

EL DISEÑO Y LA GESTIÓN DE LAS BASES DE DATOS TERMINOLÓGICAS: UN DESAFÍO PERMANENTE

Jacint BOFIAS i ALBERCH
Ariadna PUIGGENÉ i RIERA

TERMCAT. Centre de Terminologia

Desde que la informática hizo posible el almacenamiento de datos lingüísticos y posteriormente su tratamiento, la terminología no ha cesado de adaptarse a las innovaciones tecnológicas, lo que ha supuesto y supone aún hoy un desafío permanente para el trabajo conjunto de terminólogos e informáticos. En esta comunicación apuntaremos algunos de los problemas que surgen al abordar el diseño y la gestión de las bases de datos terminológicos y expondremos el enfoque que hemos adoptado en el centro de terminología TERMCAT.

Inicialmente la terminología se sirvió de la informática como un medio para almacenar la información contenida en las fichas terminológicas y para facilitar algunos de los procesos de elaboración de diccionarios: ordenación, numeración, búsqueda, actualización e incluso edición de las fichas terminológicas. De ahí que los primeros repertorios terminológicos en soporte electrónico fueran de hecho una reproducción fiel de las fichas manuales.

Posteriormente se empezaron a utilizar algunas herramientas concebidas con otras finalidades, tales como programas de gestión documental, gestores de bases de datos, macros aplicadas a procesadores de textos, etc., adaptándolas en la medida de lo posible a las necesidades concretas del trabajo terminológico.

Con el avance de los sistemas de gestión de bases de datos (SGBD) surgieron los primeros programas específicamente diseñados para gestionar datos terminológicos, que han ido mejorando sus prestaciones a medida que la informática ha evolucionado. Estos gestores permiten tratar las distintas categorías de datos de una ficha terminológica como unidades de información independientes, y en su mayoría han sido concebidos como herramientas de ayuda a la traducción que facilitan tanto la consulta de diccionarios temáticos como su gestión.¹

El reciente desarrollo de las industrias de la lengua ha puesto de relieve el papel clave de la terminología en la transmisión del conocimiento. Un número creciente de productos y aplicaciones como los verificadores ortográficos, los programas de traducción automática o los sistemas expertos, precisan terminología específica para aumentar su eficacia y precisión. Y cada uno de ellos requiere que los términos se presenten acompañados de informaciones distintas.

¹ Véase *Terminómetro*, núm. 16. Especial Terminótica.

Nos hallamos actualmente ante la doble necesidad de contar con estructuras de datos suficientemente flexibles y adaptables de modo que sea posible incorporar, combinar y reorganizar dinámicamente los datos, y de desarrollar programas que nos permitan gestionar los datos de acuerdo con esta filosofía.

Algunas preguntas clave

El diseño y la gestión de las bases de datos terminológicos plantean cuestiones tanto de índole teórica y metodológica como de índole informática, algunas de las cuales están todavía por resolver.

¿Cómo se representa un término? ¿Cuál es la información mínima necesaria que nos permite conformar una ficha terminológica? ¿Y la máxima?

Una ficha terminológica no deja de ser una representación arbitraria de un concepto basada en un modelo teórico y metodológico determinado. El mismo hecho de diseñar una ficha supone una selección y una agrupación de datos que tendrá consecuencias, tanto informáticas como terminológicas. No obstante, existe un cierto consenso a la hora de distinguir las informaciones consideradas esenciales en la constitución de una ficha, aquellas que por sí mismas pueden representar o identificar un concepto (una denominación, una definición, un contexto, una imagen, un sonido, un símbolo), del resto de informaciones. Estas últimas son informaciones complementarias que varían según las aplicaciones a que se prevea destinar la terminología y pueden llegar a ser tantas como el sistema y el equipo humano sean capaces de gestionar.

¿Cómo se clasifican los términos? ¿Deben ser gestionados por áreas temáticas o por los diccionarios de los que forman parte?

La clasificación de los términos y su agrupación para constituir un producto terminológico responden siempre a unas necesidades concretas y a una percepción determinada de la realidad, incluso cuando se trata de una base de datos de alcance general. Quizá sea ésta una de las razones por las que no existe hoy una clasificación universal o de aplicación generalizada para la terminología o la documentación. Existe, sin embargo, un modelo que pretende la clasificación jerárquica de los términos en clases y subclases, que se subdividen sucesivamente hasta alcanzar la unidad. Algunos grupos de investigadores lo han aplicado en algunas áreas o disciplinas obteniendo como resultado sistemas de clasificación y jerarquización prácticamente impecables. Probablemente uno de los éxitos de su aplicación es que se ciernen a un conjunto cerrado de conceptos de áreas no susceptibles de ser interpretadas desde puntos de vista distintos, como por ejemplo las nomenclaturas de botánica o de zoología. La generalización de tal modelo a la totalidad de los ámbitos del conocimiento no parece posible hasta ahora.

¿Qué pretendemos obtener de la base de datos? ¿Qué funciones queremos llevar a cabo? ¿Cuáles en el futuro?

El productor de bases de datos terminológicos también se enfrenta a la exigencia de definir de un modo exhaustivo los resultados que querrá obtener a partir de la base de datos así como las funciones de gestión que deberá llevar a cabo. No se trata únicamente de no olvidarse de ningún detalle, sino de prever posibilidades aún inexistentes para evitar la creación de barreras difícilmente franqueables y garantizar un cierto futuro a la aplicación.

¿Cuál debería ser la estructura de una base de datos terminológicos? ¿Qué elementos debería contemplar?

La estructura interna de los datos en distintas tablas y las relaciones que se establecen entre ellas viene determinada tanto por los datos que conformen la representación mínima y

máxima de un concepto como por las funciones de tratamiento de los datos que se decida implementar. La necesidad de optimizar las prestaciones en la gestión de los datos, sobretudo los tiempos de respuesta, puede conllevar ciertas transgresiones en el diseño de las tablas desde un punto de vista informático y en la formulación teórica de la representación de los conceptos. Por lo tanto, habrá que sopesar las limitaciones, ambigüedades y problemas de coherencia tanto informática como terminológica para adelantar posibles soluciones.

Por otra parte, la rapidez de evolución del software y del hardware obligan a tener en cuenta otros factores que inciden de un modo directo en la capacidad de mantenimiento del programa y de los datos, y en su perdurabilidad por un plazo de tiempo razonable. Nos referimos, entre otros, al uso de formatos estándar, a la transportabilidad de los datos o al desarrollo de las herramientas en diferentes versiones (monopuesto, cliente-servidor, etc.).

No hay que olvidar que, una vez implementada una solución, los costes de rediseño y migración a nuevos entornos, junto con los costes de adaptación del personal y de renovación de equipos informáticos, son un obstáculo para llevar a cabo cualquier cambio.

Un enfoque práctico

En TERMCAT, después de unos años de trabajar en la elaboración de terminología, nos hemos enfrentado a la necesidad de rediseñar la base de datos terminológicos y los programas de gestión, para preparar y tratar la información de acuerdo con las nuevas necesidades de la terminología y con los objetivos de un centro productor de terminología.

A mediados de los años 80 partíamos de una aplicación diseñada a medida sobre un gran ordenador. Diez años más tarde fue migrada a un sistema con una arquitectura cliente-servidor sobre una red de área local utilizando un programa gestor de bases de datos estándar (Paradox). La migración se aprovechó para aplicar pequeñas modificaciones a la estructura de origen, pero se mantuvo la estructura inicial por razones tanto de índole económica como de tiempo y seguridad.

La reestructuración de los datos y el rediseño de los programas de gestión, que emprendimos en el momento en que el equipo informático se completó, se fundamentaron en dos ejes.

En primer lugar, definimos una estructura vertebrada alrededor de las unidades mínimas que pueden componer los datos terminológicos. Una estructura de este tipo permite el tratamiento de los propios datos y su reutilización para obtener diferentes productos terminológicos tales como diccionarios multilingües en formato papel aptos para la edición automática, diccionarios y léxicos en soporte electrónico, vocabularios especializados para complementar verificadores ortográficos, terminología para sistemas de traducción automática, etc. A su vez, la estructura está preparada para reutilizar recursos externos, es decir, para integrar datos con características y formatos distintos. Con esta concepción pretendíamos dotar al sistema de suficiente flexibilidad para gestionar los datos y asegurar su escalabilidad de cara a facilitar futuras ampliaciones.

En segundo lugar, los programas de gestión de los datos terminológicos se han concebido de forma modular, es decir, se componen de partes independientes tan claramente diferenciadas como ha sido posible: introducción de datos, verificación de su coherencia y validez, gestión, recuperación, exportación, etc.

“Les avancées du génie logiciel et le développement des langages de programmation permettent de penser, de plus en plus, les chaînes de traitement comme des assemblages de modules autonomes, c’est à dire, comme des “objets” ou “processus”

autonomes qu'on peut appeler individuellement pour réaliser des composantes des tâches complexes."²

Esta modularidad permite, por un lado, una adaptación más rápida y menos costosa a las nuevas necesidades que plantea el trabajo terminológico (la creación de interfaces específicos según los puestos de trabajo o para la gestión más eficaz de productos distintos, pueden mencionarse como ejemplos) y, por otro lado, facilita la adaptación a las nuevas versiones de software así como la adopción de las nuevas herramientas informáticas que van saliendo al mercado.

Los nuevos desafíos

Los medios tecnológicos continuarán modificando los procedimientos de trabajo de los terminólogos, el campo de los recursos que explotan y los modos de gestión de sus datos. Habrá, pues, que continuar encontrando las vías y los mecanismos necesarios para optimizar su producción y difusión. El contexto tecnológico actual, permanentemente cambiante, empuja a constantes revisiones tanto de la estructura de las bases de datos terminológicas, como de las estrategias definidas de producción y de difusión, e incluso del mercado potencial.

Debemos estar preparados para enfrentarnos al nivel de desarrollo de las nuevas tecnologías y no quedar al margen de las nuevas demandas terminológicas que generará la sociedad de la información, entre las cuales aparece de un modo inminente la integración de los distintos tipos de bases de datos lingüísticos (bases de datos lexicográficas, corpus escritos y orales, etc.). Esto significa estar dispuestos a modificar los contenidos, el hardware, el software y la concepción misma de los gestores de bases de datos terminológicos (GBDT).

Puesto que la potencia de los programas informáticos y de los ordenadores abren posibilidades casi ilimitadas, los nuevos desarrollos dependen ahora de la capacidad de los equipos de trabajo de lingüistas e informáticos de hallar soluciones eficaces para construir un modelo metodológicamente e informáticamente consistente, válido por un plazo de tiempo razonable y adaptado a los usuarios del futuro.

LABURPENA / RESUMEN / RÉSUMÉ / ABSTRACT

Terminologiako datu-baseak diseinatzea eta kudeatzea: etengabeko erronka

Informatikak hizkuntzako datuak gorde eta, aurrerago, tratatzeko aukera eman zigunetik, terminologiak teknologi berrikuntzetara egokitu behar izan du etengabe. Horrek amaieragabeko erronka ekarri izan dio -eta oraindik ere hala dakarkio- terminologo eta informatikoen elkarlanari.

² MEUNIER, J.-G. "Vers la fiche virtuelle en terminologie". *Terminologies nouvelles* (1996), núm. 15, p.151-153.

Hala da, terminologiako datu-baseak diseinatzeak eta kudeatzeak hainbat arazo dakar bai teoria eta metodologiaren aldetik (nola adierazi terminoa? Ba al da gutxieneko adierazpenik? Nola sailkatu terminoak?), eta baita informatikaren aldetik ere (zein egitura eduki behar luke terminologiako datu-base batek? Zein erlazio jaso behar lituzke? Zein da hiztegi bateko unitatea?).

Benetako testuinguruetan ondo aplikatzeko balioko duten konponbideak aurkitu beharrak ikuspuntu jakin bat hartzera eraman gaitu, informatikako postulatuak eta metodologiak begi aurretik galdu gabe bada ere, arazo horiei aurre egin, terminologi lana bete eta emaitzak zabaltzeko aukera izan dezagun.

El diseño y la gestión de las bases de datos terminológicos: un desafío permanente

Desde que la informática hizo posible el almacenamiento de datos lingüísticos y posteriormente su tratamiento, la terminología no ha cesado de adaptarse a las innovaciones tecnológicas, lo que ha supuesto y supone aún hoy un desafío permanente para el trabajo conjunto de terminólogos e informáticos.

Efectivamente, el diseño y la gestión de las bases de datos terminológicos plantean problemas diversos tanto de índole teórica y metodológica (¿cómo se representa un término?, ¿existe una representación mínima?, ¿cómo se clasifican los términos?) como de índole informática (¿cuál debería ser la estructura de una base de datos terminológicos?, ¿qué relaciones debe contemplar?, ¿cuál es la unidad de un diccionario?).

La necesidad de encontrar soluciones que se puedan aplicar en un contexto real de un modo efectivo nos ha obligado a adoptar un determinado enfoque que, sin perder de vista los postulados teóricos y las distintas metodologías de la terminología y la informática, nos permita afrontar dichos problemas y llevar a cabo el trabajo terminológico y la difusión de sus resultados.

La conception et la gestion des bases de données terminologiques: un défi constant

Depuis que l'informatique a rendu possible le stockage de données linguistiques et ultérieurement leur traitement, la terminologie n'a pas cessé de devoir s'adapter aux innovations technologiques, ce qui a supposé, et suppose aujourd'hui encore un défi constant pour le travail conjoint des terminologues et des informaticiens.

Effectivement, la conception et la gestion des bases de données terminologiques présentent des problèmes divers aussi bien de caractère théorique et méthodologique (comment se présente un terme?, une représentation minimale existe-elle?, comment les termes sont-ils classifiés?) que de caractère informatique (Quelle devrait être la structure d'une base de données terminologiques?, quelles sont les relations qu'elle doit envisager? quelle est l'unité d'un dictionnaire?).

Le besoin de trouver des solutions pouvant être appliquées dans un contexte réel d'une façon effective nous a obligé à adopter une optique déterminée qui, sans perdre de vue les postulats théoriques et les différentes méthodologies de la terminologie et de l'informatique, nous permette d'affronter ces problèmes et de mener à bien le travail terminologique ainsi que la diffusion de ses résultats.

Design and management of terminological databases: a constant challenge

Ever since information technology first made it possible to store and then process linguistic data, terminology has had to adapt constantly to technological innovations. This has posed, and indeed continues to pose, a constant challenge for the joint work of terminologists and software specialists.

The design and management of terminological databases pose theoretical and methodological problems (how can a term be represented? Is there a minimum representation? How are terms to be classified?), as well as IT problems (How should terminological databases be structured? What relationships should be covered? What is a dictionary unit?).

The need to find solutions which can be applied effectively in a real context has led us to take an approach which does not ignore theoretical postulates and the different methods of terminology and IT, but enables us to handle problems, carry out terminological work and disseminate the results.