

BioViewMed, una herramienta visual de ayuda a la expansión de la cadena de búsqueda usando conocimiento semántico*

BioViewMed, a visual tool to support expansion of search string using semantic knowledge

Sebastián García Pérez
Escuela Politécnica Superior
Universidad de Huelva
sebastian.garcia@alu.uhu.es

Manuel de la Villa
Depto. de Tecnologías de la
Información
Universidad de Huelva
manuel.villa@dti.uhu.es

Manuel J. Maña
Depto. de Tecnologías de la
Información
Universidad de Huelva
manuel.mana@dti.uhu.es

Resumen: Presentamos una herramienta que ofrece una ayuda visual a la adición de términos en la cadena de búsqueda, en pre-recuperación, controlada por el usuario y guiada por la información contenida en ontologías.

Palabras clave: Recuperación de información, cadena de búsqueda, ontologías, UMLS, Freebase, grafos de contextos.

Abstract: We present a tool that offers a visual aid to the addition of terms in the search string in pre-recovery, user-controlled and guided by the information contained in ontologies.

Keywords: Information retrieval, search string, ontologies, UMLS, Freebase, context graphs.

1 Introducción

Son conocidos los problemas que tienen la mayor parte de los usuarios para definir en una consulta sus necesidades de información (véase, p. ej., (Spink et al., 2002)). Estos problemas se deben, en parte, a que en muchos casos las interfaces de los sistemas de búsqueda no son las adecuadas (Hearst, 1999b). Pero también, en buena medida, a que cuando el usuario inicia una búsqueda desconoce realmente qué puede serle útil y, por tanto, le resulta complicado especificar las características destacables de los elementos de información potencialmente útiles (Belkin, 2000).

El resultado de este tipo de consultas es habitualmente una lista de miles de documentos recuperados, difícilmente relacionados con las necesidades reales del usuario y de compleja navegación y asimilación.

Nuestro objetivo es ayudar al usuario a definir y concretar la búsqueda mediante la construcción de un grafo que recoja las relaciones semánticas extraídas de una ontología, a partir de un concepto biomédico. El

usuario selecciona nuevos términos directamente del grafo, que amplían la cadena de búsqueda, aumentando la capacidad descriptiva de la cadena de búsqueda, de modo que las necesidades de información expresadas se acerquen lo más posible a las necesidades reales de información.

2 Descripción del sistema

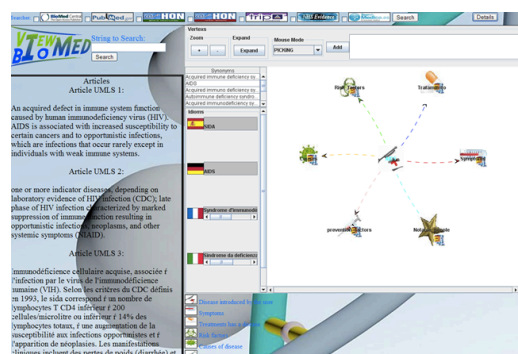


Figura 1. Interface de la aplicación completa con el grafo y definiciones

* Este trabajo ha sido parcialmente financiado por ERDF (TIN2009-14057-C03-03)

El proceso que realiza la aplicación web (Figura 1) se puede dividir en tres partes claramente diferenciadas:

2.1 Pre-recuperación del concepto.

En esta fase se realizan principalmente tres tareas:

Tarea 1. *Ayudar al usuario a definir el concepto inicial.* A medida que el usuario va introduciendo la cadena inicial de búsqueda, haciendo uso de Freebase y de AJAX, se ayuda al usuario enseñándole lista desplegable de conceptos (y su tipo) en cuya descripción aparece la cadena introducida.

Tarea 2. *Consulta a las fuentes de información.* Una vez que el usuario ha encontrado el concepto del que desea obtener información, se realizan las consultas con ese término tanto a Freebase como a UMLS.

- *Consulta a Freebase:* Utilizamos la API que nos proporciona Freebase, en concreto, una consulta MQL de la información que necesitamos, en concreto obtenemos: Tratamientos, Causas, Factores de Riesgos, Síntomas y Personas Famosas
- *Consulta a UTS (UMLS Terminology Services),* que incluye una API de Servicio Web para, de manera remota consultar y recuperar información de los recursos léxicos y semánticos de UMLS. A través de SOAP 1.2 y WDSL se describe la interfaz pública a dichos servicios Web. UTS nos proporciona la siguiente información:
 - Sinónimos.
 - Traducción del término a los principales idiomas.

Tarea 3. *Guardar la información en un archivo XML.* Esto se hace como almacenamiento intermedio con vistas a posibles comunicaciones futuras incrustado en otros sistemas.

2.2 Construcción del grafo

Cuando hemos conseguido recuperar toda la información necesaria, se construye el grafo (Figura 2) usando la librería JUNG (O'Madadhain et al., 2005).

Se muestra en el centro del grafo el concepto inicial con aristas a nodos que representan los conjuntos comprimidos de Tratamientos, Causas, Factores de Riesgo, Síntomas y

Celebridades que han padecido la enfermedad, cada uno de ellos con un icono característico.

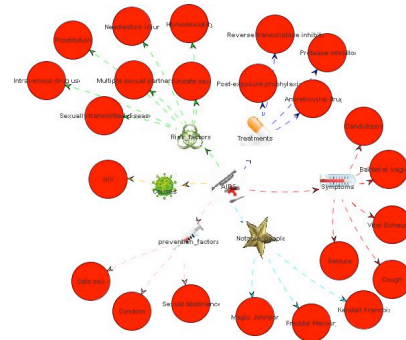


Figura 2. Grafo de conceptos expandido

2.3 Navegación por el grafo

Llegados a este punto, la interfaz permite al usuario de manera intuitiva: expandir los nodos grupales (mostrando en detalle, pej., un nodo nuevo por cada factor de riesgo asociado a la enfermedad inicial), poder ver la información asociada a cada nodo en las diferentes ontologías que utilizamos, poder añadir términos desde un nodo a la cadena de búsqueda, etc.

2.4 Recuperación con la cadena expandida

El sistema le da la posibilidad al usuario de hacer una búsqueda más detallada en buscadores médicos más especializados.

Bibliografía

- Belkin, N. J. 2000. Helping People Find What They Don't Know. *Communications of the ACM*, 43(8):58-61.
- Hearst, M. A. 1999b. User Interfaces and Visualization. En R. Baeza-Yates y B. Ribeiro-Neto, eds., *Modern Information Retrieval*, págs. 257-323. Addison-Wesley Longman Publishing Company, New York.
- O'Madadhain, J., Fisher, D., Smyth, P., White, S., and Boey, Y.-B. 2005. Analysis and visualization of network data using JUNG. *Journal of Statistical Software*, VV:1-35.
- Spink, A., Jansen, B., Wolfram, D. y Saracevic, T. 2002. From E-Sex to E-Commerce: Web Search Changes. *Computer*, 35(3):107-109.