

Vocabulario básico para la recuperación de la información en Sociología

Julián Colina
SEDIC

0.1. Resumen

El conocimiento sociológico escrito y publicado en su forma más universal y completa se encuentra recogida en diversas bases de datos en CD-ROM. Además, es necesario considerar, como una excelente herramienta de organización del conocimiento en dicho campo, la Enciclopedia Internacional de las Ciencias Sociales. Se proponen algunas reflexiones para la construcción de un tesoro de Sociología en lengua española.

Palabras clave: 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.

0.2. Abstract

Sociological knowledge is referred in various CD-Roms databases. It should also be considered, as the best knowledge organization tool in this field, the International Encyclopedia of Social Sciences. On this background, some reflections for developing a thesaurus of Sociology in Spanish are proposed.

Keywords: Thesaurus. Dictionary. Databases. Sociology. Social Sciences. Information retrieval. Knowledge organization. Socioterminology. Terminologic languages. Indexing languages. Natural language. Standards in Library and Information Science.

1. Introducción

El conocimiento sociológico escrito y publicado en su forma más universal y completa se encuentra recogido en varios CD-ROM, organizados mediante diversos tesauros. Aunque su antecesor, como mejor organización del conocimiento, es la Enciclopedia Internacional de las Ciencias Sociales. Ambos recursos se complementan para el propósito de este trabajo.

En España son en principio cuatro instituciones las que tienen algún sistema de organización del conocimiento, a saber, una relación de los conceptos fundamentales de la Sociología y sus relaciones básicas:

- CIS (Centro de Investigaciones Sociológica): *Revista Española de Investigaciones Sociológicas*.
- CINDOC (Centro de Información y Documentación Científica): *Indice Español de Ciencias Sociales y Bases de datos en CD-ROM del CSIC*.
- IESA (Instituto de Estudios Sociales Avanzados): *Revista Internacional de Sociología*.

Además, tenemos que tener en cuenta las 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.

Las fuentes bibliográficas con las que contamos en América y Europa son, fundamentalmente, las siguientes:

- *Enciclopedia Internacional de las Ciencias Sociales*, original inglés y versión española. Tomo 11: Índice.
- *Thesaurus of sociological indexing terms* de Sociological Abstracts.
- *The Contemporary Thesaurus of Social Science Terms and Synonyms: A guide for natural language computer searching* de Sara D. Knapp.
- Thesaurus de la Unesco.
- Knowledge Organization de ISKO.
- CINDOC.
- Universidad Autónoma de Barcelona.

2. Algunos antecedentes y método

Algunas fuentes importantes que pueden servir de punto de partida son:

- *Thematic list of descriptors sociology de ICSSD*. Routledge
- *Macrothesaurus para el procesamiento de la Información relativa al desarrollo*
- *Diccionario de Ciencias Sociales del Instituto de Estudios Políticos*
- *Thesaurus de la Unesco*

Respecto a las fuentes secundarias, 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*.

También hay que tener en cuenta en relación con los tesauros las aportacio-

nes de la AENOR, las Tablas CDU y sus Normas de Documentación, y algunos programas informatizados especializados en la construcción de tesauros o en bases de datos o en análisis de textos. El mismo CINDOC acaba de publicar uno., Asimismo interesan algunos Diccionarios en CD-ROM, Real Academia de la Lengua, etc., encabezamientos de materias o directorios de la Biblioteca del Congreso (USA), CEE, Universidad de Sevilla, FUNDESCO, Instituto de la Juventud, Instituto de la Mujer, etc. También algún libro y la parte expositiva de las fuentes bibliográficas, citadas más arriba; principalmente la del Tesauro de la Unesco, y, en general, la obra de Emilia Curras. Contamos igualmente con trabajos españoles concretos publicados en revistas, no demasiados, relativos a lenguajes documentales y terminológicos e indización automática y otros varios temas como algoritmos de recuperación de la información.

Deseo hacer igualmente una mención especial a los especialistas en Biblioteconomía que están trabajando en la Universidad de Colima y la Universidad Nacional Autónoma de México.

Los resultados de la evaluación de las fuentes fueron los siguientes: la *Enciclopedia Internacional de las Ciencias Sociales*, como mejor antecedente; los *Sociological Abstracts*, como el más conocido; el libro de Sara Knapp, como lo más moderno en América; el Tesauro del CINDOC (CSIC), como el único indígena encontrado, junto con el de la UAB, que lo utilizan en sus publicaciones; y el Tesauro reciente de la UNESCO, como el más preciso.

Apartir de sus metodologías y estructuras, a través de la creación de tablas de las concordancias entre unos y otros, podemos llegar a un método de optimización de un tesauro en español de Sociología y por extensión de Ciencias Sociales, pues éste último es el enfoque de la mayoría.

Me he centrado, para una somera comparación, en dos temas, para hacer una computación: las cuestiones de Metodología y de Cambio Social. Se trata de dos áreas en continuo desarrollo. He considerado los términos equivalentes, jerárquicos y asociados, pues se trata de tesauros, así como el total de páginas y términos que las componen y la fecha de edición.

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.

3. Conclusiones

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

En lengua española son pioneros la *Enciclopedia Internacional de las Ciencias Sociales*, el *Diccionario de Ciencias Sociales del Instituto de Estudios Políticos* y algunas pequeñas nomenclaturas o directorios de Bases de Datos de Organismos. El tesoro del CINDOC es un tesoro vivo, mientras que el de la UAB es el único que utiliza descriptores en una revista general de Sociología. Mientras el resto de las instituciones como CIS o AESA no tienen tesoro y no utilizan descriptores en la presentación de sus artículos. El actual tesoro de la UNESCO; en tres idiomas permite una excelente estructura y supone un verdadero tesoro, siendo sin duda el más moderno en español (además de en inglés y francés).

En mi opinión se puede compaginar la estructura rigurosa y controlada de un tesoro con un *simple* índice, vocabulario, glosario o léxico, a efecto de usar cualquiera de estas fuentes para ampliar el tesoro. La Sociología, que por definición es dinámica y con un posible vocabulario universal, común, unificado o ampliado, requiere un tesoro abierto y probablemente multidisciplinar —como las Ciencias Sociales—. Así sucede que términos que empiezan siendo asociados y cuasisinónimos, terminan finalmente como descriptores principales.

Todos han utilizado la forma clásica de los tesauros (*vease ademas, relacionado con*), lo cual hace más fácil las consultas sinópticas de los diferentes vocabularios.

Parece lógico usar lo que tenemos, además de mejorarlo, con su puesta al día o su ampliación, propuesta que ya vengo proponiendo.

El estudio de la Sociología requiere contemplar un concepto y un contexto para cada época y país. Factor que afecta a las enseñanzas de la Facultad, a la demanda que de las Ciencias Sociales hace la Sociedad y, finalmente, a los propios sociólogos. Estos aspectos se detectan aún más repasando un tesoro (y una base de datos en CD-ROM) y viendo la forma como han tratado el tema y los temas relacionados con él. Se ve la amplitud del pensamiento para cada sociedad en un tiempo determinado. Es pues también una cuestión cultural, y no debe quedar en 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 mentado.

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.

No por muy obvio dejarés de decir que hay que utilizar programas informáti-

zados y diccionarios y enciclopedias, preferiblemente en CD-ROM, para hacer o recuistruir un tesauo, todo depende de la amplitud que pensamos dar a la obra, pero piénsese que serán miles las conexiones. Probablemente lo ideal son unos 5000 conceptos principales, terminos preferentes o no, para el conjunto de las Ciencias Sociales. Consecuentemente , el posible equipo de personas, además de ser expertos en diferentes Ciencias Sociales deben de ser Bibliotecarios, Documentalistas, Informáticos, Filólogos... 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 tesauo, 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. También pienso, asímismo, que mucha terminología de la Documentación y de la Información debería estar presente en los tesauos, cualesquiera que sea su materia.

Un ejemplo de lo dicho son las dificultades que experimenté para encontrar el libro de Sara D. Knapp en diversas bases de datos de Sociología. Los usuarios inevitablemente empleamos términos del lenguaje natural y continuamos con sinónimos y cuasisinónimos, es decir, el lenguaje hablado.

Tras la valoración de los tesauos y glosarios consultados, y los finalmente seleccionados para completar la primera fase de aproximación, los clasifico en dos grupos: los mas ricos en terminología coincidieron con los americanos (*Enciclopedia Internacional de las Ciencias Sociales*, *The Contemporary Thesaurus of Social Science Terms and Synonyms: A guide for natural language computer searching* y *Thesaurus of sociological indexing terms* de los Sociological Abstracts) y el otro grupo, con un lenguaje más controlado (*Thesaurus de la UNESCO* y del CINDOC).

Un valor añadido sería la posibilidad de medir la calidad de las búsquedas del usuario —para un diseño más eficaz de la recuperación de la información— y también la calidad de la información contenida en las bases de datos, su potencial de transferencia, por medio de un algoritmo que relacionara los términos descriptores equivalentes y asociados y entre categorías; valorándolo matemáticamente con un test no paramétrico.

Debemos por lo tanto proceder, primero cotejando los tesauos indígeneas en uso con el citado Thesaurus de la UNESCO, considerando el nuevo enfoque que Sara Knapp ha dado a su enorme obra. El resultado final debería ser publicado de alguna forma para beneficio de los usuarios y para poder renovar el proceso, que, como decíamos, es dinámico.