

Analizador de leyes

Loraine Monteagudo García C-311
Ernesto Estevanell Valladares C-412
Dayany Alfaro González C-311

July 16, 2019

1 El Problema

Como resultado del proceso de Reforma Constitucional en la que se encuentra el país, las leyes son renovadas periódicamente. Por lo tanto, resulta de especial interés determinar las diferencias entre estas leyes. Una aplicación original existía, en la cual, analizándose la constitución actual, se determinaba, dado un artículo, cuáles eran los artículos de la constitución de 1976 que tenían mayor relación con él. Además, se señalaban las diferencias tanto sintéticas como semánticas de los artículos. A pesar de la utilidad de la aplicación estas diferencias eran anotadas manualmente, el objetivo principal de este trabajo era automatizar este proceso para analizar las diferencias de las leyes.

2 Solución

Se implementó un programa que recibe como entrada un **json** con una estructura propuesta en la sección siguiente conteniendo las dos leyes a comparar. A pesar de la tarea inicial se decidió comparar por párrafos por lo que se comparará un mismo artículo de ambas leyes a nivel de párrafos.

Para determinar las semejanzas entre los párrafos se utiliza un modelo **LSI** (Latent Semantic Indexing). **LSI** es un método matemático de recuperación e indexación de documentos, o en nuestro caso, contenidos, que, de manera muy sencilla de explicar, lo que hace es tratar de entender los términos y conceptos y lo que los rodea para entender los contextos y así poder relacionar de manera más inteligente distintos documentos. Después de esto se tienen los documentos representados como vectores en un espacio **LSI**, se usa la *Cosine Similarity*, que posee un producto interior con el que se evalúa el valor del ángulo comprendido entre ellos, dando como resultado un valor entre -1 y 1. El documento que tenga valor más cercano a 1 será el que esté más relacionado con el párrafo analizado.

Al no contar con un corpus de entrenamiento inicial se decidió implementar de alguna manera una fuente de conocimiento creciente. Como funcionalidad adicional se encuentra la de guardar en almacenamiento una ley determinada, luego cargarla en el programa y añadirla al corpus. Después, se permiten realizar

queries, en las que se da un artículo, luego por cada uno de los párrafos se determina el párrafo de la vieja ley más similar, devolviéndose el artículo al que pertenece dicho párrafo.

3 Detalles de implementación

Se utilizó como lenguaje de programación Python y el proyecto cuenta con las dependencias a los siguientes módulos:

- `gensim`

El proyecto cuenta con una carpeta `/src` en la que se encuentra el código fuente del programa y otra carpeta `/doc` con la documentación.

Los **json** de entrada deberán contar con una estructura similar a la presentada en el fichero `src/test/data/tlaw.json`.

4 Modo de uso

Se cuenta con el módulo principal **main.py** donde se ofrece una implementación de la clase *LawAnalyzer* que maneja el proceso entero. El programa recibirá como primer parámetro el **json** con la información de las leyes, y como segundo parámetro el artículo a comparar.

El siguiente comando muestra la funcionalidad del programa:

```
python main.py laws article
```