

La visualización de resultados de búsqueda y los gestores bibliográficos personales: una nota técnica sobre RefViz 2.

Carlos B. Amat (carbea@iata.csic.es)

Documentación Científica y Biblioteca

Instituto de Agroquímica y Tecnología de Alimentos (IATA CSIC)

Apdo. 73 46100 Burjassot (Valencia)

Resumen

Los gestores personales de bibliografía permiten la organización de las referencias; facilitan la intercalación de las referencias seleccionadas en los manuscritos; sirven de interfases para las búsquedas directas en algunos sistemas, y facilitan el trabajo colaborativo a través del intercambio de bibliografías. Se trata de programas convencionales que comparten con los sistemas generales un tratamiento lineal de las listas de referencias.

Se analiza la aplicación RefViz 2 en el contexto de las aplicaciones de visualización y análisis de resultados de búsqueda y se ilustra el cumplimiento de los objetivos adjudicados a programas similares. Las limitaciones en la gestión de referencias aisladas y la falta de conectividad con los programas de procesamiento de texto y de filtros de salida no permiten caracterizar a RefViz como un programa autónomo, sino como una aplicación modular que necesita combinar sus prestaciones con el manejo de programas más completos de gestión personal de referencias bibliográficas.

Search results visualization and personal bibliographic software: a contextual note on RefViz

Abstract

Personal bibliographic software helps organizing references. Next, these packages can create bibliography and footnotes in combination with popular word processing programs. They also allow direct searching in some data bases and sharing of bibliographic files among research colleagues. However, as search results become more and more numerous, conventional treatment of references lists presents some difficulties to scientific information users.

Along with other applications designed to visualize search results, RefViz is more a data visualization and analysis software than a bibliographic management package. Nevertheless, its use in combination with other reference managers can give users a powerful tool to explore large amount of collected references based on their content.

Introducción

La convergencia de las bases de datos estructuradas de literatura científica y las posibilidades de distribución que el espacio Web proporciona ha supuesto un reto y, al tiempo, ha ofrecido una oportunidad en el manejo de información científica. El reto está planteado desde un punto de vista cuantitativo: tanto los sistemas primarios (los grandes grupos editoriales) como los secundarios (las bases de datos de bibliografía científica) han evolucionado en sus respectivas coberturas hasta adaptarse al acelerado ritmo de

producción de documentos científicos de la actualidad. Los 36 millones de registros del Web of Science sobrepasan en poco a los 35 de Chemical Abstracts; PubMed supera los 16 millones, pero BioSIS le aventaja en 2. La consulta de las Dialog Blue Sheets revela que también las bases de datos especializadas (Historical Abstracts y PsycINFO entre otras) alcanzan registros millonarios. Además, tanto los grandes sistemas editoriales como ciertos títulos individuales han aplicado procesos de conversión retrospectiva y digitalizaciones masivas que les han permitido incorporar a sus ofertas informativas las entregas iniciales de cada serie. Como muestras, Journal of Biological Chemistry es accesible desde su primer número, publicado en octubre de 1905; los mismo que Journal of Nutrition (septiembre de 1928) y Journal of the American Chemical Society, que publicó el primer trabajo el 2 de enero de 1879. La presión de una literatura pujante originó la aparición de índices y repertorios a finales del siglo XIX e impulsó su automatización en los primeros años 60 del pasado siglo. En la actualidad, esa presión se ha trasladado a los usuarios de la información científica en forma de listas inabordables de resultados de búsqueda (Amat, 2007) y ha agravado la sobrecarga informativa, *“la percepción de sentirse abrumados y, así, confusos por la información que recibimos, de forma que se dificultan el aprendizaje o la capacidad de los usuarios para tomar decisiones informadas”* (Kim et al., 2007).

La visualización de resultados de búsqueda

Por fortuna, la abundancia de la oferta informativa ha venido acompañada por una consideración nueva de la literatura científica, que está directamente inspirada por la Web, y que comienza a ofrecer oportunidades de manejo de la información más allá de la mera ordenación lineal de grandes conjuntos de documentos resultantes de búsquedas. En este nuevo contexto, los documentos no se presentan de forma aislada, sino interrelacionada. Los indicadores de relación, además, se pueden transformar en elementos de ordenación de grupos de documentos. Google fue el primer sistema de recuperación en Internet que, superando la mera consideración textual de los documentos Web, utilizó relaciones contextuales (los enlaces) para recuperar y ordenar los resultados de búsqueda. Los principios básicos que definen este modelo de organización espacial de la información son dos: 1) todo documento está relacionado con otro y esta relación se desprende de su estructura, y 2) la cuantificación de estas relaciones permite ordenar los documentos en espacios informativos.

Cualquier profesional de la información científica conoce [CrossRef](#). Es habitual en la actualidad obtener una referencia junto con una lista de documentos relacionados. Son conocidos los *citation* y *topic maps*, diseñados por Inxight Software para el sistema [HighWire](#). Todas ellas son soluciones que transforman las relaciones entre documentos en vínculos, a veces dotados de expresión gráfica y que *proporcionan un sentido de contexto en el recorrido de grandes bases de datos*. Antes ([NorthernLight](#)) y más allá de Google se encuentran [Vivisimo](#) y su *clustered search* o la interface gráfica de resultados de [Mooter](#). Pero, en lo tocante a la información científica y su recuperación, la oportunidad que los conceptos mencionados ofrecen está representada por la aplicación de técnicas de visualización de la información a los resultados de búsqueda. La tabla 1 presenta una cronología inversa de sistemas desarrollados por equipos de organizaciones diversas que emplean la representación gráfica de elementos informativos y sus interrelaciones a través de interfases visuales.

Año	Sistema	Diana	Afiliación
2007	HistCite	Web of Science	HistCite Software LLC
2007	eTBLAST	PubMed, IOP, NASA	University of Texas
2006	HubMed	PubMed	University Health Network Toronto
2005	Dragon Plant Biology	PubMed	National University of Singapore
2005	AnaVist	Chemical Abstracts	Scientific and Technical Information Network
2005	PubNet	PubMed, GenBank	Yale University
2004	PNASLink	PubMed, WOS	Drexel University
2003	PubMatrix	PubMed	National Institutes of Health
2003	Telemakus	KB propia	University of Washington
2001	XplorMed	PubMed	European Molecular Biology Laboratory

La aplicación de estos programas a las bases de datos especificadas como dianas ofrece una representación gráfica de los documentos que se ajustan al perfil de búsqueda empleado. HistCite estructura los trabajos científicos basándose en las relaciones de citas. El resto de los sistemas recurren al modelo del espacio vectorial (Salton, 1991) para calcular los coeficientes de similitud (las distancias) entre trabajos. Los corpus suelen limitarse a los resúmenes. PubNet puede, además, visualizar las relaciones de coautoría. La figura 1 reproduce uno de los gráficos resultantes de una búsqueda en PubMed a través de eTBLAST. El recuadro muestra el grado de similitud de uno de las referencias recuperadas con el trabajo cuyo resumen se empleó como perfil de búsqueda (<http://dx.doi.org/10.1126/science.1106340>)



Sistemas convencionales de gestión bibliográfica

Los sistemas personales de gestión bibliográfica vienen siendo soluciones al reto que representa la organización de un número creciente de documentos científicos. Cumplen cuatro objetivos principales. En primer lugar, permiten la organización de las referencias bibliográficas. Además, facilitan la intercalación de las referencias

seleccionadas en los manuscritos. En tercer lugar, sirven de interfases para las búsquedas directas en algunos sistemas. Finalmente, permiten el trabajo colaborativo a través del intercambio de bibliografías. A pesar de algunos olvidos, resulta de gran utilidad el análisis comparativo que Emilio Duarte García ha publicado recientemente (Duarte García, 2007).

Cabe calificar a estos sistemas personales como convencionales, en el mismo sentido en que son convencionales las bases de datos de literatura científica. En ambos casos los registros que representan trabajos científicos se tratan de forma lineal, acumulativa y del mismo modo se presentan los resultados de búsqueda a los usuarios. Por lo que respecta a la oferta de trabajos relacionados con los resultantes de la búsqueda, siguen presentándose como listas y obligan igualmente a un examen secuencial, con frecuencia demasiado prolongado.

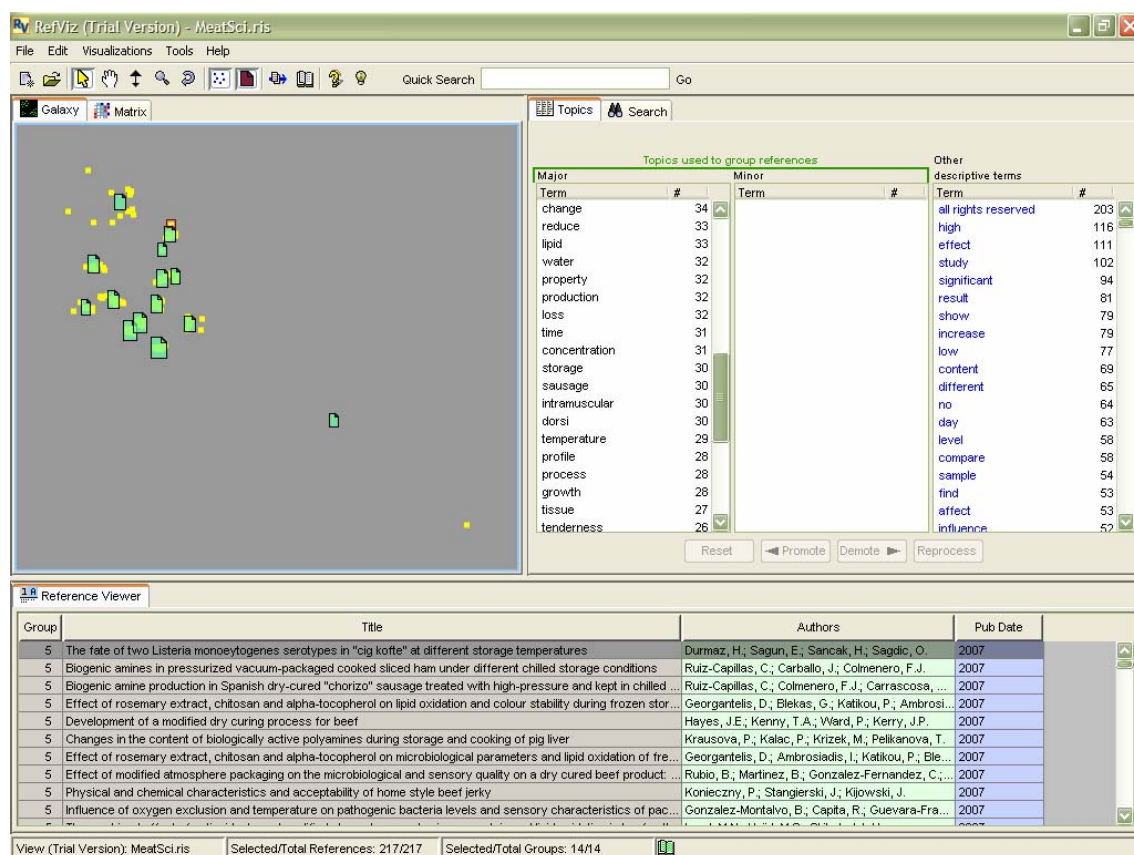
RefViz

RefViz está disponible desde julio de 2003 y su versión 2 apareció en abril de 2005. Fue desarrollado conjuntamente por Thompson ISI ResearchSoft y OmniViz (ahora BioWisdom). Conviene examinar esta aplicación siguiendo los objetivos que se adjudican a los gestores bibliográficos personales, antes enumerados.

Las particularidades de RefViz comienzan con los mecanismos de carga y visualización de referencias. No es posible añadir de forma manual referencias a los archivos, que aquí se denominan “vistas” (views). Las referencias se obtienen 1) mediante importación en lote de ficheros con formatos predefinidos (RIS, naturalmente y las salidas de varias bases bibliográficas); 2) mediante la búsqueda y recuperación desde el propio programa en PubMed, Web of Science y otros recursos, y 3) mediante la importación de archivos procedentes de bases RefereceManager, Procite o EndNote. La adición de referencias aisladas ha de producirse a través de cualquiera de estos programas para, posteriormente, a través de la herramienta de visualización de datos (tools: data visualization) poder examinarlas gráficamente en RefViz. Y es que no existen archivos o bases en RefViz: sólo visualizaciones.

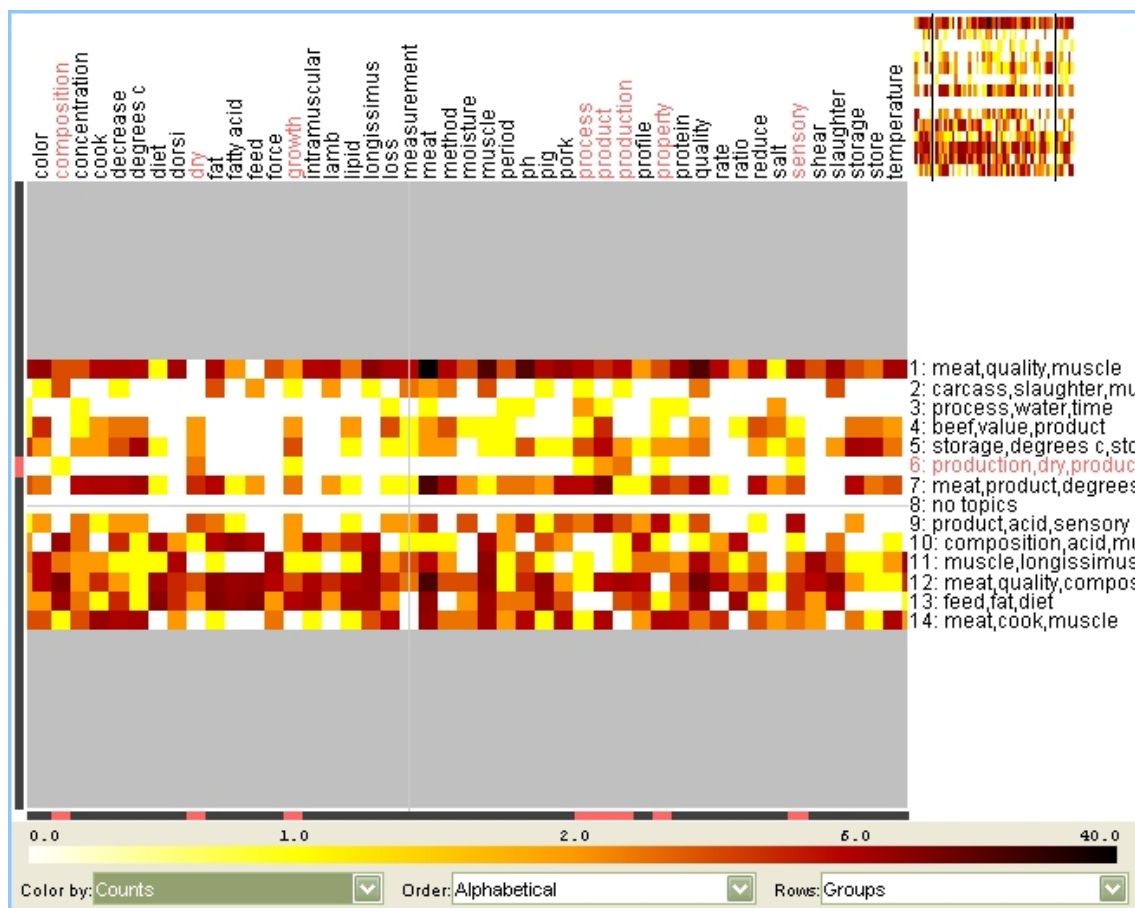
El objetivo fundamental de esta aplicación es la estructuración de grupos de documentos científicos a través del análisis de una parte de sus textos, habitualmente sus resúmenes.

La interfase del programa recuerda poderosamente a la de AnaVist (STN). La captura de la figura 2, realizada sobre un conjunto de 219 referencias, permite visualizar los documentos agrupados (galaxia), su lista y los términos principales y secundarios resultantes del análisis realizado por el programa. Son conmutables dos formatos de visualización de los conjuntos de referencias: una representación tabular (“matrix”) y otra espacial (“galaxy”).



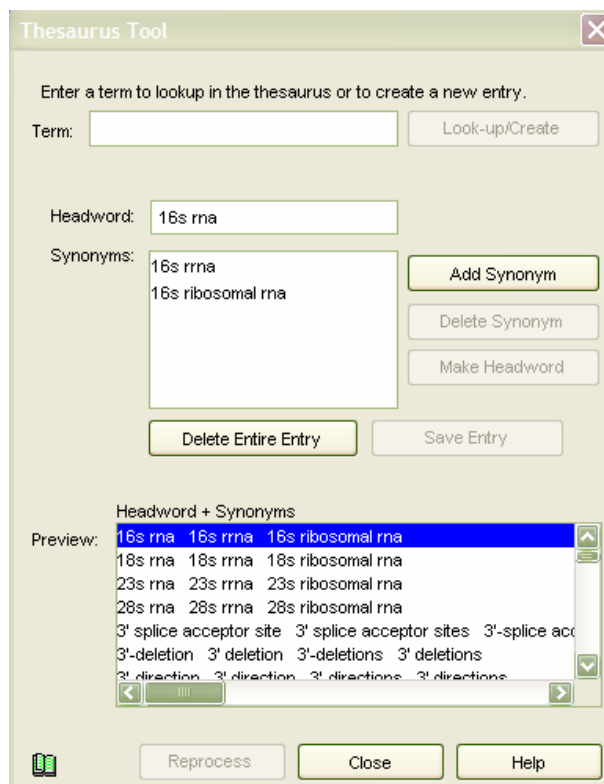
La operación básica que seguiría al establecimiento de una nueva “vista” sería su proceso adicional, mediante el descarte de alguno de los términos principales o la promoción de alguno de los secundarios. Como resultado, los grupos de documentos se reordenan y la identificación de referencias relevantes se aligera.

La visualización en modo tabular ofrece mayores posibilidades de interpretación de los datos. Así, para un grupo especializado en la investigación del color de los productos cárnicos, resulta fácil identificar en la vista que se reproduce a continuación como figura 3, que los grupos de trabajos 1, 11-12 y 14 son quienes presentan mayor “densidad” del término *color*.



Naturalmente, es posible la consulta de las referencias almacenadas en busca de aquellas que se ajusten a un perfil determinado. En respuesta a una búsqueda, no sólo se obtiene la lista de documentos relevantes; también se destaca su posición en la representación gráfica de los grupos.

Por otra parte, el control terminológico de los documentos de cada archivo y el establecimiento de relaciones se produce con mayor agilidad que en otros programas de la misma familia (figura 4) y la posibilidad de creación de términos preferentes (aquí denominados *headwords*) es mucho menos farragosa que en referente Manager, por ejemplo. Aunque las únicas relaciones admitidas parecen ser las de sinonimia. A diferencia de otros programas, no existe ni diccionario onomástico ni de nombres de publicación.



RefViz no presenta conectividad con procesadores de texto y no dispone de los habituales filtros de salida que permitan formatear las referencias de acuerdo con uno u otro estilo. De forma indirecta, las referencias seleccionadas a través de este programa se han de exportar a cualquier otro de gestión bibliográfica para, desde allí, acomodarlas en los manuscritos.

En cuanto al procedimiento de búsqueda directa, es indistinguible del habitual en este tipo de programas salvo por la aparición de una curiosa animación en forma de lebril que comunica a ladridos la obtención de documentos. PubMed y Web of Science, como es habitual, siguen siendo las dianas preferentes.

El intercambio de bibliografías es posible con RefViz. En primer lugar, a través de la exportación de conjuntos de referencias como archivos textuales. También mediante el envío a otras aplicaciones de manejo de bibliografía. Finalmente, en un formato propio de extensión .xpt que es reconocible por otros usuarios e instalaciones de RefViz. A este respecto, aunque no exista una versión “en red” sí se ofrece la distribución de varias licencias sobre una base departamental o institucional y a un precio inferior al mercado para la adquisición individual del programa: 149,95 \$ tanto en la versión para el entorno Windows como para MacOS.

Un módulo más que una aplicación

Las limitaciones en la gestión de referencias aisladas y la falta de conectividad con los programas de procesamiento de texto y de filtros de salida no permiten caracterizar a RefViz como un programa autónomo, sino como una aplicación modular que necesita combinar sus prestaciones con el manejo de programas más completos de gestión personal de referencias bibliográficas. No en vano, sus propios creadores lo conceptúan como un programa de visualización y análisis de datos. En este sentido lo ha empleado

repetidamente en fecha temprana por el grupo cubano de Juan A. Araujo (Arencibia Jorge et al., 2008) para elucidar la estructura conceptual de algunas áreas de investigación clínica. La presión de la bibliografía disponible origina archivos personales que, en los cuatro años en que se realiza una tesis o se desarrolla un proyecto de investigación, puede elevar a varios millares las referencias recopiladas. No cabe duda de que el empleo de RefViz puede rendir claves significativas para un manejo eficaz de estos volúmenes informativos. Pero su adquisición habrá de superponerse sin remedio al de gestores bibliográficos más completos.

Referencias

Amat, C. B. (2007). El fin de las listas: necesidad de bibliografías estructuradas. In T. Baiget (Ed.), *Anuario ThinkEPI* (pp. 148-151). Barcelona: EPI.

Arencibia Jorge, R., Vega Almeida, R. L., & Araujo Ruiz, J. A. (2008). La investigación sobre hiperplasia prostática benigna: construcción de una estructura conceptual a través de técnicas bibliométricas, minería de textos y análisis semántico. In Ministerio de Informática y Comunicaciones ; Ministerio de Salud Pública.

Duarte García, E. (2007). Gestores personales de bases de datos de referencias bibliográficas: características y estudio comparativo. *El Profesional de la Información*, 16, 647-656.

Kim, K., Lustria, M. L. A., Burke, D., & Kwon, N. (2007). Predictors of cancer information overload: findings from a national survey. *Information Research*, 12, paper326.

Salton, G. (1991). Developments in Automatic Text Retrieval. *Science*, 253, 974-980.