

Mejorando el acceso y visibilidad de la información multilingüe en red para la Comunidad de Madrid (MAVIR-CM)

NLP&IR UNED
GSI-UPM
BD-UC3M

PLN@UAM
GSI-UEM
CINDOC

Resumen: En la sociedad del conocimiento, Internet constituye la vía fundamental de acceso y difusión de información, y las lenguas proporcionan el soporte para ese proceso de comunicación. Para la Comunidad de Madrid, que pretende hacer del conocimiento la base de su desarrollo económico, social y humano, es imprescindible desarrollar y aprovechar la capacidad de impacto de ese canal. Para ello, además de buenas infraestructuras, productos y servicios, es preciso: 1º facilitar el acceso de los individuos y de las organizaciones a información en la web en un contexto de economía globalizada y multilingüe, y 2º facilitar la accesibilidad a información propia por parte de terceros. En este doble juego, la tecnología lingüística juega un papel fundamental: 1º como base de los sistemas de acceso a información distribuida, multilingüe y multimedia, 2º como catalizador para una presencia mayor del español en la red.

Palabras clave: Recuperación de información; multilingüismo; web semántica; sistemas de respuesta automática a preguntas; portales de información; buscadores especializados, usabilidad de la web, accesibilidad, cibermetría.

Abstract: Internet is the main tool for accessing and disseminating information in the knowledge society, and human languages provide the basic support for such communication process. The Region of Madrid intends making knowledge the basis of its economic, social and human development. For this purpose, it is necessary developing and taking advantage of the impact of such communication channel. High level infrastructures, products and services are, obviously, needed, but also: 1- Making Internet accessibility easier for individuals and organizations, in a context of a global and multilingual economy, 2- Making easier the accessibility of contents produced here by others. In this game, linguistic technologies play a fundamental role: 1- As the basis for systems providing access to distributed multilingual and multimedia contents, 2- As a catalyst for a greater importance of the Spanish language in the net.

Keywords: Information retrieval, multilingualism, semantic web, automatic question answering, information portals, specialized search engines, web usability, accessibility and cybermetrics.

1 Descripción general

MAVIR forma parte del Programa de Actividades de I+D en Tecnologías 2005-2008 de la CAM. Para la Comunidad de Madrid es especialmente importante ofrecer la mayor visibilidad de su oferta cultural, científico-técnica y de negocios, y caracterizar cuantitativamente el impacto de estas actividades con la ayuda de las técnicas de la cibermetría. Por estos motivos, un equipo interdisciplinar constituido por científicos, técnicos, lingüistas y documentalistas se han unido para desarrollar un esfuerzo integrador en los terrenos de la investigación, la formación y la transferencia de tecnología a fin de mejorar el

acceso y la visibilidad de la información en la red, especialmente en nuestra Comunidad.

La red temática propuesta incluye 25 doctores organizados en 6 grupos de investigación (UNED, UAM, UC3M, UEM, UPM, CINDOC) de la CAM que desde un perspectiva pluridisciplinar se complementan en varias dimensiones: académica vs. profesional, investigación vs. servicios, generación de recursos vs. aplicaciones.

Además de los 6 grupos anteriores hay otros grupos asociados al proyecto: Grupo de Información en PLN y Sistemas de Información de la U. de Alicante, Grupo de PLN de la UPC, Grupo de Sistemas Inteligentes y Minería de Datos de la UHU, ISTI/CNR, Dpt.

of Information Studies of the Sheffield University: Paul Clough, Collage of Information Studies and Institute for Advanced Computer Studies of the Maryland University: Douglas Oard.

En MAVIR participan, además, las siguientes entidades colaboradoras: RAE, Instituto Cervantes, Hospital de Fuenlabrada, Casa de América, Agencia EFE, DAEDALUS, BITEXT, ISOCO.

2 *Objetivos*

El objetivo general de MAVIR es mejorar el acceso y la visibilidad de la información multilingüe y multimedia en la red. Las líneas de investigación del consorcio se pueden organizar en dos grupos:

1. Tecnologías Lingüísticas: Recuperación de información distribuida, multilingüe y multimedia. Web Semántica. Sistemas automáticos de respuesta a preguntas en lenguaje natural. Clasificación y resumen automático de documentos. Representación y extracción de información lingüística.

2. Portales de información cultural, científica, técnica y empresarial. Bibliotecas y archivos digitales. Diseño centrado en el usuario: usabilidad, accesibilidad y visualización. Indicadores cibernéticos

3 *Plan de trabajo*

El plan de trabajo se articula en torno a las siguientes tareas:

Trabajo técnico en infraestructuras y recursos. La infraestructura del conjunto del programa constará de una red distribuida de servidores.

Puesta en marcha y seguimiento del programa de formación y movilidad del capital humano. Los grupos cuentan con un número consolidado de doctores con amplia experiencia en formación de investigadores, capaces de garantizar la formación de nuevos ingenieros y doctores. Se pretende realizar una propuesta de un master centrado en contenidos relacionados con “Acceso Inteligente a la Información”, de acuerdo al sistema de créditos europeo ECTS. Dicho master se podría corresponder con un título de postgrado oficial en la titulación de Informática, y se podría encuadrar dentro de “Erasmus Mundus”.

Asimismo, se pretende organizar talleres y seminarios sobre los diferentes temas de trabajo de los equipos.

Presencia y participación en eventos internacionales. Se participará con demostraciones en eventos nacionales e internacionales seleccionados, incluyendo congresos, conferencias y ferias tecnológicas. Los resultados de la investigación se contrastarán en evaluaciones científicas competitivas como CLEF y DUC.

Proyectos conjuntos en convocatorias competitivas. Se elaborarán propuestas de proyectos conjuntas y de manera coordinada en distintos programas nacionales e internacionales de financiación de la I+D.

Casos de uso. Se realizará al menos un trabajo centrado en el dominio biomédico como caso de uso. Se considerará información especializada en este dominio, así como información textual concreta, tratada de forma anónima, de especial relevancia en actividades de investigación y gestión en este ámbito. Se explorarán y evaluarán servicios avanzados tanto en un ámbito académico, de formación e investigación, como en un ámbito profesional operativo concreto.

Asesoramiento. Con el fin de que las tecnologías desarrolladas se conviertan en nuevos productos y servicios, y contribuyan a mejorar los ya existentes, los grupos participantes ofrecerán su asesoramiento a las empresas privadas y otras instituciones que así lo demanden.

Promoción y difusión. La difusión de la red se realizará través de diferentes medios como: (1) Creación de una sede Web, (2) indicadores cibernéticos y (3) elaboración de guías de buenas prácticas.

Metodología y Actividades de seguimiento científico. El modelo de trabajo de la red propuesta, es análogo a las redes de excelencia, instrumento del VI Programa Marco.

4 *IP y Página web del proyecto*

La Investigadora responsable es M^a Felisa Verdejo Mailló (UNED), felisa@lsi.uned.es.

El sitio web del proyecto:
<http://www.mavir.net/>.