

Artículos históricos

Recuperación de información: reflexiones epistémicas de una ciencia en su estado embrionario

[Lic. Alberto Camaraza Monserrate](#)

RESUMEN

La recuperación de información surgió como campo del conocimiento independiente en 1950. El contexto social y filosófico en que se enmarca tiene una singular influencia en la naciente ciencia y determina la orientación de sus presupuestos teóricos y prácticos. La relación bipolar sociedad-ciencia en este caso, provee un sentido y una guía que permite analizar y responder a incógnitas epistémicas inherentes al período de génesis de la recuperación de la información, así como a su posterior desarrollo.

Palabras clave: Recuperación de la información, historia, epistemología.

ABSTRACT

The information retrieval appeared in 1950 as a field of independent knowledge. Its philosophical and social context has a singular influence on the rising science and determines the orientation of its practical and theoretical assumptions. In this case, the bipolar society-science relation provides a sense and a guide that allows the analysis of and response to unknown epistemic questions inherent to a period of genesis of the information retrieval, as well as to its further development.

Key words: Information retrieval, history, epistemology.

Copyright: © ECIMED. Contribución de acceso abierto, distribuida bajo los términos de la Licencia Creative Commons Reconocimiento-No Comercial-Compartir Igual 2.0, que permite consultar, reproducir, distribuir, comunicar públicamente y utilizar los resultados del trabajo en la práctica, así como todos sus derivados, sin propósitos comerciales y con licencia idéntica, siempre que se cite

adecuadamente el autor o los autores y su fuente original.

Cita (Vancouver): Camaraza Monserrate A . Recuperación de información: reflexiones epistémicas de una ciencia en su estado embrionario . Acimed 2005;13(6). Disponible en: http://bvs.sld.cu/revistas/aci/vol13_6_05/aci01605.htm Consultado: día/mes/año.

Desde la antigüedad, desde la Alejandría de *Tolomeo* , hasta las bibliotecas monásticas de la Edad Media , que albergaban el acervo cultural humano, se definió con claros enunciados, la necesidad del desarrollar métodos y técnicas, que no sólo permitieran el control y conservación de las colecciones, sino también la identificación de cada de uno de sus ejemplares con el objetivo de recuperarlos correctamente. A pesar de que, desde entonces, se utilizaron técnicas para el manejo de los títulos y responder a las escasas necesidades de información de la época, no fue, sino con el desarrollo científico y la especialización, con hondas raíces en las revoluciones industriales de los siglos XVIII y XIX, el que maximizó la exigencia sobre estos métodos en aras de satisfacer necesidades cada vez más puntuales, en medio de un conocimiento científico en constante expansión, proclive tal vez en demasía a la divisibilidad nuclear de los campos del conocimiento. Así, con el propósito de adaptar la función clásica de las bibliotecas a esta nueva situación social y a la respuesta más exacta a la duda puntual, es que surgieron las bibliotecas especializadas a principios el siglo XX.

En este contexto, se desarrolla la recuperación de la información como una actividad o componente dentro del haber de archivos, oficinas y bibliotecas, cada una con sus peculiaridades en el desempeño de determinadas funciones sociales. Un punto culminante tiene lugar a finales de la primera mitad del siglo XX; por primera vez, se remueven los cimientos de esta actividad ante la emergencia, sin precedentes, de una generación de información y documentación colosal, con necesidad de un uso intensivo: la segunda guerra mundial.

¿Cuáles son las raíces que provocan este nuevo estadio informacional? Como conflicto bélico, la Segunda Guerra Mundial incorporó nuevas concepciones que tienen una influencia directa en varios ámbitos. Una de estas nuevas concepciones marca la diferencia, fundamentalmente por la sustitución de ciertos valores en el “método” de hacer la guerra. Si bien hasta entonces, las victorias militares las determinaron el poderío económico, la movilización de enormes recursos humanos y la brillantez táctica de determinados generales. Con la Segunda Guerra Mundial, el aspecto tecnológico pasó a ser un factor primordial.

LA INVESTIGACIÓN CIENTÍFICA

La creación tecnológica que durante la primera revolución industrial es fruto de la observación empírica, sufrió una modificación medular durante la segunda. En el seno de la segunda revolución industrial, la creación tiene como trasfondo fundamental la investigación científica, más allá del cociente directo de la práctica y la empiria, aunque se denominó también industrial debido a la masiva aplicación de sus

innovaciones a la industria. Esta razón permitió concebir al producto tecnológico como el engendro de la investigación científica desde los mismos albores del siglo XX.

Pero no sólo es la guerra la que genera una necesidad explícita de un desarrollo científico, sino un aspecto que define al capitalismo y al siglo XX desde sus inicios. La guerra sólo potencia este estado de necesidad, y lo lleva a instancias desmedidas, no sólo del sentido común, sino que también, las cifras lo demuestran (fig. 1).

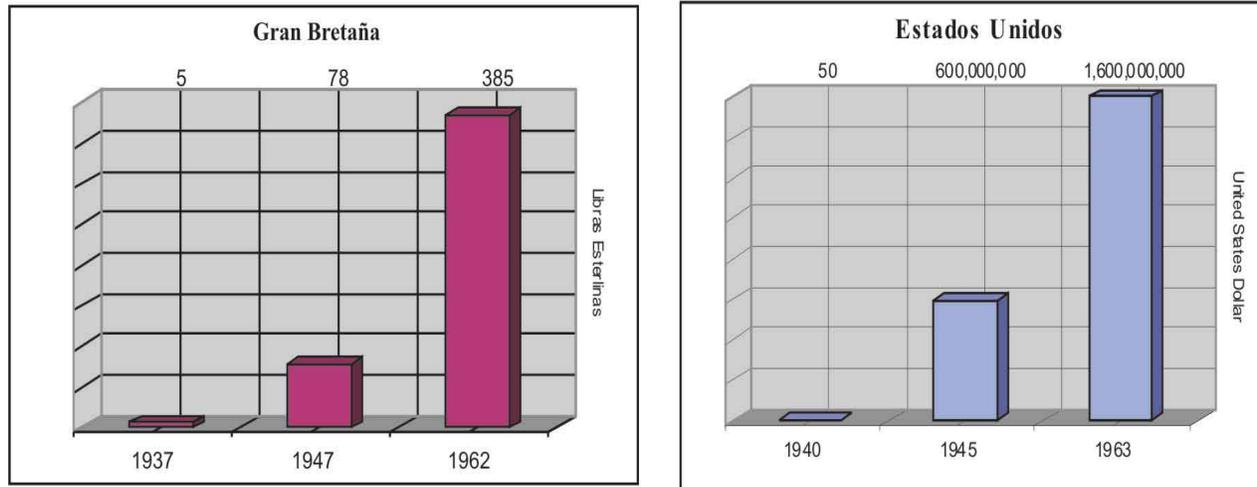


Fig. 1. Gastos en investigación

Las sumas de dinero destinados a sufragar la investigación científica en los países en conflicto aumentan exponencialmente de año en año, principalmente en el caso de los Estados Unidos, quienes alejados de la devastación ocasionada por el conflicto, poseen una economía suficientemente fuerte como para echarse al hombro el enorme gasto financiero, delirante a la razón lógica de una época previa. Esta necesidad inminente de rápidas soluciones tecnológicas que respalden la maquinaria de guerra, la inteligencia, el espionaje, y el teatro político, se afianza en el seno científico para potenciarlo significativamente, y es que “... *la guerra acabó por convencer a los gobiernos que dedicar recursos inimaginables hasta entonces a la investigación científica era factible y esencial para el futuro*”.¹ El ejemplo más claro se concreta en la bomba nuclear, dos mil millones de dólares al costo de una economía de guerra, así en un solo proyecto de naturaleza hipotética se resume el presupuesto empleado en la ciencia a lo largo de toda la historia de la humanidad.²

En septiembre de 1945, al capitular Japón y con ello, establecerse el término del conflicto, la cantidad de información y documentación acumulada durante este período de investigación científica intensiva era incomparablemente superior a cualquier otro en la historia precedente. Este argumento se manifiesta con mayor fuerza en quienes al cesar las hostilidades habían incrementado su *PIB* (producto interno bruto) en dos tercios y eran poseedores, a la vez, de casi dos tercios de la producción industrial del mundo,; así, se sentaban las bases para que su poder financiero hiciera ecos en sus resultados científicos, y donde “*la comunidad científica participó de forma muy activa en el esfuerzo militar bajo la dirección gubernamental, debido al convencimiento del importante función que la ciencia podía desempeñar en el*

desarrollo de la guerra” : 3

Los Estados Unidos

La particularidad del porqué este hecho se manifiesta con mayor impronta en los Estados Unidos que en cualquier otro país participante en la guerra, sin lugar a dudas, está determinado por dos razones fundamentales. En primer lugar, por el explosivo desarrollo económico que experimentaron en el período de conflagración, y en segundo, por el temprano desarrollo de la conciencia sobre el importancia de la ciencia para el desarrollo de la guerra, como argumenta *Fernández Molina* . Durante este período, inyecciones presupuestarias ilimitadas alimentaron la investigación, con el inusual trasfondo del crecimiento económico. En dualidad, actuaban como un catalizador de punta al desarrollo científico, que, a su vez, promovía resultados tecnológicos susceptibles de aplicarse de forma inmediata.

Sin embargo, el fenómeno de posguerra y el alivio de la distensión, posibilitó la aparición de una visión un poco más allá de lo netamente pragmático que la planteada por la conflagración bélica, y es bajo este contexto que alguien como *Vannevar Bush* comienza a cuestionarse el problema que establecía el éxodo de la información registrada, acumulada tras varios años de contienda. Tal vez, él no tuviera idea que en 1945, cuando publicó su célebre artículo “*As we may think*”, que revelaba un carácter premonitorio semejante al de *Michel de Nostredame* - nombre original de *Nostradamus* , (1503-1566), médico y astrólogo francés, autor de las “ *Centurias astrológicas* ”, una famosa colección de profecías publicada en 1555 .

Las ideas de *Bush* no sólo anunciaron el advenimiento de un nuevo campo del conocimiento, sino que más de cuarenta años después, se reconoció como partícipe directo de la idea que condujo al hipertexto. Logró intuir además, con asombrosa exactitud, lo que más tarde serían los sistemas de almacenamiento masivo (CD-ROM, DVD, etc). *Bush* , de formación matemático, fue el hombre designado por *Roosevelt* como director de la Oficina de Investigación y Desarrollo Científico de los Estados Unidos, bajo su égida se encontraba la gestión de más de seis mil científicos envueltos en la investigación para la defensa, que tal vez, le sostuvo en una posición de inigualable atractivo para afirmar:

"A la vista de las actuales inquietudes, el problema no estriba tanto en que se hagan excesivas publicaciones cuanto que éstas han superado con mucho nuestra capacidad presente de hacer un uso real de ellas (...) Profesionalmente nuestros métodos de transmitir y revisar los resultados de la investigación científica es varias generaciones antiguo y, por ahora, totalmente inadecuado en su propósito". 4

El trabajo de *Bush* va más allá de un brillante análisis de la situación informacional de la época, apunta inquietudes en cuanto al almacenamiento, el registro y la recuperación, problemas a los que ve respuesta en la aplicación de las tecnologías recién creadas y su fusión con un objetivo común; “*sus preocupaciones transparentan la fuerte relación estratégica entre investigación científica información y*

*desarrollo” (Linares Columbié R . La ciencia de la información y sus matrices teóricas: contribución a su historia. [Tesis para optar por el título de Doctor en Ciencias de la Información]. Universidad de la Habana : Facultad de Comunicación, 2004). Su visión dilucida con un gran fundamento los componentes de la cadena del desarrollo científico, y el problema que acarrearía la ruptura con los resultados de la investigación. Su mensaje señalaba el riesgo de que el legado de muchos hombres de ciencia podía volverse insignificante, perderse, debido a la enredada telaraña del conocimiento humano acumulado. Esta idea, no sólo podría hacer huérfanos de aquellos a quienes habrían podido inspirar sus predecesores, sino que retardaba severamente el avance y con ello el desarrollo mismo, así *Bush* señala:*

“La noción de Mendel sobre las leyes de la genética le estuvo ausente al mundo por una generación, tan solo porque su edición no llegó a aquellos pocos capaces de entenderla y continuarla, y esta clase de catástrofe se está repitiendo con nosotros” . 4

A *Bush* no sólo se le escucha por su inigualable posición y brillantez intelectual, desde la postura del estadista: desarrollo científico es igual a poder político y ello era una ecuación probada a lo largo de varios años de enfrentamiento. La urgencia de soluciones si bien hasta entonces no era del todo consciente, se irradió de forma explícita y su fundamento fue más que obvio.

Pero, paralelamente, al trabajo de *Bush* , y ya en el seno de la guerra se gestaba en la Universidad de Pensilvania, el “ *Calculador e integrador numérico electrónico* ” (en inglés ENIAC, Electronic Numerical Integrator and Computer) la primera computadora digital totalmente electrónica, que supuso el origen de las tecnologías computacionales .

LA RECUPERACIÓN DE LA INFORMACIÓN

“La invención en 1946 de las tecnologías computacionales fue de progresiva e inmediata aplicación en la nascente esfera de la información, especialmente para solucionar las preocupaciones dominantes en ese lapso de explosión documental, sobre como localizar y buscar información puntualmente” (Linares Columbié R. La ciencia de la información y sus matrices teóricas: contribución a su historia. [Tesis para optar por el título de Doctor en Ciencias de la Información]. Universidad de la Habana : Facultad de Comunicación, 2004).

Con este objetivo, el también matemático norteamericano *Calvin Mooers* , propuso oficialmente en 1950, la creación de un área que afrontara “ *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 Como puede notarse claramente en su enunciado, la nueva ciencia plantea una unión indisoluble con las tecnologías computacionales recién creadas, de las cuales postula un uso intensivo. “*En esencia la recuperación de información se caracteriza por la utilización de la indización para representar y recobrar la información, y por reconocer al lenguaje natural como base para la construcción de sus fórmulas identitarias*” (Linares

Columbié R. La ciencia de la información y sus matrices teóricas: contribución a su historia. [Tesis para optar por el título de Doctor en Ciencias de la Información]. Universidad de la Habana : Facultad de Comunicación, 2004).

Sin embargo, es importante apuntar qué se entendía por “información” a finales de la década de los años 40 y principios de los 50 del siglo XX. Para el lector de este tiempo, discutir sobre el término ‘información’ carece de un carácter práctico, porque la nueva era nos ha familiarizado tanto con el término que lo ha hecho parte indisoluble de nuestras vidas. A pesar de que el término información se remonta al siglo XIII, su uso se generaliza sólo a partir de los años 50 del siglo XX (*Linares Columbié R . La ciencia de la información y sus matrices teóricas: contribución a su historia. [Tesis para optar por el título de Doctor en Ciencias de la Información]. Universidad de la Habana : Facultad de Comunicación, 2004).* La diferencia no radica en la acepción de la noción información “en sí”, sino en la perspectiva desde la cual se adopta.

En busca de una definición...

Emitir un concepto en el plano científico, supone la convergencia del objeto de estudio en cuestión y una noción temporal del conocimiento. Como señalaba *Inmanuel Kant* (1724-1804), -filósofo alemán, considerado por muchos como el pensador más influyente de la era moderna-, no puede obviarse que el mundo que la ciencia investiga debe ser un mundo hasta cierto punto constituido por las ideas de aquellos que lo estudian. El mundo científico posterior a 1945 es un mundo, sin duda, fuertemente influenciado por una combinación de corrientes filosóficas que *postulan el método* científico como único método válido de conocimiento, y que rechazan toda teorización bajo otro contexto. La forma de “hacer ciencia” que dominó al mundo y particularmente a los Estados Unidos para esta época, califica de referente a aquel método científico derivado de las ciencias naturales, modélico por su método de observación y comprobación. “ *Estas corrientes filosóficas dominantes fueron: empirismo, positivismo y pragmatismo, las que entrelazadas y complementarias sirvieron de ambiente filosófico en el cual se inscribe el proceso fundacional de la ciencia de la información*” (*Linares Columbié R. La ciencia de la información y sus matrices teóricas: contribución a su historia. [Tesis para optar por el título de Doctor en Ciencias de la Información]. Universidad de la Habana : Facultad de Comunicación, 2004).*

El positivismo clásico tiene sus raíces en el siglo XIX con el francés *Auguste Comte*. En un principio, explicitaba que el conocimiento de todos los temas, de la astronomía a sociología, debía venir de la correlación de la evidencia empírica. Estas raíces modificadas por lo que a principios del siglo XX emergió como el positivismo lógico, que resaltaba la importancia de la comprobación científica y del empleo de la lógica formal -y que rechazaba que el único y verdadero conocimiento estaba en la experiencia personal-, tuvo una influencia decisiva en el rechazo de las doctrinas metafísicas por su carencia de sentido, y en la aceptación del empirismo como una materia de exigencia lógica.

El período posterior a la segunda guerra mundial se encuentra directamente precedido por el desarrollo a finales del siglo XIX y principios del XX, de lo que constituyó la primera filosofía desarrollada de forma

independiente en los Estados Unidos: el pragmatismo, según la cual la prueba de la verdad de una proposición es su utilidad práctica, el propósito del pensamiento es guiar a la acción, y el efecto de una idea es más importante que su origen. Esta doctrina fue la manera dominante de abordar la filosofía en los Estados Unidos durante el primer cuarto del siglo XX. Su origen teórico se remonta a *Charles Sanders Peirce*, quien fundamentó su teoría con el precepto de que ningún objeto o concepto posee validez inherente o tiene importancia; su trascendencia se encuentra tan sólo en los efectos prácticos resultantes de su uso o aplicación. Por lo tanto, la verdad de una idea u objeto, puede medirse mediante la investigación científica sobre su utilidad práctica.

Emitir un concepto de información en este contexto socio-geográfico sería un fruto enérgicamente permeado por el resultado de su utilidad práctica y su fundamentación empírica, por ende que “*las elaboraciones conceptuales de la noción información, derivadas de la “Teoría matemática de la comunicación”, son las más significativas en los primeros años”* (*Linares Columbié R* . La ciencia de la información y sus matrices teóricas: contribución a su historia. [Tesis para optar por el título de Doctor en Ciencias de la Información]. Universidad de la Habana : Facultad de Comunicación, 2004). ” *La Teoría matemática de la comunicación (...)* dio lugar a la única propuesta formalizada, matemática e implementada del concepto de información” . 3

Desarrollada por *Shannon y Weaver*, esta teoría vio la luz en 1948. Surgió del aumento de la complejidad y de la masificación de las vías de comunicación, como el teléfono, las redes de teletipo y los sistemas de comunicación por radio; y de la necesidad de una base teórica para la tecnología de la comunicación que las sustentaba. En esencia, la creación de *Shannon* despliega un modelo lineal que define claramente a los actores partícipes del proceso comunicativo. Su realidad perseguía dos objetivos fundamentales: determinar la cantidad de información que podía transitar por un canal de comunicación, y los efectos que sobre la decodificación podían producir las perturbaciones introducidas durante el transporte, a las que denominó ‘ruido’.

Como puede apreciarse, la información a la que hace alusión queda totalmente vacía en su apreciación de contenido semántico. “*El modelo matemático informacional pretende representar un sistema general de comunicación, porque excluye de ésta cualquier referencia a contenidos y selecciona únicamente el fenómeno que cualquier sistema de comunicación supone: la transmisión de señales”*. 6 Esta teoría es, sin lugar a dudas, por su visión netamente empírica y pragmática, además de un modelo comunicativo, un modelo de la visión científicista de la época, por lo que se le adjudicó como referente a cualquier aplicación que por ende tuviera su fundamento en el proceso comunicativo.

De otro lado, y muy a pesar de Potsdam, - Conferencia de Potsdam, reunión de los más altos dirigentes gubernamentales de Estados Unidos, la Unión de Repúblicas Socialistas Soviéticas (URSS) y el Reino Unido, celebrada a raíz de la rendición incondicional de Alemania en la Segunda Guerra Mundial- la Segunda Guerra Mundial eliminó el peligro del fascismo, pero engendró otro que sumió al mundo en el terror y la incertidumbre casi, o hasta la desintegración de la URSS en 1991: la guerra fría, situación que engendró un período de influencia decisiva en el desarrollo científico-tecnológico, del cual la recuperación de información no es una excepción, y es el ambiente sociopolítico y económico que

envuelve sus inicios reales y posterior desarrollo. En tanto, cabe señalar como condicionantes fundamentales del período de gestación de la recuperación de información de *Mooers* :

- El artículo “ *As we may think* ” de *Vannevar Bush*, publicado en 1945.
- La creación de las tecnologías computacionales en 1946.
- La formulación de la teoría matemática de la comunicación de *Shannon-Weaver* en 1948.

Y como entorno con el que tiene una relación enteramente recíproca: el ambiente derivado de la posguerra y los inicios de la guerra fría.

“Las relaciones entre la ciencia y la sociedad son enteramente recíprocas. Al igual que en el interior de la ciencia se producen transformaciones por los acontecimientos sociales también estos y en medida creciente, se producen por la influencia de la ciencia”. 2

Inicios reales de la investigación sobre la recuperación de la información

“ Los orígenes de la investigación sobre la recuperación de la información pueden remontarse al año 1953 cuando se realizaron en Gran Bretaña y los Estados Unidos, grupos de pruebas para evaluar el desempeño del entonces controvertido sistema “Uniterm” de Mortimer Taube, frente a los enfoques más convencionales en la temática de la indización y la recuperación” . 7 Sin embargo, a pesar de que estas investigaciones conforman los orígenes, no tuvieron el peso suficiente como para establecer un arquetipo emblemático de investigación en el campo.

No fue sino hasta 1957 en el Cranfield Institute of Technology y otras instituciones asociadas, cuando comienzan a realizarse una serie de pruebas que representaron el comienzo real de la investigación sobre recuperación de la información como disciplina empírica. Las pruebas Cranfield desarrolladas en dos etapas Cranfield I y Cranfield II, tuvieron una considerable influencia en el desarrollo posterior del campo, principalmente porque marcaron el giro de la investigación en un enfoque filosófico y un tanto especulativo a uno con fundamentación en lo empírico Ellis 92,7 además de que estandarizó los métodos de comprobación que eran una exigencia fundamental de la ciencia del momento.

Pero, veamos cómo se estructuraba la experimentación y que noción se tenía de un sistema de recuperación de información. Quien mejor nos muestra al respecto son las palabras del propio *Cyril Cleverdon* , director jefe del proyecto, quien definía el ambiente de prueba como:

“(…) una situación similar a la de un laboratorio, donde libre de la contaminación de las variables operacionales, el desempeño de los lenguajes de indización pudiera estudiarse de forma aislada”. 8

“(…) se tiene que crear un ambiente artificial para que se puedan realizar las pruebas experimentales y aunque tratemos de crear un ambiente lo más real posible, este sigue siendo artificial: una pieza de un

equipo similar a un túnel aerodinámico” . 9 “Partimos del criterio de que todos los lenguajes de indización son amalgamas de diferentes tipos de recursos, esos recursos caen en dos grupos: aquellos que pretenden mejorar el nivel de recobrado y aquellos que pretenden mejorar el nivel de precisión (...). El objetivo de la prueba era investigar el efecto que tendría cada uno de estos recursos, solo o en cualquier posible combinación, en la precisión y en el recobrado. Para hacer posible que esto ocurra, era esencial hacer que todo fuera constante, excepto la variable que se investigaba” . 8

No es difícil apreciar la fuerte influencia de la visión ingenieril; en realidad se importaba el método aplicado de otra ciencia a la recuperación de información. Los lenguajes de indización y su función se dan en contextos con marcadas características de rigidez, e incluso se toman como actores componentes de un sistema mecánico, a partir del supuesto que cada quien tiene su función preconcebida y una influencia objetiva. Al igual que existen las amalgamas de recursos de indización que pretenden mejorar el nivel de recobrado y precisión, existen los componentes que hacen el auto más estable, para lograr maniobrabilidad, y aquellos que lo hacen más ligero para ganar velocidad. La semejanza a un sistema físico es la característica fundamental de los experimentos de *Cranfield* .

Otro de los aportes medulares de estos experimentos fue la estandarización de los criterios de relevancia y pertinencia, futuras medidas adjuntas a la evaluación de cada sistema de recuperación. Así tuvo lugar el advenimiento de los años sesenta, con una base estructural y un intento de metodología fundada en la investigación, que por su peso, podía aceptarse tácitamente en los trabajos posteriores, tanto en pruebas experimentales como operacionales. De hecho lo fue, y es a lo que Ellis denominó “influencia indirecta” (Ellis, 1992). *“La posición adoptada por Mooers (y por muchos otros pioneros como Mortimer Taube, James Perry, Allen Kent, Hans Peter Luhn etc.) fue la de concentrarse en la construcción de sistemas. Este énfasis en los sistemas (o enfoque centrado en los sistemas) formulado en los años 50, fue el esquema exclusivo de la recuperación de información por varias décadas. Todavía predomina en buena parte de la investigación en este ámbito” . 5*

Para esta etapa, la tecnología en los sistemas empleados era muy simple. Se utilizaron ampliamente las búsquedas mecánicas en índices manuales impresos. Pero mientras la búsqueda mantuvo tales principios, las entradas o términos de indización asignados a los documentos corría con enormes limitaciones: una indización muy detallada desembocaba en el riesgo de un catálogo interminable. Ya para entonces, la búsqueda entre un sinnúmero de entradas evidenciaba que el ser humano no era capaz de lidiar con ello eficazmente.

Así, el advenimiento de las bases de datos computarizadas, donde los sistemas pudieran hacer las búsquedas con un algoritmo determinado, denotó que la indización de nuevo podía hacerse más detallada. No importaba cuantos índices existieran por cada documento, mientras más detallada fuera la indización, más efectiva sería la búsqueda; de esta forma, el sistema podía devolver un mayor número de documentos relevantes -las máquinas no tenían problema alguno con ‘buscar entrada por entrada’ de principio a fin, ellas no experimentaban cansancio ni agotamiento. Entonces, el problema dejó de radicar en realizar las búsquedas entre un sinnúmero de elementos, sino en que para que esta fuera lo más efectiva posible debía ser el producto de un intenso trabajo de indización precedente. Los indizadores tenían encima el rigor por entero de una ardua tarea, y ello incrementó notablemente los costos.

Entonces vio la luz la idea de la búsqueda a *free-text* (texto libre). Podía hacerse la búsqueda en de documentos enteros, no importaba la cantidad de ellos, a fin de cuentas se obviaban los costos de la indización. Pero a pesar de que la idea se hizo muy popular y su uso se extendió ostensiblemente (aún existe), enseguida aparecieron detractores del método, y sus razones eran fundadas. La labor de indizar no era simplemente tomar palabras al libre albedrío para describir al documento, era aquella tarea de “tomar las correctas” entre otras cuestiones que iban más allá de la ligereza del usuario, quién por intuición propia etiquetaba su interrogante. Una de las soluciones más efectivas que se dio al problema fue el denominado ‘*relevance feedback*’, o retroalimentación por relevancia. La retroalimentación por relevancia significaba de alguna forma, incrementar la utilidad de la prescripción de búsqueda (*query*) agregando términos extraídos del recobrado inicial -de documentos juzgados por el sistema como relevantes-, la realidad mostraba que el sistema podía operar mejor con una prescripción de búsqueda más extensa, mientras los usuarios en su mayoría se limitaban a pequeñas expresiones.

Para esta etapa, sin embargo, uno de los principales impedimentos que han señalado los estudiosos, era la poca disponibilidad de textos legibles por máquina (machine-readable text), lo que ataba a los investigadores en sus pretensiones de construir grandes sistemas que operaran en el universo digital.

La función del usuario en los inicios de la recuperación de la información

Volviendo a la cita anterior, pongamos atención a la semejanza sobre el ambiente de prueba que citaba *Cleverdon* con el de un túnel aerodinámico. En un túnel aerodinámico, todas las magnitudes físicas son controlables: velocidad del aire, vectores de dirección, sentido, rozamiento estimado; pero aparece la incógnita que no se aprecia en ningún lugar, ¿y la entidad usuario?, con alguna otra función, más que de receptor pasivo de información, en el mejor de los casos considerado sólo el ‘origen’ de la consulta. Sin dudas, la noción usuario implicaría un riesgo de contaminación inminente sobre el ‘ambiente de laboratorio’ que se tenía como objetivo lograr, y sobre los supuestos empiristas y objetivistas preponderantes.

“*La aparición y desarrollo de los sistemas de recuperación de información se encuadra en orientaciones teóricas con supuestos empiristas y objetivistas y con modelos basados en la teoría matemática de la comunicación*” (*Linares Columbié R* . La ciencia de la información y sus matrices teóricas: contribución a su historia. [Tesis para optar por el título de Doctor en Ciencias de la Información]. Universidad de la Habana : Facultad de Comunicación, 2004). Precisamente, el modelo *Shannon-Weaver* como se había mencionado, despliega un proceso unilateral donde la comunicación se concibe en un solo sentido, el mecanismo de retroalimentación del sistema en el que el usuario tiene una función activa no está concebido. Por lo que, con este modelo como base la función del usuario (o destino) se limita a una recepción pasiva de lo que se tiene como objetivo transmitir. Sin embargo, el logro del objetivo fundamental de la teoría matemática de la comunicación (medir la cantidad de información que puede transitar por un canal y el efecto negativo del ruido) peligra cuando no se dispone de la capacidad de evitar las distorsiones, 6 que introduciría el usuario con su subjetividad, una fuente de perturbación a la hermeticidad del sistema. “*Como se cree que el usuario introduce el*

subjetivismo y que, por tanto, coadyuva al distanciamiento de lo que debía ser una ciencia, se maximiza la necesidad de los sistemas de información, de una objetividad irrefutable; y prevalecen sus necesidades como entidad casi despersonalizada” (Linares Columbié R . La ciencia de la información y sus matrices teóricas: contribución a su historia. [Tesis para optar por el título de Doctor en Ciencias de la Información]. Universidad de la Habana : Facultad de Comunicación, 2004).

Un sistema de recuperación era un producto que funcionaba bien en la medida que mejor “engrasado” estuviera su dispositivo mecánico, por lo que su mejor funcionamiento en consecuencia, dependía del perfeccionamiento de sus partes, los lenguajes de indización y las herramientas tecnológicas tenían el lugar primordial. En este contexto, surge la tendencia de la formación de usuarios, como única forma de que estos conocieran un sistema diseñado y construido sin contar con ellos. 7 Bajo otra perspectiva histórica, se abordará el problema desde otra arista.

EL DESARROLLO DE LA CIBERNÉTICA

Si se intenta modelar esquemáticamente la recuperación de información manual, clásica de las bibliotecas, con el punto de vista de Cleverdon sobre lo que debía ser un sistema de recuperación de información, el punto de partida sería una solicitud (despersonalizada), que desencadena un proceso en el cual el bibliotecario es el eje central, y quien, por medio de los lenguajes documentarios y los sistemas de clasificación y almacenamiento, realiza la labor de identificar aquellos documentos capaces de cubrir desde la perspectiva informacional (al menos en apariencia) la necesidad inicial de donde se parte; y llevarlos a manos del usuario en cuestión.

En este caso, el bibliotecario puede, una vez identificado el título, abrir el libro e intuir si es relevante o no, de lo contrario puede remitirse a la búsqueda manual en largos índices de materia impresos para obviar este paso. Sin embargo, esto es un proceso engorroso que implica tiempo y esfuerzo. Pensemos en una labor tan extensa como la búsqueda en la versión impresa del *Chemical Abstracts* , el simple hecho de buscar artículos que mencionen dos tópicos de materia diferentes, exige la búsqueda por ambos términos independientes, luego, en la lista subsiguiente, el astronómico trabajo de identificar aquellos elementos coincidentes en ambas. En comparación con el hombre esto es algo que las máquinas pueden hacer muy bien, la cantidad no es un problema para las máquinas, ellas no experimentan fatiga, y no desisten si una actividad es aburrida.

La recuperación de la información que tiene orígenes en *Mooers* se diferencia de la recuperación clásica en el uso de técnicas o máquinas que facilitan el proceso. Si bien los primeros pasos que se dieron en este sentido, eran prácticamente manuales -los de *Batten* y *Cordonnier* que se basaban en el principio de la coincidencia óptica, el de *Mooers* que empleaba tarjetas señaladas en los bordes y, un tiempo más tarde, el de *Taube* , el sistema Uniterm - los sistemas sucesores pretendían obviar del todo la presencia humana como eje central del proceso. La función antes mencionada del bibliotecario, quien tenía el lugar principal, pretendía ser sustituido por algo o alguien capaz de lograr una eficiencia total en una

tarea que se hacía enormemente compleja con el crecimiento inusitado del universo documental.

Sin embargo, el problema fundamental que surge con la pretensión de aislar la presencia humana como eje central de la recuperación es que la actividad del bibliotecario se ubica en un 'proceso inteligente', sobre todo porque opera con las bases del lenguaje, que es complejo; y donde la acción de identificar está precedida, a su vez, por procesos mentales complejos de análisis y síntesis, sin obviar su capacidad de decisión. El logro de una 'máquina inteligente' que pueda prescindir del hombre, supone la cualidad de imitarlo en un sentido y de superarlo en otro. La cibernética, que apareció en el decenio de los años cuarenta del siglo XX, tenía entre sus objetivos imprimirle a sus logros ambas cualidades.

“ Las computadoras electrónicas que se desarrollaron a una velocidad de vértigo luego de la segunda guerra mundial, especialmente tras la invención del transistor, tenían una enorme capacidad para hacer simulaciones lo que hizo mucho más fácil desarrollar modelos mecánicos de las que, hasta entonces, se consideraban las funciones físicas y mentales básicas de los organismos, incluido el humano ”. 1 Siendo así, la cibernética se convierte en el punto de encuentro de las ciencias que se ocupaban hasta entonces del estudio de la actividad vital del organismo: la psicología y la fisiología .

Mientras que la fisiología avanzaba rápidamente en el estudio de la autorregulación de las funciones sistémicas, la psicología hacía suyo otro plano de relaciones que no eran intraorgánicas, sino que afectaban a la interacción del organismo con el medio, el nivel de la conducta de adaptación. Se trataba precisamente de la conducta del organismo íntegro respecto a su entorno, y no de funciones aisladas.

10 Se consideraba que la línea divisoria entre la fisiología y la psicología se hallaba precisamente en que la primera estudiaba funciones sistémicas aisladas; la segunda, la conducta del organismo íntegro.

“ La cibernética salió de los límites de la fisiología y la psicología, al fundamentar matemáticamente sus principios y realizarlos técnicamente en máquinas ” (Yaroshevski, 1983, p. 295). Las ideas y pretensiones fundamentales de la nueva ciencia se sintetizan con el trabajo del matemático estadounidense *Norbert Wiener* (1894-1964) , en *“Cibernética”* , en 1948. La cibernética estudia el control y la comunicación en las máquinas, los animales y las organizaciones. No comprende sólo del control automático de la maquinaria por computadoras y otros aparatos electrónicos, sino también del estudio del cerebro y del sistema nervioso humano y la relación entre los sistemas de comunicación y control.

Surge como corriente que integraba los resultados de las ciencias matemáticas, técnicas y biológicas. Así, *“ tres ramificaciones de la psicofisiología se convirtieron en premisas de la cibernética: la propiamente fisiológica, la psicofisiológica, y la propiamente psicológica ”.* 10

La ramificación propiamente fisiológica se basa en la teoría de la homeostasis o del equilibrio. La homeostasis requiere que el organismo sea capaz de detectar la presencia de cambios en el medio y de controlarlos, una pequeña variación con respecto al nivel establecido iniciará una respuesta homeostática que restituirá el estado deseado para el medio. Fue desarrollada inicialmente por *Claude Bernard* (1813-1878), un fisiólogo francés considerado fundador de la medicina experimental.

La vertiente psicofisiológica se sustenta en el mecanismo de autorregulación de la conducta, íntegramente mediante señales. Uno de sus principales exponentes fue *Iván P. Pávlov* (1849-1936), fisiólogo y Premio Nóbel ruso, conocido por sus estudios sobre el comportamiento reflejo. Sus experimentos más famosos demostraron la existencia de reflejos condicionados y no condicionados en los perros, que tuvieron gran influencia en el desarrollo de las teorías psicológicas conductistas, fisiológicamente orientadas, durante los primeros años del siglo XX

EL CONDUCTISMO

Los experimentos de *Pávlov* sobre el condicionamiento de los reflejos en animales a pesar de que fracasaron en su intento de generalizar la teoría, tuvieron una influencia fundamental en el posterior desarrollo del conductismo, importante escuela psicológica de la primera mitad del siglo XX, que ejerció una basta influencia sobre la ciencia en general. La nueva corriente estaba en concordancia con el ambiente imperante en la época. Hasta la década de 1960, la psicología estuvo imbuida de consideraciones de índole eminentemente práctica; los psicólogos intentaron aplicar la psicología en la escuela y en los negocios, se interesaron muy poco por los procesos mentales e hicieron hincapié exclusivamente en la conducta. En un principio, el movimiento fue liderado y divulgado por el psicólogo estadounidense *John B. Watson* (1878-1958), fundador de la corriente conductista en la psicología.

En 1913, en su artículo “*La psicología según un conductista*”, *Watson* expresaba que la imposibilidad de la psicología de ocupar un puesto digno entre las ciencias naturales durante su medio siglo de existencia, se debía a la falsa comprensión del objeto y del método de las investigaciones psicológicas. Pensaba que el objeto de la psicología debía ser en lugar de la conciencia, la conducta; rechazó el método subjetivo (introspectivo) y emplazó, en su lugar, lo que denominó el método objetivo. Desde el punto de vista conductista, la psicología debía centrar su análisis en la conducta observable del sujeto, como único método válido de conocimiento, el vuelco al estudio de los fenómenos psíquicos no observables, en su opinión, carecía de un sentido propio.

Y no es de extrañar que, en esta etapa, surgiera una corriente psicológica que respondiera a un modo ‘objetivo’ de estudiar al hombre, y menos en medio de las singularidades geográfico-filosóficas referidas anteriormente. La teoría de *Watson*, en concordancia con el positivismo lógico, pretendía hacer de la psicología una ciencia aislada de todo subjetivismo y voltear la mirada a aquella parte del hombre que es observable y cuantificable: la conducta. Muchos psicólogos le llamaron al conductismo la psicología sin psique.

Sin embargo, el paradigma conductista tenía por derecho propio, al aplicar claramente las bases pragmático-positivistas regentes, un puesto garantizado en la continuidad. “*Antes de que transcurriera una década, el principal oponente de Watson, Titchener - Edward Bradford Titchener*, psicólogo estadounidense profesor la Universidad de Cornell, cuya labor llevó al apogeo del enfoque introspectivo

para el estudio de la psicología -, *que había negado todo significado de los postulados conductistas para la psicología como ciencia, debía reconocer que la ola del conductismo arrastraba a toda la psicología norteamericana*” . 10

LA INVIABILIDAD DEL CONDUCTISMO PARA LA CIBERNÉTICA

“Creo que se puede escribir un tratado de psicología sin utilizar nunca los términos conciencia, estado mental, mente u otros de la misma categoría” . 11

Sin embargo a finales del siglo XIX, *Wilhelm Wundt* (1832-1920), un psicólogo alemán, considerado el fundador de la psicología como ciencia independiente, en experimentos con distintos fenómenos fisiológicos sobre las sensaciones, observó que existía “algo”, no observable empíricamente, que iba más allá de las expresiones fisiológicas del organismo. Así para él, la tarea principal de la psicología era el estudio de esa estructura no observable: la conciencia, que abarca no sólo el campo de las sensaciones sino también el de los sentimientos, las imágenes, la memoria, la atención, la percepción del tiempo y el movimiento. Debido a que el estudio se centraba en los contenidos y la dinámica de la conciencia, no es extraño que la principal herramienta de estudio fuera la introspección. El vuelco del conductismo recayó en el desconocimiento de esta estructura.

La cibernética, que se afianzó de la psicología como el método teórico para explicar la dinámica de las estructuras mentales, encontró un abismo en la teoría conductista que se hizo inviable. Si la cibernética tenía entre sus objetivos desarrollar máquinas inteligentes, capaces de dirigir procesos y a otras máquinas a su vez (la capacidad del automatismo y la autorregulación), era lógico que encontrar una estructura capaz de responder a los mecanismos y procesos mentales que brindaban esta capacidad al ser humano, fuera una de sus prerrogativas iniciales. El conductismo, que era la corriente imperante, no encontraba respuesta a ello, o no pretendía siquiera darla. Lo que mediaba entre la actividad sensorial y la respuesta motora, necesidad vital para sí, había que buscarlo en otro lugar.

Por otro lado, los planteamientos de la *Gestalt* -del alemán, *Gestalt* -, una escuela de psicología que se dedicó principalmente al estudio de la percepción y que postulaba que las imágenes se perciben en su totalidad, como forma o configuración, y no como mera suma de sus partes constitutivas, parecían apropiados, al menos, como punto de partida. La *Gestalt* situaba, en primer plano, el principio de la autoorganización y el carácter estructural de la actividad psíquica. Partía de que la organización general en un todo determinaba las propiedades y la conducta de sus componentes aislados, para la cibernética este enfoque sistémico se convirtió en determinante. No obstante, aunque la categoría imagen que brindaba la *Gestalt* era mucho más apropiada para los efectos de la cibernética, aun no resolvía positivamente la relación directa, indudable, entre la imagen y la acción. Así, los esfuerzos iban dirigidos a lograr esta conexión indispensable que no tenía más que ciertas aproximaciones teóricas de carácter especulativo.

LAS MICROCOMPUTADORAS

En el año 1971, ocurre, sin duda, un acontecimiento de singular impacto en el campo de la microelectrónica: la aparición del microprocesador. El microprocesador surgió gracias al descubrimiento de la tecnología de integración a gran escala - *VLSI* , acrónimo de *Very Large Scale Integration* - del circuito integrado que había revolucionado el mundo con su aparición en 1959. El microprocesador redujo el tamaño de la CPU al de una sola pastilla o chip de silicio. A esta época, pertenece el microprocesador Intel 4004, una computadora en un chip.

La demanda totalmente abrumadora e inesperada de la Altair 8800, la primera computadora de mesa o microcomputadora que presentó, en 1974, Micro Instrumentation Telemetry Systems (MITS), con estas características, abrió un sendero que aún hoy no parece tener límites. El acceso a los enormes mainframes prácticamente limitado a altos académicos, funcionarios del gobierno y la defensa, había delineado muy bien una clase donde por derecho podían escucharse pocas voces. La microcomputadora ofreció la posibilidad de acercamiento al ciudadano medio, después se convertiría en un fenómeno de masas. El monopolio de las herramientas informáticas –que poseían el gobierno y algunas grandes empresas con suficiente capital como para adquirir computadoras como la UNIVAC , la primera computadora comercial- y con ello, de la información, quedó en una era de oscuridad. Para entonces *“existían unos pocos profesionales, los de procesamiento de datos que eran los propietarios esenciales de las computadoras centrales, estos cófrades disfrutaban de las bendiciones de un monopolio de la información”* . 12

“Cuando las computadoras llegaron, por primera vez, a las oficinas de las grandes compañías, se habló de la llegada del "cerebro gigantesco", que contendría toda la información necesaria para gestionar la empresa”. 13 Precisamente, la utilidad de la computadora no sólo se asentaba en su capacidad de procesar, sino también de almacenar, muy a propósito la convergencia de ambas era la gran novedad. En su tiempo, el microfilm fue impactante gracias a su alta capacidad de comprimir la información en un espacio físico limitado, por lo que la revolución que supuso el almacenamiento de la información fue trascendental, pero la dicotomía capacidad de almacenamiento-procesamiento no estaba concebida aún.

Bush se había acercado a algo similar con el MEMEX, aunque su espacio histórico no le permitió llegar más allá. Aunque solo quedó en el plano teórico, MEMEX fue la solución práctica que de *Bush* , contenía un teclado y una pantalla, su modo de registro de información era el microfilm y contendría relaciones referenciales entre los documentos almacenados. La microcomputadora que se extendió a una velocidad vertiginosa a partir de los años setenta ostentaba un precio bastante módico, por lo que su masificación no tuvo tropiezos con barreras económicas significativas. Su desarrollo y el de disciplinas afines como la inteligencia artificial, hicieron no menos de la recuperación de información un proceso interactivo. El fundamento que brindaba el desarrollo de la psicología cognitiva brindaba una base estructural significativa para un cambio de enfoque en este sentido. Así, los sistemas de recuperación de información comenzaron, en perspectiva, a concebirse lejos del detrimento de la figura del usuario, al menos en una vertiente significativa.

LA PSICOLOGÍA COGNITIVA

El mentalismo, desterrado de la psicología académica desde finales del siglo XIX a causa de sus debilidades metodológicas y conceptuales, se reasumió con decisión por la nueva psicología cognitiva. *“La analogía de la computadora en que se basa prestaba a los psicólogos un vocabulario, unas directrices y una instrumentación adecuados para el estudio de la mente”*. En esencia, se ocupaba de los procesos por medio de los cuales el individuo obtiene un conocimiento del mundo y toma conciencia de su entorno, así como de sus resultados. *“Muchos de los temas tratados por la psicología cognitiva utilizan la terminología del “procesamiento de información” común para computadoras y procesos mentales. El vocabulario asociacionista se sustituyó por procesos de ‘codificación’, ‘recuperación’, ‘búsqueda de información’, y por componentes estructurales como ‘memoria operativa’ ‘memoria a largo plazo’ etcétera “*. 14

La analogía funcional mente-computadoras en la que se forman los preceptos de la psicología cognitiva, como se basa en que ambos son sistemas de propósito general -ambos codifican, retienen y operan con símbolos y representaciones internas- conllevó a la estratificación en dos vertientes de esta psicología: una débil y una fuerte. En la vertiente débil, la relación mente-computadora es de modo esencialmente instrumental; a pesar de que utiliza el vocabulario del procesamiento corresponde a lo que se denomina con propiedad psicología cognitiva. La que corresponde a la fuerte ha ido más allá de los límites de la psicología y considera que la computadora es algo más que una simple herramienta conceptual, dio lugar a una disciplina llamada ciencia cognitiva, muy pujante hoy. *“Si la computadora y la mente humana son sistemas de propósito general, el objetivo del científico es elaborar una teoría unificada sobre el procesamiento de información que englobe ambos sistemas”* . 14 *“El enfoque cognitivo fue adoptado de forma entusiasta en campos tan diversos como el de la computación, la psicología, la sociología, la inteligencia artificial, la lingüística y la ciencia de la información, y en la confluencia de los intereses de estos grupos diferentes se ha percibido el surgimiento de una disciplina, la ciencia cognitiva”* . 7

RECUPERACIÓN DE INFORMACIÓN: PSICOLOGÍA Y LENGUAJE

El conductismo, que toma como cimiento los principios evolucionistas de Darwin, concibe la universalidad de las leyes de la conducta, a partir del parentesco o tronco común entre las especies, incluido el hombre.

“(…) El conductista, en su anhelo de crear un esquema único de reacciones animales, no admite una línea divisoria entre el hombre y los animales. La conducta humana en todas sus sutilezas y complejidad, constituye tan solo una parte del esquema general de las investigaciones del conductista”.

11

Como se observa, la universalidad de las leyes de la conducta no implicaba naturalmente que los repertorios conductuales fueran iguales en todas las especies, pero las diferencias serían meramente cuantitativas. 14 No obstante al señalamiento anterior sobre la inviabilidad del conductismo para la cibernética, la similitud a que hace referencia: el hombre “como una especie más”, desecha casi por

entero un presupuesto fundamental del que sólo los seres humanos son beneficiarios: el lenguaje, y su relación bilateral con el hombre. Según Watson, *“la palabra es una reacción verbal que sustituye al objeto (o a la situación) en virtud del principio del reflejo condicionado. Pero el objeto mismo, desde el punto de vista conductista, desaparece, luego de las operaciones musculares que el individuo es capaz de realizar con él. Dicho de otro modo, el habla no puede ir más allá de los límites de la experiencia motora del individuo (...) Lo que formaba la base de la comunicación eran las señales verbales directas, y no el lenguaje como sistema histórico-social determinante de la estructura y del significado de dichas señales”* . 10

A la predominante perspectiva conductista del lenguaje, replicó el célebre *Noam Chomsky* - lingüista, profesor y activista político estadounidense, fundador de la gramática generativa, un sistema de análisis del lenguaje que ha revolucionado la lingüística moderna , en 1957 con el título: *“Syntactic Structures”* (Estructuras sintácticas), donde argumentaba que una descripción razonable sobre cómo las personas emplean el lenguaje debía ir más allá de las observaciones de la conducta, ir a las leyes mentales subyacentes. *“Mientras los conductistas se inclinaron a acentuar la naturaleza de ‘cadena de señales’ de las expresiones (una palabra seguía a la otra,) Chomsky enfatizó la estructura jerárquica de las expresiones lingüísticas”* . 15 La afirmación de *Chomsky* no hizo más que confirmar los pesares que sobrevendrían a la inteligencia artificial en materia de análisis lingüístico.

“Una persona debe tener alguna representación de la oración como un todo en la mente antes de articularla, así entonces, debe aplicar las reglas gramaticales a fin de producir su articulación. Similarmente, al escuchar primero debe percibir la oración como un todo (no una cadena de palabras separadas) para luego aplicar las reglas gramaticales que le permitirán entenderla” . 16

La recuperación de información a que nos remite, tiene una relación directa y por entero con el lenguaje, *solo los seres humanos tienen un adaptable, preciso y simbólico modo de comunicación que permite la expresión de cualquier tipo de información imaginable. “Solucionamos los problemas con efectividad mayormente porque conocemos mucho, y conocemos mucho porque aprendemos no sólo de nuestra propia experiencia en el mundo que nos rodea, sino de la experiencia de otros, que ha llegado a nosotros por medio del lenguaje, el lenguaje permite transmitir el conocimiento de generación en generación y de esta manera provee el vehículo de la cultura humana”*. 15

La función de la recuperación de información es precisamente, la de un vehículo que provee un acercamiento adecuado al acervo documental humano, al que, de un lado, aportamos nosotros mismos, y de otro, en mayor medida, la historia precedente. La actividad científica es tan intensa, sobre todo desde los últimos sesenta y cinco años que, como avizorara *Bush* en 1945, ha llegado a exceder nuestras capacidades de lidiar con el universo documental que la ciencia ha fabricado. Pero el problema no sólo estriba en que no podemos lidiar con toda la información disponible, sino que nuestra capacidad de informarnos se ha vuelto muy selectiva. No se puede simplemente señalar con el dedo, y dirigirnos a aquello que intuitivamente nos concierne, por lo que el medio que tenemos de expresar nuestra necesidad de información nos queda limitado al lenguaje. El lenguaje, permite la representación de todas las formas posibles de información, de ahí que lenguaje y recuperación tengan una doble relación,

primero, porque es el modo y el método (no absoluto) de representar la información y segundo, porque es la forma de expresión de nuestra necesidad de información, que tiene su raíz en los procesos psíquicos.

La psicología cognitiva se interesó en el lenguaje por dos razones principales: primero, porque el lenguaje es una destreza cognitiva, la habilidad para producir y entender expresiones lingüísticas es una capacidad fundamental de la mente humana, por lo que cae dentro del alcance de la psicología cognitiva. 14 En segundo lugar, el lenguaje es un entretrejo de otras habilidades cognitivas “ (...) *las palabras que adquirimos de nuestra cultura, influyen en el modo que percibimos, recordamos y pensamos* “. 15 Así, los estudios del lenguaje han concluido que el hablante se encuentra constreñido por sus intenciones, su conocimiento del mundo, su interlocutor, sus estados mentales, lo que podría resumirse en fin, en el producto de su estado cognitivo. Y tal vez, estas palabras puedan resumir la razón principal del llamado enfoque cognitivo para la recuperación de información.

El enfoque cognitivo en recuperación de la información

Los esfuerzos principales del enfoque cognitivo en la recuperación de información, recaen en consecuencia sobre si el sistema debe operar, de una u otra forma; con un modelo del estado cognitivo del usuario. La razón fundamental, en contraposición con el modelo tradicional, es considerar al usuario como parte integrante del sistema. El modelo cognitivo a diferencia del tradicional, no se pone en marcha a partir de momento en que el usuario realiza la consulta (como la traducción lingüística de la necesidad), sino que comienza incluso antes de que se produzca la necesidad de información. Son especialmente relevantes aquellos aspectos que influyen en el estado cognitivo del individuo, su estado anímico, sus creencias, su grado de motivación o áreas de interés y la forma en que estos interactúan y provocan la necesidad de información. 7

Uno de los primeros trabajos que se realizaron en esta vertiente fue el de *Oddy*, orientado a la recuperación de la información mediante el diálogo hombre-máquina. En el proyecto, que se denominó THOMAS, el usuario no trataba directamente con la estructura de la información en la base de datos, interactuaba con esta por medio de la creación de un modelo de sus percepciones o requerimientos. El modelo que creaba el sistema se utilizaba entonces para identificar el material para la recuperación, a partir de las bases de datos en un proceso inteligente mediante el cual el sistema creaba una imagen, que posteriormente se comparaba con la red de asociaciones en la base de datos, y el documento que se consideraba el más similar a esa imagen era el que se presentaba al usuario. 7

El trabajo de *Oddy* con el programa THOMAS, concluyó con las proposiciones que hiciera *Belkin* sobre las bases de los estados anómalos del conocimiento (*ASK, Anomalous State of Knowledge*). La teoría del *ASK* tenía su fundamento en el proceso que se consideraba que estaba detrás de la interacción de la recuperación de información, que en resumen comprendía los rasgos principales del enfoque cognitivo.

7

“Un receptor investiga el sistema de comunicación y reconoce la existencia de una anomalía en su

estado de conocimiento, anomalía que es afín a la separación del estado de conocimiento del generador, que identifica la estructura conceptual que debe comunicarse. El receptor convierte entonces este estado anómalo del conocimiento (ASK) en determinada estructura comunicable (por ejemplo, una solicitud), que se utiliza para recuperar del cuerpo de textos, algún texto que pudiera resultar apropiado para resolver la anomalía. El receptor interpreta el texto para descubrir la estructura conceptual que está debajo de este, esta estructura interactúa con el ASK del receptor; posteriormente el receptor toma una decisión sobre si la anomalía fue suficientemente resuelta. Si este es el caso, se cierra el sistema, si no, se reinvestiga el sistema con un nuevo ASK como base” . 17

La propuesta sobre el ASK de *Belkin* y la de otros defensores del enfoque cognitivo coincidían en las dificultades con las que se enfrenta el usuario a la hora de expresar la necesidad, esta expresión que se concreta en la solicitud o prescripción de búsqueda, es algo que no se cuestiona en el paradigma físico, pero que el enfoque cognitivo revela como un producto de especial importancia en el funcionamiento de los sistemas de recuperación. Para 1987, las palabras de *Tefko Saracevic*, editor jefe de una de las revistas científicas principales sobre recuperación de la información, evidencian la existencia real de un nuevo foco de investigación fuertemente legitimado por un número significativo de investigadores:

“El trabajo seguido por un grupo de catedráticos, no es sólo la descripción y recomendación de un modelo específico – va incluso más allá de ello – es el establecimiento del paradigma de una incipiente y deseada dirección en la investigación de los sistemas de información. Ella incorpora y mezcla en un contexto, los enfoques de la investigación en recuperación de información de un lado, dentro de la ciencia de la información, con los enfoques de la investigación en sistemas expertos dentro de la inteligencia artificial del otro. Este nuevo paradigma, recuperación de información-sistemas expertos, a pesar de que refiere dos áreas del conocimiento es, no obstante, muy diferente de aquella visión tradicional encontrada en la recuperación de información ‘clásica’, y en la tal vez también investigación ‘clásica’ sobre sistemas expertos. Esto es un cambio particularmente significativo para la comunidad de investigación de la ciencia de la información. Por supuesto, el tiempo mostrará si este nuevo paradigma llegará a concretarse. Sin embargo, lo que está demostrado es que existe un grupo fuerte de investigadores que reta la establecida definición del problema y sus enfoques en un campo de investigación cubierto de cerca por esta revista a lo largo de un cuarto de siglo.” . 18

LA INTELIGENCIA ARTIFICIAL Y LOS SISTEMAS EXPERTOS

La idea de la indización automática comenzó a acariciar la curiosidad de los científicos, desde que tuvieron que lidiar con los enormes fondos documentales. Tanto el costo de la indización como lo arduo de la tarea, y las prerrogativas acerca de que si era realmente viable o no la búsqueda a texto libre (*free text*), sedujo la imaginación de quienes hacían ciencia al pensar que si lograban un método eficiente para indizar automáticamente los documentos, todos los problemas estarían resueltos. La noción, que fue impensable en términos categóricos durante varias generaciones, cobró enormes expectativas particularmente después de la segunda mitad del siglo XX, aunque en realidad, las promesas de la

inteligencia artificial se pusieron más allá de los límites abarcables en la práctica.

El término 'inteligencia artificial' fue acuñado, en 1956, por *John McCarthy*, del Instituto de Tecnología de Massachussets. En ese año, se celebró la conferencia de Dartmouth, en Hanover (Estados Unidos), y en ella, *McCarthy*, *Marvin Minsky*, *Nathaniel Rochester* y *Claude E. Shannon* establecieron las bases de la inteligencia artificial como campo independiente dentro de la informática. En su sentido más amplio, el término inteligencia artificial, indicaría la capacidad de un artefacto de realizar los mismos tipos de funciones que caracterizan al pensamiento humano. Los primeros 'superproyectos' cuyas metas iban hasta el terreno de la ficción -máquinas capaces usar y de reconocer el lenguaje, emitir criterios, análisis...- minimizaron paulatinamente sus expectativas a medida que percibieron lo infructuoso de sus intentos. Con el decursar del tiempo, comenzaron a plantearse metas más modestas, como el intento de lograr determinadas habilidades, en determinados campos prácticos y del conocimiento. Así, cobró auge la atención sobre los sistemas expertos.

Parte del trabajo en la inteligencia artificial recayó sobre la elaboración de lenguajes para la representación precisa de la información, ellos se denominaron "*knowledge representation languages*" (lenguajes de representación del conocimiento). La idea era utilizar una notación lógica para representar hechos fácticos, de manera tal que pudiera utilizarse y procesarse por los programas. La composición de los sistemas expertos consistió en una base de conocimientos y una máquina de deducción o de inferencia. La base de conocimientos proporcionaba hechos objetivos y reglas sobre el tema, mientras que la máquina de deducción, ofrecía la capacidad de razonamiento que permitía al sistema extraer conclusiones. En una clara analogía con el ser humano, donde la capacidad de un experto de emitir soluciones ante situaciones problemáticas, se encuentra determinada en mayor medida por su experiencia práctica (expresada en la base fáctica o de conocimientos) y por su capacidad de razonamiento (la máquina de deducción). Sin embargo el problema descansa en, hasta donde los sistemas expertos son, o eran capaces, de realizar verdaderamente esta analogía.

Mientras las propuestas del análisis lingüístico inteligente, concebidas en los años fundacionales de la inteligencia artificial, cayeron en la frustración, los sistemas expertos, que no concebían metas tan lejanas, parecieron factibles en vistas de la recuperación. "*La influencia de los sistemas expertos ha cambiado la investigación sobre recuperación de la información de un paradigma relacionado en gran medida con los algoritmos de recuperación, hacia uno en el cual los usuarios, la heurística de la recuperación y la interacción conocimiento humano y computadora constituyen temas claves*". 19 Aunque es discutible que fue la influencia de los sistemas expertos y no la del enfoque cognitivo la que determinó estas pautas, 7 existen claras convergencias, si se parte de que estos sistemas que emplean de alguna forma un modelo cognitivo de usuario, claramente siguen las pautas del enfoque cognitivo.

Sin embargo, desde la posición de los investigadores de la inteligencia artificial hace varias décadas atrás, hubiera sido ridículo predecir que para el año 2001 casi todas las compañías consagradas a las tecnologías de inteligencia artificial habían dejado de existir.

"Una proyección justa sobre el índice de crecimiento de la inteligencia artificial tal vez mostraría que

pueden pasar varias décadas aún, antes de que se logren desarrollar máquinas totalmente inteligentes (...) pero hay algo más que aprender además de la histórica enseñanza sobre que no todas las curvas experimentan un crecimiento exponencial sostenido (una enseñanza que el campo de la computación tiene todavía donde sostenerse). El problema de la inteligencia artificial no fue que se llegara a la meseta de nuestra habilidad para ejecutar millones de LIPS (inferencias lógicas por segundo, ing. Logical Inferences per second), o de crear nuevos algoritmos, el problema radica en los cimientos, donde los pioneros del campo concibieron la inteligencia” . 20

Y este señalamiento parte de lo apuntado por el filósofo *Hubert Dreyfus* en los primeros días de la computación, donde señalaba que el esfuerzo principal de la inteligencia artificial descansaba, en la visión de apreciar la inteligencia humana, en la acción de asumir implícitamente que toda inteligencia podía ser producirse por mecanismos inherentes, como la manipulación lógica consciente de hechos representados mediante el lenguaje. 20

PERSPECTIVA GENERAL DE LA RECUPERACIÓN DE INFORMACIÓN

Hasta entonces, encontramos más de una manifestación concreta que guía a la investigación sobre recuperación de la información, por lo que esquemáticamente podría describirse desde varios puntos de vista. Sin embargo, parece muy adecuado en la perspectiva a nuestros propósitos, la visión general que ofrecen *Wormell e Ingwersen* para esquematizar las cuestiones elementales alrededor de las que transita la recuperación (fig. 2).

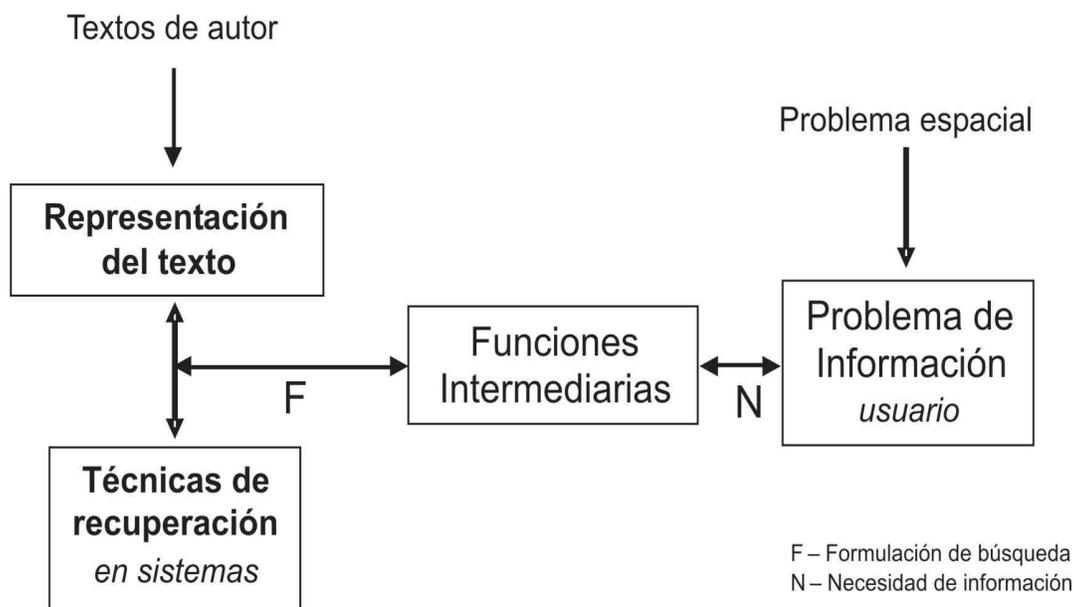


Fig. 2. Recuperación de la información

De un lado, se encuentran los documentos, o textos originales como productos directos de la creación de un autor. Esta información, contenida en el caso más típico, en textos aislados o en documentos propios, es aquella fuente con la que se supone el usuario rellene su 'laguna de información'. Los sistemas se basan, a su vez, en una forma de representación de ellos para su posterior recuperación, realizan esta instancia mediante diferentes técnicas o métodos, y son quienes de modo general hacen concreta la divergencia de un sistema a otro. Paralelamente, es una labor que tiene como punto de entrada, una solicitud de búsqueda, que en el caso 'tradicional' no hace más que decir que el usuario existe: “ *En los Proceedings de las conferencias sobre las bibliotecas digitales de la ACM en los años 97-98, de 25-29 artículos, sólo 3-4 mencionaban a las personas*”. 5

El centro del modelo describe todas aquellas funciones que median entre el sistema y el usuario, y las cuestiones envueltas en la dinámica de la transformación de su estado de carencia hacia una expresión de búsqueda concreta, independientemente de los métodos o técnicas que se utilicen para ello. Hacia la derecha, se recogen todas las cuestiones propias del enfoque cognitivo, que pretenden agregar a su universo, más allá de la formulación propiamente dicha, determinadas variables cognitivas vistas desde el contexto social en que se mueve el usuario, y en las cuales se cree, está la llave de su estado de carencia. “*De forma abreviada, el objetivo de la investigación 'tradicional' sobre recuperación de la información se centra sólo en el desarrollo de la parte izquierda del esquema (...) El enfoque 'cognitivo' se desarrolla alrededor de una unificación holística que entiende las interacciones e inferencias de todos los componentes, lo que significa, todos los estados de conocimiento, envueltos en la totalidad del modelo*”. 21

Sin embargo, puede señalarse una tercera vertiente mucho más joven aún que las antes expuestas, sin traspasar las fronteras del enfoque cognitivo. Su raíz subyace en una apreciación 'no tan' profundamente cognitiva, o tal vez en la visión de un enfoque cognitivo más ortodoxo . A pesar de la divergencia terminológica que puede prevalecer en la literatura -donde en muchos casos aparece indistintamente como enfoque cognitivo en recuperación y recuperación centrada en el usuario, o deflexiones similares, puede destacarse la denominada *por Ingwersen* como recuperación orientada al usuario ('user-oriented' approach), y cuyos esfuerzos se centran en lo que podría representarse con el mismo esquema anterior, en una mirada del centro hacia la derecha, con un objetivo en un mejor entendimiento de la conducta el usuario, principalmente mediante métodos de interacción.

RECUPERACIÓN DE INFORMACIÓN Y GUERRA FRÍA

“ *La Segunda Guerra Mundial apenas había acabado, cuando la humanidad se precipitó en lo que sería razonable considerar, una tercera guerra mundial*”. 1

Todavía la bandera soviética, no se izaba sobre el Reichstag , cuando las dos grandes potencias concluyentes de la guerra, Estados Unidos y la Unión Soviética , cada una con su bloque aliado, parecían abalanzarse una sobre la otra; se iniciaba así un período de amenaza existencial mutua que

sumió a la humanidad en el temor por más de 40 años. La capacidad armamentística para entonces superaba por mucho la que hubiera podido tener cualquier nación antes de la guerra; particularmente un nuevo tipo de medio con fines destructivos sin precedentes, hizo su aparición en agosto de 1945, para demostrar al mundo que la simple intención podía reducir a cenizas, no sólo toda forma de gobierno, sino toda forma de existencia biológica: la bomba nuclear.

La bomba nuclear fue el fruto del mayor apoyo presupuestario que conociera hasta entonces la investigación científica; 2 ningún gobierno antes de la conflagración hubiera soñado dedicar ni siquiera una fracción de los dos mil millones de dólares que se dedicaron a la bomba atómica, sobre la base de un proyecto hipotético (*Hobsbawm, 1998*). Como se dijo, “*La guerra acabó de convencer a los gobiernos de que dedicar recursos inimaginables hasta entonces a la investigación científica era factible y esencial para el futuro*”, 1 para entonces la ciencia se había convertido por mucho, en la principal fuente de poder que dictaba el discurso político.

Pero si bien la nueva arma representaba el mayor peligro conocido, fue tal vez también el mayor freno que encontró la ambición a fin de esparcir sus hedores incontenibles sobre la humanidad. En realidad, durante la guerra fría la ambición por la supremacía no dejó de existir, como lo demuestra la retórica de *John F. Kennedy*, pero la resignación por la limitada utilidad de la fuerza como arma elemental de poder, estuvo más que nunca avalada. Así, cuatro años después de Hiroshima y nueve meses después en la bomba de hidrógeno, la Unión Soviética, que siguió de cerca los pasos de su principal oponente, encontró la ansiada paridad nuclear, y con ello, cobró constancia en la mente de los políticos guerreristas la idea de que la confrontación directa, representaba más que nada, el suicidio mutuo.

A partir de entonces, la guerra se trasladó esencialmente del plano armado al plano informacional. La burda esperanza de un contraataque dependía en realidad más del tiempo de respuesta, que de la plena disposición de los medios para contrarrestarlos, un simple pestañazo era suficiente para dejar maniatado al oponente. Era entonces de suponer que la guerra armada debía traducirse en un precepto objetivo, que de alguna forma respondiera no tanto a las “hambres”, apaciguadas por la resignación de que ‘el otro estaba ahí’, sino a la tranquilidad de los nervios y al posicionamiento estratégico en una carrera por la ventaja, ello no era otra cosa que la información.

No sólo los conocidos servicios secretos como la KGB y la CIA institucionalizaron una labor desmedida e intensa, los medios empleados para ello recibieron también una fuente casi ilimitada de presupuesto. La situación llegó a un punto extremo luego de que la URSS lanzara su primer *Sputnik* el 4 de octubre de 1957, la primera nave puesta en órbita por el hombre alrededor de la tierra. La conmoción en los Estados Unidos fue tan grande que suele llamársele “el shock del *Sputnik*”. Sin duda no era una exageración, el *Sputnik* llevó a máximas instancias el cuestionamiento sobre la real validez de los planes gubernamentales afines al desarrollo tecnológico y forzó un replanteamiento de sus propósitos y métodos. Por un momento, la incipiente ciencia soviética parecía tomar la delantera.

No sólo los problemas planteados por *Bush* salieron a la palestra -se descubrió cómo en varios casos los gastos se duplicaban al realizarse dos veces una misma investigación, tan solo porque una empresa

desconocía los resultados de la otra, ello distaba de un aprovechamiento económico y temporalmente perfecto-, también se evidenció el desconocimiento, tal vez por autosuficiencia, del estado real de la ciencia soviética. Así, el cúmulo de información no sólo provenía de la investigación científica en potencia, sino también de una creciente y acumulativa labor de inteligencia, cuyo objetivo podía resumirse en una lucha constante por la supremacía o la aniquilación. Pero de que la información era la llave de estos objetivos no había duda, como lo demostrara el famoso Weinberg Report de 1963 :

"La transferencia de la información es una parte inseparable de la investigación y el desarrollo. Todos aquellos que se preocupen por dichos procesos -científicos, ingenieros, entidades de investigación académicas e industriales, sociedades técnicas, agencias gubernamentales, entre otras- deben reconocer su responsabilidad por la transferencia de la información de la misma forma que aceptan su responsabilidad por la investigación y el desarrollo" (Weinberg A. Science, Government, and Information. 1963).

En este contexto, se desarrolló todo sentido y dirección, que guió a la actividad científica durante la guerra fría (1947-1991). Los puntos focales y su prioridad monetaria variaban en dependencia de las cambiantes necesidades, y del estado de tensión o distensión de la situación internacional. Pero, si como señalara el reporte *Weinberg*, todos aquellos involucrados en los procesos productivos de diversa índole, debían reconocer su responsabilidad por la transferencia de la información derivada de su actividad, el otro extremo sería lógico que recayera sobre los métodos de acceso, registro, y recuperación de la información relevante a instancias de estas necesidades emergentes. El principio debía radicar, en que los esfuerzos dedicados a la información serían vanos de ésta hacerse inutilizable y aquí es donde la recuperación de información pasó a tomar partido.

Los años sesenta fueron una década de oro para la recuperación de información y de ella surgieron grandes promesas. 22 Los experimentos Cranfield, guía y paradigma tanto experimental como explicativo, fueron así sufragados por el gobierno en un conocido reconocimiento sobre la inviabilidad de las técnicas existentes para ello. Se establecieron instituciones como la *NSF (Nacional Science Foundation)*, que sufragaban entre otras, gran parte de la investigación en recuperación de la información. Los experimentos de *Cyril Cleverdon* y *Gerard Salton* , vieron la luz para esta temprana época.

Pero, los estimulantes tecnológicos directos que influenciaron más la actividad de recuperación también fueron producto de la “paranoia de la guerra fría”, Internet, sin duda, es el fruto más palpable de ello. Paralelamente, también se desarrollaron con especial velocidad las tecnologías computacionales, principalmente las mejoras ingenieriles introducidas sobre los aspectos de hardware, que volatizaron el estadio de las capacidades de procesamiento y almacenamiento. Las computadoras no eran sólo una necesidad práctica importante, sino también un componente esencial de la carrera aeroespacial iniciada con el *Sputnik* .

CONSIDERACIONES FINALES

La guerra fría fue un catalizador de punta al desarrollo científico, sobre todo de aquellos aspectos que confluyeron temporalmente con las políticas gubernamentales. El apoyo a la ciencia y la investigación en el contexto de la guerra fría, hubiera lucido inviable y absurdo en una época no removida por los nervios de quienes *“como el secretario de estado para la marina el presidente Truman, James Forrestal, estaban lo bastante locos, médicamente hablando, como para suicidarse porque veían venir a los rusos desde la ventana de un hospital”* . 1 Muchos de los principios e innovaciones con carácter militar desde sus inicios, se consideraron desde otra perspectiva de aplicación al pasar del tiempo.

Así, la guerra fría culminó oficialmente con la desintegración de la URSS en 1991, pero dejó consigo resultados eminentemente útiles a instancias de la recuperación, Principalmente una red mundial de información cuyo universo lejos del propósito inicial de supervivencia a un ataque nuclear, se extendió a la vida cotidiana de las personas para aportar un estímulo ilimitado a la actividad científica habitada por el propósito general de hallar información puntual, en un universo comúnmente regido por el caos.

REFERENCIAS BIBLIOGRÁFICAS

1. Hobsbawm E. Historia del siglo XX. Buenos Aires: Grijalbo Mondadori, 1998.
2. Bernal, J. D. Historia social de la ciencia II: La ciencia en nuestro tiempo. La Habana : Ciencias Sociales, 1986.
3. Fernández Molina JC. De la documentación a la Information Science : Antecedentes, nacimiento y consolidación de la "Ciencia de la Información " en el mundo anglosajón. Boletín de la Asociación Andaluza de Bibliotecarios 1993;33(9): 41-60.
4. Bush V. As we may think. Atlantic Monthly 1945;176(1):101-8.
5. Saracevic T. Information Science. JASIST 1999; 50(12):1051-73.
6. Serrano R. Teoría de la comunicación: Epistemología y análisis de la referencia. La Habana : Pablo de la Torriente Brau , 1998.
7. Ellis D. The physical and cognitive paradigms in information retrieval research. Journal of Documentation 1992;48(1):45-64.
8. Cleverdon CW. Factors determining the performance of indexing systems. Cranfield: Cranfield College of Aeronautics, 1966.
9. Cleverdon CW. Design and evaluation of information systems. ARIST 1971;6:42-73.
10. Yaroshevski MG. La psicología del siglo XX. La Habana : Pueblo y Educación, 1983.
11. Watson JB. Psychology as the behaviorist views it . Psychological Review 1913;(20):158-77.
12. Linares R. La información a través del tiempo. ACIMED 2000;8(3):228-37. Disponible en: http://bvs.sld.cu/revistas/aci/vol8_3_00/aci09300.htm [Consultado: 15 de febrero del 2005].
13. Toffler A. El cambio de poder. Barcelona: Plaza & Janes, 1990.
14. De Vega M. Introducción a la psicología cognitiva. Madrid: Alianza Editorial, 1994.
15. Gray P. Psychology. New York : World Publishers, 1991.

16. Chomsky N. Syntactic structures. Gravenhage: Mouton & Co, 1957.
17. Belkin NJ. Progress in documentation: Information concepts for information science. Journal of Documentation 1978;(34):55-85.
18. Saracevic T. Changing agenda for information retrieval: Editorial. Information Processing and Management 1987;23(5):I-II.
19. Brooks HM. Expert systems and intelligent information retrieval. Information Processing and Management 1987;(23):367-82.
20. Winograd T. From computer machinery to interaction design. 1997. Disponible en: <http://pcd.stanford.edu/winograd/acm97.html> [Consultado: 1 de abril del 2005].
21. Ingwersen P. Information retrieval interaction. London : Taylor Graham Publishing, 1992.
22. Lesk M. Seven ages of information retrieval. 1996. IFLA. Disponible en: <http://www.ifla.org/udt/op/udtop5/udtop5.htm> [Consultado: 15 de febrero del 2005].

Recibido: 25 de octubre del 2005

Aprobado: 12 de noviembre del 2005.

[Licenciado en Bibliotecología y Ciencia de la Información. Casa de las Américas. Cuba.](#)

Lic. Alberto Camaraza Monserrate. Casa de las Américas. Calle 3ra y G, El Vedado. Ciudad de La Habana , Cuba. Correo electrónico: ycamaraza@infomed.sld.cu

Ficha de procesamiento

Clasificación: Artículo histórico.

Términos sugeridos para la indización

Según DeCS 1

ALMACENAMIENTO Y RECUPERACIÓN DE LA INFORMACIÓN.

INFORMATION STORAGE AND RETRIEVAL.

Según DeCI 2

RECUPERACIÓN DE LA INFORMACIÓN /historia; EPISTEMOLOGÍA.

INFORMATION RETRIEVAL/history; EPISTEMOLOGY.

1 BIREME. Descriptores en Ciencias de la Salud (DeCS). Sao Paulo: BIREME, 2004.

Disponible en: <http://decs.bvs.br/E/homepagee.htm>

2 Díaz del Campo S. Propuesta de términos para la indización en Ciencias de la Información. Descriptores en Ciencias de la Información (DeCI). Disponible en: <http://cis.sld.cu/E/tesauro.pdf>