

# Web dragons : inside the myths of search engine technology. Amsterdam : Morgan Kaufmann, 2006.

Autor:  
Arias, Olga M.

Revista:  
Información, Cultura y Sociedad

2012, N°27, pp. 127-130



Reseña

*Web dragons : inside the myths of search engine technology* / Ian Witten; Marco Gori y Teresa Numerico. Amsterdam: Morgan Kaufmann, 2006. 261 p. ISBN : 978-0-12-370609-6.

El primer párrafo de esta obra muestra con claridad la actualidad y la vigencia del tema que aborda. Dicen los autores: “en el brevísimo tiempo transcurrido desde que cambió el milenio, las vidas de aquéllos de nosotros que trabajamos con información se han visto enormemente transformadas: mucho, la mayoría, quizás todo, está en la web, si no hoy, entonces mañana. [...] Y lo que ha hecho esto posible no es la enorme cantidad de información disponible, es una fantástica nueva tecnología: los motores de búsqueda”. No es un libro que explique la manera de operar de estas herramientas de software, sino que apunta a establecer qué ideas –o ideales– se han ido dibujando a lo largo de su reciente progreso, y qué desafíos se les presentan a sus desarrolladores. El título en sí representa una metáfora que se refiere al poder y al misterio que encierra el término “dragón”. En la mitología oriental, son buenas pero poderosas criaturas que custodian inmensos tesoros (metafóricamente, todo –o casi todo– el conocimiento de la humanidad), mientras que en la creencia occidental a menudo representan el mal. Esta dualidad de significados es utilizada por los autores para trazar un paralelo: ellos nos permiten encontrar lo que buscamos en la web, pero los que lo usan confían demasiado en ellos e ignoran sus verdaderos poderes.

Ian Witten es Doctor en Ciencias de la Computación, y profesor e investigador de la Universidad de Waikato en Nueva Zelanda. Marco Gori es profesor de Ciencias de la Computación en la Universidad de Siena, y Teresa Numerico es profesora de teoría de números y comunicación en la Universidad de Roma. Después de una breve visita de Ian Witten a Italia, surgió el proyecto de escribir este libro, que se concretó cuando el autor realizó una visita más extendida a ese país. Si bien la fecha de edición de la obra podría considerarse algo antigua para este campo del conocimiento, tiene sin embargo el valor agregado de reunir una serie de conceptos básicos acerca del surgimiento y el crecimiento de la web, y –sobre todo– del desarrollo de los motores de búsqueda. Podríamos decir que, sin entrar en tecnicismos difíciles de captar por el público en general, ofrece, no obstante, una cantidad de información útil y sumamente clara sobre esta temática.

El libro está organizado en siete capítulos, cada uno de los cuales trata una temática específica, y que pueden ser leídos secuencialmente, o no. Al final de cada capítulo los autores proponen una serie de tareas a manera de ejercicio, y también ofrecen sugerencias interesantes sobre los temas tratados.

El capítulo 1 introduce el escenario de la revolución de la información desde una perspectiva histórica y filosófica. Los autores comienzan citando

un párrafo de “La biblioteca de Babel”, de Jorge Luis Borges, y a partir de allí establecen una similitud entre “el universo” de Borges y la WWW, con una característica común: la percepción de desorden y caos informativo. Desde allí apuntan hacia otros acontecimientos que fueron igualmente hitos para la humanidad, el Iluminismo y la Revolución Francesa. Pasando por varias importantes posturas filosóficas, destacan las de Platon, Pierce y Wittgenstein. En este sentido, se hace hincapié en que la web es central para nuestro pensamiento, y la manera en que opera recuerda la manera en que pensamos, vinculando piezas de información. Hay referencias a los decisivos aportes de Ted Nelson, un investigador de Harvard que acuñó el término *hiperterxto*, y a Tim Berners-Lee, que desarrolló el software que daría inicio a web. Se explica, además, cómo está creciendo cuantitativa y cualitativamente, subsumiendo inclusive a la literatura, que hasta ahora había sido, casi exclusivamente, baluarte de las bibliotecas.

El capítulo 2 presenta un pantallazo de la larga historia de las bibliotecas, en el que se menciona el nacimiento de las principales bibliotecas nacionales. En este punto, los autores hacen una interesante comparación entre el crecimiento exponencial de la cantidad de conocimiento en los libros, la creciente miniaturización de los componentes electrónicos, y lo que sería la materialización del sueño del bibliotecario de Alejandría: ya no hay límites físicos para las bibliotecas: los dispositivos de almacenamiento vinieron a resolver el problema. Se menciona al Proyecto Gutenberg y los proyectos del millón de libros de Estados Unidos, China e India, cuyo objetivo es crear colecciones abiertas de material en el dominio público. También se mencionan otros grandes proyectos como Internet Archive, y tiendas en línea, como Amazon, que está digitalizando las colecciones de grandes bibliotecas, permitiendo también efectuar búsquedas.

El capítulo 3 describe a la web en términos de tamaño, escala, construcción y organización. Se tratan temas que tienen que ver con conceptos básicos acerca de su estructura, su crecimiento, tamaño, conectividad, como así también lo que se entiende por “web profunda”, es decir, aquellas páginas que se construyen dinámicamente a partir de consultas a bases de datos. El capítulo se completa con una descripción breve pero clara de HTTP, HTML, URIs, URLs, cómo actúan los motores de búsqueda, explicaciones sobre “cookies”, páginas dinámicas y estáticas, wikis, blogs, etc.

El capítulo 4 aborda ideas relativas a la búsqueda en texto completo, a partir del relato de cómo se generó la primera concordancia, tarea llevada a cabo por Mary Novello, a mediados del siglo XVIII sobre las obras de Shakespeare. La indización en texto libre es la corporización de la clásica concordancia, con la ventaja de que estando computarizada, funciona para cualquier documento, independientemente de si se trata de un texto banal o de escrituras sagradas o piezas maestras de la literatura. Los autores explican con bastante detalle la manera en que operan los motores de búsqueda construyendo índices y clasificando los documentos recuperados por orden de relevancia.

El capítulo 5 trata del “lado oscuro” de la web, la labor de productores de “spam”. Se describen técnicas de spam y otras modalidades de abuso por parte de firmas inescrupulosas. El capítulo comienza trazando una comparación entre la web y un ecosistema, cuyo cuerpo –Internet– sufre inevitablemente problemas técnicos, y también fallos que invaden los grandes sistemas de software. Todas estas y otras amenazas al enorme sistema de información en que se convirtió la web son consideradas desde el punto de vista de la eventual desaparición de la web, algo ciertamente impensable. Pero se equipara esta posibilidad a la destrucción masiva de libros ocurrida por lo menos en dos oportunidades en la historia de la humanidad: el caso de la biblioteca de Alejandría, y la quema de libros en la Alemania pre-nazi, en 1933. Sin embargo, la resistencia de la web a los ataques y a las fallas está dada por su estructura de red, de manera muy similar a lo que sucede con el cerebro humano. Las redes neuronales artificiales soportan el fallo de uno o varios componentes, sin que por ello se vea amenazado su funcionamiento global. La última parte del capítulo está dedicada a cuestiones existentes detrás del negocio de la visibilidad de las páginas web, en términos de cómo se impulsa artificialmente la valoración de las páginas (“page rank”), las cuestiones económicas derivadas de los avisos comerciales en la web, y la potencial guerra desatada entre facilitadores de spam y compañías que se dedican a combatirlo. Hace un llamado de atención sobre el hecho de que es difícil combatir el spam legalmente, lo cual en última instancia plantea una cuestión de ética, particularmente si queremos convertir a la web finalmente en la “biblioteca interactiva”, un repositorio universal de todo el conocimiento humano.

El capítulo 6 trata de la forma en que se maneja el flujo de información en la web. En este sentido, los motores de búsqueda juegan un rol decisivo desde muchos puntos de vista. Son los mecanismos mediante los cuales las páginas aparecen antes en los resultados de búsqueda, lo que a su vez tiene no pocas implicancias con lo que, en términos de estas tecnologías, significan prestigio, autoridad y relevancia. También se aborda el tema de la privacidad, la censura y cómo las tecnologías web han impactado sobre el derecho de autor. El hecho de que la web hoy concentra mucho poder sobre la forma de buscar de los usuarios ha llevado a algunos a proponer que debería ser nacionalizada o “internacionalizada”, en utilidades de información pública. Pero los autores disienten por dos razones: primero, por su naturaleza apolítica y segundo, porque los potentes motores de búsqueda pudieron ser desarrollados a partir de la competencia comercial intensa, y solo en una década de actividad.

El capítulo 7 aborda el aspecto social de la web y el surgimiento de comunidades en su interior, entendido este concepto en un sentido amplio, íntimamente ligado al carácter permanentemente cambiante y creciente de la web. Así, se analizan las posibilidades aportadas por el surgimiento de la arquitectura “peer-to-peer”, y se explican más o menos en detalle, entre otros, desarrollos tales como sistemas de reputación, de marcado social, y de trabajo colaborativo,

voluntario, cuyo ejemplo más destacable es Wikipedia. En los últimos párrafos se explica brevemente el futuro de las bibliotecas digitales, y el surgimiento de las aplicaciones de ofimática que ofrecen actualmente los principales motores de búsqueda. La reflexión final de los autores es interesante y provocativa: no podemos predecir cómo van a evolucionar los motores de búsqueda, de cuyo desarrollo trató el libro a través de sus capítulos, pero una cosa es cierta: ya no apuntan solo a las búsquedas, van a cambiar la forma misma en que trabajamos y actuamos.

OLGA M. ARIAS  
Biblioteca Luis F. Leloir  
Facultad de Ciencias Exactas y Naturales – UBA

*Historia de la lectura y la escritura en el mundo occidental* / Martins Lyons. Buenos Aires: Editoras del Calderón, 2012. 432 p. (Colección Scripta manent). ISBN 978-987-24602-2-8.

Este interesante libro de Lyons en la estética y cómoda edición de Editoras del Calderón, es una obra imprescindible para introducirse en el conocimiento y la discusión de la historia de la lectura, del libro y de la escritura. El mismo autor aclara que la historia de la lectura tendrá más espacio e interpretación que la de la escritura. Quizás esto se deba a que algunos momentos históricos tienen más interés en la lectura que otros o “porque refleja simplemente el estado actual de las investigaciones académicas”.

Este trabajado entrecruzamiento obliga a toda la investigación asociada, a volver sobre sus pasos y repensar la estrecha relación entre lectura y escritura casi como si de dos caras de una misma moneda textualizada se tratara. Ambas se relacionan en la demostración del autor de una manera íntima y directa, casi recíproca en plenitud comprensiva. La una encierra a la otra y cada cual se sustenta en el mismo sustrato cultural de legitimación e interpretación.

La mirada epistemológica e ideológica es contundentemente clara: “Todas las sociedades desde el antiguo Egipto han sido una “sociedad de la información”, en el sentido de que quienes controlan y restringen el acceso al conocimiento en cualquier sociedad controlan así un componente clave del poder. La relación entre conocimiento, cultura impresa y educación son los pilares de