Levante
EL MERCANTIL VALENCIANO
www.levante-emv.com

Diccionarios, bases de datos y cerebro

LAS CIENTÍFICAS CUENTAN

María Amparo Alcina Caudets
TecnoLettra, Universitat Jaume I

Los diccionarios terminológicos constituyen un recurso fundamental para diversos ámbitos profesionales del lenguaje y para las diversas disciplinas. Contienen el vocabulario, los conceptos con los que las personas se comunican. En ocasiones, se trata de palabras de uso habitual en el lenguage común, pero que toman un significado especializado en un área de conocimiento determinada; también puede tratarse de palabras solo conocidas por los expertos e incluso creadas específicamente para denominar conceptos altamente especializados. Por ejemplo, en el ámbito de la cerámica industrial en el que hemos trabajado en colaboración con nuestro entorno en Castellón, tendríamos palabras conocidas como baldosa o esmalte, o más específicas como frita. En cualquier caso, estas palabras, y a menudo expresiones de más de una palabra, reciben la denominación de términos.

En un diccionario impreso, cada entrada incluye el término, su definición y otras informaciones lingüísticas, como su categoría gramatical, género o número, que permiten conocer cómo usar la palabra en una oración. Los diccionarios bilingües o multilingües incluyen además equivalencias de traducción a otras lenguas.

Con la aparición de los ordenadores, la información de los términos comenzó a almacenarse en bases de datos terminológicas. La estructura interna de estas bases se asemeja a una tabla con filas y columnas. Cada fila o registro recoge la información relativa a un elemento de información, por ejemplo, una palabra o término. Cada columna o campo recoge un tipo de información, por ejemplo, definición, categoria gramatical, género, número, equivalencia de traducción. En el cruce entre fila y columna hallaremos el valor que corresponde a una palabra respecto a un tipo de información. Así, en el cruce entre la fila del término baldosa y la columna de información categoría gramatical, tendríamos el valor sustantivo. Esta estructuración de los datos permite al ordenador encontrar el dato buscado y devolver una respuesta concreta.

En las últimas décadas, estas bases de datos han mejorado en cuanto a aspectos técnicos (capacidad de almacenamiento, interfaces más amigables), la interacción con otros sistemas (extracción automática de términos, traducción asistida y automática) y en cuanto al desarrollo de normas y estándares que facilitan el intercambio de datos terminológicos. Su uso se ha popularizado ampliamente en el ámbito de la traducción y entidades dedicadas a la terminología. Contamos con bancos terminológicos como IATE (Unión Europea) o Termium (Gobierno de Canadá) disponibles en Intenet.

Las bases de datos, sin embargo, no permiten representar adecuadamente muchos de los datos interesantes sobre los términos, por ejemplo, sus relaciones semánticas. Por ejemplo, el concepto de mosaico contiene las características semánticas del concepto baldosa y añade otras más específicas, como el tamaño más pequeño; decimos entonces que mosaico es un tipo de baldosa o que mosaico tiene una relación semántica de hiponimia respecto a baldosa.

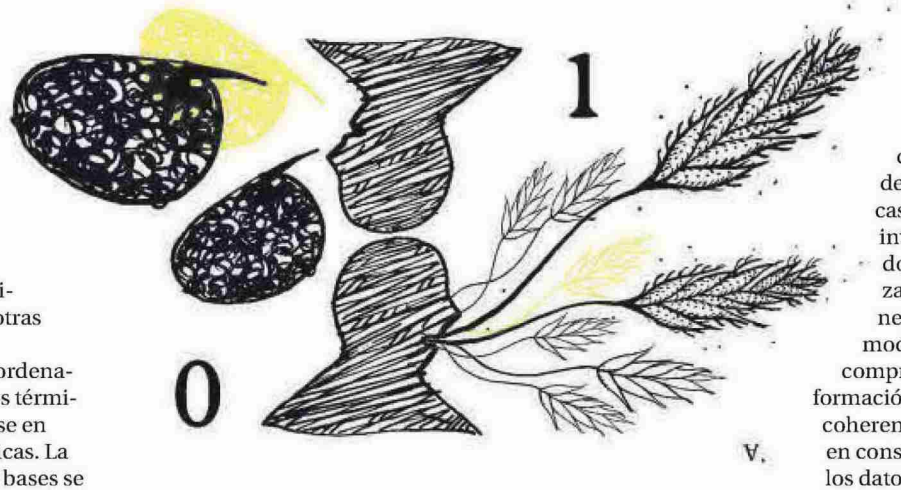
Con el fin de representar las relaciones semánticas facilitando el acceso y la navegación entre los términos según su significado (lo que llamamos consulta onomasiológica), se empezaron a utilizar redes o taxonomías, estructuras más flexibles que la tabla. Contamos así con recursos léxicos como WordNet y EuroWordNet y recursos terminológicos como EcoLexicon. Estos sistemas no carecen de problemas. Por ejemplo, dada la existencia de los fenómenos de polisemia (términos con más de un significado) o sinonimia (un significado se expresa mediante distintos términos), la navegación a través de la red puede llegar a vincular términos que tengan poco o nada en común. Esto se complica si entran en juego equivalencias de los términos en otras lenguas.

Con el objetivo de conseguir recursos digitales que permitan superar la rigidez de las bases de datos y contener la expansión de las taxonomías y redes, en los proyectos ONTODIC del grupo de investigación Tecnolettra hemos investigado la metodología y técnicas más recientes de la Ingeniería del conocimiento. En concreto, hemos estudiado los lenguajes ontológicos como OWL, que resultan muy potentes para expresar relaciones de todo tipo entre los datos y razonar eficazmente utilizando la lógica descriptiva.

Con ayuda del editor de ontologías Protégé (Universidad de Stanford), hemos diseñado un metamodelo para organizar en la ontología los datos lingüísticos, su clasificación y sus relaciones de todo tipo, no solo semánticas. Uno de los aspectos más interesantes es que el sistema, dotado de un razonador, realiza clasificaciones y deducciones automáticas a partir del modelo y los datos aportados, y comprueba la coherencia de la información representada. Si halla incoherencias, nos avisa para actuar en consecuencia, bien corrigiendo los datos o corrigiendo el modelo.

En próximos proyectos, pretendemos desarrollar estos aspectos, no solo en lo que respecta a obtener recursos digitales más flexibles y eficaces. Nos preguntamos si la representación que obtenemos de las conexiones entre los datos lingüísticos, organizados según la lógica, y cuya visualización se asemeja a redes neuronales, guarda relación con las conexiones que se producen en el cerebro, del cual el lenguaje es, al fin y al cabo, su producto más tangible y genuino.

ILUSTRACIÓN DE **ANDREA CORRALES**



Con la aparición de los ordenadores, la información de los términos comenzó a almacenarse en bases de datos terminológicas. La estructura interna de estas bases se asemeja a una tabla con filas y columnas.