



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, 17 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	IV
Índice de figuras	V
Índice de código	VI
1 Introducción	1
1.1 Justificación del proyecto	1
1.2 Objetivos	1
1.2.1 Objetivos funcionales	2
1.2.2 Objetivos docentes	2
1.3 Alcance	2
1.4 Estado del arte	3
Referencias	4

Índice de cuadros

Índice de figuras

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 una 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 deseara.

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 docentes

Bajo este punto de vista englobo tres grandes objetivos que me planteo a conseguir gracias al trabajo, desde un punto de vista docente.

El primero es el aprendizaje de la técnica web scraping. Gracias a esta se podrá obtener una noticia de cualquier periódico, incluyendo, además de los datos que obtengamos del feed, su cuerpo principal o la imagen de la noticia.

La segunda meta es la capacidad de trabajar con texto libre. A través de este trabajo quiero entrar, aunque sea un poco, dentro del Procesamiento de Lenguaje Natural. Gracias a esto, extraeré las keywords en base al texto de la noticia y podré relacionar noticias.

El tercer objetivo es la implementación de un sistema de recomendación. Gracias a esto, podré recomendar a los usuarios noticias en base a sus gustos, tal y como se ha explicado anteriormente.

1.3– Alcance

Con este proyecto quiero realizar una aplicación web capaz de leer feeds RSS. Tendrá una vista que permita la búsqueda de noticias de los diarios, en base a múltiples filtros. Además, deberá extraer los gustos de cada usuario en base al uso que haga de la aplicación, para, más tarde, recomendarle noticias. Se quiere hacer también uso de la relación que haya entre noticias, por ejemplo, agrupándolas según sus temas.

La aplicación que se desee hacer deberá ser accesible desde la web y será compatible con los principales navegadores y con los diferentes dispositivos (ordenadores, tabletas o smartphones).

Además, tendrá soporte de feeds multilingüe, en el sentido que se podrán añadir medios en diferentes idiomas, funcionando igual de bien la aplicación. Por último, la página deberá estar, como mínimo, en dos idiomas: español e inglés.

1.4– Estado del arte

Como se ha dicho a lo largo de la introducción, nuestro sistema de información viene a resolver un problema actual: navegar por la enorme cantidad de información disponible en la red. Por eso se ha dicho es un lector RSS enriquecido. Para ver qué significa ese término de enriquecido, veamos cuáles son y qué hacen los principales lectores RSS.

Referencias
