

## TECNOLOGÍAS DEL LENGUAJE HUMANO PARA LA RECUPERACIÓN, CLASIFICACIÓN, EXTRACCIÓN, ANÁLISIS Y GENERACIÓN DE INFORMACIÓN DOCUMENTAL Y WEB SOCIAL

### DESCRIPCIÓN DE LA INVENCION

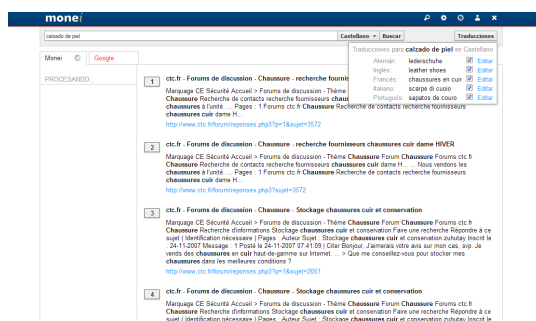
En la actualidad, el acceso a la información se ha convertido en una herramienta básica para cualquier tipo de organización, negocio e incluso para la vida doméstica. Cada vez más, la información generada en el entorno debe ser procesada para transformarse en el conocimiento que permita tomar las decisiones de futuro.

¿Qué demandan nuestros clientes potenciales? ¿Qué nuevas tendencias hay en el mercado?, ¿Qué hacen o dejan de hacer nuestros competidores? ¿Qué opinan otros usuarios sobre este hotel, lugar, producto, etc. que voy a contratar? ¿Quién y cómo ha solucionado previamente este problema que me ha surgido? ¿Quién está ofreciendo este ítem que yo demando? Son preguntas que deben realizarse día a día para poder alimentar a los actuales sistemas de inteligencia de negocio, quienes guiarán el proceso de toma de decisiones. Sin embargo, para que estos sistemas sean realmente efectivos, se requiere de respuestas precisas, es decir, información especializada al dominio, clasificada de acuerdo a la necesidad del negocio, y automatizada para garantizar el flujo continuo y actualizado.

Por otra parte, la mayor parte de las necesidades informativas ya han sido desarrolladas en algún momento, y se encuentran a disposición en algún lugar del mundo esperando a ser encontradas. Archivos documentales, páginas web, bases de datos online, redes sociales, blogs, prensa online, etc. se actualizan día a día proporcionando esta información que precisamos. Además, gracias a los medios de comunicación actuales y la digitalización de los escritos, esta tarea debería ser sencilla de alcanzar. Sin embargo, el ingente volumen de información accesible desde los actuales buscadores de información hace que el acceso preciso a información especializada sea extremadamente pobre.

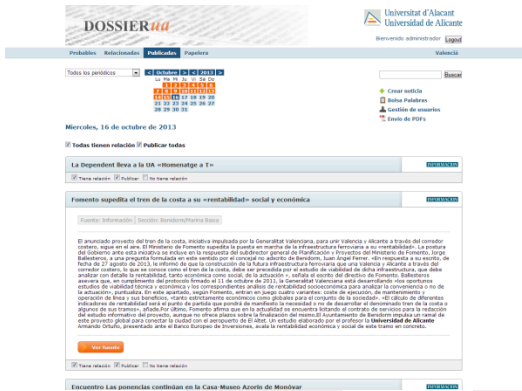
El grupo de investigación de Procesamiento del lenguaje natural y sistemas de información de la Universidad de Alicante ha desarrollado diversas técnicas para la conversión de la información documental en conocimiento preciso. La clave consiste en localizar los aspectos relevantes de la información descartando todo lo superfluo (tanto en web como en bases de datos de textos digitalizados), para posteriormente analizar, clasificar, y descubrir el conocimiento oculto en la documentación, y así generar bases de datos estructuradas y especializadas que alimentarán a nuevos procesos de vigilancia o toma de decisiones, o bien producir nuevos documentos con propósitos específicos. También se utiliza el análisis de sentimientos y la minería de opiniones sobre todo a raíz del crecimiento de las redes sociales y la web 2.0

Mediante las tecnologías comentadas, el grupo ha desarrollado los siguientes productos:



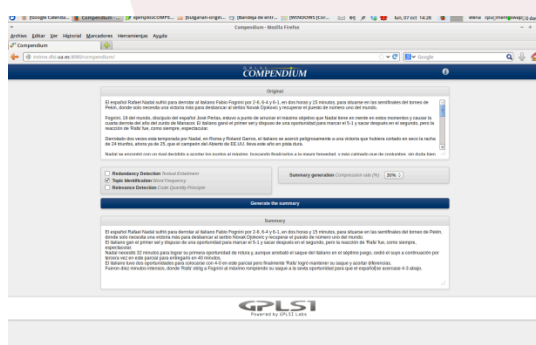
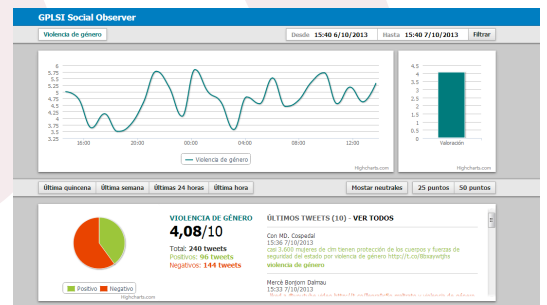
**GPLSI MONEI** es un metabuscador web inteligente que se adapta a las necesidades de la organización. Monei permite activar búsquedas multilingües simultáneas, restringir el ámbito de búsqueda a los dominios que se consideren realmente relevantes, acceder a páginas ocultas y bases de datos, clasificar y agrupar la información previamente a su visualización, y definir un ranking de resultados acorde con las necesidades específicas del negocio.

## TECNOLOGÍAS DEL LENGUAJE HUMANO PARA LA RECUPERACIÓN, CLASIFICACIÓN, EXTRACCIÓN, ANÁLISIS Y GENERACIÓN DE INFORMACIÓN DOCUMENTAL Y WEB SOCIAL



**GPLSI DOSSIER** es una aplicación que permite clasificar de forma automática la información procedente de prensa, páginas web, boletines oficiales e, incluso, redes sociales, extrayendo únicamente la información relevante al negocio y almacenándola en bases de datos estructuradas para su posterior procesamiento por sistemas de inteligencia de negocio.

**GPLSI SOCIAL OBSERVER** es una aplicación que recupera tweets de la red social de Twitter sobre un tema en concreto, una marca, un producto, una persona, etc., y de forma automática valora las opiniones expresadas por los usuarios en los mensajes, para posteriormente realizar un análisis de comportamientos y tendencias manejando comparativas, intervalos de tiempo o zonas geográficas.



**GPLSI COMPENDIUM** es un sistema de generación de resúmenes de textos automático capaz de producir resúmenes genéricos e informativos. Compendium permite obtener las ideas más relevantes de un documento (noticia, informe, página Web, etc.) sin necesidad de tener que leer todo el texto. Además, Compendium permite incorporar técnicas avanzadas de generación del lenguaje enriqueciendo el resumen con información adicional como la incorporación de información multimedia, la sustitución de términos por sus sinónimos, o la sustitución de lenguaje formal por informal y viceversa.

### SECTORES DE APLICACIÓN EMPRESARIAL

Las aplicaciones pueden darse en cualquier sector en cuya actividad se trabaje con grandes cantidades de información, que sea necesario procesar y de la cual sea necesario extraer información y datos bajo demanda. Entre otros:

- Gestión de información administrativa, de bases de datos y texto en general: recuperación, extracción y clasificación.

## **TECNOLOGÍAS DEL LENGUAJE HUMANO PARA LA RECUPERACIÓN, CLASIFICACIÓN, EXTRACCIÓN, ANÁLISIS Y GENERACIÓN DE INFORMACIÓN DOCUMENTAL Y WEB SOCIAL**

- Rastreo y seguimiento de información publicada en la WEB.
- Administración Pública.
- Periodismo y gestión de noticias.
- Marketing y elaboración de estudios socioeconómicos.
- Desarrollo de negocio. Business Intelligence.

### **VENTAJAS TÉCNICAS Y BENEFICIOS EMPRESARIALES**

- Adaptación de las aplicaciones desarrolladas a cualquier sector o área de actividad.
- Automatización de todas las necesidades de búsqueda y análisis de información de negocio.
- Posibilidad de usar las distintas aplicaciones por separado o integradas en una única aplicación.
- Construcción de herramientas para adopción de decisiones empresariales y vigilancia tecnológica.

### **ESTADO DE DESARROLLO DE LA TECNOLOGÍA**

Los productos se encuentran ya desarrollados y se podrían adaptar a las necesidades de las empresas con relativa rapidez.

### **DERECHOS DE PROPIEDAD INDUSTRIAL**

Los productos desarrollados se encuentran protegidos como propiedad intelectual bajo derechos de autor.

### **COLABORACIÓN BUCADA**

El grupo busca empresas o entidades interesadas en su know-how para desarrollar aplicaciones concretas a las necesidades y entornos de los clientes. Asimismo, busca socios para llevar a cabo otras aplicaciones y/o proyectos de investigación conjuntos.

Los acuerdos de cooperación tecnológica propuestos serían:

- Acuerdos de licencia: licencia de know-how.
- Acuerdos comerciales con asistencia técnica.
- Proyectos conjuntos de I+D

### **CONTACTO COMERCIAL**

Víctor Manuel Pérez Lozano  
SGITT-OTRI (Universidad de Alicante)  
Teléfono: +34 96 590 3467  
E-Mail: [otri@ua.es](mailto:otri@ua.es)