

# Nuevas Fuentes de Información académica

(v1a)

Lluís Codina

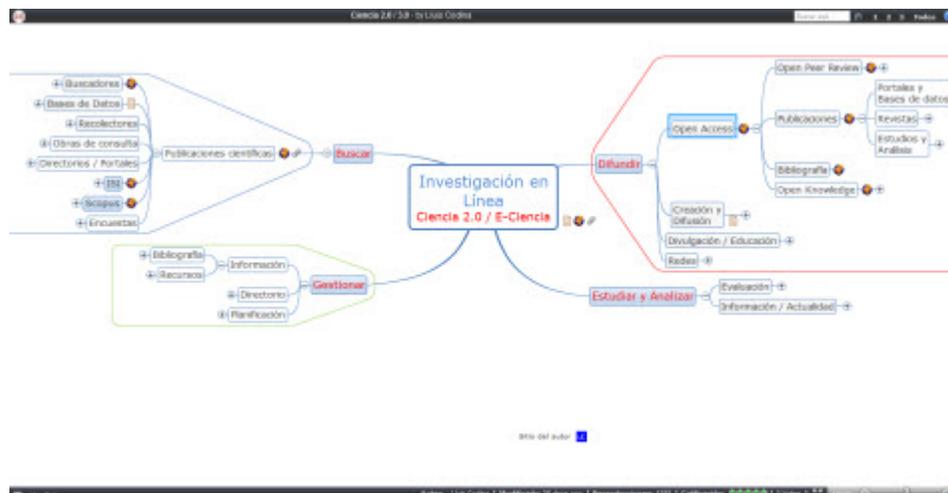
[www.lluiscodina.com](http://www.lluiscodina.com)

Universitat Pompeu Fabra

Julio 2010

## Publicación original y forma recomendada de citación:

Lluís Codina. "Nuevas fuentes de información académica". En: T. Baiget (ed.) J. Guallar (coord.). *Anuario ThinkEPI 2010: Análisis de tendencias en información y documentación*, 2010, p. 139-142 [Acceso v1a: [http://www.lluiscodina.com/fuentesInformacion\\_2010.pdf](http://www.lluiscodina.com/fuentesInformacion_2010.pdf)]



## 1. La Comunidad ATC

Un rasgo común a ciertas comunidades profesionales es que para desarrollar su trabajo necesitan obtener información académica, científica o técnica (ACT a partir de ahora). Forman un grupo característico, dentro del grupo más amplio de profesiones intensivas en información, que suele desarrollar su trabajo basándose principalmente en la denominada información cognitiva (por contraste con la información factual, administrativa, etc.).

Ahora bien, a efectos de este trabajo caracterizaremos también a estos profesionales del sector ACT bajo el marco de dos **presiones** y de dos **necesidades**. En cuanto a las presiones, la primera consiste en que los miembros de tales comunidades deben utilizar como uno de los inputs de su trabajo **información de calidad**.

Esto significa que únicamente pueden llevar a cabo su labor manejando informaciones que estén de alguna forma certificadas o garantizadas mediante esquemas de evaluación comúnmente aceptados (en general esquemas tipo *peer review*). La segunda presión es una consecuencia lógica del primero: se espera de estos profesionales que sean capaces de producir **nuevas informaciones** o nuevos conocimientos (a su vez susceptibles de ser evaluados).

Las presiones anteriores generan necesidades paralelas. La primera consiste en disponer de **yacimientos de información** donde poder acudir cuando necesitan nuevos inputs de información de calidad para su trabajo (p.e. para una nueva investigación). Este acopio de información debe hacerse forma continuada, y de poco servirá si no se dispone de medios para gestionarla de forma muy eficiente cada vez que debe ser (re)utilizada.

Pero la memoria humana necesita recurrir de forma sistemática a instrumentos auxiliares que se utilizan como memorias exosomáticas. Estas informaciones, a veces trabajosamente obtenidas, no serán memorizadas de forma indeleble sin recurrir a herramientas de almacenamiento externas.

Aunque tendemos a retener conceptos, habilidades, teorías e ideas sin que se sepa de ningún límite claro para ello, en cambio no (necesariamente) tendemos a recordarlo todo a nivel de detalle, ni a recordarlo justo en el momento en el que lo necesitamos; tampoco acertamos siempre, ni mucho menos, en detectar las relaciones entre las informaciones recopiladas y muchos menos los valores que asumen las variables con las que necesitamos trabajar.

En definitiva: identificar en un momento determinado (T1) una información o un conjunto de datos obtenidos a veces después de un trabajo intenso de búsqueda en los yacimientos a los que nos hemos referido, no garantiza que, en un momento posterior (T2) acertemos a reutilizarlo.

Todo esto nos lleva a la segunda necesidad: **sistemas para gestionar** aquella información valiosa que eventualmente deberá actuar como input en los diferentes proyectos. A diferencia de los yacimientos de información, que son de ámbito social, estos sistemas han de ser ante todo capaces de adaptarse a perfiles personales y eventualmente de grupo.

Por último, los profesionales de la ATC necesitan divulgar su producción. De nada le sirve a un joven doctor, o a un profesor de universidad haber producido un gran trabajo académico, si no disponen de herramientas para divulgarlo de forma eficiente en la era de la Web. De hecho, es una obligación o una responsabilidad legal para ciertos miembros de esta comunidad, como los profesores de universidad, cuya carrera académica está ligada a su éxito para divulgar su producción científica.

La cuestión es que, aunque para muchos todavía la Web es sinónimo de desorden y de información de baja fiabilidad, lo cierto es que en los últimos años se ha convertido en la principal fuente de todo lo contrario, es decir en el lugar privilegiado donde los profesionales del sector ATC puede encontrar recursos para:

- a) *Obtener* información con el objetivo de producir nuevas informaciones y nuevo conocimiento
- b) *Gestionar* la información obtenida para poder reutilizarla
- c) *Difundir* el nuevo conocimiento de modo que pueda ser utilizado como input de información por parte de otros investigadores

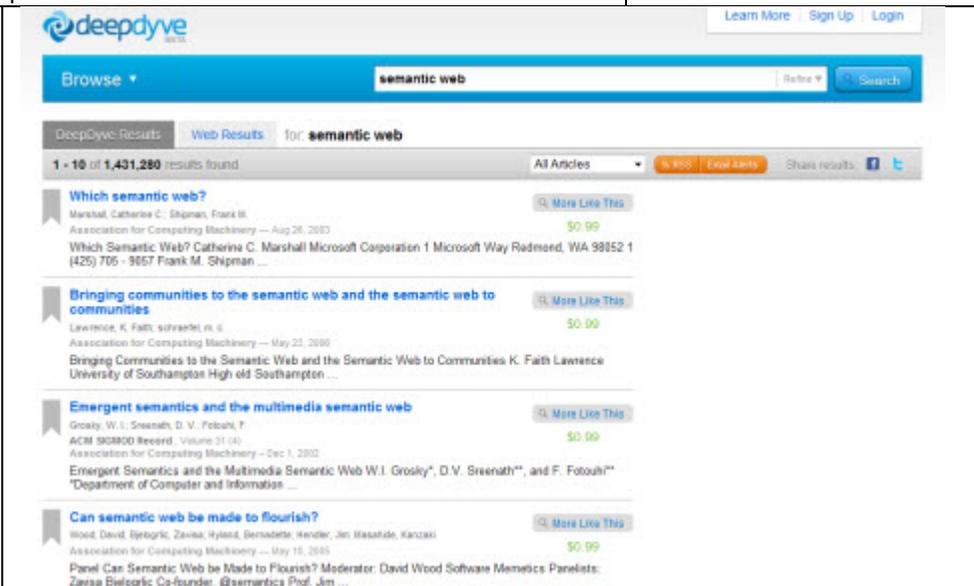
## **2. Tres recursos destacados**

A continuación presentamos tres recursos que han destacado en los últimos años en el sector que estamos considerando. La elección es inevitablemente subjetiva, pero esperamos que también sea razonable.

En lo que sigue se presentan algunas características de tres de los nuevos buscadores que, en nuestra opinión, merecen ser tenidos en cuenta por el conjunto global de sus prestaciones y por su capacidad para convertirse a medio plazo en buscadores de referencia, algunos de ellos con ventaja, al menos en algunos aspectos, sobre los dos grandes "clásicos" (Scholar y Scirus).

## 2.1. DeepDyve

www.deepdyve.com

<b>Presentación</b>	<b>Fuente</b>
<p>DeepDyve por lo que respecta a su contenido (artículos de revistas académicas de prestigio) es muy parecido a las bases de datos académicas convencionales, como por ejemplo, Science Direct, pero con una importante diferencia. Está explorando un nuevo modelo de negocio, o si se prefiere un acercamiento diferente al tradicional. Permite la búsqueda en miles de títulos de las revistas más importantes de diversas áreas, incluyendo las ciencias, las humanidades y las ciencias sociales. En total, contiene unos 32 millones de artículos. Ahora bien, en lugar de tener un enfoque de suscripción orientado a grandes corporaciones, se dirige al usuario final incluyendo una modalidad de suscripción gratuita con un precio unitario máximo por artículo completo de un dólar, esto es entre 20 y 30 veces menos de lo que suelen pedir las bases de datos académicas tradicionales en la modalidad equivalente (pago por artículo). A cambio, no se puede comprar el artículo, sino que se únicamente se "alquila". En este sentido, sería un acercamiento al modelo de los <i>e-books</i>. El alquiler puede ser de 24 horas (suscripción gratuita), siete días o permanente, (estos dos últimos en las dos modalidades de pago).</p> <p>Es importante porque ofrece un nuevo modelo de negocio que puede impactar favorablemente en entornos donde no existe la posibilidad de afrontar las suscripciones de las tradicionales bases de datos, o cuando, una corporación (empresa, hospital, biblioteca, etc.) o un investigador individual necesite puntualmente artículos de revistas no suscritas. Cabe destacar que, incluso la modalidad de suscripción más cara es parecida o inferior al precio típico de la suscripción a una sola revista académica, con la diferencia que se puede acceder al contenido de miles de ellas sin ningún límite ni de cantidad de artículos accesibles ni de tiempo de alquiler.</p>	<p>DeepDyve es una empresa norteamericana fundada en 2005 por científicos procedentes del proyecto del genoma humano.</p>
<b>Vista parcial página de resultados</b>	 <p>The screenshot shows the DeepDyve website interface. At the top, there is a search bar with the text 'semantic web' and a 'Search' button. Below the search bar, it indicates '1 - 10 of 1,431,280 results found'. The results are listed in a table-like format with columns for article titles, authors, and prices. The first result is 'Which semantic web?' by Marshall, Catherine C., Shipman, Frank B., priced at \$0.99. The second result is 'Bringing communities to the semantic web and the semantic web to communities' by Lawrence, K. Faith, priced at \$0.99. The third result is 'Emergent semantics and the multimedia semantic web' by Groszky, W. I., Sreenath, D. V., Fotouhi, F., priced at \$0.99. The fourth result is 'Can semantic web be made to flourish?' by Wood, David, Bjelogrić, Zorana, Hylsted, Bernadette, Hendler, Jim, Itasashide, Kazuaki, priced at \$0.99. Each result has a 'More Like This' link.</p>

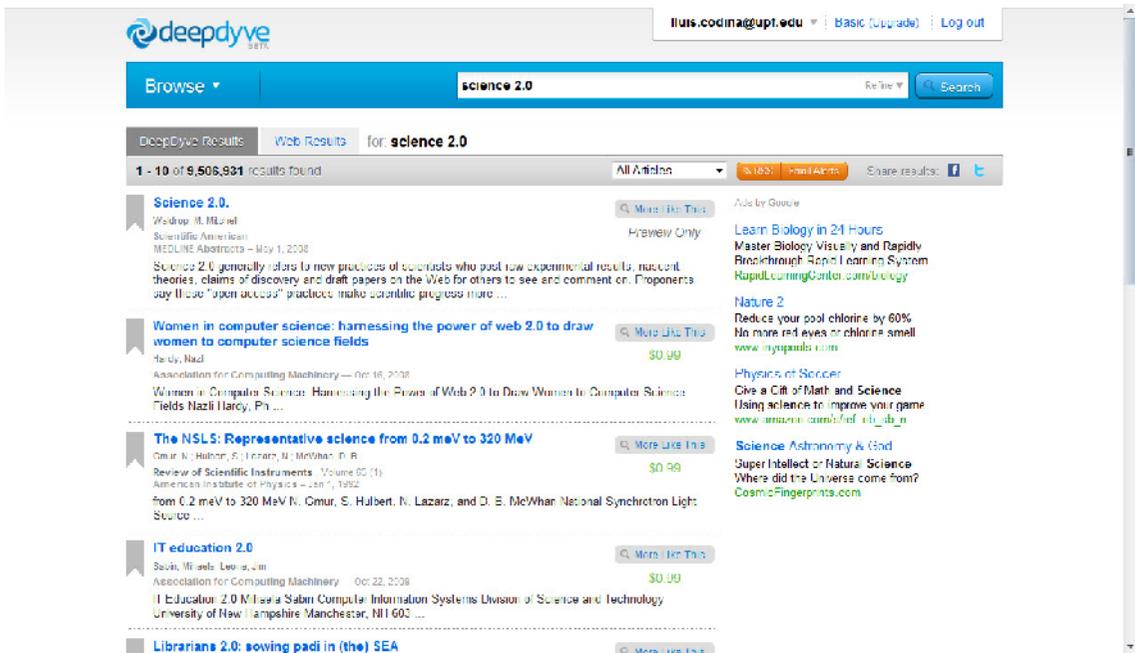


Ilustración 1 Página de resultados de DeepDyve, podemos ver los precios del alquiler de algunos documentos

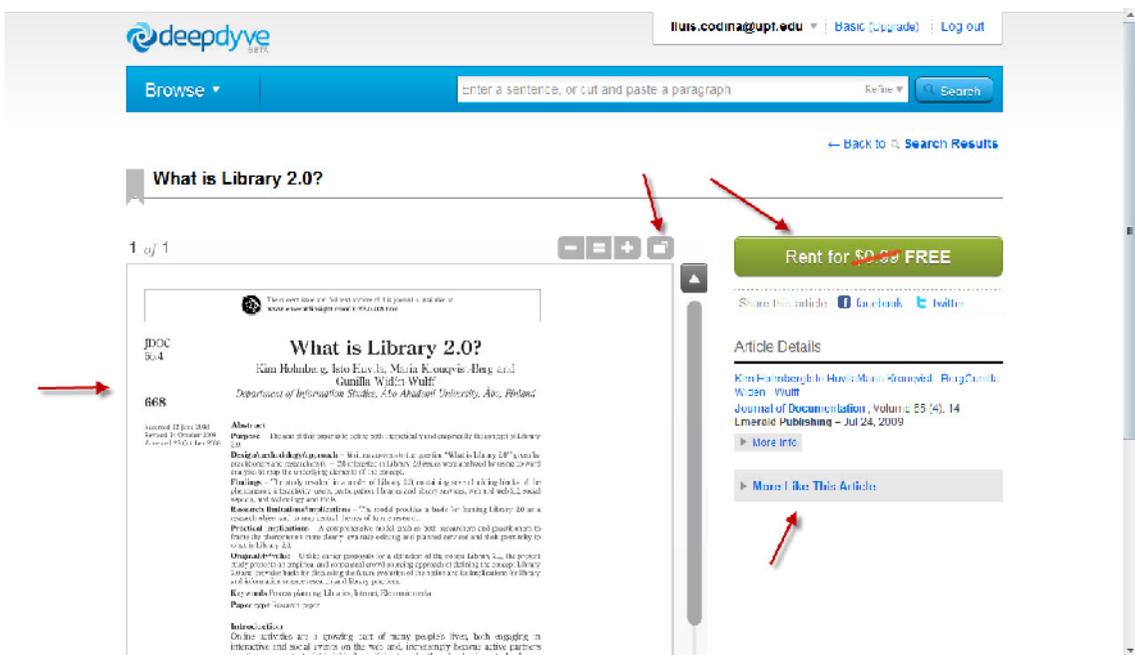
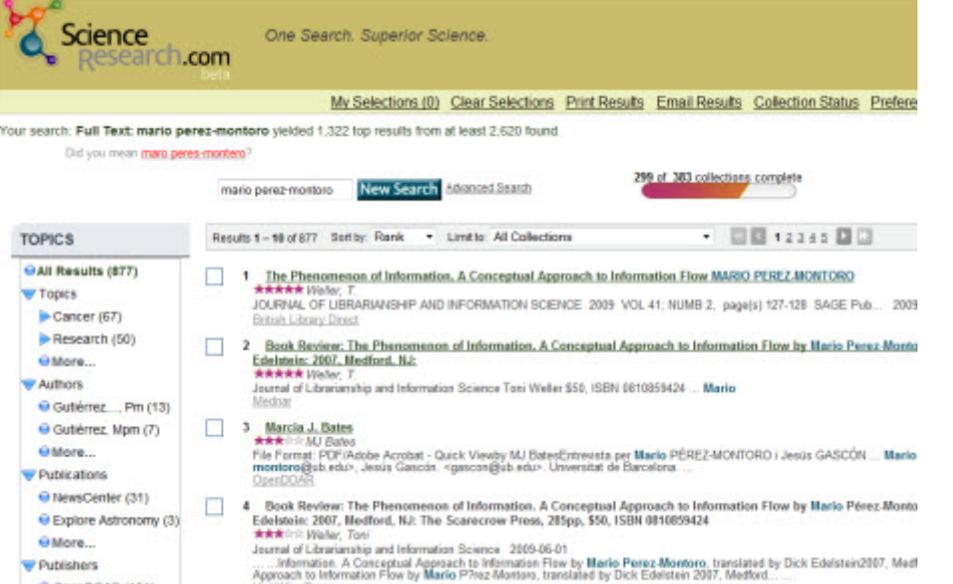


Ilustración 2: Acceso a la versión completa de un artículo. El coste era cero porque correspondía a la cuota de artículos gratuitos de la modalidad de suscripción

## 2. 2. Science Research scienceresearch.com

Presentación	Fuente
<p>Science Research es un sistema de búsqueda federada, es decir, que en lugar de indizar la información, como hace Scirus, por ejemplo, lo que hace es enviar la misma pregunta a centenares de colecciones. Las colecciones, a su vez pueden ser bases de datos, repositorios u otros buscadores. Siempre del ámbito académico. En estos momentos, Science Research es capaz de consultar unas 400 de tales colecciones, por lo cual el volumen de resultados que puede ofrecer para casi cualquier búsqueda es enorme, aunque también es muy grande el número de duplicados que aparecen a veces en sus páginas de resultados. A pesar de que el sistema intenta eliminar duplicados, no en todas las búsquedas tiene el mismo éxito.</p> <p>Es importante porque ofrece la búsqueda federada más ambiciosa, con el objetivo de ser de hecho un buscador universal que, de refinar su tecnología, podría ser en el buscador más exhaustivo.</p>	<p>Deep Web Technologies, empresa norteamericana fundada en el año 2002 especializada en el campo de las búsquedas y la gestión del conocimiento. Ha producido diversos buscadores especializados en ámbitos de la ciencia. El que comentamos aquí es el más genérico de todos ellos.</p>
<p>Vista parcial página de resultados</p>	 <p>The screenshot shows the Science Research search interface. At the top, it says 'Science Research.com One Search. Superior Science.' Below that, there are navigation links: 'My Selections (0)', 'Clear Selections', 'Print Results', 'Email Results', 'Collection Status', and 'Preferences'. The search query is 'Full Text: mario perez-montoro' which yielded 1,322 top results from at least 2,620 found. A search bar contains 'mario perez-montoro' and a 'New Search' button. A progress bar indicates '299 of 383 collections complete'. On the left, there is a 'TOPICS' sidebar with categories like 'All Results (877)', 'Topics' (Cancer, Research), 'Authors' (Gutiérrez), 'Publications' (NewsCenter, Explore Astronomy), and 'Publishers' (OpenDOAR). The main results list shows four items, including 'The Phenomenon of Information. A Conceptual Approach to Information Flow' by Mario Pérez-Montoro, with a star rating and publication details.</p>

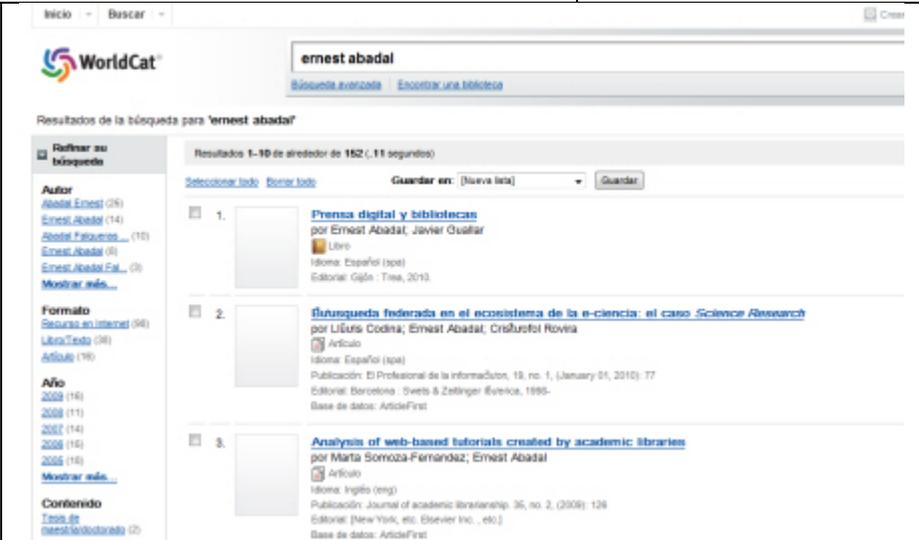
Your search: Full Text: ciberperiodismo yielded 471 top results from at least 121,859 found.

ciberperiodismo  [Advanced Search](#) 382 of 382 collections complete

**TOPICS** Results 1 - 10 of 471 Sort by: Rank Limit to: All Collections

Ilustración 3: Opciones de la página de resultados de ScienceResearch

**2. 3. WorldCat**  
**www.worldcat.org**

<b>Presentación</b>	<b>Fuente</b>
<p>WorldCat afirma ser el catálogo más grande del mundo de materiales académicos con más 1.500 millones de documentos. Es un buscador tipo recolector, es decir, introduce en su índice registros producidos por diferentes productores de bases de datos, entre ellos registros procedentes de catálogos producidos por centenares de bibliotecas de todo el mundo, así como registros de repositorios digitales también de origen universitario. WorldCat no es nuevo, pero sí lo es la integración de buscadores como OAister y así sus nuevas prestaciones tipo Web 2.0.</p> <p>Es importante porque ofrece la búsqueda más extensa de registros producidos por bibliotecas universitarias y por sus amplias y muy bien diseñadas prestaciones de tipo Web 2.0</p>	<p>OCLC es la organización más importante del mundo en servicios para bibliotecas universitarias.</p>
<p><b>Vista parcial página de resultados</b></p>	 <p>The screenshot shows the WorldCat search interface. At the top, there is a search bar with the text 'ernest abadal' and a search button. Below the search bar, it indicates 'Resultados de la búsqueda para 'ernest abadal'' and 'Resultados 1-10 de alrededor de 152 (.11 segundos)'. On the left side, there is a sidebar with filters for 'Autor', 'Formato', 'Año', and 'Contenido'. The main area displays three search results:</p> <ol style="list-style-type: none"> <li><b>Prensa digital y bibliotecas</b> por Ernest Abadal, Javier Guallar. Libro. Idioma: Español (spa). Editorial: Gijón : Trea, 2010.</li> <li><b>Investigación federada en el ecosistema de la e-ciencia: el caso SciVerse Research</b> por Lluís Codina; Ernest Abadal; Cristófol Rovina. Artículo. Idioma: Español (spa). Publicación: El Profesional de la información, 19, no. 1, (January 01, 2010): 77. Editorial: Barcelona : Sweets &amp; Zeitlinger / Elsevier, 1995-. Base de datos: ArticleFirst.</li> <li><b>Analysis of web-based tutorials created by academic libraries</b> por Marta Somoza-Fernández; Ernest Abadal. Artículo. Idioma: Inglés (eng). Publicación: Journal of academic librarianship, 35, no. 2, (2008): 126. Editorial: [New York, etc. : Elsevier Inc., etc.]. Base de datos: ArticleFirst.</li> </ol>

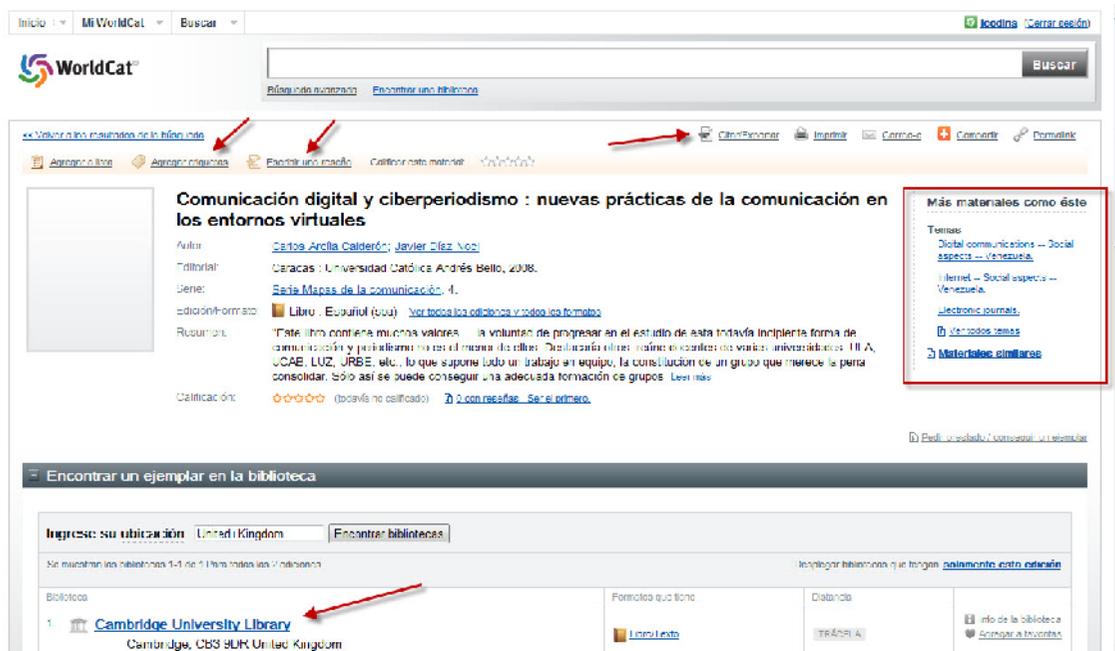


Ilustración 4: Opciones de la visualización de un registro de WorldCat

### **3. Otros recursos académicos**

La información de tipo académico, en cuanto a abundancia de recursos, está pasando por una excelente etapa. En la tabla siguiente se recogen otros buscadores y sistemas de información especializados, de creación más o menos reciente. Qué duda cabe que la selección natural actuará sobre ellos, de modo que algunas dejarán de existir en el medio plazo, pero sin duda algunos otros han llegado para quedarse, de modo que los miembros de la comunidad ACT tienen por ahora un panorama del que podemos decir que, por lo menos, es interesante.

**Tabla 1: Recursos recientes de información académica**

<b>Nombre / URL / Descripción</b>
<b>Q-Sensei</b> www.qsensei.com Buscador de la empresa del mismo nombre. Resultado de la fusión de una compañía alemana y norteamericana. Afirma indizar 24 millones de artículos y 11,5 millones de libros. Combina la indización y la búsqueda federada.
<b>Base</b> www.base-search.net Buscador tipo recolector desarrollado por la Biblioteca de la Universidad de Bielefeld (Alemania). Proporciona acceso a casi 24 millones de documentos procedentes de casi 1.600 fuentes de todo el mundo. Utiliza tecnología de la empresa FAST (adquirida por Microsoft), la misma que utiliza Scirus.
<b>Scitopia</b> www.scitopia.org Un sistema de búsqueda federada de la misma empresa creadora de Science Research. Está especializado en física, tecnología y patentes.
<b>Lumifi</b> www.lumifi.com Es una combinación de buscador, base de datos personal, sistema de análisis de la información y sistema colaborativo. Es el modelo de lo que podrían ser los auténticos sistemas de información académicos de la Web 2.0 (o de la 3.0). Su único problema es que para las búsquedas utiliza Gigablast, un motor sumamente modesto.

#### **4. Conclusiones**

Como es sabido, además de los sistemas que hemos presentado aquí y de Google Scholar y Scirus, disponemos de numerosos portales y bases de datos de editoriales académicas como los de Elsevier (Science Direct, Scopus) o Springer; de bases de datos de productores como Thomson (ISI), de bases de datos de distribuidores de bases de datos como Ebsco, etc.

Es decir, nosotros nos hemos limitado a presentar, en función de su novedad, aquí una parte *de una parte* de la oferta en recursos de información para la comunidad ACT que tiene como característica que se ofrece en abierto. Lo cierto es que acceder a grandes cantidades de información, y hacerlo a la vez de una forma selectiva, ya no es un gran problema.

Otra cosa es seleccionar en cada caso el mejor (o mejores sistemas), pero sobre todo, el problema es qué hacer con lo que encuentra el usuario de estos sistemas. En teoría hay también numerosas soluciones, desde sistemas comerciales como RefWorks hasta sistemas abiertos como Zotero o Mendeley, pero aún no parece estar bien resuelto ni mucho menos este lado del problema.

## 5. Referencias

Alonso-Arévalo, Julio. "Gestores de referencias sociales: la información científica en el entorno 2.0". En: T. Baiget (ed.) J. Guallar (coord.). *Anuario ThinkEPI 2010: Análisis de tendencias en información y documentación*, 2010, p. 275-279.

Cabezas, Álvaro, Daniel Torres, and Emilio Delgado. 2009. "Ciencia 2.0: Herramientas e implicaciones para la actividad investigadora". *El Profesional De La Información* 18 (1): 72-9.

Codina, Lluís. 2009. "Ciencia 2.0: Redes sociales y aplicaciones en línea para académicos". *Hipertext.Net* n. 7

Codina, Lluís, Ernest Abadal, and Cristòfol Rovira. 2010. "Búsqueda federada en el ecosistema de la e-ciencia: El caso science research". *El Profesional De La Información* 19 (1) (Enero-Febrero 2010): 77-85.

Codina, Lluís. *Ciencia 2.0: Diagrama interactivo y directorio de recursos*:  
<http://tinyurl.com/ciencia20directorio>

Rovira, Cristòfol, Mari-Carmen Marcos, and Lluís Codina. 2007. "Repositorios de publicaciones digitales de libre acceso en Europa: Análisis y valoración de la accesibilidad, posicionamiento web y calidad del código digital". *El Profesional De La Información* 16 (1) (Enero-Febrero 2007): p24.

---

Agradecimientos: Este trabajo ha sido financiado parcialmente por el proyecto de investigación del Ministerio de Ciencia e Innovación número CSO2009-13713-C05-04