



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA

Trabajo fin de Máster

Máster en Ingeniería Informática

Metodología Running Lean aplicada a lector de noticias inteligente

**Realizado por
(ponente): Andrés M. Jiménez Ríos**

**Dirigido por
Alejandro Fernández-Montes González**

**Departamento
Lenguaje de Sistemas Informáticos**

Sevilla, 27 de abril de 2019

Resumen

¿Por qué LT-News? Surge de la necesidad de todos de informarse, pero la enormidad de medios y el poquísimo tiempo nos hace excusarnos de la gran tarea de leer y formarse una opinión propia de lo que ocurre a nuestro alrededor.

Este trabajo se propone ser un lector RSS. Pero no uno típico, no, sino uno que te acabe conociendo tanto, que te averigüe tus gustos, te muestre las noticias en las que estés realmente interesado. Además, poseerá funcionalidades tan útiles como la de buscar noticias en los periódicos que sigas, o hacer un análisis diario de las noticias más importantes de tus secciones.

Además, podrás ir viendo gráficas en bases a tu uso de la aplicación o de la evolución de la aparición de una determinada noticia a lo largo del tiempo.

Agradecimientos

Sero te amavi, pulchritudo tam antiqua et tam nova, sero te amavi! et ecce intus eras et ego foris, et ibi te quaerebam, et in ista formosa, quae fecisti, deformis inruebam. mecum eras, et tecum non eram. ea me tenebant longe a te, quae si in te non essent, non essent. vocasti et clamasti et rupisti surditatem meam: coruscasti, splenduisti et fugasti caecitatem meam: fragrasti, et duxi spiritum, et anhelo tibi, gustavi et esurio et sitio, tetigisti me, et exarsi in pacem tuam.

Índice general

Índice general	III
Índice de cuadros	V
Índice de figuras	VI
Índice de código	VIII
1 Introducción	1
1.1 Justificación del proyecto	1
1.2 Objetivos	1
1.2.1 Objetivos funcionales	2
1.2.2 Objetivos técnicos	2
1.3 Alcance	3
1.4 Estado del arte	3
1.4.1 Feedly	3
1.4.2 Flipboard	4
1.4.3 Google News	6
1.4.4 Otras	8
2 Estudio	9
2.1 Metodología LEAN	9
2.1.1 <i>Running Lean</i>	11
2.1.2 Aplicación real	14
2.2 Primer estudio	15
2.2.1 Segmento de clientes	15
2.2.2 Problema	15
2.2.3 Propuesta de valor único	15
2.2.4 Principales hipótesis	16
2.3 Iteraciones	16
2.3.1 Primera iteración	17
2.3.2 Segunda iteración	17
2.3.3 Tercera iteración	19
2.3.4 Cuarta iteración	19
2.3.5 Quinta iteración	20
2.3.6 Quinta iteración	29
2.4 Conclusión estudio	29

3 Planificación	31
3.1 Metodología	31
3.1.1 SCRUM	32
3.1.2 Kanban	35
3.1.3 Aplicación	36
3.2 Planificación temporal	38
3.3 Roles del proyecto	39
3.4 Planificación económica	39
3.4.1 Costes directos	39
3.4.2 Costes indirectos	40
3.4.3 Presupuesto	41
4 Análisis	42
4.1 Actores del sistema	42
4.2 Modelo conceptual	42
4.3 Catálogo de requisitos	42
4.4 Diagrama de secuencia	42
5 Diseño	43
5.1 Patrones	43
5.2 Modelo de datos	43
5.3 Prototipos	43
6 Implementación	44
6.1 Tecnologías	44
6.2 Herramientas	48
6.2.1 Backend	48
6.2.2 Frontend	49
6.2.3 Servidor	51
6.2.4 Gestión	52
6.3 Estructura del proyecto	52
6.3.1 GitHub	53
6.3.2 Backend	55
6.3.3 Frontend	57
6.4 Detalles de implementación	58
7 Pruebas	59
7.1 Pruebas unitarias	59
7.2 Pruebas integradas	59
7.3 Pruebas de aceptación	59
7.4 Estándares de código	59
8 Conclusión	61
8.1 Mejoras futuras	61
8.2 Lecciones aprendidas	61
Referencias	62

Índice de cuadros

2.1	Aprendizaje entrevista de problema	14
2.2	Aprendizaje entrevista de solución	14
2.3	Aprendizaje entrevista del MVP	14
2.4	Iteraciones realizadas	16
3.1	Planificación de Sprints	38
3.2	Planificación temporal	38
3.3	Presupuesto económico	41

Índice de figuras

1.1	Logo de Feedly	4
1.2	Página inicio de Feedly	5
1.3	Logo de Feedly	5
1.4	Página inicio de Flipboard	6
1.5	Logo de Google Noticias	6
1.6	Página inicio de Google News	7
2.1	Resumen Running Lean	10
2.2	Lean Canvas	11
2.3	Adecuación de nuestra idea al mercado	12
2.4	Prueba sistemática	13
2.5	LEAN Canvas inicial	16
2.6	LEAN Canvas primera iteración	17
2.7	LEAN Canvas segunda iteración	19
2.8	LEAN Canvas tercera iteración	20
2.9	Encuesta primera pregunta	21
2.10	Encuesta segunda pregunta	21
2.11	Encuesta tercera pregunta	22
2.12	Encuesta cuarta pregunta	22
2.13	Encuesta quinta pregunta	23
2.14	Encuesta sexta pregunta	23
2.15	Encuesta séptima pregunta	24
2.16	Encuesta octavo pregunta	24
2.17	LEAN Canvas cuarta iteración	25
2.18	Landing Page 1	25
2.19	Landing Page 2	26
2.20	Landing Page 3	26
2.21	Google Analytics 1	27
2.22	Google Analytics 2	27
2.23	Google Analytics 3	28
2.24	LEAN Canvas final	29
3.1	Aprendizaje en SCRUM	32
3.2	Eventos en SCRUM	33
3.3	Equipo en SCRUM	34
3.4	Ejemplo de tablero Kanban	35
3.5	Tablero Kanban usado	37
6.1	Muestra del <i>Web Scrapping</i>	44
6.2	Muestra de Recuperación de Información	45
6.3	Muestra de sistemas de recomendación	46

6.4	Muestra de extracción del perfil del usuario	47
6.5	<i>Git Flow</i>	53
6.6	Respositorios en GitHub	54
6.7	<i>Issues</i> en Github	54
6.8	<i>Labels</i> en Github	55
6.9	<i>Milestones</i> en Github	55
6.10	Estructura en PyCharm	56
6.11	Estructura en WebStorm	57
7.1	Muestra de logs de Travis	60

Índice de código

CAPÍTULO 1

Introducción

1.1– Justificación del proyecto

Cuando decidí hacer mi Trabajo Fin de Grado, tome uno de los temas que más me apasiona: el periodismo. La idea era hacer una aplicación que superase a todas las existentes -se entiende que solo desde el punto de vista conceptual. Finalmente lo conseguí, quedando un lector RSS inteligente.

La aplicación cumplía todas las características que un usuario pide a un lector de noticias. Estas son: añadir cualquier fuente que posea sindicación de contenido, gestionar los *feeds* y leer las noticias de estos al momento, llevando un control de lectura de las mismas. Además, poseía sobre esto añadidos interesantes como guardar noticias para leerlas más tarde o agrupar los *feeds* en secciones.

Además de lo anterior, poseía características adicionales que le daban el sobrenombre de: inteligente. Una de estas era la capacidad de relacionar noticias entre sí y mostrárselas al usuario mientras leía un noticia. Otra era la de extraer el perfil del usuario y recomendarle noticias en base a cómo utilizase la aplicación y a qué tipo de artículos solía leer. La tercera característica era la búsqueda de noticias en base a los criterios que se desease.

Como se puede ver, no se encuentra actualmente un lector de noticias con estas características. ¿Por qué? El motivo es el coste de la infraestructura que ha de poseer la aplicación. Si pensamos en cualquier medio importante, veremos que su fichero RSS se actualiza de media cada diez minutos, publicando así una media de una quince noticias por hora. Si seguimos aritméticamente los cálculos, veremos que por día, sólo en España, se publican del orden de millones de artículos.

Por ello, las tecnologías, infraestructura y algoritmos que utilicé quedaron obsoletos al poco tiempo, siendo necesario profesionalizar el proyecto si se quería una aplicación que el público final pudiese utilizar. Además, faltaba en esta una parte importante, que era la monetización de la misma.

Como se puede decidir, el Trabajo Fin de Máster consistirá en rehacer la aplicación utilizando tecnologías adaptadas a esta enorme ingesta de datos y utilizando un formato más profesional, del que más tarde hablaremos. Además, de enfocar el proyecto hacia una aplicación comercial con la que sea posible ganar dinero.

1.2– Objetivos

Los objetivos se engloban en dos categorías: funcionales y técnicos. En la primera se agrupan todos los conocimientos que quiero adquirir en cuanto a metodología, conocimiento del dominio del problema y venta de producto. Segundo será el aprender un determinado número de tecnologías o infraestructuras.

1.2.1. Objetivos funcionales

El objetivo principal a adquirir en cuanto a conocimiento es la experiencia emprendedora. Esto no es más que convertir de una idea tecnológica atrayente y estimulante, a una pequeña empresa que sostenga y monetice dicha idea. Aunque pueda parecer algo nimio a simple vista, requiere todo un proceso de concepción de la idea, el cliente y el ecosistema de ventas. Gracias a las asignaturas del máster, veo posible implementar dicha idea hacia una idea, ya no puramente tecnológica, sino de negocio.

Los objetivos subyacentes a esta son el conocimiento de las metodologías que me permitan realizar este proceso. En concreto, me he basado en tres ideas actuales de emprendimiento que se pueden encontrar a lo largo del documento y que se reducen a tres autores:

- *Customer Development* de Steve Blank
- *Business Model* de Alex Osterwalder
- *Running Lean* de Ash Maurya

En estos se definen los procesos de creación de la *start-up*, creación del producto, atracción de clientes y de mejora continua. Me han parecido enormemente interesante y acertados para la aplicación en mi idea de negocio, por eso decidí aplicarlo.

El segundo objetivo es el conocimiento más férreo del dominio del problema de la idea de negocio, es decir, el periodismo. Está viviendo un proceso de cambio en cuanto al mundo de las noticias. Ya no interesa conocer lo que ocurrió el día anterior, sino lo que pasa ahora mismo en cualquier parte del mundo. Es por ello, que la prensa ha tenido que pivotar para saciar el ansia de información de sus usuarios. Además de esto, han tenido que atraer a usuarios más jóvenes, que ven más atrayente conocer el mundo a través de las redes sociales, antes que a través de periodistas.

En este proceso, además, se esconde una oscura pero cada vez más visible realidad: el mundo de la comunicación avanza varios pasos por detrás del ritmo *impuesto* por la sociedad digital, por ello podemos ver múltiple deficiencias, que se explicarán a lo largo del documento. Es por ello también objetivo del presente estudio proponer mejoras al mismo.

1.2.2. Objetivos técnicos

El objetivo técnico por antonomasia a adquirir en el Trabajo Fin de Máster es desarrollar una aplicación siguiendo las pautas marcadas por las grandes compañías, en cuanto a tecnologías, herramientas e infraestructuras se refiere. Teniendo este preámbulo se entenderán correctamente estos objetivos.

El primer objetivo es el desarrollo de una aplicación utilizando la estrategia de desacople de capa de servidor y cliente. Esto no es más que ofrecer desde el *backend* una API consumible desde cualquier cliente. Gracias a esto, se puede diferenciar las tecnologías de cliente y servidor, haciéndolas independientes.

El segundo objetivo es desarrollar la capa de presentación utilizando un framework JavaScript. Esta estrategia es enormemente usada en el mundo del desarrollo web, ya que es posible realizar toda una aplicación *frontend* con AJAX, es decir, sin recargas para extraer la información. Toda la comunicación con el servidor será asíncrona, por lo que la fluidez que percibirá el usuario será total.

El tercero será el desarrollo de una capa de inteligencia artificial que permita conocer tanto al usuario que interacciona con la aplicación así como el contenido que se muestra. Con ello, se podrá recomendar y extraer datos de ambos. Este, además de dar un valor añadido al desarrollo que se realice, lo hará enormemente interesante para un usuario con conocimiento del dominio del problema, ya que será posible extraer conclusiones.

Por último, y en la medida de lo posible, se incoará un proceso de integración y entrega continua. Para ello, se utilizarán herramientas que nos permitan agilizar en la medida de lo posible el tiempo desde que un desarrollador hace un cambio hasta que el usuario lo percibe en la web.

1.3– Alcance

Es objeto del presente trabajo es el estudio, análisis e implementación de un proceso completo de negocio que va desde la concepción de la idea hasta la venta de la misma. Para ello, se utilizará el método indicado por Ash Maurya con su idea del *Running Lean*, enormemente apoyada por emprendedores de todo el globo.

Dentro de la implementación, realizado siempre un Producto Mínimo Viable, como indica Ash, se pretende realizar una lector RSS con características adicionales que le den un valor añadido como son el análisis de perfiles de usuarios, el análisis de medios y el análisis de noticias. El hacer hincapié en el *análisis*, no es más que recordar lo que diferenciará nuestro producto del resto. El usuario, en definitiva, lo notará debido a la capacidad de realizar búsquedas en la aplicación, así como la recomendación de noticias.

Dentro del marco actual, la aplicación estará pensada para los dispositivos actuales en toda la gama de resoluciones de pantallas. Además, al permitirse la posibilidad de añadir cualquier tipo de *feed*, se añadirá a la plataforma múltiples idiomas, en la medida de lo posible.

1.4– Estado del arte

La sindicación de contenido vino a resolver un problema en 1995. Este no era otro que dar la posibilidad a un usuario de poder seguir desde una única plataforma todas las noticias de diferentes medios.

En cambio, si vemos como ha evolucionado el mundo digital, en cuanto al número de webs se refiere, podemos ver lo siguiente:

En solo 10 años, hemos pasado de tener 92 millones de sitios web en Internet a más de 1.000 millones. Internet ya no es solo un lugar dónde encontrar información o dónde encontrar ocio y entretenimiento, ni es solo una herramienta de trabajo, Internet es dinero y es cambio social.

Rana Negra (2018)

Ahora, entonces, podemos ver otro problema, no resuelto todavía. Ante la gran cantidad de medios actuales, el problema es escoger entre tanta cantidad de noticias, solo aquello que aporte valor y realmente nos interesa. Ahora que termina el segundo decenio del siglo, podemos ver cómo intenta solucionar esto cada aplicación.

1.4.1. Feedly

Aplicación nacida en 2008 con el objetivo de llegar a los usuarios que empezaban a entrar en el mundo de los smartphones. Vio su auge al desaparecer Google Reader. Uno de sus puntos fuertes que le hizo crecer, a la par del suceso anteriormente mencionado, es el tratamiento multi-idioma que se hace en la plataforma.

Una de sus mejores ventajas es la sencillez de uso a la hora de añadir nuevas fuentes a nuestro *feed*: es momentáneo. Otra es la vista que tiene dentro del apartado *Today*. Aquí encontramos las noticias nuevas de los medios a los que estamos suscritos. Puede parecer una característica baladí o demasiado popular. Sin embargo, Feedly consigue hacerlo único siguiendo una operativa que le da un valor añadido.

Primero, que este no muestra las noticias que *te has perdido* en orden cronológico puramente; entremezcla las secciones que tengas en pantallas y ahí si muestra las noticias en orden de publicación. Esto es enormemente útil ya que, por ejemplo, si tienes *El País* o el *ABC* en la sección



Figura 1.1: Logo de Feedly

Noticias, lo normal es que tengas noticias cada diez minutos. Sin embargo, si posees otra sección, *Fútbol*, donde se recogen *feeds* con análisis de los partidos más importantes, lo normal es que tengas cinco noticias a la semana.

Abstrayendo del ejemplo, podemos ver que habrá secciones con noticias cada minuto y otras, cada semana. Con esta distribución, Feedly consigue que se le da igual importancia a ambas secciones mientras que te queden por leer noticias.

Segundo, que no hay un *scroll infinito* a la hora de leer noticias. Esto hace aburrido y desmotivante al lector la lectura. Lo que hace son barridas, es decir, que va seleccionando un número determinado de noticias por cada sección. Una vez visto estas pregunta al usuario si quiere seguir viendo más.

Este detalle junto a las múltiples integraciones que posee le hace enormemente usable y querido por el usuario, por lo que hace de esta aplicación la principal a combatir en cuanto a presencia en el sector se refiere.

Puesto a poner pegas siempre se le achacan dos desventajas. La primera es que el usuario de esta aplicación suele ser una persona que ha de conocer, aunque sea levemente, el mundo de la sindicación de contenido. Esto hace que para un usuario con poco conocimiento de internet pueda parecer un mundo.

La otra característica a mejorar es la recomendación tanto de medios como de noticias. Cada vez se hace más eco en blogs especializados: Paul Bonduelle (2018). Como se puede ver, empieza a tomar cuerpo, pero falta mucho camino a recorrer.

1.4.2. Flipboard

Aplicación nacida de la mano del iPad, de la primera tableta del mercado, en 2010. Debido a su auge, se decidió desde la compañía crecer a los dispositivos móviles de Apple. Por ello, nació la aplicación para iPhone y iPod. Igual de triunfante fue tal decisión que, tras un *crowdfunding* en 2012, se decidió ir tras el mundo Android.

Su enorme ventaja es la sencillez de uso total. Cualquier tipo de usuario puede configurarse su *feed* sin necesidad de ningún conocimiento en el marco del mundo de las noticias ni tan siquiera técnico. En cuanto te registras, te dan una serie de secciones base que elegir. Una vez añadidas y dicho tus gustos, él se encargará de añadir o quitar medios en dichos feeds en cuanto a tus gustos se refiere.

Hace totalmente transparente al usuario de la tecnología que se encuentra de fondo. Por lo que este ni puede crear secciones ni seleccionar feeds que no sean populares, es decir, por la URL del RSS o Atom.

Al centrarse en el usuario, hace que la interfaz sea enormemente atractiva y el punto continuo de mejora. Es por ello que si se implementa la vista previa de la noticia en sí, es decir, en cuanto se pulsa, va directamente a la fuente original. Además, a la hora de mostrar las noticias, se muestran se hayan ya visto o no, además que en formato *scroll continuo*.

The screenshot shows the Feedly mobile application's home screen. On the left is a dark sidebar with navigation options: Home, Saved For Later, Shared Collections (with a 'NEW' badge), Add Content, All (734 items), Tech (484 items), Decoration (15 items), Food (61 items), and Competition (2 items). At the bottom of the sidebar are profile details for 'edwin@feedly.com' and links to 'via Google / Logout' and 'Team'. The main content area has a header 'Home 50x faster polling (Team Edition)' with a 'Unpin' button. Below the header are three news cards:

- Apple TV Gains Updated NFL Channel With Game Pass Integration**: A card featuring a screenshot of the NFL Now channel on Apple TV. The text notes that the channel was revamped, changed to "NFL", and supports Game Pass subscriptions. It includes a link to MacRumors and a timestamp of 100+ posts by Juli Clover 2 hours ago.
- Tep Is An Adorable Fitness Tracking App That Works Like A Tamagotchi**: A card featuring a screenshot of the Tep app interface. The text describes it as a Tamagotchi-like app for fitness tracking. It includes a link to TechCrunch and a timestamp of 400+ posts by Romain Dillet 4 hours ago.
- Apple Seeds Eighth Beta of OS X El Capitan to Developers, Sixth Beta to Public Testers**: A card featuring a screenshot of a Mac desktop screen. The text notes that Apple released the eighth beta of OS X El Capitan to developers and the sixth to public testers, nearly two weeks after the seventh beta. It includes a link to MacRumors and a timestamp of 300+ posts by Juli Clover 4 hours ago.

At the bottom right of the main content area is a 'Are you still using Apple Music?' section with a card for 'Eatsa, A Futuristic Restaurant Where Robot Cubbies Serve Quinoa'.

Figura 1.2: Página inicio de Feedly

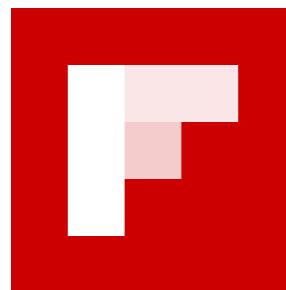


Figura 1.3: Logo de Feedly

Además de lo anterior, cabe destacar la integración con las redes sociales, en concreto Twitter. Cada noticia se redirecciona con los tuits oficiales o de cuentas con cierto conocimiento de causa de lo que se está hablando.

Otro punto fuerte de Flipboard es el concepto de *Revista*. Esto no es otro que la posibilidad de agrupar noticias que nos interesen y poder enseñarla al mundo. Tan es así, y tanto fama y utilidad le dan los usuarios, que la empresa dio el dato en 2016 que poseía 28 millones de revistas Wikipedia (2018).

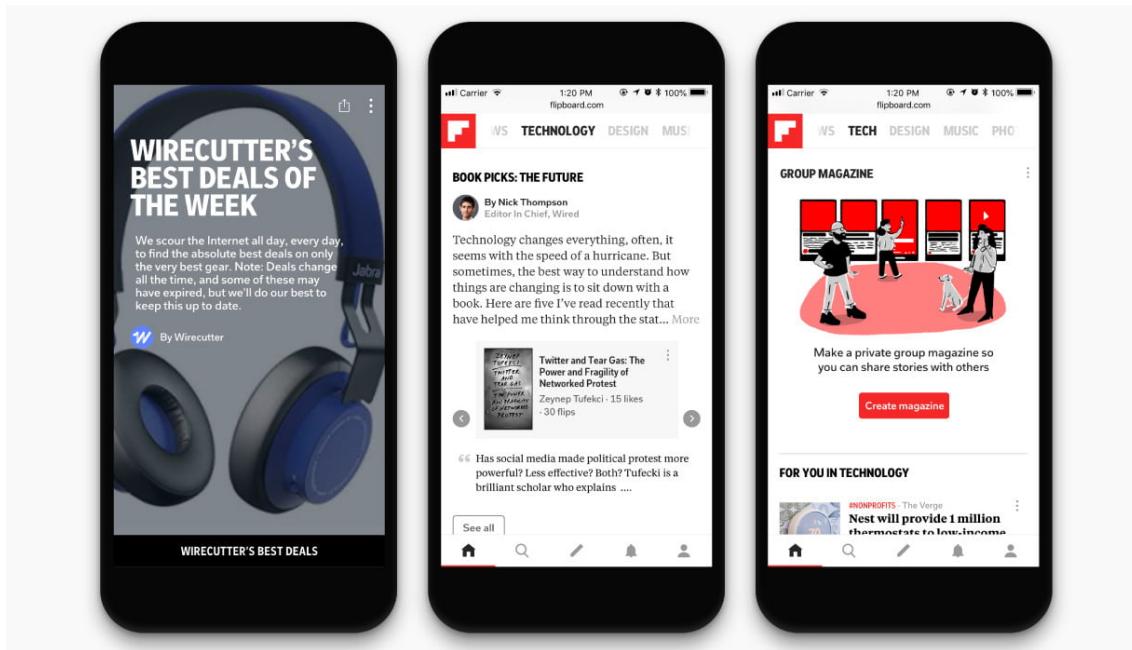


Figura 1.4: Página inicio de Flipboard

Como se ha dicho, desde el punto de vista técnico puede resultar un poco pobre, sobre todo a un usuario que posea conocimiento del tema. Además, hay poco uso de las recomendaciones de noticias o medios, y no se plantea en medio plazo el de incorporarlo.

1.4.3. Google News

Google Noticias fue lanzado en 2002. Desde la conferencia de Google del pasado año Quentyn Kennemer (2018), se le dio a la aplicación un vuelco que la hizo orientarse por completo a la inteligencia artificial con recomendaciones de noticias entre sí que den un contexto de la misma. Además, se hizo, tal y como hemos comentado con Feedly, una vista diaria.



Figura 1.5: Logo de Google Noticias

Como se puede ver, y el orden en el que se han comentado los productos digitales muestra su importancia al mundo de los lectores RSS, este es el más novedoso y el que posee un valor añadido superior.

Además, como es sabido, Google tiende a usar un diseño, hilo conductor en todos sus productos. Este era *Material Design*, y ahora *Material Theme*. Dicho esto, se podrá intuir el nivel del usabilidad del mismo, así como de su atracción visual al usuario.

Sin embargo, y este es el motivo de que aparezca en tercer lugar en el estado del arte, no se puede usar en España. Como se ha dicho, 2018 fue un cambio importante desde el punto de vista técnico. También ocurrió lo mismo en 2014, pero a menor escala. En este año, la gran G lanza Google Noticias y Tiempo. Aquí, solo usando la ubicación del usuario, se era capaz de darle unos cuantos titulares de las noticias de rabiosa actualidad.

Los grandes medios españoles se vieron afectados por esta aplicación, ya que se daba igual importancia a un medio local que a los grandes periódicos. Esto, unido al hecho de la nueva legislación española sobre la propiedad intelectual, provoca el cierre de Google News Ana Pérez Barredo (2014).

Este hecho, sin embargo, no ha hecho que disminuya el éxito de la misma en otros países. Si a Feedly o Flipboard le interesa tener su plataforma en diferentes idiomas, qué menos Google. Tan es así, que Google Noticias posee soporte para más de 40 idiomas. Esto es un paso más al tratamiento multi-idioma. Es adaptar la web y las lecturas de las noticias en diferentes abecedarios, tales como el árabe, hindi o chino.

No obstante, posee algunas deficiencias técnicas, que en realidad no son tal. Esto se debe a que Google no le interesa hacer una web de noticias, sino un simple agregador de las mismas. La diferencia radica, por ejemplo, en que a Google no le interesa ser un lugar donde consultar todas las noticias a lo largo del tiempo, sino, más bien, donde poder ver los titulares del momento.

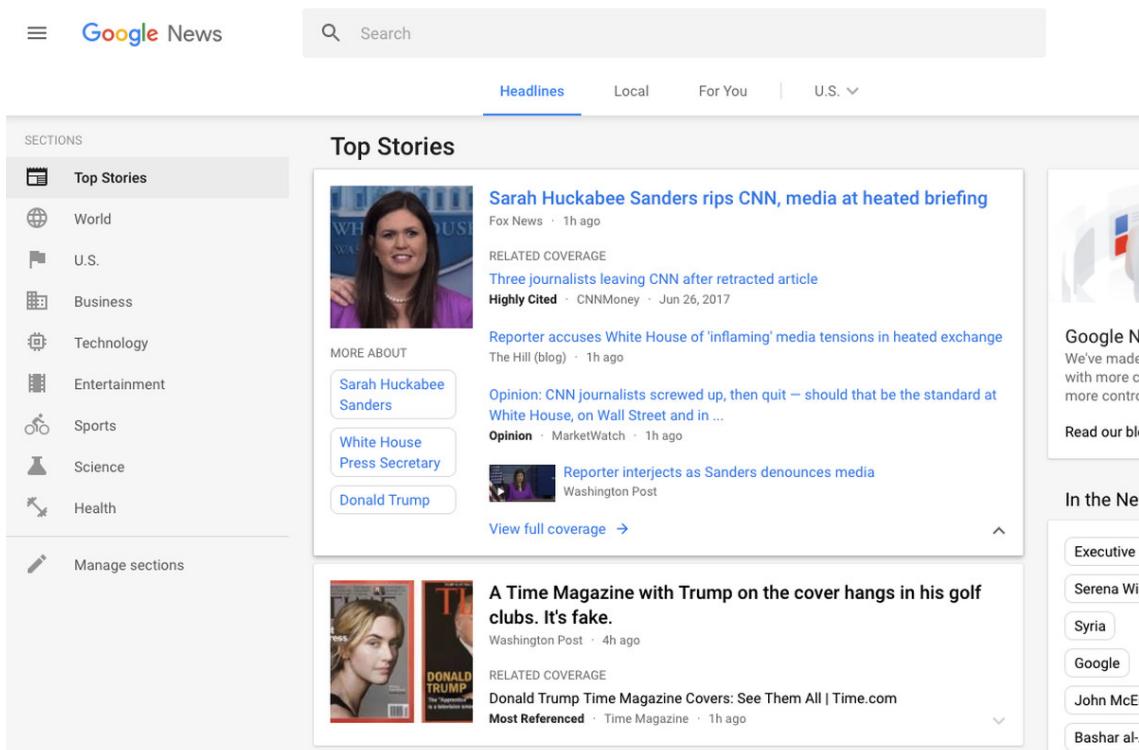


Figura 1.6: Página inicio de Google News

Pues estos detalles son los siguientes. El primero es que las noticias podremos verlas con un retraso de un cuarto de hora, es decir, que se actualizarán cada 15 minutos. El segundo es que

no se intenta hacer un resumen de la noticia en sí, como preámbulo, sino que se redirecciona a la web original para leer la noticia entera. Con esto, se ahorra todos los problemas legislativos subyacentes. La tercera no es si la ya mencionada restricción del número de noticias a guardar. Solo se podrán consultar noticias en un periodo de un mes, sino, Google ya nos las guarda.

1.4.4. Otras

Podemos encontrar otras soluciones al problema de la sindicación de contenido, pero son más simples que las ya comentadas. Algunas a destacar son Feedbin, Inoreader o Winds. Cada cual proponen algo diferente al usuario, aunque, poseen una gran deficiencia cada una que la imposibilitan de triunfar.

Primero en importancia encontramos Feedbin. Esta es una plataforma RSS que recopila noticias de todas las fuentes que selecciona el usuario, funcionando al estilo PaaS (*Platform as a Service*). Esto quiere decir, que los clientes, tanto webs, como de escritorio o de aplicaciones, que encontraremos son independientes al mismo, por lo que la funcionalidad de cada uno diferirá en cada caso.

Este ofrece las características esperadas de cualquier lector RSS en 2019, con algún añadido, propios de la una tecnología puramente de servidor tales como Newsletter, Búsqueda o Podcasts. Logicamente, con un sobrecoste adicional e inicial, ya que está casi por completo orientado al mundo iOS.

En segundo lugar encontramos Inoreader. Su enorme fama se debe a su increíble simplicidad para un usuario con previos conocimiento de sindicación de contenido. Posee muy pocas opciones y solo es accesible desde aplicaciones móviles y además es gratuito, por lo que le hace perfecto para usuarios que estén a caballo entre Feedly y Flipboard: ni la configuración de uno, ni la simpleza del otro.

En tercer lugar tenemos a Winds. Sin duda, si sigue creciendo será la aplicación del mercado en este sector, ya que con tecnologías potentes y novedosas, con una estrategia de código abierto y con una orientación clara a la inteligencia artificial, se mantiene a la vanguardia de la innovación. Actualmente solo está disponible funcionalmente para escritorio y empezando a través de una aplicación web. Sin embargo, este hecho, hace que casi ni se le mencione.

En definitiva, y sobre todo haciendo hincapié en esta tercera, a estas herramientas les hace falta un punto de cocción para ser consideradas en la primera línea en cuanto al dominio del problema se refiere. Sin embargo, o ya han llegado a un público concreto, como las dos primeras, y no se plantean seguir creciendo, o todavía les hace falta más trabajo y más llegada al público objetivo.

CAPÍTULO 2

Estudio

2.1– Metodología LEAN

A medida que va acabando el siglo XX y empezando el XXI, los empresarios se van dando cada vez más cuenta de que los métodos tradicionales para construir empresas han quedado anticuados. Ya no sirven aquellas reuniones interminables con acreedores para que financien proyectos, entre otros motivos, porque no hay sustento para las mismas.

Uno de los padres del nuevo método de emprendimiento es Steve Blank, y él es que se da cuenta y reformula todo esto. Antes, para construir un producto, ya había detrás una empresa, a la que cualquier banco estaba dispuesto a pagarle. Sin embargo, ahora, se empiezan a construir empresas para desarrollar productos, por lo que no tienen ningún aval por detrás.

Va pasando el tiempo, Steve Blank va enseñando su manera de emprender y aparecen dos personajes que van construyendo el mismo bosque aunque con diferentes árboles. Estos son Eric Ries y Alex Osterwalder. El primero aplica la metodología LEAN al emprendimiento. El segundo, convierte el plan de acción de una empresa a un gráfico sencillo de implementar. Además, Blank llega hasta un concepto revolucionario: el desarrollo de los clientes.

Pues bien, en todo este marco de trabajo nace la teoría llevada a la práctica de Ash Maurya y su Running Lean. En este, no hace más que unir los tres conceptos anteriormente mencionados: *Customer Development, Business Model Canvas* y *Lean Startup*. Sin embargo, consigue algo que ninguno de los tres si quiera ha llegado: llevarlo a la práctica.

Ash hace en su libro un rápido plan de acción para llevar a cabo el método LEAN a la construcción de una empresa que triunfe siguiendo una serie de procesos. Estos, se pueden resumir en el siguiente gráfico que aparece en su libro.

El resumen del libro se puede encontrar en la siguiente frase del mismo, aparecida en el primer capítulo del libro.

Running Lean is a systematic process for iterating from Plan A to a plan that works, before running out of resources.

Mauyria (2011)

En definitiva, no es más que documentar el plan inicial que se posee de la idea, identificar los posibles riesgos e ir sistemáticamente evolucionando nuestro plan hasta dar con el que el cliente comprará. Todo esto, como dice la cita, sin llegar a agotar los recursos. ¿Para qué es necesaria todas estas fases? Para llegar a conocer cuáles son los intereses reales del cliente y esto lleva tanto porque el mismo cliente no sabe qué es lo que quiere.

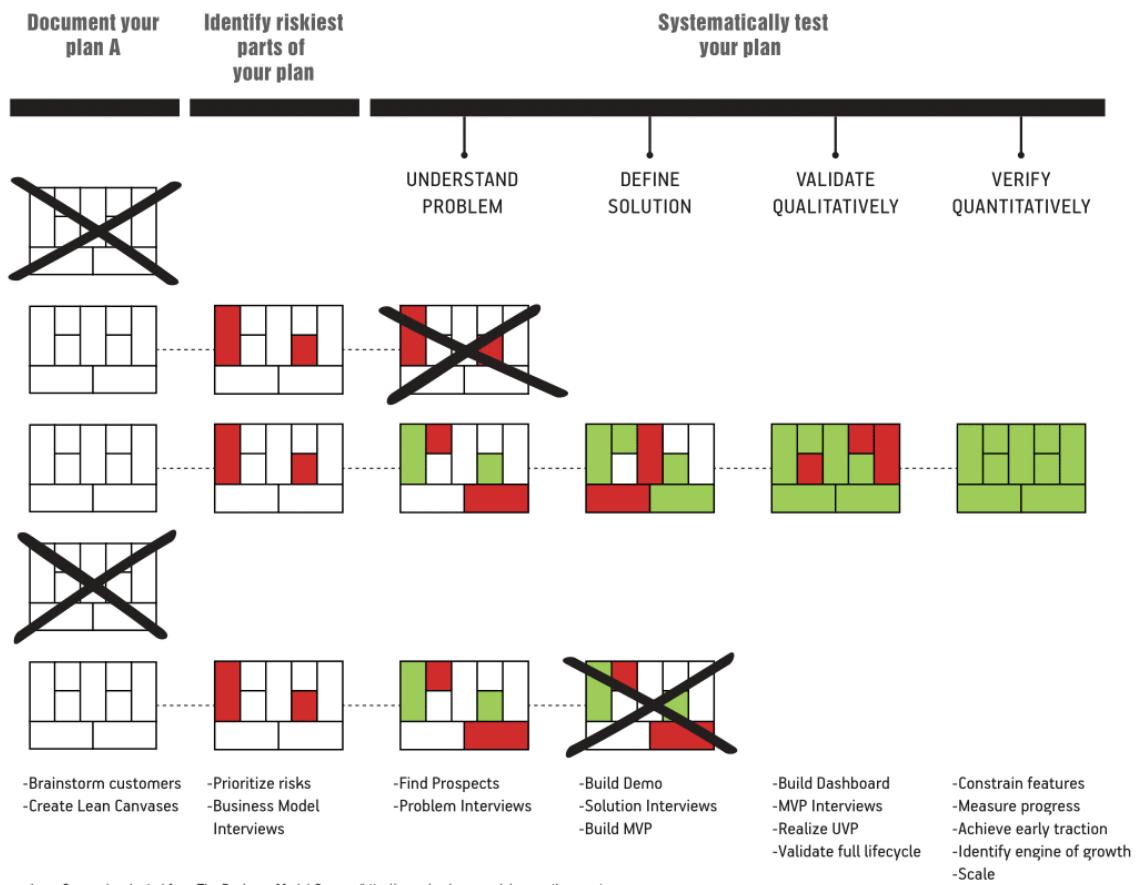


Figura 2.1: Resumen Running Lean

2.1.1. Running Lean

Una visto un enfoque general de la metodología LEAN, veamos cómo funciona operativamente el Running Lean de Ash Maurya. Como hemos dicho él distingue tres etapas, por lo que vamos a comentar brevemente cada una de las etapas.

Documentar el plan A

La primera es la documentación de un plan inicial. Esto no es más que plasmar en un papel las ideas que tenemos sobre nuestro producto. En vez de realizar un documento típico con todas sus partes y rellenando cada una de las casillas que definen un producto tipo, Ash nos proporciona un *Lean Canvas*. Este es la evolución del *Bussiness Canvas* de Osterwalder. En este rellenaremos los atributos imprescindibles que definirán nuestro producto con una condición: ha de ser un prototipo fácil de realizar y que nos quite poco tiempo.

Así, en un simple papel podemos llenar este y empezar a trabajar sobre el mismo. ¿No nos sirve? A tirarlo y a empezar de nuevo. Además que en este podemos ver de un simple vistazo las partes claves de nuestra idea de negocio y por ello compartirlo para pedir opinión. Esto es fundamental, ya que seguramente seremos de las personas que más sepan del tema, y no seamos totalmente objetivos.

Este se llena siguiendo un orden establecido, que nos ayudará a formar nuestra cabeza de emprendedor para centrarnos realmente en lo importante. Este podemos verlo en la siguiente imagen.

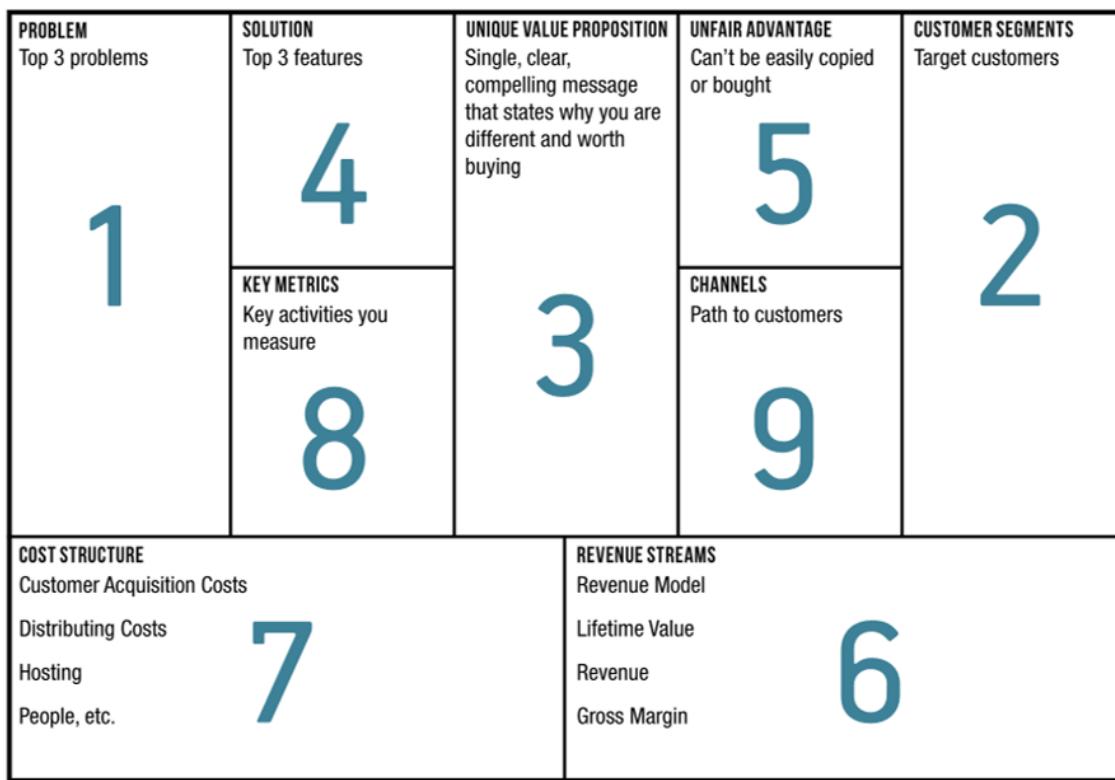


Figura 2.2: Lean Canvas

Las partes más importantes del gráfico son los tres primeros puntos: problema, segmento de clientes y propuesta de valor única. Tan es así, que Ash afirma que:

Investors, and more important, customers, identify with their problems and don't care about

your solution (yet). Entrepreneurs, on the other hand, are naturally wired to look for solutions. But chasing after solutions to problems no one cares enough about is a form of waste.

Mauyria (2011)

Es por tanto primordial encontrar el problema que queremos solucionar, algo no trivial, porque debemos clavar cuál es la necesidad real de nuestro cliente para así satisfacerla. Este gráfico en sí recoge ya toda la teoría de Maurya, que iremos explorando.

Identificar los riesgos de nuestro plan

Una vez tenemos nuestro primer esbozo de nuestra idea, con una primera aproximación a las partes esenciales de nuestro producto y cómo vamos a vender este en el mercado, tenemos que identificar unos riesgos iniciales. Sabemos que todavía no hemos llegado a la madurez de la idea, pero es conveniente conocer ya los puntos flacos de esta idea.

Ash Maurya, de nuevo, nos ayuda con esto. Lo hace enseñándonos cuales son las etapas por las que va a pasar nuestro producto. Además, en cada una de estas, para que no perdamos el foco, nos muestra qué tenemos que priorizar. Esta se puede ver en el siguiente gráfico.



Figura 2.3: Adecuación de nuestra idea al mercado

Para llegar a esos riesgos iniciales sin saber si quiera cuándo ni cómo se van a afrontar, nos ilustra Ash con las siguientes preguntas.

- **Problem/Solution Fit:** *Do I have a problem worth solving?*
 - *Is it something customers want? (must-have)*
 - *Will they pay for it? If not, who will? (viable)*
 - *Can it be solved? (feasible)*
- **Problem/Market Fit:** *Have I built something people want?*
- **Scale:** *How do I accelerate growth?*

Se centra sobre todo en el primer punto porque, como nos recomienda, hemos de pivotar para llegar al problema correcto a solucionar. Esto nos ayudará cuando testeemos nuestro producto en la siguiente fase. Pero ya, contestamos a las preguntas, podemos ver los riesgos a los que nos veremos sometidos. Estos serán mayoritariamente de cinco tipos.

El primero hará referencia al problema. Este nos dirá cuánto *dolor* le quita al posible comprador. Esto es importante, porque podemos ver un problema real, que ocurra a un número alto de personas. Sin embargo, si este lo ven como algo accesorio, nunca conseguiremos venderlo. Hay que ver, por tanto, si el problema merece la pena afrontarlo por el nivel de importancia que le da el usuario.

El segundo consiste en cómo llegamos a nuestro público, qué tipo de canales hemos de utilizar. Al igual que antes, si tenemos un buen producto, que además soluciona un problema importante, pero nuestro público objetivo no lo conoce, no venderemos nada.

El tercero se relaciona con la estructura que hayamos previsto de pérdidas y ganancias. Una vez que hayamos madurado la idea y encontrado nuestro público y cómo llegar a él, hemos de comprobar cómo nos sostendremos económicoamente. Por ello, hemos de analizar nuestro margen de beneficio y la estructura de costes e ingresos.

El cuarto consistirá en el tamaño del mercado. Tenemos que comprobar cuál es nuestro nicho de mercado y con quién hemos de competir. Con esto, aseguraremos la audiencia.

El quinto y último a analizar será comprobar si técnicamente se puede implementar nuestra solución. Solo con este criterio, se puede cambiar o alterar ligeramente nuestra idea de negocio.

Probar sistemáticamente el plan

Ya tenemos nuestro *canvas* y los posibles riesgos asociados. Ahora nos falta ir perfilándolo para llegar la madurez completa del producto. Las fases que observaremos en este testeo sistemático son las siguientes.

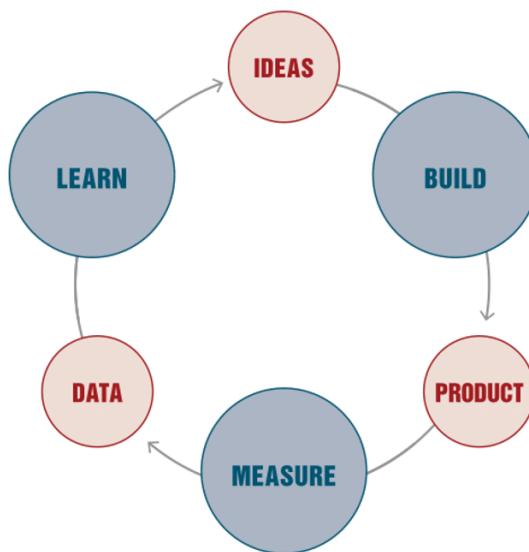


Figura 2.4: Prueba sistemática

Esta es la genialidad de Ash Maurya: el cómo convertir una metodología etérea en un algoritmo a seguir. Aquí unifica las ideas de Blank y Ries, porque es llevar la oficina a la calle. Esto no es otra cosa que preguntar y asentar ideas, preguntar y cambiar conceptos, preguntar y llegar a lo que la gente quiere.

Es por tanto primordial en esta fase salir a la calle y hacer entrevistas, muchas entrevistas. Estas nos llevarán a cambiar nuestro *canvas* y acercarnos más a la idea final.

Esta última parte contiene a su vez cuatro fases, que hemos de ir quemando para llegar a la siguiente con algo cada vez más claro. En cada una de estas, hemos de aprender algo.

Primero serán las entrevistas de problema. En estas hemos de validar las hipótesis sobre el par: problema y segmento de cliente. De manera gráfica, hemos de *aprender* en esta fase lo siguiente.

Una vez hayamos aprendido esto, podremos saltar al siguiente punto. Esto se materializa si podemos identificar a nuestro *early adopter*, es decir, nuestro primer usuario; si tenemos un problema que merezca la pena analizar; y si podemos averiguar cómo los usuarios potenciales resuelven a día de hoy este problema.

Risk	Learn
Product risk: What are you solving?	How do customers rank the top three problems?
Market risk: Who is the competition?	How do customers solve these problems today?
Customer risk: Who has the pain?	Is this a viable customer segment?

Cuadro 2.1: Aprendizaje entrevista de problema

Una vez conseguido esto, llegaremos a las entrevistas de solución. Poseen la misma dinámica que la anterior, solo que ahora cambia lo que queremos aprender. Además, hemos de tener en cuenta que estamos en la antesala de la salida a ventas del producto.

Risk	Learn
Customer risk: Who has the pain?	How do you identify early adopters?
Product risk: How will you solve it?	What is the minimum feature set needed to launch?
Market risk: What is the pricing model?	Will customers pay for a solution?

Cuadro 2.2: Aprendizaje entrevista de solución

Llegados a este punto, tendremos unas características mínimas de el producto, un precio al que el cliente está dispuesto a pagarte y un negocio alrededor del producto.

Ya tenemos problema y solución, vamos a implementarla. Ahora, con lo que sabemos, no nos interesa hacer enrolarnos en hacer una aplicación 100 % funcional, hemos de implementar solo aquello imprescindible para que sea vendida. Con esto, nos aseguramos guardar nuestro dinero y llegar a este Mínimo Producto Viable (MVP en adelante) para comprobar si nuestro cliente era lo que realmente quería. Una vez hecho, volvamos a las entrevistas, pero sobre este MVP. Ahora, lo que queremos aprender será lo siguiente.

Risk	Learn
Product risk: What is compelling about the product?	Landing Page, Activation Flow and UVP
Customer risk: Do you have enough customers?	Channels
Market risk: Is the price right?	Price

Cuadro 2.3: Aprendizaje entrevista del MVP

Ya tendremos nuestra página explicativa del producto (*Landing Page*), nuestro canal desde que el usuario llega a esta hasta que se interesa por nuestro producto (*Activation Flow*) y un MVP que cumpla el UVP definido. Si en las entrevistas vemos que convence queda solo lanzarlo al exterior y ver si tenemos los canales y precio adecuados. Esta será la última fase, continua a lo largo del tiempo: medir la adecuación entre el producto y el mercado.

Aquí veremos si convence nuestro producto y cuál es el modo de llegar a que los usuarios paguen. Ya los *early adopter* habrán entrado en el flujo de clientes, por lo que tenemos que tirar de ellos para conseguir más usuarios. Además, ahora recibiremos continuas propuestas para mejorar nuestro MVP que tendremos que analizar si compensa.

2.1.2. Aplicación real

Como se puede observar, esta metodología es totalmente practicable a un proyecto real. De una parte, porque parte de la experiencia de varios autores ilustrados sobre el tema llevando este a la práctica. De otra, porque posee una serie de pautas para aplicarlo en un entorno totalmente compartido.

Así, lo que se ha realizado en este proyecto, es la aplicación de la metodología LEAN, con su variante en el Running LEAN para proyectos de emprendimiento, a una idea propia. Esta, primero, era una idea puramente técnica, y con la ayuda y guías de este marco de trabajo, ha terminado por

convertirse en una idea de negocio, al que poder dedicarle jornadas y jornadas de trabajo para llegar de la rentabilidad básica del mismo a un beneficio alto.

La correcta aplicación y su interiorización para poder llevarla a cabo ha ido de la mano de dos hecho. El primero es la realización de las asignaturas *Fundamentos de Innovación y Emprendimiento*, FEI, y *Taller de Innovación y Emprendimiento*, TEI. A través de estas asignaturas del máster hemos estudiado y llevado a la práctica todos estos conceptos.

En FEI y TEI, en concreto, teníamos que llevar un proyecto a consecución siguiendo las claves que propone Maurya en su libro Mauyria (2011). Yo, para poder entenderlo con todas sus variantes, he utilizado este proyecto, mi TFM, como idea de negocio. Así, he ido viendo problemas y buscado resolución, he recibido feedback de parte de los profesores y alumnos, he aprendido de los proyectos ajenos. Como se ve, podría ser la fiesta del saber en cuanto a emprendimiento se refiere.

El segundo hecho importante para llevar este Estudio a término ha sido la lectura y estudio del libro tantas veces citado a lo largo de este apartado. Running Lean me ha enseñado que lo que he aprendido en las asignatura de Emprendimiento del Máster tienen bastante de fundamente real e innovador. Además, estudiado el libro he comprobado cómo llevar las teorías de Maurya hasta límites insospechados.

Así, a lo largo de los siguientes puntos, iré desgranando los pasos que se indican en Running Lean y cómo los he llevado a la práctica en mi proyecto. Además, se indicarán una serie de conclusiones sacadas de la realización del mismo.

Como se puede intuir viendo lo anterior, no tendrá sentido aplicar totalmente la metodología ya que la consecución del TFM no es otra que el aprendizaje de tecnologías y conocimientos, y no tanto la venta de productos. Esto, aunque no descartado en un marco futuro, no se realizará en las iteraciones LEAN. Por tanto, se quedará el Estudio hasta el conocimiento certero de la problema y su posterior solución.

2.2– Primer estudio

LT-News, como se ha dicho, es un lector de noticias inteligentes, capaz de detectar los gustos de los usuarios en base a su utilización de la aplicación, de relacionar noticias entre sí y de detectar las noticias más importantes del día.

El objetivo principal del proyecto es filtrar las noticias de los periódicos y quedarse solo con aquellas realmente interesantes para el usuario. El objetivo secundario y subyacente al anterior no es otro que el de no perder el tiempo.

2.2.1. Segmento de clientes

El cliente objetivo es lector actual de noticias, joven o de mediana edad con cultura e interés sobre temas de actualidad y con manejo de las nuevas tecnologías.

2.2.2. Problema

El top tres de problemas son los siguientes. Primero la gran cantidad de medios a elegir, por tanto, de noticias a leer. La segunda es que las noticias a leer noticias no son objetivas, es decir, hay las noticias falsas y el sesgo de los periódicos. La tercera es que, entre tanta cantidad de noticias, se pierden las noticias que realmente interesan.

2.2.3. Propuesta de valor único

La UPV es el algoritmo de inteligencia artificial capaz de extraer el perfil del usuario y de las noticias y de relacionarlas entre sí. Dicho de otro modo, ahorrar tiempo a la vez que estar al día.

-Gran cantidad de medios a elegir. -Gran cantidad de noticias a leer. -Fakes news y sesgo de los medios. -Perderse novedades de tus intereses.	- Recomendación de noticias. -Extracción del perfil de usuario y recomendar noticias. -Análisis diario de prensa. -Tiempo pasado en aplicación. -Número de usuarios. -Suscripciones mensuales.	-Algoritmo de inteligencia artificial capaz de extraer el perfil del usuario y de las noticias y relacionarlos entre sí. -Ahorrar tiempo y a la vez que estar al día.	-Análisis de textos de las noticias. -Webscrapping de las noticias en los medios. -Web. -Aplicaciones.	-Usuarios lectores de noticias. -Jóvenes y mediana edad. -Cultura e interés actualidad. -Personas asociadas a la tecnología.
-Coste de servidores.		-Aplicación premium: 3€/mes.		

Figura 2.5: LEAN Canvas inicial

2.2.4. Principales hipótesis

- A los lectores de noticias les preocupa la gran cantidad de noticias.
 - Es la hipótesis más arriesgada porque si la respuesta es que no, hay que pivotar completamente.
- A los lectores de noticias les preocupa que las noticias no sean objetivos.
 - Es una hipótesis que requiere ser probada para seguir validando el producto, aunque es una funcionalidad secundaria.
- A los lectores de noticias les interesa estar al día de sus temas favoritos.
 - Es otra hipótesis secundaria para comprobar.

2.3– Iteraciones

El resumen de las iteraciones realizadas sobre el producto se encuentra en esta tabla.

Iteración	Hipótesis	Fechas	Tests	Resultado
Problema 1	Primera hipótesis	S/29-09	2 entrevistas	No aplica
Problema 2	Primera hipótesis	S/06-10	4 entrevistas	Se confirma
Problema 3	Segunda hipótesis	S/13-10	4 entrevistas	Se confirma
Problema 4	Tercera hipótesis	S/20-10	143 encuestas	No se confirma
Solución 1	Soluciones	S/27-10	4 entrevistas	Se confirma

Cuadro 2.4: Iteraciones realizadas

El principal pivote realizado es simplificar el Producto Mínimo Viable, ya que hay un problema que no afecta de igual manera a los early adopters, por tanto, aunque sea importante, no se incluirá en la primera versión. Todos los demás problemas se han confirmado al cien por cien.

Con respecto a las soluciones, se han confirmado, aunque los usuarios piden más características. Dado que en la primera versión ya se ha demostrado un MVP aceptable por los usuarios, se tomarán estas características para futuras versiones.

En relación a los *early adopters*, he cambiado su percepción. Antes los consideraba como personas cultas interesadas por lo que ocurre en la actualidad y que siguen una serie de temas. Ahora, veo que es cualquier que ya use un lector RSS o aplicación de un periódico. En concreto, me centro en los que usen Feedly.

2.3.1. Primera iteración

La primera iteración consistió en confirmar la primera hipótesis, es decir, confirmar el primer problema. Para ello, se hicieron dos entrevistas. De ellas no se sacaron ninguna conclusión, simplemente se redefinió lo que se consideraba por *early adopters*. Hasta este momento eran jóvenes, por interés por la cultura y con gran uso de smartphone y ordenadores. Fue entonces cuando comprobé que estos eran más bien personas que ya usan un lector RSS o, sin más, una aplicación de información en su móvil.

Son estos mis *early adopters* por un motivo: a estos usuarios les interesa estar al día de diferentes medios y ya han buscado una solución para estar al día, que no es otra que utilizar un lector RSS. Además, si llevan tiempo utilizándolo, se habrán dado cuenta del principal problema de este: la enorme cantidad de noticias a leer.

-Gran cantidad de medios a elegir.	- Recomendación de noticias.	-Algoritmo de inteligencia artificial capaz de extraer el perfil del usuario y de las noticias y relacionarlos entre sí.	-Análisis de textos de las noticias.	-Usuarios activos de lectores de RSS.
-Gran cantidad de noticias a leer.	-Extracción del perfil de usuario y recomendar noticias.	-Análisis diario de prensa.	-Webscrapping de las noticias en los medios.	-Usuarios que suelen usar Feedly.
-Fakes news y sesgo de los medios.	-Tiempo pasado en aplicación.	-Ahorrar tiempo y a la vez que estar al día.	-Web.	
-Perderse novedades de tus intereses.	-Número de usuarios.	-Aplicaciones.		
-Coste de servidores.		-Aplicación premium: 3€/mes.		

Figura 2.6: LEAN Canvas primera iteración

2.3.2. Segunda iteración

Viendo lo anterior, se ha decidido volver a preguntar lo mismo, pero a los verdaderos usuarios potenciales. Como se dijo, fue importante confirmar la primera hipótesis, ya que de esta partía todo el problema, y, por ende, todo el proyecto. Es por ello que se realizaron cuatro entrevistas siguiendo el siguiente esquema:

1. Welcome
 - a) Presentación personal

b) Qué quiero de esta entrevista

2. Collect Demographics

- a) Nombre
- b) Email
- c) Edad
- d) Código Postal
- e) Uso de apps de noticias
- f) Veces a la semana que lee el periódico

3. Tell a Story

- a) Contexto de la cantidad de noticias en la actualidad.
- b) Leer en papel vs Leer Digital: actualidad vs concentración.

4. Problem Ranking

- a) ¿Qué problemas tienes a la hora de estar al día de todas las noticias que te interesan?
Dime en orden.
- b) ¿Piensas que las noticias que lees están dadas desde un punto de vista o pueden ser falsas? ¿Te preocupan?
- c) ¿Alguna vez has visto una noticia de un tema que te interesaba y te ha dado pena no haberte enterado antes?
- d) ¿Usas alguna aplicación de noticias?

5. Explore Customer Overview

- a) Ver su opinión sobre la manera de leer noticias

6. Wrapping Up

- a) Te gustaría una aplicación donde: ¿se recojan todas las noticias, te las relaciona y te las recomienda?

7. Document Results

- a) Pasarlo a un Google Forms

Sobre esto se ha conseguido confirmar la primera hipótesis de manera rotunda, aunque la segunda y la tercera no ha quedado del todo claras. Es claro que problema principal existe y la gente que ya lee en sindicadores de contenido ha intentado plantear soluciones en su día a día para solventarlo.

Sobre las entrevistas y el feedback recibido en clase, sí que se han modificado los canales. Antes eran un poco genéricos; ahora, intentan llegar más a los early adopter. Por tanto, se va a intentar que publiquen la aplicación en blog de aplicaciones, se van a mencionar en blogs y foros especializados en el tema y se va a intentar hacer SEO de la aplicación con los tags: aplicación, noticias, inteligente. Algunas de estas podrían ser:

- El Androide Libre - <https://elandroidelibre.elespanol.com>
- Xataka - <https://www.xatakandroid.com/>
- Android4All - <https://andro4all.com/>

-Gran cantidad de medios a elegir. -Gran cantidad de noticias a leer. -Fakes news y sesgo de los medios.	-Recomendación de noticias. -Extracción del perfil de usuario y recomendar noticias. -Análisis diario de prensa.	-Algoritmo de inteligencia artificial capaz de extraer el perfil del usuario y de las noticias y relacionarlos entre sí.	-Análisis de textos de las noticias. -Webscrapping de las noticias en los medios.	-Usuarios activos de lectores de RSS. -Usuarios que suelen usar Feedly.
-Perderse novedades de tus intereses.	-Tiempo pasado en aplicación. -Número de usuarios. -Suscripciones mensuales.	-Ahorrar tiempo y a la vez que estar al día.	<p style="color: red;">-Blogs sobre aplicaciones.</p> <p style="color: red;">-Foros especializados.</p> <p style="color: red;">-SEO de las keywords.</p>	
-Coste de servidores.		-Aplicación premium: 3€/mes.		

Figura 2.7: LEAN Canvas segunda iteración

2.3.3. Tercera iteración

La tercera iteración consistió en confirmar, de manera sistemática y similar a la anterior, las dos hipótesis restantes. Para ello, se volvieron a hacer cuatro entrevistas. De estas salió validada la segunda hipótesis y algo clara la tercera, es decir, la mitad lo afirmaron como problema y la otra mitad, no les parecía relevante.

Aquí también salieron algunas características adicionales para añadir al producto que se considerarán en el futuro, ya que no están dentro del producto mínimo viable y son las siguientes:

- Sacar de una noticia su relación al tema de fondo o contexto. Para esto es necesario llegar a contenido de calidad.
 - Una aproximación podría ser relacionar temas de las noticias con Wikipedia o fuentes relevantes y más profundas.
 - Para móviles Android, hacer que la aplicación tenga Widgets, ya que algunos usuarios estaban interesados.
 - Sobre los resultados de las entrevistas, hemos cambiado la formulación de los dos primeros problemas.

2.3.4. Cuarta iteración

Como se ha dicho, después de haber validado las dos primeras hipótesis como Producto Mínimo Viable, quedaba la tercera. Dado que las entrevistas individuales en este caso resultaban inefficientes porque no había ninguna opinión que esclareciera el asunto, se decidió realizar una encuesta y pasarla por diferentes canales. Gracias a este medio se consiguieron tres cosas, en este orden de importancia:

- confirmar la opinión sobre la tercera hipótesis,
 - saber su futuro comportamiento con respecto a la aplicación,

<p>-Enorme cantidad de noticias a leer.</p> <p>-Asegurar objetividad de las noticias.</p> <p>-Perderse novedades de tus intereses.</p>	<ul style="list-style-type: none"> - Recomendación de noticias. - Extracción del perfil de usuario y recomendar noticias. - Análisis diario de prensa. 	<ul style="list-style-type: none"> -Algoritmo de inteligencia artificial capaz de extraer el perfil del usuario y de las noticias y relacionarlos entre sí. 	<ul style="list-style-type: none"> -Análisis de textos de las noticias. -Webscrapping de las noticias en los medios. 	<p>-Usuarios activos de lectores de RSS.</p> <p>-Usuarios que suelen usar Feedly.</p>
	<ul style="list-style-type: none"> -Tiempo pasado en aplicación. -Número de usuarios. -Suscripciones mensuales. 	<ul style="list-style-type: none"> -Ahorrar tiempo y a la vez que estar al día. 	<ul style="list-style-type: none"> -Blogs sobre aplicaciones. -Foros especializados. -SEO de las keywords. 	
-Coste de servidores.		-Aplicación premium: 3€/mes.		

Figura 2.8: LEAN Canvas tercera iteración

- qué le ven en falta.

Se les preguntó, sacado de lo anteriormente, datos demográficos y confirmación de las tres hipótesis. No fueron solamente preguntas exclusivas, ya que no me darían ningún dato. Se les incluyó un lugar donde introducir texto sobre cómo intentan resolver su problema. Así se puede ver cuánto le interesa realmente eso que ha respondido.

El resumen fue el siguiente (la encuesta está en <https://goo.gl/forms/UJHY6o6lV3JxmM712>).

Como se pudo ver, no quedó confirmado la tercera hipótesis, por lo que se quitará del Producto Mínimo Viable y se añadirá como primera característica futura, ya que interesa a la mitad de los usuarios. Por otra parte, se reafirmaron las dos primeras hipótesis. Además, salieron de ahí diferentes opiniones, así como futuras características a considerar en un futuro. Algunos son:

- Hacer una aplicación que consuma poco espacio.
- Hacer cómoda la lectura del cuerpo de la noticia.
- Evitar los problemas de conexión a internet.

Con esto, nos queda el siguiente Canvas:

2.3.5. Quinta iteración

Por último, una vez confirmados los problemas, se hicieron dos entrevistas de solución a usuarios potenciales de la aplicación, dispuestos a probar todas las versiones de este. Estos son realmente *early adopters*, dispuestos a pagar en cuanto salga al mercado si ven un producto fiable.

Para esto, se procederá a enseñar una demo. Esta se hará primero con la *Landing Page*. En esta se ha intentado mostrar las futuras características del proyecto, así como recoger datos de usuarios. Esta se encuentra en varios idiomas, con idea de empezar a distribuirla por personas que habla inglesa o incluso francesa, ya que se pasó la encuesta a amigos que viven en estos países. Esta es la página resultante:

Por otra parte, se le dio tráfico, para saber la opinión de los posibles usuarios.

Con esto, se hicieron las entrevistas de solución. La estructura de la entrevista fue la siguiente:

Edad

143 respuestas



Figura 2.9: Encuesta primera pregunta

¿Cuantas horas a la semana dedicas a leer noticias?

143 respuestas



Figura 2.10: Encuesta segunda pregunta

¿Qué aplicación sueles usar para leer noticias?

138 respuestas



Figura 2.11: Encuesta tercera pregunta

¿Intentas estar al día de lo que pasa en el mundo?

138 respuestas

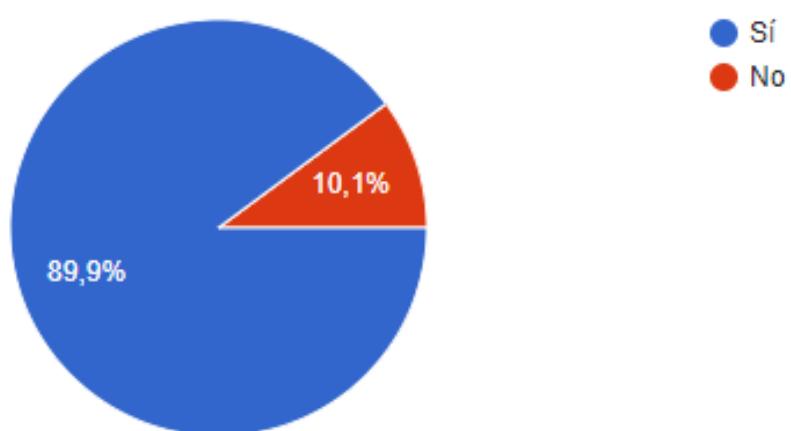


Figura 2.12: Encuesta cuarta pregunta

¿Te preocupa que las noticias que lees sean falsas?

138 respuestas



Figura 2.13: Encuesta quinta pregunta

¿Ves que muchas veces lees noticias que no sabes el contexto o la cronología de los hechos?

138 respuestas

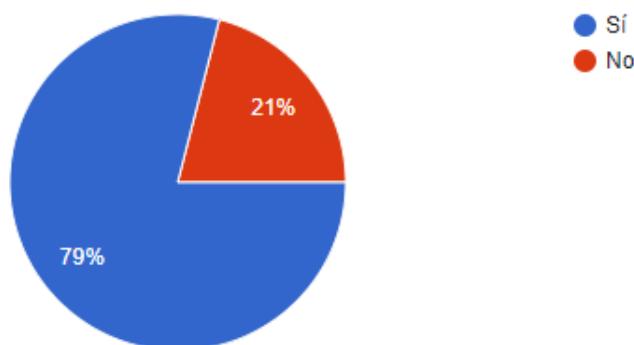


Figura 2.14: Encuesta sexta pregunta

¿Intentas estar a la última sobre los temas o noticias que te interesan?

138 respuestas

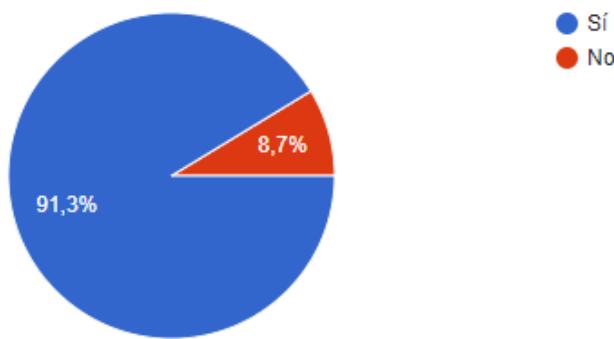


Figura 2.15: Encuesta séptima pregunta

¿Te preocupa que muchas veces te enteras de estos temas tarde o por otros, en vez de por tí mismo?

138 respuestas

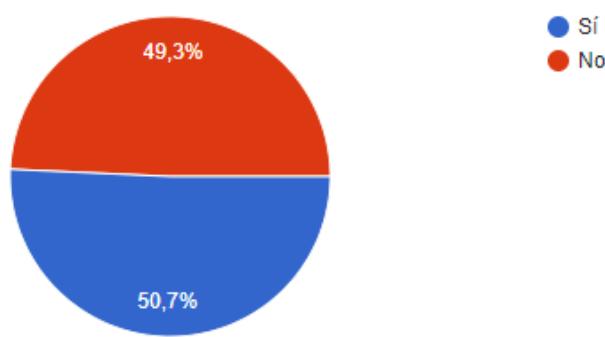


Figura 2.16: Encuesta octavo pregunta

-Enorme cantidad de noticias a leer. -Asegurar objetividad de las noticias.	- Recomendación de noticias. -Extracción del perfil de usuario y recomendar noticias. -Análisis diario de prensa.	-Algoritmo de inteligencia artificial capaz de extraer el perfil del usuario y de las noticias y relacionarlos entre sí.	-Análisis de textos de las noticias. -Webscrapping de las noticias en los medios.	-Usuarios activos de lectores de RSS. -Usuarios que suelen usar Feedly.
	-Tiempo pasado en aplicación. -Número de usuarios. -Suscripciones mensuales.	-Ahorrar tiempo y a la vez que estar al día.	-Blogs sobre aplicaciones. -Foros especializados. -SEO de las keywords.	
-Coste de servidores.		-Aplicación premium: 3€/mes.		

Figura 2.17: LEAN Canvas cuarta iteración



Figura 2.18: Landing Page 1



Figura 2.19: Landing Page 2

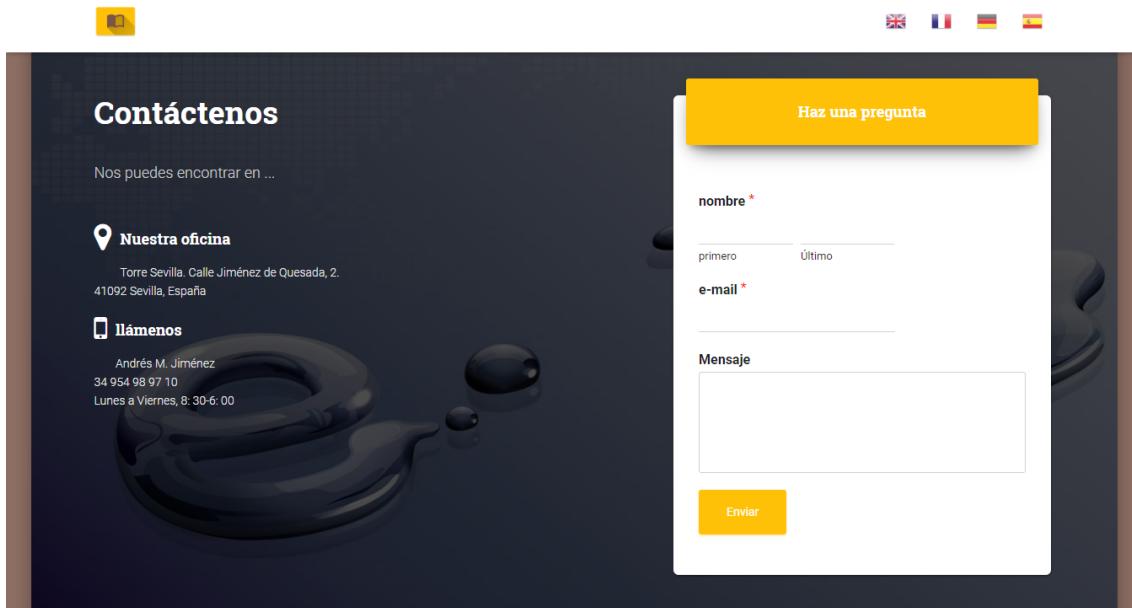


Figura 2.20: Landing Page 3

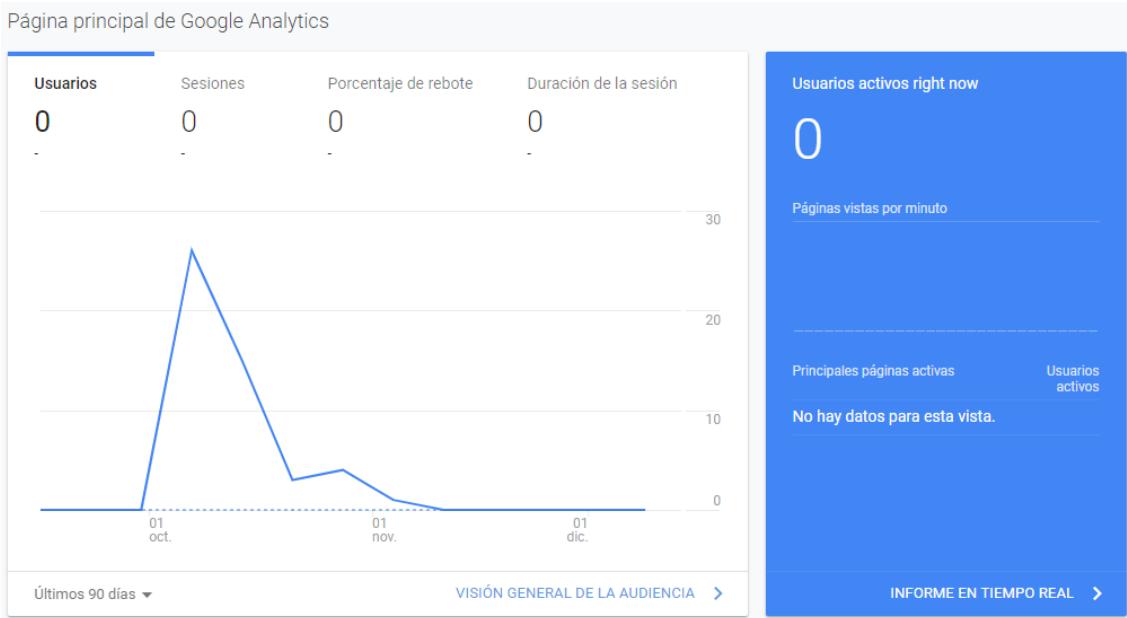


Figura 2.21: Google Analytics 1

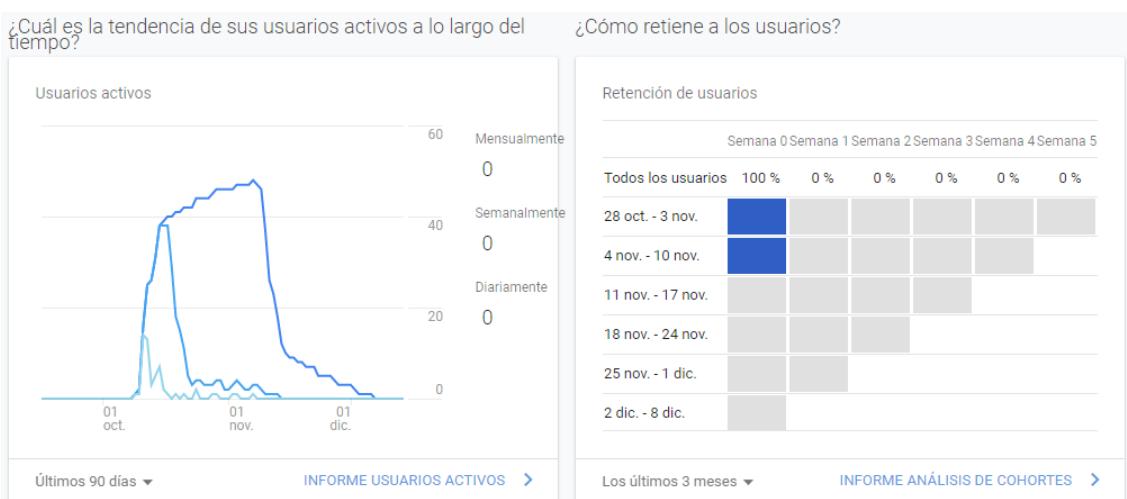


Figura 2.22: Google Analytics 2



Figura 2.23: Google Analytics 3

1. Welcome

- Presentación personal
- Qué quiero de esta entrevista

2. Collect Demographics

- Nombre
- Email
- Edad
- Código Postal
- Uso de apps de noticias
- Veces a la semana que lee el periódico

3. Tell a Story

- Contexto de la cantidad de noticias en la actualidad
- Leer en papel vs Leer Digital: actualidad vs concentración

4. Demo

- Le enseño la Landing Page
- Hablamos de cómo voy a solucionar estos problemas y le pregunta su opinión

5. Test Pricing

- ¿Qué te parece el plan de pagos?

6. Wrapping Up

- ¿Me podrías dar tu email para seguir contándote cómo sigue el proyecto?
- ¿Conoces a alguien que pueda estar interesado en el proyecto?

7. Document Results

a) Pasarlo a un Google Form

En sendas entrevistas, se volvieron a confirmar la solución, explicando lo que se iba a realizar y cómo se implementaría. Pareció correcto, estando los *early adopters* expectantes a futuras versiones.

2.3.6. Quinta iteración

Finalmente, validado ya el canvas, tanto a nivel de solución como de problemas, tanto cualitativa como cuantitativamente, ha variado en algunos aspectos que vamos a considerar. Todo lo demás, en definitiva, se ha explicado al principio del apartado.

El mayor cambio ha consistido en la variación de concepto del usuario potencial, esto gracias a las entrevistas realizadas. Otro gran cambio ha sido la formulación de problemas del usuario, que se han adaptado a un lenguaje más cercano a este, así como el evitar generalidades. De esto, como se ha visto, ha sido reducido para conseguir un MVP más conciso.

De las soluciones, se ha mantenido como estaba inicialmente, ya que ha sido validado por los usuarios. Esto se debe a que previamente a estos experimentos, se buscó las mejores soluciones posibles.

-Enorme cantidad de noticias a leer. -Asegurar objetividad de las noticias.	- Recomendación de noticias. -Extracción del perfil de usuario y recomendar noticias. -Análisis diario de prensa.	-Algoritmo de inteligencia artificial capaz de extraer el perfil del usuario y de las noticias y relacionarlos entre sí. -Tiempo pasado en aplicación. -Número de usuarios. -Suscripciones mensuales.	-Análisis de textos de las noticias. -Webscrapping de las noticias en los medios. -Ahorrar tiempo y a la vez que estar al día.	-Usuarios activos de lectores de RSS. -Usuarios que suelen usar Feedly. -Blogs sobre aplicaciones. -Foros especializados. -SEO de las keywords.
-Coste de servidores.		-Aplicación premium: 3€/mes.		

Figura 2.24: LEAN Canvas final

2.4– Conclusión estudio

La mayor conclusión no ha sido otra que el centrarse durante todo el proceso creativo y sistemático en el posible usuario. Esto ha sido posible gracias a entrevistas personales, cercanía con los posibles usuarios, adaptación del lenguaje técnico, así como otras técnicas sacadas por Ash Maurya en Running Lean.

De otra manera, personalmente, hubiese adoptado por otra técnica. Esta hubiese sido construir el producto como entendía e intentar después venderlo. Ante opiniones de usuarios, hubiese desarrollado sin ton ni son, solo por el simple hecho de vender.

Así, en cambio, sin haber construido nada, sin invertir tiempo ni cabeza en la implementación, ya vislumbro que el lo que comprará un posible usuario. Esto hace el trabajo más eficiente, así como más humilde. Humilde por el hecho de no canonizar mis ideas preconcebidas, sino pasarlas por el crisol de la opinión de los demás.

CAPÍTULO 3

Planificación

3.1– Metodología

Todo proyecto que esté englobado dentro de la ingeniería del software posee una metodología que proporciona un marco de trabajo estructurado, planificado y controlado para el proceso de desarrollo. Sin embargo, no hay una única metodología a aplicar, sino más bien múltiples y variadas. Lo importante no es tanto escoger la *perfecta*, que no existe, sino ver aquella o aquellas que mejor se ajusten a la naturaleza del proyecto en cuestión.

Como se ha podido observar anteriormente, nos hace falta una metodología que agilice el desarrollo dada la tipología del proyecto. Es por tanto imprescindible el uso de una metodología *agile* o ágil. Esta se orienta a la construcción y entrega. Para conseguir esto, reduce el tiempo de desarrollo. Dado que el producto no se puede reducir conceptualmente, habrá que construir por fases para conseguir esto.

Dicho lo anterior, se puede apreciar que los procesos de especificación, diseño y la misma implementación son concurrentes. Utiliza, por ello, un ciclo de vida evolutivo, en el que el ciclo requisitos-desarrollo-evaluación termina en una versión del proyecto. Esta puede ser probada por el cliente, dando mayor oportunidad a feedback sobre el entendimiento de los requisitos.

La metodología ha sido elegida por la naturaleza del proyecto, como se ha dicho. Además, al sector al que va dirigido es muy concreto y, a la vez, diferente. Así, no se conocían a priori todos los requisitos, porque no se tenían unos clientes concretos a los que preguntar que abarcasen la representación del conjunto. Esto se ha podido ver con propiedad cuando se ha hablado de las entrevistas con *early adopters*, en el punto 2.3.

Así, los requisitos, se han ido conociendo a lo largo del desarrollo del trabajo. Se ha intentado que esto fuera mínimo siguiendo la metodología Running Lean, aunque los métodos preventivos nunca previenen al completo sobre posibles problemas. Así, como se decía, sin todos los requerimientos por parte del usuario, no es posible comenzar un ciclo de vida clásico, ya que en este es necesario conocer, idealmente, todos los requisitos del proyecto.

Por otra parte, se necesitaba tener, en poco tiempo, una versión testeable de la aplicación, para obtener un flujo de información más claro sobre lo que se desea realmente del aplicativo. Este es el segundo argumento a favor de la metodología ágil, ya que trabaja continuamente con versiones de la aplicación, al acabar el ciclo antes mencionado, y estas pueden ser fácilmente enseñables a futuros clientes para recibir opiniones.

Sim embargo, como se ha dicho, *ninguna metodología es perfecta*. Agile presenta algunos aspectos negativos, como son la difícil planificación por los requisitos cambiantes o la falta de cobertura en todas las áreas del proceso, entre otras. No obstante, se asumen por la naturaleza del proyecto, en favor de los motivos mencionados más arriba.

Aunque se ha hablado continuamente de la metodología ágil y su aplicación al proyecto, esto

no es tan directo. Es necesario descender a marcos de trabajo que especifiquen y complementen dichas formas de trabajo. Las elegidas para este proyecto serán SCRUM y Kanban, ambas complementarias, que se explicarán en los siguientes puntos.

3.1.1. SCRUM

Dentro de la metodología anteriormente comentada, es necesario descender a los detalles del día a día. De esta manera surge SCRUM como un marco de trabajo que permite el trabajo conjunto entre miembros de un mismo equipo. Su etimología proviene del rugby, y al igual que él, a la hora de entrenar, se ha de basarse en la experiencia, organizarse para resolver los problemas y reflexionar sobre los resultados obtenidos para mejorar.

SCRUM es un marco que provee a cualquier equipo de trabajo de unas pautas para poder aplicarse. Este describe un conjunto de funciones, herramientas y reuniones que hacen trabajar de manera coordinada a los miembros del equipo. Dicho lo anterior, se puede ver que SCRUM es un marco de trabajo, mientras que la metodología ágil es un ideal. Por ello no se puede trabajar directamente con *agile*, sino que es necesario recurrir a operativas que implementan, o al menos, desarrollan dicha mentalidad.

Este marco de trabajo se basa en el aprendizaje y adaptación al medio. Esta heurística hace al equipo humilde, ya que reconoce que no posee el pleno conocimiento al inicio, y que irá evolucionando a lo largo de la realización del proyecto. Hace, de manera soterrada, que el equipo de trabajo se vaya adaptando a los cambios tanto de las condiciones como del cliente. Esto se consigue presentando a los interesados continuas versiones del producto o servicio que se va desarrollando, además de un orden por prioridad de las peticiones del cliente.

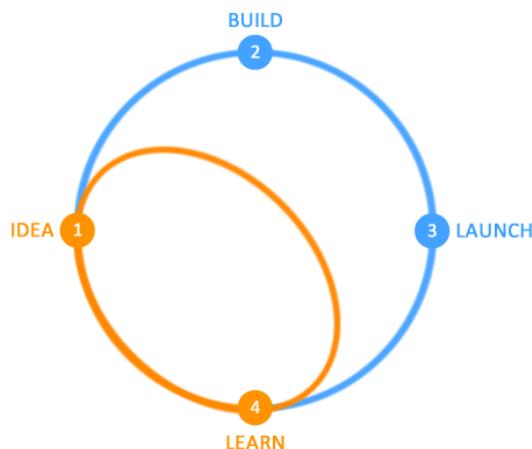


Figura 3.1: Aprendizaje en SCRUM

Como se puede apreciar en la imagen, el aprendizaje de SCRUM sigue la estructura: IDEA-CONSTRUCCIÓN-DESPLIEGUE-APRENDIZAJE. Es foco lo tiene el aprendizaje, y no tanto el desarrollo. Sin embargo, no es este un marco rígido. SCRUM se adapta sin problema a cualquier tipo de proyecto, estableciendo únicamente algunas pautas a seguir. Lo que de verdad importa, y a esto va dirigido este marco, es hacer que la comunicación sea clara, la transparencia sea real y la dedicación sea continua.

Artefactos de SCRUM

Los artefactos son aquellos desarrollos que realizamos para solucionar el problema. En SCRUM son tres: *Product Backlog*, *Sprint Backlog* e *Increment*.

El *Product Backlog* es la lista de tareas que debe realizar el equipo para desarrollar con éxito el entregable al cliente. Esta es una lista con fluctuaciones en la que se encuentran requisitos, mejoras

y correcciones a realizar sin indicar a priori una planificación de la misma. Cada tarea posee una prioridad que irá cambiando dependiendo de múltiples factores: cliente, entorno, sistema, etc.

El *Sprint Backlog* es una lista de elementos a realizar en un periodo concreto propuesto por el equipo. Esta se alimenta del *Product Backlog* y se encuentra cerrada durante el sprint. De hecho, antes de iniciar este periodo de tiempo, en la reunión de planificación del sprint, el equipo elige aquellas tareas del producto en las que trabajará.

El Increment es el producto desarrollado al final del sprint. Este está estrechamente relacionado con el concepto de finalizado. En este sólo se incluirán los elementos del *Sprint Backlog* que se hayan finalizado durante el sprint. Será el equipo de desarrollo el que indique qué entiende él mismo sobre la finalización de una tarea. Este prototipo construido, lejos del producto final, podrá ser testeable por el cliente y así proponer cambios.

Dentro de los artefactos, se suelen incluir también los gráficos de análisis de SCRUM, como *Burn-down Chart* o *Burn-up Chart*. Estos se podrán realizar tanto sobre el *Product Backlog* como por el *Sprint Backlog*. Estos son indicadores de cómo se ajusta la planificación a la realidad en base a las tareas finalizadas.

Eventos de SCRUM

Los eventos son el conjunto de protocolos, reuniones y acontecimientos secuenciales que los equipos de SCRUM realizan de manera periódica. Se encuentran seis eventos principales, que se detallarán a continuación.

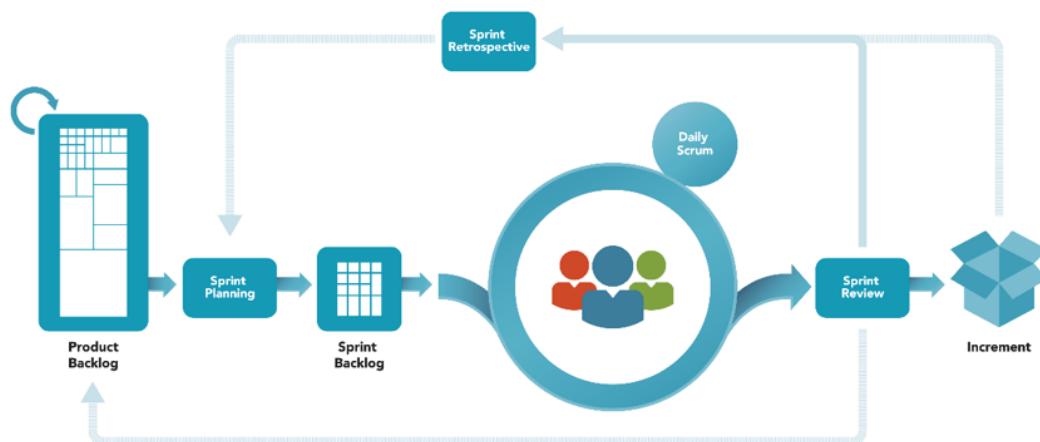


Figura 3.2: Eventos en SCRUM

El primero es la Organización del Backlog. Su función es dirigir el producto final hacia una visión comercial y tecnológica que busque el éxito del proyecto. El mantenimiento del *Product Backlog* se realiza con los comentarios de los usuarios y del equipo de desarrollo para priorizar los elementos de la misma. Gracias a esta tarea, cada elemento de la lista estará a punto para poder trabajar en cualquier momento.

El Sprint Planning es la reunión en la que el equipo de desarrollo planifica el trabajo que va a realizar durante el Sprint. En esta se establece un objetivo y se añaden al *Sprint Backlog* los elementos pertinentes. Al final de esta reunión, cada miembro debe de tener claro cuál va ser el incremento al final del sprint y la manera de lograrlo.

El Sprint es el periodo de tiempo real en el que equipo SCRUM trabaja para llevar a término el incremento. El intervalo de tiempo será elegido por el equipo, aunque autores recomiendan entre

una semana y un mes, dependiendo de la tipología del proyecto. Este será el punto de referencia para todo el marco de SCRUM.

El Daily Meeting es la reunión diaria en la que todos los miembros del equipo indican el estado de sus tareas e inquietudes que posean sobre la consecución de las mismas. La duración ha de ser breve, en torno a un cuarto de hora.

El Sprint Review es la reunión al final del sprint. En esta, el equipo comprueba el incremento realizado y lo muestra a las partes interesadas con el objetivo que obtener feedback sobre las tareas que ha ido realizando. La duración variará dependiendo del número de semanas del sprint. Se recomiendan entre una y cuatro horas.

Por último, el Sprint Retrospective. Es la reunión donde el equipo indica aquellos aspectos que han funcionado o no del sprint. Este se enfoca en todos los aspectos que rodean al equipo, desde el desarrollo hasta las personas. Su duración variará dependiendo de la duración del sprint, pero suele estar entre una y tres horas.

Funciones en SCRUM

El equipo de SCRUM debe poseer entre sus miembros de tres funciones principales: *Product Owner*, *Scrum Master* y Equipo de desarrollo. Además, como los equipos SCRUM suelen ser multidisciplinares, estos se podrán estructurar a su vez en otros equipos. Los tres roles SCRUM se concretarán a continuación.

El *Product Owner* es la persona que con mayor profundidad conoce el producto. Su misión es entender los requisitos del mercado, de los clientes y de la empresa. Estos se encargan de gestionar el *Product Backlog*, deciden si se lanzan los incrementos y lideran la mayoría de las reuniones. Este no ha de ser el Jefe de Proyecto, ya que poseen misiones diferentes.

El *SCRUM Master* es la persona con mayor conocimiento de SCRUM en el equipo. Su misión es proporcionar formación al equipo, buscar formas de afinar en su práctica y asegurar su buen funcionamiento. Este posee un nivel de detalle amplio sobre el trabajo cualquier miembro del equipo. Por ello, puede ayudar al equipo a mejorar su transparencia y flujo de entrega.

Por último, se encuentra el Equipo de desarrollo. Son los que realizan el trabajo de construcción en el día a día. Poseen el conocimiento de las prácticas orientada a un desarrollo sostenible. Estos han de tener una relación cercana, se deben encontrar en el mismo lugar y han de ser pocos. Autores expertos recomiendan entre cinco y siete.

Cada uno de los miembros del equipo, posee diferentes cualidades. Es tarea de todos conocerlas y optimizarlas, evitando cuellos de botellas. Es por ello importante sus funciones de autorganización y trabajo en equipo. Gracias a este doble aspecto, estiman la duración de sus tareas en base al histórico de tareas realizadas.



Figura 3.3: Equipo en SCRUM

3.1.2. Kanban

Kanban, al igual que SCRUM, es un marco de trabajo dentro de la metodología *agile*. Este marco pone su foco en un tablero donde aparecen las tareas que está realizando el equipo y el estado de las mismas. Esto da, de un simple vistazo, gran parte de la información del estado del proyecto. Además, debido a su simpleza, puede ser aplicable a cualquier tipo de industria.

Gracias a este doble aspecto de simpleza y diversidad, es enormemente usado dentro del marco de SCRUM. Simplemente añade algunas consideraciones al uso del tablero. En concreto, se especifican cuatro acciones y seis prácticas, todas orientadas a realizar las tareas pendientes de manera eficaz.

Principios en Kanban

David Anderson es considerado como el líder del pensamiento Lean y Kanban. Este formuló los cuatro principios para llevar a cabo de manera correcta hacia el proceso evolutivo e incremental.

El primer principio: empezar con lo que se hace ahora. Como se ha dicho, Kanban no requiere ningún tipo de configuración, pudiendo ser aplicado sobre cualquier flujo de trabajo. Así, para empezar a usarlo, no es necesario realizar cambios radicales.

El segundo principio: comprometerse a buscar e implementar cambios incrementales y evolutivos. Este marco de trabajo está diseñado para enfrentarse a cualquier resistencia al cambio. Todo está orientado a cambios continuos evolutivos e incrementales del proceso actual. Se evitan cambios drásticos en aras hacia pequeños virajes en el rumbo.

El tercer principio: respetar los procesos, las responsabilidades y los cargos actuales. Kanban reconoce que los anteriores roles y procedimientos pueden tener valor en el entorno del proyecto, por eso quiere conservarlos. Como se ve, se incita al cambio, aunque no se prescribe de manera alocada.

El cuarto principio: animar el liderazgo en todos los niveles. Recuerda a cada miembro del equipo que la consecución de los fines son exclusiva a las capas directivas. Es por ello que promueve pequeños actos día a día de mejora continua en todos los aspectos del equipo. Este se pondrá en valor cuando se consiga un rendimiento óptimo.

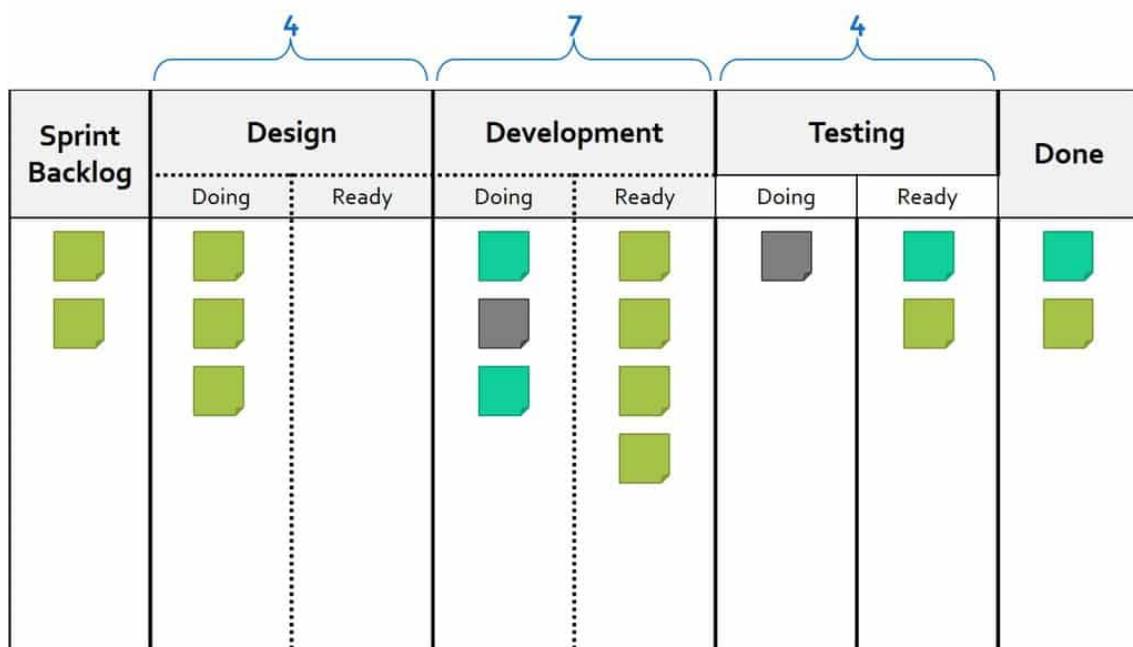


Figura 3.4: Ejemplo de tablero Kanban

Prácticas en Kanban

Una vez visto los puntos claves de la mentalidad hacia Kanban, se han de realizar las tareas pertinentes para implementar con éxito el marco de trabajo. En concreto, David Anderson indica seis prácticas para conseguir dicho fin.

La primera práctica: visualizar el flujo de trabajo. Se han de definir dos aspectos importantes. El primero es el estado por el que pasan las tareas, desde que se definen hasta que finalizan. El segundo es la información que poseerá cada tarea en el tablero.

La segunda práctica: Eliminar las interrupciones. Uno de los perjuicios más importantes que se intentan solventar con Kanban es el continuo cambio de enfoque en los miembros del equipo. Así, en cada columna del tablero, se indicará un número máximo de tareas que pueden estar a la vez. Este es el *Work In Progress* (WIP). Con esto conseguimos que solo se puede empezar con una tarea cuando alguna se termine.

La tercera práctica: gestionar el flujo. La idea de seguir la Kanban no es otra que crear un proceso continuo e ininterrumpido de las tareas. Cada uno de los estadios del flujo son las columnas por las que pasan las tareas. Teniendo esto claro, habrá que analizar la velocidad y continuidad del movimiento.

La cuarta práctica: hacer las políticas explícitas. Hay una máxima en el mundo Lean: no puedes mejorar algo que no se entiende. Así, todo proceso dentro del marco deberá estar definido, publicado y promovido.

La quinta práctica: circuitos de retroalimentación. Si se quiere que el cambio tenga éxito y sea duradero, se han de tener en cuenta las reuniones regulares para la transferencia de conocimiento. Ejemplos de estas son las reuniones diarias para ver el estado de las tareas o la reunión de revisión de entregas.

La sexta práctica: mejorar colaborando. Finalmente, si se quiere que este cambio exceda a la jurisdicción propia del proyecto que se está ejecutando y percute en la empresa, es necesario compartir ideas entre los miembros del equipo. En concreto, se ha de buscar una visión conjunta de las teorías sobre la manera de trabajar, el flujo de las tareas o procesos internos.

3.1.3. Aplicación

Una vez analizadas las metodología de trabajo, se indicarán cómo se han llevado a la práctica durante la ejecución y estudio del Trabajo Fin de Máster. Se ha utilizado un derivado entre SCRUM y Kanban, haciendo especial hincapié en el primero.

Parte importante de SCRUM es el equipo, gracias al cual se optimizan flexibilidad, creatividad y productividad. Normalmente, se trabajan con los tres roles anteriormente mencionados. Dado que solo hay un miembro que lleva a cabo todo el proceso, *Product Owner*, *Scrum Master* y *Development Team* se unen en una persona.

De la misma manera que en la definición del equipo, ha habido modificaciones en los eventos de SCRUM. La primera es el Sprint, que en nuestro caso son dos semanas. Estas son dos semanas en tiempo acumulado, no real. Esto se debe a que la dedicación al TFM ha sido fluctuante a lo largo del tiempo, además centradas en fines de semana.

Las reuniones, como se puede ya deducir, son de poco tiempo, aunque con pautas bien establecidas, como se puede ver a continuación.

- *Daily Scrum*: antes de ponerse a trabajar, se establece en el Tablero Kanban las tareas en estados *Ready* e *In Progress*.
- *Sprint Review*: se tiene con el profesor tutor del proyecto de manera formal y con un tiempo establecido (sobre una hora), y con los potenciales clientes, de manera informal y sin tiempo previamente establecido.

- *Sprint Retrospective*: se sacan las gráficas para ver resultados, se apuntan las cosas a mejorar y se toman decisiones concretas, como, por ejemplo, mejorar una tarea, cambiar una funcionalidad fruto del feedback recibido.
- *Sprint Planning Meeting*: se hacen inmediatamente después de la anterior reunión estableciendo en el Tablero los requisitos a cumplir, así como lo que haya que mejorar del anterior Sprint.

Como se ha dicho anteriormente, el punto neurálgico de SCRUM en este proyecto es el Tablero Kanban. Este es la referencia visual del estado de las tareas y, por ende, del proyecto. Es usado tanto por todos los interesados en el proyecto y, por ello, se basa en su buen uso el éxito del proyecto.

El Tablero se usa para saber en qué se está trabajando y qué falta por hacer. En este se establecen las tareas a hacer, así como un tiempo estimado. El tiempo estimado se mide en una escala tomada según la secuencia de Fibonacci, donde 1 representa una tarea muy sencilla de una media hora de duración y 20 (nuestro máximo), una tarea excesivamente larga, de unas diez horas. Se evitará siempre tener tareas de más de 8 puntos, aunque, en algunos casos, no es viable.

En el Tablero usado se poseen los siguientes flujos, en cuanto a las columnas se refiere.

- *Backlog*: al principio del Sprint, aquí aparecen todas las tareas a realizar, así como aquellas que se dejen para otros Sprints.
- *Ready*: aparecen las tareas que hay que completar en dicha semana.
- *In progress*: son las tareas que se están realizando. Como limitación, el tiempo estimado de todas las tareas que aparecen en esta columna, no podrá superar los 8 puntos, si hay más de una tarea.
- *Done*: son las tareas hechas, esto es, desarrolladas y probadas.

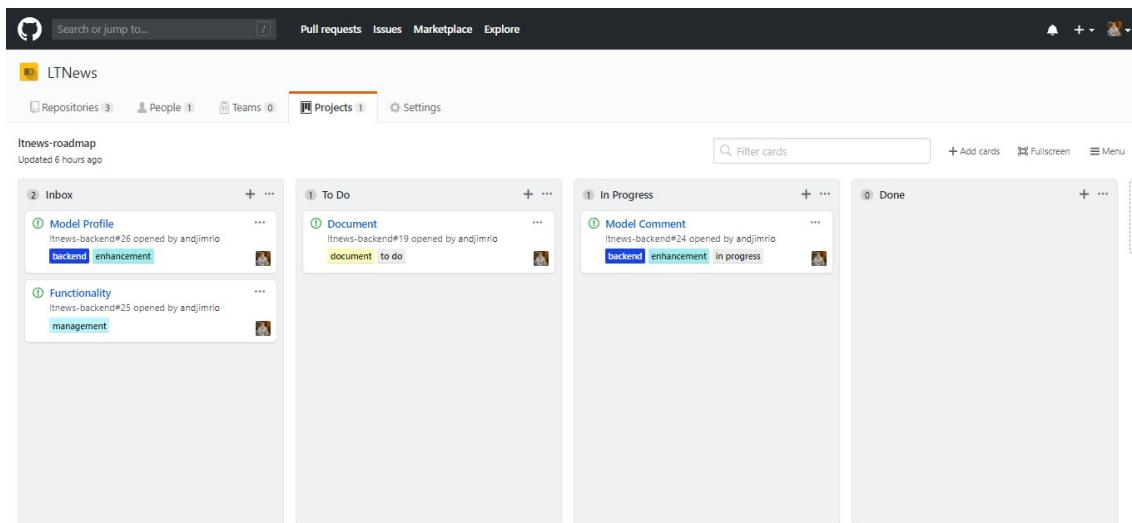


Figura 3.5: Tablero Kanban usado

Se ha usado para ello la plataforma GitHub, que ofrece a desarrolladores un tablero kanban con las características previamente mencionadas. Las tareas serán las *issues* de todos los repositorios anejos al mismo. En concreto, las tareas de los repositorios que vuelcan a este tablero son los códigos de Back-End, Front-End y documentación asociada.

3.2– Planificación temporal

En este apartado se encuentra tanto la planificación como la duración real, una vez finalizado el desarrollo. Como se ha dicho, la planificación utilizada ha sido utilizando la metodología ágil. A continuación, se recoge la planificación por sprints.

Nombre	Estimación
S0 - Fase previa	70:00:00
S1 - Investigación	30:00:00
S2 - Funcionalidad completa	100:00:00
S3 - Virtualización de los servicios	30:00:00
S4 - Capa Inteligencia Artificial	70:00:00

Cuadro 3.1: Planificación de Sprints

A su vez, los sprints se han de desglosar en fases concretas en cuanto al desarrollo del producto se refiere. Por ello, la planificación más coherente con respecto a cualquier producto software es la siguiente.

Fase	Subfase	Estimación
1. Planificación		10:00:00
2. Estudio		30:00:00
	2.1. Estudio de LEAN	10:00:00
	2.2. Realización de modelos	5:00:00
	2.3. Entrevistas	10:00:00
	2.4. Estudio de los datos	5:00:00
3. Análisis		50:00:00
	3.1. Documentación	40:00:00
	3.2. Elicitación de requisitos	5:00:00
	3.3. Modelo conceptual	5:00:00
4. Diseño		20:00:00
	4.1. Detalles técnicos	10:00:00
	4.2. Modelo de datos	5:00:00
	4.3. Mockups	5:00:00
5. Formación		30:00:00
	5.1. Django Rest Framework	5:00:00
	5.2. VueJS	10:00:00
	5.3. ElasticSearch	10:00:00
	5.4. Docker	5:00:00
6. Implementación		100:00:00
	6.1. Back-End	25:00:00
	6.2. Front-End	50:00:00
	6.3. AI-Layer	25:00:00
7. Pruebas		25:00:00
8. Despliegue		25:00:00
9. Presentación		10:00:00
TOTAL		300:00:00

Cuadro 3.2: Planificación temporal

3.3– Roles del proyecto

Tal y como dijimos antes, en nuestro proyecto se unifican los roles de SCRUM en una única persona. Sin embargo, esto no quita que esta persona tenga que asumir a lo largo del proyecto una serie de roles, en función del tipo de tareas a realizar. Sin embargo, para simplificar dichos roles, se han unificado en la responsabilidad de Jefe de Proyecto.

Por ello, las responsabilidades adquiridas y sus misiones serán las que aparecen a continuación.

- Jefe de proyecto: encargado de la gestión y planificación del proyecto, de reunirse con el cliente y de tomar decisiones sobre el equipo.
- Analista: encargado de recolectar los requisitos y necesidades del proyecto, así como la de diseñar el sistema de información.
- Desarrollador: se encarga de implementar los requisitos, es decir, de desarrollar el software.
- Tester: su misión es la asegurar la calidad de los productos obtenidos, en concreto, se encarga de elaborar y ejecutar las pruebas.

3.4– Planificación económica

Para la planificación económica se han considerado los gastos asociados a la ejecución del proyecto, sin ingresos asociados. Aunque el proyecto buscará en un futuro tener beneficios no entra este aspecto dentro del ámbito del Trabajo Fin de Máster.

Habrá que considerar, por tanto, dos tipos de costes, dependiendo si hacen referencia a las personas físicas, o a materiales e infraestructuras necesarias para llevar a cabo el proyecto.

3.4.1. Costes directos

Dentro de este tipo de costes se consideran exclusivamente los costes de personal. Tal y como se ha dicho en el apartado 3.3, para este es necesario una serie de roles, cada uno de los cuales posee un salario distinto.

Así, dentro de estos costes se van a tener en cuenta los sueldos de los distintos tipos de trabajadores durante el desarrollo del proyecto. Para esto, nos se ha de basarse en el Boletín Oficial del Estado publicado en abril de 2009, referente a salarios mínimos, (de Trabajo e Inmigración, 2009). En nuestro caso, utilizaremos los datos del sector de las TIC.

Para todos los casos, se considerará que un año posee 250 días hábiles, que supone 2.000 horas. Ahora, analicemos cada rol en concreto.

- Jefe de Proyecto: su sueldo mínimo es de 36.600\$ que a la hora hace 18,3\$
- Analista: su sueldo mínimo es de 21.555,66\$ que a la hora hace 10,78\$
- Desarrollador: su sueldo mínimo es de 15.442,56\$ que a la hora hace 7,72\$
- Tester: su sueldo mínimo es de 11.773,02\$ que a la hora hace 5,89\$

Además, a estos costes, se han de aplicar los impuestos que paga la empresa para cualquier trabajador en España. Todos son porcentajes que se aplican a la base de cotización. Estos son los siguientes:

- Contingencias comunes
 - Estos dan cobertura a bajas temporales o prestaciones de jubilación, entre otras.
 - Supone el 23,6 %

- Fondo de Garantía Social
 - Es un organismo encargado de dar garantía a los trabajadores de su salario o indemnizaciones.
 - Supone el 0,2 %
- Formación Profesional
 - Asegura la formación al trabajador.
 - Supone el 0,7 %
- Seguridad Social
 - Dependerá del tipo de contrato y de su duración. Consideramos duración determinada y a tiempo parcial.
 - Supone 7,7 %

Teniendo en cuenta todos los porcentajes, queda para aplicar a la base de cotización una suma de los anteriores porcentajes: 32,2 %.

3.4.2. Costes indirectos

En este tipo de costes, se incluyen los materiales utilizados para realizar el proyecto, que en este caso se dividen entre hardware y software. Sobre cada material, independientemente de su tipología, se aplicará una base de amortización. Esto es así ya que no se adquiere un material concreto exclusivamente para un proyecto, sino que su coste deberá aparecer reflejado en los distintos proyectos en los que se utilice hasta que termine su vida útil.

Hardware

Dentro de este apartado se consideran dos dispositivos: un ordenador portátil y un móvil. Ambos han servido para desarrollar el proyecto y para sus respectivas pruebas. El ordenador utilizado ha sido un *ASUS A55L*, valorado en 789,99\$. Se considera que la vida útil de cualquier portátil es de cuatro años. Así, cada mes utilizándolo, supondrá uno gasto de 16,46\$. Este ha sido utilizado durante todo el proyecto. El smartphone que se utilizará para comprobar la interfaz gráfica será un *Xiaomi Mi A2 Lite*, valorado en 189,99\$. Considerando en este caso una vida útil de dos años, supondrá un gasto del mes utilizado por 7,92\$. Este ha sido utilizado durante la mitad del proyecto: los meses finales para visualizar la interfaz gráfica móvil.

Software

Dentro de este, se engloban las licencias de los programas utilizados, así como los costes del servidor. Las licencias, como el hardware, requieren también de una amortización. Así, las utilizadas para llevar a cabo el proyecto y su coste han sido los siguientes:

- Windows 10 Home
 - 135\$ en pago único, amortizable en 8 años.
 - 1,41\$/mes.
- Office 365 Personal
 - Posee pago mensual.
 - 7,00\$/mes.

- PyCharm
 - 199\$ por año utilizado.
 - 16,58\$/mes.
- WebStorm
 - Al igual que PyCharm, 199\$ por año utilizado.
 - 16,58\$/mes.

Para el servidor, hemos adoptado el plan más básico de DigitalOcean. Esto hace unos 5\$/mes, que en euros son 4,45\$/mes.

3.4.3. Presupuesto

Con el conocimiento de los gastos, se procederá a dar un presupuesto del proyecto. Antes, se consideran que es necesario un colchón para pagos no previstos. Este será de un 15 % del subtotal. Además de esto, hemos de tener en cuenta el I.V.A. Por ello, la cantidad resultante será la siguiente.

	Tiempo	Coste unitario	Coste total
Personal	horas		3.755,47
Jefe de proyecto	25	24,19	604,82
Analista	100	14,25	1.425,12
Desarrollador	150	10,21	1.530,88
Tester	25	7,79	194,66
Hardware	meses		309,58
ASUS A555L	14	16,46	230,41
Xiaomi Mi A2 Lite	10	7,92	79,16
Software	meses		471,61
Windows 10 Home	14	1,41	19,69
Office 365 Professional	14	7,00	98,00
PyCharm	10	16,58	165,83
WebStorm	10	16,58	165,83
Digital Ocean	5	4,45	22,25
		Subtotal	4.536,65
		Contingencias	680,50
		Total	5.217,15

Cuadro 3.3: Presupuesto económico

CAPÍTULO 4

Análisis

- 4.1– Actores del sistema**
- 4.2– Modelo conceptual**
- 4.3– Catálogo de requisitos**
- 4.4– Diagrama de secuencia**

CAPÍTULO 5

Diseño

5.1– Patrones

5.2– Modelo de datos

5.3– Prototipos

CAPÍTULO 6

Implementación

6.1– Tecnologías

A continuación se definirán las tecnologías usadas para llevar a cabo el Producto Mínimo Viable.

Web Scrapping

El aspecto fundamental del producto es la recolección de noticias de periódicos. Esta, como se ha ido diciendo a lo largo del documento, se realiza a través de RSS. Sin esta característica, no tendría utilidad la aplicación. Sin embargo, esto solo se queda corto. Normalmente, cada medio solo publica un titular y una pequeña descripción vía RSS. Esto les sirve de gancho al usuario para que pase del lector de sindicación de contenido a la web de su medio.

Como se ha dicho, otro pilar fundamental para el buen funcionamiento del producto es el análisis de noticias. Por tanto, si no se posee más que un pequeño texto acompañado de cada noticia, poco podrá aportar este. Es por ello imprescindible conseguir el texto original del artículo para poder procesarlo correctamente. Dado que los medios no poseen esta característica, hemos de realizarla a través del *Web Scrapping*.

El *Web Scrapping* es una técnica que permite extraer información de cualquier sitio en Internet de forma automática. La información en la web se encuentra en diferentes formatos dependiendo de su misión para con el receptor. Resumidamente se encuentran dos tipos de datos: estructurados, como pueden ser ficheros XML o APIs, y no estructurados, como los ficheros HTML. Estos últimos abundan ya que, en un principio, van destinados al usuario final directamente, con la idea de ser leídos o vistos sin más.

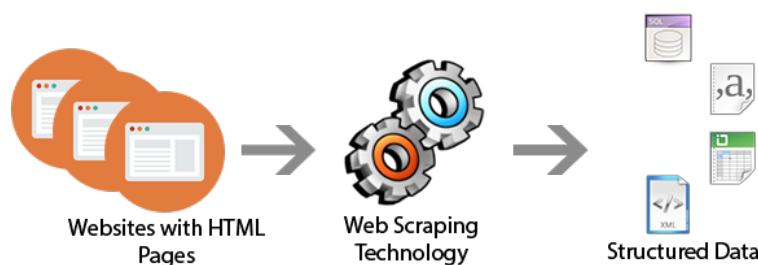


Figura 6.1: Muestra del *Web Scrapping*

Por lo dicho, esta técnica posee diferentes implementaciones, dependiendo del nivel de au-

tomatización que deseé. Un primer nivel sería el de extraer el fichero HTML y trabajar con él, utilizando, por ejemplo, expresiones regulares; otro sería, utilizar parsers o minería de datos, para extraer el contenido principal; y otro, el de analizar una página en concreto, para saber cómo se estructura el contenido realmente interesante.

Como última consideración, se ha informar que antes de trabajar con la información extraída del *Web Scrapping*, hay que saber que esta tiene implicaciones legales. Estas, en la mayoría de los casos, se encuentran en un vacío legal. En la práctica, si los datos se usan para un uso personal, no hay problema añadido. La cosa cambia si es para un uso comercial. De hecho, actualmente, hay varios juicios por un uso fraudulento de esta técnica en Australia o Estados Unidos, como se puede ver en el artículo publicado por Ainhoa Lafuente (2019).

Recuperación de información

Una vez extraída la información de los medios, es imprescindible tratarla correctamente. No basta el guardarla en una base de datos sin más, sino que es necesario usar medios de albergar la información especializado para ello. Si se usa, por ello, técnicas de Búsqueda y Recuperación de Información (ISR, por sus siglas en inglés, en adelante) podemos tratar cada noticia como un documento y así extraer metadatos de cada uno.

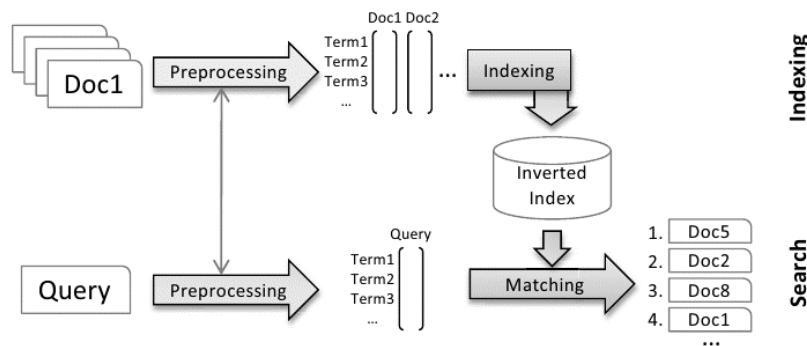


Figura 6.2: Muestra de Recuperación de Información

Esta técnica, por tanto, permite recuperar documentos en base a búsqueda de sus metadatos guardados en un índice. Esta información no suele estar estructurada y su misión es la de buscar información relevante en base a diferentes filtros. Hay diferentes técnicas, dependiendo del grado de exactitud que se requiera teniendo en cuenta la consulta. Algunas de las más importantes son las siguientes.

- **Sistemas de recuperación de lógica difusa:** las búsquedas sobre la información son de tipo fuzzy, es decir, frases normales, que el sistema transforma en frases lógicas. Así, la información que se consigue, se basa en proposiciones lógicas, teniendo en cuenta si los términos buscados aparecen o no.
- **Técnicas de ponderación de términos:** dado que hay términos de búsqueda que tienen más relevancia que otros, es lógico que se pondere dichos términos. Así, los documentos más pertinentes, serán los que contengan todos los términos y que se repita más el que más valor tenga.
- **Técnicas de clustering:** es un modelo probabilístico sobre la frecuencia de los términos de búsqueda. Se clasifican los documentos por valores que, finalmente, representan los pesos.
- **Técnicas de retroalimentación por relevancia:** una vez hecha una búsqueda utilizando una de las técnicas anteriores, se vuelve a buscar pero con términos relacionados con los primeros documentos resultantes.

- **Técnicas de stemming:** usa la técnica de Stemming para la búsqueda, es decir, trunca los términos de búsqueda quitando posibles prefijos o sufijos. En fin, evita ambigüedades léxicas.

Como se puede observar, estas técnicas se pueden combinar entre sí dentro de una implementación concreta, aunque siempre el funcionamiento es el presentado en la imagen 6.2: se procesan los documentos, se indexan y se busca sobre el índice.

Como se ha dicho al inicio del apartado, interesa que este índice creado trate correctamente la información guardada. Esta es, por tanto, la mayor ventaja de usar un ISR: se pueden elegir técnicas y tratamientos concretos por cada campo, así como por cada índice de documentos. En el ejemplo de guardar noticias se puede decir que extraiga las palabras claves, que de más importancia al cuerpo de la noticia, que obvie mayúsculas y tildes en el índice y para el análisis, que no tenga en cuenta artículos y preposiciones.

Esto dará una versatilidad a esta *base de datos no relacional* mucho mayor, ya que estará especializada en el dominio del problema concreto. Mientras tanto, en una base de datos clásica, la información se guarda siempre de manera homogénea y genérica para cualquier dato que se quiera albergar.

Sistemas de recomendación

Una vez obtenidas las noticias y analizadas y albergadas en el sistema habrá que dar uso a los metadatos albergados. Esta, como se ha dicho a la hora de explicar el producto, será el valor diferencial a cualquier lector de noticias: la capacidad de extraer el perfil del usuario y recomendar noticias en base a sus intereses. Esto, por tanto, se consigue con un sistema de recomendación especializado.

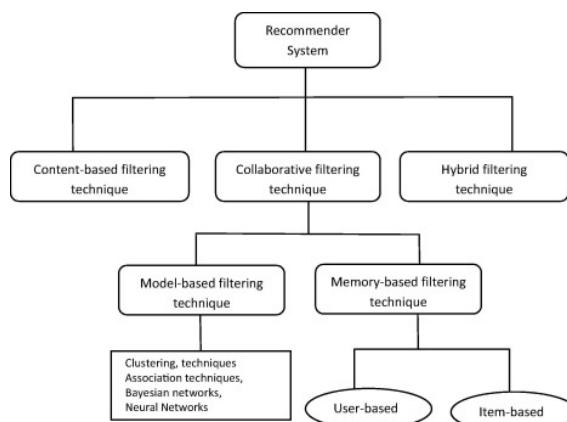


Figura 6.3: Muestra de sistemas de recomendación

Es la técnica de filtrado de información dependiendo del interés del usuario en los mismos. En definitiva, se encarga de comparar el perfil del usuario con los ítems en concreto para, finalmente, recomendarle uno que no haya visto previamente. Para realizar esta técnica, tenemos la capacidad de elección entre diversas implementaciones. Como se puede apreciar, es importante analizar tanto a usuarios como ítems a recomendar.

Analizar el perfil del usuario significa, en definitiva, extraer características de su interacción con los ítems. El método de extracción de sus características puede ser implícito o explícito. Dentro del primero encontramos aquellos sistemas a los que el sistema solicita al usuario que valore un determinado producto, o, directamente, seleccione sus intereses. Dentro del segundo, a aquellos en los que, en base a visitas a un ítem, al tipo de interacción con el mismo, al tiempo que está viéndolo, u otras técnicas, saca información del usuario.

Ambos tienen aspectos positivos y negativos. En el primero, el usuario dice al sistema expresamente qué le gusta o disgusta, mientras el segundo, la aplicación ha de suponerlo. Esto último puede llevar a numerosos fallos de interpretación frente a la certeza que proporciona el primero. Por otro lado, el método implícito extrae dicha información sin necesidad de la intervención del usuario, cosa que el primero no puede. Esto puede hacer que en el método explícito el usuario nunca complete sus preferencias, por el poco uso de la aplicación.

Como se podrá intuir, la mejor solución depende de la naturaleza de la aplicación y el enfoque que ponga esta en el usuario objetivo. No obstante, una buena solución podrá ser un aproximación a ambas, de forma híbrida. Con esto, se poseerá las ventajas de ambas de manera sencilla.

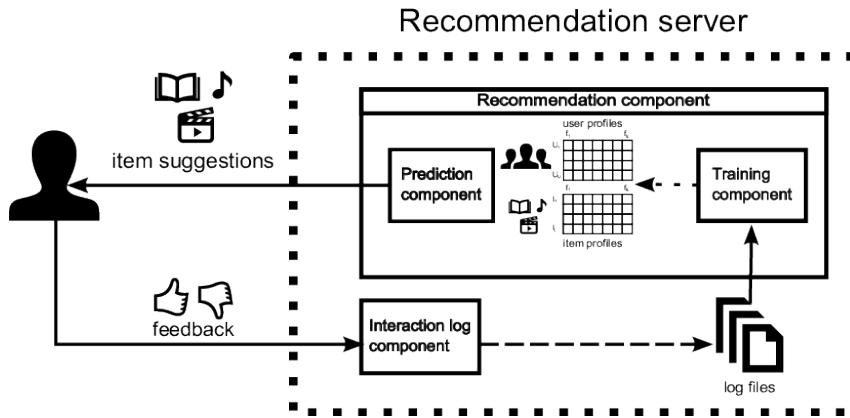


Figura 6.4: Muestra de extracción del perfil del usuario

Una vez obtenido el perfil del usuario, hemos de poder recomendarle ítems. En este apartado encontramos dos enfoques mayoritarios (y un tercera, de la unión de ambos) como se puede apreciar en la imagen 6.3.

La primera implementación son las recomendaciones basadas en contenido. Esta técnica consiste en analizar los ítems que tenemos en nuestros sistemas en base a una o varias características. Así, sabiendo los ítems con los que el usuario ha interactuado, podemos recomendarle otros que posean características similares. El principal problema de este sistema de recomendación es la falta de serendipia, es decir, al usuario siempre se le muestra el mismo tipo de productos. Esto se debe a es la información que se tiene y sobre eso se recomienda. Esto conllevará un hartazgo de las recomendaciones, resultando finalmente inútiles.

La segunda implementación es el filtrado colaborativo. Esta consiste en mostrar productos que han resultado atractivos para otros usuarios con gustos similares y que todavía no ha visto. Para esto, se han de calcular similitud entre los usuarios, y no sobre los ítems, como vimos en las recomendaciones en base al contenido. El principal problema del este es que funcionará bien si se poseen diferentes y numerosos tipos de usuario y si se tiene del usuario en cuestión de bastante información para tener un perfil complejo que comparar con otros usuarios. Esto es un problema, para aplicaciones que se empiezan desde cero sin usuarios.

Aquí, igual que anteriormente con el perfil del usuario, es posible combinar ambas técnicas haciendo un filtrado híbrido. Esto hará que el se contrarresten los defectos de ambas, centrándose en sus virtudes. Con respecto a la falta de serendipia, el filtro colaborativo lo contrarresta haciendo similitud entre usuarios, por que el usuario puede verse sorprendido al ver que se le recomienda algo que puede ser nuevo para él. Con respecto a un gran número de usuarios necesarios inicialmente, las recomendaciones basadas en contenido lo contrarrestan por la similitud de ítems, por lo que, por ejemplo, lo puede haber un solo usuario y funcionar bien el filtro colaborativo.

6.2– Herramientas

Para el desarrollo del producto y las tecnologías previamente mencionadas, se han usado las siguientes herramientas.

6.2.1. Backend

Para el desarrollo del servidor se han usado Python con el framework web Django, ambos en sus últimas versiones, además, de librerías que permitan implementar los requisitos de la aplicación.

Python

Python es un lenguaje interpretado y multiparadigma. Esto último se debe a que soporta tanto programación orientada a objetos como programación imperativa, además de programación funcional. Actualmente está administrado por la fundación *Python Software Foundation* y posee una licencia de código abierto denominada *Python Software Foundation License*. Esta nos permite hacer modificaciones de su código fuente y la creación todo trabajo derivado de este, sin necesidad de que este sea también código abierto.

Se ha decidido utilizar este lenguaje de programación en el lado del servidor debido a la cantidad y facilidad de uso de herramientas que implementan las tecnologías previamente mencionadas. Es la referencia en temas de *Web Scrapping* e Inteligencia Artificial, como se puede apreciar en el artículo de Ian Buckley (2018).

Django

Django es un framework de desarrollo web, también open source y que posee el patrón de diseño Modelo-Vista-Controlador. Ya no se habla de *Model-View-Controller*, sino de *Model-Template-View*. Sin embargo, hacen referencia a los mismos conceptos. El Modelo es la capa de accesos a los datos; la Plantilla, es la capa de presentación; la Vista posee la lógica de negocio de la aplicación y, por tanto, hace de puente entre ambos. Vemos, por tanto, que es en esencia lo mismo.

Su misión es la facilitar la creación de aplicaciones web complejas. Pone el acento en la conectividad, reutilización y la extensión de todos y cada uno de sus componentes. Tiene, además, como principal característica desde el punto de vista del desarrollo, el que todo esté hecho en Python.

Se ha decidido utilizar este framework por la facilidad de uso, y la potencia que posee sin configuración previa, es decir, posee integrado desde el primer momento aspectos avanzados de seguridad (evita inyecciones SQL, validaciones de los formularios, autentificación en cada página, entre otros), gestión en cada momento las sesiones, así como la transparencia en multitud de aspectos: la base de datos utilizada, la localización de los archivos estáticos, el despliegue...

Librerías

Los módulos son gestionados por Python a través del CLI *pip*. Estos se añaden automáticamente al software a través de este. Las librerías usadas son las siguientes:

- **django-rest-framework**: es la herramienta que permite ofrecer una API usando Django. Esta convierte las vistas de Django en un endpoint para publicitar contenido a través de los verbos HTTP.
- **django-rest-auth**: ofrece sobre la API implementada una capa de autenticación usando el modelo de usuarios de Django usando diferentes técnicas, como extracción completa de la información de usuarios o registro por redes sociales.

- **django-cron**: crea operaciones que se ejecutan sobre el CLI propio de Django para poder ser ejecutadas de manera recurrente. Estos métodos interaccionarán con todos los módulos de Django.
- **gunicorn**: es un servidor HTTP que soporta WSGI. Este ofrece a través de peticiones HTTP las aplicaciones de Django.
- **beautifulsoup**: es una herramienta que permite tratar con archivos HTML para extraer información.
- **feedparser**: permite descargar y parsear todo tipo de Feed RSS. Además, es compatible con muchísimos formatos: desde RSS 0.9 hasta RSS 2.0; desde Atom 0.3, hasta Atom 1.0.
- **newspaper3k**: es una herramienta de obtención de noticias. Analiza una web y calcula la localización del contenido principal. Además, soporta la mayoría de lenguajes principales y posee herramientas de Procesamiento de Lenguaje Natural en algunos idiomas.
- **python-dateutil**: facilita el tratamiento con fechas en Python y su conversión a diferentes formatos.

Para elegir cada una, se ha realizado un pequeño estudio entre sus posibles alternativas. Dado que son módulos enormemente usados, son la referencia en su tema, y rara vez hay alguna alternativa de la misma embergadura que haga dudar al desarrollador.

6.2.2. Frontend

Para el desarrollo de la interfaz de usuario se ha escogido JavaScript y el framework de desarrollo VueJS.

JavaScript

Desde el inicio del desarrollo web hasta hoy en día solo han existido tres lenguajes en el mundo del frontend: HTML, CSS y JS. El primero para la estructuración de la información, el segundo para el diseño y el tercero para la interacción del usuario. Es sobre este tercero donde mayores avances ha habido.

Ni que decir tiene que JavaScript se ha convertido en el lenguaje de referencia en el desarrollo web si se quiere una interfaz desacoplada, asíncrona y elegante, como revela la encuesta mundial de The State of JavaScript (2019). Son estos los motivos por los que se ha escogido.

De hecho, si se quiere realizar una *Single Page Application* (SPA en adelante), no hay otra opción plausible. Esta permite tener un simple fichero HTML e ir interactuando con él a través de JavaScript exclusivamente. Esto da al usuario una visión limpia y rápida de la aplicación en cuestión.

Se ha utilizado este sobre otros lenguajes de mayor alto nivel basados en JavaScript como TypeScript o Flow por su sencillez y recomendaciones de las herramientas utilizadas.

VueJS

Una vez decidido que se pretende realizar un frontend desacoplado del lado del servidor, solo queda la elección del framework a utilizar. Realmente, esta decisión en nuestros días se reduce a tres posibles alternativas: React, Angular y VueJS. Esta afirmación se basa en los datos y dirección de la comunidad, como se puede observar en el ya citado The State of JavaScript (2019).

Si se observa, aparecen, junto a estos tres, dos framework en cuanto a utilización actual: VanillaJS y AngularJS. Sin embargo, estos dos van quedando en desuso en cuanto a framework frontend de una aplicación web como tal. VanillaJS utiliza JavaScript en su estado puro, por lo que, realmente, no se le puede considerar un framework; consiste en una librería se usa cuando se

quiere tener en cuenta velocidad y espacio. AngularJS, por su parte, es el predecesor de Angular (también conocido como Angular 2). Cambia por completo su funcionamiento, siendo esta primera versión una librería y la segunda y posteriores, un framework. Google, autor del mismo, ha dejado de actualizar AngularJS, por lo que no ha de tenerse en cuenta para desarrollos actuales.

Visto lo anterior, comparemos los tres frameworks mencionados para ilustrar el motivo de elegir VueJS sobre los demás, perteneciente al estudio realizado por Miguh Ruiz (2018).

- **React:** es la librería de Facebook centrada en la creación de vistas. Su gran virtud son los patrones de eventos, ya que estos permiten actualizar en tiempo real las vistas con los datos. Realmente no es un framework, por lo que es posible utilizarlo anexo a Angular o VueJS. Si embargo, debido a su potencial, puede ser usado exclusivamente para realizar la capa de presentación.
- **Angular:** es el framework de Google diseñado sobre su primera versión y hecho para ser totalmente independiente y funcional. Usando HTML, CSS y TypeScript consigue crear vistas, componentes independientes y rutas, todas relacionados entre sí. Trabaja sobre el concepto *Single Page Application* (SPA) utilizado para ello Webpack, esto es, una vez desarrollada la aplicación, se compila en un fichero HTML (*index.html*), un fichero CSS y varios ficheros JS.
- **VueJS:** es el framework creado por la comunidad que une las ventajas de los desarrollos anteriores, a destacar: el ciclo de vida de los componentes propio de React y las directivas de Angular. La curva de aprendizaje es más reducida que Angular y el proyecto poseerá menos ficheros que este, ya que aglutina HTML, CSS y JS en un fichero *.vue*. Además, es totalmente versatil: se puede utilizar como librería para un desarrollo pequeño hasta como framework en un aplicación SPA o SSR completa.

Son estos motivos por los que se ha optado por VueJS, aunque sinceramente, debido a su similitud con Angular, se podría haber optado por cualquiera. Además, el ecosistema que ofrece Vue, como se verá en el siguiente punto, es único.

Librerías

Los módulos usados están relacionados con el ecosistema que ofrece Vue así como por la comunidad de JavaScript. Estos se gestionan a través de *npm*.

- **vue-router:** gestiona la navegabilidad entre las vistas, compartiendo datos entre sí y haciendo las accesibles desde la URL del navegador.
- **vuex:** administra los estados de determinados objetos a lo largo de toda la aplicación. Ofrece de manera general un almacén de estados que pueden ser mutados y consultados desde cualquier módulo.
- **vuetify:** implementa un ecosistema basado en Material Design, el diseño propuesto por Google, usando componentes Vue.
- **axios:** gestiona las peticiones HTTP en JavaScript ofreciendo una interfaz completa y simple.
- **moment:** convierte cualquier campo de tipo fecha al formato deseado por el usuario.

Todas estas librerías son usadas por la gran mayoría de la comunidad y no poseen alternativas claras y robustas: los desarrolladores prefieren mejorárlas a enfrentarse a ser competencia directa.

6.2.3. Servidor

En este apartado se mencionarán las tecnologías usadas en la parte del servidor, de manera conjunta a las de Cliente y Servidor previamente mencionadas.

PostgreSQL

Es imprescindible usar en este tipo de proyectos una base de datos consistente. Además, dada la naturaleza de la aplicación en cuestión, esta ha de ser Relacional. De todas las posibilidades que ofrece este gran mundo, Django recomienda usar esta debido a su naturaleza y fuerte interconexión.

Postgres es un sistema gestor de bases de datos de tipo relacional, como se ha dicho. Posee una gran escalabilidad, al usar una infraestructura propia de paralelización: multiprocesos en vez de multihilos. Esto implica que desacopla los posibles errores propios de la gestión de hilos, como son las condiciones de carrera.

Elasticsearch

Para implementar la Búsqueda y Recuperación de Información, descrito en el punto 6.1, referente a las tecnologías, es necesario usar un Índice.

Elasticsearch es un servidor de búsqueda, distribuido y accesible a través de una interfaz de tipo RESTful. Está desarrollado en Java, basándose en Lucene y es de código abierto. Propone una sintaxis basada en JSON para las búsquedas.

Se ha decidido usar sobre Índice como Algolia o Solr por dos motivos principales. El primero se debe a su coste, nulo en proyectos con implantación propia de dicha herramienta. El segundo es que se está poniendo a la cabeza en el desarrollo de índices generales en la comunidad.

Nginx

Una vez albergado y desarrollada la aplicación, es necesario hacerla accesible a través de Internet de manera segura y consistente. Para ello nos hace falta un servidor web o proxy inverso. Las dos grandes alternativas en este punto son Nginx y Apache.

Nginx es un tipo de servidor web ligero y con alto rendimiento. Es open source y permite, además, integración de servidor SMTP, así como soporte de HTTP, HTTP2, IPv6 y demás protocolos.

Se ha decidido usar este por el gran rendimiento que ofrece sobre el Servidor HTTP Apache. Por otra parte, ofrece una gran comunicación y facilidad en la configuración en aplicaciones Full Stack.

Docker

Para hacer toda esta arquitectura independiente a cualquier sistema y con la configuración similar a entornos de desarrollo y producción es necesaria la virtualización de los servicios. Esto ha marcado un nuevo hito en el mundo del software: la abstracción entre sistemas operativos y la concentración de la configuración en un único fichero. Esto es Docker.

Docker automatiza el despliegue de cualquier software en contenedores. Estos poseerán toda la configuración que necesita el desarrollo para funcionar. Es la evolución de la máquina virtual: se convierten en instancias Linux ligeras y aisladas.

Actualmente, no hay una alternativa clara a Docker. De hecho, tecnologías de orquestación de contenedores, tales como Swarm o Kubernetes, se basan en esta.

6.2.4. Gestión

En cuanto a la gestión del proyecto se refiere, podemos hablar de dos grandes ámbitos: desarrollo y seguimiento del proyecto.

Desarrollo

Para el desarrollo del proyecto han sido necesarias las herramientas que se mencionarán adelante. La preferencia general de dicha elección se basa en dos premisas. La primera es la orientación de la comunidad del desarrollo y la segunda es la disponibilidad real de dichas herramientas. Además de utilizar tecnologías gratuitas, la universidad nos ofrece plataformas y programas en periodo de prueba.

En primer lugar son necesarios los programas para el desarrollo de las diferentes partes de la aplicación. Dado que cada módulo se desarrolla en un lenguaje concreto, se ha apostado por usar IDEs personalizados. Así, y debido a su potencial, se han usado todos aquellos necesarios creados por JetBrains. En el caso de no existir, se ha decidido acudir a Visual Studio Code.

Se ha usado también una herramienta para la realización de diagramas llamada Astah. Esto permite analizar lo que se necesita antes de implementarlo. Gracias a esta se harán todo tipo de diagramas: desde estados hasta clases. Se ha usado esta gracias a su gratuidad a estudiantes.

Para mantener una trazabilidad y respaldo del código, se han usado tecnologías de control de versiones. En concreto, la elegida ha sido Git, debido a su posición en la comunidad del desarrollo y a su funcionalidad y robustez. La plataforma remota donde se alojará dicho código sobre Git es GitHub. Ha sido elegida por ser la principal usada por la comunidad.

En cuanto a la comunicación y despliegue por el servidor, se ha elegido como VPS: Digital Ocean. Posee un enorme potencial y un periodo de prueba más que suficiente para la distribución de la aplicación. Además, para la comunicación con este, se ha usado WinSCP y PuTTY, como herramientas SSH principalmente usadas en Windows.

Seguimiento

Una vez llegado a las decisiones convenientes para desarrollar el producto, es necesario seguir una serie de pautas, marcarse unos hitos y tener la posibilidad de comprobar el avance del proyecto. Así, han de usarse una serie de herramientas para medir la evolución y seguir con acierto.

Se han usado grosso modo dos: Toggl para la gestión del tiempo y Waffle para la gestión de las tareas. La primera permite comprobar las horas que lleva cada tarea. Esta permite etiquetar cada una, por lo que se podrá comprobar el tiempo dedicado a cada tipo de tarea (gestión, documentación, desarrollo, despliegue o investigación), así como el total dedicado al proyecto. Con la segunda se consigue agrupar las tareas en hitos y asociarlos a funcionalidades dentro del código. Además, ambas están relacionadas, por lo que simplifica dicha gestión, siendo esta la principal motivación para su uso.

Waffle es la implementación concreta del tablero canvas mencionado en el punto 2.1. Si se aprecia la imagen 3.5, vemos que es GitHub la plataforma mencionada para el tablero. Lo ocurrido se puede apreciar en el artículo de Waffle (2019). Esta herramienta deja de mantenerse por falta de recursos y recomienda su migración a GitHub.

6.3– Estructura del proyecto

Como se ha ido diciendo a lo largo del documento, el proyecto utiliza Git. Así, se ha estructurado el trabajo en ramas según la metodología Git Flow. Este estructura el cualquier proyecto software con cinco ramas principales:

- **master:** el código en esta rama es la versión estable y funcional que se encuentra funcionando en producción.

- **hotfix**: cualquier fallo detectado en producción se arregla aquí y se lleva a *master* una vez solucionado: no antes.
- **release**: sobre los cambios agrupados del desarrollo, se prueba integralmente antes de ser subidos a *master*.
- **development**: agrupa los cambios en las diferentes características que se van trabajando.
- **feature**: cada una de las funcionalidades a añadir en el sistema.

Esto se puede ver gráficamente en la siguiente imagen.

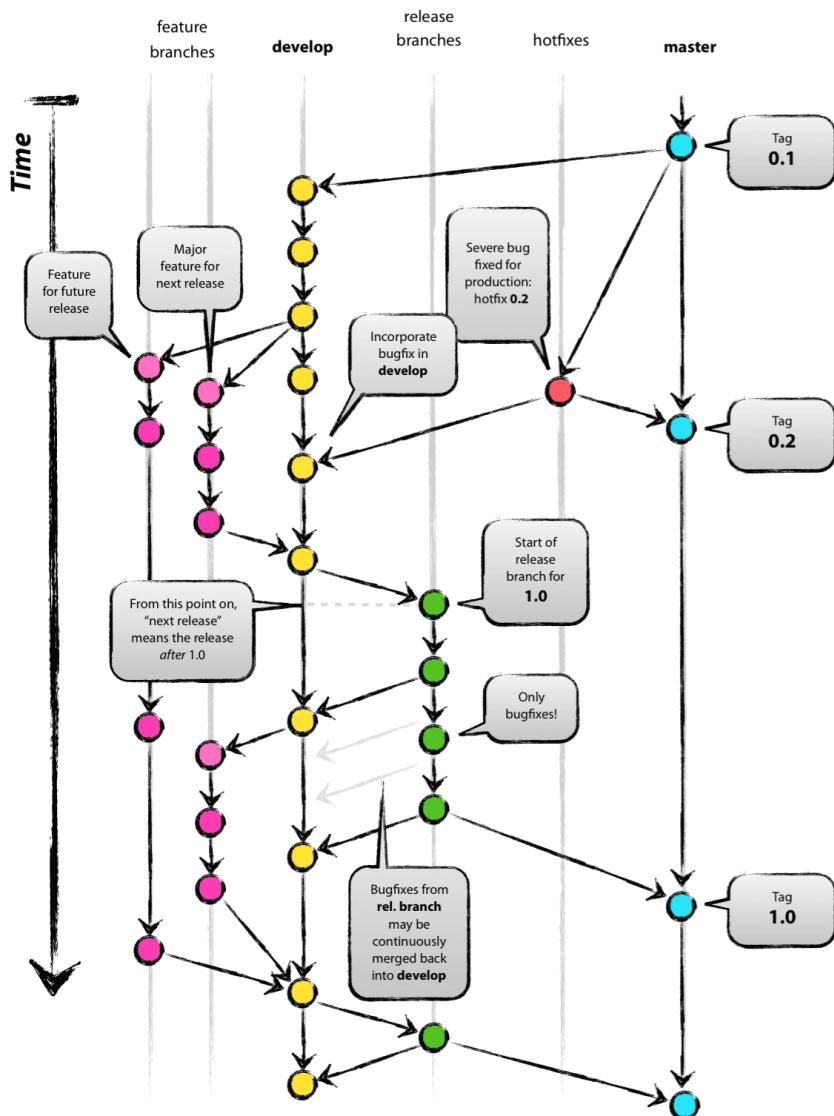


Figura 6.5: *Git Flow*

Teniendo esto en cuenta, se mostrará cómo se ha estructurado cada repositorio de manera general en GitHub, para luego descender, de manera individual, a cada uno.

6.3.1. GitHub

De manera general, se poseen estos repositorios de proyectos en la plataforma de código abierto GitHub. Cada uno de los cuales alberga un desarrollo concreto.

The screenshot shows the GitHub dashboard for the LT-News organization. At the top, there's a yellow icon with a book and the text 'LTNews' followed by the subtitle 'Tu lector inteligente de noticias'. Below this, there are tabs for 'Repositories 4', 'People 1', 'Teams 0', 'Projects 1', and 'Settings'. A search bar says 'Find a repository...', and filters for 'Type: All' and 'Language: All'. A green button labeled 'New' is visible. On the right, there's a sidebar for 'Top languages' (Python, TeX, Vue, Dockerfile) and a section for 'People' showing one user named 'andjimrio' (Andrés M. Jiménez). The main area lists four repositories:

- ltnews-documents**: TeX, Updated 25 minutes ago.
- ltnews-backend**: Python, GPL-3.0, Updated 12 days ago.
- ltnews-frontend**: Vue, Updated 12 days ago.
- ltnews-docker**: Dockerfile, Updated 13 days ago.

Figura 6.6: Respositorios en GitHub

Los repositorios que muestra la imagen son los siguientes:

- **ltnews-documents**: es el proyecto LaTeX donde se desarrollan los documentos a presentar, tanto la presente memoria como la presentación del trabajo.
- **ltnews-docker**: alberga la arquitectura de la aplicación en docker y se relaciona con los demás repositorios de código.
- **ltnews-backend**: es el proyecto de Django para el servidor de la aplicación.
- **ltnews-frontend**: es la capa de presentación de la aplicación realizada en VueJS.

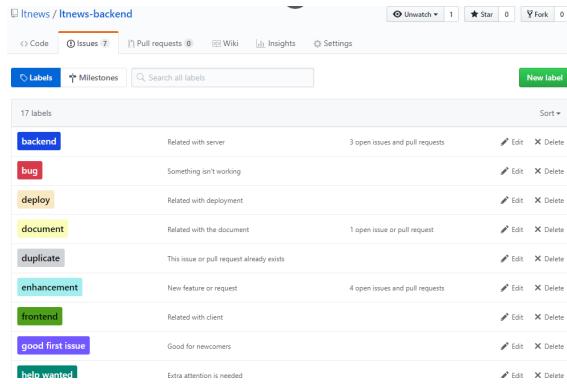
Gracias a la integración que realiza GitHub, se ha añadido metacontenido a los commits que se van realizando. En concreto, se han utilizado tres entidades: *issues*, *labels* y *milestones*.

The screenshot shows the 'Issues' tab for the 'ltnews-backend' repository. At the top, there are buttons for 'Code', 'Issues 7', 'Pull requests 0', 'Wiki', 'Insights', and 'Settings'. Below this, a message encourages labeling issues for new contributors. The main area displays a table of issues:

	Author	Labels	Projects	Milestones	Assignee	Sort
1. 🌟 Newspaper extraction backend enhancement	#31 opened 13 days ago by andjimrio	54 - Capa Intelig...				
2. 🌟 Daily analysis backend	#32 opened 13 days ago by andjimrio	54 - Capa Intelig...				
3. 🌟 Profile extraction backend enhancement	#33 opened 13 days ago by andjimrio	54 - Capa Intelig...				
4. 🌟 Document document	#30 opened 13 days ago by andjimrio	53 - Virtualizaci...				
5. 🌟 News article extraction enhancement social	#35 opened 13 days ago by andjimrio	53 - Social				

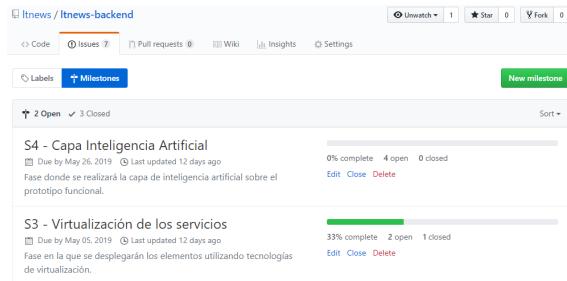
Figura 6.7: Issues en Github

Del primero, los *issues*, comentar que GitHub permite usar esta como una herramienta integrada de gestión de incidencias. Como aparece en la siguiente imagen, cada *issue* viene relacionado

Figura 6.8: *Labels* en Github

con: proyectos, hitos, encargado o etiqueta. Además, posee una traza de comentarios y commits relacionados. Es este el motivo de su uso como gestor de tareas del proyecto.

A la vez, como se ha dicho, se pueden utilizar diferentes etiquetas para relacionar las incidencias. Estas son las que aparecen en la imagen 6.8. La idea general es poder clasificar cada tarea según su entidad correspondiente: plataforma relacionada o tipo de tarea. De esta manera, es fácilmente clasificable cada una de las tareas.

Figura 6.9: *Milestones* en Github

En la última entidad, se agrupan también las tareas o *issues* en hitos, llamadas *milestones*. Esto sirve para indicar las fases del proyecto y la fecha de finalización de cada sprint.

A nivel más operativo, que un repositorio esté en GitHub conlleva a que estén presentes los siguientes archivos:

- .gitignore: es el archivo que contiene los elementos que deberá ignorar git en su proceso de control de versiones.
- LICENCE: es la licencia del proyecto, que en el presente caso es MIT Licence: esta es una licencia de software permisiva, que posee pocas limitaciones en la reutilización.
- README.md: contiene la información imprescindible del proyecto que aparecerá reflejada en GitHub.

6.3.2. Backend

En cuanto al código del servidor, se ha estructurado el código siguiendo las pautas recomendadas tanto por Python, como por Django. Así, la estructuración en carpetas es la que aparece en la imagen 6.10. Estará se irá desgranando para explicar su lógica.

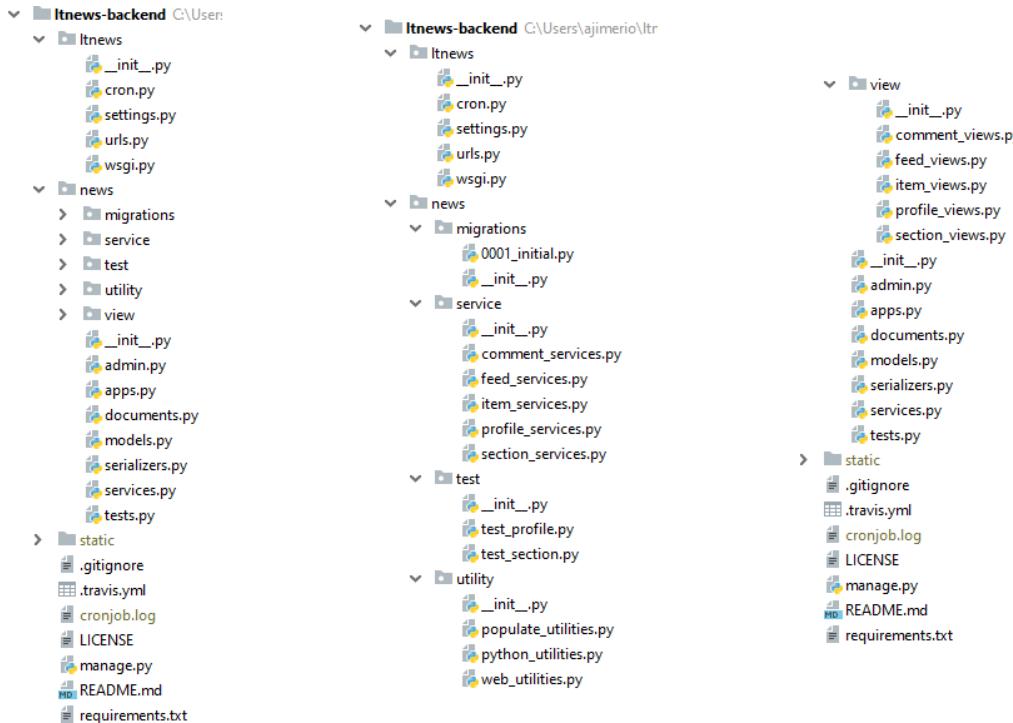


Figura 6.10: Estructura en PyCharm

- **Itnews:** contiene la lógica de funcionamiento de Django.
 - cron.py: archivo que contiene las tareas recurrentes de la aplicación, entre ellas, la que actualiza las noticias y la que calcula el perfil del usuario.
 - settings.py: archivo que contiene las propiedades principales de nuestra aplicación.
 - urls.py: archivo con los accesos a las direcciones accesibles a nuestro proyecto; relaciona controlador con url.
 - wsgi.py: archivo que tiene la información y propiedades del despliegue del servidor Django.
- **news:** contiene la lógica de la aplicación implementada.
 - **migrations:** contiene las migraciones hechas del modelo de Django.
 - **service:** contiene los servicios de las entidades por cada entidad del modelo y son intermediarios entre el modelo y el controlador, ya que extraen la información de la base de datos ya ordenada y útil de utilizar en los controladores.
 - **test:** contiene los tests agrupados en las entidades del sistema.
 - **utility:** contiene las clases de utilidad que han sido necesarias desarrollar la el correcto funcionamiento del sistema.
 - **view:** contiene los controladores por entidad, en concreto, la implementación de cada API publicada.
 - admin.py: archivo en el que se declara el nombre de las entidades del sistema.
 - apps.py: archivo con nombre y características de la aplicación.
 - documents.py: archivo que contiene la comunicación de Elasticsearch con la aplicación así como la implementación de los documentos.
 - models.py: archivo que contiene las entidades del modelo.

- serializers.py: archivo que contiene la implementación de serialización de cada entidad del modelo a través de la API.
- services.py: archivo con los servicios genéricos que usan todas las entidades.
- tests.py: archivos que contiene las pruebas generales del sistema.
- **static**: alberga los ficheros estáticos generados para el panel de administración.
- .travis.yml: los comandos que se ejecutarán en la herramienta de pruebas de integración continua Travis.
- manage.py: es el archivo principal de Django, el archivo de ejecución principal.
- requirements.txt: contiene todos los componentes que usa el proyecto de Python, así como sus versiones.

6.3.3. Frontend

El código del cliente se estructura siguiendo las directrices del CLI de Vue. Estas no son más que una implementación concreta y ordenada de las pautas de Node y Webpack. Así, la organización enseña muestra en la imagen 6.11.

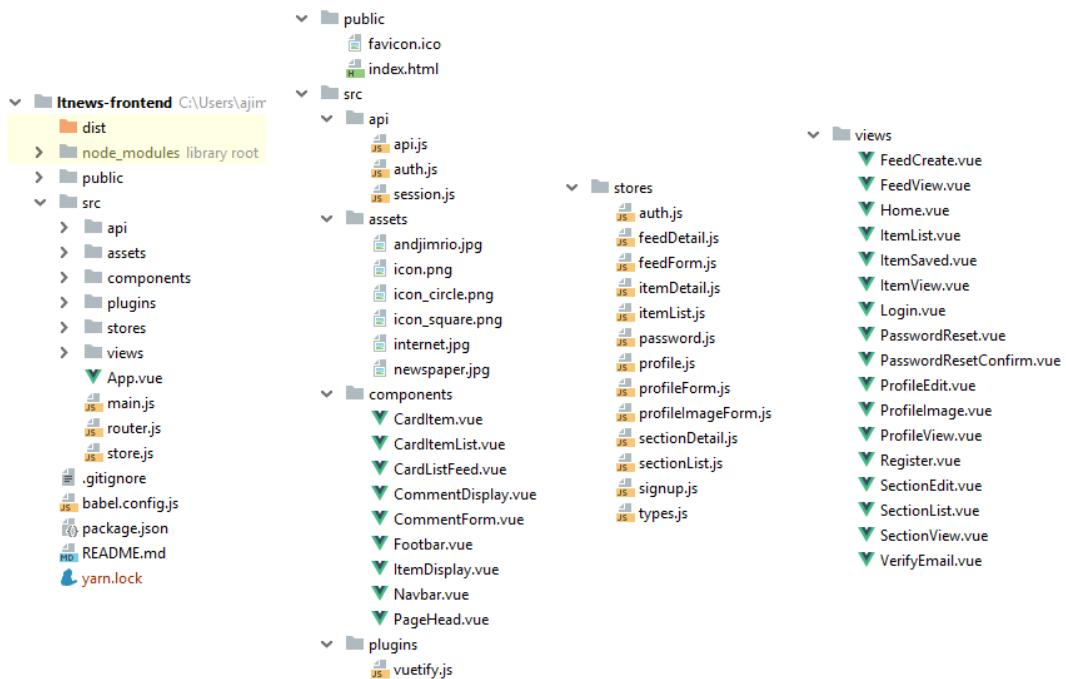


Figura 6.11: Estructura en WebStorm

- **dist**: contiene la construcción de la aplicación en formato SPA realizada por *npm* o *yarn*.
- **node_modules**: contiene los paquetes necesarios por *npm* para la construcción y ejecución de la aplicación.
- **public**: contiene los ficheros globales de la aplicación tales como el índice o el favicon.
- **src**: contiene la lógica general de la aplicación siguiendo la estructuración de Webpack.

- **api**: contiene las llamadas generales a la API y la configuración de estas por *axios*.
 - **assets**: contiene las imágenes y ficheros estáticos usados por la aplicación.
 - **components**: contiene los componentes de la aplicación: estos funcionan de manera independiente y pueden ser usados en cualquier vista.
 - **plugins**: contiene la configuración de algunos módulos usados por la aplicación, como en el caso de *Vuetify*.
 - **stores**: contiene la configuración de estados de las diferentes entidades creadas; don gestionadas por *vuex*.
 - **views**: contiene las vistas que posee la aplicación con la lógica implementada en cada una.
 - App.vue: archivo con la arquitectura visual que presenta la aplicación.
 - main.js: archivo con la configuración general de la aplicación Vue.
 - router.js: archivo con la navegabilidad entre las diferentes vistas así como su relación con las URLs.
 - store.js: archivo con la relación de estados y mutaciones especificados en la carpeta **stores**.
- babel.config.js: archivo con la configuración de Babel.
 - package.json: archivo con el fichero de configuración de la aplicación con la información general y de librerías necesarias y uso específico de estas.
 - yarn.lock: archivo con la configuración generada por las herramientas de node antes mencionadas.

6.4– Detalles de implementación

CAPÍTULO 7

Pruebas

7.1– Pruebas unitarias

Los test unitarios prueban una funcionalidad en toda su completitud y tipología de casuísticas. Debido a la complejidad y gran tiempo que conlleva la realización de las pruebas, se han recudido estas a una parte representativa y en la parte del servidor.

Para desarrollarlas, por tanto, se ha utilizado el framework de pruebas que proporciona Django. Con él, solamente haciendo uso de la función *Test*, se pueden ejecutar todas las pruebas a la vez. Se han comprobado principalmente los usos principales sobre una parte de la aplicación.

En total se han realizado en torno a 20 prueba unitarias de dos entidades exclusivamente: *Section* y *Profile*. Estas poseen toda la lógica que se ofrece por la API probadas, haciendo uso de una nueva base de datos para ello. Además, contienen tanto las pruebas positivas y negativas. Esto quiere decir que de estas entidades de comprueban el correcto funcionamiento, así como los posibles que podrían acontecer.

7.2– Pruebas integradas

Como se ha dicho, los tests unitarios comprueban una funcionalidad concreta del sistema. Aunque hubiera pruebas sobre cada característica concreta del sistema, esto no implicaría la comprobación total de la aplicación. Es por ello importante la comprobación de todas las funcionalidades en su conjunto. Estos son las pruebas integradas.

Estas se realizan en la aplicación utilizando un sistema externo: Travis. Este, con un fichero de configuración, visto en el apartado 6.3, compone la arquitectura del sistema y realiza todo el conjunto de pruebas de manera global. Así, se consigue integrar toda la funcionalidad en una misma ejecución.

La siguiente imagen muestra los diferentes logs de las últimas ejecuciones de Travis. Se puede ver que hay algunos en error. Esto es debido a el cambio de composición de la aplicación: pasa de el montaje de cada componente individualmente a una servicios interconectados y virtualizados con Docker.

7.3– Pruebas de aceptación

7.4– Estándares de código

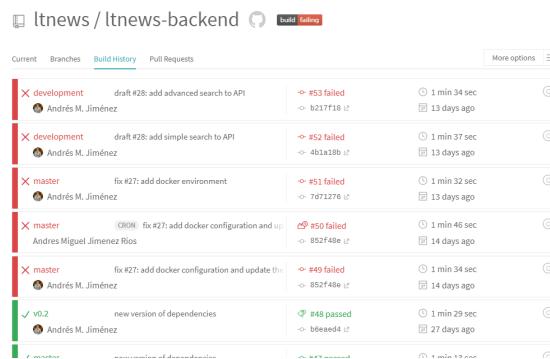


Figura 7.1: Muestra de logs de Travis

CAPÍTULO 8

Conclusión

8.1– Mejoras futuras

8.2– Lecciones aprendidas

Referencias

- Ainhoa Lafuente. (2019). *Qué es el web scraping*. Descargado 2019.04.20, de <https://aukera.es/blog/web-scraping/>
- Ana Pérez Barredo. (2014). *España se queda sin google news*. Descargado 2019.02.24, de https://elpais.com/politica/2014/12/16/actualidad/1418718308_671454.html
- de Trabajo e Inmigración, M. (2009). Resolución de 18 de marzo de 2009, de la dirección general de trabajo, por la que se registra y publica el xvi convenio colectivo estatal de empresas de consultoría y estudios de mercado y de la opinión pública. *Boletín Oficial del Estado*, 3, 28–32. Descargado de <https://www.boe.es/boe/dias/2009/04/04/pdfs/BOE-A-2009-5688.pdf>
- Ian Buckley. (2018). *6 reasons why python is the programming language of the future*. Descargado 2019.04.21, de <https://www.makeuseof.com/tag/python-language-future/>
- Mauyria, A. (2011). *Running lean*. O'Reilly.
- Miguh Ruiz. (2018). *Los 5 frameworks de javascript para frontend más usados en 2018*. Descargado 2019.04.23, de <https://openwebinars.net/blog/los-5-frameworks-de-javascript-para-frontend-mas-usados-en-2018/>
- Paul Bonduelle. (2018). *The data science behind recommendations in feedly*. Descargado 2019.02.24, de <https://blog.feedly.com/data-science-behind-recommendations-in-feedly/>
- Quentyn Kennemer. (2018). *What's new in google news at google i/o 2018*. Descargado 2019.02.24, de <https://www.androidcentral.com/whats-new-google-news-google-io-2018>
- Rana Negra. (2018). *¿cuántos sitios web hay en internet?* Descargado 2019.02.17, de <https://www.rananegra.es/blog/cuantos-sitios-web-hay-internet>
- The State of JavaScript. (2019). *The state of javascript 2018*. Descargado 2019.04.21, de <https://2018.stateofjs.com/introduction/>
- Waffle. (2019). *Farewell from waffle*. Descargado 2019.04.26, de <https://blog.waffle.io/farewell-from-waffle-%EF%B8%8F-794da4a72851>
- Wikipedia. (2018). *Flipboard*. Descargado 2019.02.24, de <https://es.wikipedia.org/wiki/Flipboard>