

Tipi Ciudadano es una página web cuya fuente de información es la página del Congreso de los Diputados. Permite a la ciudadanía fiscalizar la actividad de todos los grupos parlamentarios presentes en dicha cámara realizando búsquedas específicas sobre las intervenciones de todos los parlamentarios.

Lo hace de acuerdo con 21 taxonomías relacionadas con temáticas y políticas de interés público definidas de acuerdo a criterios consensuados por sus desarrolladores con organizaciones independientes, colectivos de activistas y expertos en cada materia tratada.

Etiqueta cada acto parlamentario de acuerdo a esas 21 taxonomías e incluye 1.700 expresiones regulares que se emplean en la Cámara. Gracias a ello cualquier ciudadano puede averiguar prácticamente al instante cuáles han sido los grupos parlamentarios más activos en según qué materias, qué políticos hablan o elevan a la cámara propuestas, preguntas al gobierno, etc.

[Ellos mismos ofrecen una breve, clara y concisa definición de lo que ofrece la web.](#)

ALGUNAS CARACTERÍSTICAS GENERALES DE TIPICIUDADANO.es

Entrando en cuestiones de tipo más técnico, XXX enumera algunas de las características que definen www.tipiciudadano.es : es una web orientada a datos que dispone de un buscador básico y uno avanzado; permite consultar información relativa a la actividad parlamentaria de acuerdo con criterios temáticos muy definidos y exportarla en formato CSV para poder trabajar con ella.

Ofrece una experiencia de navegación completa, lo que significa que además permite inter-relacionar los resultados que brinda tras cada búsqueda realizada. A esto suma que posibilita la ver una gráfica resumen mediante la que se puede percibir “a vista de pájaro” la información que contiene la web.

Asimismo, elabora rankings mediante los cuales se puede averiguar, por ejemplo, qué diputado o grupo parlamentario es el más activo en relación a temas como la dependencia, conductas adictivas, libertad de expresión, políticas de igualdad, etc.

Para terminar, cualquier ciudadano puede solicitar recibir alertas con el fin de que llegue a su mail información actualizada sobre la actividad parlamentaria en torno a temáticas de su interés.

TIPI res una web que funciona realizando cuatro tareas fundamentales:

Scrap/Labeling/Alerts/Stats, que podrían traducirse como Escraqueo/Etiquetado/Alertas/Estadísticas)

XXXX comenta que la primera tarea que se plantea un equipo a la hora de desarrollar una herramienta de estas características es cómo resolver la problemática de la obtención de datos. En ese sentido, el orden en el que decidieron solventar es:

1-Limpiar los datos

2-Procesarlos. En esta fase se aplican a los conjuntos de datos algoritmos que permiten analizarlo a un nivel básico a partir del cual se pueden aplicar a posteriori otros algoritmos con el fin de generar estadísticas más precisas sobre lo que se desee averiguar.

BASE DE DATOS ESCOGIDA:RAZONES

A la hora de elegir el tipo de base de datos a emplear, la disyuntiva se planteó entre si escoger una [SQL ó No SQL](#). La principal virtud del primer tipo de BBDD está en que enlaza todos los datos. Por ejemplo: Historial médico, con médico de cabecera, Centro de salud que corresponde al paciente, etc. El inconveniente es que se trata de una BBDD poco eficiente que requiere invertir mucho en recursos en caso de que se manejen cantidades ingentes de datos.

Por esa razón optaron por emplear una BB.DD. NoSQL. Éstas se caracterizan por permitir introducir los datos de forma desagregada, en bruto, para que posteriormente puedan ser organizados a gusto de quien diseña la herramienta. Es la mejor opción si se va a manejar una gran cantidad de datos. Su BBDD se llama [MONGODB](#). Lo escogieron por su gran comunidad, por disponer de un gran ecosistema de archivos, por poderse integrar con muchos lenguajes ([Python](#), [Javascript](#)...), por tener una API potente (lo que permite desarrollar aplicaciones empleando los datos que alberga).

LENGUAJE EMPLEADO, CRONES, SCRAPEO Y CÓDIGO ABIERTO

METEORJS es ¿El mecanismo de publicación y suscripción? que escogieron porque permite meter código en el cliente y en el servidor. Hace [bindings](#) sobre elementos HTML y Datos, lo que significa que cada vez que se introduce un nuevo dato en la BB.DD se actualiza en el navegador.

Para poder tener la web actualizada de forma permanente es necesario programar tareas, automatizar chequeos. Por ejemplo, se puede solicitar al sistema que “todos los viernes, calcule las estadísticas de la web”. Establecer estas tareas automáticas es posible gracias a los “crones”. Hasta ahora había un problema, no permitían saber cuándo se terminaba de hacer una tarea y se comenzaba con la siguiente.

Esto fue solventado por Spotify, que desarrolló Luigi, un “[cron](#)” que permite automatizar tareas determinando también su sucesión: La orden que recibe la máquina gracias a Luigi es más concreta: “Haz A y cuando termines de hacer “A”, pasa a hacer “B”..y así sucesivamente. Esto hace posible que la web esté siempre actualizada, lo que es importante teniendo en cuenta que la actividad parlamentaria es constante.

SCRAPEO WEB-La web tipiciudadano [SCRAPEA](#) la del Congreso para rescatar los datos valiosos y facilitarlos de manera más accesible y estructurada. Recomienda a los asistentes que quieran más sobre este proceso utilizar Beautiful Soup, de Phytion. También hace mención a SCRAPY [www.scrapy.org](#), una librería hecha en Python para obtener información de una web. Scrapy es ideal si la web de la que quieres extraer información te banea la IP y te ves en la tesitura de hacer un Proxy para solventar esa barrera.

Etiquetado de textos- La web lo realiza mediante una minería de Datos básica basada en el reconocimiento de patrones. Como se ha expuesto al principio, se etiqueta cada acto parlamentario de acuerdo a los 21 diccionarios temáticos preestablecidos y a 1.700 expresiones regulares.

Otra de las capacidades a destacar es que utilizando la información que brinda TIPI se pueden desarrollar Apps, columnas en páginas web, Widgets para Wordpress en los que aparezca información relativa a la actividad parlamentaria sobre aquellos temas que el usuario considere más importantes para su blog o página web.

Todas las personas que quieran trabajar con el código que han empleado para desarrollar esta herramienta pueden hacerlo visitando la web [www.github.com/ciecode-madrid/tipi](#).