

16:55 [Seminario C] PRESENTACIÓN: **Grupo de Gramática del Español**

Uso de información lingüística refinada para a extracción de léxicos bilingües a partir de corpus comparables

PONENTE: Pablo Gamallo

Un dos principais obxectivos de investigación do grupo é mellorar os sistemas de análise automática e de extracción de información fornecendo elaborados modelos lingüísticos. Con esta filosofía, imos describir un sistema de extracción de léxicos bilingües a partir de corpus comparables. Os corpus comparables conteñen textos en diferentes linguas, que non son traducións uns dos outros, mais que versan sobre temáticas semellantes. Centrarémonos en textos de artigos de xornais españois e galegos.

Os léxicos extraídos poden ter variadas aplicacións. Poden servir para mellorar sistemas de tradución automática, de recuperación de información multilingüe, memorias de tradución, ou ben para construír bases terminolóxicas multilingües.

Moderador: Manuel Vilares

17:20 [Seminario C] PRESENTACIÓN: **Telémaco, Información, Documentación y Sistemas**

Sistema de respuestas, sobre texto estructurado, basado en casos

PONENTE: José Rando

Telémaco, fue constituída en 1995 con la misión de "diseñar, desarrollar e implementar sistemas de información, la producción, promoción y distribución de bases de datos de interés general y particular". Desde entonces trabaja en la localización y captura automática de las fuentes de información, la estructuración de los textos, así como en la extracción de la terminología específica apoyándose en tesauros. Sobre esta base se pretende continuar el trabajo de enriquecemento e organización de la información de cara a dar soporte a los tipos de servicios de valor añadido que presta a sus usuarios.

Moderador: Manuel Vilares

17:45 PAUSA CAFÉ

18:10 [Seminario C] PRESENTACIÓN: **Grupo de Investigación sobre Gramática y Léxico**

Léxico, sintaxis y Semántica

PONENTE: José M. García-Miguel

Moderador: Álvaro Barreiro

18:35 [Seminario C] PRESENTACIÓN: **CIS Galicia**

Vixia, Un sistema de Vigilancia basado en una plataforma tecnológica

PONENTE: Xoan Porto

Moderador: Álvaro Barreiro

19:00 CLAUSURA – Guillermo Rojo

Primeiras xornadas empresas-universidade
**Rede Galega de Procesamento da Linguaxe e
Recuperación de Información**

Facultade de Filoloxía, Santiago de Compostela
26 Outubro 2007

PROGRAMA

9:30 RECEPCIÓN DE ASISTENTES E ENTREGA DE DOCUMENTACIÓN

9:50 BENVIDA – Manuel Vilares

10:00 [Aula 16] PRESENTACIÓN: **Imaxin Software**

Imaxin software: 10 anos traballando nas tecnoloxías lingüísticas

PONENTE: José Ramom Pichel

Percorreremos todos os traballos que leva realizado a nosa compañía durante os seus dez anos, no campo do procesamento documental e das tecnoloxías lingüísticas en xeral. Esta traxectoria converteunos na única empresa galega e entre as vinte de Europa, provedora de PLN e localización para Microsoft (Vendor).

Faremos tamén fincapé nos desenvolvementos I+D+i financiados mediante programas sectoriais TIC da Xunta e PROFIT do Ministerio español de Industria, en colaboración con universidades e outras empresas. Dentro destes destacaremos o proxecto de tradución automática de código aberto OpenTrad, expoñendo a nosa experiencia con vistas a futuros proxectos coa Rede PLN e RI das universidades galegas.

Moderador: Pablo Gamallo

10:25 [Aula 16] PRESENTACIÓN: **Grupo Lengua y Sociedad de la Información**

Recuperación de Información inteligente en español

PONENTE: Jesús Vilares

El principal problema al que han de hacer frente los sistemas de Recuperación de Información es el de las limitaciones impuestas por la variación lingüística propia del lenguaje humano: cómo el mismo concepto puede ser expresado de muy diferentes formas. Esta presentación hará un breve recorrido por el trabajo desarrollado en nuestro grupo de investigación acerca de la aplicación de técnicas de Procesamiento del Lenguaje Natural para el tratamiento de este fenómeno en el caso de la Recuperación de Información sobre textos en español.

Moderador: Pablo Gamallo

10:50 [Aula 16] PRESENTACIÓN: **Serindoc**

Servicios para la Información y la Documentación de Archivos y Bibliotecas

PONENTE: Raquel Morano

Moderador: José M. García-Miguel

11:15 [Aula 16] PRESENTACIÓN: **Information Retrieval Lab**

NowOnWeb: Recuperación de noticias en la web con técnicas avanzadas de IR

PONENTE: Javier Parapar

El grupo IRLab - Information Retrieval Lab es un grupo de investigación de la UDC cuya línea de investigación principal es la recuperación de información en la web y en repositorios de documentos. La recuperación eficaz ante las necesidades de los usuarios y la construcción de sistemas eficientes y escalables son los objetivos perseguidos por el grupo.

Como ilustración del trabajo aplicado realizado por el grupo presentamos NowOnWeb (<http://nowonweb.dc.fi.udc.es>), un sistema para la recuperación de noticias en la web con características diferenciadas de los sistemas genéricos que ofrecen Google, Yahoo! y Microsoft y con potencialidad para adaptarse a necesidades específicas de las empresas y administraciones.

Moderador: José M. García-Miguel

11:40 PAUSA CAFÉ

12:20 [Aula 16] PRESENTACIÓN: **Thera, Centre de Llenguatges i Computació**

Un modelo dinámico de Transferencia de Tecnología

PONENTE: Enric Castellón

Expondremos la experiencia de Thera, como Spin Off de la Universitat de Barcelona en un proceso de inversión y desarrollo de I+D+i a medio y largo plazo, la dinámica de comunicación entre la Universidad y en general los generadores de conocimiento y el mercado, y el papel de las SpinOff en este proceso.

Moderador: Margarita Alonso

12:45 [Aula 16] PRESENTACIÓN: **Compiladores y Lenguajes**

Extracción de conocimiento a partir de colecciones documentales

PONENTE: Milagros Fernández

A pesar de su creciente disponibilidad en soporte electrónico, la mayoría del conocimiento humano es todavía sólo accesible en formato de texto. En una aparente contradicción, esta circunstancia, que facilita su comprensión, supone una limitación en el marco del actual proceso de globalización en el acceso a la información.

En este contexto, una de nuestras líneas de trabajo actual hace referencia al proceso de adquisición automática de conocimiento a partir de corpus. Nuestro trabajo combina técnicas de procesamiento del lenguaje natural con la inteligencia artificial, en una propuesta de generación semi-automática de ontologías cuyo desarrollo y validación se ha realizado sobre corpus enciclopédicos de carácter técnico. Las aplicaciones se extienden al campo de la recuperación y extracción automática de información, así como a la búsqueda de respuestas.

Moderador: Margarita Alonso

13:10 [Aula 16] PRESENTACIÓN: **Lingway**

Lingway HR Suite: herramientas avanzadas para el e-empleo

PONENTE: José A. Coch

Lingway es una empresa especializada en motores de búsqueda y text-mining verticales. La principal verticalización es al mundo de los recursos humanos y en especial el *e-empleo*: se trata de Lingway HR Suite, que contiene analizadores automáticos y motores de búsqueda semánticos de currículums y de ofertas de empleo, detectores automáticos de ofertas de empleo discriminatorias, módulos de matching ofertas-currículums y anonizadores de currículums.

Lingway emplea 30 personas, entre las cuales 12 son ingenieros lingüistas especializados en diversas lenguas.

Moderador: Fco. Mario Barcala

13:35 [Aula 16] PRESENTACIÓN: **Grupo de Sistemas Inteligentes**

Detección de novedad en textos y su aplicación para la producción automática de resúmenes orientada a dispositivos móviles

PONENTE: David E. Losada

Una de las líneas de investigación del Grupo de Sistemas Inteligentes es la de recuperación de información. Si bien dentro de este área el grupo ha realizado contribuciones en varias subáreas, esta presentación estará enfocada en la detección de novedad en textos y su aplicación para la producción automática de resúmenes orientada a dispositivos móviles. Estas técnicas tienen gran importancia en la actualidad para procesar la ingente cantidad de documentos disponibles en formato electrónico de modo que se enfoque a los usuarios a las partes novedosas de los textos.

Moderador: Fco. Mario Barcala

14:00 XANTAR – Restaurante Auditorio de Galicia

16:30 [Seminario C] PRESENTACIÓN: **Mister Doc**

Archivos Digitales Mister Doc: Vulgarizar el concepto como plataforma tecnológica de desarrollo

PONENTE: José Filipe Proença

Mister Doc consiste en una entidad creada deliberadamente para intentar llevar a la mayoría de la sociedad de información el concepto de archivos digitales y gestión de documentos, soportados en bases de datos tipificadas por sector de actividad. El resultado de su investigación y desarrollo consiste en una serie de productos que pretende ver utilizados por cualquier entidad, independientemente de su tamaño. Su experiencia proviene de proyectos de más de 25 millones de documentos todos diferentes, capturados y clasificados en cadenas superiores a 60.000/día.

Nuestra investigación y desarrollo está orientada a herramientas de recuperación de información que posibiliten respuestas a consultas, dentro de la superficie demarcada por el máximo de productividad, el máximo cambio de mentalidad y el mínimo cambio en las metodologías probadas.

Moderador: Manuel Vilares