

## NOTICIA AMPLIADA



## Buscar información Internet

Universidad de Jaén

La Universidad investiga nuevas técnicas que hagan más flexible la búsqueda y recuperación de información en Internet.

21/7/2006



### Buscar información Internet

La Universidad de Jaén participa en el macroproyecto *"Recuperación de Respuestas en Documentos Digitalizados"*, encargado por el Ministerio de Ciencia y Tecnología, con el que se pretende desarrollar nuevas técnicas

y herramientas que sean capaces, no sólo de recuperar información específica en documentos digitales, sino que puedan localizar respuestas a preguntas concretas en grandes volúmenes de información digitalizada, según informa Andalucía Investiga.

Se trata de un proyecto coordinado en el que participan, junto con la Universidad de Jaén (UJA), la Universidad Nacional a Distancia (UNED), la Universidad Politécnica de Valencia (UPV) y la Universidad de Alicante (UA).

El estudio se plantea debido a la creciente información digitalizada y no estructurada que se ha producido en los últimos veinte años, y que hace necesaria la investigación en sistemas de información textual que faciliten la localización, acceso y tratamiento de toda esta ingente cantidad de datos.

A esto hay que añadir el crecimiento exponencial de información digital en diferentes lenguas, lo cual requiere sistemas que permitan buscar o recuperar información solicitada en el idioma no sólo origen (es decir en el que se formula la pregunta) sino también en el idioma destino (es decir en el que está escrita la pregunta) en grandes volúmenes de información digitalizada.

### Sistemas multilingües

El grupo de Jaén ha centrado sus esfuerzos en este punto, a través de uno de los cuatro subproyectos que componen este macroproyecto coordinado.

Con el subproyecto *'Desarrollo de sistemas de recuperación de información multilingües: técnicas y herramientas'*, el grupo de *'Sistemas Inteligentes de Acceso a la Información'* de la Universidad de Jaén, trata de desarrollar un sistema de búsqueda de información multilingüe.

Las premisas marcadas en el diseño son básicamente tres: en primer lugar, el sistema debe ser escalable a, al menos ocho idiomas: español, inglés, francés, alemán, italiano, danés, finlandés y sueco. En segundo lugar, debe soportar modelos de recuperación de información basados en documentos, incluso, en imágenes. Por último, debe permitir una fácil integración en un sistema de preguntas y respuestas multilingüe.

Todos estos sistemas de búsqueda y recuperación de información multilingüe, desarrollados con la puesta en marcha de este ambicioso proyecto, se están evaluando a nivel internacional en la campaña internacional CLEF (Cross- Language Evaluation Forum, siendo éste otro de los objetivos del macroproyecto.

Este forum internacional se centra en el desarrollo de infraestructura necesaria para la experimentación y evaluación de sistemas de recuperación de información que trabajen sobre las lenguas europeas en contextos monolingües y translingües, y en la creación de conjuntos de datos reutilizables por los sistemas desarrollados.

En definitiva, con este trabajo se pretende "aportar técnicas y herramientas que faciliten la accesibilidad a la información, y, sobre todo, que mejoren la calidad de la información digitalizada, incluso la existente en otros idiomas", según indica L. Alfonso Ureña, investigador responsable del subproyecto de Jaén.

En cuanto a la aplicabilidad de los resultados, L. Alfonso Ureña resalta que se limitan a una posible transferencia tecnológica de las técnicas y recursos desarrollados, para la creación de un buscador en Internet que pudiera recuperar en lenguaje natural información en distintos idiomas

Universia

