

CAPÍTULO 8

LA RECUPERACIÓN DE INFORMACIÓN: LA REPRESENTACIÓN DE LO SUBJETIVO

EDBERTO FERNEDA
Universidade Estadual Paulista

RESUMEN

La recuperación de información designa la operación mediante la cual se seleccionan los documentos de una colección en función de una demanda determinada de información. Se recupera un documento si su representación coincide total o parcialmente con la representación de la necesidad del usuario. La correcta interpretación de tales representaciones es fundamental para la eficiencia de un sistema de información, que involucra procesos cuya formalización y automatización solo son posibles por medio de simplificaciones de conceptos típicamente subjetivos. Estas simplificaciones afectan directamente en la eficiencia de los sistemas de información. Este trabajo presenta y evalúa las formas de representación computacional de los conceptos y operaciones que hacen parte del proceso de recuperación de información. La automatización del proceso de recuperación de información viabiliza la operación de grandes cantidades de datos de forma rápida y ágil, pero no necesariamente proporciona resultados consistentes o satisfactorios. El juicio de relevancia de la información implica procedimientos basados en las capacidades y habilidades humanas de abstracción, aprehensión y representación de su significación.

Palabras clave: recuperación de información, representación de la información, significación de la información, relevancia, subjetividad.

ABSTRACT

Information retrieval designates the operation by which documents are selected from a collection based on a specific informational demand. A document is retrieved if its representation totally or partially coincides with the representation of the user's need.

The correct interpretation of such representations is fundamental for the efficiency of an information system, which involves processes whose formalization and automation are only possible through simplifications of typically subjective concepts. These simplifications directly or indirectly affect the efficiency of information systems. This work presents and evaluates the forms of computational representation of concepts and operations that are part of the information retrieval process. The automation of the information retrieval process makes it possible to operate large amounts of data in a fast and agile way, but it does not necessarily provide consistent or satisfactory results. Judging the relevance of information implies procedures based on human capacities and abilities of abstraction, apprehension and representation of its meaning.

Keywords: information retrieval, information representation, meaning of information, relevance, subjectivity.

INTRODUCCIÓN

LA BÚSQUEDA POR INFORMACIONES utilizando algún tipo de recurso tecnológico se convirtió en una actividad común en la sociedad contemporánea. Al hacer una búsqueda en la web estamos buscando por una información que venga a satisfacer una determinada necesidad. Una información es considerada relevante si ella trae el conocimiento que necesitamos en un determinado momento, en una determinada situación.

La idea de la utilización de dispositivos electrónicos en la búsqueda por información tuvo su génesis con los trabajos de Paul Otlet (1934) y posteriormente con el artículo intitolado “As We May Think”, de Vannevar Bush (1945). Las ideas vehiculizadas en esos trabajos abrieron camino para diversas investigaciones en las décadas siguientes. En el inicio de los años de 1950, el matemático, físico y científico de la computación Calvin Northrup Mooers (1951) creó el término “Information Retrieval”, inaugurando un área de la investigación que se consolidó y se fortaleció a lo largo de los años. La popularización de la Internet y el surgimiento de la web trajeron nuevos desafíos y gran interés en la investigación y en el desarrollo de técnicas para auxiliar la búsqueda y recuperación de información en esa colección mundial.

Desde las primeras investigaciones hasta los días actuales, el papel de los sistemas de recuperación de información pasó de simple herramienta experimental a sistemas de uso cotidiano, útiles a todos que necesiten la información para sus actividades. En ese periodo, el acelerado avance tecnológico e innumerables ideas, conceptos y técnicas fueron propuestos y desarrollados. Sin embargo, la búsqueda de informaciones relevantes y útiles es todavía una tarea ardua. Esa dificultad lleva a la reflexión sobre los principales elementos involucrados en el proceso de recuperación de información que aparentemente son ajenos a los avances tecnológicos, o por lo menos a las tecnologías actualmente disponibles (Ferneda, 2013).

La recuperación de información es la operación por la cual se selecciona documentos de una colección en función de una determinada demanda informacional. En su esencia, la recuperación se da por medio de la comparación entre las representaciones de los documentos y de la representación de la necesidad de información del usuario. Un documento es recuperado si su representación coincide total o parcialmente con la representación de la necesidad del usuario. Recuperar información implica, por lo tanto, operar selectivamente sobre un conjunto de ítems de información, lo que involucra procesos cuya formalización y automatización solamente son posibles por medio de simplificaciones de conceptos típicamente subjetivos (Ferneda, 2003).

El objetivo de este texto es evaluar las formas de representación computacional de los conceptos y operaciones inherentemente subjetivos que hacen parte del proceso de recuperación de información.

LA RELEVANCIA DE LA RELEVANCIA

El concepto de relevancia es crucial en la Recuperación de Información, siendo muchas veces utilizado en la propia enunciación de los objetivos de esa área. Es una cuestión fundamental y una preocupación central para el funcionamiento y evaluación de los sistemas de recuperación de información (Saracevic, 2017; Mizzaro, 1997; Cooper, 1971).

El término “relevancia” es generalmente utilizado para identificar un elemento que sobresale o se destaca en un determinado conjunto. Es utilizado también para distinguir un objeto “de gran valor o interés” o incluso hacer referencia “a aquello que importa o es necesario”¹.

El concepto de relevancia puede ser expresado por diferentes términos. Vannevar Bush (1945) utilizó la expresión “ítem de importancia momentánea”; Mooers (1951) se refería a “informaciones útiles”. Términos como “pertinente”, “valioso”, “útil”, “significativo” son utilizados en connotaciones diferentes, pero generalmente con significado subyacente a la relevancia.

Como la mayoría de las nociones fundamentales, la relevancia es intuitivamente bien comprendida - nadie necesita explicarla para alguien en el mundo. Esa es su fuerza. Es por eso que los sistemas que tienen por objetivo la recuperación de informaciones relevantes para los usuarios, incluyendo mecanismos de búsqueda y una variedad de aplicaciones de investigación en los medios sociales, son muy bien aceptados globalmente – las diferencias en culturas, sociedades y costumbres no interesan. Sin embargo, la relevancia es una noción humana, no técnica. Esa es su debilidad. Como todas las nociones humanas, la relevancia es confusa. La relevancia engloba muchas variables que son difíciles de controlar e incluso comprender formalmente. La relevancia siempre, repito siempre, involucra también un contexto. Todos los algoritmos de búsqueda en todos los sistemas del mundo están intentando aproximar,

1

Ferreira, A. B. de H. Novo Dicionário da Língua Portuguesa.

con varios grados de éxito, la noción humana de la relevancia. Es para eso que sirven, es por eso que existen². (Saracevic, 2015, p. 27, traducción propia).

La relevancia siempre involucra una relación. Siempre existe un “para” asociado a la relevancia que se refiere a un contexto, un asunto en cuestión. Algo é relevante *para* alguien o *para* un determinado contexto. El concepto de relevancia no es necesariamente binario, existen gradaciones que se modifican en la medida que las intenciones y los horizontes cognitivos cambian, o cuando el asunto en cuestión es alterado (Saracevic, 2017, p. 17). Según Sperber y Wilson (2005, p. 224, traducción propia), “intuitivamente, relevancia no es una cuestión de todo o nada, sino una cuestión de grados”. La atribución de esos grados de relevancia es un proceso inherentemente subjetivo.

Saracevic (1975), Swanson (1986) y Harter (1992) distinguen dos tipos de relevancia: “relevancia objetiva” y “relevancia subjetiva”. La *relevancia objetiva* está relacionada a los sistemas, en cuanto a la *relevancia subjetiva* se relaciona a la operación y uso de tales sistemas por sus usuarios. Según Swanson (1986), en un sistema de recuperación de información siempre cabrá al usuario juzgar la relevancia de las informaciones resultantes de una búsqueda. Ese arbitraje tiene carácter individual, una “experiencia mental” basada en las características de cada usuario. Saracevic (2017, p. 24) argumenta que los sistemas son creados por diferentes *designers*, que utilizan diferentes enfoques y diferentes métodos de desarrollo. Así, de alguna manera los sistemas también son subjetivos. Por lo tanto, según el autor, no existe relevancia “objetiva”. Toda relevancia es subjetiva, aun cuando formalizada en un algoritmo.

De forma similar a la clasificación dicotómica objetiva/subjetiva, diversos autores utilizan la terminología “relevancia del sistema” y “relevancia del usuario” (Mizzaro, 1997). La *relevancia del sistema* es una relevancia potencial supuesta, definida y formalizada a partir de hipótesis o conjeturas relacionadas con la estructura de representación de los ítems de información, la forma de organización de esos ítems y el grado de similitud de cada ítem en relación a la expresión de búsqueda. En el entorno web, por ejemplo, considerando su estructura formada por un conjunto de páginas conectadas por enlaces, el algoritmo básico de Google (PageRank) parte de la idea que la cantidad de enlaces que una página web recibe

2 *As most fundamental notions, relevance is intuitively well understood – nobody has to explain it to anybody in the world. That is its strength. That is why the systems aiming at retrieval of relevant information to users, including search engines and a variety of search apps in social media, are so well accepted globally – differences in cultures, societies, and mores do not matter. However, relevance is a human, not a technical, notion. That is its weakness. As all human notions, relevance is messy. Relevance encompasses many variables that are hard to control and even fathom formally. Relevance always, repeat always, involves a context as well. All the search algorithms in all the systems in the world are trying to approximate, with various degrees of success, the human notion of relevance. That is what they are all about, that is why they exist.*

de otras páginas puede servir como una medida de su relevancia (Brin & Page, 2012). Los sistemas bibliotecarios utilizan criterios de relevancia adaptados a la estructura de representación de los ítems de su colección. El sistema Primo³ clasifica los resultados de una búsqueda con base en los siguientes criterios de relevancia (Ex Libris, 2015):

- 1. El grado de correspondencia de un ítem en relación a la consulta.** Un ítem es considerado más relevante si los términos de la consulta ocurren en campos de metadatos específicos del registro del ítem (autor, título, asunto) y si los términos del registro aparecen en la misma secuencia de la consulta;
- 2. La importancia académica de un ítem.** La significancia académica del ítem es calculada a partir de factores no relacionados a la consulta. Para calcular la importancia académica de un ítem es llevado en consideración si fue publicado en una revista con revisión por pares, números de citas, entre otras características;
- 3. La relevancia de un ítem para el tipo de búsqueda.** El sistema infiere el tipo de búsqueda que el usuario está conduciendo. En una búsqueda por un tema amplio o genérico el sistema añade en sus resultados artículos de referencia. En investigaciones por ítems más específicos, el sistema considera autores, títulos u otras características para colocar algunos ítems en lo alto de la lista de resultados.
- 4. La actualidad de un ítem.** Se supone que los usuarios generalmente prefieren materiales recientes.

Un sistema de recuperación atribuye relevancia siguiendo criterios formalizados por sus algoritmos. Esos algoritmos tienen la función principal de comparar la representación de cada documento del acervo con la expresión de búsqueda enunciada por el usuario. El resultado de esa comparación es un valor numérico que representa el grado de relevancia de cada documento en relación a la búsqueda. Ese grado de relevancia es generalmente utilizado para ordenar (*hacer un ranking*) el conjunto de documentos resultante de una búsqueda.

A partir del conjunto de documentos resultantes de su búsqueda, el usuario juzga la relevancia de los ítems recuperados (*relevancia del usuario*) utilizando su conocimiento sobre el asunto investigado. Borlund (2003) argumenta que la relevancia es un concepto cognitivo multidimensional cuyo significado es ampliamente dependiente de las percepciones de los usuarios y de sus necesidades. El juicio de relevancia del usuario es delimitado inicialmente por su necesidad de información. Sin embargo, la importancia dada a ciertas dimensiones de relevancia

3 Ex Libris Primo es un conjunto de herramientas desarrollado y comercializado por la empresa Ex-Libris que implementan recursos de búsqueda y recuperación en acervos de objetivos digitales.

puede cambiar dinámicamente, conforme el usuario avance en el análisis de los resultados.

El concepto de relevancia ha desempeñado un papel importante en el desarrollo de los sistemas de recuperaciones de información. Si la eficacia de un sistema está en su capacidad en recuperar documentos relevantes, esa eficacia puede ser medida por la proximidad entre la relevancia del sistema y la relevancia del usuario. La relevancia del sistema puede ser formalizada utilizando características relacionadas a la organización y a la representación de los ítems de un acervo. Sin embargo, la relevancia del usuario escapa a cualquier tipo de formalización o representación.

LA INSUFICIENCIA DE LA REPRESENTACIÓN

... En aquel Imperio, el Arte de la Cartografía logró tal Perfección que el mapa de una sola Provincia ocupaba toda una Ciudad, y el mapa del imperio, toda una Provincia. Con el tiempo, esos Mapas Desmesurados no satisficieron y los colegios de Cartógrafos levantaron un Mapa del imperio, que tenía el tamaño del imperio y coincidía puntualmente con él. Menos adictas al Estudio de la Cartografía, las Generaciones Sigüientes entendieron que ese dilatado Mapa era inútil y no sin Impiedad lo entregaron a las Inclemencias del Sol y de los Inviernos. En los desiertos del Oeste perduran despedazadas Ruinas del Mapa, habitadas por Animales y por Mendigos; en todo el País no hay otra reliquia de las Disciplinas Geográficas. (Suárez Miranda, Viajes de varones prudentes, libro cuarto, cap. XLV, Lérida, 1658).

Jorge Luis Borges – del rigor en la ciencia

Toda representación es incompleta. Si no fuera incompleta no sería una representación. Una representación es comúnmente más breve o menor que el objeto representado, restringiéndose a las características consideradas más relevantes. Por lo tanto, la creación de una representación involucra elecciones sobre lo que a ella le será incluido y lo que le será excluido. Algo del original es siempre perdido. Una representación siempre será una versión distorsionada del real, aunque sea por su insuficiencia (Saracevic, 1991).

El proceso de recuperación de información involucra dos instancias de representación: la representación de cada ítem de información de una determinada colección y la representación de la necesidad de información del usuario por medio de una expresión de búsqueda (consulta). Según Belkin, Oddy y Brooks (1982a), esas dos representaciones son de naturaleza distintas. Un ítem de información (documento) es la representación de un “estado de conocimiento

coherente”, mientras que una consulta es la representación de un “estado anómalo de conocimiento”. Hay situaciones en que el usuario es capaz de especificar exactamente qué informaciones son necesarias para resolver un determinado problema. Sin embargo, la situación más común es aquella en que el usuario no tiene conocimiento previo de la información que necesita, ni tampoco consigue formalizarla en una expresión de búsqueda.

La representación de un documento incluye los elementos descriptivos que lo identifican y lo caracterizan en un acervo, así como los elementos indicativos de su contenido informativo. La figura 1 ilustra los elementos del proceso de representación documental definidos por Mizzaro (1998).

Figura 1.

Representación del proceso de representación de un documento



Fuente: elaborado por el autor.

Un *documento* es la representación física de un conocimiento, la materialización de una *información*. Es la entidad que el usuario de un sistema de recuperación de información obtiene en respuesta a su búsqueda. El “sustituto” (*surrogate*) es la representación del documento, es constituido de elementos que lo distinguen entre los demás ítems del acervo. Es el elemento que será comparado con la expresión de búsqueda, responsable por la recuperación del documento. Mizzaro (1998) ordena esos tres elementos de la siguiente forma:

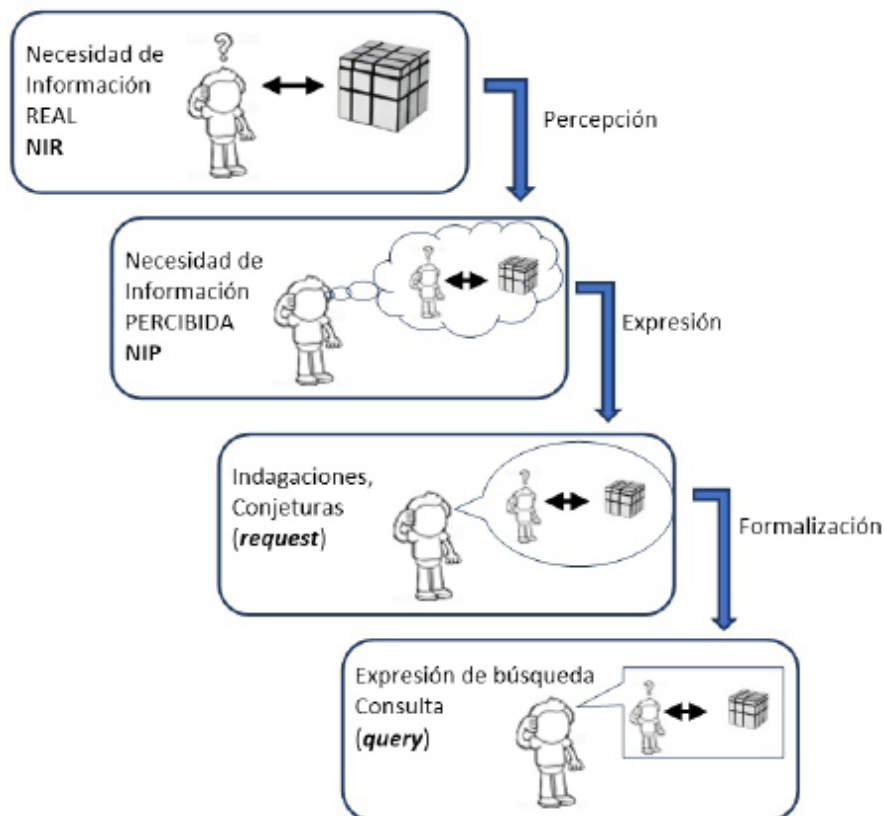
$$\textit{sustituto} < \textit{documento} < \textit{información}$$

En un sistema de recuperación de información, el acervo documental es constituido *a priori*, siendo susceptible de ser procesado por técnicas automatizadas tales como indexación automática, minería de textos, entre otras. Por otro lado, la necesidad de información del usuario solo se percibe tras su enunciación por medio de una expresión de búsqueda y su interpretación es dificultada por el número reducido de términos que normalmente es utilizado. Sin embargo, a partir de su definición, la expresión de búsqueda puede ser utilizada en procesos interactivos que visan resolver posibles ambigüedades o que permitan su enriquecimiento semántico (Pansani, 2021).

La figura 2 ilustra el proceso de representación de una búsqueda definido por Mizzaro (1998). Posee cuatro entidades – *Necesidad de Información Real (NIR)*; *Necesidad de Información Percibida (NIP)*, *Indagaciones o conjeturas (request)* y *Expresión de búsqueda o consulta (query)* – y tres operaciones: *Percepción*, *Expresión* y *Formalización*.

Figura 2.

Representación del proceso de búsqueda por información



Fuente: adaptado de Mizzaro (1998).

Según Belkin *et al.* (1982a; 1982b), una necesidad de información surge a partir de una reconocida anomalía en el estado de conocimiento del usuario sobre algún asunto o situación problemática de la cual no consigue especificar con precisión lo que es necesario para resolverla. Mizzaro (1998) denomina esa necesidad inicial de *Necesidad de Información Real* (NIR). El usuario percibe su necesidad y construye una representación mental, posiblemente incompleta o incorrecta en relación a la NIR: la *Necesidad de Información Percibida* (NIP). Por medio de conjeturas o indagaciones (*request*), el usuario expresa su necesidad en un lenguaje humano, un lenguaje natural. Finalmente, el usuario formaliza sus indagaciones en una consulta (*query*) utilizando el lenguaje proporcionado por el sistema de recuperación de información. A cada nivel de representación ocurre una pérdida o una distorsión en relación al nivel anterior. Los elementos constituyentes de ese proceso pueden ser ordenados de la siguiente forma (Mizzaro, 1998):

$$query < request < NIP < NIR$$

Por lo tanto, la consulta (*query*) es la materialización lingüística posiblemente incompleta de una necesidad de información, tras una secuencia de representaciones mentales.

En esencia, el proceso de recuperación de información se efectúa por la comparación entre representaciones: la representación de la necesidad de información del usuario y la representación de cada documento de una colección. El resultado de esa comparación será generalmente un número que representa el grado de relevancia del documento en relación a la búsqueda e irá posicionarlo en el listado de resultados.

IMPRECISIÓN MATEMÁTICA

[...]

- De acuerdo - dijo Pensamiento Profundo -. La Respuesta a la Gran Pregunta...

- ¡Sí...!

- de la Vida, del Universo y de Todo... - dijo Pensamiento Profundo.

- ¡Sí...!

- Es - dijo Pensamiento Profundo, haciendo una pausa.

- ¡Sí!

- Es...

- ¡¡¡¿Sí...?!!!

- Cuarenta y dos - dijo Pensamiento Profundo, con calma y majestad infinitas.

Douglas Adams – Guía del autoestopista galáctico

Los primeros ordenadores pesaban varias toneladas y su programación era hecha por medio de la conexión directa de sus circuitos. En la década de 1950 la programación era hecha a través de la transmisión de instrucciones en código binario por medio de tarjetas o cintas perforadas. Con el surgimiento de los lenguajes de programación, el código binario quedó restringido al núcleo del ordenador y la comunicación con el mundo externo era hecha por una nueva capa de programa. “Aquello que ayer fuera interfaz se convierte en órgano interno” (Lévy, 1993, p. 101, traducción propia). Actualmente los ordenadores son constituidos por un conjunto de dispositivos y capas de programas que se comunican unas con las otras, permitiendo gran distanciamiento de su núcleo binario.

¿Binaria, la informática? Sin duda, en un cierto nivel de funcionamiento de sus circuitos, pero hace tiempo que la mayoría de los usuarios ya no tiene cualquier relación con esta interfaz. ¿En qué, un programa de hipertexto o de *design* es “binario”? (Lévy, 1993, p. 102, traducción propia).

En respuesta a la cuestión colocada por Pierre Lévy, podemos confirmar que actualmente utilizamos un ordenador sin necesitar que tengamos conocimiento del funcionamiento de sus circuitos, así como utilizamos cualquier otro aparato electrónico. Sin embargo, el alma binaria de un ordenador atraviesa todas sus capas de programas y limita su capacidad de efectuar tareas que la mayoría de los seres humanos hacen con relativa facilidad.

En el proceso de recuperación de información, recursos computacionales viabilizan la operación de grandes acervos documentales, como es el caso de la web. Sin embargo, la naturaleza de los ordenadores exige la matematización de conceptos y procesos típicamente subjetivos. La relevancia, ahora destituida de su subjetividad, pasa a ser un número. La estrategia primaria de la automatización del proceso de representación documental (indexación) es el simple conteo de palabras. Las palabras con mayor número de ocurrencias en la superficie textual de un documento son electas como representantes de su contenido intelectual. La necesidad de información es representada por un conjunto de palabras destituidas de sus significados.

La automatización del proceso de recuperación de información impone una lógica en la cual la información debe ser numéricamente definida en el interior de un sistema cerrado, que desconsidera algunos factores humanos involucrados en ese proceso.

CONSIDERACIONES FINALES

El término “subjetivo” es definido como el “que pertenece al sujeto pensante y a su íntimo”; “pertinente a lo característico de un individuo; individual, personal, particular”²⁴. Subjetivo es todo aquello que es propio del sujeto o a él relativo. Es algo que está basado en una interpretación individual.

Los conceptos involucrados en el proceso de recuperación de información son típicamente subjetivos. El desarrollo de sistemas computacionales exige simplificaciones de tales conceptos para que sea posible formalizarlos y representarlos por medio de algoritmos y programas. Esas simplificaciones afectan directa o indirectamente la eficiencia de los sistemas de información. Se observa que gran parte de las investigaciones en Recuperación de Información está concentrada en la búsqueda de formas más eficientes de representar la subjetividad involucrada en ese proceso.

La automatización del proceso de recuperación de información viabiliza la operación de grandes cantidades de datos de una forma rápida y ágil. Sin embargo, no necesariamente proporciona resultados consistentes o satisfactorios. La información, tomada por su significado de sentido común, está directamente relacionada a su significado, lo que implica procedimientos basados en las capacidades y habilidades humanas de abstracción, aprehensión y representación de su significación.

REFERENCIAS

- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982a). ASK for Information Retrieval: Part I. Background and Theory. *Journal of Documentation*, 38(2), 61-71.
- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982b). ASK for information retrieval: Part II. Results of a design study. *Journal of Documentation*, 38(3), 145-164.
- Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology*, 54(10), 913–925.
- Brin, S. & Page, L. (2012). Reprint of: The anatomy of large-scale hypertextual Web search engine. *Computer Networks*, 56(18), 3825–3833.
- Bush, V. (1945). As We May Think. *Atlantic Monthly*, 176, 101–108.
- Cooper, W.S. (1971). A Definition of Relevance for Information Retrieval. *Information Storage and Retrieval*, 7, 19-37.

- Ferneda, E. (2003). *Recuperação da Informação: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação* (Tesis doctoral). Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, SP, Brasil.
- Ferneda, E. (2013). *Ontologia como recurso de padronização terminológica em um Sistema de Recuperação de Informação* (Relatório de Pós-Doutorado). Programa de Pós-Graduação em Ciência da Informação, Universidade Federal da Paraíba. João Pessoa, PB, Brasil.
- Ex Libris. (2015). Primo Discovery: Search, Ranking, and Beyond.
- Harter, S. P. (1992). Psychological relevance and information science. *Journal of the American Society for Information Science*, 53(4), 257–270.
- Mizzaro, S. (1997). Relevance: The Whole History. *Journal of the American Society for Information Science*, 48(9), 810-832.
- Mizzaro, S. (1998). How many relevances in Information Retrieval? *Interacting with Computers*, 10(3), 303–320.
- Mooers, C. N. (1951). Zatoncoding applied to mechanical Organization of Knowledge. *American Documentation*, 2(1), 20-32.
- Otlet, P. (1934). *Traité de documentation: le livre sur le livre - théorie et pratique*. Bruxelles: Mundaneum.
- Pansani, E. A., Jr. (2021). *Contextualização e Expansão de Consultas em Sistemas de Recuperação de Informação: um método baseado em ontologias de domínio* (Tesis doctoral). Faculdade de Filosofia e Ciências, Universidade Estadual Paulista-UNESP, Marília, SP, Brasil.
- Lévy, P. (1993) *As tecnologias da Inteligência: o futuro do pensamento na era da informática*. São Paulo: Editora 34.
- Saracevic, T. (1975). Relevance: A review of and framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, 26(6), 321-343.
- Saracevic, T. (2017). *The notion of relevance in Information Science: everybody knows what relevance is. But, what is it really?* Williston: Morgan & Claypool.
- Saracevic, T. (1991). Information science: origin, evolution and relations. En International Conference on Conceptions of Library and Information Science: historical, empirical and theoretical perspectives, 1991, Finland. Proceedings... Helsink.
- Sperber, D., & Wilson, D. (2005). Teoria da relevância. *Linguagem em (Dis)curso* 5, 221-268.
- Swanson, D. R. (1986). Subjective versus objective relevance in bibliographic retrieval systems. *Library Quarterly*, 56(4), 389-398.