



TIPO DE RESULTADO I+D

Nueva tecnología
Nuevo producto
Nuevo servicio
(Nuevo conocimiento o capacidad)



GRADO DE MADUREZ COMERCIAL

Modelo o idea conceptual
Prueba de concepto (diseño)
(Validado en un entorno controlado)
Validado en un entorno real
Implantado entorno real con éxito



PROTECCIÓN

(No aplica)
Patente
Software
Know how
Modelo de utilidad

Descripción de la solución. Problema que resuelve

En las últimas décadas se ha ido implementando la digitalización y el almacenamiento virtual de materiales como artículos y documentos en repositorios y bases de datos. Este gran volumen de información digital ha creado la necesidad de desarrollar herramientas de búsqueda que permitan encontrar la información de manera precisa y rápida. Aunque en la actualidad existen multitud de motores de búsqueda disponibles, la gran mayoría están limitados a búsquedas rígidas y no disponen de funcionalidades suficientes que permitan dar libertad al usuario. Con el objetivo de ofrecer más posibilidades a los usuarios y extraer un mayor valor de las bases de datos documentales, investigadores del Instituto Universitario de Análisis y Aplicaciones Textuales (IATEXT) perteneciente a la Universidad de las Palmas de Gran Canaria (ULPGC) han desarrollado un motor de búsqueda capaz de procesar corpus de documentos digitalizados para realizar búsquedas textuales partiendo de frases o palabras y empleando diversos filtros.



Los investigadores han desarrollado un motor de búsqueda que actúa sobre un conjunto de documentos (denominado corpus), el cual se procesa para entrenar al motor para que puede realizar, posteriormente, las búsquedas y la contextualización de las respuestas. De este modo, se crean bases de datos de documentos digitales sobre las cuales actúa el motor desarrollado.

Ámbitos de aplicación comercial

Creación de motores de búsqueda avanzados en grandes bases de datos de documentos textuales, tales como boletines oficiales, registros mercantiles. En general, de interés para organismos con bases de datos y fondos documentales potentes con necesidad de gestión.

Sistemas informáticos para empresas de gestión documental.

Oportunidades de mercado

La tecnología desarrollada proporciona los siguientes beneficios:

- Es un software versátil; dado que emplea un léxico amplio que permite realizar búsquedas empleando diversos criterios de discriminación.
- Es una tecnología adaptable y dinámica, de manera que es posible ir incorporando nuevos textos al corpus y además implementar nuevas funcionalidades que mejoren el motor.
- Es aplicable a cualquier corpus textual, como se ha comprobado en diferentes buscadores ya desarrollados por los investigadores.

Características y funcionalidades del motor:

- Más de 4 millones de términos únicos.
- Rapidez a la hora de procesar nuevos documentos, alcanzando velocidades de hasta 10.000 palabras por minuto.
- Permite hacer búsquedas con inteligencia lingüística.
- Adaptable a las necesidades y el tipo de búsqueda que necesite el usuario.
- Dispone de algoritmos adicionales para hacer búsquedas más amplias, incluyendo formas con uno o varios pronombres enclíticos y prefijos.

Ventaja competitiva y aspectos innovadores

El software emplea un servicio de lematización que también ha sido desarrollado por los investigadores, en el que cada término único representa un lema. Su léxico cuenta con más de 4 millones de palabras distintas y el servicio de lematización emplea este léxico para reconocer con exactitud las palabras en un texto en español. Además, el servicio cuenta con dos algoritmos adicionales: uno de ellos es capaz de detectar hasta tres pronombres enclíticos y el otro puede detectar prefijos en cualquiera de las palabras incluidas en el léxico.

Asimismo, también se han implementado otras funcionalidades que emplean la inteligencia lingüística y que incrementan el potencial del software desarrollado: búsqueda por lemas, por categoría gramatical, con comodines y con distancia entre palabras; así como filtros de contextualización propios de cada corpus documental. Además, es posible adaptar estas funcionalidades a las necesidades de cada cliente.

AUTORÍA

AUTORÍA
Francisco Carreras Riudavets;
Gustavo Rodríguez Rodríguez

CONTACTO

Oficina Transferencia de
Resultados de Investigación
(OTRI) otri@fpct.ulpgc.es
928 45 99 56 / 43
<https://otri.ulpgc.es/>