

# *Vocabulario básico para la recuperación de la información en Sociología*

Julián COLINA  
*SEDIC*

## **RESUMEN**

El conocimiento sociológico escrito y publicado en su forma más universal y completa se encuentra en las Bases de Datos y en los CD-ROM; y su antecesor, como mejor organización del conocimiento, en la Enciclopedia Internacional de las Ciencias Sociales. Aquí se consideran algunas reflexiones para la reconstrucción de un tesoro de Sociología en lengua española, que ahora, con las fuentes propuestas, es ya posible.

**Descriptores:** Tesoro, Diccionario, Bases de Datos, Sociología, Ciencias Sociales, Recuperación de la Información, Organización del Conocimiento, Socioterminología, Lenguajes Terminológicos, Lenguajes Documentales, Lenguaje natural, Normas de Documentación.

## **INTRODUCCION**

El conocimiento sociológico escrito y publicado en su forma más universal y completa se encuentra ahora en las Bases de Datos, y su antecesor —como mejor organización del conocimiento— en la Enciclopedia Internacional de las Ciencias Sociales. Ambos recursos se complementan para el propósito de esta Nota. Si se dispusiera de un vocabulario lo más universal posible para la recuperación exhaustiva de información, se aprovecharía enormemente todo lo publicado, principalmente en Revistas y Documentación en general.

En España son cuatro instituciones las que trabajan en temas de Sociología, con algún Tesouro sistemático:

- CIS Centro de Investigaciones Sociológica. Revista Española de Investigaciones Sociológicas.
- CINDOC Centro de Información y Documentación Científica. Índice Español de Ciencias Sociales. Bases de Datos en CD-ROM del CSIC.
- IESA Instituto de Estudios Sociales Avanzados. Revista Internacional de Sociología.
- Facultades de Ciencias Políticas y Sociología y otras Instituciones de formación, asistencia y/o investigación, que desde hace más de veinte años están publicando más de una decena de REVISTAS.

## FUENTES BIBLIOGRAFICAS EN AMERICA Y EUROPA

- *Enciclopedia Internacional de las Ciencias Sociales*, original inglés y versión española. Tomo 11 : Indices.
- *Thesaurus of sociological indexing terms de Sociological Abstracts*. Bases de Datos en línea y en CD-ROM.
- *The Contemporary Thesaurus of Social Science Terms and Synonyms. A guide for natural language computer searching*, Sara D. Knapp. Oryx.
- Thesaurus de la UNESCO. Parte de Temas: Ciencias Sociales y Humanas.
- *KNOWLEDGE ORGANIZATION* de ISKO. Frecuentes trabajos y reseñas de Documentación en Sociología.
- CINDOC. *Tesouro de Sociología*. Parte alfabética, jerárquica y alfabética KWIC.
- Universidad Autónoma de Barcelona, en la Revista *Papers de Sociología*. Términos usados para los Artículos.
- AENOR. *CDU*. Parte de Tablas de Ciencias Sociales. Edición Media.
- AENOR. *UNE 50-106-90. Directrices para el establecimiento y desarrollo de tesauros monolingües*.
- Emilia Currás. *Tesauros. Lenguajes terminológicos*. Paraninfo.
- Georges Van Slype. *Los lenguajes de indización. Concepción, construcción y utilización en los sistemas documentales*. Fundación GSR.
- *Diccionario de Ciencias Sociales* del Instituto de Estudios Políticos y UNESCO.

Actualmente comienzan a aparecer aquí, en Revistas y en Congresos, más trabajos que tratan estos temas. En general son los citados en la Revista Española de Documentación Científica del CINDOC y en la Revista Knowledge Organization de ISKO.

Hay otras cuestiones asociadas con los tesauros que es necesario o útil

conocer: algunos programas informáticos especializados en la construcción de tesauros o en bases de datos o en análisis de textos (el mismo CINDOC acaba de publicar uno); asimismo, diccionarios en CD-ROM, como el de la Real Academia de la Lengua; y también son interesantes los encabezamientos de materias.

La parte expositiva en el Tesoro de la UNESCO y en las demás obras completan un buen manual-guía.

Una mención especial para la Universidad de Colima y la Universidad Nacional Autónoma de México y para las personas que están trabajando en Biblioteconomía, en Iberoamérica.

## METODO Y CRITERIOS DE LA SELECCION

De las fuentes citadas se han seleccionado, para la formación de un VOCABULARIO BASICO, las siguientes: la *Enciclopedia Internacional de las Ciencias Sociales*, como mejor antecedente; el *Sociological Abstracts*, como más conocido; el libro de Sara Knapp, lo más moderno en América; el thesaurus del CINDOC (CSIC), como el único indígena encontrado, junto con el de la UAB; el Tesoro reciente de la UNESCO, como el más preciso, y la edición Media de las Tablas CDU, de inminente aparición en español.

Aquí están algunas instrumentos empleados. Pueden hacerse las concordancias entre unos y otros para llegar a un método de optimización de un tesauro en español de Sociología, y por extensión de Ciencias Sociales; éste es el enfoque de la mayoría, o al menos la opción de un enfoque más amplio de la Sociología.

Existen otras clasificaciones internacionales para la bibliografía que aquí no se han contemplado, ya que el análisis se reduce a las expuestas.

Nos hemos centrado, para una somera comparación, en dos temas, para hacer una evaluación de la cantidad y calidad de los términos descriptores: las cuestiones de Metodología y de Cambio Social, ambas siempre en desarrollo. Y así pues se han considerado sus términos equivalentes, jerárquicos y asociados, pues se trata de un tesauro, y otros parámetros como el total de páginas y número de términos e índices en la obra y la fecha.

Finalmente, no fue posible una comparación estricta entre todos ellos, pues no todos tienen los tres tipos de índices del CINDOC, o están en cierta forma comprendidos en uno alfabético o les falta el índice jerárquico o no son propiamente tesauros de vocabulario controlado. Así pues, una mera tabulación de las veces que unos términos se relacionan con sus equivalentes o sus asociados no era verdaderamente descriptiva de una calidad del vocabulario o era irrelevante.

De haber dispuesto del material citado en soporte informático, el análisis podría haber sido muy preciso, pero de todas formas el estudioso puede

hacer sin mucho esfuerzo comparaciones manuales en los términos de su interés.

Un objetivo de este trabajo era, además, detectar todas las fuentes bibliográficas de tesauros en español, aunque no conocemos los trabajos hechos en Iberoamérica. La Norma y tratados citados, junto con la parte expositiva, que acompaña a los tesauros, creemos que es suficiente para la parte doctrinal del proyecto propuesto en esta Nota.

## CONCLUSIONES

Lo primero que se constata, es que la labor que todos ellos han desarrollado es abrumadora, y el texto muy completo, por el número de conceptos, términos equivalentes y asociados. Un nuevo enfoque, lenguaje natural y páginas, es el de Sara Knapp, en lengua inglesa.

En lengua española, la Enciclopedia; y pioneros, el Diccionario y algunas pequeñas nomenclaturas o directorios de Bases de Datos de Organismos.

El del CINDOC es un thesaurus vivo, actualizado, bien diseñado y único general de Sociología.

La UAB es la única institución que está utilizando descriptores en una revista general de Sociología.

El resto de instituciones, CIS e IESA, no emplean Descriptores en la presentación de sus Artículos.

El thesaurus de la UNESCO, de excelente estructura, es el más moderno en español, en inglés y francés, en cuanto a la parte de Ciencias Sociales.

El de las Tablas CDU –la parte de Sociología– ha renovado en calidad ediciones anteriores, hasta tal punto que es mejor la edición Media que la edición anterior completa de Ciencias Sociales.

Creemos que se puede compaginar la estructura rigurosa, controlada, de un tesoro con un «simple» índice, vocabulario, glosario o léxico; dicho esto a efecto de usar cualquier fuente para ampliar un tesoro.

Un posible vocabulario universal o común o unificado o ampliado, deberá ser abierto y probablemente multidisciplinar, dado el dinamismo de la Sociología y las Ciencias Sociales.

Términos que empiezan siendo asociados y cuasisinónimos, terminan finalmente como descriptores principales. Deberían ser ya incorporados en beneficio del usuario, que sí los emplea.

Todos han utilizado la forma clásica de los tesauros de «vease», «además», «relacionado con», etc. Ello hace más fácil las consultas sinópticas de los diferentes vocabularios.

Parece lógico usar lo que tenemos, y también mejorarlo, con su puesta al día o su ampliación, que es lo que se propone.

Existe, por tratarse de Sociología, un concepto y un contexto para un de-

terminado tiempo y país. Tiene que ver con las enseñanzas de las Facultades, la demanda que de las Ciencias Sociales hace la sociedad, y los usos que hacen los sociólogos. Estos aspectos se detectan más aún repasando un tesoro (y un CD-ROM) y viendo la forma como han tratado un tema y qué temas han relacionado con él. Se ve la amplitud del pensamiento para cada sociedad en un tiempo determinado.

Es, pues, también una cuestión cultural y no debe ser una valoración. Es como el lenguaje mismo. Pero el lenguaje o el vocabulario de cada uno de nosotros es un reflejo de nuestras experiencias y, al menos en su formulación, debe ser muy amplio, añadiendo, en el caso que tratamos, más palabras sinónimas y términos relacionados a los términos principales —que deben completarse—, para la construcción (o mejor, la reconstrucción) de nuestro tesoro mencionado.

No entramos en el tema de la imposibilidad de hacer natural el lenguaje científico, por su especificidad. Ni tampoco en la conflictividad propia de las Ciencias Sociales, que dificultaría su universalidad.

Hay trabajos españoles concretos publicados en revistas —no demasiados— relativos a lenguajes documentales y terminológicos e indización automática y otros varios temas asociados, como algoritmos de recuperación de la información.

Algunos de los tesauros incluyen un vocabulario español-inglés, que es lo correcto. Sirve para usar los descriptores de los artículos en ambos idiomas, que es lo habitual. Lo mismo cabe decir de los resúmenes.

No por muy obvio dejaremos de decir que hay que utilizar programas informatizados, y diccionarios y enciclopedias —preferiblemente en CD-ROM— para hacer o reconstruir un tesoro. Todo depende de la amplitud que se piense dar a la obra, pero piénsese que serán miles las conexiones. Así, probablemente lo ideal serían unos 5000 conceptos principales, términos preferentes o no, para el conjunto de las Ciencias Sociales.

Consecuentemente, el posible equipo de personas, además de ser expertos en diferentes Ciencias Sociales, debe estar compuesto de bibliotecarios, documentalistas, informáticos, filólogos... y artistas y pensar en los usuarios, en los autores y en los editores.

Las Bases de Datos deben hacer muy asequible la terminología que están empleando, al menos a nivel de palabras clave o descriptores, incluyendo la posibilidad de consultar el tesoro, aunque sólo sean las relaciones jerárquicas. Consecuentemente, los documentalistas en Sociología deberían emplear descriptores, además de resúmenes —esto generalmente sí se hace— para describir los artículos de las revistas y otras literaturas. Asimismo, mucha terminología de la Documentación y de la Información debería estar presente en los tesauros, cualesquiera que sea su materia.

La valoración de los tesauros y glosarios consultados y los finalmente seleccionados para completar una primera fase de aproximación concluyó en

dos grupos: los más ricos en terminología fueron los americanos (Enciclopedia Int. Cs. Ss., Sara y SocioAbs), y el otro grupo, más preciso, más controlado, formado por CDU, UNESCO y CINDOC.

Un valor añadido sería la posibilidad de medir la calidad de las búsquedas del usuario, en aras de un diseño más eficaz para la recuperación de la información y también para la calidad de la información. También la capacidad de su transferencia, contenida en las Bases de Datos, por medio de un algoritmo que relacione los términos descriptores equivalentes y asociados y entre categorías, valorándolo matemáticamente con un test no paramétrico.

¿Qué debería hacerse? Cotejar los tesauros indígenas en uso con los citados UNESCO y CDU. Considerar el nuevo enfoque que Sara Knapp ha dado a su enorme obra, esto es importante. Añadir aquellos glosarios especializados de nuestro interés, ya existentes. Cuidar más los términos asociados. Eventualmente, incorporar los trabajos hechos en Ibeoramérica. El resultado final debería de ser publicado de alguna forma, para beneficio de los usuarios y para poder renovar el proceso, que, como sabemos, es dinámico.

Debemos dar las gracias a la Biblioteca de Referencia del Cape Cod Community College, a la Biblioteca Pública de la Ciudad de Falmouth y a la empresa ANILOC de Sandwich, todas de los Estados Unidos. En España al CINDOC, a ISKO (que propició la difusión del Proyecto) a DOC6, al GAD, a la librería DELSA, a la Universidad Autónoma de Barcelona, a las Hemerotecas de la Biblioteca Nacional, a la revista Educación y Biblioteca, por la difusión del proyecto mexicano y que obviamente está referido a personas concretas que han hecho posible hacer esta nota.

---