



Servicios Lingüísticos SL  
<http://www.star-transit.com/es>  
<http://www.star-spain.com/>

## **TECNOLOGÍA PARA LA GESTIÓN TERMINOLÓGICA MULTILINGÜE: TERMSTAR Y WEBTERM**

**LIDIA CÁMARA**

*STAR Servicios Lingüísticos*

### **LA GESTIÓN TERMINOLÓGICA**

Cada vez se hace más acuciante la necesidad de gestionar de una forma optimizada los recursos lingüísticos y muy especialmente los recursos terminológicos multilingües (RTM). Los ámbitos de uso y aplicación de los RTM en la disciplina de la Ingeniería Lingüística (IL) son el reflejo de la importancia que se concede a la creación, organización, mantenimiento y manipulación en este entorno. Los sistemas expertos en su primera fase de concepción y diseño, los sistemas que funcionan como prototipo y los sistemas ya comercializados relacionados con los procesos implicados en la manipulación de recursos lingüísticos requieren de bases de datos terminológicas para la optimización de los diferentes flujos de trabajo. Para ello se debe partir de unos criterios cuyo intento de formalización dan cuenta de la complejidad que entrañan. Inspirados en parte del catálogo de problemas asociados a la gestión terminológica que proponen Budin y Melby (2000) y los suscitados por nuestra propia experiencia, exponemos a continuación los más relevantes para nosotros y las posibles soluciones técnicas si las hay.

#### **1. Identificación y detección de términos**

Aparte de las divergencias de identificación de términos originadas por la decisión de si una pieza léxica o varias constituyen un término, cuya motivación responde a planteamientos teóricos y funcionales dispares, la detección de los términos ya identificados puede originar problemas. Si bien no se pone en duda que las entradas terminológicas deben responder a formas canónicas en los diccionarios orientados a humanos, se deben tener en cuenta las dificultades que pueden generar las formas flexionadas en el momento de detectar esos términos. La incorporación de un

analizador (*parser*) morfológico para cada lengua -como lo hace TermStar- soluciona el problema de la detección de términos flexionados.

## **2. Exhaustividad**

Debido a la propia dinámica de la lengua, no podemos tener bancos de datos terminológicos exhaustivos, ni aun cuando delimitemos el dominio o subdominio conceptual. Ahora bien, si contamos con una implementación con la suficiente flexibilidad y accesibilidad que ofrezca la posibilidad de retroalimentar las bases de datos constantemente mediante sugerencias internas y externas con una plataforma de bases de datos del tipo ODBC, que hace posible el entendimiento con diferentes plataformas (SQL, Oracle, Access, etc.), posibilitaremos la actualización inmediata de los datos y un acceso inmediato a los mismos. TermStar, junto con WebTerm nos ofrece estas ventajas. Mediante el cuadro de comentarios en la interfaz vía Internet, obtenemos comunicación inmediata, con sugerencias a candidatos a términos, correcciones en la definición de los términos, propuesta para equivalencias en otras lenguas, variaciones corporativas, regionales, prescriptivas, etc. La actualización de los datos no requiere ninguna formación en ingeniería. Al aceptar XML, WebTerm, a través de su interfaz independiente de plataformas, proporciona el intercambio de datos con TermStar en cualquier punto del mundo. La ubicación del gestor de terminología es completamente independiente de la ubicación del servidor. Además, gracias al uso de formatos estándar, como HTML/XML, servlets de Java y de JDBC, WebTerm ofrece la libertad necesaria para diseñar un sistema propio sin la obligación de utilizar bases de datos o servidores web específicos. (Scholand: 2001)

## **3. Clasificaciones por categoría**

Para poder operar no sólo con nuestros datos, sino posibilitar también tanto la reutilización en un ámbito diferente al de éstos como su intercambio, se precisan de estándares que unifiquen las etiquetas de marcación para las categorías gramaticales, género, número, etc. Estas categorizaciones, que quizá a priori no deberían representar demasiada dificultad, responden a criterios de concepción y aplicación de entornos diferentes. De ahí las discrepancias. TermStar permite introducir listas de valores predeterminadas para garantizar la unificación y homogeneidad del almacenamiento.

## **4. Clasificaciones ontológicas**

Casi todas las aplicaciones de la IL y muchas del ámbito del procesamiento del lenguaje natural (PLN) y de la Tecnología de la Información (TI) en general precisan de la formalización de descriptores o identificadores. La tarea no tiene fácil solución. En primer lugar, porque no todos ordenamos nuestro mundo conceptual de igual manera, y mucho menos aún en un entorno multilingüe. Y, en segundo lugar, porque existen muchos grupos diferentes con fines específicos también discrepantes, aunque se reaprovechen recíprocamente de los RTM de grupos vinculados directamente a su área de especialidad, o vinculados de forma indirecta por establecer otro tipo de relaciones, por ejemplo, comerciales. Por este motivo aunque muchos organismos pueden hacer uso de tesauros convencionales de carácter universal, necesitan además un catálogo de descriptores específicos que cubran sus necesidades. TermStar permite introducir una estructura jerárquica de hasta ocho niveles. Estas relaciones conectan los descriptores de mayor rango a menor, de genérico a específico. Automáticamente conecta también determinadas relaciones meronímicas vinculadas a un término. El sistema permite, además, establecer relaciones semánticas asociativas por afinidad, por ser una relación instrumental o por oposición al propio término.

## 5. Usuarios finales

Para que diferentes usuarios puedan acceder a los RTM de una forma económica hay que tener en cuenta sus diferentes necesidades. Así como ya es normal encontrar filtros en los bancos de datos terminológicos que discriminan los conceptos en función de la combinación de lenguas, así como en función del dominio conceptual al que se deseen asociar durante el proceso de búsqueda, todavía no existe un banco de datos que haya tenido en cuenta el usuario al que puede ir destinado. TermStar permite establecer vistas pertinentes según el usuario. De este modo, a la información que necesita un documentalista se le asignará una vista determinada, diferente a las que se asignen al terminólogo o al traductor. De igual manera, mediante Webterm se pueden hacer búsquedas no sólo por combinación de lenguas o por dominios conceptuales, sino también por predeterminación de usuario.

## 6. Modelo de bases de datos y formatos

El modelo de base de datos relacional es para unos el idóneo; para otros, es el de base de datos orientada a objetos para la producción de un diseño modular de componentes cerrados, que se pueden reutilizar una y otra vez, y modificar sin afectar a los demás. Más allá de elegir un modelo u otro, lo principal es que éste sea flexible y que utilice formatos independientes del fabricante de los productos utilizados para el almacenamiento de datos. Además, estos formatos deben permitir el intercambio y la integración en otras aplicaciones para que los recursos puedan integrarse en otros sistemas de trabajo. Las herramientas que nos ocupan admiten la integración y el intercambio de datos en XML. Así se puede integrar la base de datos en otras aplicaciones para que los sistemas de gestión de contenidos, por ejemplo, tengan acceso a la misma terminología desde cualquier lugar del mundo.

Los principales representantes de la IL cuentan con el consorcio LISA (*Localisation Industry Standards Association*). Allí se dan cita los fabricantes y usuarios más importantes vinculados a la TI y la IL para llegar a un consenso que agrupe las necesidades actuales para el intercambio de recursos lingüísticos. Para el intercambio de RTM se ha adoptado la especificación MARTIF (*Machine-Readable Terminology Interchange Format*), el formato estándar (ISO 12200) para el intercambio de bases de datos terminológicas. Se trata de un conjunto de etiquetas específicas con funciones determinadas escritas mediante XML. Estas etiquetas con funciones específicas para un grupo de interés se denominan *Definición de Tipo de Documento* (DTD), (Cámara: 2001). Conscientes de la necesidad de trabajar con bases de datos terminológicas que integren no sólo términos, sino también lexemas para preparar los sistemas que procesan, y que se alimentan de recursos terminológicos y lexicográficos genéricos, como los sistemas de traducción asistida y los sistemas de traducción automática, los miembros del consorcio LISA están dedicando sus esfuerzos al establecimiento de un formato de intercambio de bases de datos terminológicas, denominado TBX (*Term Base Exchange*). Esta DTD, también escrita en XML, reúne la DTD estándar MARTIF para el intercambio de bases de datos terminológicas orientadas a humanos, en la que se inscribe el formato OLIF (*Open Lexicon Interchange Format*). Esta DTD nació del proyecto Otelos con el fin de convertirse en estándar para el intercambio de bases de datos de lexicones orientados a la traducción automática, (Warburton: 2000): . TermStar y WebTerm aceptan el formato estándar MARTIF y sus desarrolladores trabajan también con el consorcio LISA en la integración del formato OLIF.

## LOS RETOS EN LA GESTIÓN DE LA TERMINOLOGÍA

Desarrollos tecnológicos como los aquí presentados están preparados para asumir los retos que plantea la gestión de RTM. Sin embargo, éstos no son la panacea que dan respuesta a todos los problemas planteados, pero sí que permiten el almacenamiento sostenible de los recursos lingüísticos con los que contamos.

La falta de consenso para establecer los criterios de procesamiento requiere de un esfuerzo común que debe llevarse a cabo en varias etapas:

### a) Estandarización técnica

Actualizar las bases y bancos de datos terminológicos para su estandarización, lo que requiere el uso consecuente de los formatos de intercambio establecidos por la Industria de la lengua y correr el riesgo de pérdida de datos, especialmente aquellos codificados en formatos no susceptibles de ser manipulados.

### b) Acuerdos metodológicos

Llegar a acuerdos de método que unifiquen criterios de catalogación y clasificación de los datos asociados a un término. Si bien la armonización de la catalogación de información relacionada con la categoría gramatical, el género, la variante lingüística, la transcripción fonética, el grado de validación y aceptación, etc. de los términos podría alcanzar un consenso a corto y medio plazo, la clasificación mediante ontologías parece casi una ilusión.

### c) Publicación de criterios adoptados

Conseguir el compromiso de los diferentes profesionales vinculados muy estrechamente a los RTM. Todos ellos, usuarios no sólo de estamentos públicos sino de empresas dedicadas a todas aquellas actividades vinculadas a la creación, consulta, manipulación y gestión en general de RTM, deben comprometerse a compartir este conocimiento; es decir, a compartir criterios de catalogación y clasificación haciéndolos públicos para que toda la comunidad pueda aprovecharse y mejorar paulatinamente esos criterios.

## CONCLUSIONES

Somos conscientes de las enormes dificultades que entraña todo el proceso que planteamos, tan complejo que cabe pensar que se trata de un reto insalvable. No obstante, creemos que el esfuerzo continuado y compartido da frutos y muchos. Aprovechamos esta ocasión para animar a las partes implicadas que todavía no lo hayan hecho a reflexionar sobre las posibles implementaciones tecnológicas y sobre el reaprovechamiento y circularidad de los datos elaborados.

## BIBLIOGRAFÍA

BUDIN, G y A. MELBY (2000): «Accessibility of Multilingual Terminological Resources —Current Problems and Prospects for the Future» [en línea] en *Translation, Theory, and Technology*, <[http://www.ttt.org/Salt/Athens\\_Paper.doc](http://www.ttt.org/Salt/Athens_Paper.doc)>

- CÁMARA, L. (2001): «El papel de las herramientas TAO en la documentación técnica multilingüe» [en línea], *Tradumática*, 0, <http://www.fti.uab.es/tradumatica/revista/articles/lcamara/art.htm>
- SCHOLAND, M. (2001): «TermStar/WebTerm» [en línea] en *Star Language Technology*, <http://www.star-transit.com/products/termstar/es/>.
- WARBURTON, K. (2000): «LISA Terminology Survey Report» [en línea] en *LISA (Localisation Industry Standards Association)*, <<http://www.lisa.org/2001/termsurveyresults.html>>.