

VOTING

Sistema de orientación partidaria

INTRODUCCIÓN

En este trabajo se utilizan expresiones regulares para reconocer las entidades de interés, las cuales son:

Personas, Votos, Partidos políticos, Expedientes y Leyes.

Y luego se arman relaciones entre estas entidades, las de mayor interés son:

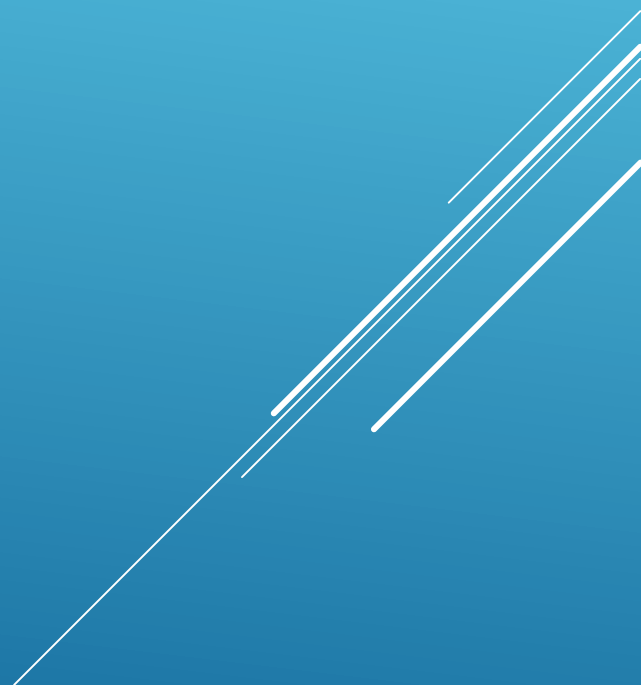
Persona -> Voto, y Persona -> Partido político.

Several white lines of varying lengths and slopes are positioned on the right side of the slide, extending from the middle to the bottom right corner.

CARACTERÍSTICAS Y OBJETIVOS

El objetivo del trabajo es extraer información de documentos en forma de tabla, que indican el tipo de voto emitido por cada diputado con respecto a las leyes indicadas en esos documentos.

Utilizando IEPY es bastante simple separar las oraciones, y formar reglas para relacionar las entidades. Pero al no poder formar relaciones ternarias ni poder relacionar entidades en distintas oraciones, se dificulta extraer relaciones como, las leyes que se están votando con el voto emitido por una persona.



HERRAMIENTAS

En este trabajo se utiliza IEPY como herramienta principal, se le hicieron cambios en la parte de pre procesamiento para adaptarlo a las necesidades del trabajo.

Cabe aclarar que se modifico la instalación de IEPY ya que tenia algunos errores que surgieron al personalizarlo.

Para la interfaz web que realiza y almacena las preguntas se utilizo una aplicación de Django.

DESARROLLO

En este trabajo se necesito implementar un script para pasar los datos a formato csv y así poder cargarlos. Luego durante el pre procesamiento de estos documentos se implemento un tokenizer y un sentencer para poder personalizar las oraciones. Después para concluir con el pre procesamiento se implemento un lemmatizer, POStagger y syntactic parser triviales ya que no son de interés en este trabajo, pero son necesarios para el core de IEPY.

Con la información pre procesada se construyeron reglas para extraer las relaciones entre persona -> voto, y persona -> partido.

También se modifiko el script `iepy_rules_runner` de IEPY para poder correr las reglas de forma personalizada.

CONCLUSIONES

Extraer información de un documento no resulta ser una tarea sencilla, y computacionalmente es muy costosa. Esta aplicación se volvería muy lenta si se cargan grandes cantidades de datos, por lo tanto se debería optimizar el algoritmo para mostrar los resultados.

Con respecto a los resultados obtenidos, a mi me sorprendió bastante, aunque hay que considerar que no se cargaron datos suficientes para obtener valores cercanos a la realidad.

A futuro también se pueden mejorar las reglas para extraer las relaciones o tratar de utilizar el learning core de IEPY con un gran corpus de entrenamiento.