

ITBA - Proyecto Final

Propuesta funcional - Visualización de noticias

Agustina Fainguersch | Juan José Marinelli | Lucas Moscovicz

Tutor: Ariel Aizemberg

1 - Objetivo

Los objetivos principales del trabajo a presentar son:

- La optimización del sistema existente para mejorar la velocidad y calidad de respuesta del sistema demostrando lo propuesto mediante mediciones concretas y métricas detalladas.
- El desarrollo de diversas visualizaciones de noticias recolectadas de varios medios usando la API descrita previamente junto con librerías adecuadas para lo mismo como lo son, entre otras:
 - d3.js
 - dc.js
 - chart.js
 - crossfilter

2 - Propuesta de implementación

Se propone la implementación del sistema en 2 partes, por un lado la mejora del tiempo de respuesta del servidor de provisión de datos, implementado hoy en día en **PHP** junto con una base de datos **Postgres**. Para ello se agregarán los siguientes módulos al proyecto:

- Una instancia de **ElasticSearch**, un indexador de bases de datos que permite optimizar la búsqueda de texto mejorando así la calidad de las visualizaciones ya existentes en la aplicación del tutor.
- Una API hecha en **Java**, paralela a la API existente en PHP para realizar algunas de las consultas extra para obtener información para las nuevas visualizaciones a realizar.

Por otro lado, se desarrollarán nuevas visualizaciones interactivas basadas en las noticias de la base de datos existente usando tanto el backend existente como el que se desarrollará durante el transcurso del proyecto.

Las visualizaciones a agregar son:

- Cantidad de noticias por medio por fecha: Permite, seleccionando un rango de fechas, visualizar en un gráfico de puntos unidos por líneas, cuantas noticias son publicadas diariamente por cada medio seleccionado.
- Ranking de medios por categoría en Google News: Permite, seleccionando una determinada categoría, visualizar cuales son los medios que mayor cantidad de noticias publican en gráficos de barra, torta y radar para el distribuidor de noticias **Google News**.
- Categorías de noticias por medio: Permite, seleccionado uno o varios medios, visualizar la cantidad de noticias por categoría que los mismos publican en un gráfico de radar.
- Word Cloud: Permite ver las palabras más mencionadas en un rango de noticias variando el tamaño de palabra según la cantidad de ocurrencias.

3 - Documentación del trabajo

Una vez implementadas las optimizaciones sobre el servidor, se realizará un trabajo para demostrar y documentar los beneficios del trabajo realizado. Para eso se realizará lo siguiente:

- Recolección de datos: Tiempo medio de respuesta del servidor para obtener:
 - Palabras para generar el word cloud con diferentes fechas y medios. Con y sin **ElasticSearch**.
 - Tiempo de respuesta de la API desarrollada en Java para realizar las nuevas visualizaciones.
- Documentación del proyecto: Descripción detallada de las implementaciones realizadas para facilitar su reproducción una vez terminado el proyecto.
 - API Desarrollada en Java
 - Nuevas Visualizaciones
 - Instancia de ElasticSearch con las mejoras en la API desarrollada en PHP

4 - Entregables

- Código fuente y documentación de la API realizada en Java, así como el instructivo para ponerla en funcionamiento.
- Código fuente y documentación del módulo realizado de **ElasticSearch**, así como el instructivo para ponerlo en funcionamiento.
- Código fuente y documentación de las nuevas visualizaciones agregadas al sistema.
- Informe completo del desarrollo así como de las iteraciones realizadas para llegar al mismo.
- Informe de métricas de la optimización de la API existente.

5 - Avances del proyecto

Los avances del proyecto se registrarán semana a semana en el siguiente documento:

<https://docs.google.com/spreadsheets/d/1ea7Zxd131l5qb9y7IPKjTQize1cw8bNrXkfV3vC7cUw/edit#gid=0>

En el mismo se informará de la planificación para el trabajo a realizar así como del trabajo efectivamente realizado en los períodos de tiempo detallados.