

El Big Data al servicio de la gestión pública

David Nieto Antón

Responsable del Área Healthcare en la Dirección de Aplicaciones.

En un mundo conectado, el volumen de datos que se producen, gestionan y almacenan crece de forma exponencial, y en la misma proporción, las necesidades de obtener información relevante y significativa. Los beneficios de una explotación adecuada de estos datos mediante la aplicación de la tecnología Big Data son múltiples, y la información que se puede llegar a obtener extremadamente valiosa y reveladora.

Las Administraciones Públicas no son ajenas a esta realidad y no son pocas las que son conscientes de la importancia y necesidad de explotar de manera eficiente la ingente cantidad de datos que atesoran (fondo documental, en información sanidad, en economía y hacienda, etc)

De forma específica, la Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información está apostando firmemente por estas tecnologías no sólo realizando jornadas específicas y destinando ayudas a proyectos de I+D en el ámbito temático *Soluciones TIC para la empresa - Soluciones para la ayuda a la toma de decisiones de negocio. Internet v3: Big Data. Convertir el problema del “Big Data” en una ventaja competitiva para las organizaciones*, sino desarrollando un proyecto para la aplicación de “Big Data” en sus propios procesos de negocio.

La SETSI vio una clara oportunidad de aplicación en el ámbito de las tareas de evaluación de las solicitudes de ayuda dirigidas a incentivar la puesta en marcha de proyectos de alto potencial y que permitan incrementar la competitividad de la industria TIC española. El escenario de la evaluación de ayudas supone un elevado volumen de documentación no estructurada apoyada por un conjunto limitado de datos estructurados, que debe ser revisada y contrastada por un conjunto de evaluadores, por lo que la aplicación de técnicas Big Data se identificó como un excelente mecanismo para dar soporte y facilitar la labor de estos profesionales.

SATEC fue seleccionada a finales de 2014 para la implementación del sistema de información para el apoyo a la evaluación de ayudas que debía estar basado en el uso de herramientas y tecnologías Open Source, de modo que permitiese realizar sobre el conjunto de datos disponible: procesos de extracción; transformación y carga (ETL) de información, procesamiento de lenguaje natural, incluyendo tokenizado, lematizado y reconocimiento de entidades, análisis de tópicos y análisis semántico, cálculos de similitud entre documentos basados en dichos análisis, descarga de información de sitios web, búsqueda textual y facetada sobre el corpus documental, etc.

De la confianza y colaboración entre SETSI y SATEC, se está finalizando la implementación de la herramienta que facilita:

- El análisis general de las temáticas sobre el conjunto completo de las ayudas solicitadas, incluyendo qué temáticas se tratan y en qué número son tratadas, para una planificación más eficiente de la asignación de evaluadores.
- Una visión específica de las temáticas que trata cada una de las solicitudes de forma individual
- El análisis de las jerarquías de temas que se han detectado en dichas solicitudes

- Una presentación de la evolución temporal de las temáticas detectadas en las solicitudes de ayudas de las distintas convocatorias.
- Ayuda en la determinación de las combinaciones temáticas que suelen darse más frecuentemente en la documentación presentada.
- Funcionalidades de búsqueda para ayudar en la identificación de documentos relevantes a la hora de realizar la evaluación de un proyecto innovador, incluyendo la búsqueda textual, el filtrado por metadatos específicos (años, empresas, CNAE, provincias...), y la posibilidad de realizar búsquedas temáticas e identificar documentos similares.

Los resultados obtenidos hasta el momento están alineados con los objetivos marcados con la SETSI y han servido para identificar nuevas posibilidades de aplicación como pueden ser:

- La identificación de qué empresas que están trabajando en determinadas temáticas innovadoras.
- La posibilidad de ampliar el conocimiento del I+D español relacionando las solicitudes de ayuda con artículos científicos y patentes presentadas.
- El impacto que tienen los proyectos financiados sobre el I+D y la competitividad
- La extensión de los servicios ofrecidos por la herramienta a otros organismos que conceden ayudas dentro de la AGE y las Administraciones de las CCAA.

En consecuencia, podemos destacar que la aplicación de Big Data en la gestión pública es una realidad que se encuentra avalada por el impulso que estas tecnologías están teniendo desde todos los niveles de la Administración, existiendo un sinfín de posibilidades de aplicación y otras muchas que se derivarán de las mismas. El aprovechamiento de estas tecnologías debe dar sus frutos en la mejora del conocimiento de la ingente cantidad de información, la optimización de los procesos, el apoyo en la elaboración de políticas públicas, el control del gasto, la prevención del fraude, la colaboración entre administraciones, etc.