

Semandal

Capítulo 1

Introducción

1.1. State of the art

Actualmente en la web podemos encontrar publicada bastante información sobre nuestros municipios y sus ayuntamientos, a veces el acceso a esta información es complicado, ya sea por el formato en el que se encuentra o lo complicado que se le hace al usuario llegar hasta dicha información. Podríamos decir que las páginas webs de los ayuntamientos carecen de un modelo estructural a seguir por lo que buscar cualquier cosa en ellas puede ser un caos.

Por ello se debería buscar una forma de facilitar al usuario el acceso a dichos datos dotándolos de una estructura ya que una noticia es una noticia en todos los ayuntamientos, pero, ¿qué pasa con la palabra actualidad? ¿Viene a significar lo mismo que noticia?

Realmente, una noticia de actualidad es una noticia que ha sucedido recientemente, pero al fin y al cabo, no deja de ser una noticia. En la misma línea nos podemos encontrar que se denomina a un apartado, que contiene la misma información, de forma distinta en 2 pueblos. Y si ésta es distinta en 2 pueblos, la variedad que encontramos a la hora de analizar todas las páginas webs andaluzas es bastante amplia.

Para ello debemos dotar a la información de una estructura, de un significado, de una semántica. Por lo cual debemos crear una semántica con una estructura común y adaptar la información de los ayuntamientos a dicha estructura.

Lógicamente la variedad de contenido en una página web al ser tan grande debemos ir paso a paso con lo que iremos analizando la información en distintas etapas.

A parte del análisis y obtención de información, también queremos que nuestro trabajo no sea en vano y darle al usuario una forma de acceder a dicha información de una forma simplificada, de tal forma que éste pueda consultar algo de la misma forma en cualquier ayuntamiento y además permitir la interacción entre usuarios a partir de una aplicación que tendrá nuestro proyecto.

1.2. Proyecto SEMANDAL

primero justificar la necesidad del proyecto... que no está hecho

Dadas la necesidad de jerarquización de datos, Semandal nace como un proyecto orientado a resolver este problema. Como hemos dicho anteriormente se necesitan tres partes claramente diferenciadas en el proyecto que son las siguientes:

Extracción de información

Para poderle ofrecer al usuario ciertos datos sobre los pueblos y municipios andaluces a parte de capturar la información de éste de su página web, también debemos recopilar otra serie de datos como su propia página web, localización, deuda, población, etc . . . Por tanto debemos recopilar información desde varias fuentes. Como tenemos que obtener datos de fuentes tan diversas como Wikipedia, Google o el INE, salvo esta última, todas tienen una forma de extracción de datos propia, y también debemos usar varios recursos de Google como la geolocalización. Realizaremos dicha obtención de datos con técnicas de Aprendizaje automático.

Poner referencia al dibujo de la arquitectura

Tratamiento y almacenamiento de la información

Aquí es donde realmente SEMANDAL pone su nombre en práctica y se realiza la semantización y jerarquización de datos. Le damos a la información un significado y la almacenamos de tal forma que esta pueda ser fácilmente recuperada de forma automática por un software, que es lo que usaremos para la publicación de la información.

Publicación de la información

El último pero no menos importante apartado de nuestro proyecto ya que éste será el que verá el usuario final. En este apartado es donde recopilaremos la información de la jerarquía creada y la mostraremos al usuario de una forma sencilla y concisa.

Aquí lo que habría que decir es... que hay que recoger información de forma automática(ya se verá cómo)... y publicarla de forma que sea accesible por las máquinas

1.3. Objetivos

El primer paso es ver la estructura interna de las páginas que vamos a visitar, con el tiempo nos hemos dado cuenta que la mayoría de las páginas webs de los ayuntamientos están basadas en plantillas u otros gestores de contenido, la mayoría de ellas porque son encargadas a una empresa. Dicha empresa usa una plantilla normalmente creada por ellos mismos para construir la web de uno o varios ayuntamientos.

en realidad era toda la info de la página... pero empezamos con un ejemplo que son las noticias.

La intención sería intentar recopilar toda la información de las páginas webs de los ayuntamientos pero como por algo debemos empezar, lo hacemos con un ejemplo para ver la dificultad de la extracción de dichos datos como son las noticias de todas las páginas webs de los pueblos de Andalucía indiferentemente de la forma en la que se encuentre el contenido estructurado en dicha página. La aplicación también deberá actualizar las noticias nuevas que vayan saliendo en cada web de los ayuntamientos. También extraemos información relativa a los municipios de la página de cada pueblo de Wikipedia y de documentos obtenidos en la web del Instituto Nacional de Estadística.

¿OpenCMS?

Una vez tenemos almacenada la información, el siguiente paso es su publicación. Para ello crearemos un plataforma web donde se permitirá al usuario registrarse y recibir información de manera personalizada según su localización y preferencias. Nuestra aplicación también tendrá que tener interopera-

bilidad con Facebook para así permitir al usuario iniciar sesión en nuestra plataforma a través de Facebook y utilidades de Me gusta y compartir. También se implementará un sistema de registro de usuarios e inicio de sesión usando el protocolo OAuth2. Así mismo, nuestra aplicación para Android tendrá un sistema donde cada usuario podrá valorar de forma positiva o negativa un comentario. Como también queremos que los datos que tenemos almacenados sean accesibles por las máquinas, se creará una API, es decir una serie de funciones que otros programadores pueden incluir en sus programas, para descargar información de nuestra base de datos de forma libre a través de una determinada dirección web que da una respuesta con la información seleccionada en formato JSON.

Todo list

primero justificar la necesidad del proyecto... que no está hecho	4
Poner referencia al dibujo de la arquitectura	4
Aquí lo que habría que decir es... que hay que recoger información de forma automática(ya se verá cómo)... y publicarla de forma que sea accesible por las máquinas	5
en realidad era toda la info de la página... pero empezamos con un ejemplo que son las noticias.	5
¿OpenCMS?	5