Tarea E12

Alexander Zapata Galindo – Código 201425426

E12 - Natural Language Processing Overview

Research for a commercial product which uses NLP as a core (E.g. google assistant) and describe, in general words, which NLP techniques it uses. (If two of you have the same commercial product, both of your grades will be penalized --- Be creative! ) Due date: 23/07/2021 5:59 pm

<https://forms.gle/N2QLwvy9gxtEthQ48>

En términos generales, el Lenguaje Natural (LN) es el medio que utilizamos de manera cotidiana para establecer nuestra comunicación con las demás personas. Por su parte, el Procesamiento de Lenguaje Natural (PLN) es un campo de las ciencias de la computación, de la inteligencia artificial y de la lingüística que estudia las interacciones entre las computadoras y el lenguaje humano; en otras palabras, consiste en la utilización de un lenguaje natural para comunicarnos con las computadoras, debiendo estas entender las oraciones o frases que le sean proporcionadas.

Las aplicaciones del PLN son muy variadas; se destacan, entre otras: i) traducción automática, ii) recuperación de la información, iii) extracción de información, iv) sistemas de búsquedas de respuestas, v) generación de resúmenes automáticos, vi) reconocimiento de voz, y vii) análisis de sentimientos.

Management Solutions y mrHouston Tech Solutions, tras un proceso de investigación y desarrollo de varios años, desarrollaron una solución integral para la gestión y control del riesgo reputacional, llamada Hatari™, que emplea técnicas de PLN con base en información de medios de comunicación y redes sociales.

La herramienta Hatari™ extrae los temas (‘tópicos’) contenidos en un corpus de textos y determina de forma estadística su impacto sobre las métricas de interés para la organización. Con ello, permite medir el impacto a lo largo del tiempo de eventos reputacionales sobre las ventas, la liquidez, el valor de la acción o las altas y bajas de clientes, o cualquier otro indicador relevante.

Hatari™, en esencia, da respuesta a cuatro preguntas:

* ¿Qué dicen las noticias y redes sociales que sea relevante para la compañía?
* ¿Qué conceptos clave aparecen en esas noticias?
* ¿En qué medida impactan en la compañía?
* ¿Y cómo afectan a los indicadores clave (valor de la acción, ingresos, depósitos, etc.)?

Para ello, de forma esquemática, Hatari™:

* Se alimenta de un corpus de textos en lenguaje natural, como noticias en prensa digital, menciones en redes sociales, textos provenientes de páginas web u otras bases de datos documentales.
* Pre-procesa este corpus mediante técnicas de minería de texto, eliminando la información sin carga semántica (preposiciones, conjunciones, puntuación, etc.), y lo transforma en conjuntos de palabras aptos para la modelización.
* Mediante técnicas de NLP, obtiene los tópicos presentes en los textos, sin necesidad de proporcionarle ninguna información a priori (e.g. palabras clave), y construye una serie temporal con su importancia a lo largo del tiempo.
* Utilizando modelos econométricos, calcula la correlación de la serie temporal de aparición de cada tópico con otras series temporales proporcionadas por el usuario, lo que permite determinar la influencia de un determinado tópico sobre un indicador.
* Adicionalmente, mediante análisis de sentimientos, determina si el tono en el que está escrito cada texto del corpus utilizado es positivo, neutro o negativo.
* Aplicado a la cuantificación del riesgo reputacional, Hatari™ permite determinar qué eventos reputacionales concretos, caracterizados a través de tópicos, impactan en el desempeño de una compañía a lo largo del tiempo, y en qué medida lo hacen.

Información tomada de:

<https://www.managementsolutions.com/es/microsites/soluciones-propietarias/hatari>