



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, 12 de enero 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	2
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 del proyecto es hacer un lector RSS enriquecido; esto es, que de un valor añadido a los que ya existen. En concreto, me propongo tres características a alto nivel que, creo, puedan dar este nuevo valor:

- Ver las noticias de los periódicos que elija el usuario.
- Buscar las noticias en base a distintos filtros que el usuario quiera.
- Recomendar noticias a los usuarios.

El primer objetivo, como vemos, es, y debe ser, común a todos los lectores RSS. Es por ello, que nuestro sistema debe ser capaz de extraer las noticias regularmente de los medios que estén en la base de datos y de mostrar dichas noticias de una manera clara y atractiva. El segundo objetivo es claro: nuestra aplicación debe analizar las noticias para permitir su búsqueda.

Por último, se propone un valor añadido sobre los demás lectores: la capacidad de recomendar noticias a los usuarios. Para ello, es necesario un doble aspecto. El primero es que, gracias al punto anterior, tenemos analizadas las noticias que poseemos. Gracias a esto, podemos relacionar noticias entre sí según los temas que traten. El segundo aspecto es el análisis de los usuarios de la aplicación: se extraerá un perfil del mismo según sus gustos. Gracias a esta relación entre usuarios-temas, y noticias-temas, podemos relacionar usuarios con noticias, es decir, recomendar.

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
