

D.2. Información = Matemáticas

Por José Ramón Pérez Agüera

Pérez Agüera, José Ramón.

“Información = Matemáticas”. En: *Anuario ThinkEPI*, 2007, pp. 127-129.



“Los temas documentales más en boga como son los buscadores, data-mining, datawarehouse, web semántica, ontologías, cienciometría..., tienen un sólida base matemática”

“¿Puede el gestor de información desconocer la naturaleza matemática de la misma?”

EL PLENO SIGLO XVIII Michael Faraday sentó las bases del electromagnetismo a partir de muchas horas de laboratorio. Desgraciadamente, este famoso científico carecía de conoci-



Michael Faraday.



James Maxwell.

mientos matemáticos suficientes como para formalizar sus observaciones de manera que pudiesen ser expresadas en términos matemáticos. Hizo falta esperar 100

años para que otro eminente científico, James Maxwell, formalizara las observaciones de Faraday en sus ya famosas ecuaciones del campo electromagnético. El desconocimiento matemático de Faraday le costó, como poco, la tediosa tarea de describir en lenguaje natural sus conclusiones, con la consiguiente ambigüedad que esto conlleva, además del gasto desmesurado de papel, y con el consiguiente retraso que, hasta cierto punto, esto supuso para su disciplina.

Me disculpará el lector este comienzo tan poco usual, pero teniendo en cuenta la pertinencia del ejemplo con el tema que aquí nos ocupa, junto con el hecho de que cuando escribo esto nos encontramos en el Año Mundial de la Física, he pensado que era una buena forma de comenzar esta mini-reflexión sobre la importancia de las Matemáticas en la Documentación.

Y es que desde mi perspectiva, siempre parcial, de lo que es la Documentación hoy en día, creo que gran parte de los documentalistas se encuentran en una situación muy similar a la del afamado científico inglés al que hacíamos referencia en la introducción, lo cual no deja de ser curioso y hasta cierto punto paradójico, si seguimos un poco la pista al desarrollo de la Ciencia de la Información en los últimos 60 años.

Hace ya un buen de décadas, Shannon definió matemáticamente el concepto de información gracias a formalismos matemáticos (recordemos que definió el bit, la unidad elemental de información). Investigaciones posteriores como la de Dekang Lin dejan claro que la propuesta de Shannon iba más allá de la mera transmisión física de información. Algunos años después, Salton formalizó una colección documental gracias a un espacio vectorial n-dimensional, donde cada documento se correspondía con un vector. Hoy en día, los tesoros, una herramienta

típicamente documental, se pueden representar como un grafo, el cual está definido mediante una función matemática, o como una matriz, donde cada término corresponde a un vector cuyas componentes pueden ser documentos u otros términos, de cara a establecer las relaciones que existen dentro del tesoro a partir de cálculos de similitud entre los vectores que representan a los términos que forman parte del tesoro.

Los ejemplos de tratamiento matemático de la información y el conocimiento no se detienen aquí e inundan nuestro área de actividad miremos donde miremos. Uno de los temas más recurrentes entre la comunidad de documentalistas hoy en día, como son los buscadores de internet, tiene una sólida base matemática, centrada en operaciones probabilísticas de cierta complejidad como son el *teorema de Bayes* para probabilidad condicionada y los *modelos de Markov*. El análisis de información en bases de datos, léase *datamining* y *datawarehouse*, no pueden ser entendidos sin conocimientos de álgebra. La Web semántica y las ontologías, otro de los temas en boga, no pueden entenderse en profundidad sin conocimientos sólidos de álgebra, matemática discreta y lógica matemática (en particular lógica descriptiva si nos centramos en ontologías).

Peligroso rechazo a las matemáticas

A la vista de estos ejemplos ¿puede el gestor de información desconocer la naturaleza matemática de la misma?

Desde mi perspectiva, todos los elementos con los que trabaja el documentalista y/o gestor de información hoy en día pueden y deben ser representados en función de abstracciones y formalismos matemáticos, y no sólo esto, sino que la formalización de estos elementos ayuda enormemente a su tratamiento, hasta el punto de mejorar en muchos aspectos las tareas de gestión de información. Pese a esto, se encuentra un rechazo innato, en muchos casos incomprensible, a la inclusión real de las matemáticas en la Documentación. Los planes de estudio son un buen ejemplo de esto, pues son escasas las asignaturas referentes a los aspectos matemáticos de la información, y cuando se



Richard Feynman.

encuentran, los contenidos no son quizá todo lo pertinentes que deberían.

Sin duda existe una buena parte de científicos de nuestro área que dominan perfectamente el lenguaje matemático, pero mi reflexión se centra más en un sentir general que, creo

yo, impera en la comunidad de científicos y profesionales de la Ciencia de la Información. Este sentimiento tiene su origen en nuestra herencia archivística y bibliotecaria, pero a la larga puede suponer una rémora que perjudique profundamente a las futuras generaciones de gestores de la información que se están formando ahora. Hemos de ser conscientes que existen asignaturas como "Estructura de datos" y "Estructura de la información", que no son impartidas en facultades de Documentación sino, en facultades de Informática.

Este enfoque cientifista-matemático de la Documentación no será del agrado de muchos, pero dado que nos encontramos en plena Sociedad de la Información no estaría de más reflexionar sobre este hecho de cara a adaptar nuestra profesión a las necesidades de la sociedad. Creo que es hora de olvidarse un poco de **Otlet** y **La Fontaine** y centrarse más en **Shannon**, **Feynman**, **Salton**, **Sparck-Jones**, **Zipf**, e incluso en **Cutter**, si queremos retrotraernos al apogeo positivista en el que se desarrolló esta disciplina.

Muchos se preguntarán que en base a esta reflexión se difumina la diferencia entre informáticos y documentalistas, como si las fronteras entre disciplinas se pudieran marcar tan claramente como para decir que un conocimiento es



Gerard Salton.

Documentación y otro es Informática. Si esto fuera así deberíamos dejar de hablar de web

semántica, de recuperación de información, de bases de datos, etc.

No creo que este sea el caso, y opino que la Documentación tiene mucho que aportar al resto de ciencias que trabajan con información, y que forman la verdadera Ciencia de la Información. Eso sí, creo que debemos jugar con todas las reglas del juego, inclui-



Claude Shannon.

das las matemáticas, ya que de esa forma nuestra aportación será mucho más palpable y no tendremos que esperar 100 años a que alguien formalice matemáticamente lo que nosotros ya demostramos en base a un empirismo centrado en la práctica documental.

José Ramón Pérez Agüera, Universidad Complutense de Madrid, Facultad de Informática, Departamento de Ingeniería del Software e Inteligencia Artificial

jose.aguera@fdi.ucm.es

http://multidoc.rediris.es/joseramon

Lo que otras disciplinas pueden aportar a la documentación

Por Jorge Serrano Cobos



Aprovecho la nota de **José Ramón** (y otras sobre tags, folksonomies, sociología, **Luhmann**, etc.) para decir que todos esos temas son muy interesantes.

Tengo un compañero de trabajo que es físico, y con él he aprendido la entropía de la información:

<http://club.telepolis.com/ohcop/bb7.html>

y la teoría cuántica de la información:

http://www.tendencias21.net/index.php?action=article&id_article=67988

Todo es ponerse las pilas con el tema a ver qué nos aporta.

La cuestión es aprender qué podemos aplicar de forma práctica de esas disciplinas

en nuestro trabajo y luego, como dice **José Ramón**, qué podemos aportarles desde la Documentación (como han hecho **Rosenfeld** y **Morville** en el campo del diseño web, por ejemplo).

Lo único que necesitaríamos es que se tenga en cuenta que muchos lo miraremos desde la barrera, y que no habremos hecho carreras específicas de cada temática (hacen falta unos cuantos clones, vaya) ni tenemos mucho tiempo. Pero el saber, al menos en *Gmail*, no ocupa –tanto– lugar.

Jorge Serrano Cobos

jorgeserrano@gmail.com

http://www.masmedios.com

http://trucosdegoogle.blogspot.com