



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN BIBLIOTECOLOGÍA Y ESTUDIOS DE LA INFORMACIÓN

FACULTAD DE FILOSOFÍA Y LETRAS

**INSTITUTO DE INVESTIGACIONES BIBLIOTECOLÓGICAS Y DE LA
INFORMACIÓN**

**LOS CRITERIOS DE RELEVANCIA EN LA RECUPERACIÓN DE
INFORMACIÓN: SU APORTACIÓN AL DISEÑO DE UNA INTERFACE**

TESIS

QUE PARA OPTAR POR EL GRADO DE

MAESTRO EN BIBLIOTECOLOGÍA Y ESTUDIOS DE LA INFORMACIÓN

PRESENTA:

JORGE ALBERTO MEJÍA RUIZ

DIRECTOR DE TESIS:

DR. EGBERT J. SÁNCHEZ VANDERKAST

MÉXICO, D. F., 2012

Para Leticia y José Carlos que son mi luz

A mis padres José e Irene

A mis hermanos José y Carlos

A mi tía Ceo

Agradecimientos

Mi agradecimiento total al Dr. Egbert J. Sanchez por su paciencia, guía y apoyo en la elaboración de esta tesis.

A los síndicos de tesis Dra. Brenda Cabral, Dra. Araceli Torres, Dr. Ariel Rodríguez, Mtro. Miguel Gama, por el apoyo y tiempo invertido en la revisión de este trabajo.

Tabla de contenido

INTRODUCCIÓN	i
LA RECUPERACIÓN DE INFORMACIÓN Y LOS SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN ...	1
1.1. INTRODUCCIÓN	1
1.2. ANTECEDENTES Y DEFINICIÓN.....	1
1.3. LOS SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN	7
1.4. TENDENCIAS EN LA RECUPERACIÓN DE INFORMACIÓN	10
1.4.2. EL ENFOQUE CENTRADO EN EL USUARIO	12
1.5. EVALUACIÓN DE LOS SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN.....	13
LA RELEVANCIA	17
2.1. INTRODUCCIÓN	17
2.2. LA DEFINICIÓN DE RELEVANCIA	17
2.3. LOS ESTUDIOS SOBRE LA RELEVANCIA	19
2.4. RELEVANCIA Y PERTINENCIA.....	28
2.5. TENDENCIAS	29
2.6. TIPOLOGÍA.....	33
2.7. JUICIO DE RELEVANCIA	41
2.8. MEDICIÓN DE LA RELEVANCIA	42
CRITERIOS DE RELEVANCIA COMO BASE PARA EL DISEÑO DE UNA INTERFAZ DE RECUPERACIÓN DE INFORMACIÓN	46
3.1. INTRODUCCIÓN	46
3.2. PROCESO DE BÚSQUEDA DE INFORMACIÓN	47
3.3. PROCESO DE BÚSQUEDA DE INFORMACIÓN DE LOS PROFESIONALES	52
3.4. EL MODELO PROPUESTO DE BÚSQUEDA DE INFORMACIÓN DE LOS PROFESIONALES.....	53
3.5. METODOLOGÍA	56
3.6. PARÁMETROS DE BÚSQUEDA POR ÁREA DE CONOCIMIENTO	58
3.7. PARÁMETROS DE BÚSQUEDA EN LAS ÁREAS DE INGENIERÍA Y TECNOLOGÍA Y CIENCIAS DE LA SALUD	68
3.7.1. PERFIL DEL ÁREA DE INGENIERÍA Y TECNOLOGÍA.....	68
3.7.2. PERFIL DEL ÁREA DE CIENCIAS DE LA SALUD.....	70
3.8. PARÁMETROS DE BÚSQUEDA DE INFORMACIÓN VS. ELEMENTOS DE JUICIO DE RELEVANCIA: PROPUESTA	72
3.8.1. ÁREA DE INGENIERÍA Y TECNOLOGÍA	73
3.8.2. ÁREA DE CIENCIAS DE LA SALUD	75
3.9. INDICADORES PARA EL DISEÑO DE UNA INTERFAZ DE BÚSQUEDA DE INFORMACIÓN.....	76
PROPUESTA PARA LA SOLUCIÓN AL PROBLEMA.....	78
CONCLUSIONES	82
BIBLIOGRAFÍA	85

INTRODUCCIÓN

El presente estudio tiene como finalidad ofrecer una visión general del valor de la relevancia* en la recuperación de información, mediante la identificación de aquellos criterios que establezcan los puntos principales a considerar en el diseño de una interfaz** que permita al usuario recuperar aquellos documentos que pueda considerar relevantes, a fin de dar respuestas concretas a sus necesidades de información.

El principal objetivo de un sistema de información, lo constituye la recuperación de información relevante para quien la busca y al mismo tiempo descartar aquellos materiales no relevantes.

Identificar aquello que sería relevante para los usuarios se convierte en muchos casos en un problema. La relevancia la tenemos que relacionar con la efectividad, eficiencia y eficacia que cualquier sistema de información debe incluir. Sin embargo, la podemos centrar principalmente en la relación entre la información recuperada por el sistema y por el usuario.

Podemos considerar, de acuerdo con autores como Borlund (2003)¹, Froehlich (1994)², Saracevic (1975)³ y Vakkari (2001)⁴ a la relevancia desde dos puntos de vista principales.

* Entendemos la relevancia como los juicios de valor que se dan al resultado de una búsqueda en un sistema de recuperación de información.

** Por interfaz nos referimos al intermediario entre el usuario de la información y el sistema de recuperación de información, principalmente en la forma en la que debe funcionar para satisfacer la pregunta formulada.

¹ Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for information Science and technology*, 54 (10), 913.

² Froehlich, T. (1994). Relevance reconsidered Towards an agenda for the 21st century: Introduction to special topic issue on relevance research. *Journal of the American Society for information Science*, 45 (3), 124

³ Saracevic, T. (1975). Relevance: a review of and framework for the thinking on the notion in information. *Journal of the American Society for Information Science*, 26 (6), 321

⁴ Vakkari, P. (2001). A theory of the task-based information retrieval process: a summary. *Journal of Documentation*, 57 (1), 44

El primero, desde el punto de vista del usuario del sistema, donde se calificaría la respuesta del sistema a partir de la pregunta que formula el usuario y que surge a partir de una necesidad de información.

El segundo, desde el punto de vista del sistema, al cual lo podemos reconocer como el análisis que hace el sistema con relación a la información que se encuentra almacenada y a la pregunta formulada por el usuario.

La relevancia ha sido estudiada desde la perspectiva de diferentes disciplinas, y en particular por lo que se refiere a la recuperación de información, tenemos, como se verá más adelante, que ha sido objeto de estudio desde la década de 1950 hasta la fecha, donde se ha considerado como parte importante en el diseño de sistemas de recuperación de información.

Considerando la relevancia como un tema poco tratado en nuestro país, se hace necesaria la investigación del mismo, de ahí el interés por ahondar más en su estudio con la finalidad de apoyar los desarrollos nacionales de sistemas de recuperación de información en el futuro.

En este estudio se planteó la siguiente hipótesis de trabajo:

“Si se identifican los criterios de relevancia por una comunidad dada nos permitirá diseñar y evaluar las herramientas de información, como los sistemas de recuperación de información documental y se facilitará la recuperación”.

Para la propuesta metodológica se hizo una investigación documental con la finalidad de apuntalar el marco teórico sobre la temática de la relevancia en la bibliotecología. Se tomaron algunas comunidades, con el fin de detectar los

criterios que utilizaron para diseñar estrategias de búsqueda y su juicio sobre la relevancia de la información recuperada.

Los objetivos que se pretenden lograr con esta investigación son:

Objetivo general

A partir de la revisión teórico-conceptual sobre un sistema de recuperación de información se proponen los criterios de relevancia con los cuales los usuarios juzgan la información.

Objetivos particulares

- Analizar las diversas teorías sobre la relevancia en la información.
- Señalar los factores que permiten a los usuarios de un sistema de recuperación de información, determinar la relevancia de un objeto de información.
- Identificar los criterios de relevancia que son empleados con mayor demanda por los usuarios de un sistema de recuperación de información.

Esta tesis está dividida en tres capítulos:

En el primero, se da un panorama general acerca de la relevancia y como ha sido estudiada a lo largo del tiempo, así como el modelo de relevancia que se utilizará en este trabajo.

En el segundo capítulo, se presentan diferentes aspectos acerca de *la recuperación de información y los sistemas de recuperación de información (SRI)*, tales como características, tipos y la evaluación de los mismos.

En el último capítulo, se aborda el estudio para determinar qué criterios de relevancia son considerados por una comunidad y su aplicación en la propuesta para la utilización como elemento en el diseño de una interface.

Durante el desarrollo de esta tesis se busca analizar la SRI para verificar si lleva a la valoración o relevancia, ni para manifestar aspectos de diseño gráfico o de usabilidad de una interface, así tampoco se busca alcanzar una investigación experimental; más bien lo que se pretende en este trabajo es lograr un primer acercamiento al tema, el cual en nuestro país es poco lo investigado.

CAPÍTULO 1

La recuperación de información y los sistemas de recuperación de información

1.1. Introducción

Recuperación de información implica la representación, almacenamiento, organización y el acceso, de una manera sistemática e íntegra mediante una interface que permita al usuario plantear sus necesidades de información y obtenga aquella que se ha publicado sobre una temática en particular, que responda al interés del usuario y que se encuentre en cualquier soporte o en fuentes documentales y en su obtención para la consulta. Mediante la utilización de un sistema de recuperación de información (SRI) que se encargue de la recuperación relevante para quien la solicite.

Hay otras formas de concebir la recuperación de información, como: la operación que permite obtener o recuperar de un determinado acervo, la información que cumpla con la demanda de información que se realice al efectuar una búsqueda.

1.2. Antecedentes y definición

Según Saracevic, Calvin Mooers, fue el forjador del concepto *information retrieval* y base de la perspectiva informativa del concepto de documentación. Definía el problema de la información como "*los aspectos intelectuales de la descripción de información y sus especificaciones para la búsqueda, además de cualquier sistema, técnica o instrumento que se utilice en la operación: la recuperación de información*".⁵

⁵ Saracevic T. (1999). Information Science. *Journal of the American Society for Information Science and Technology*, 50, (12), 1053

Junto con esta definición debemos considerar la coincidencia con la creación de la primera computadora digital durante la década de 1940, con lo cual inició la evolución de la ciencia de la información, debido a la generación de ésta en formato digital, dando como resultado un cambio en la noción de cómo se almacenaba, manejaba, recuperaba y distribuía la información, pasando de sistemas manuales a automatizados.

En relación con su desarrollo encontramos lo que Baeza-Yates señala al respecto de la evolución de la recuperación de información: las bibliotecas fueron las primeras instituciones que adoptaron sistemas de recuperación de información para recuperar información. Comúnmente los sistemas usados en las bibliotecas fueron desarrollados primero por instituciones académicas y después por vendedores comerciales. Este desarrollo lo enmarca en tres generaciones:

“En las primeras generaciones, tales sistemas consistían básicamente en la automatización de las tecnologías previas, como los catálogos en tarjetas, y básicamente permitían búsquedas basadas en el nombre del autor y el título de la obra.

En la segunda generación, se incrementó la funcionalidad de las búsquedas permitiendo agregar búsquedas por encabezamientos de materia, palabras claves y algunas búsquedas más complejas.

En la tercera generación, que es la más recientemente desarrollada, el foco principal es la implementación de interfaces gráficas, formatos electrónicos, características hipertexto y arquitecturas de sistemas abiertos.”⁶

⁶ Baeza-Yates, R. (1999). Modern information retrieval. Reading, Mass. : Addison-Wesley Longman.

En cuanto a la definición del término, encontramos tres tendencias principales:

- 1) aquellas que dan de forma general una definición
- 2) las que se refieren a la dificultad para poder definirla
- 3) aquellas que la consideran como sinónimo o que su principal interés es diferenciarla de la recuperación de datos.

En lo que se refiere a aquellas pertenecientes al grupo de las definiciones generales encontramos que Lancaster la define como el proceso de búsqueda en una colección de documentos, que tienen como meta identificar aquellos documentos que estén relacionados con un tema en particular.⁷ El autor ha intentado deslindar la recuperación de la documentación automatizada, enfocándose más hacia los aspectos lingüísticos y cognitivos principalmente en la definición de la necesidad de información, de lenguaje, de indización, de la estrategia de búsqueda, de la visualización y de la evaluación de la recuperación.

Salton la define como “la recuperación de información tiene que ver con la representación, almacenamiento, organización y acceso a los ítems de información”.⁸

Podemos considerar la definición de Salton a diferencia de la de Lancaster con una tendencia más encaminada hacia el manejo que se le da a la información sin considerar las condiciones que la rodean y que pueden influir en la recuperación de la misma.

Dentro del área de la bibliotecología encontramos que el *Diccionario de bibliología y ciencias afines*, señala a la recuperación de información como: “*Conjunto de procedimientos, generalmente automáticos, mediante los cuales los datos*

⁷ Lancaster, F. (1979). *Information retrieval systems : characteristics, testing, and evaluation*. New York : Wiley.

⁸ Salton, G. (1983). *Introduction to modern information retrieval*. México : McGraw-Hill.

*contenidos en los documentos concernientes a una materia dada pueden ser indizados, reservados y extraídos por el usuario”.*⁹

En esta definición encontramos un elemento a destacar que se refiere al impacto que tienen las tecnologías informáticas, aunque también está considerando a los documentos y a los usuarios como elementos en donde empieza y termina el proceso de recuperación de información.

En cuanto a lo que se refiere a aquellos que mencionan la problemática para poder definir la recuperación de información al igual que el de relevancia se ha prestado a confusiones en el momento de establecer su significado dentro del campo de la información.

Existen autores, entre ellos Van Rijsbergen,¹⁰ que menciona que es un término que ha sido definido en un sentido muy amplio. Otro autor que coincide con esta idea es Hert,¹¹ este señala que el término se ha prestado a confusión debido principalmente a la cantidad de definiciones existentes.

Lo expresado, tanto por Lancaster como por Hert indica que la recuperación de información ha pasado por un proceso evolutivo a lo largo del tiempo, a partir de los sistemas de recuperación manuales a los automáticos, de enfoques dirigidos hacía el sistema o hacia el usuario, de tal forma que se vuelve difícil encontrar una definición que cubra toda esta diversidad o que se limite a estar lo más actual posible.

Por otro lado Chowdhury¹² menciona que el concepto es muy extenso desde el punto de vista terminológico, sin embargo presupone que existen ciertos

⁹ Martínez de Sousa, J. (2004). Diccionario de bibliología y ciencias afines. Gijón, Asturias : Trea, 2004.

¹⁰ Rijsbergen, C. J. Van. (1979). Information retrieval. London : Butterworth, 1979. En línea <http://www.dcs.gla.ac.uk/Keith/Chapter.1/Ch.1.html> [Consultado: 20/05/07]

¹¹ Hert, C. A. (1997). Understanding information retrieval interactions: theoretical and practical implications. Greenwich, Connecticut : Ablex.

¹² Chowdhury, G. (2004). Introduction to modern information retrieval. London : Facet.

documentos o registros que contienen información la cual debe de estar organizada bajo un orden dado para su fácil recuperación.

Dentro del último grupo: los que la comparan con la recuperación de datos, podemos diferenciar dos tipos distintos de información a recuperar que los pueden definir como sinónimos: a) recuperación de información y b) recuperación de datos.

Estos términos podemos considerarlos como diferentes ya que representan por un lado la recuperación por ejemplo de datos estadísticos en los que no existe una interacción con una persona son simplemente el vaciado de datos en una base, por el contrario, la recuperación de información busca la generación de ideas a partir de aquello que fue creado por un autor.

En este sentido habría que considerar lo que nos presenta Martínez¹³ cuando menciona que un ejemplo de este desacierto lo hallamos en el *Glosario ALA* que define “*information retrieval*” como “recuperación de la información en primera acepción” y como “recuperación de datos” en una segunda acepción, considerando ambos términos sinónimos en lengua inglesa.

También encontramos que en el *Diccionario enciclopédico de ciencias de la documentación*, considera a la recuperación de información como el conjunto de “procedimientos y técnicas de representación, ordenación, búsqueda, presentación y evaluación de la información en sistemas automatizados, con el objetivo de facilitar el acceso eficaz y eficientemente a la misma”.¹⁴

García, define la recuperación de información como la “*acción y efecto de recuperar, previa búsqueda y localización, los datos o información concreta que se*

¹³ Martínez, M. (2004). *Recuperación de información : modelos, sistemas y evaluación*. Murcia: Kiosko.

¹⁴ *Diccionario enciclopédico de ciencias de la documentación*. (2004). Madrid : Síntesis, 371

desea de entre la almacenada en un fondo documental o en una memoria de ordenador".¹⁵

De las definiciones anteriores podemos mencionar que el manejo terminológico recuperación de datos y de información obedece a una visión principalmente informática y de acceso y recuperación en bases de datos principalmente, sin ocuparse de resolver problemas de recuperar información sobre un tema.

Por otra parte, Van Rijsbergen,¹⁶ menciona que podemos establecer las diferencias entre recuperación de datos y recuperación de información de acuerdo con el cuadro 1 de la siguiente forma:

Cuadro 1: Diferencias entre recuperar datos o información

	Recuperación de datos	Recuperación de información
Acierto (correspondencia)	Exacta	Parcial, la mejor
Inferencia	Algebraica	Inductiva
Modelo	Determinístico	Posibilístico
Lenguaje de consulta	Fuertemente estructurado	Estructurado o natural
Especificación de la consulta	Precisa	Imprecisa
Error en la respuesta	Sensible	Insensible

Podemos mencionar que la recuperación de datos se basa en distribuciones más rígidas y estructuradas, mientras que la recuperación de información se enfrenta a lo contrario al tratar de recuperar elementos de una pregunta que no tiene esa normalización, dando como resultado que en la recuperación de datos la respuesta va a corresponder exactamente con lo buscado mientras que en la

¹⁵ García, L. (2000). Diccionario del archivero bibliotecario : terminología de la elaboración, tratamiento y utilización de los materiales propios de los centros de documentación. Gijón, Asturias : Trea.

¹⁶ Rijsbergen, Van. Op.cit. [consultado: 23/05/07] <http://www.dcs.gla.ac.uk/Keith/Chapter.1/Ch.1.html>

recuperación de información se refiere al valor que se le da al resultado de la búsqueda y que va a depender de la utilidad que el usuario le asigne.

La estructuración de los datos es siempre más fácil que la estructuración de la información y por tanto, es más fácil su recuperación, la recuperación de información, frente a la recuperación de datos, consiste en el tratamiento y procesamiento de documentos, no datos o registros.

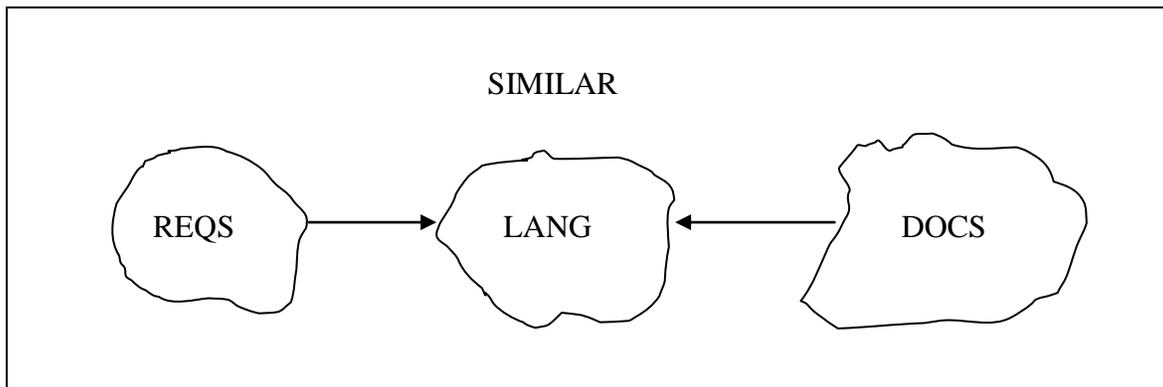
1.3. Los sistemas de recuperación de información

En la actualidad la recuperación de información ha tomado una gran importancia, esto se debe en gran medida a que los buscadores de internet están situados como el primer método utilizado para obtener cualquier tipo de información, sea para el uso académico, lúdico o empresarial, por mencionar algunos.

Los sistemas de recuperación de información según Salton (1983) *“cualquier sistema de recuperación de información puede ser descrito como un conjunto de ítems de información (DOCS), un conjunto de peticiones (REQS) y algún mecanismo (SIMILAR) que determine que un ítem satisface las necesidades de información expresadas por el usuario en la petición”* a esta definición Salton agregó *“los documentos se pueden convertir a un formato especial, mediante una clasificación o un lenguaje de indización que llamo LANG”*.¹⁷

¹⁷ Salton, G. ; McGill, M. (1983). Introduction to modern information retrieval. New York :McGraw Hill.

Figura 1: Sistema de recuperación de información (Salton)



Salton, G. ; McGill, M. Introduction to modern information retrieval. New York :McGraw Hill, 1983. p. 10

La figura 1 muestra que el proceso entre REQS y SIMILAR representa la formulación de la búsqueda y el proceso de SIMILAR y DOCS es el de recuperación de documentos, de esta forma SIMILAR realiza la función de determinar la similitud que se da entre la pregunta y la representación de los recursos de información.

Los sistemas de información al igual que el mismo manejo de la información han tenido que evolucionar a lo largo del tiempo, creando modelos de recuperación, entre los más comunes podemos encontrar a los booleanos, probabilísticas y vectoriales.

En los últimos años y debido a los intereses económicos derivados de buenos posicionamientos en los distintos buscadores se está produciendo una explosión en todos los campos relacionados con este tema, por tanto es necesario conocer como se estructuran los modelos de recuperación.

Según Dominich,¹⁸ podemos encontrar los siguientes modelos para la recuperación de información:

¹⁸ Dominich, S. (2000). A unified mathematical definition of classical information retrieval. *Journal of the American Society for Information Science*, 51, (7), 614-624.

1. Clásicos: que incluyen los tres más comúnmente citados: booleano, espacio vectorial y probabilístico.
2. Lógicos: desarrollados en la década de los 90, basados en la lógica formal. La recuperación de información se entiende como un proceso inferencial a través del cual se puede estimar la probabilidad de que una necesidad de información de un usuario, expresada como una o más consultas, sea satisfecha ofreciendo un documento como "prueba".
3. Basados en la interactividad: que incluyen posibilidades de expansión del alcance de la búsqueda y hacen uso de retroalimentación por la relevancia de los documentos recuperados.
4. Basados en la inteligencia artificial: que tiene su base en la teoría del conocimiento, las redes neuronales, los algoritmos genéticos y el procesamiento del lenguaje natural.
5. Alternativos: que están basados en la *lógica fuzzy* o *lógica difusa*.

Este último modelo, utiliza la lógica difusa como un medio para representar y manipular datos que no son precisos. A la vez puede ser vista como una extensión de los sistemas de la lógica clásica que proporciona un marco de trabajo para tratar con el problema de la representación del conocimiento en un entorno de incertidumbre e imprecisión. Donde la lógica difusa y las formas de razonar de los seres humanos convergen ya que todo razonamiento y pensamiento del ser humano son aproximados, lo que nos lleva a la imprecisión.

En los sistemas de información la lógica difusa agrega un cierto grado de verdad o de falsedad, de tal forma que encontramos que se acerca a los criterios que aplicamos en la vida diaria, donde podemos calificar una cosa en términos medios, como cuando juzgamos un resultado de no tan relevante a la necesidad de información, podemos utilizar un grado de imprecisión.

Ramírez¹⁹ presenta un ejemplo de cómo pretende introducir un grado de incertidumbre en las sentencias y menciona que con la *lógica difusa*, las proposiciones pueden ser representadas con cierto grado de verdad o cierto grado de falsedad, a esta situación se le conoce como grado de pertenencia, correspondiendo así a una realidad en la que no sólo existe el negro o el blanco sino una infinita gama de matices de gris. De esta manera los valores *ceros* y *uno* se subdividen en valores infinitos entre ellos. Por medio de la *lógica difusa* pueden formularse matemáticamente nociones como “poco caliente” o “muy frío”, para que sean procesadas por computadoras y cuantificar expresiones humanas vagas, tales como “no muy alto” o “un poco rápido”.

La lógica difusa nos ayuda para poder dar ese grado de pertenencia de las cosas dentro de un razonamiento acercado a lo que un usuario pudiera buscar, considerando el grado de imprecisión que pudiera existir en el lenguaje que utilice en la consulta y permitiendo que un sistema de recuperación de información se acerque a la realidad.

1.4. Tendencias en la recuperación de información

El desarrollo de la recuperación de información se ha visto influenciada por dos tendencias, en su uso por:

- ❖ el enfoque centrado en el sistema
- ❖ el enfoque centrado en el usuario

1.4.1. El enfoque centrado en el sistema

El enfoque centrado en el sistema está fundamentado en la actividad de los sistemas, en los cuales los usuarios llevan a cabo una consulta y el sistema arroja

¹⁹ Ramírez, A. (2007, octubre) La lógica difusa en los sistemas automatizados de recuperación de información. *XIV reunión de bibliotecarios de la Península de Yucatán*, Mérida, Yucatán.

una lista con los resultados obtenidos.²⁰ En este sentido se podría considerar que bajo este punto de vista tanto el usuario como el sistema son vistos de manera independiente.

En este enfoque el centro principal de atención es la construcción de sistemas, entendido como *“un conjunto de elementos que interactúan, y de sus interacciones surge un comportamiento como un todo. Por eso, es importante tener en cuenta que los sistemas presentan un cierto carácter de totalidad más o menos organizada.”*²¹

La principal analogía que realiza este enfoque es que los sistemas de recuperación de información deben ser vistos como un sistema físico, tanto en su naturaleza como en las diferentes técnicas o métodos que se utilicen para su estudio.²² Se concebía a los sistemas de recuperación de la información, literalmente iguales a los sistemas físicos, donde se consideraba apropiado emplear las pruebas físicas o mecánicas, pero mejor que eso, se consideraba legítimo suponer, y apropiado aplicar técnicas similares en su forma a las técnicas apropiadas para los sistemas físicos o mecánicos.

En la actualidad el paradigma físico es el más establecido ya que ha logrado, desde sus concepciones, resolver determinados problemas en la recuperación de información y además ha logrado llevar a la práctica y poner en marcha diversos sistemas. Sin embargo este punto de vista ha recibido disímiles críticas, las cuales principalmente recaen en el hecho de que éste ignora el papel que desempeña el usuario. Algunas de las deficiencias que se le señalan al mismo son: la ausencia de técnicas de retroalimentación o la falta de estudio acerca del concepto de relevancia y de los aspectos cognitivos provenientes de los usuarios y lo que se

²⁰ Vargas - Quesada, Benjamín; Moya Anegón, Félix de; Olvera Lobo, María Dolores. (2002). Enfoques en torno al modelo cognitivo para la recuperación de información: análisis crítico. *Ciência da informação*, 32 (2)

²¹ Ponjúan Dante, Gloria; et. al. (2004). *Sistemas de información: Principios y aplicaciones*. Ciudad de la Habana: Félix Varela.

²² Ellis, David. (1992). The Physical and Cognitive Paradigms in Information Retrieval Research. *Journal of documentation*, 48 (1)

puede considerarse como un elemento fundamental: no se considera al usuario como parte del sistema.²³

1.4.2. El enfoque centrado en el usuario

El enfoque centrado en los usuarios no es resultado de un hecho al azar, sino que es producto de una revolución cognitiva que nace en la psicología con su perspectiva de la psicología cognitiva y que se extiende a variadas disciplinas y campos de investigación.

El modelo cognitivo engloba lo que constituye la base del modelo tradicional, dando lugar a lo que Ingwersen (1992)²⁴ denomina la teoría cognitiva de la recuperación de la información, que implica un intento de globalizar la recuperación de información, por medio de la representación de todos sus componentes, como representaciones estructurales con distintos niveles de complejidad que cooperan en un proceso de comunicación.

Este modelo, a diferencia del tradicional *"no se pone en marcha a partir del momento en que el usuario realiza una consulta al sistema de recuperación de información, sino que comienza incluso antes de que se produzca la necesidad informativa del usuario. En este punto de vista se tiene en cuenta la extensa gama de factores implicados en el proceso de recuperación de información; especialmente aquellos aspectos que influyen en el estado cognitivo y contextual del individuo, por ejemplo, su estado anímico, sus creencias, los objetivos planteados, su grado de motivación o sus áreas de interés, entre otros."*²⁵

Bajo esta vertiente centrada en los usuarios, *"el objetivo de la recuperación de información es suministrar información que pueda ser un suplemento para las*

²³ Saracevic, Tefko. (1999). Information Science. *Journal of documentation*, (50) 12

²⁴ Ingwersen, P. (1992). Information retrieval interaction. London : Taylor Graham.

²⁵ Op. Cit. [7]

*condiciones mentales humanas conscientes o inconscientes en una situación dada.*²⁶

Las principales diferencias entre el enfoque centrado en los sistemas y el enfoque centrado en los usuarios son:

Los usuarios se conciben como un componente más del SRI, por lo que es necesario su estudio y el del entorno en el que se desenvuelven.

El término relevancia como indicador para la evaluación de los SRI es vista desde una dimensión más cualitativa. Sólo aquellos documentos que modifiquen las estructuras de conocimiento de los usuarios pueden ser considerados realmente relevantes, esto es conocido como relevancia cognitiva.

La representación de la información contenida en los SRI, que es llevada a cabo a través de los lenguajes documentales, tiende a acercarse más a las realidades concretas de los usuarios para lograr una mayor comunicación entre el sistema y los usuarios. En la construcción de los lenguajes documentales existe una tendencia de aproximarse más al lenguaje natural y así lograr una mejor representación de la información contenida en los documentos y una mejor decodificación por parte de los usuarios.

Al estudiar los aspectos cognitivos y socioculturales, no se restringe sólo al estudio de estos aspectos en los usuarios, sino también se debe incluir a los profesionales de la información y a los creadores de los documentos.

1.5. Evaluación de los sistemas de recuperación de información

La historia de la evaluación de los sistemas de información comienza en 1957 con los proyectos Cranfield, desarrollado en dos etapas Cranfield I y II, los cuales

²⁶ Op.cit. "[11]

deben su trascendencia y repercusión a que crearon un modelo, una metodología y unas pautas para la evaluación de los modelos aún vigentes.

Su primera etapa se llevó a cabo de 1957 a 1962 y dio pie a las herramientas que se utilizarían en años posteriores para la evaluación:

- Una colección de documentos (fuente) de los que se obtuvieron las preguntas que se le plantearían al sistema.
- Una serie de juicios de relevancia de los documentos recuperados y expresados en tres niveles: muy útil, útil, nada útil.
- El uso de las medidas de exhaustividad y precisión para analizar los resultados.

La segunda parte del proyecto se realizó de 1963 a 1966, estos estudios se centraron en la eficacia de los lenguajes de indización, se determinó la relevancia de los documentos en relación con las preguntas formuladas, posteriormente se aplicó una escala de varios niveles, de uno a cuatro, y no un sistema de relevancia binaria (relevante-no relevante).²⁷

La importancia de estos proyectos consiste en que fueron los primeros en los que se estableció una metodología para la evaluación, así como las herramientas que debían utilizarse, considerando a éstas como la colección de documentos de donde se van a extraer las preguntas, además de los juicios de relevancia. Mediante estas herramientas podemos calcular, tanto la precisión como la exhaustividad de los resultados. Dicha metodología es hasta la actualidad la que se utiliza en la evaluación de la recuperación.

²⁷ Cleverdon, C. (1997). *The Cranfield tests on index languages devices. Readings in information retrieval*. San Francisco: Morgan Kaufmann.

Los sistemas de recuperación de información se pueden evaluar en función de distintos parámetros Salton²⁸, Baeza-Yates²⁹ siendo los más utilizados la *Precisión (Precision)* y la *Cobertura (Recall)*.

El parámetro de precisión refleja la exactitud del resultado obtenido para una consulta, la proporción de los documentos relevantes recuperados en relación a todos los obtenidos. Un sistema que recupera más información incorrecta que correcta trabaja con una menor precisión que aquel que no genera información incorrecta.

Mientras que el parámetro de cobertura muestra la proporción de documentos relevantes obtenidos. Esta medida muestra la cantidad de documentos correctos y relevantes devueltos por un sistema respecto a la cantidad total de documentos relevantes y correctos presentes en la colección.

La medida de F (F-measure) es una medida que combina *R* y *P* para evaluar los sistemas de RJ del siguiente modo:

NR = el número documentos recuperados relevantes

NB = el número documentos relevantes de la colección y,

NE = el número documentos recuperados

Se define:

$$P = \frac{NR}{NE}$$

$$R = \frac{NR}{NB}$$

²⁸ Salton, G. (1998). Automatic text processing. Addison-Wesley.

²⁹ Op. Cit. [7]

NB

$$F\text{-measure} = \frac{(\beta^2 + 1) \cdot P \cdot R}{\beta^2 \cdot P + R}$$

Un F-measure alto, indica una buena eficacia global de recuperación de documentos relevantes.

Existen medidas para evaluar otros aspectos de un sistema de recuperación de información, tales como:

Fallout. Refleja la tendencia de un sistema a asignar datos incorrectos.

Overgeneration. Representa la cantidad de información irrelevante que es generada por el sistema, o sea los documentos recuperados que no son relevantes para la consulta realizada.

Tiempo de respuesta. Tiempo promedio que requiere el sistema de recuperación de información para responder a una consulta.

Eficiencia de almacenamiento. Calcula la sobrecarga de espacio como la relación entre el tamaño de los índices y documentos con respecto al tamaño de la colección.

Tanto los sistemas de recuperación de información como su evaluación han ido desarrollándose a la par una de la otra, y en el caso particular de la evaluación se ha buscado la manera de poder realizar este proceso de forma efectiva de tal suerte que se pueda conocer el proceder de la recuperación, y así hallar las formas que ofrezcan el mejor modo de recuperación de información de un usuario frente a una necesidad dada. Esto nos llevará sin lugar a dudas a una continua investigación, ofreciendo nuevas formas de recuperación frente a las ya utilizadas o en otros casos implementando mejoras a las ya existentes.

CAPÍTULO 2

La relevancia

2.1. Introducción

Este capítulo se centra en la relevancia. Se presentan algunos estudios realizados a lo largo del tiempo, principalmente de la década de 1950 a la fecha, se maneja la definición de la relevancia y algunas de las opiniones encontradas respecto a este concepto.

Se continúa con las principales tendencias encontradas sobre las diversas clasificaciones propuestas por diferentes autores acerca de la relevancia, y los varios tipos de ella.

También se refiere a los juicios de relevancia, así como a la medición de la misma, considerando las dos principales medidas utilizadas: la manual y la conocida como *polling*.

Finalmente se elaboran algunas consideraciones respecto a la temática.

2.2. La definición de relevancia

El concepto de relevancia ha sido estudiado desde diferentes campos disciplinarios tales como la:

- Lógica: que establece que la relevancia es usada para establecer cuales son las condiciones necesarias para la validación de una inferencia de A para B es como A es relevante a B.³⁰
- Filosofía: ésta ha estado envuelta en forma general en la noción de relevancia en la explicación del “acerca de” y en la teoría del significado. Al tratar de explicar que hace que funcione el mundo. Como no es un asunto homogéneo lo divide en diferentes sistemas de relevancia:
 - a) la relevancia temática: que envuelve la percepción de algo que inicia la problemática;
 - b) la relevancia interpretacional: que incluye el grupo de conocimientos que se tienen a mano sobre el significado de que cosa se percibe;
 - c) relevancia motivacional: se refiere al curso de acción para ser adoptado.³¹
- La Bibliotecología y Estudios de la Información: la relevancia representa el grado sobre el cual la información que es recuperada de búsqueda en una colección de la biblioteca o en otro recurso, tal como catálogo en línea o base de datos bibliográfica, es juzgada por el usuario para ser aplicable a (“sobre”) un tema de una pregunta. La importancia depende de la opinión subjetiva del investigador del grado al cual el documento satisface la necesidad de información, que se puede o no se puede haber expresado completamente o con precisión en la búsqueda.³²

³⁰ Saracevic, T. (1975). Relevance: a review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, 26: 321–343.

³¹ Schutz, A. (1970). *Reflections on the problem of relevance*. New Haven, CT: Yale University Press.

³² Online Dictionary for Library and Information Science. [en línea] 2007. [consultado: 28/feb/2007]. disponible en: http://lu.com/odlis/odlis_r.cfm

2.3. Los estudios sobre la relevancia

El impulso que se viene percibiendo en el desarrollo de las bibliotecas está llevando a muchas instituciones a ofrecer su información de forma electrónica. La manera que se elija para recuperar y presentar al usuario la información es tan importante como el propio contenido de la misma, ya que la forma en la que se haga accesible va a establecer que la información obtenida cumpla con la finalidad de ser útil o por el contrario se quede simplemente almacenada.

En la actualidad los sistemas de recuperación de información* han dado en mayor o menor medida satisfacción a las necesidades que de ellos se solicitan, sin embargo, tendríamos que considerar cuales son los elementos que influyen en la recuperación de información.

Uno de los puntos que deberíamos considerar para la creación y diseño de sistemas de recuperación de información debería ser la relevancia, ofrecer a quien solicita información, documentos relevantes, considerar que cualquier usuario va a querer que la respuesta que reciba de un sistema de información sea de información realmente útil para sus necesidades, logrando que los sistemas de información sean funcionales.

Identificar aquello que sería relevante para los usuarios se convierte en muchos casos en un problema. La relevancia la tenemos que relacionar con la efectividad, eficiencia y eficacia que cualquier sistema de información debería tener. Sin embargo, la podemos centrar principalmente en la relación entre la información recuperada por el sistema y por quien busca información.

Los estudios sobre el tema de la relevancia tienen como antecedente algunas investigaciones que mencionaremos a continuación. Conocer estos trabajos nos

* Entendiendo al sistema de recuperación de información como proceso donde se accede a una información previamente almacenada, mediante instrumentos informáticos que permiten la recuperación de información relevante a una necesidad

permite tener una referencia hacia las características que presenta este tema en la actualidad.

Dentro de estas investigaciones encontramos a Saracevic (1975)³³ que describe como uno de los primeros estudios realizados que intentaba determinar cuáles son los juicios que un usuario utilizaba para considerar a los documentos relevantes al trabajo realizado por Bradford en la década de 1930, donde señala en un documento la referencia hacia los artículos relevantes de una materia en publicaciones periódicas donde se podían encontrar títulos con menor relevancia para una disciplina o tema, pero con mayor importancia para otro tópico, mientras que el núcleo esencial de las revistas forma la base de la literatura para todas las disciplinas, por consiguiente, la mayoría de los trabajos importantes son publicados en pocas revistas.

Saracevic, considera la relevancia como uno de los puntos centrales de la documentación, las ciencias de la información y de la recuperación de información, cuya meta es poder lograr que un sistema de recuperación de información provea de información relevante a los usuarios. Asimismo, la relevancia debe ser considerada como un elemento principal al momento de diseñar y evaluar cualquier sistema de recuperación de información.³⁴

Es a finales de la década de los cincuenta cuando Brian Vickery³⁵ inicia la discusión sobre la relevancia y la define como la decisión en el óptimo nivel de discriminación. Rees, en 1966, entiende por relevancia “los criterios usados para cuantificar el fenómeno involucrado cuando un usuario juzga la utilidad, importancia, grado de coincidencia, proximidad, apropiación, igualdad, pertinencia

³³ 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), 324-325.

³⁴ Saracevic, T. (1996). Relevance reconsidered. *Information science: Integration in perspectives. Proceedings of the Second Conference on Conceptions of Library and Information Science. Copenhagen (Denmark)*, 207.

³⁵ Vickery, B. (1959). Subject Analysis for Information Retrieval. *Proceedings of the international conference on scientific information*, 2, 863. [consultado 27/02/07]. Disponible en: http://books.nap.edu/openbook.php?record_id=10866&page=855

o valor de un documento o representación de un documento contra una necesidad de información, pregunta, investigación, etcétera”.³⁶

Bajo esta perspectiva podemos observar que se considera el juicio de valor que realiza el usuario al momento de enfrentarse a dar respuesta a una necesidad de información, elementos que darán pie a lo que veremos más adelante como el enfoque dirigido hacia el usuario, donde es éste el principal elemento a considerar al momento de determinar la relevancia de la recuperación de información.

Cooper³⁷ en 1971 delimitaba la relevancia con base en nociones lógico-matemáticas, llamadas:

- a) Vinculación (*entailment*) o consecuencia lógica: se refiere a que una sentencia (conclusión) está determinada a ser la consecuencia lógica de una serie de sentencias (premisas) siempre y cuando se cumplan ciertas condiciones.
- b) Minimalidad (*minimality*): se refiere a que una sentencia es lógicamente relevante a una necesidad de información, sí y solo sí es mínimamente miembro de un algún conjunto de las oraciones almacenadas para algún componente de la declaración de la necesidad.

Determinó también que una sentencia:

s es relevante a otra sentencia **r** si **s** corresponde a un mínimo conjunto de premisas **M** vinculando a **r**

$$(s, r) \text{ iff } \exists M (s \in M \wedge M \models r \wedge M - s \not\models r)$$

La delimitación que se hace sobre la relevancia bajo estas nociones lógico-matemáticas establecen que existe una relación entre la relevancia y una

³⁶ Rees, A. M. (1966). The relevance of relevance to the testing and evaluation of document retrieval system. *Aslib proceedings*, 18 (11), 316-324.

³⁷ Cooper, W. (1971). A definition of relevance for information retrieval. *Information storage and retrieval*, 7 (1), 22, 24.

consecuencia lógica, que podemos considerar para otro de los principales enfoques que encontramos sobre la relevancia, que es el enfoque dirigido al sistema, donde es la parte del sistema de recuperación de información mediante el uso de algoritmos la que tiene más peso para determinar la relevancia sin considerar al usuario.

En 1990 Schamber, Eisenberg y Nilan, desarrollaron una teoría sobre la relevancia orientada hacia el usuario y que se definía como el intercambio dinámico de información y comunicación que depende de la relación entre la información y la necesidad de información del usuario.

Establece tres ejes sobre la naturaleza de la relevancia y su rol ante la conducta de la información:

- a) *“La relevancia es un concepto cognitivo multidimensional que depende en gran medida de la percepción del usuario sobre la información y de sus propias necesidades de información.*
- b) *La relevancia es un concepto dinámico que depende del juicio del usuario, sobre la calidad de la relación entre la información y la necesidad de ella en un cierto punto.*
- c) *La relevancia es un concepto complejo, pero sistemático y medible, si se aproxima de manera conceptual y operacional desde la perspectiva del usuario”.*³⁸

Estos autores, como ya se mencionó, se centran en uno de los principales enfoques que se adoptan para la relevancia: el del *usuario*. Además identificaron que la relevancia es un concepto complejo que se compone de diferentes elementos, y que es dinámico pues al depender del juicio de quien busca, se va a

³⁸ Schamber, L., Eisenberg, M. B., Nilan, M. S. (1990). A re-examination of relevance : toward a dynamic, situational definition. *Information processing & management*. 26 (6), 755-775

mantener siempre cambiante y situacionalmente ligado al conjunto de relaciones o factores que envuelvan la necesidad de información.

Sin embargo, debemos hacer mención especial de lo realizado por Tefko Saracevic, quien planteó a la relevancia de manera más completa, esquemática y metódica, pues fue el primero en reunir todos los trabajos sobre la relevancia, publicados entre los años 1959 a 1976 y propone un marco para clasificar las diferentes nociones que se tenían sobre la relevancia hasta ese momento.

Saracevic es considerado uno de los principales autores sobre el tema de la relevancia, al inicio de sus trabajos su enfoque estaba dirigido a la relevancia en el sistema y considera a la relevancia como elemento principal en la recuperación de información. Además de que es citado en la mayoría de los trabajos que hablan del tema de la relevancia; sus escritos comienzan en 1969 con un estudio acerca de los efectos de los juicios de relevancia en títulos, resúmenes y textos completos, a partir de dicho estudio ha trabajado con la relevancia desde la definición del concepto hasta la experimentación con el mismo, para la comprobación de su aplicación.

Otra perspectiva es la de Martínez,³⁹ quien señala que un documento recuperado se puede considerar relevante cuando el contenido del mismo posee alguna significación o importancia con motivo de la pregunta realizada por el usuario, es decir, relativa a la necesidad de información.

Lo anterior establece que no sólo basta que la información recuperada de un sistema de información coincida con la pregunta realizada, sino que debe de ofrecer, para ser considerada relevante, elementos nuevos a los conocidos por quien necesita la información.

³⁹ Martínez, F. J. (2004). Recuperación de información: modelos sistemas y evaluación (p. 45). Murcia: Kiosko.

La investigación acerca de la relevancia tiene un fuerte período de avance en la década del 70, donde se establecen los primeros fundamentos y posteriormente se han venido efectuando nuevos desarrollos y aplicaciones.

Al respecto de la historia de la relevancia, Mizzaro,⁴⁰ realizó un trabajo sobre los diferentes documentos que han tratado el tema de la relevancia, estudió alrededor de 160 documentos, los dividió en tres grandes períodos: a) antes de 1958, b) 1959-1976 y c) 1977-1997. Dentro de cada periodo diferenció siete aspectos que van desde fundamentos metodológicos, clases de relevancia, hasta cómo se establecen los juicios de relevancia. De dicho documento destacan los periodos siguientes:

- a) El preestablecimiento de la temática de **1930 a 1958**. Este periodo concluye con el reconocimiento del concepto en la Conferencia Internacional de Información Científica (*International Conference for Scientific Information (ICSI)*).
- b) Consolidación y de bases para la investigación **1959-1976**. Se destacan los documentos de corte conceptual y experimental; las distinciones entre relevancia y pertinencia; sobre los juicios de relevancia humana, los cuales son afectados por las condiciones que los rodean, con lo que hace cuestionable la confianza en ellos; el establecimiento de escalas para medir los juicios de relevancia, de igual forma se trató de estudiar la naturaleza dinámica de los juicios de relevancia.
- c) Tendencias de la investigación sobre la relevancia (**1977 a 1997**) liderado por:

⁴⁰ Mizzaro, S. (1997), Relevance: The whole history. *Journal of the American Society for Information Science*, 48, 810–832.

- I. Schamber, Eisenberg y Nilan quienes estudian en 1990 la tendencia del trabajo en relevancia y la clasifican en varios apartados bajo las etiquetas 1) multidimensional, 2)cognitiva y 3) dinámica.

- II. Froehlich en 1994, establece seis puntos respecto a la relevancia que a continuación se mencionan:
 - 1) La relevancia de los usuarios no pueden ser definida en un sentido preciso.
 - 2) La relevancia es una categoría natural, derivada de la experiencia.
 - 3) La distinción entre la relevancia y pertinencia no es lo mismo para usuarios y bibliotecarios.
 - 4) La temática es el núcleo de la relevancia.
 - 5) Los juicios de relevancia se basa en un conjunto finito de criterios.
 - 6) La hermenéutica pueda darnos una base para modelar los criterios del usuario y del sistema de recuperación de información.

- III. Schamber en 1994 propone tres argumentos fundamentales en el estudio de la relevancia a saber:
 - 1) Qué factores contribuyen a los juicios de relevancia.
 - 2)Cuál es el rol de la relevancia y la evaluación de sistemas de recuperación de información.
 - 3) A qué se le puede nombrar relevancia, o puede hablarse de diferentes tipos de relevancia.

- IV. O'Brien⁴¹ establece la importancia de la relevancia en la evaluación de catálogos de acceso público en línea (OPAC, por sus siglas en inglés), señala que la relevancia es un rasgo central, pero a la vez el de mayor incertidumbre y dinamismo en el mundo de los OPACs, y para la parte

⁴¹ O'Brien, A. (1990) . Relevance as an aid to evaluation in OPACs. *Journal of Information Science*, 16, 265–271.

de los sistemas de recuperación de información, en dar relevancia de la información para responder a la necesidad de información resulta difícil obtener una medida.

Finalmente en este período hubo un cambio, se pasó de los estudios orientados al sistema hacia los estudios orientados al usuario.

Como se ve, los estudios sobre la relevancia no son nuevos. Se ha tenido interés en conocer cómo es y de qué forma podemos recuperar todo aquello que sea relevante para una necesidad de información dada, cómo lograr que los sistemas de recuperación de información respondan a esa necesidad, o cómo se llevan a cabo por parte de los usuarios, los juicios, para determinar cuándo una información es considerada relevante o no relevante.

Consideramos que la integración de ambos enfoques, tanto el dirigido al sistema como el dirigido al usuario deben de ser puestos en juego al momento de diseñar una interfaz de recuperación de información.

Al ser la relevancia una de las plataformas sobre las cuales se basan las medidas que se aplican en la recuperación de información, es importante definirla como concepto. Sin embargo, nos encontramos con autores como Saracevic y Harter que veremos más adelante, que señalan la dificultad de definir la relevancia.

Ya que este concepto es entendido intuitivamente y difícil de definir, tal como lo expresa Harter *“el principal problema de los estudios sobre los factores que afectan a la relevancia es que se han hecho de manera intuitiva, tal vez esto sea debido la variedad de interpretaciones de este término”*.⁴²

⁴² Harter, S. P. (1971). The *Cranfield II* relevance assessments: a critical evaluation. *Library quarterly*, 41 (3), 229-243.

Por otro lado Saracevic menciona que *“nadie ha podido explicar a los usuarios de los sistemas de recuperación de información que es la relevancia, incluso si luchan (en ocasiones en vano) por encontrar cosas relevantes. Por ello la relevancia se entiende intuitivamente”*.⁴³

Por su parte Mizzaro señala que *“uno de los principales problemas en la recuperación de información, son la diversidad de interpretaciones de algunos conceptos, como es el caso de la definición de relevancia”*.⁴⁴

Al intentar definir “relevancia” nos enfrentamos a un concepto que está marcado por la subjetividad, ya que podemos entenderlo y explicarlo de diferentes maneras dependiendo de cada individuo; ya que un documento puede resultar relevante para una persona y para otra no, dependiendo de la necesidad de información que tenga cada uno, en otro caso, un mismo documento en cierto momento puede resultar relevante para una persona y en otro momento ya no, o en el último de los casos podemos encontrar documentos que sólo en cierta parte de los mismos pudieran resultar relevantes para una disciplina y el resto de su contenido no.

Así nos enfrentamos al hecho de que la mayoría de las personas pueden determinar cuando un documento es relevante para satisfacer su necesidad de información, sin embargo les resulta difícil poder definir lo que significa “relevante”.

Tomando en cuenta lo anterior, resulta complejo dar una definición de la relevancia sin embargo para fines de este trabajo, la definiremos como: **los juicios de valor que se dan al resultado de una búsqueda en un sistema de recuperación de información.**

En suma, consideramos que sin disminuir valor a la determinación del concepto de relevancia, será mejor simplemente realizar la identificación de la misma. Esto con

⁴³ Saracevic, T. (1996). op. cit., 215.

⁴⁴ Mizzaro, Stefano. (1997). Relevance: the whole history. *Journal of the American Society for Information Science*. 48, 810-832.

base en la dificultad que han señalado los autores presentados, además de que es finalidad de este estudio buscar aquellos criterios que son utilizados por un usuario para valorar como relevante el resultado de una búsqueda de información en un Sistema de Recuperación de Información (SRI).

2.4. Relevancia y pertinencia

Un término que se relaciona con el de relevancia es pertinencia, los que comúnmente son tratados de forma ligada y en algunos casos como sinónimos.

Al respecto Foskett, definió lo relevante como el *“documento perteneciente al campo/materia/universo del discurso delimitado por los términos de la pregunta, establecido por el consenso de los trabajadores en ese campo”*,⁴⁵ por otra parte define como pertinente a *“aquel documento que añade nueva información a la previamente almacenada en la mente del usuario, que le resulta útil en el trabajo que ha propiciado la pregunta”*.⁴⁶

Lancaster y Warner señalan sobre esta cuestión: *“aunque puede usarse otra terminología, la voz relevancia parece la más apropiada para indicar la relación entre un documento y una petición de información efectuada por un usuario, aunque puede resultar erróneo asumir que ese grado de relación es fijo e invariable, siendo mejor decir, que un documento ha sido juzgado como relevante a una específica petición de información”*.⁴⁷

Por otra parte marcan que la relevancia de un documento hace relación a una necesidad de información que plantea un usuario, no tiene por qué coincidir con la que emitan muchos expertos sobre el contenido del documento sino con la satisfacción de ese usuario y la *“utilidad”* que estos contenidos van a tener para él,

⁴⁵ Foskett, D. J. (1972). A note on the concept of ‘relevance’. *Information Storage Retrieval*, 8 (2), 77-78.

⁴⁶ Foskett, D. J. (1972). Op. cit., 77-78.

⁴⁷ Lancaster, F. W., Warner, A. J. (1993) *Information Retrieval Today*. Arlington, Virginia: Information Resources.

opinando que “*es mejor, en este segundo caso, hacer uso de la palabra pertinencia*”.⁴⁸

Según Korfhage, la relevancia es la medida de “*como una pregunta se ajusta a un documento*”, y la pertinencia “*la medida en la que un documento se ajusta a una necesidad informativa*”.⁴⁹

La valoración de pertinencia de un documento recuperado es mucho más difícil de realizar ya que es el propio usuario el único que sabe si un documento se ajusta a su necesidad o no. Además de que la pertinencia cambia de un momento a otro, ya que la información conocida no es pertinente, puesto que no resuelve la necesidad informativa.

Por otra parte para la relevancia es más común establecer valores binarios en su cálculo, es decir, un documento es relevante, si sirve como respuesta a una pregunta le corresponde el valor 1 ó no sirve que sería el valor 0, aunque esto sería tajante ya que podemos encontrar una relevancia parcial en la que podemos clasificar información que pudiera resultar por un lado muy relevante, medianamente relevante o poco relevante.

La relevancia estaría ligada con el concepto de la relación efectiva entre los contenidos de un documento con una temática determinada, y la pertinencia delimitada a la utilidad evidente entre un documento recuperado y una necesidad de información individual.

2.5. Tendencias

El significado de la relevancia ha evolucionado con el paso del tiempo, de tal forma que podemos reducirlo en dos tendencias principales: 1) la relevancia

⁴⁸ Lancaster, F. W., Warner, A. J. (1993) Op. cit

⁴⁹ Korfhage, R. (1997). *Information Storage and Retrieval*, New York: John Wiley.

orientada al sistema o relevancia objetiva y 2) la orientada al usuario o relevancia subjetiva.

De tal forma encontramos a Vickery quien propició en 1959 un debate sobre la distinción entre “*relevancia de un tema*” que se refería a que los sistemas de recuperación de información juzgaban “*relevante*” y “*relevancia del usuario*” respecto de las necesidades del mismo.⁵⁰

A partir de entonces se inicia la división de perspectivas, por un lado encontramos la relevancia orientada al sistema como “*el resultado de la coincidencia entre los términos temáticos buscados en una consulta y los términos temáticos asignados a los documentos*”.⁵¹ Desde este punto de vista podemos considerar que todos los documentos recuperados por el sistema, por definición, serían relevantes para el usuario.

Por otro lado tenemos lo expresado por Reid,⁵² quien se refiere a la relevancia orientada al sistema, como la relevancia del sustituto de un documento a una pregunta, juzgada de acuerdo a un tema, cuando la pregunta es realizada en un sistema de recuperación de información. Para el caso tendríamos como ejemplo las búsquedas que se hacen en un catálogo de biblioteca, donde el registro bibliográfico es la respuesta que se obtiene y sería el sustituto, pues es sólo una representación del mismo y la búsqueda que se realiza sería por medio de un tema, entonces lo que realiza el sistema es la comparación entre el tema buscado y los temas registrados en la base de datos, el resultado de esa comparación para el sistema sería lo relevante a la pregunta realizada.

⁵⁰ Vickery, B. (1959). The structure of information retrieval systems. *Proceedings of the international conference on scientific information*, 2, 1275-1290.

⁵¹ Saracevic, T. (1970). The concept of “relevance” in information science: A historical review. En T. Saracevic (Ed.), *Introduction to information science (111-151)*. New York: R. R. Bowker.

⁵² Reid, J. (1999). A new task-oriented paradigm for information retrieval: implications for evaluation for information retrieval systems. *Digital libraries: interdisciplinary concepts, challenges and opportunities. Proceedings of the CoLIS3 Conference*. Zagreb, 97-108

A partir de 1980 se desarrollan teorías de relevancia orientadas hacia el usuario o relevancia subjetiva, Schamber, Eisenberg y Nilan quienes la definen como “*la utilidad, o potencial uso de los materiales recuperados, con relación a la satisfacción de los objetivos, el interés, el trabajo o los problemas intrínsecos del usuario*”.⁵³

Bajo esta perspectiva Boyce⁵⁴ señala que la relevancia subjetiva se estudia desde el punto de vista del usuario que consigue nueva información en un documento. De acuerdo con esta perspectiva, la información conocida no es relevante.

Estas dos tendencias subjetiva y objetiva se contraponen, mientras la primera se basa solamente en las propiedades de los mecanismos internos del sistema para poder cruzar los términos de la pregunta realizada contra la información de los temas asignados a los documentos para definir lo relevante, ignorando el papel de quién formula la pregunta en la otra encontramos al usuario como punto principal y es sobre el cual se plantea la respuesta que debe de dar el sistema de recuperación de información considerando el contexto individual y sus necesidades personales de información.

Por otra parte existen autores que están a favor de las dos tendencias, y señalan que la relevancia se compone tanto de aspectos objetivos como subjetivos.

En este sentido debemos considerar que los esfuerzos actuales se basan en lograr que tanto los usuarios como los sistemas de recuperación de información funcionen juntos para el beneficio de ambos, dando paso a modelos integrales.

Barry por su parte enmarca a la relevancia en relación a siete criterios:

⁵³ Schamberg, L., Eisenberg, B. and Nilo, S. (1990). A re-examination of relevance: toward a dynamic, situational definition. *Information Processing and Management*, 6, 755-775.

⁵⁴ Boyce, B. (1992). Beyond Topically: A two storage view of relevance and retrieval process. *Information processing and Management*, 18, 105-109.

- 1) *“Información que contiene un documento: se refiere a las características propias de la información contenida en el documento, si es extensa, precisa, efectiva, clara, reciente.*
- 2) *Experiencia previa del usuario: cuál es la experiencia del usuario en el tema, su formación, habilidad de entender la información, si la información es novedosa para él.*
- 3) *Creencias y preferencias del usuario: se refiere a que tanto el usuario está de acuerdo con la información, de acuerdo con su punto de vista, o también con la respuesta emocional que el documento cause en el usuario.*
- 4) *Otras informaciones y fuentes: pertenece a la relación entre el contenido de la información y la información o fuentes del área*
- 5) *Fuentes del documento: a esta categoría pertenecen las evaluaciones de las fuentes de los documentos.*
- 6) *Documento como entidad física: se refiere a la observación del documento como una entidad física, como un objeto a ser obtenido, así como el costo que tendría de acuerdo con su extensión o ubicación.*
- 7) *Situación de los usuarios: el grupo final se encarga de examinar cuales son los factores situacionales que afectan al usuario”.*⁵⁵

De estos criterios, tres se puede considerar como objetivos: 1) Información que contiene un documento, 4) Otras informaciones y fuentes y 5) Fuentes del documento, ya que atienden a la información que se obtiene de un sistema de información; y cuatro subjetivos 2) Experiencia previa del usuario, 3) Creencias y preferencias del usuario, 6) Documento como entidad física y 7) Situación de los usuarios, que atienden a puntos de vista relacionados con ellos.

Por su parte Ingwersen propone un modelo cognitivo para la recuperación interactiva de información el cual se compone de tres elementos:

⁵⁵ Barry, C. L. (1994). User -defined relevance criteria: An exploratory study”. *Journal of the American Society for Information Science* 1994, 45, 149-159.

- 1) *“Sistema: comprende los documentos u objetos de información (que pueden ser representados de varias formas), que están organizados en un archivo, que a través de un algoritmo se prepara para combinarse con una consulta vía un mecanismo de interfaz.*
- 2) *Usuario y entorno: el usuario comúnmente tiene un problema o tarea por realizar que deriva en una necesidad de información y*
- 3) *Entorno socio-organizacional: va a proveer el contexto o la base situacional que va a influenciar las actividades del usuario”.*⁵⁶

En el enfoque orientado al sistema, la relevancia es considerada como una propiedad del mismo sistema. Mientras que en la relevancia orientada al usuario y en los procesos cognitivos, se da a través del proceso cognitivo del usuario y su actualización de conocimiento y la necesidad respecto a la información estimulada por el contexto.

Es claro que el concepto cubre una gran área del conocimiento y tal vez por eso los últimos estudios se concentran en la interacción entre el usuario y el sistema para tratar de establecer qué es finalmente.

2.6. Tipología

A partir de las tendencias enfocadas al usuario o al sistema, podemos considerar partir de la existencia de diferentes tipos de relevancia, tales como:

- 1) Relevancia lógica: basada en lo establecido por Cooper,⁵⁷ sobre la aplicación de la lógica matemática, mencionada anteriormente.
- 2) Relevancia temática: que se caracteriza por la relación entre la temática de la solicitud de información y los objetos de información, puede ser medida

⁵⁶ Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of documentation*, 52, 3-50.

⁵⁷ Cooper, W. S. (1971). A definition of relevance for information retrieval. *Information Storage & Retrieval*, 7, 19-37.

en términos de la aproximación entre la información y la pregunta. Es un tipo de relevancia orientada al sistema ya que depende totalmente de los procesos que se realizan dentro del sistema de recuperación de información, tanto en la indexación de la información y de los sistemas de búsqueda para recuperar información relevante. Entre los autores que han trabajado este tipo de relevancia encontramos a Cooper⁵⁸ y Boyce⁵⁹.

- 3) Relevancia cognitiva: que es medida con base en la relación entre el conocimiento o necesidad de información del usuario y como interpreta la información que encuentra, si es novedosa, informativa, adecuada. En este tipo de relevancia están involucrados ambos enfoques tanto el del sistema como el del usuario, pues involucra, por un lado, como en el caso anterior tanto la indexación como el sistema de búsqueda para recuperar información relevante y por el otro depende de que el usuario sepa como formular su petición, su conocimiento previo y al entendimiento de su necesidad de información.
- 4) Relevancia situacional: que se mide en términos de la relación entre la situación percibida y la utilidad de la información percibida por el usuario.⁶⁰

La relevancia la podemos clasificar en distintos tipos de acuerdo con criterios diferentes, pero dentro de todos consideramos que merece un poco más de atención la relevancia situacional, pues considera elementos que permitirán una delimitación más concreta.

La relevancia ocupa un espectro multidimensional y dinámico, que podemos entender como la unión de elementos cognitivos y factores situacionales que dan como resultado procesos dinámicos de juicios de relevancia, aprendizaje y

⁵⁸ Cooper, W. S. (1971). Op. cit.

⁵⁹ Boyce, B. (1982). Beyond topicality: A two stage view of relevance and the retrieval process. *Information Processing & Management*, 18,105-109.

⁶⁰ Hersh, W. (1994). Relevance and retrieval evaluation: Perspectives from medicine. *Journal of the American Society for Information Science*, 45, 201-206.

desarrollo de necesidad de información. El tipo de relevancia que cumple con estos requerimientos es la relevancia situacional.

El primero que manejó el concepto relevancia situacional fue Patrick Wilson, que introdujo la idea de que los juicios de relevancia varían con la situación o la tarea. El define la relevancia situacional como la relación entre el objeto de información y quien recibe la información y su punto de vista del mundo y su situación en él. Un objeto de información es situacionalmente relevante si trae consigo un cambio en quien recibe la información.⁶¹

Para Cosijn, la relevancia situacional se refiere a la relación entre la situación que se percibe, tarea o problema que se tiene y la utilidad de los objetos de información como son percibidos por el usuario.⁶²

De tal forma podemos considerar a la relevancia situacional, como la utilidad o beneficio, que se toma de la información que fue solicitada, con relación a una serie de elementos, tales como los intereses, finalidad o problemas, los cuales establecieron por principio la búsqueda de la información. Sin embargo esta relevancia se ve afectada por dos elementos, el primero sería la capacidad del sistema en la indexación de la información y el segundo la habilidad de recuperar información relevante.

A partir de la información anterior podemos considerar como la propuesta más apropiada, la de Mizzaro.⁶³ Éste establece que la relevancia debemos clasificarla dentro de un espacio de cuatro dimensiones.

En la primera dimensión tenemos uno de los siguientes tres elementos y podemos identificarlo como fuentes de información:

⁶¹ Wilson, P. (1973). Situational relevance. *Information Storage and Retrieval*, 9 (8), 457-471.

⁶² Cosijn, Erica; Ingwersen, Peter. (2000). Dimensions of relevance. *Information processing and management*, 36, 533-550.

⁶³ Mizzaro S. (1998). How many relevances in information retrieval?. *Interacting with Computers*, 10, 303-320.

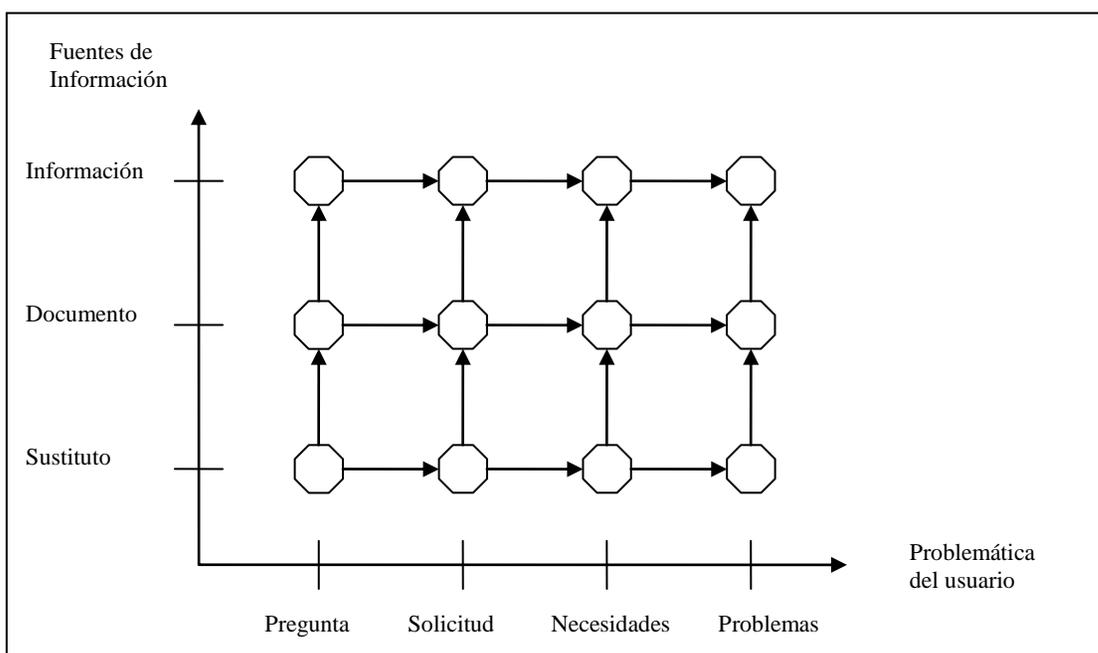
- ☑ Documento: la entidad física que el usuario de un sistema de recuperación de información (SRI) obtiene después de una búsqueda de información.
- ☑ Substituto: la representación de un documento. Puede asumir diferentes formas y puede estar formado por una o más de las siguientes: título, lista de palabras, nombre(s), de autor(es), datos bibliográficos (datos de publicación, editores, páginas, entre otras), resúmenes, extractos (partes de un documento) y otras.
- ☑ Información: la entidad (no física) que el usuario recibe/crea cuando lee un documento.

En la segunda dimensión tenemos cuatro elementos que representan la problemática del usuario:

- ▶ Problemas. Aquel que un humano enfrenta y que requiere información para resolver.
- ▶ Necesidades de información. Representación del problema en la mente del usuario. Es diferente del problema porque el usuario puede no percibir de forma correcta su problema.
- ▶ Solicitud. Representación de las necesidades de información del usuario en un lenguaje “humano”, usualmente en lenguaje natural; y
- ▶ La pregunta. Formulación de búsqueda de información en un lenguaje de “sistema”, por ejemplo booleana.

La relevancia puede ser vista como una relación entre estas dos entidades, una de cada grupo: la relevancia de un sustituto para una pregunta, o la relevancia de un documento para una solicitud, o la relevancia de la información recibida por el usuario para sus necesidades de información, etcétera, que se presentan en el cuadro 1.

Cuadro 1. Dos dimensiones de la relevancia.



(Fuente: Mizzaro S. (1998). How many relevances in information retrieval?. *Interacting with Computers*, 10, 303-320.

En éste podemos apreciar la relevancia vista como un punto en un espacio de dos dimensiones. En el eje horizontal tenemos los elementos de la problemática del usuario: Pregunta, Solicitud, Necesidad, Problemas, mientras en el eje vertical tenemos los elementos de las fuentes de información; cada relevancia es representada por un octógono y las flechas representan un orden parcial entre las relevancias. Este orden implica lograr que la relevancia esté cerca de la relevancia de la información obtenida y de la problemática del usuario, llegando al punto más alto.

A estas dimensiones debemos agregarle la tercera dimensión que sería el tiempo que transcurre desde el inicio del problema hasta su solución, la interacción entre

estas dimensiones cambiará a lo largo del tiempo, un documento, información o sustituto puede ser relevante a una solicitud en cierto punto de tiempo o lo contrario en otro tiempo.

A partir de aquí tenemos la cuarta dimensión (cuadro 2) que puede ser dividida en los siguientes tres componentes:

❖ Tema:

Se refiere al área temática en la que el usuario este interesado. Por ejemplo; el concepto de relevancia en bibliotecología y estudios de la información.

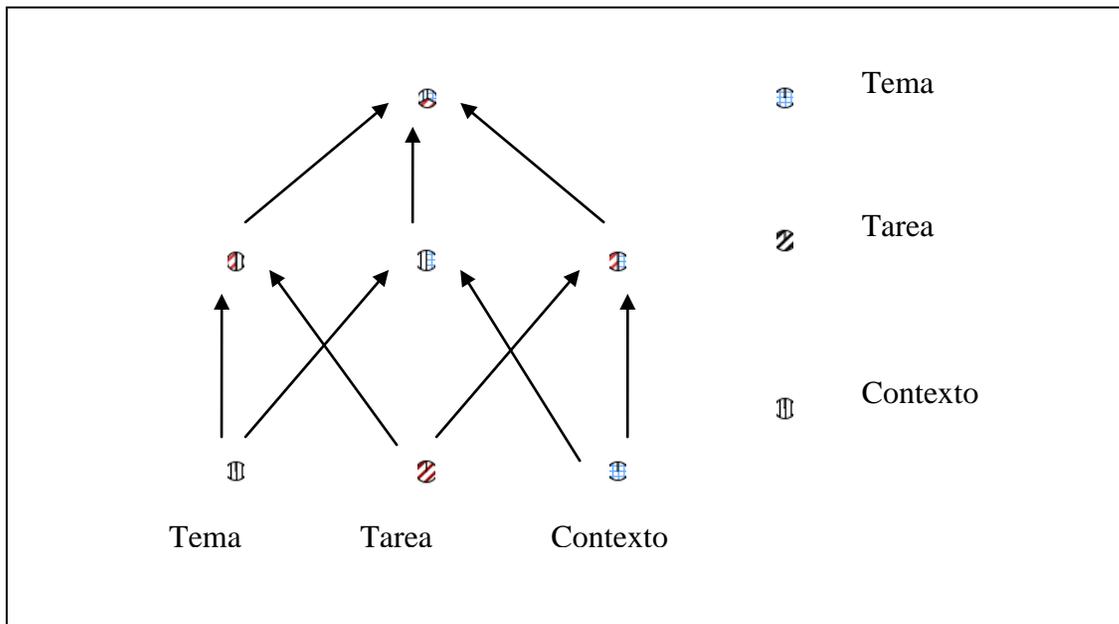
❖ Tarea:

La actividad que el usuario va a ejecutar con los documentos recuperados. Por ejemplo; para escribir un documento.

❖ Contexto:

Aquello que va a incluir todo lo no pertinente al tema y tarea, pero que sin embargo afecta la forma en que la búsqueda toma lugar y la evaluación de resultados. Por ejemplo; los documentos que ya son conocidos por el usuario (así como los que no valdría la pena recuperar), tiempo y/o dinero disponibles para la búsqueda, etcétera.

Cuadro 2. Cuarta dimensión de la relevancia.



(Fuente: Mizzaro S. (1998). *How many relevances in information retrieval?*. *Interacting with Computers*, 10, 303-320.

Usualmente, la investigación en recuperación de información se concentra en el componente temático, pero al usuario no está interesado en obtener información inservible para la tarea que él tiene que ejecutar, o de un documento que ya conoce.

En otras palabras, un sustituto, un documento, alguna información, es relevante a una pregunta, solicitud, necesidad de información, problema que requiere información, con respecto a uno o más de los componentes, es posible para hablar de “relevancia de un sustituto a una pregunta para que refiera el tema, tarea, contexto” o “la relevancia de un documento para una solicitud para que corresponda el tema y tarea” o “la relevancia de la información recibida por el usuario a la necesidad de información de todos los tres componentes”, etcétera.

A partir de este punto la situación de búsqueda de información toma un lugar en un intervalo de tiempo: el usuario tiene un problema, lo percibe, interactúa con un SRI (y con apoyo de un intermediario humano), expresa su necesidad de información en una solicitud, formalizándola en una pregunta, examina los

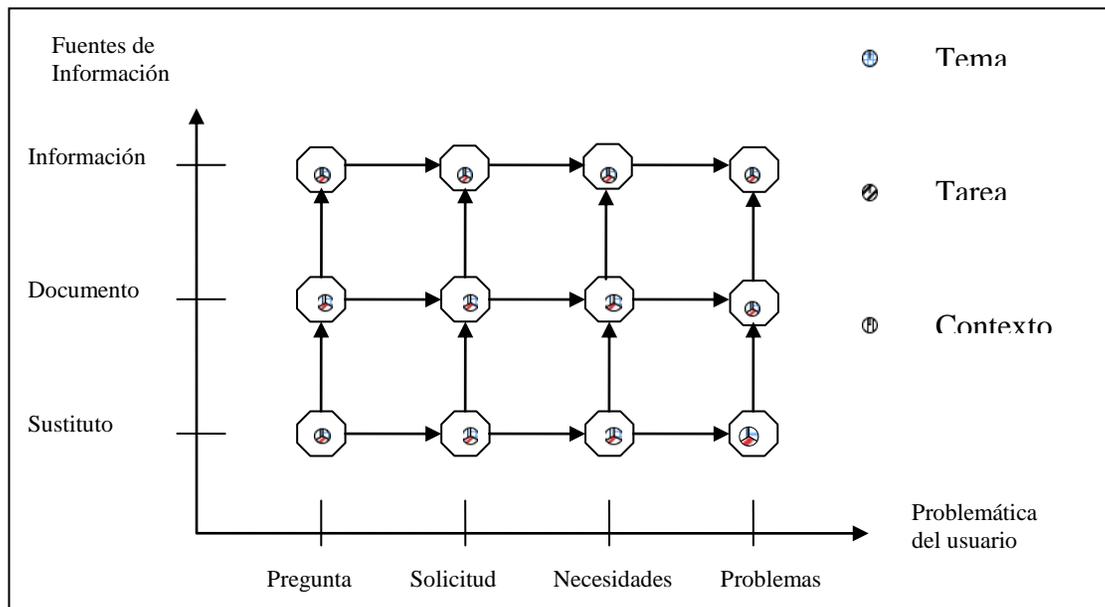
documentos recuperados, reformula la pregunta, reexpresa sus necesidades de información, percibe el problema de diferente manera, etcétera.

También se debe tomar en cuenta el tiempo que: un sustituto (un documento, cierta información) puede ser no relevante a una pregunta (solicitud, necesidad de información, problema) en un cierto punto de tiempo, y ser relevante después, o viceversa, esto sucede, por instantes, si el usuario aprende algo que le permita entender un documento, o si el problema del usuario cambia, etcétera.

Por lo tanto, la relevancia puede ser vista como un punto en un espacio de cuatro dimensiones (cuadro 3), los elementos de cada una de las cuatro dimensiones serían:

- 1) Sustituto, documento, información
- 2) Pregunta, solicitud, necesidad de información, problema
- 3) Tema, tarea, contexto y la combinación de ellos
- 4) Los varios instantes de tiempo surgidos de un problema que requiere información hasta su solución

Cuadro 3. Dimensiones de la relevancia.



Fuente: Mizzaro S. (1998). How many relevances in information retrieval?. *Interacting with Computers*, 10, 303-320.

A partir del modelo de Mizzaro tendríamos que ir colocando los diferentes elementos como son: al usuario, el tipo de información que está solicitando, problemas que busca solucionar con la información requerida, como un punto dentro del modelo, de tal forma que podríamos ir seleccionando el tipo de información que se solicita dentro del SRI, sumando cada uno de los puntos se buscaría que la información recuperada coincida con la pregunta y que sea de utilidad para quién realiza la búsqueda.

Se trata de que el sistema se adapte a la necesidad de la persona o al grupo de personas, esto cuando se trate de individuos que compartan alguna característica, logrando que se obtenga toda la información relevante que se encuentre en el SRI, que dicha información sea nueva, de tal forma que se evite traer información poco relevante por ser ya conocida.

2.7. Juicio de relevancia

Un juicio de relevancia es la asignación de un valor a lo relevante de un documento por un individuo en cierto punto de tiempo. Existen muchas clases de juicios de relevancia, pero podemos agruparlos en cinco categorías según Mizzaro:

1. La clase de relevancia juzgada.
2. La clase de juez (usuario o quién juzga).
3. Que es lo que usa el juez (sustitutos, documentos o información) para expresar su juicio de relevancia, en este punto debemos conocer pues el juicio puede por ejemplo ser dado con base en un sustituto.
4. Si utiliza preguntas, solicitudes, necesidades de información o problemas, para expresar su juicio de relevancia.

5. El tiempo en el que el juicio es expresado, esto es necesario pues un juicio puede cambiar entre un cierto tiempo y otro.⁶⁴

El juicio de relevancia se convierte entonces en la quinta dimensión pues combina los elementos de las cuatro dimensiones vistas anteriormente para formar el juicio que determine si se tiene relevancia o no relevancia.

2.8. Medición de la relevancia

Se encuentran dos principales medidas de la relevancia, por un lado encontramos un método manual y por el otro uno conocido como “polling”.

El método manual consiste en la revisión de los documentos uno a uno para saber si se adecuan o no como respuesta a una pregunta formulada. En la mayoría de las veces es relevante establecer si un documento para una pregunta determinada resulta, por ello es conveniente que los juicios los haga más de una persona y de ser posible un número impar de especialistas.

El principal problema que presenta este método, es que en colecciones muy grandes, hay que invertir gran cantidad de recursos informáticos, además algunas bases de datos son más especializadas que otras, lo que implica contar con un número mayor o menor de especialistas, de tal forma que para librar estos inconvenientes se implementan colecciones experimentales, en las cuales se fijan de antemano qué documentos son relevantes para cada pregunta.

El segundo método es el conocido como “*polling*” el cual es utilizado en bases de datos muy grandes, donde se hace imposible evaluar cada uno de los documentos para saber cuáles de ellos son relevantes y cuáles no, funciona mediante el análisis de una cantidad determinada de documentos arrojados de la búsqueda en diferentes sistemas, partiendo del hecho de que todo lo recuperado es relevante y

⁶⁴ Mizzaro, S. (1998), op.cit.

aquello que no fue recuperado no lo es, de estos se eligen los primeros de cada resultado y es sobre estos sobre los cuales se realiza el análisis por parte de los investigadores, de tal forma que son ellos quienes determinan si los resultados fueron relevantes o no.⁶⁵ Se podría consultar constantemente un SRI para indicarle al usuario cuando lleguen nuevos contenidos a la misma, sin embargo esto implica un consumo de recursos que puede afectar al final al SRI.

Las medidas que se basan en el juicio acerca de la relevancia que alguien hace ya sea usuario del documento recuperado o en algunos casos la representación de éste, de igual manera cuando la búsqueda no la hace directamente un usuario encontramos desacuerdo entre los investigadores a la hora de considerar qué documentos eran o no relevantes.

Considerando que depende del juicio de una persona tenemos un gran número de variables y factores que pueden afectar al juicio sobre la relevancia, Schamber,⁶⁶ en una revisión de la literatura, identifica hasta ochenta factores diferentes, comentando además que no es una lista completa. Entre algunas de las que señala están: inteligencia, educación formal, concepto de relevancia, importancia, pertinencia, facilidad para detectar la relevancia, tiempo para realizar su juicio, entre otras.

De todas estas variables, algunas son de tipo anímico y psicológico, que dependen del estado cognitivo del usuario y de las circunstancias o situación del momento, lo que agrega todavía más complejidad a cómo realiza sus juicios de relevancia.

⁶⁵ Kowalski, G. *Information retrieval systems theory and implementation*. Boston: Kluwer Academic Publishers, 1997. p. 156

⁶⁶ Schamber, L. "Relevance and information behavior". En: *Annual Review of Information Science and Technology*. Meford, NJ: American Society for Information Science, 1994, vol. 29, p. 19.

Consideraciones últimas

Hablar de la búsqueda de la relevancia nos remonta a los principios mismos de la búsqueda de cualquier tipo de información, pues siempre se ha necesitado contar con información que aporte nuevos elementos a los que ya se tenían, de tal manera que la relevancia la podemos asociar con la naturaleza misma de la inquietud humana de conocer nuevas cosas y por lo tanto de un fenómeno que en su principio no ha cambiado pero si en las herramientas que usa para lograrlo.

La relevancia es un elemento dinámico y por lo mismo estudiado por diferentes disciplinas, sin embargo, encontramos la dificultad que existe para definirla, como nos señalan los autores consultados, dado que solamente las personas pueden asignar aquellos juicios de valor sobre lo que les resulta relevante o no relevante al momento de determinar respecto de una información obtenida.

La tecnología que se utilizan para la búsqueda de la relevancia ha tenido una evolución a la par de los cambios tecnológicos, desde la creación de las primeras computadoras, los sistemas de recuperación de información, la Internet y actualmente sobre aplicaciones Web, por lo que encontramos como han cambiado los enfoques desde el orientado al sistema en el que la consideración de los usuarios de los sistemas no era tomada en cuenta, el enfoque hacia el usuario donde ya es visto como elemento principal y finalmente la relevancia situacional donde se trata de considerar todos aquellos elementos que influyen para la determinación de ese valor de relevancia.

Realizar estudios de relevancia trae consigo la reconsideración de los sistemas de recuperación de información pues con base en ese conocimiento estamos en posibilidad de diseñar o solicitar que nuestros sistemas manejen niveles de relevancia en la recuperación.

Considerando que la relevancia es un tema poco explorado en nuestro país, se hace necesaria la investigación del tema, surge el interés por ahondar más en el mismo y que se considere tema de estudio que ayude a resolver y a apoyar los desarrollos de sistemas de recuperación de información.

Como ya hemos mencionado, el objetivo de un sistema de recuperación de información es obtener toda aquella información relevante para el usuario ya que éste espera una respuesta exacta y perfecta a su pregunta. Los bibliotecólogos, entonces, deben considerar la aplicación de modelos de recuperación de información basados en criterios de relevancia a fin de lograr dicho objetivo.

Debemos considerar, también, el hecho de cómo han cambiado los sistemas de recuperación de información, donde el acceso a la misma se está realizando en la actualidad directamente por el usuario, por lo tanto deben diseñarse interfaces para realizar las búsquedas de información más amigables y donde no importe el formato, pues ahora tenemos la posibilidad de poner al alcance de cualquiera información no sólo en forma de documentos de texto, sino que ahora se utilizan recursos multimedia.

Sin embargo, nos enfrentamos a un hecho particularmente importante que aun con todo el peso que tiene la relevancia en la recuperación de la información, en nuestro país, al parecer, este tema no ha tenido la importancia suficiente o no está explícita, cómo para que realicemos estudios sobre ella, o por lo menos con la formalidad de llevar a cabo estudios o experimentos para determinarla, podría parecer que con la simple asignación temática que hacemos a los documentos es suficiente para lograr satisfacer las necesidades informativas de los individuos que consultan nuestros sistemas de recuperación de información.

CAPÍTULO 3

Criterios de relevancia como base para el diseño de una interfaz de recuperación de información

3.1. Introducción

En este capítulo se presenta la metodología que se seguirá para llevar a cabo el estudio de criterios de relevancia en recuperación de información de determinadas comunidades epistémicas.

De acuerdo con la clasificación de las áreas de conocimiento de la ANUIES se presentan las distintas disciplinas que se estudian en nuestro país.

A partir de las áreas del conocimiento se determinarán cuáles son los parámetros que se siguen al momento de realizar la búsqueda de información en un sistema de recuperación de información.

Se contrastarán elementos de información entre los parámetros de búsqueda contra los criterios de relevancia para obtener cuáles son los indicadores que servirán para la propuesta.

Se establece una propuesta de indicadores de relevancia de búsqueda de información de las disciplinas de Ciencias de la Salud e Ingeniería y Tecnología.

3.2. Proceso de búsqueda de información

El objetivo de cualquier biblioteca o servicio de información es proveer el acceso a los recursos, información e ideas. Este acceso por su parte requiere de un proceso de búsqueda de información con la finalidad de poder ofrecer a quien busca información relevante, por ello debemos hacer mención acerca de cómo se lleva a cabo este proceso.

Para poder determinar cómo se lleva a cabo encontramos que Carol Kuhlthau, uno de los investigadores del proceso de búsqueda de información señala que el proceso se puede dividir en seis etapas:

- 1) *“Iniciación, que empieza cuando una persona comienza la búsqueda de conocimiento o entendimiento crean incertidumbre y aprehensión común.*
- 2) *Selección, una área general, temática o problema es identificado y la incertidumbre inicial da el paso a un sentido de optimismo y una preparación para comenzar la búsqueda.*
- 3) *Exploración, inconsistencias, información incompatible es encontrada y la incertidumbre, la confusión y la duda frecuentemente se incrementa.*
- 4) *Formulación, se enfoca una perspectiva y la incertidumbre disminuye por un lado y la confianza comienza a aumentar.*
- 5) *Colección, la información pertinente a la perspectiva buscada es obtenida y la incertidumbre se reduce dentro del aumento del interés en el proyecto.*
- 6) *Presentación, la búsqueda es completada, se tiene un nuevo entendimiento de la persona para explicar su conocimiento a otros, o de alguna forma poner en uso su aprendizaje”.*⁶⁷

⁶⁷ Kuhlthau, C. (2005) Kuhlthau’s information search process. *Theories of information behavior*. Medford, New Jersey : American Society for Information Science and Technology by Information.

Conforme pasan las etapas mencionadas por Kuhlthau encontramos que las personas que llevan a cabo la investigación de una temática se tienen que enfrentar en primera instancia con el aumento de las dudas sobre el tema que está buscando. Pero conforme se relaciona con la información encontrada la incertidumbre va disminuyendo y aumenta la confianza, así como el conocimiento sobre el tema de estudio aumenta, hasta llegar a una asimilación de esta información.

David Ellis por su parte interpreta las actividades de búsqueda de información por etapas y las describe de la manera siguiente:

- 1) *“Inicio – actividades características del inicio de la búsqueda de información*
- 2) *Encadenamiento – siguiendo la cadena de citas u otras formas de conexiones referenciales entre el material*
- 3) *Hojear – búsqueda semidirecta en un área de potencial interés*
- 4) *Diferenciación – usando las diferencias entre recursos como un filtro de la naturaleza y calidad del material examinado*
- 5) *Supervisión – mantener un seguimiento en el desarrollo en un campo por el monitoreo de un recurso en particular*
- 6) *Extracción – trabajo sistemático sobre un recurso particular para localizar materiales de interés”*.⁶⁸

Al comparar las actividades propuestas tanto por Kuhlthau como por Ellis encontramos que ambos coinciden en que al desarrollar cualquier búsqueda de

⁶⁸ Ellis, D. (2005). Elli's model of information-seeking behavior. *Theories of information behavior*. Medford, New Jersey : American Society for Information Science and Technology by Information. pp. 138-139

información se debería partir de lo general a lo específico para establecer la relevancia de la información.

En ambas propuestas la persona se acerca a las fuentes de información por la incertidumbre relativa a una temática, por lo que busca el entendimiento o la comprensión de la misma.

La siguiente etapa según Kuhlthau es la de seleccionar, mientras que para Ellis es hojear. Aunque los autores utilizan términos diferentes en esencia es lo mismo, pues se identifica el área de interés.

La exploración y la diferenciación coinciden de igual manera en la determinación de filtrar aquella información que es encontrada.

También encontramos que la formulación y la supervisión concuerdan en enfocar el interés de la búsqueda en un tema; del mismo modo la colección y la extracción realizan la misma acción de recuperación de información sobre el tema en particular.

Finalmente Kuhlthau incluye la *Presentación* que no considera Ellis y por su parte Ellis incluye el *Encadenamiento* y Kuhlthau no. Estos dos puntos no considerados tanto por uno como por el otro debemos pensarlos como complementarios, toda vez que al ordenarlos se integran al flujo que se presenta en la búsqueda de información.

De las ideas de los autores mencionados podemos concebir una búsqueda de información de manera general. Por lo que se propone el orden siguiente:

- 1) Inicio de la búsqueda de información para satisfacer una necesidad de información.

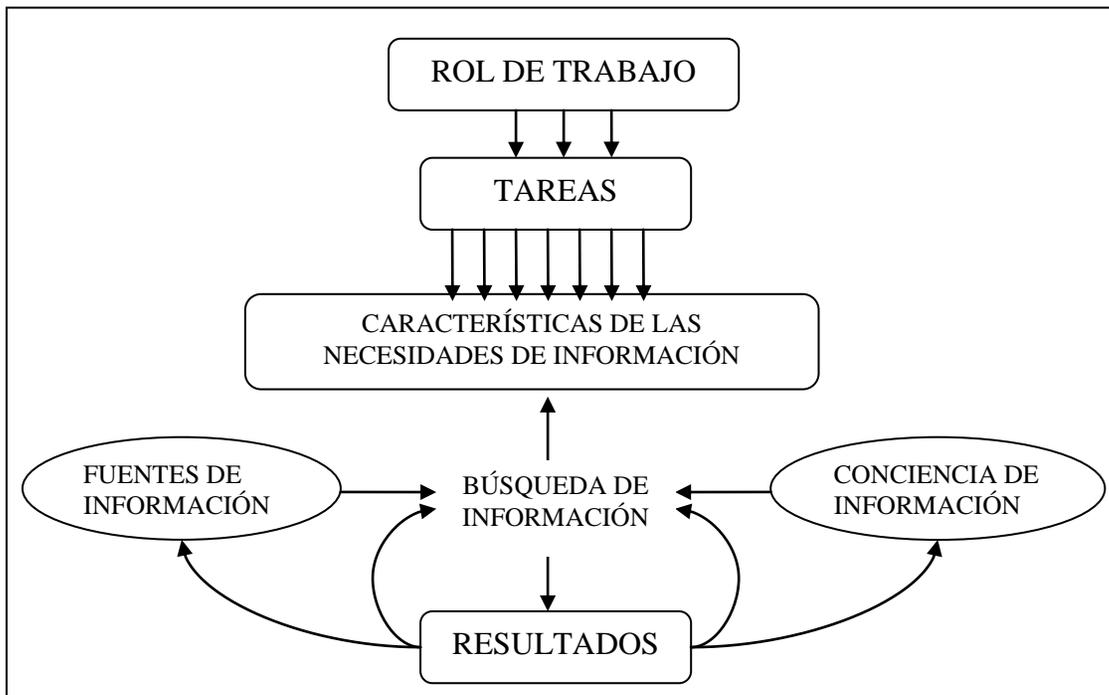
- 2) Encadenamiento de las referencias relacionadas con la información que se está buscando.
- 3) Selección del área general de potencial interés.
- 4) Filtrado de la información a fin de descartar aquella que no resulta relevante con el interés de quien busca.
- 5) Desarrollo de la búsqueda enfocada en el perfil particular de interés.
- 6) Obtención de aquellos materiales de interés.
- 7) Presentación de la información resultante una vez completada la búsqueda, donde la persona puede exponer el conocimiento obtenido.

Esta propuesta es considerada para un nivel general donde a los usuarios podemos ubicarlos en nivel medio superior.

Ahora bien, en cuanto a áreas o disciplinas muy particulares y como se desarrolla la búsqueda de información por parte de ciertos grupos, en este caso por profesionales que realizan búsquedas de información para resolver las tareas que se presentan para llevar a cabo sus actividades encontramos que G. Leckie, K. E. Pettigrew y C. Sylvan⁶⁹ desarrollaron un esquema sobre como se lleva a cabo la búsqueda de información por parte de profesionales.

⁶⁹ Leckie, G., Pettigrew, E., Sylvain, C. (1996) Modeling the information seeking of professionals. En: Library quarterly, 66 (2), p. 180

Figura 2: Modelo de búsqueda de información de profesionales



Leckie, G., Pettigrew, E., Sylvain, C. (1996) Modeling the information seeking of professionals. *Library quarterly*, 66 (2), p. 180

El esquema de la figura 2 sugiere que solo a través del entendimiento de los roles de trabajo y sus tareas asociadas seremos capaces de realmente entender por qué, cómo y cuándo la búsqueda de información puede ocurrir. Los autores concluyen que la búsqueda de información y la información relacionada con la práctica donde se vuelven más similares a través de diversas profesiones

El esquema asume que el trabajo provoca que los roles y tareas toman lugar dentro de un contexto específico para una posición de trabajo. El contexto fue deliberadamente dejado indefinido y se anticipan que factores contextuales, como son la ideología y las relaciones de poder de la organización; tienen un impacto sobre el esquema de trabajo; en los sitios y lugares particulares que comienzan a ser estudiados.

3.3. Proceso de búsqueda de información de los profesionales

Diferentes autores mencionan que las actividades de búsqueda de información de diferentes tipos de profesionales son tan diversas como los profesionales que representan.

Algunos módulos consideran al individuo como punto de partida otros miran al contexto organizacional y otros más presentan solo la información buscada y su relación con una o dos claves de la actividad de trabajo.

Sin embargo la gran diversidad de trabajos conceptuales y representaciones gráficas, llevan a supuestos que son comunes en muchos modelos.

Los puntos importantes dentro de este modelo de información serían:

- I. Proveer de varios tipos de servicios a sus clientes sería la principal actividad para todos los profesionales.
- II. Un entendimiento sólido del entorno de trabajo no es suficiente en si mismo. Para desarrollar ideas del proceso de búsqueda de información de profesionales, los detalles de ese trabajo individual deben ser examinados en profundidad la necesidad de cierto tipo de información, el proceso de recuperación y los usos que se hacen de ella se derivan de las formas en que la práctica profesional se lleva a cabo a diario.
- III. En relación con el punto anterior, una revisión de los elementos detallados del trabajo profesional, revela que los profesionales tienen responsabilidades complejas y se espera que para ejecutarlas satisfactoriamente interactúen con diversos roles, tales como: administrativos, interpersonales que a menudo son muy diferentes y se complican aún más cuando se asocian a las tareas que deben llevar a cabo.

Juntos estos roles y tareas generan necesidades de información que deben ser satisfechas para avanzar con la tarea que se tiene establecida

Finalmente, muchos estudios de diversos grupos profesionales concluyen que frecuentemente los profesionales se frustran en sus búsquedas de información relevante y necesaria. Dicha frustración surge del gran número de variables complejas e interactivas que influyen en los procesos de búsqueda de información.

3.4. El modelo propuesto de búsqueda de información de los profesionales

Considerando los puntos anteriores, el supuesto básico en el modelo establece que los roles y tareas relacionadas llevadas a cabo por profesionales en el curso de su necesidad diaria de información particular dan a su vez lugar a un proceso de búsqueda de información.

Existen diversos estudios que demuestran que los profesionales deben asumir diferentes roles en el curso de su día de trabajo.

Se habla de cinco principales roles: proveedor de servicios, administrador o manejador, investigador, educador y estudiante.

Dentro de estos roles existen tareas específicas, constituyendo una segunda capa en el componente rol/tarea, en el modelo los investigadores muestran que la búsqueda de información está altamente relacionada con la realización de un rol particular y su tarea asociada.

Sin embargo, las características de información en cada caso son diferentes, debido a que están conformados por una serie de variables que intervienen. Estas diferencias pueden afectar el resultado de la búsqueda de información.

Algunas de las influencias en las necesidades de información de los profesionales incluyen:

- ◆ demografía individual (edad, profesión, especialización, localización).
- ◆ contexto (situación específica de la necesidad, impulso interno o externo).
- ◆ frecuencia (necesidad recurrente o nueva).
- ◆ predictibilidad (anticiparse a la necesidad, o inesperada).

Más allá de la naturaleza de la necesidad misma, la forma en la que la información es influenciada por otros factores importantes. En el modelo, el primero de esos factores ha sido etiquetado como recursos.

La búsqueda de información de los profesionales abarca un interminable número de recursos: colegas, bibliotecarios, manuales, artículos de revistas y su propio conocimiento personal y experiencia. Estos recursos de información pueden ser en términos generales caracterizados en tipos de canales o formatos. Incluyen: formales (por ejemplo una conferencia, una revista) o informales (una conversación), internas o externas (recursos dentro de una organización o fuera). Los recursos disponibles y las preferencias individuales por ciertos canales, pueden afectar la búsqueda de información.

El segundo grupo de factores que influyen en la búsqueda de información han sido etiquetados como conocimiento indirecto de varios recursos de información (como son colegas, una base de datos en línea o un manual) y la percepción formada de el proceso o acerca de la recuperación de información, tienen un rol crucial en el proceso general de búsqueda de información.

Por lo tanto, el conocimiento general acerca de las fuentes de información y/o contenido, pueden determinar el camino que la búsqueda de información tome.

Algunas de las variables que están involucradas en la convivencia individual de lo que debe de ser: accesible (relativamente fácil de acceder), puntualidad (que se halle cuando se necesite) y costo (relativo al costo-efectividad).

El resultado de la búsqueda de información puede ser etiquetado como el punto final del trabajo-relativo, requerimientos de roles específicos y de tareas específicas.

El resultado que se tiene es que cuando se logra reunir la información que se necesita, el profesional logra la tarea en cuestión, tal como diagnosticar una dolencia o completar un reporte técnico.

Algunos de los resultados comunes de la búsqueda de información identificados en la literatura incluyen el proveer un servicio y/o un producto; realizar las metas operativas y lograr las metas para el desarrollo profesional.

Sin embargo es posible que los resultados de la búsqueda de información, así mismo como la necesidad no sean satisfechos y más búsqueda de información sea requerida, situación que se ha caracterizado en el modelo como una retroalimentación.

De esta forma se realiza una segunda ronda de búsqueda donde son ajustados aquellos factores que se puedan modificar a fin de obtener resultados diferentes. Es como completar diferentes tipos de recursos y factores involucrados.

La retroalimentación también ilustra que el resultado de la búsqueda de información no es un evento en una dimensión. Esto hace posible que el resultado de una tarea asociada con un rol particular (tal como asesorar clientes en el rol de asesoramiento).

3.5. Metodología

Para poder determinar los criterios que se utilizan para juzgar la relevancia de la información recuperada, se decidió recurrir a la tabla de áreas de estudio elaborada por la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES), que es la asociación que ha participado en la formulación de programas, planes y políticas nacionales y que agremia a las principales instituciones de educación superior del país.

La ANUIES en la década de 1970 diseñó dicha tabla para agrupar las áreas del conocimiento, la cual es utilizada hasta nuestros días para realizar la estadística de la educación superior del país.

Se seleccionó esta división del conocimiento para sistematizar la información de las áreas del conocimiento con el fin de encontrar los parámetros de semejanzas y/o diferencias que se presenten entre las diferentes disciplinas.

Las áreas de conocimiento en la educación superior mexicana son seis en total.

Áreas de conocimiento:

1. Ciencias Agropecuarias:

Considera, entre otras, a dos disciplinas principalmente que le dan sentido y representatividad al nombre del área Agronomía y Veterinaria.

2. Ciencias de la Salud:

Considera a las disciplinas que se interrelacionan para manejar problemas tanto individuales como colectivos del hombre, respecto a deficiencias físicas y mentales. En esta área encontramos Medicina, Enfermería, Odontología, entre otras.

3. Ciencias naturales y exactas:

Considera a las primeras como aquellas creaciones fuera del alcance del hombre pero que sin embargo este puede transformar para su beneficio. Se tiene un objeto físico, material, palpable y que pertenece al mundo real. Por otro lado tenemos las ciencias formales donde los objetos no son cosas ni procesos, sino formas en las que se puede verter contenidos tanto fácticos como empíricos. Incluye Biología, Física, Matemáticas, entre otras.

4. Ciencias sociales y administrativas:

Se consideró agrupar estas dos disciplinas con el fin de evitar el problema de delimitar donde termina lo social y empieza lo administrativo. Incluye Administración, Ciencias Políticas y Administración Pública, Sociología, Derecho y otras.

5. Educación y Humanidades:

Consideran al humanismo contemporáneo a partir de un contexto real y realizado, configurado por los diversos fenómenos del hombre en sus propias obras. Incluye Pedagogía, Letras, Historia, Filosofía y otras.

6. Ingeniería y Tecnología:

Se basa en la necesidad, acentuada en los últimos años, de fomentar el desarrollo tecnológico del país en diversas áreas. Comprende la ingeniería (con todas sus facetas).

Esto con base en los parámetros obtenidos de ellas para confrontarlos con los que Mizzaro propone y así determinar los juicios de relevancia, del cual se obtendrán los indicadores que se proponen como base para el diseño de una interfase de recuperación de información.

Mizzaro⁷⁰ propone que la relevancia se debe considerar a partir de cuatro dimensiones que comprenden:

- a) La problemática del usuario
- b) Las fuentes de información
- c) El tiempo, tarea, tema y contexto
- d) El juicio de relevancia

3.6. Parámetros de búsqueda por área de conocimiento

Conocer las necesidades y comportamiento de las disciplinas académicas de las que provienen los usuarios puede ayudar a diseñar y desarrollar sistemas y servicios de recuperación de información, con el fin de ofrecer información y servicios confiables.

Para tener servicios confiables de recuperación necesitamos conocer bajo que parámetros están trabajando las diferentes comunidades epistémicas, donde consideramos el término parámetro como *“aquellos límites establecidos cuyo valor afecta la ejecución o resultado de un proceso u operación”*.⁷¹

Estos parámetros obedecen a que cada comunidad epistémica tiene sus particulares objetos y fenómenos de estudio y establece con base en ellos como realiza sus investigaciones, su búsqueda de información, o los documentos que se generan de todo ello.

⁷⁰ Mizzaro S. (1998). How many relevances in information retrieval?. *Interacting with Computers*, 10, 303-320.

⁷¹ Online Dictionary for Library and Information Science. http://lu.com/odlis/odlis_p.cfm [consultado: 10-11-07]

La realización de una búsqueda de información dentro de una comunidad tiene como resultado el resolver una necesidad, explica Krikelas *“como cualquier actividad dirigida a identificar un mensaje que satisfaga una necesidad percibida como tal. En otras palabras, la búsqueda de información comienza cuando alguien se da cuenta de que el nivel de conocimientos que posee es menor del que necesita para afrontar una cuestión”*.⁷²

Este comienzo que menciona Krikelas, lo podemos observar en la actividad diaria donde nos encontramos en cualquier centro de información donde podemos observar que los usuarios, tanto a nivel individual como de una comunidad, acuden en búsqueda de información que satisfaga sus necesidades y es en este sentido que debemos ofrecer aquello que sea inicialmente relevante a dichas necesidades.

Para responder a la necesidad de información, el sistema de información debe considerar si existen rasgos que diferencien o agrupen a quienes buscan documentos relevantes, pues si lo considera se podrá responder a la pregunta: ¿cuál puede ser el comportamiento de búsqueda y anticiparse para dar información relevante para la satisfacción de la necesidad?

Al crear sistemas basados en la conducta de búsqueda de información se debe pensar en la relación entre diferentes características tanto particulares como generales, de tal forma que se obtenga un patrón particular de búsqueda de información.

En el caso de las comunidades epistémicas debemos pensar que cada una de ellas tiene sus necesidades específicas de información, las cuales se generan a partir de cada disciplina y actividad.

⁷² Krikelas, J. (1983). Information Seeking Behavior: patterns and concepts. En: *Drexel library quarterly* 19, pp. 6-7

Un punto a considerar, respecto de lo anterior, consiste en que al momento de realizar la búsqueda de información, quien la realiza de acuerdo a sus características puede combinar los elementos o en su caso no llevarla a cabo de la misma manera que otros individuos.

Sin embargo en algunos casos quienes pertenecen a comunidades con intereses particulares podríamos considerar que establecen procedimientos de búsqueda de información similares.

En este sentido encontramos a Wenger⁷³ quien llama a estos grupos, comunidades de práctica, y que junto con McDermont y Snyder las definen como *“Un grupo de personas que comparten una preocupación, un conjunto de problemas o un interés común acerca de un tema, y que profundizan su conocimiento y pericia en esta área a través de una interacción continuada”*.⁷⁴

Wenger afirma que para que las comunidades de práctica existan deben de estar presentes tres elementos:

- 1) *“Que los miembros entiendan que la comunidad es para, esto es, que los miembros deben sentirse unidos a la empresa y responsables.*
- 2) *Mutuo compromiso cuando los miembros tienen tiempo para construir confianzas y relaciones sobre cualquier otra interacción regular.*
- 3) *Los miembros deben de desarrollar un repertorio compartido de historias, lenguajes, etc. Que envuelvan un conocimiento distintivo de la comunidad y permita a los miembros negociar su significado”*.⁷⁵

A partir del último punto que menciona Wenger podemos decir que las comunidades de forma general deben presentar características similares al momento de realizar búsquedas de información, pues comparten conocimientos,

⁷³ Wenger, E. (2000). Communities of practice: the key to knowledge strategy. Knowledge and communities. Boston, MA : Butterworth Heinemann

⁷⁴ Wenger, E.; McDermott, R.; Snyder, W.M. (2002). Cultivating communities of practice. Boston: Harvard Business School Press.

⁷⁵ Wenger, E. (2000). Op cit.

que los lleva a información relevante, así como el compartir estrategias de búsqueda a fin de mantenerse en una misma línea de información.

Podemos asumir que las personas pueden seguir ciertos pasos a fin de encontrar información que satisfaga sus necesidades de información para poder resolver los problemas que se les presentan.

Además podemos estimar el agrupar a estos individuos de acuerdo a ciertas características que presenten. Agruparlos de acuerdo a su profesión o actividad, lo que permitirá saber cuáles son los elementos que toma en cuenta la comunidad para poder establecer los juicios de valor, para posteriormente determinar la relevancia de la información que obtienen de un sistema de recuperación de información.

De manera general podemos señalar que para cada área del conocimiento, algunos autores manifiestan elementos claves que se podrían tomar en cuenta cuando se pretende construir interfaces de búsqueda.

En el Área de Ciencias Agropecuarias. Bigdeli⁷⁶ realizó un estudio de las necesidades y de los hábitos de búsqueda de información y encontró los elementos siguientes:

- a. Experiencia personal ganada con la participación en conferencias locales e internacionales
- b. Uso de bibliotecas personales
- c. Consulta con otros expertos del área
- d. Uso de reportes, proyectos, tesis, así como de publicaciones periódicas seguido del uso de libros
- e. Uso de centros de información y otras bibliotecas

⁷⁶ Bigdeli, Z. (2007). Iranian engineers' information needs and seeking habits: an agro-industry company experience. En: *Information Research*, 12 (2)

- f. Facilidad de obtener información de la red y de bases de datos

De lo encontrado podemos agregar a Álvarez⁷⁷ quien en el documento que realizó aunque es de fecha anterior, considera un elemento que no menciona Bigdeli; comúnmente se desarrollan y actualizan sus sistemas de información aun sin darse cuenta de forma espontánea y/o inconsciente, sin tener un procedimiento formal a seguir con esta finalidad.

Para el Área de Ciencias de la salud. Encontramos a Leckie⁷⁸ que señala como principales las siguientes:

- a. Uso principal de revistas y libros de texto
- b. Preferencia por información actualizada ofrecida por revistas.
- c. Consulta con colegas
- d. Bibliotecas médicas locales
- e. Uso o impresión de recursos electrónicos
- f. Predomina la comunicación oral.
- g. Preferencia por archivos, conocimientos y experiencia personales.
- h. Preferencia por buscar de manera directa, sin ayuda del bibliotecario.
- i. Seguimiento de las referencias que se encuentran en los artículos de revistas.

En el Área de Ciencias Naturales y Exactas. Los elementos que distinguen a estas áreas los encontramos en los trabajos de Hernández⁷⁹ y Villanueva⁸⁰ donde hallamos coincidencias al señalar los siguientes elementos:

- a. Preferencia por información actualizada ofrecida por revistas.

⁷⁷ Álvarez J. (2000) La gestión del negocio agropecuario y el manejo de la información. Universidad de la República

⁷⁸ Leckie, G., Pettigrew, E., Sylvain, C. (1996) op cit.

⁷⁹ Hernández, P. (2001). La producción del conocimiento científico como base para determinar perfiles de usuario. En: Investigación bibliotecológica, 15 (30)

⁸⁰ Villanueva, A. (2007). La comunidad de matemáticos en México. Anales de documentación (10)

- b. Predomina la comunicación oral.
- c. Preferencia por archivos, conocimientos y experiencia personales.
- d. Intensidad en el uso de recursos electrónicos de recuperación de información (bases de datos).
- e. Seguimiento de las referencias que se encuentran en los artículos de revistas.
- f. Cuando utilizan fuentes escritas estas suelen ser manuales, reportes técnicos, papeles de conferencias, entre otros.
- g. Menor uso de los catálogos de bibliotecas.
- h. Preferencia por buscar de manera directa, sin ayuda del bibliotecario.

Del Área de Ciencias Sociales y Administrativas. En estas áreas consideramos de igual manera lo encontrado por Hernández:⁸¹

- a. Preferencias por información de monografías antes que de revistas pero mayor aceptación de revistas que en el caso de las humanidades.
- b. Falta de familiaridad con el uso de sistemas electrónicos de recuperación de información, que se convierte en un impedimento para la búsqueda de información.
- c. Frecuentan la biblioteca para buscar información y consideran que brindan aportes a su investigación.
- d. Consultan a los bibliotecarios.

En el Área de Educación y Humanidades. Owen⁸² y Albornoz⁸³ señalan:

- b. Satisfacción de necesidades de información principalmente a través de monografías.

⁸¹ Hernández, P. (2001). Op cit.

⁸² Owen, D. (1991) Conceptual organization and retrieval of text by historians: the role of memory and metaphor. En: Journal of the American Society for Information Science, 42 (9)

⁸³ Albornoz, S., et al. (2007) Búsqueda de la información y uso de la biblioteca por parte de los investigadores de humanidades. En: revista interamericana de bibliotecología, 30 (1)

- c. Familiaridad con el uso de sistemas electrónicos de recuperación de información.
- d. Notas personales y manuscritos.
- e. Otros artefactos de información como cintas magnéticas y materiales audiovisuales.
- f. Uso del catálogo de biblioteca.
- g. Recurren a pedir ayuda al bibliotecario.
- h. Consultan a los colegas.

Para el Área de Ingeniería y Tecnología. Para esta área tomamos también lo que menciona Leckie y colaboradores⁸⁴ que consideran:

- a. Preferencia por información actualizada ofrecida por revistas.
- b. Predomina la comunicación oral.
- c. Preferencia por archivos, conocimientos y experiencia personales.
- d. Intensidad en el uso de recursos electrónicos de recuperación de información (bases de datos).
- e. Seguimiento de las referencias que se encuentran en los artículos de revistas.
- f. Cuando utilizan fuentes escritas estas suelen ser manuales, reportes técnicos, papeles de conferencias, entre otros.
- g. Menor uso de los catálogos de bibliotecas.
- h. Preferencia por buscar de manera directa, sin ayuda del bibliotecario.

De manera general podemos resumir la información anterior en el cuadro 4.

⁸⁴ Leckie, G., Pettigrew, E., Sylvain, C. (1996) Op cit.

Cuadro 4. Concentrado de coincidencias en las preferencias de las distintas áreas.

Á r e a	C i e n c i a s	C i e n c i a s	C i e n c i a s	C i e n c i a s	C i e n c i a s	C i e n c i a s	C i e n c i a s	C i e n c i a s	C i e n c i a s
Elementos para recuperación de información	Ciencias Agropecuarias	Ciencias de la Salud	Ciencias Naturales y Exactas	Ciencias Sociales y Administrativas	Educación y Humanidades	Ingeniería y Tecnología			
Quando utilizan fuentes escritas estas suelen ser de tipo manuales, reportes técnicos, papeles de conferencias, entre otros	✓	✓	✓			✓			
Intensidad en el uso de recursos electrónicos de recuperación de información (bases de datos)	✓	✓	✓			✓			
Uso de centros de información y otras bibliotecas	✓	✓		✓	✓				
Uso de los catálogos de bibliotecas				✓	✓				
Predomina la comunicación oral	✓	✓	✓			✓			

Preferencia por archivos, conocimientos y experiencia personales	✓	✓	✓	✓	✓	✓	✓
Preferencia por buscar de manera directa, sin ayuda del bibliotecario	✓	✓	✓	✓	✓	✓	✓
Preferencia por información actualizada ofrecida por revistas	✓	✓	✓	✓	✓	✓	✓
Preferencias por información de monografías					✓	✓	
Seguimiento de las referencias que se encuentran en los artículos de revistas	✓	✓	✓	✓	✓	✓	✓
Consulta con colegas del área	✓	✓	✓	✓	✓	✓	✓

Elaborado por el autor

De los elementos agrupados en el cuadro anterior podemos decir que:

Las áreas de Ciencias Sociales y Administración, Educación y Humanidades son las únicas que no utilizan fuentes escritas de tipo manuales, reportes técnicos y de conferencias etc., esto debido a la naturaleza de las mismas áreas.

De los recursos electrónicos de recuperación de información vemos que la mayoría de las áreas los utiliza, mientras que son poco utilizados en el área de Ciencias Sociales y Administrativas es la única que no hace uso de este recurso.

Respecto al uso de centros de información y las bibliotecas encontramos que son utilizados por personas de las áreas siguientes: Ciencias Agropecuarias, Ciencias de la Salud, Ciencias Sociales y Administrativas, y Educación y Humanidades.

En cuanto al uso del catálogo las áreas de las Ciencias Sociales y Administrativas, Educación y Humanidades utilizan los catálogos de las bibliotecas. En estas dos áreas no predomina la comunicación oral; tienen preferencia por buscar de manera indirecta con ayuda del bibliotecario.

Del uso de archivos, conocimientos y experiencia personales, solo el área de Ciencias Sociales y Administrativas no los utiliza.

Por lo que se refieren a la utilización de información de revistas o de monografías, observamos que ciencias sociales y administrativas, educación y humanidades tienen preferencia por la información de materiales monográficos, sin embargo, en las demás áreas predomina el uso de revistas.

De igual manera sucede con el seguimiento a las referencias pues las dos áreas al no usar revistas no realizan el seguimiento de las referencias.

Finalmente hallamos que en la interacción con colegas del área se dividen las preferencias dando como resultado que la mitad de las área lo realiza y la otra no.

3.7. Parámetros de búsqueda en las áreas de Ingeniería y Tecnología y Ciencias de la Salud

Los especialistas de la información durante mucho tiempo han estado interesados en saber cómo los individuos/usuarios encuentran y usan información durante su trabajo diario. Entender cómo la información y la búsqueda de información envuelve el trabajo de los profesionales en cuestión.

Podemos encontrar múltiples trabajos que investigan diferentes disciplinas, pero para la finalidad de este trabajo nos ocuparemos dentro del marco de las áreas del conocimiento de Ciencias de la Salud y de Ingeniería y Tecnología.

A partir de ese conocimiento se realizará la confrontación de lo expresado por Mizzaro para proponer cuáles elementos deben considerarse para poder establecer las características que debe reconocer para ofrecer una interfase de recuperación basada en la relevancia.

3.7.1. Perfil del Área de Ingeniería y Tecnología

El área de Ingeniería y Tecnología es el primer grupo de profesionales considerados dentro de la presente investigación.

A quienes participan en Ingeniería y Tecnología los podemos considerar especialistas entrenados a nivel universitario en numerosas subespecialidades: ingenieros civiles, electrónicos, mecánicos, petroleros, entre otros.

El campo de acción de la Ingeniería y Tecnología abarca un rango amplio de medios ambientes de trabajo, desde pequeñas firmas de consultoría hasta grandes consorcios.

Desde este punto de vista podemos asumir lo que menciona Kemper⁸⁵ que en las diferentes etapas de su carrera pueden ocupar diferentes funciones incluyendo investigación y desarrollo, diseño, manufactura y construcción, ventas, consultoría, enseñanza, entre otras.

En lo que se refiere al uso y búsqueda de información, los ingenieros se ubican en la búsqueda de solución de problemas técnicos que resultan en productos, procesos y/o servicios y que son plasmados como nuevos conocimientos en forma de artículos o reportes.

En este sentido Shuchman⁸⁶ menciona que los ingenieros necesitan más información de la que generan, es por ello que la importancia de obtener información relevante para satisfacer su necesidad de información adquiere mayor importancia.

Una de las características sobresalientes que predomina entre los ingenieros es la utilización de la comunicación oral, comparten el conocimiento con sus compañeros de trabajo o supervisores.

Las bibliotecas por su parte son poco usadas en oposición de los archivos personales, el conocimiento y la experiencia personal. Cuando consultan fuentes escritas estas consisten en libros de texto, técnicos, catálogos y revistas.

⁸⁵ Kemper, J. (1990). Engineers and their profesión. Philadelphia : Saunders, p. 35

⁸⁶ Shuchman, H. (1981). Information transfer in engineering. Glastonbury, Conn. : Fututres group, p. 25

King y Griffiths mencionan que los ingenieros como otros profesionales emplean alrededor del 20 al 80 por ciento de sus horas de trabajo en la búsqueda y manipulación de información.

Las posiciones que pueden ocupar los ingenieros pueden llevarlos a diferentes necesidades de información. Por ejemplo, aquellos que se dedican a la investigación y desarrollo para los cuales el objetivo es un producto o aparato para ser entregado a un cliente.

Cuando se trata de encargos que incorporan pequeños grupos de trabajo los cuales ejecutan tareas asociadas con un rol o función particular. Dependiendo del rol va a ser el conjunto de tareas que se va a ejecutar y los requerimientos de información necesaria para llevar a cabo la tarea.

3.7.2. Perfil del Área de Ciencias de la Salud

Este grupo incluye a todos aquellos que podemos considerar como profesionales de la salud tanto médicos como dentistas y enfermeras.

Una de las características que presenta esta área es la necesidad de información que se relaciona con la enfermedad y el cuidado del paciente.

Stinson y Mueller⁸⁷ señalan que los recursos de información utilizados por esta área van relacionados entre otros factores con el tipo de práctica que se realiza, la especialidad, el lugar y tamaño de donde se lleva la actividad y la edad profesional del individuo. Además de que en el caso de aquellos profesionales que se encuentran en zonas urbanas recurren a colegas para obtener información que apoye a la solución de sus necesidades de información.

⁸⁷ Stinson, E. y Mueller, D. (1980). Survey of health professional's information habits and needs. En: Journal of the American Medical Association, 243 (January)

Por su parte Gruppen y otros⁸⁸ observaron que esta área recurre a revistas y libros de texto más comúnmente y relegan la consulta a otros profesionales en casos de consultas informales.

Por su parte Borden y otros⁸⁹ encontraron que no hay una diferencia significativa respecto del uso de fuentes de información primaria que usan regularmente dentro de lo que consideran colecciones de libros y revistas personales o de su oficina y la consulta a los colegas. De igual manera mencionan que las bibliotecas que se encuentran cerca de aquellos profesionales reportan un gran uso de sus servicios.

⁸⁸ Gruppen, L.; et al. (1987). Information-Seeking strategies and differences among primary care physicians. *Mobius* (7)

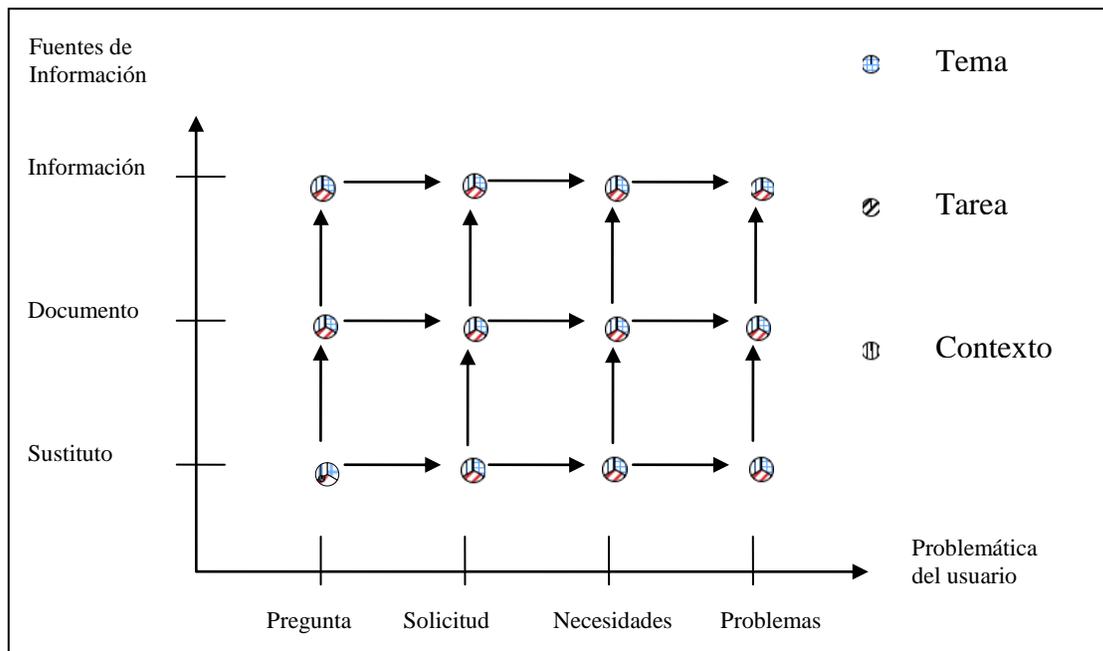
⁸⁹ Borden, V.; et al. (1994). Assessment of physicians information needs in five Texas counties. En: *Bulletin of the Medical Libray Association* (82, april) p. 190

3.8. Parámetros de búsqueda de información vs. elementos de juicio de relevancia: propuesta

A partir de los elementos que se determinaron son utilizados por las áreas de Ciencias de la Salud e Ingeniería y Tecnología, y considerando las dimensiones de la relevancia en la recuperación de información propuestas por Mizzaro de tal manera que se identifiquen cuales son los puntos que deben considerarse para determinar los juicios de relevancia que utilizan dichos profesionales, a fin de proponer en una interfase de búsqueda.

Iniciando con las dimensiones que propone el modelo de relevancia de Mizzaro (cuadro 5) tenemos:

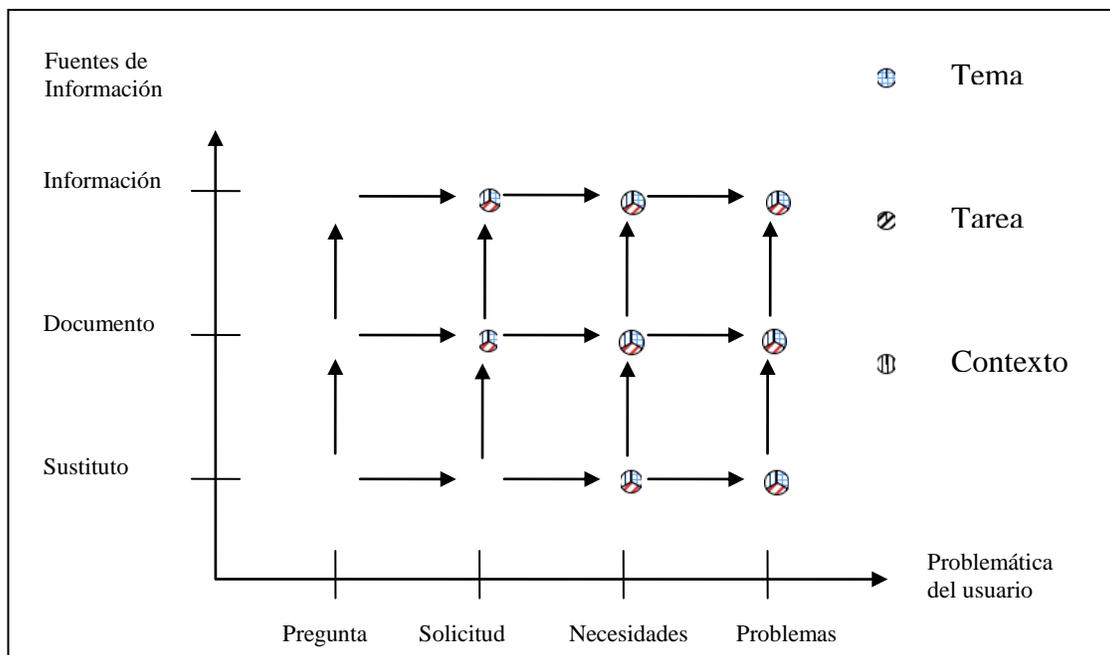
Cuadro 5. Modelo de relevancia



3.8.1. Área de Ingeniería y Tecnología

En este grupo al relacionarlo con las características que presenta el modelo de relevancia encontramos que la ingeniería y la Tecnología (cuadro 6) prácticamente ocupa todos los elementos que menciona el modelo con excepción de cuatro, no utilizan centros de información, por lo tanto no hacen uso de los catálogos de las mismas, no utilizan información de monografías y no consultan con colegas del área.

Cuadro 6. Modelo de relevancia aplicado a Ingeniería y tecnología.



Elaborado por el autor

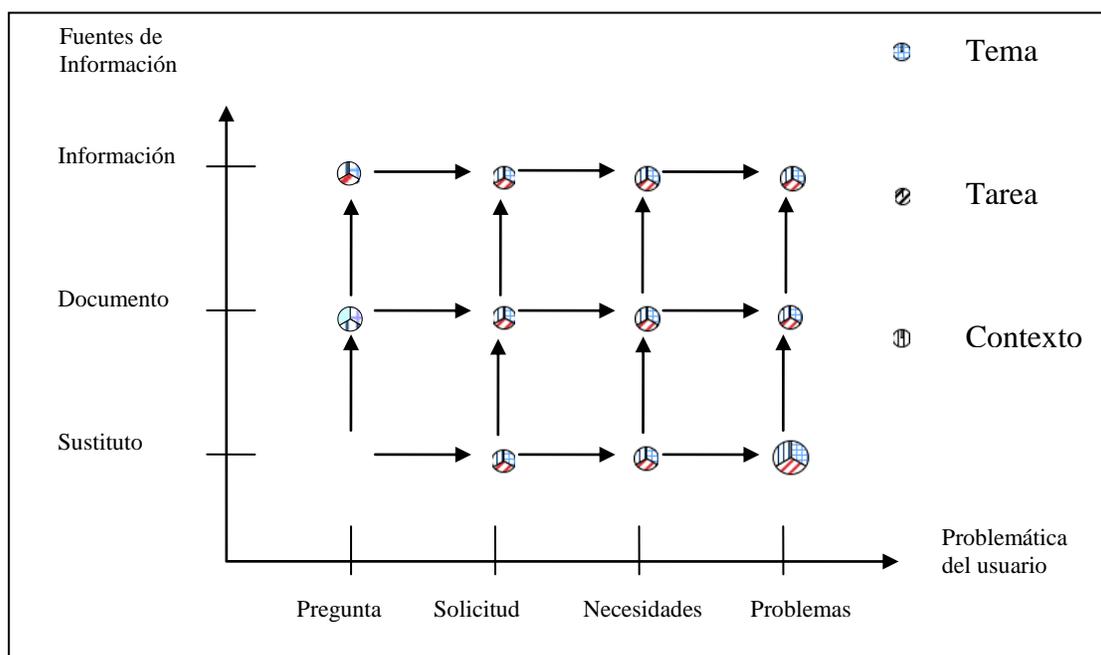
Como se mencionó la búsqueda de información por parte de los ingenieros abarcó prácticamente todo el modelo y aquellos puntos donde no fue aplicado aun podemos considerar que están influidos por parte de otras dimensiones, por ejemplo cuando se menciona que no utilizan centros de información y sus catálogos, tendríamos que considerar el contexto bajo el que se desarrolla la tarea de búsqueda de información o el que no consulten monografías, estos elementos nos tendrían que hacer pensar que entonces existe una falta de difusión hacia lo que se puede encontrar en un sistema de información o que en el tiempo en el que

se tuvo la necesidad de información no existieron las condiciones ideales para ello.

3.8.2. Área de Ciencias de la Salud

En este grupo al relacionarlo con las características que presentan el grupo de las ciencias de la salud tenemos que ocupa todos los elementos que menciona el modelo con excepción de dos, el de uso de los catálogos de bibliotecas y el de preferir información de monografías.

Cuadro 7. Modelo de relevancia aplicado a ciencias de la salud



Elaborado por el autor

En el caso de las ciencias de la salud (cuadro 7) al igual que en la ingeniería tendremos que considerar que aquellos puntos que no fueron considerados se ven influenciados por otros elementos del modelo, el punto de coincidencia entre los dos de no usar la biblioteca tendría que llevarnos a la reflexión acerca de mejorar los sistemas de recuperación de información a fin de que resulten efectivos para estos grupos epistémicos.

La capacidad para definir y estructurar una necesidad de información, identificando conceptos claves y términos que describan el perfil de búsqueda de

un grupo en particular tendría que ir más allá de su simple reflejo en un esquema determinado ya que como mencionamos existen diversos factores involucrados en la búsqueda de información en un sistema de recuperación de información.

Así, tendríamos que considerar que aunque cada grupo puede tener sus propios objetos de estudio y puede llevar a cabo diferentes estrategias a fin de hacerse de información para responder a sus necesidades, tendríamos que considerar todos los puntos del modelo al momento de considerar el desarrollo de una interfaz de recuperación.

3.9. Indicadores para el diseño de una interfaz de búsqueda de información

En la actualidad podemos encontrar interfaces que muestran diferentes niveles de desarrollo y complejidad en cuanto a su funcionalidad, terminología utilizada o propiedades de ayuda para los usuarios. Por regla general podemos encontrar que no existen bibliotecas que no ofrezcan a sus usuarios catálogos de acceso a sus colecciones en forma automatizada.

Sin embargo, una situación que encontramos a lo largo de la historia del desarrollo de estos catálogos automatizados es la poca intervención de diseños enfocados a los usuarios, más bien encontramos que el diseño está enfocado hacia los sistemas, que son resultado del trabajo de diseñadores de sistemas y programadores.

En este sentido debemos considerar la Conferencia Internacional sobre Principios de Catalogación⁹⁰ de donde surgen los conocidos “Principios de París” de 1961, que señalan que un catálogo debe de cumplir con:

⁹⁰ International Conference on Cataloguing Principles (Paris : 1961). *Report.* – London : International Federation of Library Associations

- ❖ Encontrar todo aquello que corresponda a la búsqueda del usuario.
- ❖ Identificar un documento.
- ❖ Seleccionar lo que corresponda a las necesidades del usuario.
- ❖ Obtener o adquirir los documentos que recuperamos.

Cuando un usuario se enfrenta a realizar la búsqueda de información en un sistema de recuperación, espera encontrar una cierta información que puede estar representada en dos o tres documentos o en el peor de los casos en cientos de registros, como resultado de sistemas que solo consideran una recuperación por palabras por ejemplo.

En el caso de encontrarse en una situación como la anterior el usuario debe recurrir a la delimitación mediante la utilización de términos para refinar los resultados o la combinación de términos mediante operadores booleanos, aunque debemos reconocer que sólo aquellos usuarios que conocen el sistema son los que llegan a utilizar estas técnicas de búsqueda.

Si consideramos, sin embargo, la utilización de criterios de relevancia para la ordenación de los resultados, entonces podríamos estar en condiciones de que sin importar la cantidad de registros recuperados, los usuarios sabrían de antemano que los primeros que revisaría serían los que por relevancia son los más importantes o relevantes, y los restantes serían los menos relevantes y podría decidir si los revisa o no.

Propuesta para la solución al problema

Con el fin de dar respuesta a la problemática de la aplicación de los criterios de relevancia que permitan a los usuarios la recuperación de información de acuerdo a lo que ellos consideren relevante, se plantea el diseño de una interfase para la recuperación de la información que considere dichos criterios.

Entendemos como interfaz la presentación en pantalla, que un sistema informático ofrece a un usuario para que pueda interactuar con él.

El diseño elegido para un sistema debe ser el apropiado para la comunidad de usuarios a la que va dirigido y las tareas que van a realizar. Un mismo diseño puede ser apropiado para unos usuarios y desaconsejable para otros, eficiente para unas tareas e ineficiente para otras.

La aplicación de la interfaz en la recuperación de información, debe tener en cuenta tanto los componentes físicos, como son los dispositivos de entrada y salida y la presentación de la información en pantalla, como aquellos más conceptuales tales como los mecanismos de selección y de retroalimentación y las reglas para la recuperación de la información.

De igual manera la interfaz debe considerar en su diseño la usabilidad, es decir, del grado de satisfacción de usuario en su manejo traducida en facilidad de uso y rapidez en obtener los resultados deseados.

Dentro de las características que se ofrecerían al usuario estarían las siguientes:

Ofrecer una terminología común a los diseñadores y a los usuarios, lo que evita problemas de entendimiento.

Proporcionar una mayor facilidad para el mantenimiento ya que todos los programas comparten una estructura, con el consecuente ahorro en tiempos y trabajo.

Dar a los sistemas una identidad común y hacer que sus elementos sean más fáciles de reconocer.

Reducir la necesidad de formación para manejar el programa porque lo que se conoce sobre uno se puede aplicar a otros.

Proporcionar seguridad al usuario, que puede predecir la actuación del sistema al interactuar con él y evitar sorpresas.

Lo importante es que los usuarios estén seguros, es decir, que controlen el sistema, que puedan aprender fácilmente a manejarlo y que les invite a explorar sus opciones y capacidades. Para que se den estos resultados es necesario que los principios de diseño atiendan a los componentes clave del "iceberg" (Mandel, 1997) un 10% es lo que se ve (presentación), un 30% lo que el usuario siente (interacción) y un 60% cómo interactúa el usuario a través del diseño, de la metáfora empleada, y como se acomoda el modelo del sistema al modelo mental del usuario (relación persona-ordenador).

Quiere decir que la parte que más ve el usuario, la de la apariencia, no debe recibir la mayor dedicación a la hora de diseñar el sistema, en cambio las relaciones que éste no percibe directamente se llevarían un 60% de la atención por parte del diseñador.

Se propone un modelo en el que se trate de resolver la necesidad real del usuario a lo largo del tiempo ofreciéndole información que tenga en cuenta tanto el tema sobre el que se busca como las tareas para las que se va a emplear la información y otros datos de contexto. Que se presente la información clasificada en función

de su relevancia de manera que el usuario pueda arrastrar los documentos en razón de los factores de relevancia que más le interesen en cada momento.

Es fundamental que el usuario pueda reformular la consulta una vez obtenidos los resultados y así afinar el planteamiento para obtener documentos que se ajusten mejor a sus necesidades.

Estas técnicas son las de retroalimentación por relevancia, el ciclo de interacción en el que el usuario selecciona los documentos que considera relevantes, el sistema utiliza las características de los documentos seleccionados para revisar la consulta y vuelve a ejecutarla. Esta retroalimentación introduce posibilidades en el diseño, entre ellas las que permiten al usuario iniciar y controlar el proceso.

La forma de que el usuario indique en la interfaz de resultados su valoración de la relevancia con la intención de mejorar su consulta, consistirá en marcar en la lista de títulos los que considera relevantes (incluso indicar el grado en que los considera relevantes) y también puede tener la posibilidad de marcar como no relevantes aquellos que no se lo parezcan.

Estos datos son muy útiles para los algoritmos que aprenden a obtener documentos relevantes. Una vez que el sistema recoge las marcas hechas por el usuario, reformula la consulta modificando los pesos de los términos utilizados y puede hacer dos cosas: formular directamente la consulta u ofrecer una lista de términos al usuario para que marque los que le interesa buscar. Por último, da los nuevos resultados.

La ordenación por grado de relevancia será tan eficiente como lo sea el algoritmo que se encuentre tras ella, si este sólo tiene en cuenta el número de ocurrencias de cada documento, dará resultados muy pobres e incluso equivocados.

La ordenación de relevancia no sustituye, la utilización de operadores booleanos, sino que es un medio para entregar los resultados ordenados por un criterio mucho más riguroso que el simple azar.

Conclusiones

La relevancia como elemento de valor para el diseño de un sistema de recuperación de información está cobrando mayor importancia hoy en día, ya que podemos encontrar sistemas que incluyen la opinión de los usuarios como parte del diseño y con esto conocer cuáles son los puntos importantes para ellos.

Al identificar estos criterios e incluirlos dentro de los sistemas de recuperación, los usuarios serán los más beneficiados pues podrán lograr resultados exitosos para atender a las necesidades de información que tengan que resolver.

Si consideramos un sistema de recuperación de información y la respuesta que ofrece a un usuario, dicha respuesta estaría relacionada con la evaluación de la relevancia que ese usuario le da al resultado obtenido aunque podríamos pensar que esa valoración sería aplicable a ese usuario en particular, sin embargo las características que envuelven la necesidad de información de una comunidad y sus integrantes, nos permite creer que la respuesta de la determinación de la relevancia es semejante para los integrantes de dicha comunidad.

En la actualidad uno de los problemas al que nos enfrentamos es que si bien se está considerando más el enfoque hacia el usuario, los sistemas de recuperación de información, en el caso de libros por ejemplo, no pueden ofrecer más recursos que los de la descripción bibliográfica de una obra conocida o una materia general para poder integrar cuestiones de relevancia.

Una de las soluciones que encontramos actualmente es el uso de la retroalimentación por relevancia, la cual utiliza la información que se va generando en los procesos de búsqueda y recuperación de información realizados y que son juzgados por parte de los usuarios los que califican cuales son relevantes y con esta evaluación se establecen nuevos valores a los documentos. Sin embargo aunque se considera al usuario para asignar una medida de relevancia a los

documentos, siguen siendo los sistemas los que determinan la relevancia de los documentos que son recuperados, de tal forma que se mantiene el problema de hacer coincidir por una parte la opinión del usuario y por otra la que resulta del valor que es asignado por parte del sistema.

En cuanto podamos identificar las necesidades de información y las podamos combinar con la forma en la que la información es almacenada empezaremos a dar respuesta a la búsqueda de información. Aquí es donde el nuevo modelo conceptual Requisitos Funcionales de los Registros Bibliográficos (FRBR por sus siglas en inglés)⁹¹ está comenzando a cambiar la forma en que los catálogos de las bibliotecas ofrecen información y cuyo objetivo es la definición de las funciones que deben cumplir los registros bibliográficos respecto a los diferentes medios, las diferentes aplicaciones y necesidades de los usuarios lo que llevará a los catálogos a ofrecer información relevante y pertinente cuando se les solicite y de igual manera influirán en el diseño de interfaces amigables para el usuario.

En la actualidad seguimos padeciendo la total falta de inteligencia de los sistemas actuales, si bien se trata de darles algo de la experiencia y del conocimiento de los profesionales de la bibliotecología, nada sustituye a un operador humano, que sí puede interpretar y reconocer temas dentro de la estructura de un documento.

Sin duda será cuestión de tiempo para que llegemos a la integración de elementos tecnológicos y humanos que permitan, por un lado, conocer el perfil de cada usuario de tal forma que el resultado de la búsqueda de información ofrezca resultados relevantes, y por el otro, que la información este organizada y relacionada de tal manera que siempre se mantenga actualizada y la interacción con los sistemas de recuperación de información se dé con la seguridad de obtener resultados que sean juzgados de forma favorable.

⁹¹ IFLA. Study Group, *Functional Requirement for Bibliographic records: final report*, Munchen: K.G. Saur, 1998. 136 p. Consultado en enero, 2012 del sitio Web http://archive.ifla.org/VII/s13/frbr/frbr_current_toc.htm

Coincidimos con Mizzaro que señala que se pueden mejorar los sistemas de recuperación de información y como interactúan con los usuarios siempre y cuando se considere que la relevancia no es solamente una, sino que se pueden encontrar diferentes tipos de ella en el juicio que lleva a cabo el usuario, de igual forma consideramos que el modelo que presenta de las diferentes dimensiones de la relevancia, representación de la información, representación del problema del usuario, tiempo y los componentes de tema, tarea y contexto, es adecuado para la determinación de puntos a considerar en el diseño de interfaces de sistemas de recuperación de información eficientes. Es necesario entender cómo los usuarios utilizan las estrategias y términos de búsqueda y cómo son juzgados los resultados relevantes para el mismo problema de búsqueda. Una cuestión importante en esta situación es la construcción de perfiles de usuarios que mantengan asociados la información previamente recuperada junto con las necesidades de los mismos, lo que permitirá mejorar los algoritmos de búsqueda que combinados den, como señalamos, información relevante a quien requiere responder a una necesidad de información.

La relevancia ha sido uno de los temas más importantes en el desarrollo de sistemas de recuperación de información, pero también ha generado discusión en cuanto a llegar a un acuerdo sobre su naturaleza y determinación y como implementarla en un SRI. Sin embargo la evaluación de SRI en los diferentes momentos de tiempo y como es medida su eficacia se sigue basando en el juicio que alguien realiza sobre la relevancia de la información recuperada, esto nos lleva a considerar que hasta la fecha no se ha encontrado un sustituto de la relevancia en lo que se refiere al criterio para medir esa eficacia y para evaluar a los SRI.

Bibliografía

Albornoz, S., et al. (2007). Búsqueda de la información y uso de la biblioteca por parte de los investigadores de humanidades. *revista interamericana de bibliotecología*, 30 (1).

Álvarez J. (2000). La gestión del negocio agropecuario y el manejo de la información. Universidad de la República.

Baeza-Yates, R. (1999). *Modern information retrieval*. Reading, Mass. : Addison-Wesley Longman.

Barry, C.L. (1994). User -defined Relevance Criteria: An Exploratory Study. *Journal of the American Society for Information Science*, 45, (3).

Bigdeli, Z. (2007). Iranian engineers' information needs and seeking habits: an agro-industry company experience. *Information Research*, 12 (2).

Borden, V.; et al. (1994). Assessment of physicians information needs in five Texas counties. *Bulletin of the Medical Libray Association*, 82, (april).

Borlund, P. (2003). The Concept of Relevance in IR. En *Journal of the American Society for Information Science and Technology*, 54, (10).

Boyce, B. (1982). Beyond Topically: A two storage view of relevance and retrieval process. *Information procesing and Management*, (18).

Chowdhury, G. (2004). *Introduction to modern information retrieval*. London : Facet.

Cleverdon, C. (1997). *The Cranfield tests on index languages devices Readings in information retrieval*. San Francisco: Morgan Kaufmann.

Cooper, W. (1971). A definition of relevance for information retrieval. *Information storage and retrieval*, 7 (1).

Cosijn, Erica; Ingwersen, Peter. (2000) Dimensions of relevance. *Information processing and management*. (36).

Cuadra, C. A., & Katter, R. V. (1967). *Experimental studies of relevance judgements* (NSF Rep. TM-3520/001, 002, 003, 3 vols.). Santa Monica, CA: Systems Development Corporation.

Diccionario enciclopédico de ciencias de la documentación. (2004). Madrid : Síntesis.

Dominich, S. (2000). A unified mathematical definition of classical information retrieval. *Journal of the American Society for Information Science*, 51, (7).

Ellis, D. Elli's. (2005). Model of information-seeking behavior. Theories of information behavior. Medford, New Jersey : American Society for Information Science and Technology by Information.

Ellis, David. (1992) The Physical and Cognitive Paradigms in Information Retrieval Research. *Journal of Documentation*, 48, (1).

Foskett, D. J. (1972). A note on the concept of 'relevance'. *Information Storage Retrieval*, 8, (2).

Froehlich, T. (1994). Relevance reconsidered : towards an agenda. *Journal of the American Society for Information Science*, 45, (3).

García, L. (2000). Diccionario del archivero bibliotecario : terminología de la elaboración, tratamiento y utilización de los materiales propios de los centros de documentación. Gijón, Asturias : Trea.

Gruppen, L.; et al. (1987). Information-Seeking strategies and differences among primary care physicians. *Mobius*, 7, (3).

Harter, S. P. (1971). The *Cranfield II* relevance assessments: a critical evaluation. *Library quarterly*, 41, (3).

Hernández, P. (2001). La producción del conocimiento científico como base para determinar perfiles de usuario. *Investigación bibliotecológica*, 15 (30).

Hert, C. A. (1997). Understanding information retrieval interactions: theoretical and practical implications. Greenwich, Connecticut : Ablex.

Hersh, W. (1994). Relevance and retrieval evaluation: Perspectives from medicine. *Journal of the American Society for Information Science*, 45, 201–206.

IFLA. (1998). Study Group, *Functional Requirement for Bibliographic records: final report*, Munchen: K.G. Saur. En línea: http://archive.ifla.org/VII/s13/frbr/frbr_current_toc.htm

Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of documentation*, 52, (1).

Ingwersen, P. (1992). Information retrieval interaction. London : Taylor Graham.

International Conference on Cataloguing Principles (Paris : 1961). *Report*. – London : International.

- Kemper, J. (2000). *Engineers and their profesion*. Philadelphia : Saunders.
- Korfhage, R. (1997). *Information Storage and Retrieval*, New York: John Wiley.
- Kowalski, G. (1997). *Information retrieval systems theory and implementation*. Boston: Kluwer Academic Publishers.
- Krikelas, J. (1983). Information Seeking Behavior: patterns and concepts. *Drexel library quarterly*, 19, (2).
- Kuhlthau, C. (2002). Kuhlthau's information search process. *School Library Media Activities Monthly*, 19, (2).
- Lancaster, F. (1979). *Information retrieval systems : characteristics, testing, and evaluation*. New York : Wiley.
- Lancaster, F. W. and Warner, A.J. (1993). *Information Retrieval Today*. Arlington, Virginia: Information Resources.
- Leckie, G., Pettigrew, E., Sylvain, C. (1996). Modeling the information seeking of professionals. *Library quarterly*, 66 (2).
- Martínez, F. J. (2004). *Recuperación de información: modelos sistemas y evaluación*, Murcia: Kiosko.
- Martínez, M. (2004). *Recuperación de información : modelos, sistemas y evaluación*. Murcia: Kiosko.
- Martínez de Sousa, J. (2004). *Diccionario de bibliología y ciencias afines*. Gijón, Asturias : Trea.
- Mizzaro, Stefano. (1997). Relevance: the whole history. *Journal of the American Society for Information Science*, 48, (9).
- Mizzaro, Stefano. (1998). How many relevances in information retrieval?. *Interacting with Computers*, 10.
- O'Brien, A. (1990) . Relevance as an aid to evaluation in OPACs. *Journal of Information Science*, 16.
- Online Dictionary for Library and Information Science. En línea http://lu.com/odlis/odlis_r.cfm
- Owen, D. (1991). Conceptual organization and retrieval of text by historians: the role of memory and metaphor. *Journal of the American Society for Information Science*, 42 (9).

Ponjúan Dante, Gloria; et. al. (2004). *Sistemas de información: Principios y aplicaciones*. Ciudad de la Habana: Félix Varela.

Ramírez, A. (2007) La lógica difusa en los sistemas automatizados de recuperación de información. *XIV reunión de bibliotecarios de la Península de Yucatán*, Mérida, Yucatán.

Rees, A. M. (1966). The relevance of relevance to the testing and evaluation of document retrieval system. *Aslib proceedings*, 18, (11).

Reid, J. A. (1999). New task-oriented paradigm for information retrieval: implications for evaluation for information retrieval systems. *Digital libraries:interdisciplinary concepts, challenges and opportunities. Proceedings of the CoLIS3 Conference. Zagreb*.

Rijsbergen, C. J. van. (1979). *Information retrieval*. London : Butterworth. En línea: <http://www.dcs.gla.ac.uk/Keith/Chapter.1/Ch.1.html>

Salton, G. (1998). *Automatic text processing*. Adisson-Wesley.

Salton, G. ; McGill, M. (19839. *Introduction to modern information retrieval*. New York :McGraw Hill.

Saracevic T. (1999). *Information Science. JASIST*, 50, (12).

Saracevic, T. (1975). Relevance : a review of and framework for the thinking on the notion in information. *Journal of the American Society for Information Science*. 26, (6).

Saracevic, T. (1996). Relevance reconsidered. *Information science: Integration in perspectives.Proceedings of the Second Conference on Conceptions of Library and Information Science*. Copenhagen (Denmark).

Saracevic, T. (1975). Relevance: a review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, 26, (6).

Saracevic, T. (1970). The concept of “relevance” in information science: a historical review. T. Saracevic (Ed.), *Introduction to information science*, 1970

Saracevic, T. (1999). *Information Science. Journal of Documentation*. 50, (12).

Schamber, L. (1994). Relevance and information behavior. *Annual Review of Information Science and Technology*. Meford, NJ: American Society for Information Science, 29.

Schamber, L., Eisemberg, M. B., Nilan, M. S. (1990). A re-examination of relevance : toward a dynamic, situational definition. *Information processing & management*, 26, (6).

Schultz, A. (1970). Reflections on the problem of relevance. New Haven, CT: Yale University Press.

Shuchman, H. (1981). Information transfer in engineering. Glastonbury, Conn. : Fututres group.

Stinson, E. y Mueller, D. (1980). Survey of health professional's information habits and needs. *Journal of the American Medical Association*, 243.

Vakkari, P. (2001). A theory of the task-based information retrieval process: a summary. *Journal of Documentation*, 57, (1).

Vargas - Quesada, Benjamín; Moya Anegón, Félix de; Olvera Lobo, María Dolores. (2002). Enfoques en torno al modelo cognitivo para la recuperación de información: análisis crítico. *Ciência da informação*, 32, (2).

Vickery, B. (1959). Subject Analysis for Information Retrieval. En Proceedings of the international conference on scientific information, 2. En línea: http://books.nap.edu/openbook.php?record_id=10866&page=855

Vickery, B. (1959). The structure of information retrieval systems. *Proceedings of the international conference on scientific information*, 2.

Villanueva, A. (2007). La comunidad de matemáticos en México. *Anales de documentación*, (10).

Wenger, E. (2000). Communities of practice: the key to knowledge strategy. En Knowledge and communities. Boston, MA : Butterworth Heinemann.

Wenger, E.; Mcdermott, R.; Snyder, W.M. (2002). Cultivating communities of practice. Boston: Harvard Business School Press.

Wilson, P. (1973). Situational relevance. *Information Storage and Retrieval*, 9, (8).