

Actualidades

NC-ISO 2788: 2000. Directrices para el establecimiento y desarrollo de tesauros monolingües

Actualmente, universidades de todo el mundo laboran en el desarrollo de las denominadas ontologías, así como de la Web semántica. Aunque se han logrado ciertos avances, todavía queda mucho por hacer en áreas como la inteligencia artificial aplicada a los sistemas de información, con el fin de que sean capaces de cubrir las expectativas de los especialistas que laboran en esta clase de organizaciones.

La Web semántica, como esperamos, llegará por su propio peso, en el momento y espacio convenientes, por el nivel de los resultados acumulados en las investigaciones y en las innovaciones tecnológicas correspondientes. En estas circunstancias, al menos, estamos en la obligación de laborar en el perfeccionamiento de sistemas lingüísticos artificiales, como los tesauros, para su aplicación en el quehacer de nuestras organizaciones de información, con el objetivo de contribuir a la correcta indización de los documentos que ingresan a sus fondos y, más importante aún, su recuperación.

Para cumplir con estos objetivos, la consulta de la Norma Cubana – *International Standard Organization*, NC-ISO 2788: 2000, denominada *Directrices para el establecimiento y desarrollo de tesauros monolingües*, puede ser un instrumento muy útil. Las recomendaciones establecidas en esta norma tienen por objeto garantizar una práctica coherente en el espacio de uno o varios puntos de indización, como ocurre en el caso de una red con procesamiento descentralizado.

Sus autores advierten, sin embargo, que la norma no se debe considerar como una instrucción de cumplimiento obligatorio y, en muchos casos, describen procedimientos opcionales, por ejemplo, para la presentación de las relaciones entre los términos sin designar ninguna de ellas como técnica preferente. La elección de un procedimiento variará de unos centros a otros en dependencia de decisiones administrativas que escapen al campo de aplicación de esta norma nacional.

Las técnicas descritas en esta norma están basadas, en gran parte, en principios generales, aplicables a cualquier campo temático.

Sin embargo, se reconoce que los indizadores que trabajan en campos temáticos específicos necesitan apartarse ocasionalmente de estas recomendaciones generales, y esto se indica en los lugares apropiados.

En lo posible, las técnicas descritas en esta norma no se limitan a un método determinado de indización, sea pre o poscoordinado. Esta norma está sujeta, no obstante, a las siguientes restricciones:

- Trata de la presentación y organización de los términos que constituyen un subconjunto controlado del lenguaje natural. No sugiere procedimientos, organización o presentación de fórmulas matemáticas o químicas.
- Se basa generalmente en el concepto de “términos preferentes”.

- Su aplicación se limita a los centros que emplean indizadores humanos para analizar documentos y expresar su contenido mediante un lenguaje de indización controlado. No es aplicable en aquellas unidades que emplean técnicas de indización totalmente automáticas, en los que los términos que aparecen en los textos se organizan en conjuntos, según criterios, que puede establecer una computadora a partir de un algoritmo, por ejemplo, por su frecuencia de aparición o por su proximidad en el texto. Se considera, sin embargo, que un tesoro monolingüe bien construido puede ser de utilidad cuando se consultan sistemas de indización de texto libre.
- Trata principalmente de procedimientos para indizar colecciones de documentos incluidos en catálogos o bibliografías. No se ocupa de la preparación de índices de contenido para libros, aunque algunos de los procedimientos recomendados puedan resultar útiles a este propósito.

La norma en cuestión contiene recomendaciones relativas a tesauros monolingües, sin referencia a los requisitos propios de los tesauros multilingües, es decir, aquellos tesauros en los que las equivalencias conceptuales se expresan mediante términos seleccionados a partir de más de un lenguaje natural. La norma UNE 50-125 trata de la construcción y mantenimiento de estos tesauros multilingües.

Por motivo de que esta norma se basa en principios independientes, tanto cultural como lingüísticamente, dichos principios se han aceptado como bases de la norma para tesauros multilingües generales, aplicables indistintamente a los dos tipos de tesauros, que se describen en extenso solo en esta norma nacional, y no se repiten en la norma UNE 50-125 (Oficina Nacional de Normalización. Norma NC-ISO 2788: 2000. Directrices para el establecimiento y desarrollo de tesauros monolingües. La Habana: Oficina Nacional de Normalización; 2000).

En la norma se ofrecen las definiciones de documento, lenguaje de indización, tesoro, término de indización, término preferente, término no preferente, término compuesto e indicador clasificatorio; se aclaran las abreviaturas y símbolos utilizados en dicha norma y se exponen los dos métodos principales que se emplean en los tesauros para controlar al vocabulario.

Los conceptos representados por los términos de indización pueden pertenecer a las categorías siguientes: entes concretos y entes abstractos; entre estos últimos se encuentran:

1. Actividades y acontecimientos.
2. Entidades abstractas y propiedades de los objetos, los materiales o las acciones.
3. Disciplinas o ciencias.
4. Unidades de medida.

Se aclara que el creador de un tesoro necesita conocer la existencia de estas clases, porque afectan a algunos de los procedimientos descritos más adelante, por ejemplo, la elección entre las formas singular y plural y la aplicación de la prueba de relación jerárquica. Se añade que los términos de indización deben ser preferentemente sustantivos a frases sustantivadas.

Además, se tratan los hológrafos o términos polisémicos, la elección de los términos, la ortografía, los extranjerismos y barbarismos, la transliteración, los términos de jerga y argot, las denominaciones comunes y comerciales, las denominaciones coloquiales y científicas, los topónimos, los nombres propios de instituciones y personas, las notas de aplicación y definiciones, los términos compuestos, los términos que deben mantenerse compuestos, los términos que deben fraccionarse sintácticamente, el orden de las palabras en los términos compuestos, la presentación sistemática de un tesoro, así como la organización primaria de un tesoro sistemático, la incorporación y la supresión de términos.

Por último, se exponen aspectos relacionados con cuestiones generales como la notificación del proyecto, la eliminación de la duplicación de esfuerzos, el depósito en centros de referencia y los procedimientos de revisión.

Procesado por los licenciados *Javier Santovenia Díaz* y *Rubén Cañedo Andalia*