



Mesa Redonda: Resultados Científicos del Consorcio MAVIR

IV Jornadas MAVIR

Presentación

- ▶ Ana García Serrano: UNED
- ▶ José Miguel Goñi Menoyo: UPM
- ▶ Paloma Martínez Fernández: UC3M
- ▶ Manuel de Buenaga Rodríguez: UEM
- ▶ Isidro Aguillo Caño: CSIC-CINDOC
- ▶ Antonio Moreno Sandoval: UAM



Guión de la Mesa Redonda

- ▶ I) Presentación de los grupos participantes: perfil y líneas de investigación
- ▶ II) Resultados científicos más relevantes y beneficios obtenidos del consorcio
- ▶ III) Planes futuros en el consorcio MA2VICMR
- ▶ IV) Preguntas del público



UNED:

- ▶ Orígenes y perfil investigador del grupo “NLP&IR”
 - ▶ Creado en 1993 por Prof. Dra. Felisa Verdejo
 - ▶ 20 investigadores, 9 de ellos doctores permanentes en el Departamento de Lenguajes y Sistemas Informáticos en la Escuela Superior de Ingenieros de Informática
- ▶ Líneas de investigación, áreas en las que ha trabajado
 - ▶ **Acceso Inteligente a la Información:** asistentes de búsqueda translingües, sistemas de búsqueda de respuestas, visualización y navegación de resultados, clustering multilingüe de documentos y clasificación automática de páginas web
 - ▶ **Adquisición y Representación de Conocimiento**
- ▶ **Léxico y Gramatical:** bases de datos léxicas, redes semánticas multilingües, tesauros multilingües y minería de la

UPM:

- ▶ Orígenes y perfil investigador del grupo
 - ▶ Dept. de Ingeniería de Sist. Telemáticos (ETSIT)
 - ▶ Dept. de Matemática Aplicada a las T.I. (ETSIT)
 - ▶ Dept. de Inteligencia Artificial (FI)
 - ▶ Dept. de Ingeniería y Arq. Telemáticas (EUITT)
- ▶ Líneas de investigación, áreas en las que ha trabajado
 - ▶ Creación de recursos lingüísticos
 - ▶ Desarrollo de herramientas eficientes
 - ▶ Recuperación de Información
 - ▶ Búsqueda de Respuestas



Grupo de Sistemas Inteligentes

- ▶ Grupo reconocido por la UPM
- ▶ Surge de la fusión de GSI (LIA) e ISYS con la incorporación de otros investigadores
- ▶ <http://www.upm.es/observatorio/vi/index.jsp?pageac=grupo.jsp&idGrupo=265>



Principales líneas de investigación

Desarrollo de herramientas:

- ▶ **Librería TRIELIB:** acceso eficiente a bancos de datos textuales (léxicos, índices).
- ▶ **Herramientas de procesamiento morfológico:** basado en continuaciones e integrado con TRIELIB para tratamiento muy eficiente.
- ▶ **Herramientas y recursos específicos para tareas de IR y QA:** (stemmers, NER, segmentadores, stopwords lists).
- ▶ **IDRA: InDexing and Retrieving Automatically collection of texts**

▶ (<http://sourceforge.net/projects/idraproject/>)

UEM:

Orígenes y perfil investigador del grupo

- ▶ Primeros proyectos de I+D: 1999
- ▶ Interdisciplinar con base Informática:
participación de departamentos
 - Sistemas Informáticos
 - Arquitectura de Computadores y Automática
 - Especialidades Médicas
 - Periodismo y humanidades / Traducción
- ▶ Situación actual:
16 investigadores, 7 de ellos doctores



http://www.esp.uem.es/gsi/

The screenshot shows a web browser window displaying the homepage of the GSI (Grupo de Investigación en Sistemas Inteligentes) website. The browser's address bar shows the URL <http://www.esp.uem.es/gsi/>. The website header features the logo of the Universidad Europea de Madrid and the text "Grupo de investigación en Sistemas Inteligentes". Below the header, there is a navigation menu with links for "Inicio", "Contactanos", "Noticias", and "Enlaces". A search bar is located in the top right corner, set to "Español". The main content area is divided into several sections: "Menú Principal" with a list of links; "Inicio" with a title "GSI: Grupo de Investigación en Sistemas Inteligentes" and a brief description of the group's research focus; "Recursos GSI" with links to documentation, SVN, and a search tool; "Sindicación" with RSS and OPML feeds; "Noticias" with a list of recent news items; "Acceder" with a login form; and "Usuarios Conectados" showing one user is online. The footer contains a list of research lines.

Universidad Europea de Madrid

Grupo de investigación en Sistemas Inteligentes

Laureate International Universities

Inicio | Contactanos | Noticias | Enlaces

Español | buscar...

Menú Principal

- Inicio
- Noticias
- Elog
- Personal
- Proyectos
- Publicaciones
- Transferencia de Tecnología
- Contacto
- Becas ofertadas

Inicio

GSI: Grupo de Investigación en Sistemas Inteligentes

escrito por Administrator
Tuesday, 27 de November de 2007

El grupo de Investigación de Sistemas Inteligentes de la UEM, centra su actividad investigadora en la actualidad especialmente en Sistemas Inteligentes de Acceso a la Información, en los que se integran elementos de Análisis del Contenido Textual (operaciones como recuperación de información, categorización, generación de resúmenes) con uso de ontologías y elementos específicos de dispositivos móviles (e.g. acceso por contexto). Tiene un centro de interés actual y experiencia significativa en la utilización y evaluación de esta investigación y tecnologías en entornos de salud.

El grupo ha participado en los últimos años en diversos proyectos de investigación financiados relacionados con estos temas, en ocasiones participando en consorcios formados por la universidad, empresas de desarrollo de Software como Gesfor, Bitext, Answare Tech e ISOCO, y clientes o usuarios finales como la Agencia Efe, el Diario ABC y el Hospital de Fuenlabrada, y otras instituciones en el ámbito europeo como el Instituto di Linguistica Computazionale (Pisa, Italia), University of Sheffield (Sheffield, Reino Unido) y Telefónica Investigación y Desarrollo (España)

En la financiación de estos proyectos han participado diferentes entidades como el Ministerio de Educación y Ciencia, el Ministerio de Industria, la Comunidad de Madrid y la Comisión Europea, a través de diferentes programas de I+D.

Recursos GSI

- Documentación GSI
- Servidor SVN
- BuscadorBioV2 (Beta)

Sindicación

- RSS 0.91
- RSS 1.0
- RSS 2.0
- ATOM 0.3
- OPML SHARE IT!

Noticias

- IV Jornadas MAVIR. Acceso a Información en Redes Sociales. 19 y 20 de Noviembre de 2009
- 1st International Workshop on Mining Social Media Programme
- Workshop on Adaptive Multimedia Retrieval (AMR2009)
- BuscadorBioV2
- Beca para el congreso anual de la SEPLN

Acceder

Usuario

Contraseña

Recordarme

[¿Recuperar contraseña?](#)
[¿Quieres registrarte? Hazlo aquí](#)

Usuarios Conectados

Hay 1 invitado en línea

Lineas de Investigación

- Sistemas de análisis del lenguaje natural: generación automática de resúmenes, categorización automática de documentos, desambiguación de términos
- Sistemas basados en recuperación de información, multilingüismo, sistemas de respuesta automática a preguntas, portales de información; buscadores especializados, modelado de

▶ Líneas de investigación

- ▶ Procesamiento del Lenguaje Natural
- ▶ Minería de Datos
- ▶ Computación ubícua
- ▶ Redes Sociales
- ▶ Casos de uso o dominios aplicación: media, seguridad, salud

(foco en acceso a información en Internet)

▶ Proyectos I+D recientes representativos

- ▶ Medical Miner (CICYT): Universidad de Vigo, Universidad de Huelva, UEM
- ▶ Naviga (Eurostars) Answare, Encore, UEM, Technosens, MindAutonomy, IONAN
- ▶ ISSE (PROFIT): ISOCO, Hospital de Fuenlabrada, Universidad Carlos III, UEM



UC3M: Bases de Datos Avanzadas

- ▶ **Orígenes y perfil investigador del grupo**
 - ▶ Creado en 1994, perteneciente al Departamento de Informática. Seis doctores y siete ayudantes
 - ▶ Desde 2000 trabajando en Aplicación de las Tecnologías del Lengua a los SI.
- ▶ **Líneas de investigación:**
 - ▶ Extracción de información en dominio biomédico: farmacología y notas clínicas
 - ▶ Sistemas de Interacción
 - ▶ Búsquedas de respuestas (reconocimiento entidades y tratamiento información temporal)
 - ▶ Integración de recursos léxicos y ontologías en aplicaciones



CSIC: Laboratorio de Cibermetría

- ▶ Orígenes y perfil investigador del grupo
 - ▶ Desde 1992 se trabaja en los estudios métricos de Internet
 - ▶ En 1997 se crea la revista electrónica *Cybermetrics*
 - ▶ Desde 2001 a raíz de la participación en proyectos europeos pasa a llamarse Laboratorio
- ▶ Líneas de investigación, áreas en las que ha trabajado
 - ▶ Desarrollo de indicadores cibernéticos
 - ▶ Evaluación de instituciones académicas
 - ▶ Publicación de los Ranking Webs de Universidades, Centros de Investigación,

Ranking Web de Universidades

Se edita desde **2004**, a partir de trabajos cibernmétricos iniciados en 1996, inspirado por el **Ranking de Shanghai**

Es el Catálogo de Universidades, con sus dominios web, más completo y actualizado: **17.000 universidades** de todo el mundo

Se publican los datos de las **6.000 primeras** universidades, incluyendo las “World-Class” y las que tienen importancia regional para los países en vías de desarrollo

Hay dos ediciones al año (**Enero y Julio**) y se publica conjuntamente con los Rankings de Centros de Investigación (+70.00) y de Repositorios

Los datos web son **indicadores** de la **calidad** de la investigación, docencia y transferencia de conocimiento y del **impacto** global de las actividades de las universidades

The screenshot shows the website's main page with a header 'Ranking Web de Universidades del Mundo' and a date 'Julio 09'. The page is divided into several sections: 'Otros del Ranking' (Other Rankings) with a list of categories like 'Acercas de nosotros', 'Acercas del Ranking', 'Top 6000 Universidades', etc.; 'Acceso abierto a la información académica' (Open Access to Academic Information) with a sub-section 'Nueva edición Julio 2009 del Ranking Web de Universidades' containing introductory text about the ranking's methodology and goals; 'Rankings Web del Cybermetrics Lab' with links to 'Repositorios', 'Hospitales', 'Escuelas de Negocios', and 'Centros de Investigación'; 'Gráficos' (Charts) featuring a pie chart; 'Noticias' (News) with several articles; and a search bar at the bottom left. The footer includes 'CybermetricsLab', a logo, and the text 'CERN-CSE 2009'.

UAM: NLP@UAM

- ▶ Orígenes y perfil investigador del grupo
 - ▶ Dept. de Lingüística: Laboratorio de Lingüística Informática
 - ▶ Dept. de Ingeniería Informática: Enrique Alfonseca
- ▶ Líneas de investigación, áreas en las que ha trabajado
 - ▶ Creación de recursos lingüísticos multilingües:
 - ▶ Compilación de corpus y bases de datos
 - ▶ Anotación morfosintáctica en español y árabe
 - ▶ Recuperación de Información
 - ▶ Búsquedas de respuestas

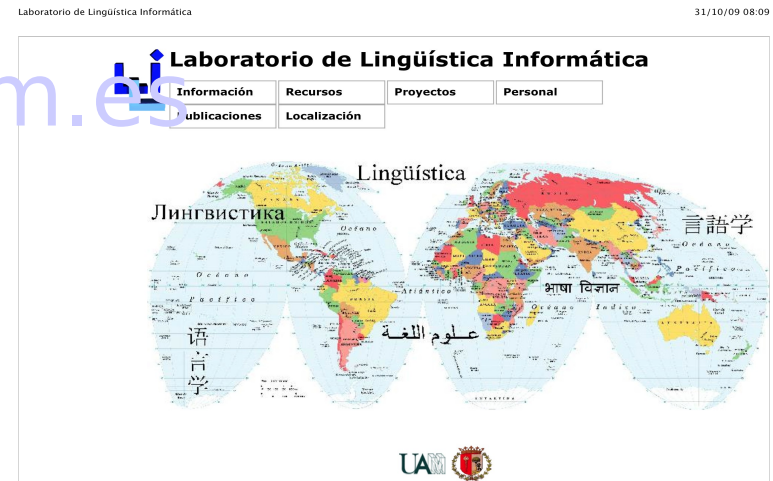
Laboratorio de Lingüística Informática (LLI-UAM)

▶ <http://www.llif.uam.es>

▶ Fundado en 1988

▶ Parte del Dept. de Lingüística

▶ Grupo de investigación reconocido por la UAM



Principales líneas de investigación

Creación de recursos lingüísticos multilingües para:

- ▶ **Sistemas PLN simbólicos:** gramáticas y lexicones para MT y IR.
- ▶ **Sistemas PLN estadísticos:** corpus de entrenamiento (escritos y orales, treebank), bases de datos acústicas para QA
- ▶ **Herramientas:** taggers morfosintácticos (español y árabe), transcriptor fonológico (español), chunker (español)
- ▶ **Aplicaciones para la Lingüística y aprendizaje de lenguas:** “learner corpora”

Resultados científicos más relevantes

- ▶ 1. Proyectos y colaboraciones con empresas
- ▶ 2. Recursos, catálogos y servicios
- ▶ 3. Infraestructura
- ▶ 4. Seminarios y máster
- ▶ 5. Captación de investigadores y estancias
- ▶ 6. Participación y organización de evaluaciones



Proyectos y colaboraciones con empresas

- ▶ 12 PYMES de Consultoría TIC y del Lenguaje asociadas a MAVIR
 - ▶ ALMA, Answare, Bitext, Corex, Daedalus, EFE, INFO I+D, iSOCO, Thera, Vocali, TMT Factory, Wipley
- ▶ **Proyectos de I+D (2005-2009)**
 - ▶ 40 proyectos I+D, 4.5M€, más de 30 organismos participantes
 - ▶ incluyendo empresas asociadas.mavir
 - ▶ Esquemas/Programas de financiación: VIPM, VIIPM, MEC-CICYT, Cenit, Profit, Avanza
 - ▶ MedIEQ: quality labeling of medical web content
 - ▶ Multimatch: multimedia and multilingual search engine on heritage content
 - ▶ ISSE: Interoperabilidad Basada en Semántica para la Sanidad Electrónica
- ▶
 - ▶ NAVIGA Adaptable Platform for elderly people and persons with disability to access the Information Society

Recursos, catálogos y servicios

- ▶ Corpus MAVIR: multimodal y bilingüe (español e inglés)
 - ▶ <http://www.llf.uam.es/ESP/CorpusMavir.html>
 - ▶ Más de 10 horas y 100.000 palabras (las 3 primeras Jornadas)
 - ▶ Catálogo de Recursos Lingüísticos, Sistemas y Software de propósito general en Ingeniería Lingüística, compilado conjuntamente por TIMM y MAVIR (febrero 2009).
 - ▶ Servicios proporcionados
-
- ▶ ▶ Compilación y etiquetado de corpus

Infraestructura

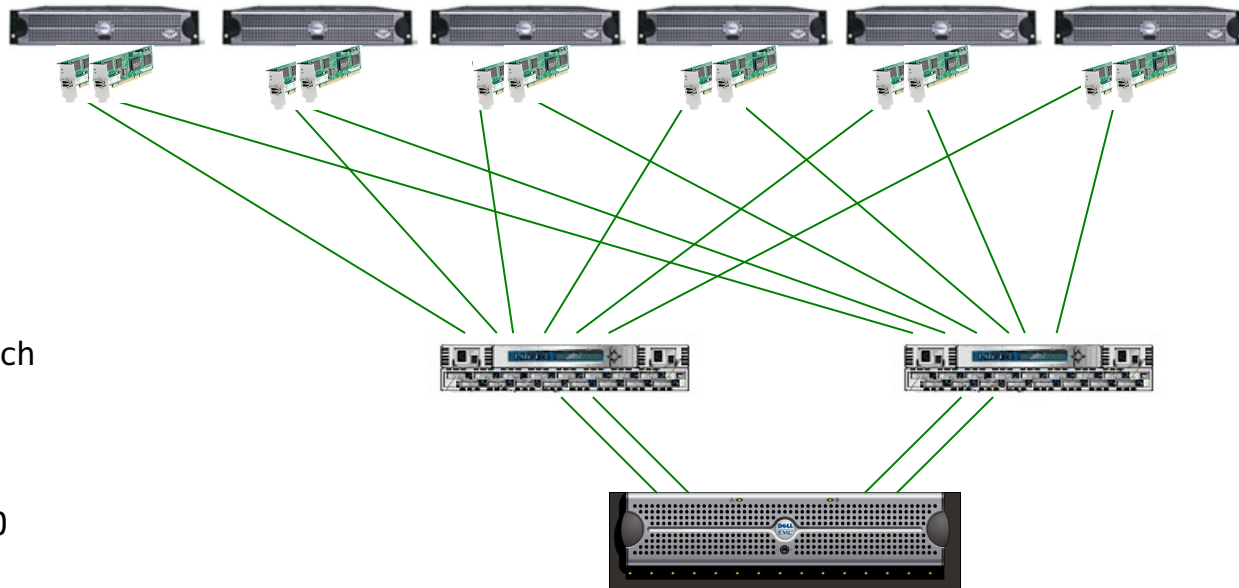
- ▶ Actividad liderada por la UPM y la UNED
 - ▶ Convenio de copropiedad
- ▶ Creación de recursos de computación comunes:
 - ▶ Compartir capacidad de procesamiento
 - ▶ Compartir almacenamiento:
 - ▶ Repositorio de colecciones textuales (ej. CLEF, NTCIR, etc.)
 - ▶ Repositorio de herramientas de uso común
 - ▶ Intercambio de ficheros de resultados, experiencias, etc.
- ▶ Aprobadas unas normas de (buen) uso



Granja (clúster) de servidores

6 x PowerEdge 2950

Red Hat Linux ES



2 x Dell|EMC Switch
MData 4400

Dell|EMC CX300
15x300 GB FC



Formación: Seminarios y máster

- ▶ 18 seminarios (9 con al menos 10h lectivas) cofinanciados por MAVIR y Ayudas a la Movilidad para postgrados de calidad
- ▶ Propuesta de Master en “Tecnologías de la lengua”
 - ▶ Título oficial de 60 ECTS con contenidos sobre “Acceso Inteligente a la Información”
 - ▶ Interuniversitario: UNED, UC3M, UPM + Empresas tecnológicas (estancias, seminarios, dirección de trabajos).
 - ▶ Dos normativas oficiales diferentes (Reales Decretos de 2005 y 2007) ... EEES ... Sucesivos cambios en la gestión de la propuesta...
- ▶ Estrategias universidades divergentes y con calendarios distintos

Captación de investigadores y estancias

- ▶ Captación de personal:
 - ▶ Incorporación doctores: 8 investigadores
 - ▶ Predoctorales: 3 como Investigadores de Apoyo (comunidad de Madrid)
- ▶ Cofinanciación estancias investigadores en otros grupos (grupos asociados MAVIR y otros)
 - ▶ 12 investigadores predoctorales en grupos internacionales y nacionales (3-6 meses) algunas cofinanciadas por Ayudas de Movilidad de Doctorado con mención europea.
 - ▶ 10 investigadores postdoctorales



Participación y organización de evaluaciones

- ▶ CLEF 2006-2009:
 - ▶ Participación en casi todas las tareas (WebCLEF, ImageCLEF, GEOclef, QA-CLEF, iCLEF)
 - ▶ Organización de alguna tarea (QA-CLEF, iCLEF)
 - ▶ Lenguas: español, inglés, francés, italiano, portugués, alemán, holandés, búlgaro, ruso, húngaro, checo
 - ▶ Organización WePS (Web People Search, WWW 2009)
 - ▶ NTCIR 2008: Participación en inglés y japonés
 - ▶ TAC 2009 (Text Analysis Conference)
 - ▶ ACE 2007 (Automatic Content Evaluation).
-
- ▶ DUC 2006 (Document Understanding Conference)

Plan de trabajo en MA2VICMR

▶ Objetivos científicos-tecnológicos:

- ▶ Acceso a la información multilingüe y multimodal
- ▶ Extracción de información, análisis semántico y pragmático
- ▶ Metodologías de evaluación y métricas

▶ Nuevos socios:

- ▶ URJC: Grupo de Algoritmia aplicada a la visión artificial y biometría
- ▶ UPM: Grupo de Tecnología del Habla
- ▶ UAM: NETs (web semántica) y ATVS (tecnología de habla)

▶ Empresas tecnológicas asociadas: Daedalus, iSOCO, BiText

Turno de preguntas



GRACIAS

