



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA

Trabajo fin de Máster

Máster en Ingeniería Informática

Lector de noticias inteligente

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

**Dirigido por
Alejandro Fernández-Montes**

**Departamento
Lenguaje de Sistemas Informáticos**

Sevilla, 24 de febrero de 2019

Resumen

¿Por qué LT-News? Surge de la necesidad de todos de informarse, pero la enormidad de medios y el poquísimos 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 anhelum 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	VII
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
2.2 Primer estudio	9
2.3 Iteraciones	9
2.4 Conclusión estudio	9
3 Planificación	10
3.1 Metodología SCRUM	10
3.2 Planificación temporal	10
3.3 Roles del proyecto	10
3.4 Planificación económica	10
4 Análisis	11
4.1 Actores del sistema	11
4.2 Modelo conceptual	11
4.3 Catálogo de requisitos	11
4.4 Diagrama de secuencia	11

5	Diseño	12
5.1	Patrones	12
5.2	Modelo de datos	12
5.3	Prototipos	12
6	Implementación	13
6.1	Tecnologías	13
6.2	Herramientas	13
6.2.1	Backend	13
6.2.2	Frontend	13
6.2.3	Servidor	13
6.2.4	Gestión	13
6.3	Estructura del proyecto	13
6.3.1	Backend	13
6.3.2	Frontend	13
6.4	Detalles de implementación	13
7	Pruebas	14
7.1	Pruebas unitarias	14
7.2	Pruebas integradas	14
7.3	Pruebas de aceptación	14
7.4	Estándares de código	14
8	Conclusión	15
8.1	Mejoras futuras	15
8.2	Lecciones aprendidas	15
	Referencias	16

Índice de cuadros

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

Índice de código

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 múltiples 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*.

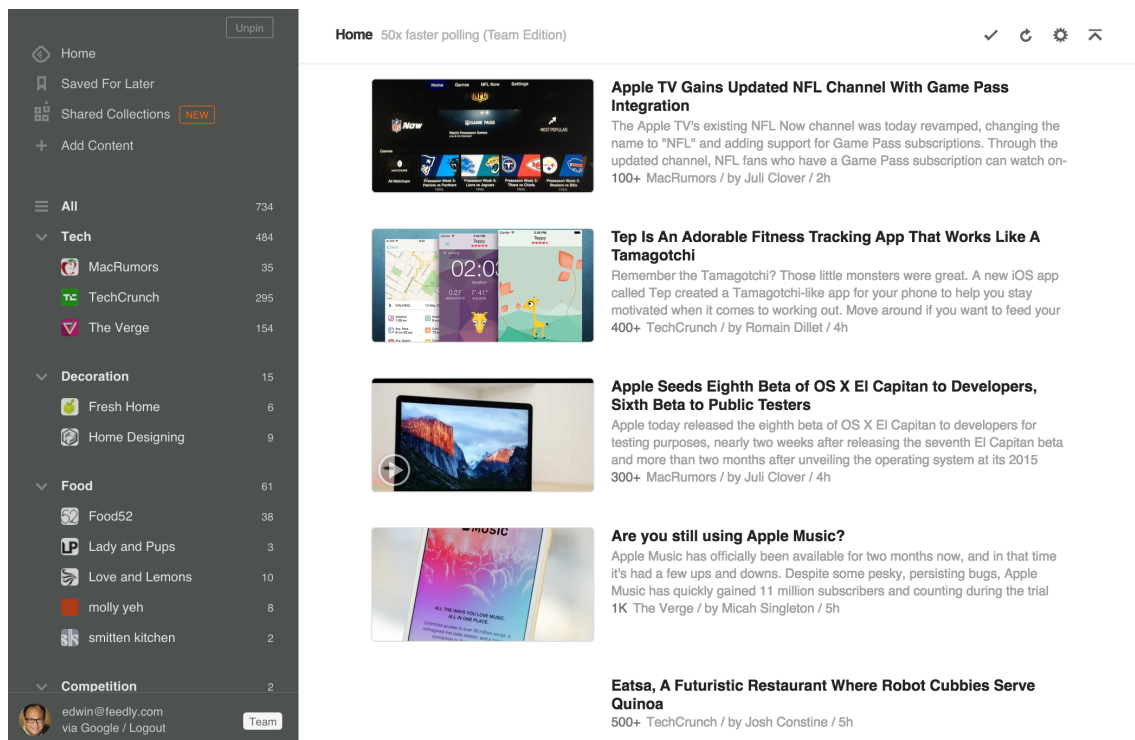


Figura 1.2: Página inicio de Feedly

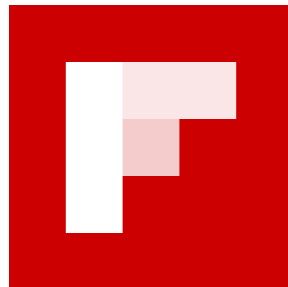


Figura 1.3: Logo de Feedly

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).

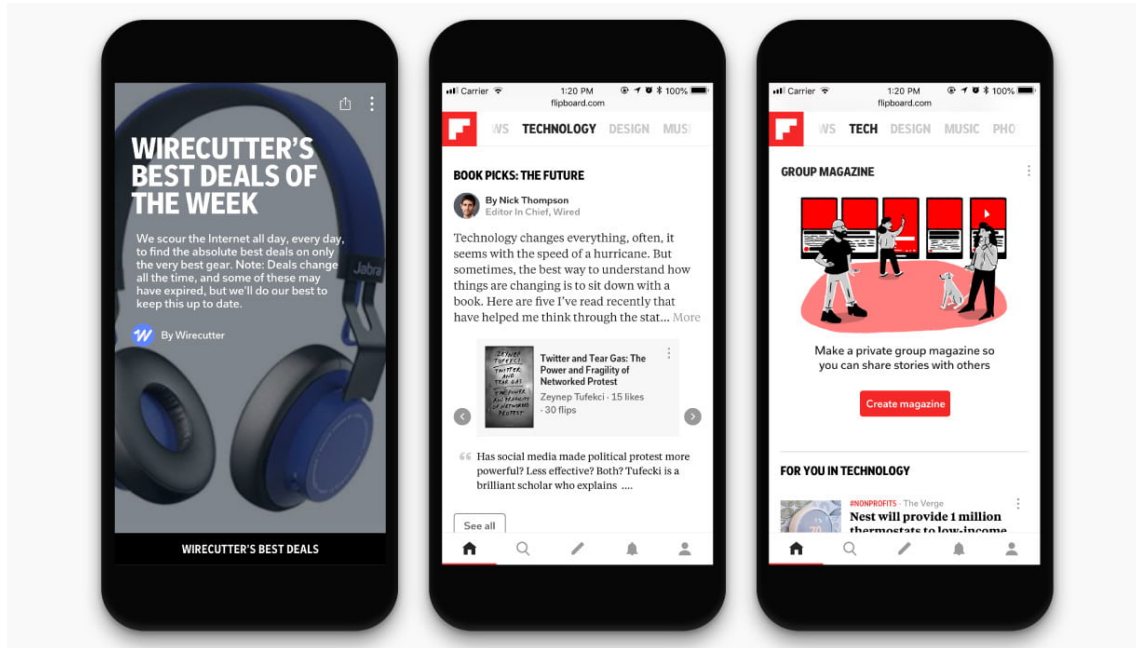


Figura 1.4: Página inicio de Flipboard

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 Kenemer (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.

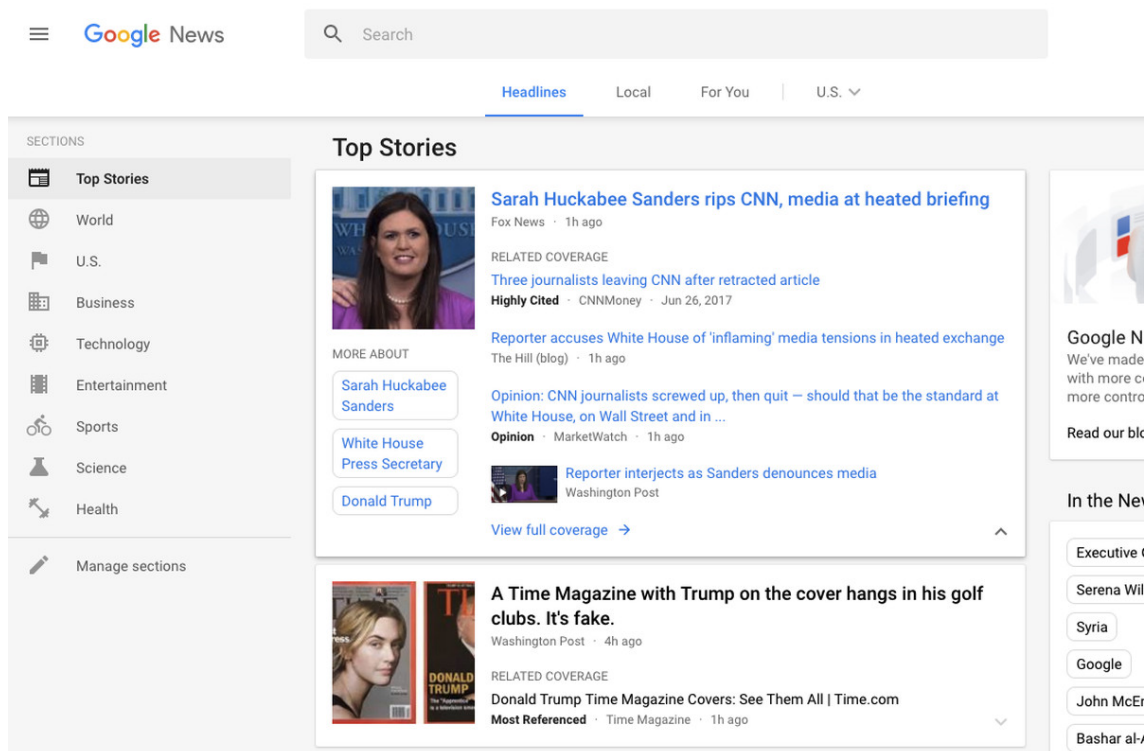


Figura 1.6: Página inicio de Google News

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 seleccionen el usuario, funcionando al estilo PaaS (*Plataforma as a Service*). Esto quiere decir, que los clientes, tanto webs, como de escritorio o de aplicaciones, que encontremos 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 sobre coste 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 coacción 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

2.2– Primer estudio

2.3– Iteraciones

2.4– Conclusión estudio

Planificación

3.1– Metodología SCRUM

3.2– Planificación temporal

3.3– Roles del proyecto

3.4– Planificación económica

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

Implementación

6.1– Tecnologías

6.2– Herramientas

6.2.1. Backend

6.2.2. Frontend

6.2.3. Servidor

6.2.4. Gestión

6.3– Estructura del proyecto

6.3.1. Backend

6.3.2. Frontend

6.4– Detalles de implementación

Pruebas

7.1– Pruebas unitarias

7.2– Pruebas integradas

7.3– Pruebas de aceptación

7.4– Estándares de código

CAPÍTULO 8

Conclusión

8.1– Mejoras futuras

8.2– Lecciones aprendidas

Referencias

- 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
- 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>
- Wikipedia. (2018). *Flipboard*. Descargado 2019.02.24, de <https://es.wikipedia.org/wiki/Flipboard>