

TEORIA DE LA INDIZACIÓN: NUEVOS PARÁMETROS DE INVESTIGACIÓN

Dr. Antonio García Gutiérrez *

RESUMO

GARCIA GUTIÉRREZ, A. *Teoría de la indización: nuevos parámetros de investigación. Trans-in-formação, 1(2), maio/ago. 1989.*

O autor define a indexação como um dos processos modulares da documentação. Ela está exposta não só à interdisciplinaridade no plano epistemológico, como também às manipulações ideológicas. A indexação chega até as áreas da linguística, semiótica, semântica, psicologia social e tecnologia, fato que torna necessário uma definição da própria perspectiva teórica. Rejeita-se o modelo tradicional ou "coincidente" da indexação, devido a que ele causa continuismo, isto é, significantes idênticos e gerais passíveis de um reconhecimento linear, fora do contexto. O autor propõe uma alternativa, a indexação vetorial a qual é de projeção conceitual e resultado de uma cuidadosa análise documental. Por não se produzir uma indexação baseada na linguística documental, capaz de interagir com discos óticos, full text, etc., corre-se o risco de separar irreversivelmente a tecnologia da documentação.

Unitermos: *Indexação – Análise documentária – Lingüística documental.*

La indización es un ejercicio hermenéutico inseparable de la capacidad humana de la congñición e inserto en la dialéctica que da vida al binomio, documentalmente interactivo, hombre-cultura. Esta operación está presente en el acceso cotidiano a los mensajes, a las informaciones que circulan en torno al ser humano, por lo que debe destacarse su dimensión psicosocial. En realidad, el acto indizador es lo que activa la reacción y la respuesta en un proceso general de comunicación como un poderoso gatekeeper.

Podemos detectar la operación indizadora, desvinculada por el momento del ámbito consciente y utilitarista de la Documentación, en las actitudes selectivas, en primer lugar, y valorativa, a continuación, que definen al ser humano en el canal receptivo. Pero acaso, el cerebro del emisor también se impregna y activa merced a datos y hechos que lo

* Professor Titular de Documentación – Facultad de Ciencias de la Información, Universidad Complutense de Madrid.

in-forman y, en consecuencia, lo convierten en efmero receptor conmutado en emisor por impulsos autoinducidos (autoindizados).

No pertenecen a la Documentología las conexiones de una actividad propiamente documentológica, como la indización, con los procesos psicocríticos y, por tanto, no es su deber definir estas relaciones o identificar cómo y cuándo actúan, aunque sí es importante para el documentólogo tomar conciencia de que la indización, y el amplio corpus de estudio que comporta, es un fenómeno ligado a las capacidades cognitiva, intelectual (inferencial y heurística) y memorística (acumulativa) humanas y que, en consecuencia, la propuesta de un método unívoco está sujeta a constantes fluctuaciones debidas, precisamente, a las variaciones que acontecen en la psicología humana. A pesar de ello, y aún pareciendo paradójico, la indización es básicamente una operación inviable fuera de un marco estrictamente normativo impuesto por la objetivación del análisis efectuado por colectivos para el aprovechamiento de colectivos o por el sistema de algoritmos de una computadora.

La indización no descarta, en parte, la aproximación semiótica. En su mismo lexema lleva incorporada la indización el concepto de indicio. En su interacción con la realidad, el ser humano busca vestigios que le permitan deducir o actuar por lo que indización y toma de decisión son procesos cercanos y complementarios. Pero el hombre que escruta su entorno para extraer indicios no realiza tal actividad sistemáticamente, ni las captaciones son uniformes, ni los resultados son asimilables, ni siquiera la casuística de cada acto es homologable. En el acercamiento instintivo del ser humano a los indicios estriba la primera diferencia con respecto a la indización, desde la perspectiva documentológica, que presupone una actitud científica: el instinto se transforma en imaginación instintiva y la aproximación desordenada deviene método científico.

Otras dos diferencias me parecen básicas: la localización de indicios (humo significa fuego, las huellas indican el paso de un ser vivo, etc.) no se dirige a objetos concretos (documentos) sino a situaciones más amplias que incluyen los procesos y las condiciones de producción y sus efectos. Por ello, los corpus de trabajo de la indización general, semiótica y documental van sustancialmente siendo universales en los contenidos pero restringidos por el soporte en el campo documentológico. Esta restricción ofrece la ventaja de poder aplicar los modelos de inventario, formalización y categorización del Análisis de Contenido en determinados campos del conocimiento que se hallan documentalizados (fragmentado el discurso por unidades físicas de contenido). Los avances de la Semiótica benefician notablemente a la indización la cual utiliza el filtro del pragmatismo para adaptar sus métodos y frenar unas aspiraciones por lo denotado y lo connotado que no se encuentran entre sus objetivos inmediatos. Podemos aseverar, en suma, que en la indización el interés por la información

prepondera sobre el interés por la ideología subyacente. Esta frontera, compleja de trazar, también se la plantean, como se sabe, muchos lexicólogos que investigan en el borde del lenguaje. Su definición exacta es vital, en Documentación, para planificar la formación de especialistas y configurar el alcance y las responsabilidades de sus tareas.

Los objetivos de la búsqueda de huellas y pistas del semiólogo, del periodista o del detective con respecto al documentalista difieren. Los primeros persiguen la observación, explicación o demostración de un fenómeno. El último desearía objetivar su acceso a las fuentes para derivar los resultados hacia otros usuarios pero durante un lapsus el propio documentalista se convierte en investigador, actividad que abandona inmediatamente tras conocer el mensaje porque su método persigue reducir o simplificar grandes masas de información, urdidas en macroestructuras, en una exigua lista de conceptos o sistema mosaico que representa el discurso general soportados en los documentos. Lista lo suficientemente precisa como para poder representar exhaustiva y específicamente cualquier enunciado. Tal síntesis conceptual, insisto, es sólo una etapa intermedia, su canal de expresión, porque la finalidad última del documentalista es la comunicación diferida. Al pues, método de aproximación, corpus de acción, sistema de representación y comunicación masiva son los pilares sobre los que sitúa la Documentología uno de sus fundamentales ámbitos operativos.

1. La indización, distintivo documentológico

Observamos como la indización, siendo condicionante esencial del objeto de estudio de la Documentación, se sitúa en una de las puertas externas de la misma disciplina, esto es, se erige como una zona expuesta no solamente a la articulación interdisciplinar, en el plano epistémico, sino también a la contaminación y manipulación de la ideología (de los ejes subyacentes y del metatexto), en el plano pragmático o profesional.

En la consideración de lo que denomino Análisis documental de contenido (para excluir, en consecuencia, todas las operaciones externas erróneamente fusionadas en los planes de estudio), brillará la indización como operación medular, quizá única, con la transformación del proceso analítico que comportará la irrupción de tecnologías inteligentes en la gestión del conocimiento. En efecto, junto al deslizamiento de los métodos y valoración e los productos documentales utilizados hoy día ha de surgir una alternativa de referencia que centralice las tradicionales palabras claves y los resúmenes en una misma operación. Tal operación debe configurarse como una proceso de captación textual dentro del discurso general de la Ciencia en cuestión mediante mecanismos que produzcan una traducción reductora, es decir, simplificadora del original dando lugar a enunciados

sintéticos de representación. La indización bien puede seguir representando ese rol dentro del análisis documental aunque sus modos hayan de sufrir una adecuación.

Precisamente por ser la indización la zona de mayor riesgo en Documentación, recuérdese que trasciende en los niveles lingüísticos, semántico, psicosocial, semiótico, tecnológico, etc., del mensaje documentalizado, nos parece el más claro marcador de los límites del corpus teórico de la Documentología. Allí donde la Documentología interacciona con la Lingüística, la Lógica, la Teoría de la Información o la Bibliotecología, la indización aparece como concepto-guía del campo específico y exclusivo de la Documentación. Por tanto, definir adecuadamente la indización desde una perspectiva teórica supone el mayor paso para la consolidación de la Ciencia de la Información/Documentación en la que se inscribe como principal valedora.

A pesar de que la nueva metodología de la indización que propugnamos teóricamente para el próximo decenio se basa en categorías distintas a las establecidas por la indización coordinada en los años cincuenta, no cabe duda de que fue precisamente el empeño de algunos investigadores de la información, llamémosles con términos actuales documentólogos, en buscar esquemas de representación más flexibles y naturalizados (más cercanos de la lengua que de la nomenclatura), lo que dió pábulo a la irrupción documentológica entre los que hace tres décadas dudaban entre Bibliotecología y Documentación. La indización, y su connotación de análisis de contenido, suponía la entrada en este terreno de lingüistas, algunos semiólogos y tecnólogos ante la mirada impávida de los documentalistas (personas muy profesionalizadas pero desvinculadas de la observación teórica de sus propios problemas) siendo una de las principales causas del despegue de la Documentación las aportaciones de investigadores curiosos que habían tomado la indización como campo de pruebas de otras disciplinas, i.e. Lingüística aplicada, Estadística informática, Cálculo de Probabilidades, Teoría de la Comunicación entre otras.

A mi juicio, fue también la indización lo que aceleró la entrada del término y del concepto "información" en el seno de los estudios y actividades de organización documental. Con la perspectiva informativa, la Documentación sufre un nuevo giro, escorando hacia la Comunicología, ámbito epistemológico en el que se acomoda perfectamente, la cual introduce nuevas y decisivas aportaciones como la teoría de modelos comunicativos (provenientes, a su vez, de la teoría matemática de la información), los procesos, actores, canales, señal, códigos que intervienen también en la comunicación mediante documentos. Todo ello no sólo refuerza la presencia del lenguaje y del medio en Documentación sino también el papel de la propia indización como operación de flujo informativo en una y otra dirección. Y es este rol regulador el que imprime mayor

carácter socio-comunicativo a la Documentación. En cualquier caso, la indización no es, aunque implica, comunicación ya que prevalece en ella una operación intelectual sintética y estilística previa a la formulación de interrogantes y respuestas.

En consecuencia, la aparición del informador científico, técnico o especializado en general es fruto directo de la precisión de las técnicas indizadoras como filtro entre la masa de conocimientos y unos usuarios abrumados por la misma. La intensidad y el nivel de indización es lo que determina, hoy día, la especialidad y la profundidad tanto de los analistas como de las memorias documentales.

II. Fracaso del método coincidente

Como ha sido apuntado, y a la luz de la evolución tecnológica, es necesario observar la enfermedad que aqueja a los métodos de indización actuales e intuir las líneas que debe seguir la investigación teórica tendentes a la búsqueda y propuestas de nuevos métodos.

A mi modo de ver, desde la indización libre a la controlada, preo coordinada, ya usen encabezamientos de materia, unitérminos o descriptores métodos, todos ellos, que ofrecen un indudable perfeccionamiento cuanto más se vertebra el análisis de contenido, aún no se ha abordado el principal problema, causante del inmovilismo aparente de la teoría de la indización: uso del "método coincidente" como referente incuestionable en todas las alternativas citadas, lo que por un lado provoca continuismo y no reforma en las investigaciones y, por otro, frena la introducción de elementos morfosintácticos imprescindibles para completar los avances semánticos.

El método coincidente consiste en basar el éxito de la comunicación documental en el uso de etiquetas, formas o significantes idénticos, que han de coincidir, reconocerse linealmente, en el día logo diferido que mantienen un emisor/intermediario remoto y un receptor/usuario a través de un canal y un medio tecnológico. El principio de la coincidencia debe ser formulado en el marco general de la Lingüística documental y dirigido hacia los significados, los contenidos o los mensajes, más que a las formas de representación conceptual. Esta afirmación categórica, contra el anacronismo en que se envuelven los lenguajes documentales, se fundamenta en la propia teleología de la Lingüística documental y se ajusta al espíritu de coherencia que domina en la misma disciplina: la "naturalización" de los procesos indizadores y de los léxicos documentales. Tras el abandono de los encorsetados códigos alfanuméricos por parte de los primeros thesaurus y la adopción de términos de la lengua natural se abra un horizonte que no ha sido suficientemente aprovechado. En realidad tal cambio no era más que una conversión estrictamente facial: mutar conceptos en códigos o palabras no entraña una

profunda naturalización de un sistema nomenclator sino tan sólo la transformación de códigos-alfanúmeros en códigos-vocablos. Un thesaurus podría conmutar los descriptores en números y combinarlos con el mismo resultado. Mientras ello sea posible, tendremos que hablar de inventarios y no de lenguajes, de etiquetado y no de indización (tal vez protoindización).

Piénsese en dos hombres que necesitan, para comunicarse, una coincidencia de formas: los sinónimos no recogidos, los grafemas o fonemas mal empleados, las polisemias o las metáforas no previstas producirían graves distorsiones en el diálogo, máxime si éste se compone de unos pocos conceptos no redundantes y se produce en unos segundos. Si la comunicación entre dos individuos es posible mediante la lengua es porque los significantes que utilizan representan conceptos que se son redefinibles por otros significantes, es decir, no tienen la exclusividad de tal representación; otros significantes pueden ajustar el concepto, otros conceptos pueden explicarlo. Ello permite la deducción y la lectura "entre líneas" parámetros exigibles a la tecnología superior. Hacia este método de coincidencia semántica deben tender los principios de indización y tal avance supone una mayor interactividad de usuario/sistema/productor pero todo ello comporta una gran complejidad en la preparación del eslabón intermedio, el sistema. En efecto, no debemos olvidar que la comunicación documental es una comunicación mediatizada y convencionalizada tanto en el sentido tecnológico como ideológico (gatekeeper). En el plano estrictamente técnico se requiere un doble esfuerzo en la elaboración del sistema con el fin de hacer posible, a través de una computadora inteligente, una mayor libertad del receptor en su interrogatorio manteniéndose la lengua natural en superficie tanto en el nivel léxico o de vocabulario como en el combinatorio o morfosintáctico.

Si la principal condición de la eficacia del sistema es la satisfacción del usuario tras la comunicación establecida, la indización debe focalizar su objeto de interés en la obtención de un método y un producto que permitan nuevas formas y filosofía de la comunicación documental de la misma manera que la Informática ha supuesto no sólo un avance en la cantidad y rapidez de transmisión de datos sino también en los comportamientos psico-sociales de los científicos usuarios abriendo nuevas perspectivas para la investigación.

III. Indización y representación

La función aprehendedora de mensajes de la indización es la parcela más conectada con las disciplinas mencionadas aún sin el trasfondo que prevalece en los objetivos de las mismas. La vertiente puramente documentológica de la indización es la que concierne al sistema de representación del que se desarrolla la potencialidad de comunicación.

Tras la indización por palabras aislada surgió la indización por conceptos (descriptores) problemática desde sus inicios por la dificultad de establecer los límites sémicos de tales conceptos con reglas no válidas para todos los casos. La improvisación y un exceso de convencionalismo se apoderaron de la articulación conceptual a pesar de la racionalización del sistema mediante thesaurus. Mi propuesta, desarrollada en una reciente investigación (Estructura de la Documentación), se basa en la relación y podríamos denominarla indización vectorial o indización de proyección conceptual. Es la relación viva en un sistema de representación documental, concebido teóricamente, lo que devuelve su integridad al discurso. Un mosaico de descriptores estáticos no resuelve la dialéctica usuario/memoria ni aún contando con la existencia de referentes que prevean esas posibles extensiones relacionales de los conceptos. Sin embargo, los lenguajes documentales realizan esta función a priori, en laboratorio, estando lejos de la optimización real. Los conceptos desprendidos de miles de enunciados que forman los contenidos proposicionales de los documentos interaccionan entre sí, provengan o no de la misma fuente. Insisto, es la reintegración del discurso científico en el resultado del análisis documental.

Los operadores booleanos no colman las aspiraciones de los enunciados de complejidad media o alta y en cualquier caso nada tienen que ver con la gravitación conceptual ya que sólo funcionan en el nivel de expresión. Las necesarias normas sintácticas que regulan los enunciados de conceptos provenientes de la indización precisan una mayor versatilidad o transportabilidad de las representaciones (verbos, adjetivos, adverbios, etc.) facultad no recogida por la indización efectuada sobre base sustantiva.

La representación por sintagmas prefabricados o por atomización sémica es el ángulo sobre el que pueden girar las tendencias de la investigación. Coherentemente con nuestra propuesta de naturalización del proceso documental y al hilo del fracaso de los lenguajes documentales sintácticos de los años sesenta (Syntol, Semantic code) debido a su complejidad y dificultad de prever los mensajes naturales, opto por la "atomización sémica" de los conceptos y por la descentralización del descriptor como núcleo significante. La irrupción, en los sistemas de representación, de articulaciones léxicas, prefijos, sufijos e infijos independientes de los lexemas junto a todas las categorías léxicas y a un número básico de operadores lógico-sintácticos (sintactores) supone acceder a un nuevo y natural método de síntesis de los documentos.

El producto de la indización es, en el nivel de superficie o expresión el elemento esencial de comunicación. En Documentación, la comunicación debe producirse mediante un acto de atomización del mensaje en manos del analista y la recomposición del mismo mensaje por un usuario final situado en coordenadas cronoespaciales y socioculturales distintas. La existencia de un lenguaje referente o la posibilidad, en ausencia de aquél, de dialogar

mediante intuiciones es lo que aporta a la indización su caracter social aunque en el primer plano de actuación de los interlocutores (mensaje/atomización/expresión) podemos considerarla como elemento individual rigido por la psicolingüística del ser humano.

En los cuatro modelos comunicativos tradicionales, sean estos simultáneos, personales, diferidos o anónimos, la indización pertenece al que combina el anonimato y la transmisión diferida aunque, con la irrupción de la interactividad en las modernas bases de datos, la comunicación puede considerarse simultánea en cuanto al medio aunque diferida en cuanto al emisor que ejecutó el análisis. Es precisamente el carácter integrador de la indización entre las distintas disciplinas que conciernen al documento el factor esencial de contacto entre el productor y el consumidor de información.

Para construir un modelo integrado de indización y lenguaje deberán hacerse minuciosos estudios conceptuales y gramaticales sobre la estructura de determinados discursos para llegar a tal estado de la cuestión. Y estos avances, soportados necesariamente por sistemas expertos e inteligentes, repercutirán en el modo de aproximación analítica al mensaje original, en su transformación ponderada, en los mecanismos de sugerencia del sistema a potenciales receptores y en la estructura y configuración de los lenguajes documentales. Habremos pasado, entonces, de los esclerotizados procesos con escritores a una comunicación flexible entre individuo y Cultura mediante la auténtica interactividad del diálogo interpersonal bajo la presencia reguladora y referencial del discurso.

IV. Evolución de la indización automática

El conocido test de Cranfield llegó a conclusiones muy ilustrativas en lo que se refiere al papel preponderante de la computadora en el análisis documental de textos. Sin embargo, aquel trabajo fue realizado en los albores de la irrupción de la Informática en la gestión de documentos a principio de los sesenta. Lo que, obviamente, no se podía prever era la evolución vertiginosa que fueron adquiriendo los programas de indización automática en los siguientes veinte años, evolución que, en mi opinión, ya ha llegado a culminar los límites de la Informática tradicional. La necesidad de mutar los soportes computacionales usuales sugiere, también, un cambio en la filosofía de las perspectivas de estudios teóricos y metodológicos en indización.

El método frecuencial aparece, en primer lugar, destinado a textos breves escritos en lenguaje unívoco, clásico de laboratorios experimentales, basándose en el recuento convencional de las formas. Según códigos previos, los significantes repetidos varias veces en un mismo párrafo, página, documento o colección de documentos son extraídos como

descriptores. Sin embargo, la gran cantidad de vocabulario ordinario que frecuentan los textos científicos puso pronto en evidencia el primer defecto del recuento: muchas repeticiones eran meramente lingüísticas o coyunturales, válidas en el nivel de enunciado pero sin representación alguna del contenido.

Este problema de orden lingüístico hizo surgir lo que podemos denominar los métodos indizadores basados en el lenguaje como referente. Según el grado de incidencia lingüística en la acción indizadora del ordenador distinguiremos la simple lista de stop terms que no alcanza unos pocos centenares de términos vacíos y que aparece incorporada a ciertos tipos de software de gestión documental y word processors. De esta manera, adverbios, preposiciones, conjunciones y otros elementos desprovistos de significado propio en los enunciados no son extraídos en el recuento frecuencial con una ventajosa reducción de términos pero también con una importante pérdida de información y de sentido.

A pesar de este indiscutible aunque tímido avance, que ya supone una aproximación lingüística a la indización automática, decenas de palabras desprovistas de interés, si bien significativas, seguran siendo consideradas como claves por el ordenador. Esta laguna fue pronto cubierta por los antidiccionarios (para algunos llamados, erróneamente, thesaurus negativos): vocabularios de palabras no interesantes que adicionaba a la lista de stop terms cualquier categoría léxica significativa (sustantivo, adjetivo o verbo) cuyo concepto es, con toda seguridad, irrelevante para el centro de documentación. Exhaustivos listados fueron introducidos en las memorias de las computadoras con objeto de que los softwares comparasen, en una tosca lectura lineal, el significante del texto indizado y los términos del antidiccionario. Desde luego, solamente los ausentes en el listado podrían ser considerados descriptores. El fracaso fundamental de este método radica en la desconexión del usuario con las palabras llenas, con los términos extraídos por la computadora, de manera que la búsqueda se establece sobre la "adivinanza" on line: / aumento de precios/ = / subida de precios/ = /inflación/, etc.

La aportación del último método lingüístico que voy a reseñar, el que incorpora un thesaurus, venía a solucionar la indicada lacra: con un thesaurus, el documentalista y el usuario final poseen un instrumento puente que les permite elaborar estrategias de búsqueda con términos conocidos. La posibilidad de un mítico diálogo con la computadora se restablece porque empleamos rúbricas familiares para el software, a pesar de la consolidación de la ideología que supone el thesaurus con respecto al azar de la indización libre.

Acumulando los recursos descritos hasta aquí, puede pensarse que la compensación a varias décadas de investigaciones está cercana. Nada más lejos de la realidad. Hemos debido llegar a esta situación para constatar que

los procedimientos de perfeccionamiento sobre las mismas bases teóricas se han agotado. Las razones se desprenden de los dos problemas apuntados a lo largo de esta reflexión que aún no han sido abordados, al menos, en niveles pragmáticos o de aplicación:

- * el método de la coincidencia formal en la indización es insuficiente para sostener un proceso comunicativo interactivo complejo.

- * la indización automática se basa en la lectura lineal, método que ha de ser rechazado. Las investigaciones en la materia deben versar sobre la lectura automática contextual.

Precisamente los métodos que apuntamos seguidamente han adoptado, teóricamente, estas consideraciones. La indización ponderada incorpora códigos de valoración del descriptor en el contexto. Tales códigos, integrados por el documentalista sobre términos extraídos por la computadora dan un peso específico al concepto en el contexto. No obstante, mientras esta labor deba ser realizada por el ser humano no podemos referirnos con plenitud a indización automática integral.

En el momento en que la propia máquina sea capaz de seleccionar términos no por su presencia en el texto sino por su valor contextual y pueda denominar conceptos y temas con significantes ausentes del texto estaremos inmersos en la indización contextual. Sin rechazar los métodos anteriores, por el contrario los adscribe a sus capacidades, la indización contextual se basa en el mantenimiento de una memoria con cientos de miles de enunciados acumulados y revisados y un mecanismo inteligente de inferencia. Descartando textos excesivamente complejos y, por supuesto, literarios, una inteligencia artificial podría aproximar su proceso indizador, razonablemente, al realizado por el ser humano. Algunos éxitos ya han sido logrados en Ciencias experimentales y aplicadas y en un campo más cercano a la Documentación, la traducción automática de textos científicos y técnicos.

La incorporación de la indización en el objeto de aplicación de las inteligencias artificiales repercutiría positivamente en las empresas de bases y bancos de datos, generalmente escasas de expertos en documentación, las cuales podrían destinar al personal liberado, gracias a los nuevos recursos del sistema inteligente, a tareas menos rutinarias. El producto de la indización contextual no podría ser considerado como tradicional: secuencias de palabras claves, resúmenes, extractos, etc. sino más bien como la esencia del documento, en solitario o acompañando al full-text, que he denominado en otros trabajos "síntesis contextual": fragmento del discurso general de una Ciencia representativo de una unidad física portadora de una porción del mismo discurso (documento). Por tanto, el nuevo producto documental no restringe su información a un documento exclusivamente ya que está impregnado por el contexto en el que el mismo documento se halla inserto.

Las memorias masivas de tecnología óptica abarantan fulminantemente las discutidas memorias magnéticas por lo que el full-text no debe ser un gran inconveniente para los centros de documentación del siglo XXI. Pero la solución de las dos incapacidades que afectan a la esencia de la Documentología, insuficiencia del análisis documental actual y dificultades para el acceso al original, pasa por la prevención teórica y la participación de documentólogos en los equipos interdisciplinarios que trabajan, hace más de una década, en las relaciones de la tecnología punta con el análisis de contenido. Esta misma precaución ha de someter a discusión los actuales métodos de evaluación y operatividad de los sistemas de indización. De no ser así, los documentalistas verán impuesta una pesada herencia de ingenieros y científicos experimentales, más preocupados por la transmisión que por el contenido, que afectará una vez más a los campos social y humanístico, produciéndose una profunda y acaso irreversible disociación entre Tecnología y Documentación.

V. Consideración final

La teoría de la Indización debe establecer sus parámetros de investigación actuales sobre la aproximación semántica al texto/discurso redefiniendo los módulos de comportamiento de los usuarios. La nueva metodología de la indización que propugno estará teóricamente condicionada por los siguientes supuestos:

1º Transformación de los principios de la indización tradicional asentados sobre la representación formal de los conceptos y la comunicación basada en la coincidencia de esas mismas formas. Incluimos en la expresión tradicional, los métodos de indización formal por computadora.

2º Disponibilidad de recursos tecnológicos ad hoc, tanto en lo que respecta al hardware como a los lógicos, capaces de asimilar las propuestas teóricas que suplanten la actual fuerza del significante por un mayor protagonismo del significado.

3º Propuesta de una nueva estructura de lenguaje combinatorio que actúe de marco referente de la indización.

4º Establecimiento de un cuadro especulativo de tipos y formas de interrogantes de usuario y respuestas plausibles por parte del sistema.

5º Creación de un nuevo producto referencial que limine las fronteras descriptor/resumen (y sus funciones respectivas recuperación/información) unificando los objetivos y considerando las ventajas de las tecnologías ópticas para conserva y difusión de originales.

BIBLIOGRAFIA

- BARDIN, Laurence: **El análisis de contenido**. Madrid: Akal, 1986, 183 p.
- BELKIN, N.; SEEGER, T. and WERSIG, G.: Distributed expert problem treatment as a model for information system analysis and design. **Journal of Information Science**, nº 5, 1983, p. 153-167
- BONVILLE, Jean de: Application du paradigme de Shannon à la Bibliotéconomie et à la Documentation. **Revue canadienne des Sciences de l'Information**, vol. 3, mai 1978, p. 13-27
- BROOKS, H.M.: Expert Systems and Intelligent Information Retrieval. **Information Processing & Management**, vol. 24, nº 4, 1978, p. 367-382
- CROFT, V.A. and THOMPSON, R.H.: I3R: A New Approach to the Design of Document Retrieval Systems. **Journal of the ASIS**, nov. 1987, p. 389-404
- CHIORNII, A.I.: **Introducción a la teoría de la búsqueda de información**. La Habana: Instituto de Documentación e Información científica y técnica, 1975, 332 p.
- DIJK, Teun Van: **Texto y contexto**. Semántica y pragmática del discurso. Madrid: Cátedra, 1980
- GARCIA GUTIERREZ, Antonio: **Lingüística documental**. Barcelona: Mitre, 1984, 279 p.
- IDEM: **Estructura lingüística de la Documentación**. Teoría y método. Madrid, 1989 (en prensa)
- INGWERSEN, Peter: Towards a New Research in Information Retrieval. In **Knowledge Engineering: Expert Systems and Information Retrieval** / ed. Irene Wormell. London: Taylor Graham, 1987, p. 151-169.
- LOPEZ YEPES, José; SAGREDO, Félix et al.: **Estudios de Documentación general e informativa**. Madrid: Seminario Millares Carlo y UNED de las Palmas, 1981
- PECHEUX, Michel: **Hacia el análisis automático del discurso**. Madrid: Gredos, 1978, 374 p.
- SALTON, G. & MCGILL, J.M.: **Introduction to modern information retrieval**. New York: Mc Graw-Hill, 1983
- SMIT, J. (coord.): **Análise documentária: a análise da síntese**. Brasília: IBICT, 1987, 133 p.

ABSTRACT

GARCIA GUTIERREZ, A. *Theory of indexing: new trends and approaches*. **Trans-in-formação**, 1(2), maio/ago. 1989.

The author sets indexing as a vital process and research field in Documentology. Theory of indexing does not only grow from interdisciplinarity at the epistemological level but it is also involved in deep ideological bias. Content analysis in Documentation must be

considered from different points of view: Linguistics, Semiotics, Semantics, Psychology or Technology are important fields which feed the theoretical conception of indexing. Traditional methods of indexing based on "coincidence" of significant must be revised having in mind the poor results obtained from matching, disregarding both concepts and speech context. A review on automatic indexing is given: Statistical and linguistic approaches are not sufficient to meet information growing and complexity. Vectorial indexing based in non-traditional documentary languages (using vectors as main elements of concept structure) is proposed as a theoretical position for current research on the topic.

Recebido em 24.04.89.