



UN NUEVO MODELO PARA LA GESTIÓN DOCUMENTAL

23 de Julio de 2008

Un equipo científico de la Universidad de Granada, dirigido por María Amparo Vila Miranda, desarrolla un sistema de gestión semántica para datos documentales utilizando *soft computing*, con aplicación al Control de Autoridades, lo que implica el desarrollo y la adaptación de modelos de representación y técnicas de extracción de conocimiento, así como de filtrado de datos. La Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía ha subvencionado este proyecto con más de 140.000 euros.

Alicia Barea Lara

La gestión semántica de datos documentales se caracteriza porque los formalismos de representación y recuperación de la información están basados en el uso de entidades estructuradas con propiedades complejas, relacionadas entre sí y que representan el significado del documento (ontologías). En definitiva, la que genéricamente se propone en el paradigma de la 'Web Semántica'.

Es un problema conocido que los datos documentales están afectados de incertidumbre y vaguedad, que pueden tener dos orígenes: la inherente a los propios documentos y la producida por una imperfecta representación. Esta vaguedad debe ser contemplada y formalizada si se quiere llegar a una gestión que pueda calificarse plenamente de semántica. La Computación Flexible y, más concretamente, la Lógica Difusa son herramientas introducidas en el ámbito de la Inteligencia Artificial para el manejo de tales problemas; ambos se usarán en el desarrollo de este proyecto.

Con este planteamiento, los elementos que caracterizan el proyecto y que delimitan las áreas donde se encuadra son tres. En primer lugar, la representación estructurada de la información documental, utilizando *ontologías* para representar su esquema conceptual y *lógica difusa* para poder contemplar vaguedad en los datos. En segundo lugar, las técnicas de *minería de textos* utilizadas para establecer tanto los procesos de filtrado de la información como la instanciación de documentos. Por último, los problemas de *control de autoridades*, que se intentan automatizar utilizando el conocimiento y las representaciones antes establecidas.

La hipótesis de trabajo del proyecto es que, dada una base de datos bibliográfica, es posible mejorar el uso de la misma empleando una representación estructurada con un esquema conceptual semántico basado en dos tipos de ontologías: una ontología general del documento, y una ontología de temas y contenidos. Esta representación permitirá, además, la generación automática de registros de autoridades que podrán ser refinados por los expertos documentalistas. Utilizando esta representación es posible instanciar una colección previa de documentos mediante un proceso de limpieza y filtrado, utilizando reglas previamente obtenidas por medio de técnicas de minería de datos.



Miembros del grupo de investigación

Objetivos del proyecto

El objetivo general del estudio es el desarrollo de un sistema de gestión de datos documentales que contemple una representación de datos en la que se incluya la semántica de los mismos.

Este sistema introducirá grandes mejoras en dos tareas clásicas dentro del campo de la documentación, ya que permitirá, por un lado, una recuperación de documentos más ajustada a los requerimientos de los usuarios, de acuerdo con los principios de la *Web Semántica*, es decir, evitando la sobre-información y permitiendo que la información sea compartida; y por otro, automatizar en gran medida el control de autoridades locales, entendiendo éste como el proceso de garantizar que ciertos datos (entradas) del documento, referentes a elementos tales como autor, lugar de publicación, etc., son correctos y se corresponden con una entidad establecida (autoridad). Esto evitará gran parte de esta costosa tarea que habitualmente es realizada por el documentalista.

El logro de este objetivo implica el desarrollo y la adaptación de modelos de representación y técnicas de extracción de conocimiento (así como de filtrado de datos) y la aplicación de dichos modelos y técnicas tanto a la generación de una representación del documento mucho más rica y estructurada que la que actualmente se maneja, como al proceso de control de autoridades.

Tanto la representación de la información documental mediante ontologías que reflejen la estructura interna de los documentos, como la extracción de conocimiento a partir de información textual, han recibido mucha atención, constituyendo esta última un área de trabajo muy amplia dentro de la *minería de datos (text mining)* ampliamente desarrollada en los últimos cinco años.

Innovaciones

La novedad de este estudio radica en el uso conjunto de las técnicas de minería de textos en datos semiestructurados y de una representación basada en una ontología del documento; en la introducción del paradigma de *soft computing* en el proceso de estructuración del conocimiento y en algunas técnicas y modelos de *minería de textos*, lo que proporcionará una gran flexibilidad en la representación unido a una gran potencia de computación; y por el desarrollo de procesos de extracción de conocimiento y toma de decisiones para automatizar, dentro de lo posible, el control de autoridades locales, lo que tradicionalmente es una tarea hecha por expertos humanos.

Más información:

María Amparo Vila Miranda
Departamento de Ciencias de la Computación e Inteligencia Artificial
Universidad de Granada

E-mail: vila@decsai.ugr.es

[« VOLVER](#)
[\[IMPRIMIR\]](#)
[\[ENVIAR NOTICIA\]](#)
[\[MÁS NOTICIAS\]](#)
[\[HEMEROTECA\]](#)


Este portal se publica bajo una [licencia de Creative Commons](#).

Area25
Diseño web

[Quiénes somos](#) : [Contáctanos](#) : [Boletín electrónico](#) : [Innova Press](#) : [Andalucía Innova](#) : [Mapa web](#)