

Transferencia de tecnología:
dominio médico y cultural
(Medicina – Grupo I)

Extracción de Información de los historiales médicos

Manuel de Buenaga (UEM)
Antonio Moreno (UAM)
Enrique Torrejón (Bitext)
Francisco García Lombardía (H Fuenlabrada)
José Luis Martínez (Daédalus)
Paz Iglesias Casarrubios (H Fuenlabrada)

*I Jornadas MAVIR
Tecnologías de la lengua en la WWW: retos y
mercados potenciales
Madrid, 16-17 de Noviembre de 2006*



www.mavir.net

Contenido



1. Definición inicial del grupo de trabajo
2. Esquema de elementos inicial
3. Resultados de la reunión
4. Papel de los participantes
5. Análisis de las fuentes de información
6. Planteamiento de propuesta Profit

1 Definición inicial del grupo de trabajo



- Grupo 1: Médico I. Extracción de Información de los historiales médicos
- Objetivos:
 - Búsqueda de respuestas sobre historiales médicos
 - Compleción de formularios de registro de paciente a partir de su historial médico (necesarios para auditoría)
- Usuarios: Hospital de Fuenlabrada
- Empresas: BITEXT y DAEDALUS
- Universidades: Europea de Madrid, Autónoma de Madrid

2 Participantes y esquema de elementos inicial de reunión



- Ficha de reunión
 - 29 de septiembre de 2006
 - Hospital de Fuenlabrada. Sala 1 de Gerencia
- Asistentes
 - Antonio Moreno (UAM)
 - Enrique Torrejón (Bitext)
 - Francisco García Lombardía (H Fuenlabrada)
 - José Luis Martínez (Daédalus)
 - Manuel de Buenaga (UEM)
 - Paz Iglesias Casarrubios (H Fuenlabrada)

2 Participantes y esquema de elementos inicial de reunión



- Esquema de elementos iniciales
 - Definición de objetivos
 - Presentación en taller Mavir mitad de Noviembre (16,17)
 - Objetivo del trabajo (e.g. definición básica propuesta Profit)
 - Definición del papel de las entidades participantes
 - Interés
 - Equipo
 - Escenarios/Usos/Funcionalidades
 - Enfoques complementarios o alternativos

3 Resultados de la reunión



- Revisión de los objetivos y la situación general del programa MAVIR (e.g. relación programa de postgrado)
- Planteamiento de dos funcionalidades o sistemas diferentes:
 - Sistema de extracción de datos de informes de alta para introducción en plantilla interna del hospital para gestión de estadísticas
 - Sistema de búsqueda de información multimodal, incluyendo informes e imágenes de radiografías, con facilidades de análisis de texto e imágenes.
- Análisis y puesta en común sobre conjunto significativo de situaciones concretas y ejemplos de utilización
 - tratamiento de pacientes: procesamiento de placas con diagnóstico de neumonía
 - con informe inicial del radiólogo y posterior tratamiento por un pediatra
 - interpretación puede diferir de la un segundo
 - también, el uso frecuente de datos estadísticos por parte de los médicos para investigaciones y presentaciones en congresos.
- Continuación básica de trabajo:
 - Descripción básica del papel de cada entidad en un posible proyecto Profit.
 - Hospital de Fuenlabrada: profundización en el análisis y definición de las fuentes de información concretas que podrían ser objeto de trabajo en los proyectos.

4 Papel de los participantes (Bitext)



- Las dos líneas de investigación que se propusieron en la reunión en el Hospital de Fuenlabrada:
 1. Buscador multimodal
 2. Ayuda automática para rellenar formularios
- 1. Buscador multimodal
 - Con la tecnología lingüística de Bitext, se puede desarrollar la parte lingüística del buscador multimodal para realizar consultas en lenguaje natural de la parte textual de la base de datos que se interroge.
 - Esta parte lingüística podría incorporar la información léxica de las ontologías que desarrollara la UAM. Este desarrollo sería continuación del ya iniciado en el proyecto ISIS
- 2. Ayuda automática para rellenar formularios
 - Con la tecnología lingüística de Bitext, se puede desarrollar aplicaciones que pueden "leer" textos y buscar patrones o expresiones específicas que puedan ser candidatos para rellenar los campos de los formularios.
 - Este desarrollo se basaría en la tecnología de NaturalMail que ha desarrollado Bitext para la categorización automática de la información contenida en correos electrónicos.

4 Papel de los participantes (Daedalus)



- Tres necesidades detectadas:
 1. Localización de información en expedientes médicos combinando datos textuales con datos en formato de imagen
 - Explotación al máximo de ambas fuentes de información (texto o imagen) por separado así como de su posible combinación. Por ejemplo, se podría dar acceso a todos los ficheros textuales que hacen referencia a una misma patología, haciendo lo mismo con las imágenes correspondientes, que complementarían a la información textual.
 - Ninguno de los actuales miembros de MAVIR tiene experiencia y conocimientos para un análisis de imagen detallado (reconocimiento de formas), pero podría buscarse fuera.



1. Desarrollo de herramientas para apoyo al proceso de diagnóstico
 - Agrupamiento de diagnósticos previos
2. Apoyo en el proceso de compleción de expedientes
 - centrada en la explotación de documentos de texto libre para rellenar una serie de formularios predefinidos.
- DAEDALUS tiene experiencia en el desarrollo de aplicaciones para la búsqueda de información. Ej.:
 - TPI Páginas Amarillas, funcionalidad 'Usted quiso decir ...' (búsqueda borrosa)
 - StockPhotos, desarrollando taxonomías para clasificación de imágenes y herramientas de ayuda a su etiquetado
 - SDAE, desarrollando LUCAS, un Localizador Universal de Contenidos Audiovisuales

4 Papel de los participantes (Hospital de Fuenlabrada)



- Actividades en proyectos
 - Prescripción de los requisitos de información clínica
 - Colaboración la definición de diccionarios o contenidos para comparación
 - Facilitar datos de muestra desde el sistema de información hospitalario
 - Ser probador de la solución de recuperación de información desarrollada
 - Participar en el ciclo de mejora continua
- Aporte de fuentes de información electrónica
 - Información demográfica de pacientes (anonimizada)
 - Informes de alta hospitalaria
 - Informes de continuidad de cuidados
 - Notas de evolución
 - Formularios con indicadores
 - Informes de pruebas diagnósticas (placas, analíticas, biopsias, etc.)
 - Posibilidad de facilitar imágenes y video diagnóstico (placas RX, endoscopias, ecografías...)

4 Papel de los participantes (UAM)



- En torno a las dos funcionalidades
 1. Buscador multimodal
 2. Ayuda automática para rellenar formularios
- Doctores del grupo de la UAM
 - Moreno Sandoval (Laboratorio de Lingüística Informática)
 - Alfonseca Cubero (Departamento de Ingeniería Informática) y sus investigadores asociados
- Función 1: Buscador multimodal
 - Desarrollo de un **extractor terminológico** aplicado al dominio médico, y más en concreto a los datos disponibles.
 - Este programa incluiría las siguientes funcionalidades:
 1. Extracción automática de posibles términos a partir de los datos
 2. Búsqueda automática de definiciones y ejemplos para cada término, tanto en los datos originales como en otras fuentes externas
 3. La validación final de la base terminológica con expertos médicos.



- Función 1: Buscador multimodal (sigue)
 - **Creación de una ontología del dominio médico**, a partir de los términos extraídos, que servirá de interfaz para los otros módulos de recuperación desarrollados por los otros equipos del proyecto. El trabajo en la ontología sería compartido con otros equipos.
 - Extensión del modelo a un **buscador multilingüe**. La UAM tiene experiencia en recuperación de información en lenguas como el inglés y el árabe. Una de las funcionalidades es aplicar el extractor terminológico y la ontología a estas lenguas o a otras como el francés o alemán.
- Función 2: Ayuda automática para rellenar formularios
 - La UAM tiene experiencia previa en elaboración de resúmenes a partir de documentos periodísticos y educativos.
 - Su contribución en este proyecto podría situarse en la ayuda a la composición automática de resúmenes a partir de informes médicos de cierta extensión.

4 Papel de los participantes (UEM)



- Las dos funcionalidades propuestas en la reunión en el Hospital de Fuenlabrada:
 - 1. Buscador multimodal
 - 2. Ayuda automática para rellenar formularios
- El grupo de Sistemas Inteligentes de la UEM (6D+6 TS – Informática y Ciencias Salud) podría jugar un papel relacionado con los siguientes aspectos de un posible proyecto Profit, bien relacionado con una sólo de las funcionalidades, o en ambas:
 - Coordinación de proyecto, si es necesario, por su experiencia general, y particular en este programa
 - Análisis de métodos y sistemas existentes: incluyendo proyectos en marcha y productos en el mercado relacionados.
 - Recogida de datos (en colaboración directa con el Hospital de Fuenlabrada, al menos, necesariamente)
 - Desarrollo de biblioteca de componentes software para categorización de documentos y reconocimiento de entidades nombradas
 - Desarrollo de prototipo con interfaz de usuario
 - Evaluación del sistema sobre un entorno médico, incluyendo experiencia sobre usuarios de perfil académico (profesores y alumnos)



- Buscar oportunidades de colaboración o sinergias con los proyectos relacionados en los que participa, ISIS y SINAMED

Título del proyecto: "**SINAMED**: Diseño e integración de técnicas de generación de resúmenes y categorización automática de textos para el acceso a la información bilingüe en el ámbito biomédico"

Entidad financiadora: Ministerio de Educación y Ciencia. Programa Nacional de Tecnologías de la Información del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica 2004-2007. TIN-2005-08998-C02-01

Entidades participantes: Fundación General de la Universidad Europea de Madrid, Universidad de Huelva

Duración, desde: 07/10/2005 hasta: 07/10/2008

Subvención total del proyecto: 128.500 euros Subvención del proyecto FG.UEM: 69.800 euros

Número de investigadores participantes: 7(UEM), 4(UHU)

Título del proyecto: **ISIS**: Sistema Inteligente de Acceso Integrado a la Información del Historial Clínico del Paciente y Documentación Médica Relacionada (Fase 1)

Entidad financiadora: Ministerio de Industria Turismo y Comercio. Programa de fomento de la Investigación Técnica (PROFIT). Programa Nacional de Tecnologías de la Sociedad de la Información. Subprograma nacional de e-contenidos FIT-350200-2005-16

Entidades participantes: Prouniversidad (Universidad Europea de Madrid), Universidad de Huelva, Hospital de Fuenlabrada, Bitext S.L.

Duración, desde: 01/04/2005 hasta: 31/12/2005

Presupuesto del proyecto: 118.000 euros Cuantía de la subvención: 42.000 euros

Presupuesto del proyecto UEM: 76.600 euros Cuantía subvención UEM: 19.158

Número de investigadores participantes: 7(UEM), 4(UHU), 2(Bitext), 2(HFU)

5 Análisis de las fuentes de información



- Información demográfica de pacientes (anonimizada)
- Informes de alta hospitalaria
- Informes de continuidad de cuidados
- Notas de evolución
- Formularios con indicadores
- Informes de pruebas diagnósticas (placas, analíticas, biopsias, etc.)
- Posibilidad de facilitar imágenes y video diagnóstico (placas RX, endoscopias, ecografías...)

6 Planteamiento de una propuesta Profit



- Fecha objetivo?
 - Convocatoria 2006 (TSI): BOE 03-03-06
- Consorcio?
 - Abierto
 - Análisis y resolución de posibles conflictos
- Tipo de proyecto?
 - Estudio técnico previo a la investigación industrial
 - Proyecto de investigación industrial
 - Estudio técnico previo de desarrollo precompetitivo
 - Proyecto de desarrollo precompetitivo
- Funcionalidades?
 - Selección de funcionalidades?
 - Integración con nuevos elementos/funcionalidades
- Presupuesto?
 - e.g. ISIS
 - fase1: 118.000€ / 42.000€
 - fase2: 234.000€ / 170.000€