

Aplicación de tecnologías de Procesamiento de lenguaje natural y tecnología semántica en Brand Rain y Anpro21

Natural language processing and semantic technologies. The application on Brand Rain and Anpro21

Oscar Trabazos, Silvia Suárez, Remei Bori y Oriol Flo

Anpro21 y Brand Rain
Carrer de Barcelona, 2
17002, Girona
{otc,ssb,rem}@anpro21.com
oflo@bluemapconsulting.com

Resumen: Este artículo presenta la aplicación y resultados obtenidos de la investigación en técnicas de procesamiento de lenguaje natural y tecnología semántica en Brand Rain y Anpro21. Se exponen todos los proyectos relacionados con las temáticas antes mencionadas y se presenta la aplicación y ventajas de la transferencia de la investigación y nuevas tecnologías desarrolladas a la herramienta de monitorización y cálculo de reputación Brand Rain.

Palabras clave: Procesamiento de lenguaje natural, web semántica, ontologías, minería de datos, análisis de reputación, análisis de sentimiento, aprendizaje automático.

Abstract: This paper presents the application and results on research about natural language processing and semantic technologies in Brand Rain and Anpro21. The related projects are explained and the obtained benefits from the research on this new technologies developed are presented. All this research have been applied on the monitoring and reputation system of Brand Rain.

Keywords: Natural language processing, semantic web, ontologies, data mining, text mining, reputation analysis, sentiment analysis, machine learning.

1 Introducción

Somos una empresa de tecnología que investiga técnicas de Inteligencia artificial (IA) y de procesamiento del lenguaje natural (PLN) para desarrollar herramientas y servicios para el campo del marketing, la comunicación y el branding.

Con una trayectoria de casi 10 años a las espaldas, en 2010 lanzamos al mercado uno de los software de análisis de la reputación de marca más completos de Europa, se trata de Brand Rain.

Una meta-herramienta de monitorización y análisis de la reputación en la cual se aplica la labor de investigación que desarrolla Anpro21 y que ya incorporan en su día a día centenares de gabinetes de comunicación y empresas de sectores bien diversos.

Brand Rain permite capturar, monitorizar y analizar toda la información que se publica respecto a una marca, entidad o persona, medir la reputación y el sentimiento de estas informaciones y determinar la reputación de que goza la marca en los medios off-line (prensa, radio y televisión) on-line y en las redes sociales. Precisamente, esta es una de las particularidades de Brand Rain, tres herramientas en una, para analizar todos los impactos de una marca mediante un software de uso sencillo e intuitivo.

En lo que se refiere al campo de I+D+i centramos nuestras líneas de investigación en el área de la Inteligencia Artificial, PLN, técnicas de indexación, recuperación de la información y Big data. Este artículo pretende exponer los proyectos que llevamos a cabo, las

tecnologías que usamos y el cómo aplicamos las técnicas a nuestro software Brand Rain.

2 *Estado del arte*

En los últimos años, las aportaciones que se han hecho desde el PLN han mejorado sustancialmente, permitiendo el procesamiento de ingentes cantidades de información en formato texto con un grado de eficacia aceptable. Muestra de ello es la aplicación de estas técnicas como una componente esencial en los motores de búsqueda web, en las herramientas de traducción automática, o en la generación automática de resúmenes.

Para el desarrollo de PLN se han usado diferentes herramientas y analizadores sintácticos, semánticos y de dependencias. Un analizador sintáctico (o parser) es una de las partes de un compilador que transforma su entrada en un árbol de derivación. Algunos ejemplos de analizadores sintácticos y de dependencias son Freeling [Lluís Padro, 2013], Gate [Sonal Gutap, 2014], Stanford Parser [H. Cuniham, 2012].

3 *Proyectos de I+D+i de Anpro21*

Actualmente trabajamos en 3 grandes proyectos, en colaboración con la Universitat de Girona, la Universidad de Salamanca y consolidadas empresas del sector tecnológico como Ibermática y Verbio Technologies.

3.1 *Vídeo-reputación*

El proyecto de vídeo-reputación persigue el objetivo final de desarrollar un sistema de análisis de los contenidos multimedia que sea capaz de interpretar imágenes, vídeos y voz para determinar su temática, su significado y su tono para finalmente poder cuantificar y determinar la reputación corporativa de las empresas en internet.

Esta plataforma permite monitorizar y analizar el contenido audiovisual mediante técnicas de interpretación de la voz, el audio, los gestos y el tono que se desprenden de los mensajes.

Para este proyecto se aplican técnicas de IA relacionadas con reconocimiento de patrones en imágenes y vídeos, reconocimiento automático de voz y tono o modulación de la voz y procesamiento de lenguaje natural.

El desarrollo de técnicas de análisis audiovisual nos permitirá también acercarnos a la detección de la ironía y el sarcasmo, uno de los grandes retos de las herramientas de monitorización y análisis de la reputación.

3.2 *Redes complejas*

El proyecto de redes complejas tiene que ver con la detección de crisis de reputación o amenazas en la red. Mediante tecnologías de redes complejas llegamos a establecer, localizar y definir una red para la detección de personas que influyen la reputación de una marca hasta dar con el núcleo de la red, el influenciador clave.

Esto será de gran ayuda para la gestión de crisis de reputación y también tendrá otras aplicaciones como pueden ser la detección de amenazas en la red, como el acoso escolar o la violencia de género.

Para este proyecto se aplican tecnologías de minería de texto, machine learning, boosting de autor y procesamiento de lenguaje natural.

3.3 *Find Your Fund (FUF)*

Find your Fund está diseñado para enlazar emprendedores y empresas en búsqueda de financiación con inversores mediante técnicas de procesamiento del lenguaje natural. Para llevarlo a cabo se usan analizadores sintácticos (Freeling, Gate), ontologías y minería de textos.

4 *Aplicaciones de las tecnologías en Brand Rain*

Los distintos proyectos de investigación desarrollados por Anpro21 nutren continuamente el sistema de Brand Rain, y es el mismo software el que muestra las necesidades de seguir investigando para perfeccionar y afinar en el análisis de la reputación y en la detección de la ironía y el sarcasmo, el gran caballo de batalla del sector.

Por lo tanto, las tecnologías desarrolladas en Anpro21 se incorporan al software Brand Rain como nuevas funcionalidades que permiten actualizar y perfeccionar el servicio que da esta herramienta día a día a los profesionales de la comunicación.

Cálculo de la Reputación

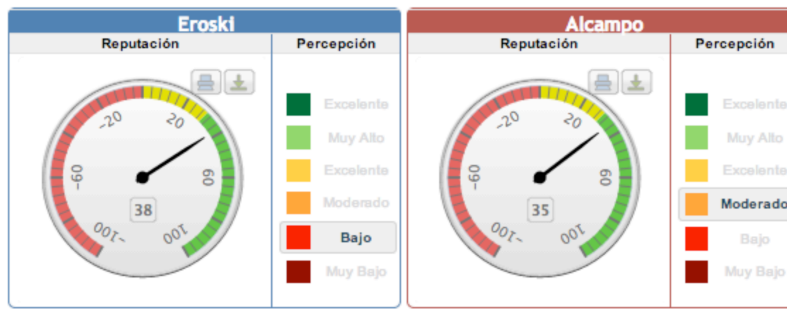


Figura 1: Cálculo de la reputación tal y como se muestra en Brand Rain.

Dimensiones de la reputación

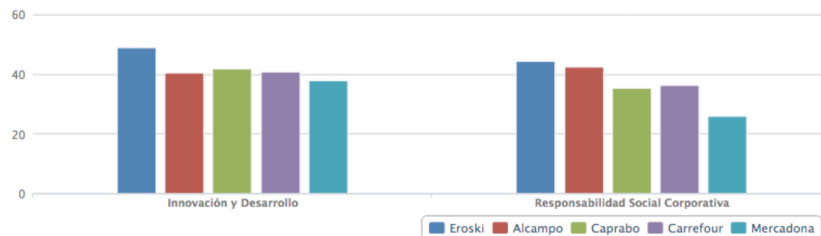


Figura 2: Dos dimensiones de la reputación. Innovación y desarrollo y responsabilidad social.

En este capítulo vemos las aplicaciones de las tecnologías mencionadas en el capítulo anterior en dos de las funciones más destacadas de Brand Rain, el cálculo de la reputación y la detección de influenciadores.

4.1 Cálculo de la reputación y análisis de sentimiento

En Brand Rain el cálculo de la reputación (Figura 1) de una marca se realiza a partir del análisis de las menciones, noticias y conversaciones capturadas sobre la misma. Para llevar a cabo este análisis contamos con un sistema de diseño propio que se encarga del estudio semántico y del contexto que envuelve cada una de las noticias. Para el análisis de las menciones utilizamos algoritmos de análisis de texto, de PLN (Analizadores sintácticos y morfológicos) y patrones lingüísticos.

Estos algoritmos permiten examinar semánticamente las frases, mensajes o noticias

capturadas y detectar dónde se habla de la marca que queremos estudiar y sobretodo el cómo se habla.

Mediante el análisis de sentimiento sabemos si una mención es positiva (se habla bien de la marca), negativa (se habla mal de la marca) o neutra (el mensaje no se puede considerar ni positivo ni negativo para la marca). Para llevarlo a cabo contamos con una serie de ontologías de sentimiento que se basan en un estudio científico realizado a más de 20000 personas y adaptado en el cual se hace un ranking de las palabras con una puntuación de -10 a 10, siendo -10 muy negativo y 10 muy positivo. Esta ontología cuenta con mas de 15.000 palabras rankeadas y etiquetadas con una puntuación para cada sentimiento. Aun así, está ontología se adapta y personaliza a las necesidades de cada cliente y sector con métodos de machine learning.

Para ir más allá del análisis de sentimiento, Brand Rain contempla en sus cálculos otros

análisis semánticos que tienen en cuenta el universo corporativo de cada marca. Este universo abarca conceptos que entendemos como dimensiones de la reputación de marca como el desempeño financiero, el respeto al medio ambiente, la calidad de los productos, la inversión en I+D+i o la responsabilidad social corporativa. (Figura 2). Las dimensiones tienen alto valor para objetivar la reputación de marca y intervienen en el cálculo final.

El análisis se realiza utilizando técnicas de análisis de sentimiento, PLN, machine learning y ontologías de sentimiento.

4.2 Detección de influenciadores y redes complejas

La detección de la red de influenciadores de una marca se realiza mediante tecnología de redes complejas. En primer lugar, se reconocen los autores que publican contenido sobre las marcas y todos sus alias posibles. Esto se hace mediante técnicas de boosting y reconocimiento de patrones en texto.

Una vez identificados los autores y medios más destacados para la marca, el software aporta datos sobre el autor y su capacidad de impactar al público.

En un futuro, está previsto que ofrezca también datos sobre el grado de peligrosidad del autor dependiendo de cómo habla de la marca y que llegue a dibujar toda la red de personas que influyen la marca hasta dar con el influenciador clave.

De esta forma, analizando la actividad del influenciador clave y el sentimiento de los mensajes que emite se podrá llegar a detectar el foco de una crisis de reputación y actuar en consecuencia para evitarla.

5 Conclusiones y valoraciones

La investigación desarrollada en PLN, representan una ventaja competitiva para Brand Rain. Nuestros clientes se han visto beneficiados al poder utilizar una herramienta cada vez más inteligente, y adecuada a sus necesidades, permitiendo un análisis de reputación y de marca con características muy avanzadas, que reúne los beneficios de la IA y de la aplicación de PLN.

Sin embargo, este es un tema retador y existen tópicos como la ironía y aspectos más complejos relacionados con la lingüística, que

nos impulsa a querer investigar cada vez más para la mejora continua de nuestros sistemas.

En definitiva, son muchos los beneficios obtenidos tras la aplicación del PLN en nuestros sistemas. Sin duda los más destacados son aquellos que nos han permitido abrir nuevas vías de negocio. Entre éstos está el análisis de sentimiento, con el que hemos podido analizar aspectos cualitativos de una marca, como el tono contextual del que está rodeada, o la reputación de la misma tal y como la entenderíamos las personas.

Nuestra herramienta analiza diariamente más de 600.000 menciones y hace un seguimiento exhaustivo de más de 1.000 marcas, la aplicación del PLN ha supuesto poder realizar análisis semántico, que junto con los algoritmos de big-data representan una gran mejora para empresas como la nuestra que gestionan un enorme volumen de datos.

Bibliografía

- H. Cunningham, V. Tablan, A. Roberts, K. Bontcheva. 2013. Getting More Out of Biomedical Documents with GATE's Full Lifecycle Open Source Text Analytics. *PLoS Comput Biol*, 9(2):e1002854. doi: 10.1371/journal.pcbi.100284.
- Lluís Padró and Evgeny Stanilovsky. 2012. FreeLing 3.0: Towards Wider Multilinguality. En *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*. ELRA. (Istanbul, Turkey).
- Sonal Gupta and Christopher D. Manning 2014. SPIED: Stanford Pattern-based Information Extraction and Diagnostics. En *Proceedings of the ACL 2014 Workshop on Interactive Language Learning, Visualization, and Interfaces ACL-ILLVI*.