



LOS TESAUROS CONCEPTUALES COMO HERRAMIENTA DE PRECISIÓN EN LOS SISTEMAS DE ORGANIZACIÓN CIENTÍFICA*

Miguel-Ángel López Alonso**

.....

RESUMEN

A partir de la aplicación del modelo conceptual a la recuperación de la información, se analiza su estructuración mediante los tesauros conceptuales, como elemento de precisión en las recuperaciones con lenguaje natural, y su integración con las reglas de la base de conocimientos de un agente experto. Se explica además, la aplicación de la tecnología hipertextual al diseño y desarrollo de estos agentes especializados en la organización de la información científica.

Palabras clave: RECUPERACIÓN DE INFORMACIÓN
TESAUROS CONCEPTUALES
TECNOLOGÍA HIPERTEXTUAL

.....

INTRODUCCIÓN

Se detecta una tendencia creciente, apoyada en el modelo conceptual, para proporcionar al usuario soluciones de acceso a la recuperación de la información, que mejoren las preguntas iniciales, mediante sesiones de búsqueda en las que éste comparta sus propios conocimientos con los existentes en las terminologías a su

* Artículo derivado de la investigación: *Aplicación del Modelo Conceptual a la Recuperación de la Información*. Madrid: Universidad Carlos III, 1998. Recibido en mayo de 1999

** Prof. Universidad Carlos III de Madrid – España. Doctor en Documentación. Vallehermoso, 55 – 28015, Madrid – España. Email: cuevaslo@bib.uc3m.es.



alcance. *En las nuevas aplicaciones de las industrias de la lengua se hace uso de los tesauros conceptuales como elemento de precisión para las búsquedas en lenguaje natural, dentro del vocabulario mixto usado en los sistemas de recuperación de la información.*

En una primera fase del procesamiento del lenguaje natural¹, la lingüística automatizada desarrolla diferentes analizadores morfológicos, sintácticos o semánticos, para evitar la análisis de las palabras y de sus relaciones².

El análisis por el significante, característico de los analizadores morfosintácticos, obliga a descomponer los descriptores complejos en sus componentes. La utilización de los analizadores semánticos implica el análisis de las palabras para su descomposición en el mayor número de ideas posible (ej.: "retroalimentación", se forma a partir de dos términos simples, alimentación y retro = volver a).

En una segunda fase, la terminología reúne los conceptos obtenidos del análisis contextual de los documentos científicos de las distintas ramas del conocimiento, para la creación de bancos de datos terminológicos.

Finalmente, se utilizan los procedimientos de la inteligencia artificial en la rama del procesamiento del lenguaje natural³ para facilitar la creación de macrodiccionarios electrónicos o bases de conocimientos terminológicos, que permitan búsquedas automatizadas en diferentes bases multilingües, en línea, trabajando cooperativamente a través de la red Internet.

La *normalización del lenguaje* puede aplicarse fácilmente al conocimiento codificado de los vocabularios controlados, pero, limitada sólo a pequeños grupos de usuarios (ej.: en la morfología restringida de las nomenclaturas especializadas, en

-
1. Lo que se ha dado en llamar «Modelización de los Procesos Lingüísticos Naturales».
 2. Conociendo los complicados mecanismos del razonamiento humano se podrá resolver la ambigüedad léxica u oracional, profundamente difíciles de programar.
 3. Mediante el *agente experto*, dotado de un «motor de inferencia» que controla cómo y cuándo deberá aplicarse la información recogida (en las dos primeras fases de la modelización), al corpus lingüístico de una lengua viva.



la sintaxis especial de las patentes o de los contratos legales, etc.). Puesto que muchos textos técnicos se almacenan en soporte digital, se les puede convertir a un formato adecuado para su análisis terminológico con las técnicas desarrolladas por la lingüística automatizada. Si se les analiza y compara con los vocabularios electrónicos, almacenados en los bancos terminológicos, se pueden obtener listados de elementos nuevos, erróneos o necesitados de equivalente en otras lenguas.

En el *diseño de los sistemas integrados de gestión de la información*, los documentalistas deben fijarse como meta la investigación pormenorizada de la base de conocimientos del sistema, dejando el diseño del motor de inferencia y de la interfase de usuario a los ingenieros del conocimiento. La habilidad del sistema para aplicar la heurística de los distintos profesionales científicos al proceso de recuperación de la información, proporciona respuestas a las siguientes preguntas de los usuarios:

- ¿Cuándo usar un tesoro?
- ¿dónde empezar a usarlo?
- ¿qué relaciones seguir?
- ¿qué nuevos conceptos seleccionar?
- ¿cómo incluir los nuevos descriptores en la pregunta?

Esto les permitirá ampliar o restringir la búsqueda, mediante interacciones con el sistema, del tipo:

- Use términos más generales y términos relacionados para un más amplio tratamiento de la búsqueda.
- Muévase más arriba o más abajo en la jerarquía del tesoro para modificar la especialización.
- Use sólo vocabulario controlado para una mayor especificidad en la búsqueda.⁴

4. JONES, Susan et al. Interactive Thesaurus Navigation: Intelligence Rules Ok?. *Journal of the American Society for Information Science*, 1995, 46(1), p. 58-59.



Como punto de partida para el diseño de sistemas perfeccionados, se considera imprescindible la fusión de las teorías de la recuperación de la información (IR) y de la organización del conocimiento (KO) que aporta a la primera sus potentes estructuras de conocimientos.

Se han establecido como postulados mínimos para su integración:

- 1° Que la recuperación de la información, en general, no sea una actividad aislada sino que se integre con la indización previa y esto les permita mantener un alto nivel de correspondencia⁵, y
- 2° que la recuperación de la información científica, en particular, se produzca en el marco del trabajo habitual del profesional y de sus habituales sistemas informáticos de tratamiento documental (ej.: archivos documentales, procesadores de texto, correo electrónico, etc.).

LOS VOCABULARIOS CONTROLADOS COMO SUPERACIÓN CONCEPTUAL DE LOS SISTEMAS DE CLASIFICACIÓN

Los sistemas de clasificación de estructura enumerativa aparecieron cuando predominaba la preocupación por clasificar el contenido de los documentos temática y físicamente, mediante su representación sintética. Con el posterior crecimiento exponencial de la información, la capacidad de estos vocabularios controlados para representar y recuperar el contenido de los documentos se desborda y obliga a realizar nuevas investigaciones.

Para intermediar en la comunicación entre la ciencia de la documentación y la lingüística documental, aparecen los lenguajes de estructura combinatoria o vocabularios controlados⁶, que tratan los documentos prescindiendo de los sistemas de

5. DOMINICH, S. Interaction Information Retrieval» *Journal of Documentation*, 1994, 50 (3), pp. 197-212.

6. MOREIRO GONZALEZ, J. A. De la Documentación a la Ciencia de la Información: evolución de los conceptos y aplicaciones documentales. Seminario de Humanidades Agustín Millares Carlo, separata Homenaje a Antonio de Bethencourt Massieu, 1995, p. 20.



clasificación universales y se especializan por áreas del conocimiento. La clasificación, como ayuda a la catalogación y a la recuperación por materias, se sustituye por la indización mediante descriptores tomados de un tipo de estos vocabularios, los tesauros. Los sistemas integrados de gestión de la información recurren a los descriptores de dichos tesauros específicos, para indizar los documentos por sus títulos o resúmenes y, tras la ponderación de los conceptos más repetidos, proceden a utilizarlos posteriormente como descriptores postcoordinados en la recuperación, mediante combinación con los operadores booleanos.

Los vocabularios controlados constituidos sobre lenguajes de estructura combinatoria buscan un acercamiento de los conceptos documentales y los lingüísticos que mejore el análisis y la indización de la información almacenada en las bases de datos documentales, para su exacta recuperación en los sistemas anteriores. Aunque los estudios de Markey, Atherton y Newton⁷ ya habían señalado que las recuperaciones con lenguaje natural producían mayor exhaustividad y menor precisión que las realizadas con un vocabulario controlado, será Boyce quien afirme:

«Dado que cada documento recuperado mediante una búsqueda, a partir de un término descriptor, es uno de los documentos recuperados por otra búsqueda basada en sus términos componentes en el índice inverso primitivo, será imposible que la primera búsqueda, con vocabulario controlado, recupere más documentos que la segunda con lenguaje libre», lo que, en consecuencia le induce a afirmar:

«el vocabulario controlado, usado en los grandes sistemas comerciales de gestión de la Información, constituye un elemento de precisión en virtud de la propia estructura por defecto de sus registros informáticos, y nunca será un elemento de exhaustividad».⁸

En esta línea, podemos afirmar que la menor exhaustividad y mayor precisión inherente a las recuperaciones realizadas con vocabularios controlados, depende-

7. MARKEY, K., ATHERTON, P. y NEWTON, C. An analysis of controlled vocabulary and free text search statements in online searches. *Online Review*, 1980, 4, pp. 225-236.

8. BOYCE, B. R. y McLAIN, J. P. Entry Point Depth and Online Searching using a controlled vocabulary. *Journal of ASIS*, 1989, 40 (4), p. 273.



rán preferentemente del hecho intrínseco que las caracteriza: "*ser una reducción en la cantidad de términos utilizados para la búsqueda*". Y, en segundo lugar, del tipo de vocabularios controlados utilizados (tesauros, clasificaciones, etc.), del número de términos asignados a cada descriptor, de sus capacidades para evitar homónimos, etc.⁹ Con su integración se logra mayor exhaustividad y precisión en las recuperaciones, al eliminar el «ruido documental» derivado de la utilización de los sistemas de clasificación tradicionales.

LA INCORPORACIÓN DE UN TESAURO CONCEPTUAL COMO ELEMENTO DE PRECISIÓN EN LAS RECUPERACIONES CON LENGUAJE NATURAL

Entre los procedimientos propuestos para evitar la falta de precisión del lenguaje natural, se destaca la incorporación de los conceptos obtenidos del análisis de contenido de los documentos en un tesauro conceptual que facilite las recuperaciones desde la interfase sistema - usuario. De esta forma se incrementa la variedad de las ecuaciones de búsqueda del usuario, o al menos le muestra la variedad de las existentes; lo que, de acuerdo con el *principio de variedad de requisitos* de Ashby¹⁰, incrementa simultáneamente la cantidad de respuestas del sistema consultado.

Los tesauros conceptuales diseñados para ayudar en la enunciación de las preguntas en la fase de recuperación de la información en grandes bases de datos, propuestos por autores como Bates¹¹, Lancaster¹² o Schmitz-Esser¹³, atenuan la indeterminación de las búsquedas en lenguaje natural, especialmente en aquellas

9. SOERGEL, D. Indexing and Retrieval Performance: The Logical Evidence. *Journal ASIS*, 45 (8), 1994, pp. 589-599.

10. ASHBY, W. R. *An Introduction to Cybernetics*. London: Methven, 1973, pp. 202-212.

11. BATES, M. J. «Subject Access in Online Catalogs: A Design Model. *Journal of the ASIS*, 1986, 37 (6), p. 361.

12. LANCASTER, F.W. *Vocabulary control for information retrieval*. Information Resource Press, 1986 (XVII ed), 270 p.

13. SCHMITZ-ESSER, W. New Approaches in Thesaurus Application. *International Classification*, 48 (3), 1991, pp. 142-147.



bases de datos cuyos documentos con texto completo no han sido previamente indizados con ningún otro tesoro.

Diversos experimentos con ecuaciones de búsqueda, en las que los términos del lenguaje natural de los usuarios se apoyan con términos adicionales extraídos de un tesoro conceptual¹⁴, han llegado a doblar la precisión en el número de documentos recuperados¹⁵. Otros, que tratan de medir los efectos de la ampliación de las búsquedas mediante este tipo de tesoros, obtienen los mejores resultados cuando los usuarios:

- a. pueden elegir de entre una gran cantidad de términos del campo específico tratado, sean suyos, de la base de datos o incluso de indizaciones previas, y
- b. controlan interactivamente el proceso de navegación por el tesoro, en vez de utilizar procedimientos automáticos¹⁶.

Metodología para la compilación automática de Tesoros Conceptuales

La modelización de los principios teóricos que presiden la estructura de los tesoros ha seguido preferentemente los modelos lingüísticos o matemáticos; sin embargo, las recientes teorías giran alrededor de la noción del motivo o materia de la que tratan los textos, es decir, del concepto semántico o cognitivo.

Maniez propugna (1976) un modelo de tesoro en el que las relaciones no sean lingüísticas: paradigmáticas (pertenecientes a la lengua, fuera de todo contexto), o sintagmáticas (pertenecientes al discurso, integradas en su contexto), sino

14. CROFT, W. B. y DAS, R. Experiments with query acquisition and use in document retrieval systems. En: Proceedings of the 13th Conference on Research and Development in Information Retrieval, Brussels, Belgium, 9/1990. KRISTENSEN, J. Expanded End-user's Query statements for free text searching with a search-aid thesaurus. *Information Processing & Management*, 1993, 29 (6), pp. 733-744.

15. EKMEKCIOGLU, F. C., ROBERTSON, A. M. Y WILLET, P. Effectiveness of query expansion in ranked-output document retrieval systems. *Journal of Information Science*, 1992, 18, pp.139-147.

16. *Ibid.* Cit. nº 4.



extrasemánticas o asociativas; de forma que aúnen términos y conceptos reales por su similitud de sentido, en el contexto específico del usuario¹⁷.

Dewèze formaliza (1981) la representación de las relaciones semánticas con la adopción de una teoría semántica extralexical que sitúa a un nivel superior al de los lenguajes naturales, y la perspectiva de construir tesauros multilingües. En esta teoría, un significado es definido como «un conjunto de semas a los que se pueden atribuir posteriormente relaciones lexicales en varios idiomas»¹⁸.

La red semántica conceptual facilita:

- el encaminamiento de las hojas hacia una raíz y viceversa, gracias a la función de orientación presentada por las clases de discriminantes o facetas,
- la transición de un árbol a otro atravesando los nudos polijerárquicos o relaciones asociativas,
- la extensión a un árbol completo, la restricción a un subárbol, o incluso la restricción a árboles parciales, o
- la exploración transversal de la red buscando configuraciones de una o varias facetas en una reunión o en una intersección de árboles.

A partir de este concepto de red semántica de Dewèze, Schaüble (1989) propone una nueva estructura de la información, el espacio conceptual. Y construye una teoría de los tesauros conceptuales, modelada como un sistema mediante la lógica matemática del dominio algebraico, que revela una estrecha relación entre los tesauros y el modelo espacial, y en la que las relaciones entre términos son definidas con más precisión que en los tesauros jerárquicos¹⁹.

17. MANIEZ, J. Los lenguajes documentales y de clasificación. Madrid: Pirámide, 1993, p. 214.

18. DEWÈZE, A. Réseaux sémantiques: essai de modélisation; application à l'indexation et à la recherche documentaire. Lyon: Universidad Claude Bernard. Tesis doctoral, 1981, p. 68.

19. SCHAÜBLE, P. Information Retrieval Based on Information Structures. Informatik-dissertationen eth Zürich, Verlag der Fachvereine, 1989, 135 p.



Durante las dos últimas décadas se han propuesto diversos métodos para la construcción automática de los tesauros conceptuales, pero todos ellos descansan básicamente en análisis estadísticos de la covarianza de las palabras que componen los textos, no alcanzando a utilizar una función realmente apropiada²⁰. Aunque no existe una clara demostración de la utilidad de los términos generados mediante dicho análisis, una investigación reciente ha sugerido que la relevancia de los documentos recuperados puede doblarse si el usuario añade a sus propios términos, los sugeridos por un tesoro autogenerado mediante dicho análisis estadístico²¹.

Chen sugiere que para la autogeneración de un tesoro conceptual automatizado, se deben crear las condiciones previas adecuadas para el desarrollo de algoritmos potentes y capaces de producir:

- listas de términos que actúen como filtros para la identificación inmediata de los conceptos más importantes de los documentos,
- indizadores automáticos para la identificación por materias del resto de los conceptos, y
- analizadores semánticos que agrupen los conceptos relacionados, mediante el procesamiento de los diferentes documentos de la base de datos y la obtención de sus índices de covarianza correspondientes²².

En la misma línea conceptual, Paice propone que tanto las preguntas como los documentos sean representados por «*entidades condensadas ideales*» o situaciones sacadas del texto, generadas mediante «*expansión activada*»²³ de los paneles conceptuales:

20. La mayoría de las funciones de análisis de covarianza existentes (ej.: coseno, Dice, de Jaccard, etc.) son simétricas por naturaleza, habiendo producido efectos colaterales no deseados y poca exactitud en los resultados.

21. Ibid. Cit. n° 15.

22. CHEN, H., YIM, T. y FYE, D. Automatic thesaurus generation for an eElectronic community system. Journal of the ASIS, 1995, 46 (3), p. 179.

23. Formados por la distribución («mapping») de los términos tomados de un área del conocimiento específica y agrupados en forma de red neuronal.



«Si somos capaces de hacer equivalentes una alusión particular de un texto con un tema conocido a través de dichas entidades, entonces una entidad almacenada podrá ser utilizada como base para la representación de la idea implicada»²⁴

Para la construcción de estas entidades, capaces de ser comprendidas, descritas y almacenadas como base para la representación de las ideas implicadas en las preguntas y en los documentos²⁵, se precisa disponer de una base de conocimientos terminológicos suficientemente profusa y exacta, que permita abarcar:

- los términos del tesoro que estén en el fragmento de texto analizado,
- los términos no mencionados en dicho fragmento que aparecen interrelacionados con los anteriores en el tesoro,
- los términos que se interrelacionan con ambos términos anteriores dentro de la red semántica, y
- las propiedades de todos ellos, junto con las relaciones de cualquier tipo que los unan (asociativas, de relevancia, etc.).²⁶

Con una exacta medición del grado de solapamiento entre cada dos de estas «situaciones ideales»²⁷ se puede incrementar la relevancia de los sistemas de recuperación y hacer posible la utilización de un espacio conceptual que refleje relaciones asociativas más estrechas entre los datos e ignore las menos importantes. Las recuperaciones se realizarán colocando un nuevo objeto, correspondiente a una pregunta, dentro de la red neuronal, de manera que pueda localizar sus documentos más próximos conceptualmente²⁸.

24. PAICE, C.D. «A Thesaural Model of Information Retrieval». *Information Processing Management*, 1991, 27(5), p. 435.

25. Existentes separadamente de los textos y de los tesauros terminológicos autogenerados.

26. GREEN, R. Topical Relevance Relationships. Why topic matching fails?. *Journal of the ASIS*, 1995, 46 (9), p. 648.

27. Ej.: incorporando pesos a los vectores que unen los conceptos.

28. DEERWESTER et al. Indexing by latent semantic analysis. *Journal of the ASIS*, 1990, 41 (6), p. 394.



APLICACIÓN DE LA TECNOLOGÍA HIPERTEXTUAL EN LOS SISTEMAS DE ORGANIZACIÓN DE LA INFORMACIÓN

Mientras la organización de la información contenida en *los trabajos impresos* está dispuesta secuencialmente (en capítulos y párrafos con una variedad de índices), la información contenida en *las bases de datos de texto completo* precisa de una estructuración que permita una recuperación no tan rígida, mediante referencias cruzadas entre sus distintos documentos asociados. La organización hipertextual permite moverse a través de una red conceptual de macroestructuras, que se define a partir de las asociaciones de ideas del usuario, derivadas de su nivel de conocimientos de cada disciplina y de su filosofía personal.

Los sistemas hipertextuales permiten mejorar la organización escasamente estructurada de estas bases de datos, y atenuan la débil relevancia en la recuperación de los documentos científicos, derivada de las características propias de su discurso, del considerable volumen de los textos, del almacenamiento masivo en bases de datos en texto completo, y de la integración de tipos muy diversos de documentos, entre otros factores.

Estos sistemas se caracterizan por:

- Proporcionar una gran potencia conceptual para la gestión del texto de los documentos y de su red de relaciones no estructuradas, y
- Favorecer que la colección documental sea explorada utilizando criterios asociativos definidos por los enlaces hipertextuales entre las diversas partes de sus documentos.

La creación de una ontología hipertextual:

Desde que en 1945 Vannevar Bush (asesor científico de Roosevelt) propusiera su sistema de recuperación Memex mediante referencias cruzadas²⁹, se ha considerado repetidamente la idea de crear sistemas hipertextuales para el procesamiento

29. BUSH, V. As We may think. The Atlantic Monthly, July 1945



APLICACIÓN DE LA TECNOLOGÍA HIPERTEXTUAL EN LOS SISTEMAS DE ORGANIZACIÓN DE LA INFORMACIÓN

Mientras la organización de la información contenida en *los trabajos impresos* está dispuesta secuencialmente (en capítulos y párrafos con una variedad de índices), la información contenida en *las bases de datos de texto completo* precisa de una estructuración que permita una recuperación no tan rígida, mediante referencias cruzadas entre sus distintos documentos asociados. La organización hipertextual permite moverse a través de una red conceptual de macroestructuras, que se define a partir de las asociaciones de ideas del usuario, derivadas de su nivel de conocimientos de cada disciplina y de su filosofía personal.

Los sistemas hipertextuales permiten mejorar la organización escasamente estructurada de estas bases de datos, y atenuan la débil relevancia en la recuperación de los documentos científicos, derivada de las características propias de su discurso, del considerable volumen de los textos, del almacenamiento masivo en bases de datos en texto completo, y de la integración de tipos muy diversos de documentos, entre otros factores.

Estos sistemas se caracterizan por:

- Proporcionar una gran potencia conceptual para la gestión del texto de los documentos y de su red de relaciones no estructuradas, y
- Favorecer que la colección documental sea explorada utilizando criterios asociativos definidos por los enlaces hipertextuales entre las diversas partes de sus documentos.

La creación de una ontología hipertextual:

Desde que en 1945 Vannevar Bush (asesor científico de Roosevelt) propusiera su sistema de recuperación Memex mediante referencias cruzadas²⁹, se ha considerado repetidamente la idea de crear sistemas hipertextuales para el procesamiento

29. BUSH, V. As We may think. The Atlantic Monthly, July 1945



de las fuentes documentales. Se ha desarrollado dicha idea con el avance de la tecnología informática, siendo Ted Nelson uno de sus más fervientes defensores en el ámbito científico actual, a partir de que en 1981 definiera su sistema Xanadu³⁰. La tecnología hipertextual es la mejor herramienta para estructurar enlaces significativos entre los documentos científicos, y para relacionar los descriptores de los tesauros y los códigos clasificatorios.

Para mejorar la precisión de las recuperaciones documentales se propone el diseño de ontologías especializadas por áreas del conocimiento que conduzcan hasta las preguntas específicas del área tratada. Deberán ser reutilizadas, en las bases de conocimientos de los agentes expertos de la IA, para la autogeneración de tesauros conceptuales internos que permitan la distinción de los sinónimos, la supresión de los homónimos y la inducción de relaciones asociativas entre los descriptores.

Una ontología para una base de conocimientos de la IA debe abarcar los diferentes tipos de documentos, las descripciones conceptuales, las relaciones entre dichos documentos (citas), y las de éstos con los diferentes problemas científicos; además de índices, descripciones bibliográficas, tesauros, códigos clasificatorios, formalizaciones de validez, información terminológica, etc. Su aplicación debe proporcionar una metavisión de la estructura y de la terminología del dominio que facilite recuperaciones altamente relevantes.

Superación del carácter estático de la tecnología hipertextual:

Se discute la aplicación de esta organización en los agentes expertos, como apoyo directo en la resolución de problemas del trabajo científico profesional, debido a la complejidad de la tarea intelectual para crear las bases de conocimientos de soporte. Sin embargo, una buena parte de los documentos científicos están especialmente indicados para su organización hipertextual, al estar repletos de referencias cruzadas, expresas o tácitas (ej.: las leyes y sus desarrollos reglamentarios, la doctrina jurídica y sus definiciones o comentarios, los casos legales de las diferentes jurisdicciones, etc.)³¹.

30. NELSON, T. Replacing the printed word: A complete literary system. Information Processing '80, 1980.

31. BENCH-CAPON, T. An incremental approach to legal drafting support. Proceedings of JURIX'94, 1994. Lelystad: Kroninklijke Vermande, Netherlands.



Dado que las bases de datos documentales no fueron diseñadas para la resolución directa de problemas, la organización de su conocimiento requiere el análisis de cada documento y el resumen de sus notas más destacadas, mediante dos fases consecutivas: primera: la extracción previa de los datos menos relevantes (clasificación de los documentos), y segunda: el posterior examen especializado para estructurar el conocimiento referido a temas específicos.

- a. Una de las limitaciones de estos sistemas es el "*desbordamiento cognitivo*" o *desorientación de los usuarios* durante su navegación a través de los textos o de las clasificaciones hipertextuales para encontrar documentos relevantes o conceptos apropiados, respectivamente³².

La unión de la información de numerosas fuentes con un único nodo conduce a una sobresaturación de enlaces en dicho nodo. La introducción de muchos de éstos dificulta la navegación, reduce el deseo de los usuarios de explorar su información e incrementa las posibilidades de perderse. Se trata de utilizar sólo enlaces significativos e incluir en cada nodo una idea conceptual completa que sintetice el conocimiento contenido en las fuentes enlazadas.

- b. Otra de las limitaciones de estos sistemas se deriva de *la necesidad de controlar manualmente* los enlaces creados durante el trabajo automático, dado que éstos no contienen *información conceptual directa* "a priori" sobre el contenido de los documentos relacionados.

Al no poder procesar dinámicamente la información existente en sus fuentes, dichos sistemas recuperan solamente los documentos que tienen enlaces predefinidos con un nodo principal o mantienen relaciones dinámicas con otros. Se trata de proporcionar al usuario herramientas globales de ayuda a la navegación: mapas cognitivos, historia semántica de los nodos visitados, opción de retorno mediante el marcado de nodos, etc.³³

32. ROVIRA, C. Entornos hipertextuales de aprendizaje. Anuario SOCAD, 1997, pp. 121-126.

33. CODINA, L. El libro digital. Barcelona: Centre d'investigació de la Comunicació, 1996.



CONCLUSIONES Y SOLUCIONES PROPUESTAS

- La normalización de la terminología de los documentos es uno de los problemas cruciales para el desarrollo de los sistemas hipertextuales. Se estima que en éstos se precisa de un *conocimiento terminológico mínimo* de 80.000 términos normalizados. Para ello se propone el desarrollo de aplicaciones hipertextuales que utilicen las tecnologías de la IA para automatizar la compilación de hipertesoros a partir de las bases de conocimientos terminológicos, capaces de abarcar una buena parte del corpus lingüístico de cada una de las lenguas más utilizadas³⁴.

Entre las soluciones investigadas atenuar el carácter estático de la tecnología hipertextual, se destaca la utilización del tesoro Conceptual³⁵, concebido como una red semántica en la que en cada nodo hay un único concepto semántico que puede contener una serie de descriptores asociados, que también pueden ser identificados en la red de descriptores relacionados según las típicas relaciones de los tesoros: preferenciales, jerárquicas o asociativas.

Para evitar la falta de precisión del lenguaje natural, se propone la incorporación de los conceptos obtenidos del análisis de contenido de los documentos científicos en un tesoro conceptual que facilite las recuperaciones desde la interfase sistema usuario del profesional. De esta forma se incrementa la variedad de las ecuaciones de búsqueda del usuario, o al menos le muestra la variedad de las existentes; lo que, de acuerdo con el *Principio de Variedad de Requisitos* de Ashby³⁶, incrementa simultáneamente la cantidad de respuestas del sistema consultado.

34. Electronic Dictionary Research Project. Japan Key Technology Center. JTEC/WTEC Hyper-Librarian, 1995, 30/6. http://www.itri.loyola.edu/kb/c5_s2.htm.

VIVALDI, J. Seminario Procesamiento Lenguaje Natural (SEPLN'95). Univ. de Deusto, Bilbao, 20-23 sepre., 1995.

35. LÓPEZ ALONSO, Miguel-A. Un tesoro conceptual para la recuperación de la información jurídica comercial. *Rev. Espa. Doc. Cient.*, 1998, 21 (2), pp. 164-173.

36. *Ibid*, Cit, nº 10.



- La perfección de los agentes expertos del campo de la IA permitirá mejorar de las bases de conocimientos de dichos sistemas, facilitando su viabilidad como herramientas para la organización y recuperación del conocimiento en la actividad científica habitual³⁷, apoyados en la tecnología hipertextual integrada con las redes semánticas³⁸. Estas últimas deberán proporcionar un mapa de las variables introducidas en la búsqueda y de las obtenidas en la recuperación, y ser capaces de almacenar información de forma distribuida, aprendiendo de los problemas cotidianos, tras un período de validación que actualice las respuestas³⁹.

Para su optimización, *los sistemas integrados de gestión de la información deberán contar con una interfase hombre-máquina, capaz de relacionar los conceptos científicos a partir de las ontologías preestablecidas*. El usuario tendrá acceso a ellos a través de una representación gráfica de la red (tabla de contenidos gráficos con las conexiones visibles mediante enlaces hipertextuales), por la que podrá rastrear los documentos siguiendo su línea de interés⁴⁰.

Se deberá priorizar el desarrollo de asociaciones de redes neuronales que de alguna manera conviertan en dinámicos los nodos hipertextuales, permitiendo que todos los conceptos situados en cada uno de ellos puedan ser consultado simultáneamente (no sólo secuencialmente) de acuerdo con las necesidades cognitivas de cada usuario en un momento dado. Para ello será preciso utilizar neurordenadores, con procesamiento "masivamente paralelo", que trabajen con los múltiples nodos hipertextuales del mismo modo que el cerebro humano trabaja con múltiples neuronas.⁴¹

37. CORTEZ, E. M. et al. The hybrid application of an inductive learning method and a neural network for intelligent information retrieval. *Information Processing and Management*, 1995, 31 (6), p.790.

38. BERGER, F.C.; VAN BOMMEL, P. Augmenting a characterization network with semantic information. *Information Processing and Management*, 1997, 33 (4), pp. 453-479.

39. COHEN, P. R. y KJELDSEN, R. Information retrieval by constrained spreading activation in semantic networks. *Information Processing and Management*, 1987, 23 (4), p. 257.

40. ALLEN, L.E. *Language, law and logic; plain legal drafting for the electronic age*. Computer Science and Law, Cambridge U. Press, 1980., pp. 75-100.

41. MOYA, F.de, HERRERO, V. y GUERRERO, V. La aplicación de redes neuronales artificiales a la recuperación de la información. *Anuario SOCADI 1998*, pp. 147-164.