.3. Propuesta de valor.....	22
4.1.4. Solución .....	22
4.1.5. Canales de distribución.....	23
4.1.6. Fuente de ingresos .....	23
4.1.7. Estructura de costes .....	23
4.1.8. Métrica clave .....	23
4.1.9. Ventaja diferencial .....	24
4.2. Datos abiertos.....	24
4.3. Test de preguntas inicial .....	26
4.4. Desarrollo del sistema de recomendación .....	27
4.5. Diseño de la aplicación .....	32
4.5.1. Diagrama de flujo principal.....	33
4.5.2. Página inicial .....	35
4.5.3. Carga de datos inicial .....	36
4.5.4. Cuestionario para el alumnado.....	36
4.5.5. Proceso de selección .....	39
4.5.6. Finalización .....	42
5. Posibles mejoras.....	44
6. Conclusiones .....	45
Glosario de términos.....	46
Bibliografía .....	47
Anexo I: Lean Canvas vs Business Canvas.....	49

Anexo II: Respuestas detalladas del test inicial .....52

Anexo III: Lista de titulaciones de la UA.....58



# Índice de figuras

Figura 1: Esquema de la recomendación colaborativa .....	13
Figura 2: Esquema de la recomendación basada en contenido .....	14
Figura 3: Esquema de la recomendación basada en conocimiento .....	14
Figura 4: Esquema de la recomendación híbrida .....	15
Figura 5: Esquema del proceso de generación de una recomendación .....	15
Figura 6: Esquema global de gradUAdvice .....	18
Figura 7: Indexador de Lucene .....	27
Figura 8: Ejemplo de selección con Lucene .....	31
Figura 9: Diagrama de flujo .....	34
Figura 10: Pagina principal .....	35
Figura 11: Ejemplo de respuesta servidor .....	36
Figura 12: Cuestionario inicial .....	37
Figura 13: Elección del bachillerato.....	37
Figura 14: Lista de asignaturas elegidas .....	38
Figura 15: Condiciones a aceptar .....	38
Figura 16: Primera iteración .....	39
Figura 17: Segunda iteración.....	40
Figura 18: Tercera iteración.....	41
Figura 19: Resultado final.....	42
Figura 20: Formulario final .....	43
Figura 21: Modelo Lean Canvas.....	50
Figura 22: Modelo Lean Canvas vs Business Model Canvas .....	51
Figura 23: Gráfica de tipos de bachillerato.....	52
Figura 24: Gráfica de año cursado las PAU .....	53
Figura 25: Gráfica de decisión de la carrera universitaria.....	53
Figura 26: Gráfica de explicación de planes de estudio.....	54

Figura 27: Gráfica de planes de estudio .....	55
Figura 28: Gráfica de dudas con el plan de estudios .....	55
Figura 29: Gráfica de acudir a la universidad en busca de información .....	56
Figura 30: Gráfica de utilización de herramienta online.....	57

## Indice de tablas

Tabla 1. Modelo Lean Canvas .....	20
Tabla 2: Ejemplo JSON de la API de datos UA.....	25
Tabla 3: Similitudes coseno del Grado en Administración y Dirección de Empresas.....	30

# 1. Introducción y justificación del proyecto

La cantidad de datos abiertos disponibles hoy en día es cada vez mayor, resultando indispensable la creación de aplicaciones que hagan uso de los datos y aporten valor. Una de las categorías de datos abiertos más relevantes hoy en día son los datos del sector público, es decir, aquellos datos que las administraciones públicas poseen y que en el contexto de un gobierno abierto deberían ser accesibles a todo el mundo para su reutilización.

En la vida cotidiana se presentan situaciones en las que hace falta tomar decisiones que a simple vista se pueden considerar sencillas. Ejemplos claros como elegir una película para ver, un restaurante para cenar, un libro para leer o planear unas vacaciones pueden ser tareas bastante complicadas por una razón principal: la gran cantidad de libros, restaurantes, películas y destinos vacacionales que existen.

Los Sistemas de Recomendación han surgido para intentar paliar las dificultades de tratar con tantas opciones. En términos generales, los sistemas de recomendación generan sugerencias sobre productos (o acciones) dentro de un determinado dominio en el cuál está interesado el usuario [1]. Son técnicas de filtrado de información que producen sugerencias personalizadas para un conjunto de datos iniciales. Tienen como objetivo el predecir el comportamiento de los usuarios que lo usan para proporcionarle datos que con una alta probabilidad de ser útiles. Se han hecho muy populares en los últimos años [1]. Entre los ámbitos más extendidos se encuentran las recomendaciones en tiendas online, películas (Netflix), música (Spotify), libros (Amazon), productos (eBay) o recomendaciones de perfiles a los que seguir en redes sociales (Facebook, Twitter, LinkedIn, etc...).

Este trabajo se enfoca en la unión de una fuente de información disponible y reutilizable, y el cometido de un sistema de recomendación. Recibir unos datos como pueden ser las titulaciones universitarias de la Universidad de Alicante y junto a un sistema de recomendación basado en contenido con lo que implica, diseñar una aplicación web que aconseje a los potenciales estudiantes de la UA qué posible o posibles carreras podrían cursar en un futuro.

Proyecto que surge de haber observado como la tasa de abandonos en la universidad cada año crece y la decepción de los alumnos por no encontrar la carrera profesional que más se ajusta a sus gustos. Problema que se remonta a una mala decisión de la titulación universitaria que se cursa, y a su vez falta de información que pueda serle atractiva al potencial estudiante universitario. Una cuestión que ha llevado a este proyecto analizar el problema con otros ojos, divisar una posible solución, desarrollarla y experimentar con los resultados.

De ahí sale **gradUAdvice**, la unión de tres términos: *Grado*, *UA* y *Advice* (recomendación en inglés), una forma elocuente y atractiva de darle un título a una aplicación web desarrollada para realizar una función clara: recomendar titulaciones universitarias a un usuario utilizando sus gustos y pasando por un proceso de selección de términos, con el fin de obtener una o varias titulaciones que le puedan ser atractivas, sin necesidad (aunque condición no suficiente) de indagar en los planes de estudios para distinguir entre titulaciones.

Como todo SR, se necesita una base de datos fiables de donde se pueda extraer información relevante, para ello se ha elegido el Portal de Datos Abiertos de la Universidad de Alicante ([www.datos.ua.es](http://www.datos.ua.es)). Proyecto empezado en 2015 con la idea de abrir el mundo del Open Data (explicado detalladamente en las secciones siguientes) a todos los públicos, en particular a los desarrolladores, que aplicando sus conocimientos, puedan obtener beneficios y darle un renombre a la universidad.

Por lo tanto, del portal de datos de la UA, se extraerán datos acerca de las asignaturas de las titulaciones universitarias cursadas en la universidad, para alimentar correctamente al SR.

En cuanto a la utilización de datos abiertos, se definen como datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen.

Se puede justificar su uso en este trabajo mencionando sus características principales:

- Disponibilidad y acceso: la información está disponible como un todo y a un costo razonable de reproducción, preferiblemente descargándola de Internet. Además, la información está disponible en una forma conveniente y modificable, como por ejemplo formato JSON.

- Reutilización y distribución: los datos están provistos bajo términos que permitan reutilizarlos y redistribuirlos, e incluso integrarlos con otros conjuntos de datos.
- Participación universal: todos deben poder usar, reutilizar y redistribuir la información, sin ninguna discriminación en términos de esfuerzo, personas o grupos.

¿Cómo recomienda gradUAdvice? El SR devuelve unos términos relevantes de cada titulación y las clasifica según similitud, es entonces cuando en la parte visual, al usuario se le muestran en un formato intuitivo y atractivo visualmente: Tag Clouds o nubes de tags.

Término creado para representar una nube hipotética donde dentro alberga palabras que puedan interesar al usuario, de esta forma es más funcional para percibir los términos más prominentes.

De esta forma se puede recomendar al usuario qué términos engloban ciertas titulaciones y que él mismo se decante por un grupo u otro según le resulte más llamativo o sea de su interés. Y así tantas veces hasta que gradUAdvice le arroje un resultado o varios que equivalga a la titulación universitaria que más se ajuste a sus necesidades.

Una parte muy importante de precede a un desarrollo e incluso a un análisis de especificaciones es es el estudio de mercado y competencias organizacionales.

En este trabajo se considera la opción de adjuntar un estudio de mercado para asegurar la existencia de una vía de mercado fructífera. Crear un modelo de negocios basado en una plantilla universal llamada Model Lean Canvas y analizar cada uno de los puntos para llegar saber cuales son los puntos claves sobre los que se tiene que sustentar gradUAdvice.

Además, sustentar el modelo con un cuestionario real lanzado para comprobar que la comunidad universitaria nueva de este curso 2016-2017 tendría la iniciativa de usar una herramienta online web que les permitiese aconsejar titulaciones universitarias a cursar según sus gustos. Con la esperanza de llegar a la conclusión afirmativa a esta pregunta:

*¿Utilizarían los usuarios gradUAdvice para elegir titulación universitaria?*



## 2. Estado de la cuestión

Antes de la llegada de Internet, un consumidor de cualquier tipo de producto o servicio tenía acceso limitado o carente a veces a la información relacionada tanto con el mismo producto/servicio en sí como con otras posibles opciones. La única forma de hacerse llegar era a través de la publicidad, problema al que tenía que hacer frente el usuario: ¿la información que me proporcionan es veraz, o solo es más spam? Actualmente la situación se ha revertido, pasar de escasear la información a tener que elegir bien entre la información que interesa y la que no [3].

Una aproximación para intentar ofrecer a cada persona lo que busca es mediante el análisis del contenido. Para ello se hace una representación del contenido de cada elemento del conjunto y se compara con una representación del contenido que el usuario está buscando.

Este filtrado basado en contenido es efectivo principalmente para encontrar documentos textuales en función de unos criterios de búsqueda. Sin embargo, es más difícil parametrizar de forma automática contenidos multimedia.

Existen varias técnicas estándar de recomendación:

- **Recomendación colaborativa:** si los usuarios han compartido algunos de sus intereses en el pasado, tendrán gustos similares en el futuro. Véase la figura 1.

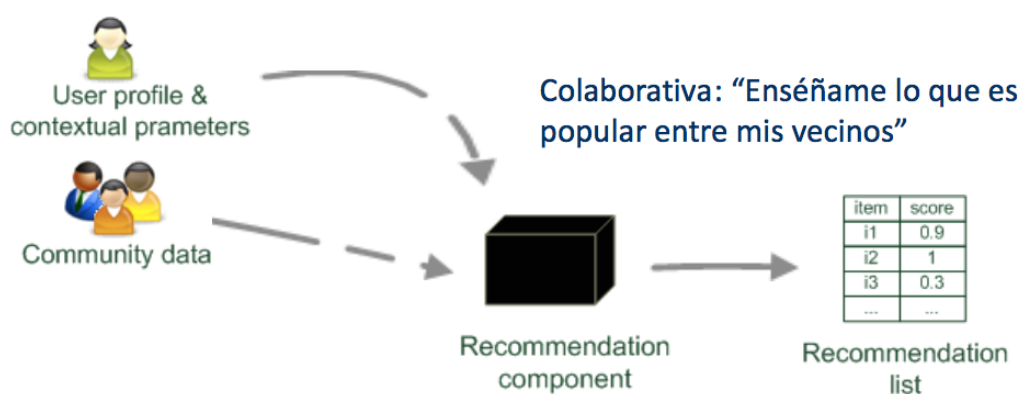
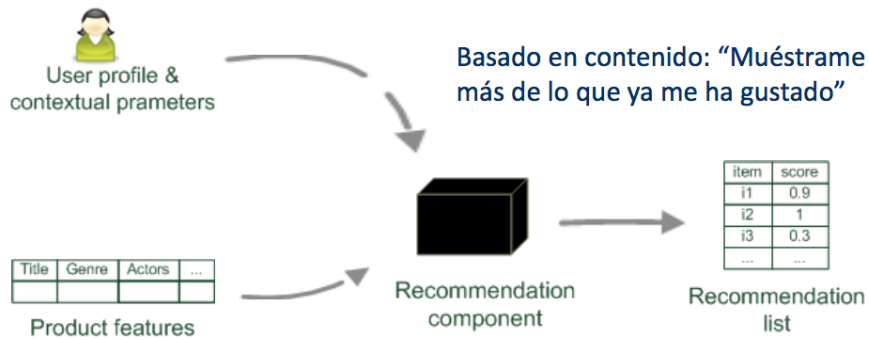


Figura 1: Esquema de la recomendación colaborativa

- **Recomendación basada en contenido:** partiendo de una base de elementos con características y



un perfil con asignaciones de importancia a esas características. Emparejar aquellos elementos que mejor cumplan las preferencias del usuario. Véase la figura 2.

Figura 2: Esquema de la recomendación basada en contenido

- **Recomendación basada en conocimiento:** recomendar a usuarios novatos (de los que no se tiene historial y/o preferencias) sugerencias que están más centradas en el dominio de aplicación. Se necesita información adicional para poder realizar las recomendaciones (información sobre el usuario y/o sobre los elementos). Véase la figura 3.

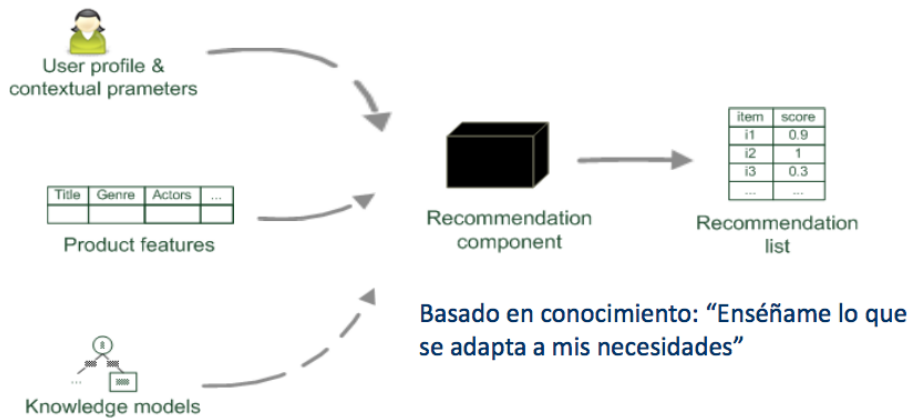


Figura 3: Esquema de la recomendación basada en conocimiento



- **Recomendación híbrida:** combinación de varias técnicas anteriores para obtener mejores recomendaciones. Véase la figura 4.

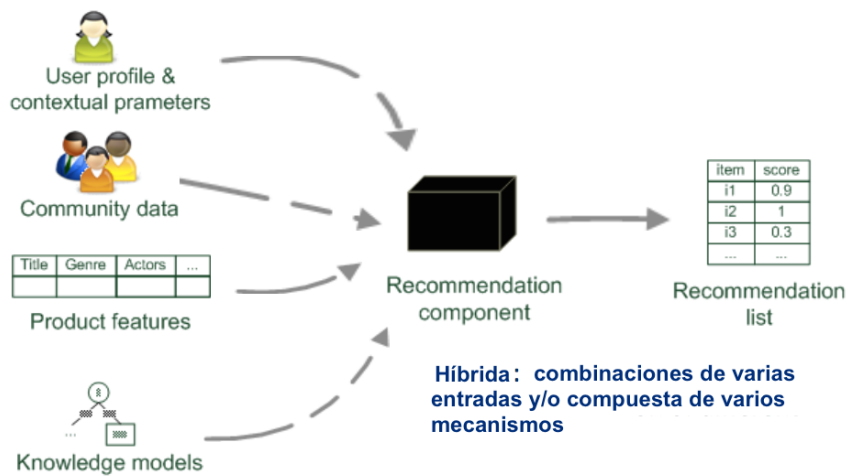


Figura 4: Esquema de la recomendación híbrida

La información sobre los usuarios puede venir dada de dos formas [4] que no tienen porqué ser mutuamente exclusivas: por extensión o intencionalmente. Puede verse la figura 5 para entender el funcionamiento completo de un sistema de recomendación.

La primera referencia a información que se tenga sobre las experiencias pasadas del usuario con respecto a los ítems encontrados. Es lo que también se conoce como navegación implícita ya que el usuario no es consciente de estos seguimientos. Mientras que la segunda, alguna especificación se entiende de los ítems deseados por los usuarios y aporta intencionalmente al sistema cualquier información sobre sus preferencias.

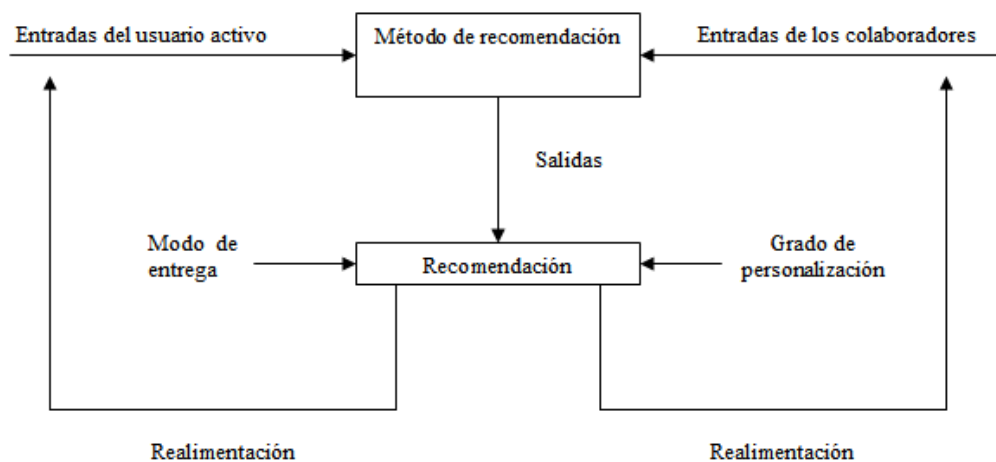


Figura 5: Esquema del proceso de generación de una recomendación

La salida del sistema está constituida por las recomendaciones generadas, que varían dependiendo del tipo, cantidad y formato de la información proporcionada al usuario. Algunas de las formas más comunes de representar la salida son las siguientes:

- Sugerencia o lista de sugerencias al usuario de una serie de ítems.
- Presentar al usuario predicciones del grado de satisfacción que se asignará al ítem concreto. Estas estimaciones pueden ser presentadas como personalizadas al usuario o como estimaciones generales del conjunto de colaboradores.
- Cuando la comunidad de usuarios es pequeña o se conocen bien los miembros de dicha comunidad, podría ser útil visualizar las valoraciones individuales de los miembros que permitiría al usuario activo obtener sus propias conclusiones sobre la efectividad de una recomendación.

Independientemente de estos formatos de salida, puede resultar muy interesante incluir una breve descripción o explicación sobre el ítem recomendado a modo de justificación del porqué de dicha recomendación.

En cuanto a las soluciones que el mercado aporta en términos de “aconsejar titulaciones universitarias”, haciendo una indagación en Google<sup>1</sup> los resultados que arroja con más frecuencia son foros donde se discuten experiencias en titulaciones universitarias, tipos de profesorado, formas de examinación y lo más importante, salidas profesionales de una carrera universitaria u otra.

Además de buscar en los repositorios de las plataformas móviles más importantes como pueden ser la App Store<sup>2</sup> de iOS y el PlayStore<sup>3</sup> de Android, las únicas aplicaciones que destacan en un barrido simple son apps que conciernen a cómo mejorar la productividad, gestión del tiempo y de recursos en la universidad y estudios en general.

Por lo tanto, la oportunidad de mercado tanto en web como en aplicación de mercado están presentes para que gradUAdvice sea pionera y represente a la Universidad de Alicante.

---

<sup>1</sup> <https://www.google.es>

<sup>2</sup> <https://www.appstore.com/>

<sup>3</sup> <https://play.google.com/store>

### 3. Objetivos

El objetivo principal del proyecto es crear una herramienta que sea capaz de facilitar al usuario la tarea de elegir titulación universitaria que quiera cursar, mediante el uso de un sistema de recomendación.

Para hacer efectivo el propósito de crear una herramienta online y que sea exitosa entre los alumnos de la Universidad de Alicante, hace falta diseñar un modelo de negocio (explicado en mayor detalle en el punto 4.1 del documento). Este modelo se refleja mediante una representación gráfica de un análisis de competencias y organizacionales de una empresa o startup (véase el glosario de términos). Ésta representación se puede ejemplificar en una técnica bautizada con el nombre de *Model Lean Canvas*, y es una plantilla de modelo de negocio que pretende evaluar estratégicamente los puntos a tener en cuenta para que gradUAdvice tenga éxito en el mercado.

Otro de los objetivos es obtener provecho de los datos abiertos que gratuitamente proporciona la Universidad de Alicante para el desarrollo de aplicaciones web o móviles. Información que alimenta el sistema de recomendación basado en contenido que se quiere implementar, y con ello devolver al usuario, un potencial alumno de la universidad, un resultado que optimice y dirija sus necesidades hacia un futuro mejor.

Este trabajo acapara tres ámbitos globales como son:

- la **investigación** en los sistemas de recomendación y open data,
- el **análisis** de las competencias de mercado desde el enfoque empresarial y de negocio,
- y el **desarrollo** del sistema en conjunto para obtener una herramienta útil.

En resumen, la aplicación web que se propone, quiere garantizar que un alumno pueda elegir correctamente una titulación universitaria de su gusto. Y además, nutrir a la universidad con datos nuevos a utilizar en otra aplicaciones que consideren oportunas<sup>4</sup>.

---

<sup>4</sup> Los resultados resultado de ejecutar gradUAdvice y finalizar el proceso de selección de una titulación, se almacenan como fuente de información relevante y enriquecida para otros proyectos que se deseen desarrollar.

## 4. Metodología

En esta sección se hablará del cuerpo del documento como tal, tratar todos los puntos comentados como el **modelo de negocio**, como se tratan los **datos abiertos** en gradUAdvice, desvelar y analizar los resultados del **test inicial** realizado, la implementación del **algoritmo del sistema de recomendación** y las iteraciones que conlleva a devolver un resultado u otro, como se ha **desarrollado la aplicación** en sí con las tecnologías mencionadas y finalmente la **ejecución** y explicación de la funcionalidad de la aplicación web.

A modo de resumen global del sistema, se expone como gráficamente se representaría el funcionamiento de gradUAdvice y qué componentes hacen que funcione como un único bloque:

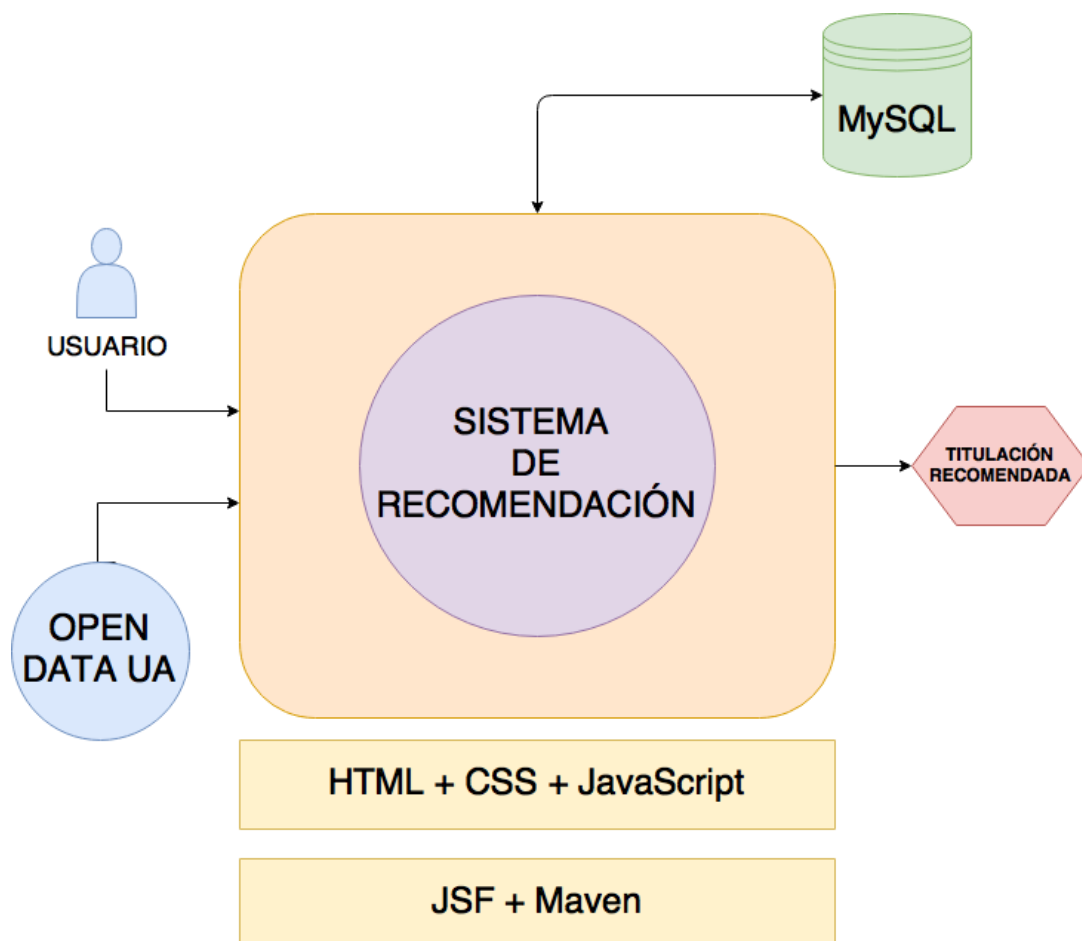


Figura 6: Esquema global de gradUAdvice

## 4.1. Modelo de negocio

Una parte muy importante del análisis a la hora de realizar cualquier inversión ya sea un producto o un servicio, es tener un modelo de negocio bien estructurado.

El modelo de negocio es una representación tanto gráfica como textual utilizada para un amplio conjunto de descripciones que representan aspectos centrados de un negocio, entre ellos: el propósito, el proceso de negocio, el mercado objetivo, las ofertas, las estrategias, la infraestructura, las estructuras organizacionales, las prácticas de comercio, así como los procesos operacionales y las políticas. Con el cual un negocio se sustenta en busca de una fuente de ingresos y beneficios.

A lo largo de la historia, se han ido implementando los modelos de negocio como diseños de las estructuras organizacionales para aprovechar una oportunidad comercial [5].

Las plantillas más conocidas a la hora de realizar un modelo de negocio, en cualquier búsqueda que se pueda realizar en la WEB 2.0., son dos: Business Lean Canvas y Model Lean Canvas. (en el Anexo I se explica con más detalle la diferencia).

En este caso se ha decidido por el Model Lean Canvas ya que se ajusta en mayor medida al proyecto desarrollado como gradUAdvice.

El diseño del sistema quedaría como se muestra más abajo en la tabla 1, mientras que la explicación está en los puntos siguientes. Es conveniente especificar que al ser un Trabajo Final de Master (*TFM*), el proyecto no contempla el aspecto financiero, por lo que se ha obviado.

<p><b>PROBLEM</b></p> <p>Carrera profesional poco clara</p> <p>Malas elecciones de titulación</p> <p>Falta de asesoramiento personalizado</p> <p>No lectura de los planes de estudio por parte de los usuarios</p> <p>Folletos de carreras breves o muy sobrecargados de información</p> <p>Mucha Información en internet y muy tediosa</p>	<p><b>SOLUTION</b></p> <p>Solventar el déficit de aprobados</p> <p>Reducir el número de abandonos</p> <p>Éxito en la carrera profesional</p> <p>Metodología en forma de juego para auto-guiarse</p> <p>Los datos generados se guardan para la retroalimentación de otros proyectos de la Universidad (Machine Learning)</p>	<p><b>UNIQUE VALUE PROPOSITION</b></p> <p>Saber cual es la titulación que más se acerca a su propósito según sus gustos</p> <p>Evitar la decepción o falta de motivación por la falta de determinación</p> <p>Auto-concienciar el alumno a priori de la titulación que más se ajusta a él</p> <p>Alternativa al asesoramiento humano</p>	<p><b>UNFAIR ADVANTAGE</b></p> <p>Servicio intuitivo sin el esfuerzo de leer todos los planes de estudio</p> <p>Herramienta desarrollada personalizada para la Universidad de Alicante</p> <p>Retroalimentable y escalable según la información con la que se alimenta la aplicación</p> <p>Personalizada y refinada con el tiempo</p>	<p><b>CUSTOMER SEGMENTS</b></p> <p>Alumnos con PAU finalizadas</p> <p>Usuarios con intención de matriculación en la Universidad</p>
<p><b>KEY METRICS</b></p> <p>Alumnos registrados</p> <p>Alumnos que han finalizado el proceso</p> <p>Opinión positiva</p>				
<p><b>COST STRUCTURE</b></p> <p>No contempla</p>		<p><b>REVENUE STREAMS</b></p> <p>No contempla</p>		

Tabla 1. Modelo Lean Canvas

A continuación se explica punto a punto los apartados de forma detallada de la plantilla Model Lean Canvas expuesta en la tabla 1.

#### **4.1.1. Segmento de clientes**

El usuario objetivo son aquellos futuros estudiantes de la Universidad de Alicante que estén dudosos a la hora de elegir la titulación que quieren realizar.

Dentro de este amplio abanico, son particularmente interesantes aquellos estudiantes que hayan finalizado sus estudios de bachillerato y realizado las PAU (Pruebas de Acceso a la Universidad). Esto es debido a que parte de los datos necesarios para el proceso de descarte los proporcionan los propios alumnos (asignaturas con mejor nota, preferencias según gustos de las asignaturas de bachillerato, etc.)

También puede albergar otros usuarios, que se quedarían con una probabilidad un tanto más reducida a la hora de obtener resultados finales óptimos.

#### **4.1.2. Problema**

Muchos jóvenes llegan a aprobar las PAU sin tener muy clara la carrera que quiere desempeñar, de hecho, muchos de los cambios de titulación universitaria se efectúan debido a una mala elección al principio del propio estudiante. Este hecho se debe al escueto asesoramiento que puede recibir y no conocer muy a fondo las bases de las titulaciones a las que quiere aspirar.

Pocos estudiantes han llegado a leerse por completo los planes de estudio de las carreras que mas le interesan, eso si ha llegado a plantearse dos, tres o más carreras a seleccionar. Los folletos repartidos por las universidades suelen ser muy breves o muy sobrecargados, y las descripciones de tanto la titulación como las asignaturas de la misma suelen incluir contenidos muy específicos y que no suelen ayudar a la comprensión del usuario.

El anterior es en el mejor de los casos, pero si el estudiante aún no sabe por cual rama (p. ej. ciencias específicas o ciencias generales) decantarse, el problema es aún mayor ya que tiene que buscarse la vida con la información que le den amigos o familiares, internet o la misma universidad. Esta cantidad información suele ser abrumadora y no siempre viene de la mano del mejor resultado.

#### **4.1.3. Propuesta de valor**

Con gradUAdvice el usuario, en este caso un futuro estudiante de la Universidad de Alicante, pueda de antemano saber cual o cuales son las titulaciones que más se acercan a su gusto. Consiguiendo así que una vez titulado, no caiga en la espiral de decepción típica o pérdida de la motivación, y además evite, entre otros factores negativos, el perder un curso escolar por falta de determinación.

De esta forma, el estudiante puede auto-concienciarse de cuales son sus capacidades y gustos personales sin tener que acudir al asesoramiento humano de la universidad y/o los programas de acción tutorial para orientarse.

#### **4.1.4. Solución**

El gran objetivo de GradUAdvice es el de aumentar el número de aprobados en las titulaciones universitarias, además de asegurar un éxito en la elección del futuro profesional y vocacional de cada alumno. Una herramienta con la que pueda orientarse a la hora de seleccionar la titulación universitaria que más le convenga.

Mediante una metodología en forma de procesamiento de selección o descarte, los futuros estudiantes pueden auto-guiarse para saber cuales son sus gustos, metas y vocación profesionales.

El procedimiento que haya seguido cada alumno se guarda de tal forma que sirve para “alimentar” otros proyectos futuros de la Universidad. Además del cuestionario inicial con el cual cada alumno aporta datos anónimos.



#### **4.1.5. Canales de distribución**

Como canal principal de distribución, el uso de las redes sociales con el perfil de la universidad en cuestión. Las redes sociales son el método de comunicación más efectivo y más barato para llegar a los estudiantes y cualquier público en general.

La Universidad de Alicante se encargaría de publicar en las redes sociales más populares (Facebook, Twitter, Instagram, etc...) para que el usuario objetivo se percate de la existencia de la herramienta y le de el uso que crea conveniente.

El otro canal a tener en cuenta es el propio canal oral. La difusión en las excursiones que realizan los centros de educación secundaria a la Universidad para informarse sobre las titulaciones y salidas profesionales. Además de la divulgación por parte de los gestores en las secretarías de los centros que constituyen el campus.

#### **4.1.6. Fuente de ingresos**

No contempla

#### **4.1.7. Estructura de costes**

No contempla

#### **4.1.8. Métrica clave**

Las tres métricas más significativas para el proyecto serían las siguientes:

- Alumnos registrados: hacer un análisis en la base de datos de cuantos alumnos han finalizado del cuestionario inicial.
- Alumnos que han finalizado el proceso: análisis en la base de datos de cuantas soluciones se han proporcionado y cuantos alumnos han quedado satisfechos.
- Opinión positiva: al finalizar la prueba, se proporciona un apartado para opiniones, habría que examinar y distinguir entre las consideraciones positivas y negativas.

#### **4.1.9. Ventaja diferencial**

GradUAdvice ofrece un servicio lo más intuitivo posible al alumno sin el exhausto esfuerzo de recorrer todos los planes de estudios para acercarse a una titulación que le pueda interesar. De este modo, es la primera herramienta desarrollada en y para la Universidad de Alicante que ofrece esta posibilidad.

Este es el análisis realizado a priori, es posible que en un futuro, y debido a futuras mejoras, la ventaja diferencial se vea modificada y refinada derivando de la demanda más específica y especializada de los alumnos.

#### **4.2. Datos abiertos**

Uno de los objetivos principales de este trabajo es la obtención y uso de los datos que proporciona la Universidad de Alicante mediante su API DatosUA (<http://datos.ua.es/es>).

Entre el catálogo de datos que ofrece la API de la Universidad de Alicante, se realizó una petición expresa para liberar<sup>5</sup> la información en referencia a las titulaciones universitarias, en este caso grados para poder tratarla en esta aplicación.

En un principio la idea era obtener el máximo número de información disponible para ir filtrando y moldeando según sea conveniente, pero por temas burocráticos solo se ha podido obtener la información en referencia a la lista de asignaturas de los grados universitarios de la UA<sup>6</sup>.

Para la aplicación, se ha procedido a la adquisición de una API key rellenando el formulario con los datos del alumno (nombre y email), para así poder realizar peticiones al servidor y obtener los datos necesarios (<https://dev.datos.ua.es/apikey.html>).

---

<sup>5</sup> No todos los datos de la Universidad son ofertados por la misma, en caso de que no existan de entre el catálogo, se pueden pedir y una vez aprobada la petición, se ponen a disposición del público.

<sup>6</sup> Se puede consultar el listado de asignaturas en este link: <http://datos.ua.es/es/ficha-datos.html?idDataset=1004>

Desde la aplicación se ha diseñado un algoritmo que es capaz de obtener los datos en formato JSON, para luego tratarlos y posteriormente almacenarlos en base de datos, que es de donde se consumirán para el sistema de recomendación.

Un ejemplo de la sintaxis que devuelve el standard JSON para mostrar las asignaturas a la petición realizada es la siguiente:

```
{
  "crdtsects": "6",
  "nomasicorto": "MATEMATICAS 1",
  "crdtspra": "0",
  "ofertada": "S",
  "NOMEST": "GRADO EN INGENIERÍA INFORMÁTICA",
  "estadoasi": "V",
  "tfg": "N",
  "crdtsteo": "0",
  "CODCEN": "3013273",
  "nomasi": "MATEMATICAS 1",
  "complform": "N",
  "duracion": "S",
  "codasi": "34002",
  "CODEST": "C203",
  "caca": "2015-16",
  "tipo": "T",
  "docencia": "S",
  "ciclo": "9",
  "descurso": "1",
  "SUBTIPEST": "GRA"
}
```

Tabla 2: Ejemplo JSON de la API de datos UA

De esta cantidad de datos, se han elegido los siguientes para formar parte de los registros en BD:

- **NOMEST**: nombre de la titulación universitaria
- **tfg**: si la asignatura es un Trabajo Final de Grado (S/N).
- **CODCEN**: código del centro o facultad de la UA.
- **nomasi**: nombre de la asignatura
- **codasi**: código de la asignatura
- **tipo**: tipo de la asignatura (T: troncal, B: básica y O: optativa).

Los demás datos pueden ser o no relevantes a la hora de tratarlos en el SR según convenga, es decir, este trabajo final de master tratará estos datos, aunque los más importantes sean el nombre de la asignatura y el grado al que pertenecen, los otros atributos aportados, sirven para clasificar y/o recomendar posteriormente como se explicará más adelante.

### **4.3. Test de preguntas inicial**

Para asegurar que existe una oportunidad real en el mercado para la aplicación, se ha realizado un estudio de mercado desarrollando un pequeño cuestionario. La finalidad del mismo es hacerse a la idea de cuales son las vías de estudio de bachillerato que más predominan entre los usuarios muestreados, y si con la llegada de una herramienta online que les facilite la elección de una carrera universitaria, la usarían.

La encuesta se puede encontrar en el siguiente enlace: <https://goo.gl/forms/FOfG9kI5qsSmKO2d2>

Cabe mencionar que el estudio fue respondido por un total de 88 personas.

El periodo en el que estuvo la encuesta online ha sido desde el 03/06/2016 hasta el 16/06/2016, fecha en el que el último cuestionado aportó su granito de arena. Es importante destacar que las PAU (Pruebas de Acceso a la Universidad) se realizaron entre el 6 y el 8 de junio de 2016.

Los resultados se exponen de forma más detallada en el Anexo II del documento.

En reglas generales, la encuesta podría haberse refinado un poco más como se han ido comentando en las pequeñas reflexiones a cada gráfica. Pero en cómputo general, la iniciativa de una herramienta online para recomendar titulaciones universitarias sería aceptada a priori. De esta forma, y aunque la cantidad de alumnos muestreada sea pequeña, se puede llegar a la conclusión de que es necesaria cualquier herramienta o aplicación que de cobertura en los asuntos universitarios.

Por lo tanto, como primera aproximación, el estudio de mercado ha resultado ser un éxito favorable.

## 4.4. Desarrollo del sistema de recomendación

Para la implementación del sistema de recomendación, se ha optado por la librería de código abierto Apache Lucene<sup>7</sup>. Una herramienta de recuperación de información originalmente desarrollada en Java e implementada para otros lenguajes como C#, Python, Ruby, PHP, entre otros.

De entre los varios proyectos que engloba Apache Lucene, se ha optado por Apache Lucene Core<sup>8</sup>, una librería java especializada en ser un motor de búsqueda de texto.

El funcionamiento de Lucene es metódico y empieza siempre por el mismo paso: Indexar para poder tratar los documentos a tratar en la solución. Existen dos formas de hacer esto, almacenar el índice en memoria o en almacenamiento masivo como puede ser el disco duro del ordenador. Se ha optado por la segunda debido a que se hacen varias iteraciones y podría ser necesario acudir a alguna indexación previa.

Se realizan varias indexaciones porque, y como se explicará más adelante, el proceso de selección de la titulación universitaria se realiza por descarte o poda. Por lo tanto, es necesario volver a indexar aquellos documentos que no han sido descartados, para que el proceso sea más fiable y los resultados mas acotados.

La información que Lucene indexa en este caso el contenido de los documentos y la frecuencia de las palabras que aparecen en el documento, ya que es la información más relevante. Se descarta almacenar información como la localización del archivo original y el directorio del índice (datos ya disponibles en un fichero de configuración en java).

La configuración del indexador queda de la siguiente forma:

```
Document doc = new Document();
FieldType fieldType = new FieldType();
fieldType.setIndexed(true);
fieldType.setIndexOptions(FieldInfo.IndexOptions.DOCS_AND_FREQS_AND_POSITIONS_AND_OFFSETS);
fieldType.setStored(true);
fieldType.setStoreTermVectors(true);
fieldType.setTokenized(true);
Field contentField = new Field(fieldName, getAllText(listFiles[i]), fieldType);
doc.add(contentField);
writer.addDocument(doc);
```

Figura 7: Indexador de Lucene

---

<sup>7</sup> <https://lucene.apache.org>

<sup>8</sup> <https://lucene.apache.org/core/>

El siguiente paso en la ejecución consiste en preparar el índice para ser leído. Obtener la cuenta de términos, frecuencias (mediante TF-IDF) y número total de documentos indexados. Posteriormente se obtienen los términos almacenados en el índice para generar los vectores a tratar en el sistema de recomendación.

## **TF-IDF**

Term Frequency - Inverse Document Frequency [6] son las siglas en inglés para la frecuencia de término - frecuencia inversa de documento, o sea, la frecuencia de ocurrencia del término en la colección de documentos. Esto es una medida numérica que expresa la relevancia de una palabra para un documento en una colección de los mismos.

El valor tf-idf crece proporcionalmente al número de veces que un término aparece en el documento, pero es compensada por la frecuencia de la palabra en un conjunto de documentos, en este caso, se entiende por conjunto de documentos las titulaciones universitarias.

La fórmula de obtención del valor TF-IDF consta de tres pasos:

1. Obtención de la frecuencia de un término en un documento.

$$TF(t, d) = \frac{\text{Número de apariciones del término en el documento}}{\text{Número de documentos que incluyen el término}}$$

2. Obtención de la frecuencia inversa de documento:

$$IDF(t, D) = \log \frac{\text{Número total de documentos}}{\text{Número de documentos que incluyen el término}}$$

3. Obtención del valor TF-IDF:

$$TFIDF(t, d, D) = TF(t, d) * IDF(t, D)$$

Una vez obtenidos los términos y la frecuencia con la que aparece cada con respecto al documento y a la colección de documentos, se procede a la comparación mediante la similitud coseno de todos los documentos entre sí.

Una distancia empleada asiduamente en la búsqueda y recuperación de información representando documentos en un espacio vectorial [7].

## Similitud coseno

La similitud coseno es una medida que se obtiene como resultado de una semejanza entre dos vectores en un espacio que posee un producto interior con el que se evalúa el valor del coseno del ángulo comprendido entre dichos vectores. Esta función devuelve un valor igual a 1 si el ángulo comprendido entre dos vectores es cero, es decir, si dos vectores son idénticos el resultado es 1.

En caso de que el ángulo de coseno difiera, el resultado sería inferior a uno. Entonces, la métrica se comprende en el rango [0,1].

$$\text{distance} = \frac{\cos^{-1}(\text{similarity})}{\pi}$$
$$\text{similarity} = 1 - \text{distance}$$

Cabe mencionar que el algoritmo desarrollado con Lucene proporciona la similitud o afinidad no solo un documento con los demás, sino el de todos los documentos entre sí también, es decir, comparar el primer grado con todas las 42 demás, posteriormente el segundo grado con todas las demás, y así hasta tener una tabla de afinidades de 1764 elementos.

En la siguiente tabla se puede ver la similitud coseno entre el documento 0 que equivale al “Grado en Administración y Dirección de Empresas” con las otras 42 titulaciones totales (incluyendo la comparación del grado consigo mismo):

Grado en Administración y Dirección de Empresas = 1.0
Grado en Arquitectura Técnica = 0.8276325888794807
Grado en Arquitectura = 0.48664400746395897
Grado en Biología = 0.6128235192938828
Grado en Ciencias de la Act. Física y del Deporte = 0.6094862770989284
Grado en Ciencias del Mar = 0.6915234625049057
Grado en Criminología = 0.720783223773775
Grado en Derecho = 0.664337658551351
Grado en Economía = 0.816314562711601
Grado en Enfermería = 0.5980596373354633
Grado en Español/Lengua y Literaturas = 0.7289092523524464
Grado en Estudios Árabes e Islámicos = 0.6605664796079234
Grado en Estudios Franceses = 0.6779433308288187

Grado en Estudios Ingleses = 0.42995374166385936
Grado en Fundamentos de la Arquitectura = 0.7234650811616876
Grado en Geografía y Ordenación del Territorio = 0.5355232910177031
Grado en Geología = 0.7517204854176754
Grado en Gestión y Administración Pública = 0.7597311249144158
Grado en Gestión y Administración del Territorio = 0.7927788973115265
Grado en Historia = 0.7384542932052708
Grado en Humanidades = 0.7949805302658688
Grado en Ingeniería Civil = 0.8087641834369984
Grado en Ingeniería en Sonido e Imagen en Telecomunicación = 0.8436514322762538
Grado en Ingeniería Informática = 0.7320321978575125
Grado en Ingeniería Multimedia = 0.824774852725529
Grado en Ingeniería Química = 0.7759137656926319
Grado en Ingeniería Robótica = 0.7711440751058524
Grado en Maestro de Educación Infantil = 0.8033808960330673
Grado en Maestro de Educación Primaria = 0.7484819372705457
Grado en Matemáticas = 0.7766685211339833
Grado en Nutrición Humana y Dietética = 0.3814078440066919
Grado en Óptica y Optometría = 0.8650903633121879
Grado en Publicidad y Relaciones Públicas = 0.5464616686807686
Grado en Química = 0.8175962234882237
Grado en Relaciones Laborales y Recursos Humanos = 0.8436736436448765
Grado en Sociología = 0.8234152549929838
Grado en Tecnologías de la Información para la Salud = 0.38898445010055266
Grado en Filología Catalana = 0.1231307284779798
Grado en Trabajo Social = 0.6617787555909379
Grado en Turismo = 0.6932686189882021
Programa Conjunto Decrim. Grado en Derecho y en Criminología = 0.743410628229839
Programa Conjunto DADE/Derecho y Adm. y Dir. de Empresas = 0.8803005564180784
Programa Conjunto TADE/Turismo y Adm. y Dir. de Empresas = 0.8009808375889549

Tabla 3: Similitudes coseno del Grado en Administración y Dirección de Empresas



Con la lista de afinidades total obtenida, el objetivo es clasificar las titulaciones en dos grandes grupos de términos más relevantes. Se construye extrayendo las dos titulaciones que tengan la tasa de similitud más baja entre si, esto es el valor más bajo de similitud coseno entre dos documentos de toda la colección, de esta forma se puede iniciar la clasificación.

El procedimiento parte de dos documentos y se van comparando los siguientes 41. Teniendo el documento o titulación A y la titulación B, el documento siguiente a compararse C, si la afinidad del documento C con A es mayor que con B, entonces C pertenece al grupo de A, y así hasta tener dos montículos claros.

Es así como se consigue la clasificación en la primera iteración del algoritmo de recomendación (tal como se expresa gráficamente en la figura 8), y poder exponerlos en la interfaz para que el usuario elija entre un grupo de términos más relevantes u otro.

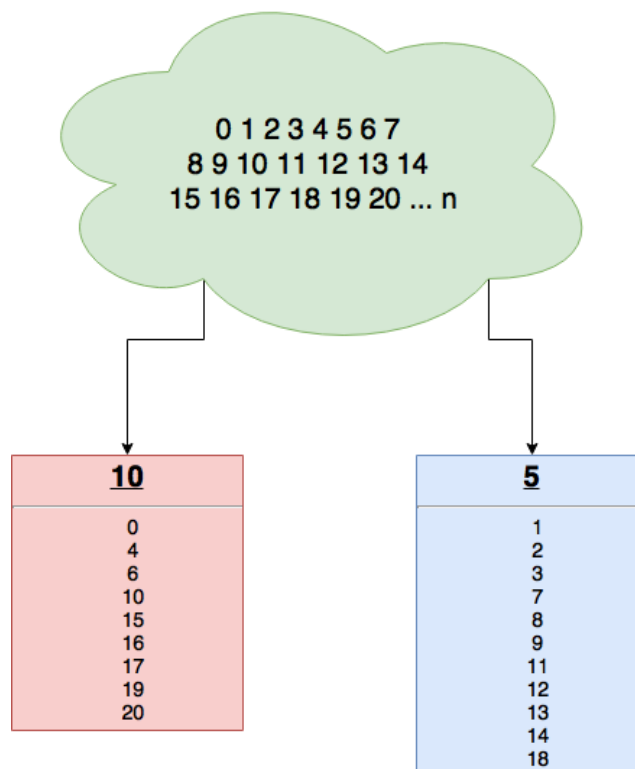


Figura 8: Ejemplo de selección con Lucene

En la figura anterior se observa que hay dos grandes grupos de un hipotético documento 10 y otro documento 5. Ambos tienen una lista de documentos que son más afines a ellos que al compañero. De esta mencionada lista se obtiene el vector de términos.

Entonces, se construye “la nube de términos” de cada grupo y se exponen en la interfaz para que el usuario se decante por una o por la otra. En cualquiera de los dos casos, el algoritmo de recomendación volvería a empezar indexando, obteniendo los ficheros pero solo los seleccionados (p.e.: se ha seleccionado la parte izquierda, se indexan los documentos: 0, 4, 6, 10, 15, 16, 17, 19 y 20), y volviendo al proceso de dos nuevos montículos.

## 4.5. Diseño de la aplicación

Como ya se menciona anteriormente, gradUAdvice al crearse para ser una herramienta online, se ha decidido por una implementación web. Dicho esto, la tecnología o framework utilizado para la parte de interfaz es HTML5 con los estilos de CSS3 y limadas las interacciones con JavaScript. Mientras que para la parte del servidor se ha elegido JavaServlet Faces (JSF) dado que permite la ejecución de código java, que es el lenguaje de programación más conocido y del cual se puede obtener mayor cantidad de información<sup>9</sup>. Además, es una tecnología que en el mercado laboral destaca bastante y las empresas que desarrollan en plataformas J2EE (también conocida como Java Platform o JavaEE -Java Enterprise Edition-), implementan JSF u Oracle ADF<sup>10</sup> para las aplicaciones web. Una de las ventajas de usar JSF es que el framework en sí, permite la navegabilidad entre páginas, manejar eventos desde el servidor con Java, validaciones a nivel de servidor y lo más importante: permite desarrollar interfaces para aplicaciones web.

---

<sup>9</sup> <https://hipertextual.com/archivo/2014/10/lenguajes-programacion-mas-populares/>

<sup>10</sup> Oracle Application Development Framework, un framework comercial desarrollado por Oracle para la creación de aplicaciones empresariales basada en patrones de diseño listos para usar y a su vez basada en J2EE.  
<http://www.oracle.com/technetwork/developer-tools/adf/overview/index.html;jsessionid=6R39V8WhqTQ7HMb2vTQTkzbP5XRFgs4RQzyxQ7fqxH9y6p6vKXk4!-460884186>

A modo de resumen, para el diseño de gradUAdvice se han desarrollado las siguientes facetas:

- **Front-end o lado del cliente** diseñado en *JSF*<sup>11</sup> 2.2 (*JavaServer Faces*) y *HTML5*<sup>12</sup> (*HyperText Markup Language*) con ayuda de hojas de estilo *CSS3*<sup>13</sup> y el lenguaje de programación interpretado *JavaScript*<sup>14</sup> para algunas secciones funcionales.
- **Back-end o lado del servidor** programado en el lenguaje de programación *Java* junto a *servlets* en el mismo lenguaje y para la construcción del proyecto *Apache Maven* [8].
- **SGDB (Sistema Gestor de Bases de Datos)** o simplemente base de datos, se ha optado por *MySQL*<sup>15</sup>.

Por lo tanto, la tecnología usada para la aplicación es una tecnología actual, respaldada por el ámbito profesional y laboral, con mucha información disponible en la web y tiene una comunidad de desarrolladores lo suficientemente amplia y además, está hecha para poder escalar en cualquier momento.

#### 4.5.1. Diagrama de flujo principal

El diagrama de flujo de la aplicación, como bien se explica en la figura 9, presenta un inicio donde arrancaría la página principal de gradUAdvice. Una vez en ese estado, se presentan dos opciones: cargar los datos de la API de datosUA o empezar con el cuestionario inicial. Al terminar el cuestionario el proceso de selección de titulación empezaría. Un proceso que está desarrollado a modo de juego: unas iteraciones a realizar de descarte marcadas por nubes de tags<sup>16</sup>. Se realizarán tantas iteraciones como sean necesarias para llegar a la titulación recomendada por el sistema. Finalmente se propone al usuario que introduzca una opinión constructiva de lo que le ha parecido la aplicación y finalmente se le expondrán los resultados.

---

<sup>11</sup> [https://es.wikipedia.org/wiki/JavaServer\\_Faces](https://es.wikipedia.org/wiki/JavaServer_Faces)

<sup>12</sup> [http://www.w3schools.com/html/html5\\_intro.asp](http://www.w3schools.com/html/html5_intro.asp)

<sup>13</sup> [https://en.wikipedia.org/wiki/Cascading\\_Style\\_Sheets#CSS\\_3](https://en.wikipedia.org/wiki/Cascading_Style_Sheets#CSS_3)

<sup>14</sup> <https://es.wikipedia.org/wiki/JavaScript>

<sup>15</sup> <https://www.mysql.com>

<sup>16</sup> Nube de tags: representación visual de las palabras que conforman un texto, en donde el tamaño es mayor para las palabras que aparecen con más frecuencia. [https://es.wikipedia.org/wiki/Nube\\_de\\_palabras](https://es.wikipedia.org/wiki/Nube_de_palabras)

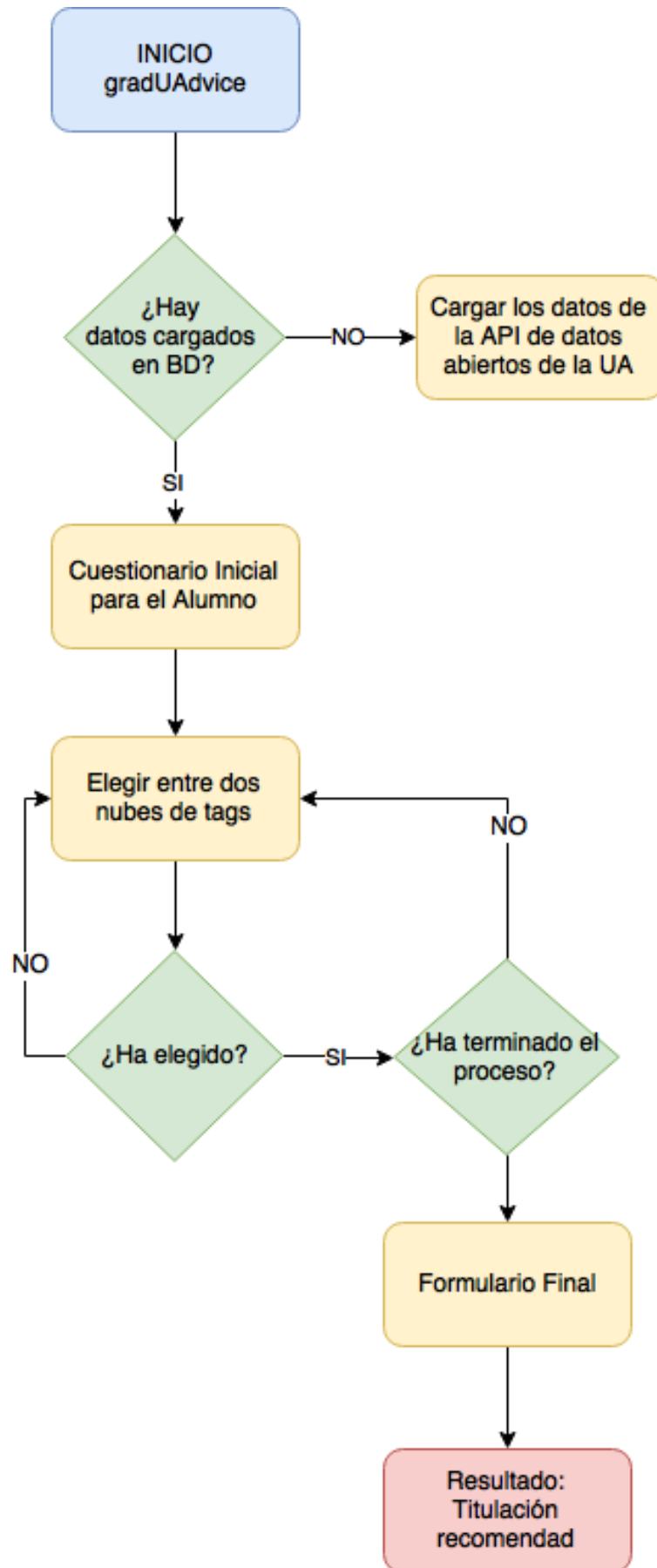


Figura 9: Diagrama de flujo

## 4.5.2. Página inicial

En la página inicial o master page de la aplicación (véase la figura 10) se verá un contenedor principal que hace de introducción a la funcionalidad. Existen dos posibles salidas:

- **Empezar:** redirecciona al cuestionario inicial de la herramienta (punto 5.4 de la memoria) y así empezar el proceso de selección.
- **Actualizar la BD del Portal de Datos Abiertos:** (punto 5.3 de la memoria) este proceso carga los datos JSON recibidos del portal de datos abiertos de la Universidad de Alicante en la base de datos, para luego poder alimentar el sistema de recomendación. En caso de primera ejecución y la BD esté vacía, es necesario pasar por este paso. En caso contrario, actualizaría los registros.

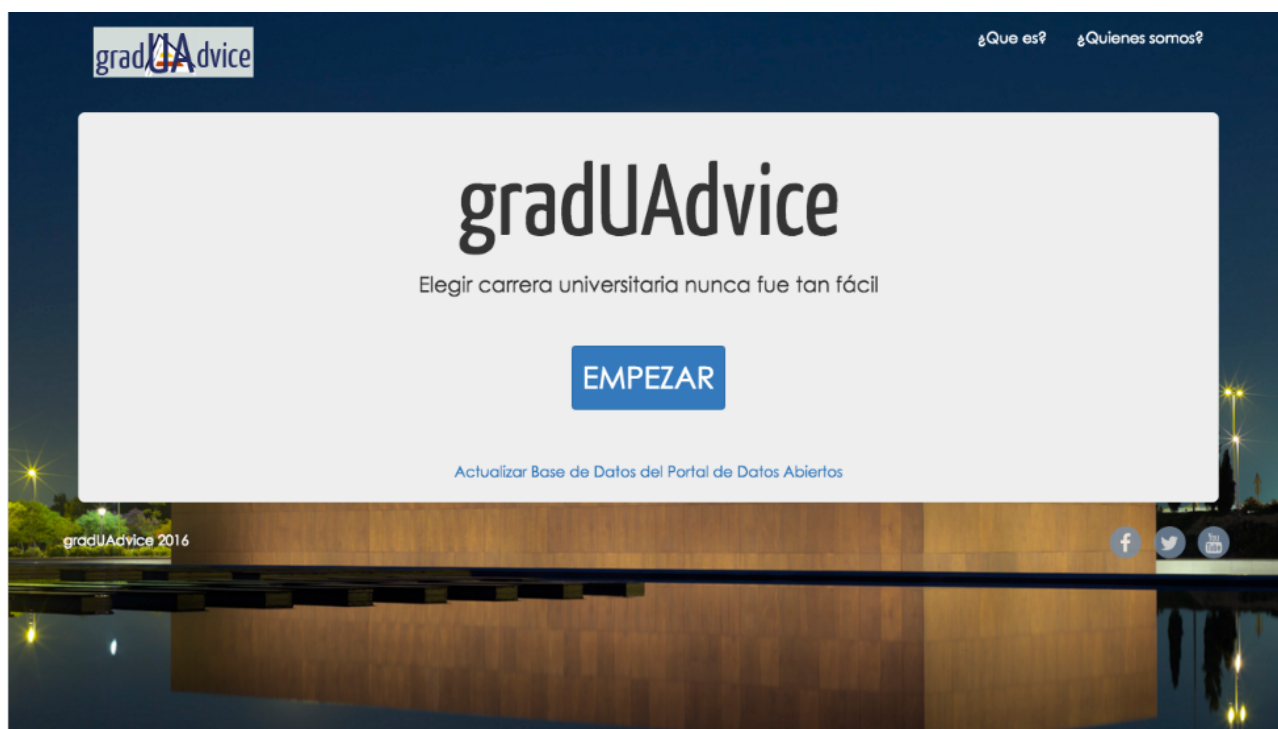


Figura 10: Pagina principal

### 4.5.3. Carga de datos inicial

```
INFORMACIÓN: Server startup in 3837 ms
iniciando sesión en gradUadvice
Leyendo datos de la API...
Se han devuelto 4470 registros del JSON.
Borrando los registros de ASIGNATURAS...
Introduciendo datos en BD...
Se han introducido 2629 asignaturas en BD
Creando ficheros de grados...
Redireccionando a la página principal...
```

Figura 11: Ejemplo de respuesta servidor

Una vez iniciada la aplicación y seleccionado la opción 2 del punto anterior, procede:

1. **Lectura de datos de la API:** se hace una consulta de los datos a la API de la universidad.
2. **Obtención de los resultados:** los datos se transforman para que la aplicación pueda utilizarlos.
3. **Borrar los registros de BD:** en caso de existir datos ya en la tabla en base de datos, se borrarían.
4. **Introducir los nuevos registros en BD:** cargar los nuevos registros en la base de datos.
5. **Crear ficheros de asignaturas:** para que el sistema de recomendación utilice el contenido de las titulaciones -sus asignaturas-, se crean ficheros por cada titulación, en concreto 43 (véase Anexo II para la lista de grados de la Universidad de Alicante).
6. **Redireccionar a página principal:** vuelta a la página principal para poder empezar el proceso.

### 4.5.4. Cuestionario para el alumnado

Uno de los objetivos de gradUAdvice es proporcionar datos relevantes para posibles futuros proyectos de la Universidad de Alicante. Con este cuestionario inicial, a parte de servir para detectar los gustos de los potenciales alumnos universitarios, sirven para ser almacenados y tratados.

Los datos más relevantes, manteniendo el anonimato lo más cuidadosamente posible son:

- El género biológico
- la edad en intervalos: 18 a 21, 22 a 25 y mayor de 25.
- En caso de haber realizado las PAU este año
- El tipo de bachillerato que ha realizado
- Un número de asignaturas ordenadas por preferencia que más le han gustado de bachillerato

gradUAdvice ¿Que es? ¿Quienes somos?

## Nuevo Estudiante

**Sexo:**  
 Hombre  Mujer

**Edad:**  
 18 - 21  22 - 25  Mayor de 25

**Has cursado PAU este año?**  
 Sí  No

**Bachillerato:**

Aceptar las condiciones ?

gradUAdvice 2016 f t

Figura 12: Cuestionario inicial

En cuanto a la lista de asignaturas de bachillerato que más le gustaron al alumno, primero ha de elegir un tipo de bachillerato de los tres disponibles (ciencias y tecnología, humanidades y ciencias sociales, y arte). Automáticamente se cargarán dos pequeños contenedores donde el usuario usando la metodología drag&drop (seleccionar un elemento y arrastrarlo al otro contenedor), elige las cuatro asignaturas que más le gustaron. Esto es debido a que basta con cuatro para terminar el estudio y el resultado final al término de la ejecución sea intuitivo.

**Sexo:**  
 Hombre  Mujer

**Edad:**  
 18 - 21  22 - 25  Mayor de 25

**Has cursado PAU este año?**  
 Sí  No

**Bachillerato:**

**Asignaturas de bachillerato de Ciencias y Tecnología:**

- 
- 
- 
- 
- 
- 
- 
- 

Figura 13: Elección del bachillerato

Quedando por ejemplo de esta manera (las asignaturas elegidas pueden ser máximo 4, en caso de elegir más, se descartarán):

18 - 21  22 - 25  Mayor de 25

Has cursado PAU este año?  
 Si  No

Bachillerato:  
Ciencias y Tecnología

Asignaturas de bachillerato de Ciencias y Tecnología:

Física
Dibujo Técnico
Historia de España
Historia de la Filosofía
Lengua Castellana y Literatura
Lengua Cooficial y Literatura
Primera Lengua Extranjera
Matemáticas
Ciencias de la Tierra y Medioambientales
Biología
Química

Figura 14: Lista de asignaturas elegidas

El usuario debe aceptar las condiciones para poder seguir con el proceso de selección de la titulación universitaria que el sistema le recomiende. En la figura se pueden ver las condiciones que aceptaría y por lo tanto acceder a dar sus datos al proyecto:

Condiciones

Este es un sistema de información de la Universidad de Alicante y su uso está permitido a cualquier persona, en adelante usuarios.

El usuario se obliga a utilizar este sistema de información de forma diligente, correcta, lícita y de conformidad a la Ley y al presente aviso. En particular, se compromete a no utilizar los contenidos con una finalidad contraria a la Ley, moral u orden público; a no reproducir o copiar, no distribuir y no permitir el acceso público, a excepción de que cuente con autorización expresa, y en cualquier tipo de soporte, ya sea físico o lógico.

Se recopilarán datos de carácter anónimo para la Universidad de Alicante con los que podrá tratarlos como fuente fiable de información y con ello la obtención de nuevos recursos, de los cuales proporcionar nuevos servicios a la comunidad. Con ello el usuario aceptaría ceder sus datos proporcionados en el cuestionario para dicha finalidad

En virtud de lo anterior, es obligación del usuario notificar de forma inmediata a los gestores del sistema acerca de cualquier hecho que permita el uso indebido de los datos aportados y proporcionados, tales como le robo, extravío, o el acceso no autorizado a los mismos, con el fin de proceder a su inmediata intervención judicial.

Close

Figura 15: Condiciones a aceptar



Una vez pulsado el botón de guardar y almacenado el nuevo usuario en la base de datos, continua el proceso de selección.

#### 4.5.5. Proceso de selección

La siguiente pantalla, como se ve en la figura 16, mostraría la primera iteración del sistema de recomendación. El análisis de todas las titulaciones y clasificación en dos grandes grupos para que el usuario pueda elegir, usando el concepto de *Cloud Tag*<sup>17</sup> para clasificar los dos grupos de términos más relevantes. Una visualización de texto intuitiva y que incluye términos relevantes de todo un texto, con la característica de aumentar de tamaño un término o palabra proporcionalmente a las veces que dicho término aparezca en el texto.

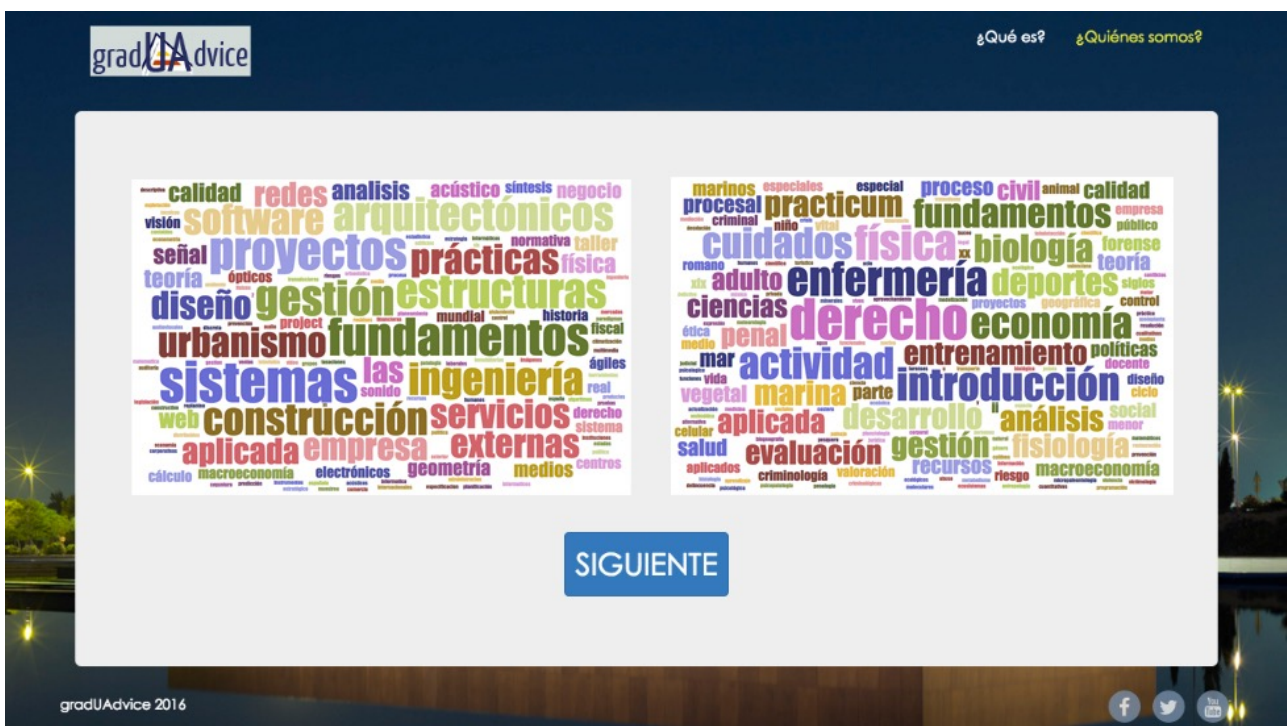


Figura 16: Primera iteración

Y así tantas veces como sea necesario hasta un máximo de 4 (limitado por el programa), en caso de llegar a la 4ª iteración y no obtener una titulación recomendada, el sistema dará como resultado la lista de titulaciones del último grupo elegido.

<sup>17</sup> [https://en.wikipedia.org/wiki/Tag\\_cloud](https://en.wikipedia.org/wiki/Tag_cloud)

Limitar el número de iteraciones a cinco viene de realizar una acotación intermedia, en la que en algunos casos pueda ser un número menor de iteraciones necesarias o mayor:

- En la primera iteración dos grandes grupos de 21 titulaciones.
- La segunda iteración grupos de 11 y 10 titulaciones respectivamente.
- La tercera iteración grupos de 5 y 5 titulaciones.
- La cuarta iteración grupos de 2 y 3 titulaciones

Con esta limitación se asegura que devolverá al menos una o dos titulaciones y dará juego a que el usuario pueda reflexionar sobre ese resultado.

Los grupos de términos se han filtrado por frecuencia de palabras y se han eliminado las palabras redundantes más conocidas como stopwords. Y se ha utilizado una librería de código abierto de JavaScript [9] para el diseño tan llamativo.

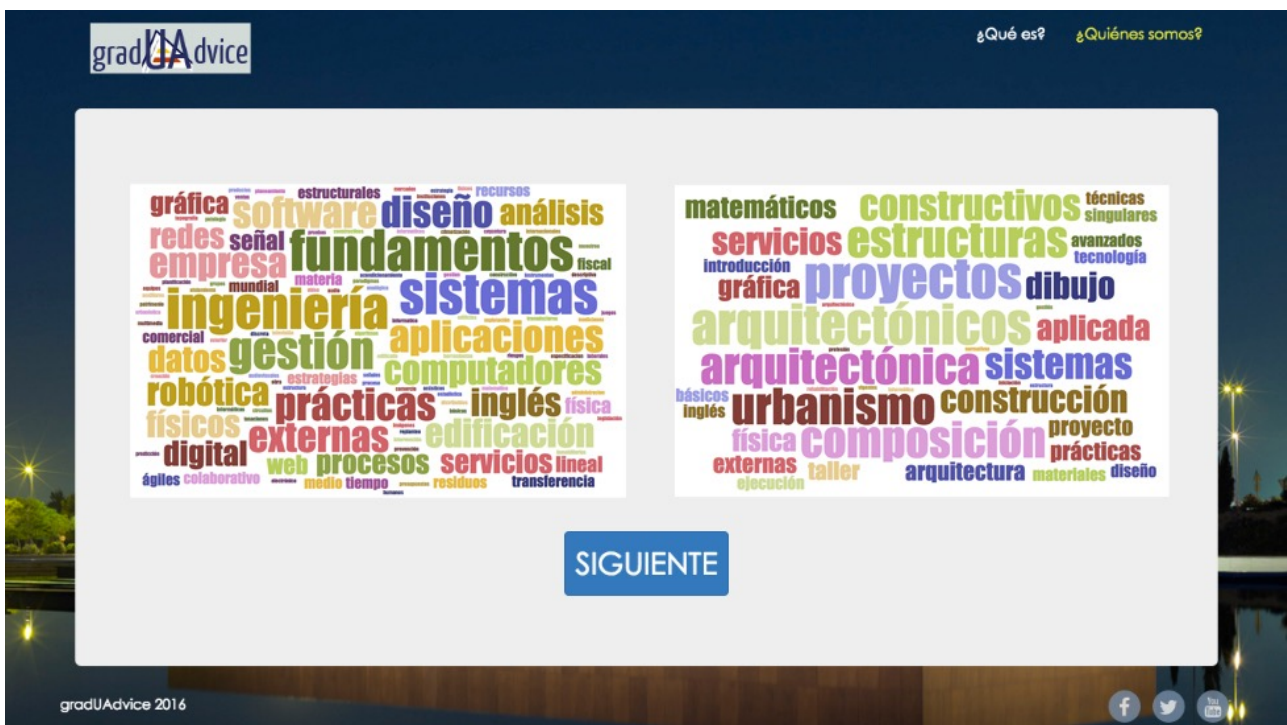


Figura 17: Segunda iteración

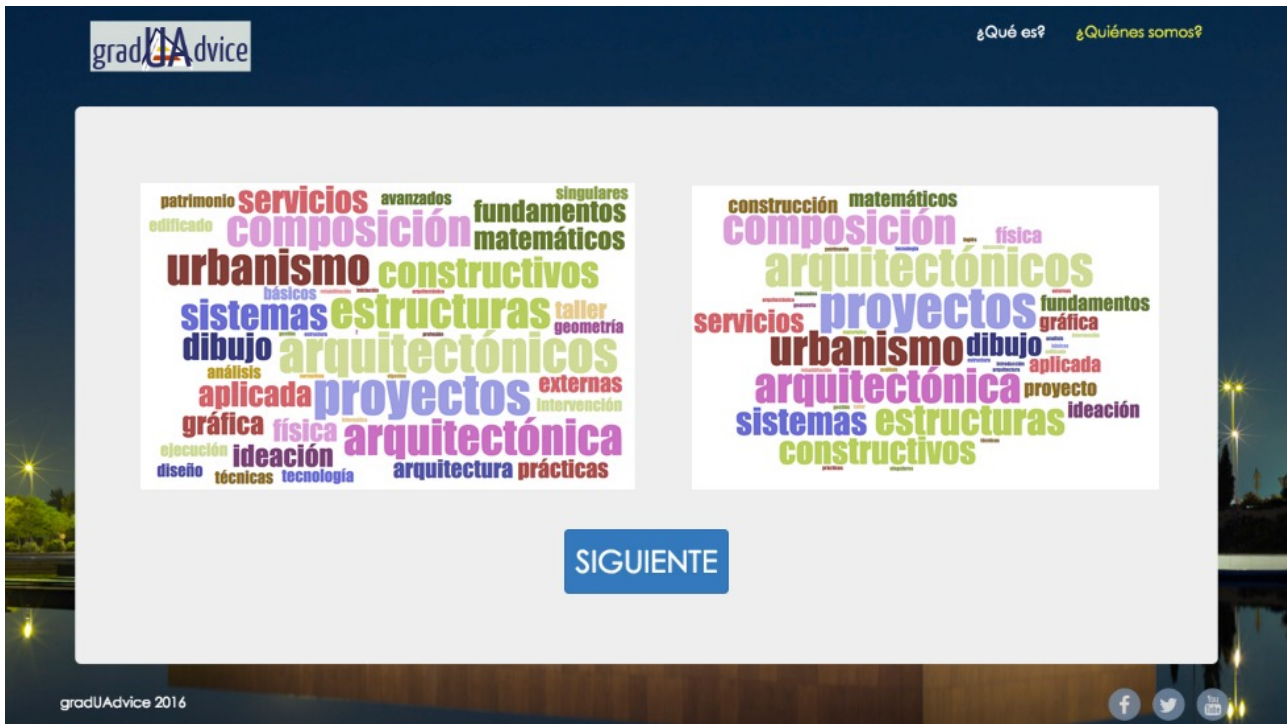


Figura 18: Tercera iteración

Se puede observar como a cada iteración el número de términos en los grupos decrece debido al acotado de la información, además de ver como las grandes familias como las ingenierías, las ciencias de la salud y letras se van separando.

En la primera iteración se ha elegido el grupo izquierdo, en la segunda el derecho y finalmente en la tercera el izquierdo nuevamente, dando resultado a la titulación “Grado en Arquitectura”.

#### 4.5.6. Finalización

Por último se obtiene la titulación obtenida como resultado y las asignaturas de bachillerato que el alumno eligió como preferencia de gusto. De aquí el usuario reflexiona si el resultado es correcto y/o debería replantearse la carrera profesional que en un momento quiso hacer suya.

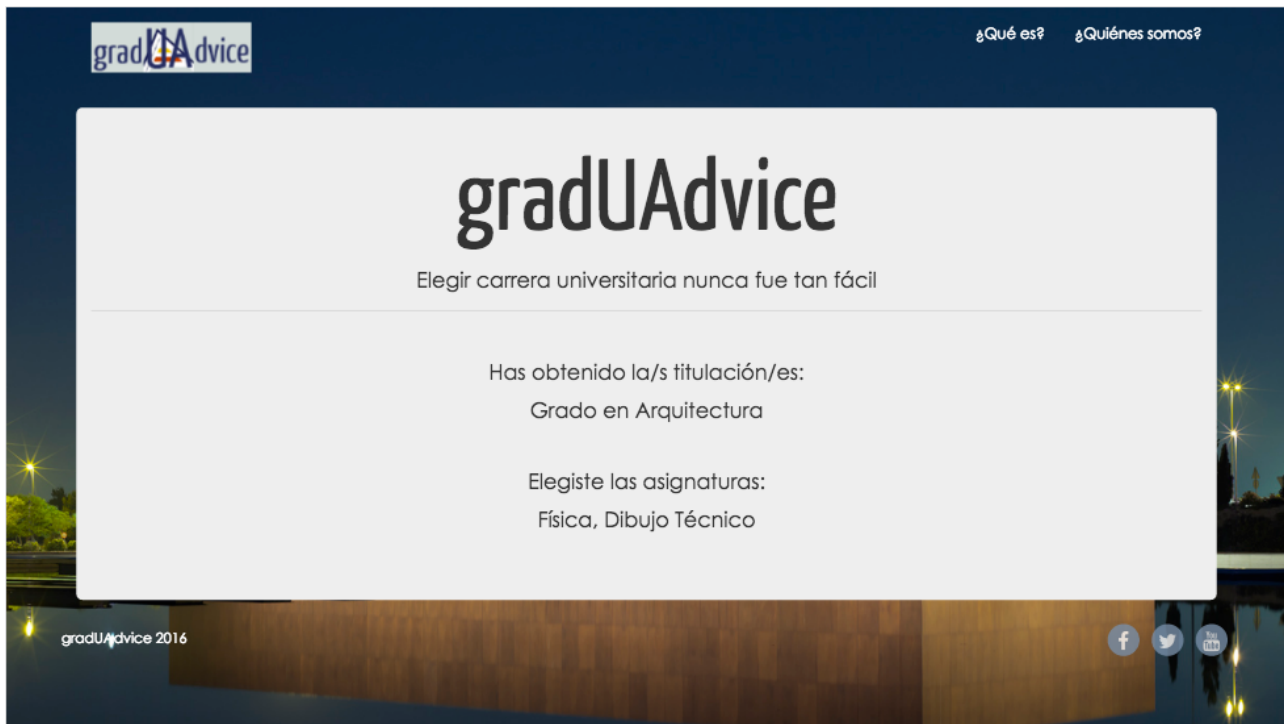


Figura 19: Resultado final

El último paso de la ejecución para un usuario es el de introducir una opinión constructiva de lo que le ha parecido el proceso. De esta forma se pueden aplicar nuevas mejoras, y además vale para tener un feedback para nuevas aplicaciones desarrolladas en la UA. Información a día de hoy muy importante ya que las aplicaciones se implementan orientadas e intuitivas para el usuario.



Figura 20: Formulario final

## 5. Posibles mejoras

A lo largo de la ejecución y como se ha podido comprobar con la obtención de datos por parte del portal de datos abiertos de la Universidad de Alicante, solo se han tratado las asignaturas de las titulaciones universitarias. Una posible mejora es el tratado de más información, el sistema de recomendación diseñado, conforme más términos reciba, más preciso es. Por lo tanto, con información como los objetivos y competencias de las titulaciones, se podría conseguir más precisión en gradUAdvice.

Otra posible mejora es el desarrollo de la aplicación para dispositivos móviles, de tal forma que pueda abarcar más usuarios y llegar a ellos con mayor facilidad, dado que prácticamente la mayoría de la población tiene en su propiedad un dispositivo móvil (teléfono o tablet).

En caso de que la Universidad de Alicante interviniese en el desarrollo de la aplicación, se podría integrar el login del campus virtual y con ello ampliar la seguridad de gradUAdvice. Es sabido que al realizar la preinscripción en la UA, la misma facilita al potencial alumno de su campus un usuario y contraseña para que pueda realizar los trámites pertinentes. Podría ser una herramienta bastante útil con el objetivo de sensibilizar al estudiante de si ha decidido correctamente su futuro estudiantil y salida profesional.

Como última mejora, se podría plantear un estudio estadístico para el tratado de las asignaturas de bachillerato para proponer pesos y plantearlas frente a las asignaturas de una titulación, teniendo en cuenta las obligatorias, troncales, básicas y optativas, así conseguir un porcentaje más sutil de cara a mostrar los resultados. Referido al sentido de obtener los resultados y extraer un porcentaje de similitud entre la lista de asignaturas de bachillerato frente a la/las titulación/es obtenidas como resultado final de la ejecución de gradUAdvice.

## 6. Conclusiones

Después del trabajo realizado en este TFM, y el planteamiento inicial, puedo decir que se ha asemejado bastante a lo que es el análisis, implementación y entrega de un proyecto laboral cualquiera. Incluyendo los problemas iniciales de un análisis ambicioso y superficial.

A parte de esos pequeños obstáculos que marcan un poco el proceso del trabajo, las sensaciones finales al ver la herramienta finalizada es de satisfacción. El haber aprendido y meterme de lleno en los sistemas de recomendación, esa incógnita para muchos y toda la computación que engloba.

Lo que más me ha gustado al tratar sistemas de este tipo y documentarme sobre ellos es que son totalmente moldeables a la necesidad que uno quiera aplicarle, es decir, pueden ser lo precisos que uno quiera que lo sean. Quizá el punto que más puede marcar el desarrollo de un sistema de recomendación es el estudio estadístico previo, con el cálculo de métricas y dar prioridad a un conjunto de datos, puede elevar la precisión del algoritmo.

En cuanto a los datos abiertos, es una forma de obtención de datos muy valiosos que para el organismo que los emite en su condición de “conjunto de datos” carecen de valor. Pero con una idea que se ajuste a las necesidades del mercado y un desarrollo óptimo, se pueden revalorizar de una forma impresionante. En este caso ha sido bastante fructífera la impresión ya que no había tratado de darle valor a tanta cantidad de datos.

En reglas generales en cuanto al modelo de negocio, cabe mencionar que es un área que de haberse extendido habría llegado a ser el objetivo principal. Ya que es una parte del análisis de mercado, competencias y enfoque del producto muy importante para que el mismo salga a flote. Como ya me dijo un profesor en el máster: “un producto vale lo que cuesta por su análisis de mercado, y su éxito queda marcado por un análisis bueno o malo”.

Como conclusión final, y más en concreto una valoración final del proyecto realizado, he aprendido que el procesamiento de datos en un sistema de recomendación es una área de investigación muy laboriosa y que sus frutos se deben a años de investigación, métodos de prueba-error y necesita de elevados conocimientos estadísticos y lingüísticos.

# Glosario de términos

1. **Start-up:** término anglosajón utilizado para referirse a las compañías emergentes y buscan arrancar en el mercado emprendiendo con un producto o servicio innovador.
2. **Open data (datos abiertos):** A parte de referirse a un conjunto de datos, la palabra open data hace referencia a una práctica que busca que determinados datos estén disponibles de forma libre para todo el mundo, sin restricciones de derechos de autor, de patentes o de otros mecanismos de control. Son datos que históricamente han estado bajo la jurisdicción de un ente privado o público y que por lo tanto su acceso estaba restringido (copyright y/o patentes).
3. **API key:** es un código con el que se identifica a un programa o persona que la usa para acceder a un servicio. Este servicio ofrece este código único para prevenir usos maliciosos o de carácter destructivo, en cuyo caso ese código identificador quedará invalidado de consumir el servicio.
4. **JSON:** abreviatura de JavaScript Object Notation, es un formato standard de texto ligero para el intercambio de datos. Es considerado un símil del formato XML presentando la ligera ventaja de ser más sencillo un analizador de JSON que uno de XML.
5. **Servlet:** clase en Java utilizada para extender la funcionalidad de un servidor. Utilizados frecuentemente para aplicaciones web en las que se ejecutan en el contexto del navegador web. En otras palabras, generar páginas web de forma dinámica a partir de los parámetros de la petición que envíe el navegador web.
6. **Indexar:** ordenar una serie de datos o informaciones de acuerdo a un criterio común a todos ellos para facilitar su consulta y análisis.
7. **Stopword:** en castellano palabra vacía, hace referencia a aquellos términos que carecen de significado como artículos, pronombres, preposiciones, etc. Palabras que son filtradas antes o después del procesamiento de datos en lenguaje natural.



# Bibliografía

1. Luis M. de Campos, Juan M. Fernández-Luna, Juan F. Huete, Miguel A. Rueda-Morales. (2007) Uso de Conocimiento estructurado en un sistema de recomendación basado en contenido. Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada.
2. Sergio Manuel Galán Nieto. (2007). Filtro Colaborativo y Sistemas de Recomendación. Consultado el 29 de agosto de 2016, en <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/31.pdf>
3. Cosine similarity. Consultado el 10 de agosto de 2016. [https://en.wikipedia.org/wiki/Cosine\\_similarity](https://en.wikipedia.org/wiki/Cosine_similarity)
4. R.R. Yager. Fuzzy Logic Methods in Recommender Systems. Fuzzy Sets and Systems, Volume 136, Issue 2 (2003), 133-149.
5. George, G and Bock AJ. 2012. Models of opportunity: How entrepreneurs design firms to achieve the unexpected. Cambridge University Press, ISBN 978-0-521-17084-0
6. What does tf-idf mean. Consultado el 10 de septiembre de 2016 en <http://www.tfidf.com>
7. Singhal, Amit (2001). "Modern Information Retrieval: A Brief Overview". Bulletin of the IEEE Computer Society Technical Committee on Data Engineering 24 (4): 35–43.
8. Repositorio de librerías Maven. Consultado 10 de agosto de 2016 en <https://mvnrepository.com>
9. Jason Davies d3-cloud. Consultado el 10 de septiembre de 2016 en <https://github.com/jasondavies/d3-cloud>



# Anexo I: Lean Canvas vs Business Canvas

El Business Model Canvas es una plantilla (véase la figura 21) de gestión estratégica para el desarrollo de nuevos modelos de negocio o documentar los ya existentes. Su autor, Alexander Osterwalder lo desarrolló en base a su trabajo anterior sobre la ontología de los modelos de negocio<sup>18</sup>.

Surgieron varios modelos más específicos, entre ellos cabe destacar el Model Lean Canvas, sugerido por Ash Maura<sup>19</sup>. Éste aporta una visión que consiste en darle un enfoque de negocio para start-ups que tratan de lanzarse al mercado con un producto estrella, mientras que el modelo convencional expone alianzas, recursos y relaciones con el cliente que una start-up no tendría:

*“El Lean Canvas se centra fundamentalmente en entender el problema o necesidad del cliente para luego poner el foco en el diseño del producto. Este enfoque no está tan presente en el lienzo de Alex Osterwalder (creador del Business Model Canvas), sino más bien en su segundo lienzo, el de la propuesta de valor. Al igual que en éste lienzo, en el Lean Canvas el lado derecho representa al mercado, mientras que el lado izquierdo muestra la solución o producto. Para Ash Maurya una startup debe concentrar primero sus recursos en encontrar soluciones a los problemas o necesidades de un determinado segmento de clientes.”*

---

18 <http://www.innovacion.gob.cl/reportaje/metodologia-canvas-la-nueva-forma-de-agregar-valor/>

19 <https://leanstack.com/why-lean-canvas/>

En cuanto a los cambios en forma de bloques que incorpora el modelo Lean Canvas frente al Business Canvas, se citan a continuación:

- **Problema:** un aspecto clave para una start-up es el de identificar correctamente el problema al que se enfrenta y definir las funcionalidades principales del producto.
- **Solución:** una vez entendido el problema, el siguiente punto es el de exponer las posibles soluciones
- **Métricas clave:** es crítico definir desde el principio los 3 ó 4 primeros indicadores para medir el modelo de negocio. Actividad clave para no incurrir en objetivos erróneos y fracaso prematuro.
- **Ventaja diferencial:** conforme la start-up va avanzando hacia el éxito es inevitable la aparición de competencia y copiadores, por lo que la defensa es indispensable, lo que diferencia el producto de otros “falsos”.



Figura 21: Modelo Lean Canvas

Mientras que los bloques eliminados son:

- **Actividades y recursos clave:** la explicación que da Ash Maurya para eliminar estos dos bloques era que ayudaban a visualizar lo que la start-up hacía, pero hacía que se perdiera foco en lo verdaderamente importante.
- **Relaciones con los clientes:** en las primeras etapas de una startup, es más importante encajar la solución en las necesidades y problemas de los clientes, que establecer relaciones fuertes. Para las relaciones más estrictamente hablando, está el bloque de Canales.
- **Socios clave:** el máximo enfoque de una startup es el de mirar hacia la estructura de costes y en los canales, no en la búsqueda aliados

Los cambios comentados en los puntos anteriores se pueden ver reflejados en la siguiente figura:

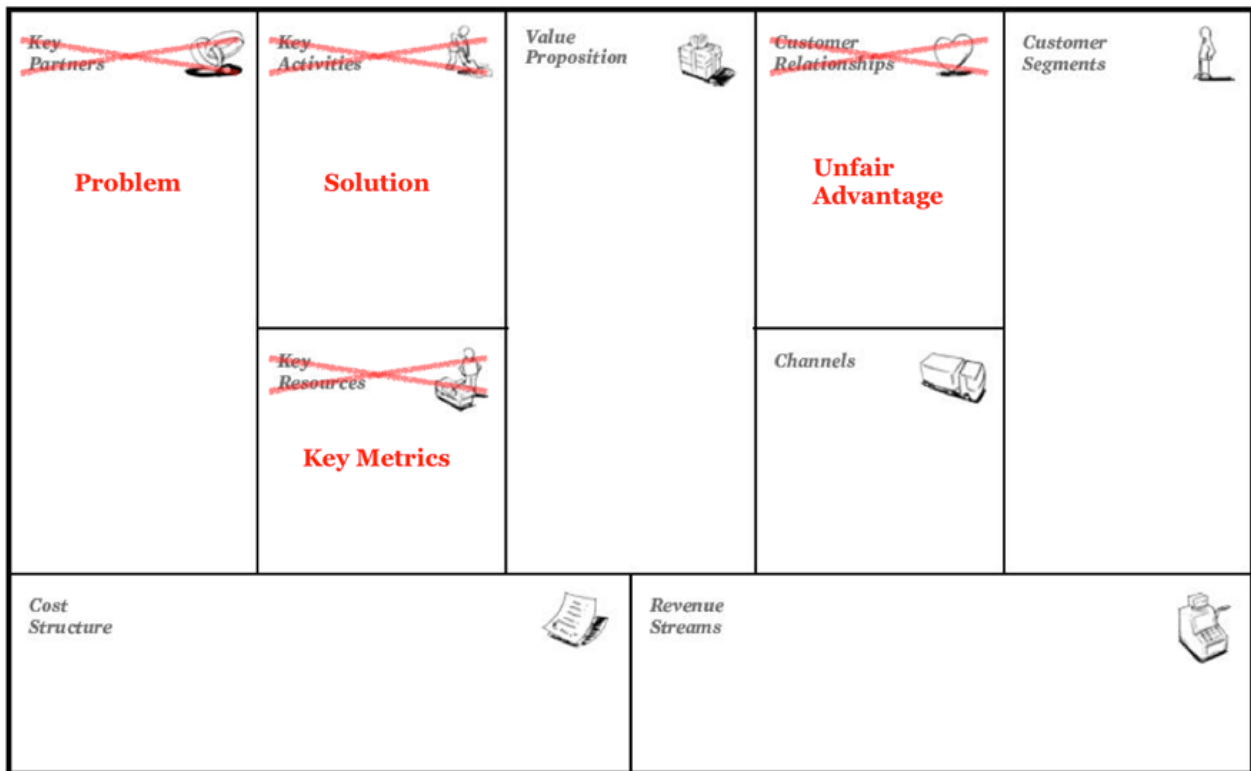


Figura 22: Modelo Lean Canvas vs Business Model Canvas

# Anexo II: Respuestas detalladas del test inicial

El cuestionario empieza preguntando el tipo de bachiller que el encuestado ha cursado:

## 1. ¿Qué tipo de bachiller has cursado?

- *Ciencias y Tecnología (48)*
- *Humanidades y Ciencias Sociales (38)*
- *Artes (2)*

Se puede apreciar en la figura 23 que las mayores porciones se dividen entre “ciencias puras” y “letras”, donde habría sido conveniente haber separado esos dos grandes grupos, para al final tener seis grupos (bachiller tecnológico, ciencias de la salud, humanidades, ciencias sociales y artes).

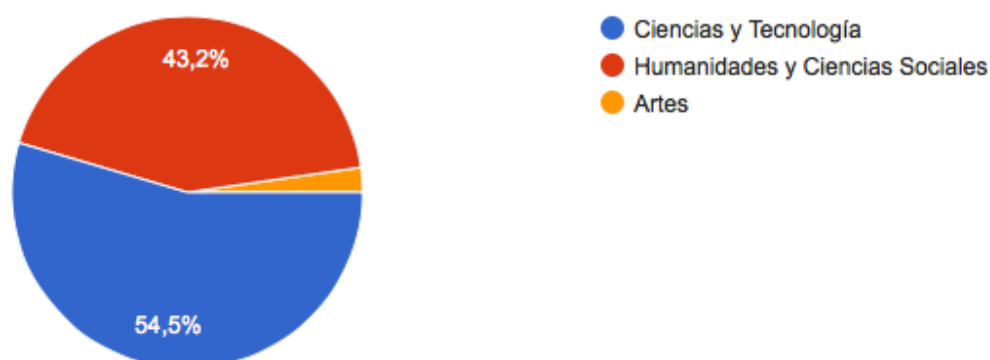


Figura 23: Gráfica de tipos de bachillerato

## 2. ¿Este es el primer año que haces selectividad?

- *Si (72)*
- *No (16)*

El estudio iba dirigido a alumnos supuesto primer año en la universidad, aunque se abrió el abanico a estudiantes de otros años, ya que podrían aportar información relevante y opiniones más sofisticadas. Véase los resultados en la figura 24.

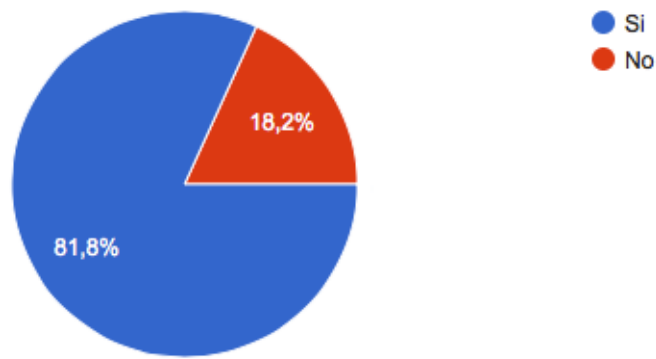


Figura 24: Gráfica de año cursado las PAU

### 3. ¿Sabes ya qué carrera vas a estudiar?

- *Si* (58)
- *No* (30)

Esta pregunta es muy importante, ahí es donde gradUAdvice viene a actuar, a paliar la indecisión del alumnado (34,1% de los encuestados no saben que quieren estudiar a tres días de las PAU), eso sin olvidar que algunos de ellos pueden llegar a haber decidido la carrera universitaria que van a cursar, sin saber a ciencia cierta que es su verdadera vocación u objetivo final. Véase los resultados en la figura 25.

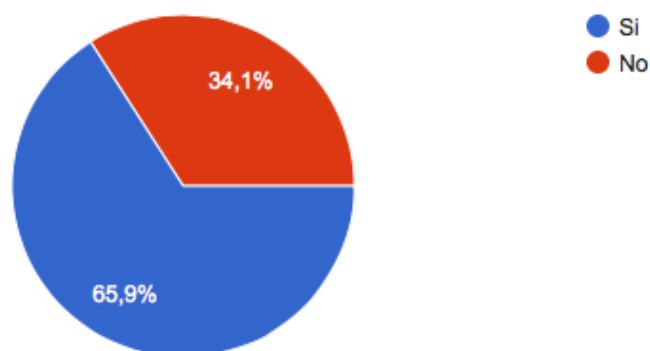


Figura 25: Gráfica de decisión de la carrera universitaria

#### 4. ¿Crees que los planes de estudios están bien explicados?

- *Si* (60)
- *No* (28)

Otra pregunta muy importante y queda patente en el modelado de negocio Lean Canvas: Los planes de estudios de la universidad o no están muy bien explicados, o los alumnos escatiman en leerlos y comprenderlos. La verdadera respuesta se ofrece en la 6ª pregunta a continuación. Véase los resultados en la figura 26.

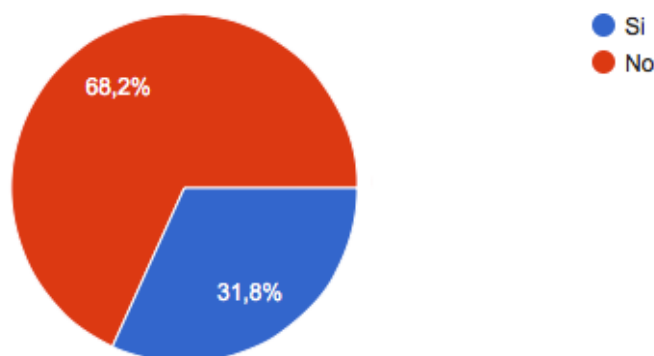


Figura 26: Gráfica de explicación de planes de estudio

#### 5. ¿Has leído algún plan de estudios o al menos el de la carrera que te gustaría cursar?

- *Si* (76)
- *No* (12)

Se puede observar por el número de afirmaciones, que el plan de estudios es una fuente de información relevante para los potenciales alumnos universitarios, pero queda ese pequeño resquicio de que la información proporcionada por un plan de estudios no es suficiente y a veces incompleta. Véase los resultados en la figura 27.



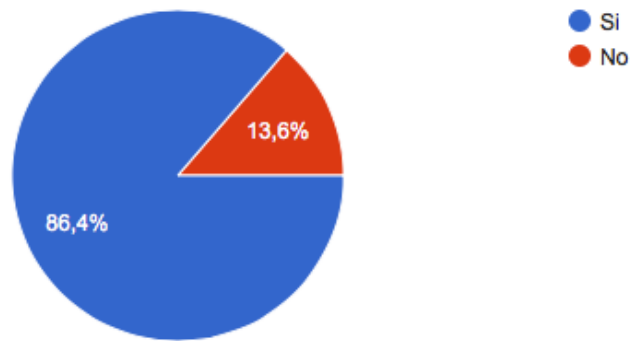


Figura 27: Gráfica de planes de estudio

**6. En ese proceso de lectura, te han surgido dudas, ¿Has tratado de resolverlas?**

- *Si* (54)
- *No* (34)

Queda aún muy patente ese elevado porcentaje (38,6%) de alumnos que aún teniendo dudas con la carrera que quieren estudiar, prefieren quedarse con la duda, que resolverla en cuanto antes. Véase los resultados en la figura 28.

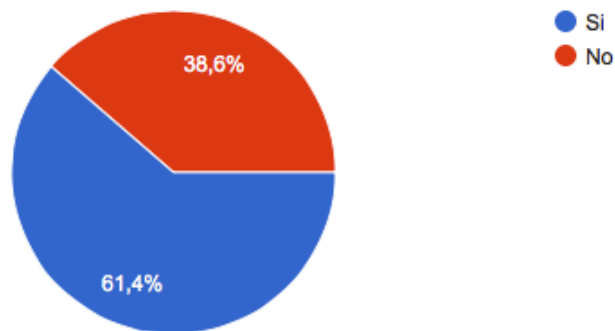


Figura 28: Gráfica de dudas con el plan de estudios

### 7. ¿Has acudido a alguna facultad en la universidad para preguntar o informarte?

- *Si (41)*
- *No (47)*

Se puede observar con diferencia y más claridad en esta pregunta como está el panorama. La información que proporcionan las facultades no son consideradas valiosas por parte de los alumnos. Eso o que a posteriori de realizar la encuesta, se observa que faltó una pregunta de especificar el porqué. Puede ser debido al poco valor o vergüenza de acercarse la secretaria de una facultad, el esfuerzo que implica moverse a la universidad en verano o incluso que la información proporcionada en internet es de más fácil acceso. Véase los resultados en la figura 29.

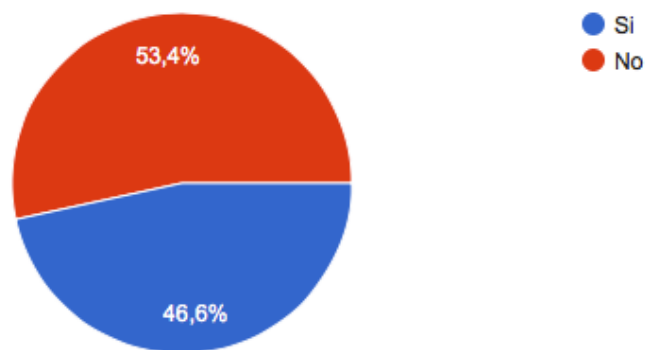


Figura 29: Gráfica de acudir a la universidad en busca de información

### 8. ¿Utilizarías una herramienta online que te ayudara con la elección de tu carrera universitaria?

- *Si (72)*
- *No (16)*

Definitivamente esta es la pregunta clave de la encuesta. Se puede observar que más del 81% de los encuestados estaría dispuesto a utilizar una herramienta online, como es gradUAdvice, para hacer de guía o instructor acerca de las posibles titulaciones universitarias que les gusten. Véase los resultados en la figura 30.

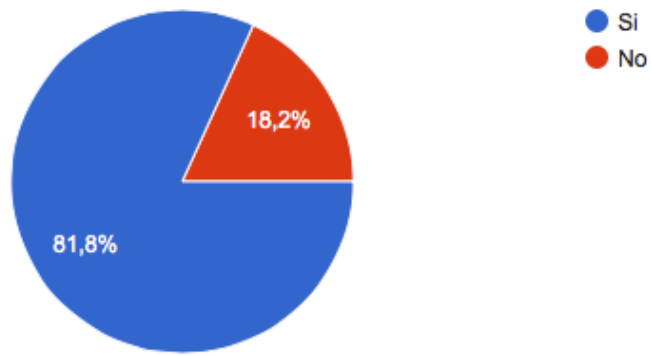


Figura 30: Gráfica de utilización de herramienta online

## **Anexo III: Lista de titulaciones de la UA**

1. Grado En Geografía Y Ordenación Del Territorio
2. Grado En Historia
3. Grado En Humanidades
4. Grado En Turismo
5. Grado En Estudios Árabes E Islámicos
6. Grau En Filologia Catalana
7. Grado En Estudios Franceses
8. Grado En Español: Lengua Y Literaturas
9. Grado En Estudios Ingleses
10. Grado En Traducción E Interpretación
11. Grado En Geología
12. Grado En Matemáticas
13. Grado En Química
14. Grado En Biología
15. Grado En Ciencias Del Mar
16. Grado En Óptica Y Optometría
17. Grado En Gestión Y Administración Pública
18. Grado En Derecho
19. Grado En Criminología
20. Grado En Relaciones Laborales Y Recursos Humanos
21. Programa Simultáneo Dade: Derecho Y Administración Y Dirección De Empresas
22. Programa Conjunto Decrim. Grado En Derecho Y En Criminología
23. Grado En Sociología
24. Grado En Administración Y Dirección De Empresas
25. Grado En Economía
26. Grado En Publicidad Y Relaciones Públicas
27. Grado En Trabajo Social
28. Programa Simultáneo Tade: Turismo Y Administración Y Dirección De Empresas
29. Grado En Ingeniería En Sonido E Imagen En Telecomunicación
30. Grado En Arquitectura
31. Grado En Ingeniería Informática
32. Grado En Ingeniería Química
33. Grado En Ingeniería Multimedia
34. Grado En Arquitectura Técnica
35. Grado En Fundamentos De La Arquitectura
36. Grado En Ingeniería Civil
37. Grado En Tecnologías De La Información Para La Salud
38. Grado En Ingeniería Robótica
39. Grado En Maestro En Educación Infantil
40. Grado En Maestro En Educación Primaria
41. Grado En Ciencias De La Actividad Física Y Del Deporte
42. Grado En Enfermería
43. Grado En Nutrición Humana Y Dietética