

Representación de la información terminológica en OntoTerm®: un sistema gestor de bases de datos terminológicas basado en el conocimiento

Antonio Moreno Ortiz
amo@uma.es
Universidad de Málaga

1 Introducción

Además de extraer y recopilar la información terminológica, la misión fundamental del terminógrafo es la de representar, es decir, hacer explícita, la información terminológica del ámbito de especialidad que le ocupa. En la realidad cotidiana, esta representación puede ser tan detallada o tan básica como exija el trabajo en curso o la finalidad última que se persigue con la recopilación terminológica:

Representation is about making things explicit, about resolving ambiguities and above all, particularly in the context of artificial intelligence, about creating a surrogate of a class of things that exist in the real world on a computer system.

POINTER Report (sección 4: 97)

La forma en la que se representa la información recopilada por los terminógrafos es un aspecto vital de la gestión terminológica y puede también condicionar los usos (y usuarios) de un repositorio terminológico. En este artículo mostramos las herramientas necesarias para integrar la representación del conocimiento terminológico en el ámbito general de la representación del conocimiento (*Knowledge Representation: KR*), rama de la Ingeniería del Conocimiento y la Inteligencia Artificial, en la que se han desarrollado mecanismos formales y altamente estructurados de representación del conocimiento. Pensamos que la adopción de estos mecanismos de representación beneficia en gran medida la calidad de los repositorios terminológicos y puede servir para plasmar de forma explícita la orientación onomasiológica de la terminografía. En esta línea, proponemos uno de estos mecanismos de representación, una *ontología de conceptos*, como medio para representar el conocimiento que el terminógrafo adquiere sobre el área de especialidad de modo formalizado y reutilizable. Además, la estructuración del dominio de especialidad por medio de una ontología de conceptos permite que se haga patente la perspectiva desde la que se aborda el trabajo terminográfico y los criterios seguidos en el establecimiento de jerarquías, relaciones conceptuales y atributos asignados a los conceptos.

Mostraremos a continuación el sistema gestor de base de datos terminológica basada en el conocimiento que hemos desarrollado, denominado OntoTerm®, que ha permitido a los miembros del proyecto OncoTerm introducir, modificar y consultar la ontología, así como su posterior desarrollo¹. En este sistema la representación de información atañe a los tres aspectos fundamentales de las unidades terminológicas. Es decir, se representa de forma explícita:

- información conceptual del ámbito de especialidad;
- información lingüística relevante a la unidad terminológica;
- información sobre el uso que los especialistas del área hacen de los términos, tal y como se desprende del estudio de los textos que éstos producen o por medio de consultas específicas con especialistas.

Antonio Moreno Ortiz 62//00 9:41 PM
Con formato: Numeración y viñetas

Este sistema gestor se compone de dos módulos principales: el editor de ontologías y el editor de bases de datos terminológicas. El primer módulo permite la representación de la información conceptual, mientras que en el segundo se representa la información terminológica de carácter lingüístico, de uso y administrativa. Este sistema, además, permite que la información representada cumpla dos características que nos parecen esenciales para una buena gestión terminológica:

- estar organizada de forma sistemática;
- contar con un formato estándar de representación, que asegure, no sólo la reutilización y el intercambio de la información, sino también la consistencia y coherencia interna de los datos.

Antonio Moreno Ortiz 62//00 9:41 PM
Con formato: Numeración y viñetas

2 Resituando el concepto de *concepto*

El papel central de los conceptos en la gestión terminológica no admite hoy en día discusión alguna, lo que viene siendo así al menos desde Wüster: “all modern work on the problems of terminology takes as its starting point the *concept* itself” (Wüster 1968: xii). Sin embargo, cuando se dice que el análisis terminológico debe siempre partir del concepto estamos dando por supuesto que sabemos lo que es un *concepto*. Pero ocurre que el concepto de *concepto* es, posiblemente, uno de los más escurridizos y difíciles de definir, aunque deberíamos tener una

¹ Información detallada sobre este sistema gestor de base de datos puede encontrarse en su sede web: <http://www.ontoterm.com>.

noción muy clara de lo que es, pues de otro modo estaríamos construyendo un gigante con pies de barro, al basar toda una disciplina en algo que no se sabe muy bien lo que es.

Sager (1990: 23) recoge las siguientes definiciones de *concepto* que varios comités estandarizadores habían formulado hasta la fecha:

- Los conceptos son constructos mentales, abstracciones que se pueden emplear para clasificar los distintos objetos del mundo exterior e interior. (Recomendación Estándar Británica para la selección, formación y definición de términos técnicos).
- Los objetos de todos los campos de conocimiento y actividades humanas, las cosas, sus propiedades, cualidades, fenómenos, etc. se representan mediante conceptos. (Propuesta de revisión del Reino Unido para el documento de la ISO R 704).
- Un concepto es un constructo mental para la clasificación de objetos individuales del mundo exterior e interior por medio de una abstracción más o menos arbitraria. (Borrador de 1968 del estándar ISO 704).
- Un concepto es una unidad de pensamiento, generada mediante la agrupación de objetos individuales relacionados entre sí por características comunes. (Borrador de documento DIN alemán).
- Un concepto es un grupo coherente de juicios sobre un objeto cuyo núcleo se compone de aquellos juicios que reflejan las características inherentes del objeto. (Propuesta de la Unión Soviética para la revisión del documento ISO 704).
- Un concepto es una unidad de pensamiento.

Y añade las siguientes puntualizaciones:

- Un concepto se usa para estructurar el conocimiento y percepción del mundo circundante y no necesita ser expresado.
- Distintas escuelas de pensamiento tienen definiciones diferentes del concepto “concepto”. (Versión final del Draft International Standard ISO/DIS 704, 1985).

A continuación el autor propone que, dada la gran diversidad de opiniones, para los propósitos de la terminología, “[...] ‘concept’ be considered another axiomatic primitive, like ‘word’ or ‘sentence’, conveniently left undefined.”

Sin embargo, no pensamos que sea apropiado basar una metodología completa de trabajo en una primitiva axiomática, porque se corre el peligro de que el terminógrafo, aun siendo consciente de la premisa fundamental de su disciplina (partir del concepto para llegar al término), deje de pensar en estos términos, ya que, para empezar, no cuentan con una definición clara. Además, los creadores de herramientas informáticas, sin duda un aspecto clave en la gestión terminológica moderna, también deben contar con esta definición, de modo que sus productos se adapten a lo que los terminólogos esperan de ellos y no al contrario. Por ello, pensamos que es fundamental contar con una definición adecuada, cuya aplicación se adapte a las necesidades del terminógrafo, facilitándole su trabajo en lugar de obstaculizarlo.

Antonio Moreno Ortiz 62//00 9:41 PM
Con formato: Numeración y viñetas

En efecto, es muy difícil definir un concepto que, como se deduce de lo expuesto por Sager (1990: 23) tiene un carácter axiomático. Sin embargo, no estamos tratando de definir lo que es un concepto sin un contexto y aplicación concretos, lo cual nos llevaría a multitud de problemas de índole filosófica, cuya discusión queda fuera de nuestro ámbito de investigación. En realidad, lo que nos interesa es adoptar una definición que sea válida para nuestros propósitos, fundamentalmente la estructuración de un cuerpo de conocimiento de naturaleza abstracta que se manifiesta en forma de unidades léxicas en las diversas lenguas naturales.

Una característica común a todas estas definiciones es que son individualizadas, es decir, no tienen en cuenta lo que probablemente hace que un concepto sea tal: las *relaciones* con los demás conceptos. Efectivamente, si pensamos en cualquier campo de conocimiento humano, lo que hace que un determinado concepto sea distinguido de los demás son tanto sus propiedades distintivas como las relaciones que guarda con los demás conceptos. Estas relaciones, normalmente consideradas en terminología son por supuesto, las que, con diversos nombres, también se consideran en las ciencias cognitivas, la IA, la lingüística, etc., es decir, relaciones ISA, AKO, HAS-A, etc. En nuestra opinión, es imprescindible incluir el carácter relacional de los conceptos en su definición.

Otra característica a la que las definiciones recogidas por Sager no hacen mención, pero que es extremadamente relevante para la terminología, es el hecho de que un concepto lleva normalmente asignada una representación física concreta, normalmente una realización fónica u ortográfica con la que se manifiesta en una lengua determinada. En definitiva, esta realización es el resultado más palpable del trabajo terminológico, luego también debería formar parte de nuestra definición.

Según lo que hemos expuesto, por tanto, podríamos reformular, o mejor dicho resituarse el concepto de *concepto* en la siguiente definición provisional:

Un concepto es una abstracción de un conjunto de objetos, propiedades o eventos existentes en el mundo real o un mundo posible, que puede poseer una realización física en una lengua natural o sistema de representación determinados, al cual se puede hacer referencia mediante un símbolo arbitrario, aunque necesariamente único, dentro de un sistema representacional. Como constructo, posee ciertas propiedades distintivas de los demás conceptos, con los que guarda diversos tipos de relaciones. Tanto sus propiedades intrínsecas como sus relaciones con los demás conceptos deben ser evidentes, y por tanto susceptibles de ser especificados de forma explícita.

Esta definición, además de ser situada en el contexto de la terminología, incorpora algunas características que la hacen compatible con la concepción que se tiene en otras disciplinas, notablemente la inteligencia artificial y, dentro de ésta, la ingeniería del conocimiento. En primer lugar, incluye las conceptualizaciones de las propiedades y los eventos como conceptos (y no sólo los objetos), haciendo énfasis en la posibilidad de que sean representados y explicitados por medio de sus propiedades y sus relaciones. En efecto, hace referencia a “sistema representacional”, “propiedades” y “relaciones”, términos empleados desde hace mucho tiempo por los investigadores de la IA y que han servido como base para la implementación de sistemas informáticos cuya misión es la de representar y procesar conocimiento.

Posiblemente, la mayor diferencia con respecto a las definiciones que mostramos anteriormente venga dada por la última afirmación: la posibilidad de especificar de forma explícita (y formalizada) las propiedades y relaciones de los conceptos. El objeto de incluirla en nuestra definición es plasmar en ella nuestra concepción de lo que debe ser el tratamiento del sistema conceptual subyacente a toda base terminológica.

Los fabricantes de herramientas informáticas consideran que un sistema gestor de bases de datos terminológico está basado en el concepto si cumple la condición de que cada concepto corresponda a un sólo término, lo cual deja al terminógrafo la decisión de si esto será realmente así en su base terminológica. Sin embargo, en estos sistemas no se representan explícitamente dos aspectos determinantes para la diferenciación entre conceptos: las propiedades o características distintivas de los conceptos y las relaciones conceptuales. En los sistemas que permiten establecer relaciones entre conceptos, éstas no son más que punteros o enlaces a otros conceptos, sin que el tipo de relación sea considerada en sí misma un concepto y por tanto, sólo queda especificada por la definición en lenguaje natural que contiene el enlace, y por tanto necesariamente interpretable.

Podríamos decir que, aunque el terminólogo cumpla con la máxima de ir desde el concepto hasta el término, las herramientas que utiliza no imponen que esto deba ser así, ni permiten la representación de las propiedades de los conceptos y las relaciones entre ellos de una manera coherente, formalizada y recuperable. La siguiente cita del grupo de la Universidad de Ottawa resume muy bien la idea que estamos tratando de exponer aquí:

[...] we need to recognize that while the final results of terminology research may be *based* on knowledge, they are not systematically *encoded* as knowledge. Rather, most of the subject-field knowledge so laboriously acquired by the terminologist unfortunately stays where it was first stored, namely *in the terminologist's head*.

Meyer, Eck & Skuce (1997: 98)

Efectivamente, existe una gran diferencia entre simplemente dar por hecho que la base terminológica está basada en conceptos porque cada término o conjunto de términos equivalentes en varios idiomas se corresponden con un sólo concepto, y contar con un sistema conceptual complejo, estructurado y formalizado, es decir, el tipo de sistema representacional que, además de contener conocimiento sobre el ámbito en cuestión es capaz de extraer generalizaciones sobre el mismo, deducir información que no se encuentra plasmada de forma explícita y llevar a cabo otros procesos más o menos “inteligentes”.

3 Terminología basada en ontologías

En la actualidad, coexisten dos usos diferenciados del término *ontología*, que corresponden a dos ramas del saber diferentes y, por tanto, le atribuyen características y propiedades distintas. El término *ontología* se origina en el campo de la filosofía y la epistemología. Como ciencia, la Ontología es una rama de la metafísica que se ocupa del estudio de la naturaleza de la existencia, de los seres y de sus propiedades transcendentales; en filosofía, por tanto, una ontología se considera como una explicación sistemática de la Existencia. La magnitud de esta rama del saber y sus conexiones con la epistemología (rama de la filosofía que estudia la naturaleza y las fuentes del conocimiento) hacen que el estudio de la ontología desde el punto de vista filosófico quede fuera de los alcances de nuestra investigación.

Derivado de su significado original, aunque con un entendimiento mucho más pragmático y aplicado, el término *ontología* se usa en el ámbito de la ingeniería del conocimiento para referirse a un conjunto de conceptos organizados jerárquicamente, representados en algún sistema informático cuya utilidad es la de servir de soporte a diversas aplicaciones que requieren de conocimiento específico sobre la materia que la ontología representa. A éste segundo significado de ontología, entendida como un cuerpo estructurado de conocimiento, es al que queremos circunscribir nuestra discusión sobre ontologías y su aplicación y utilidad en la gestión terminológica.

3.1 Definición de ontología como especificación del conocimiento

Ya hemos mencionado que una definición de diccionario típica del término ontología la identifica con “la rama de la metafísica que estudia la naturaleza de la existencia”. En las aplicaciones reales, sin embargo, una ontología es una entidad computacional, y no ha de ser considerada como una entidad natural que *se descubre*, sino como recurso artificial que *se crea* (Mahesh 1996). Una ontología ha de entenderse como un entendimiento común y compartido de

Antonio Moreno Ortiz 62//00 9:41 PM
Con formato: Numeración y viñetas

un dominio, que puede comunicarse entre científicos y sistemas computacionales. Ésta última característica, el hecho de que puedan compartirse y reutilizarse en aplicaciones diferentes, explica en parte el gran interés suscitado en los últimos años en la creación e integración de ontologías (Steve et al. 1998a, b).

El sinónimo más usual de ontología es *conceptualización*. Según la definición de Gruber (1993:199), una ontología constituye “a formal, explicit specification of a shared conceptualization”. En esta definición, convertida ya en estándar, *conceptualización* se refiere a un modelo abstracto de algún fenómeno del mundo del que se identifican los conceptos que son relevantes; *explícito* hace referencia a la necesidad de especificar de forma consciente los distintos conceptos que conforman una ontología, lo cual imponíamos como condición de representación conceptual en nuestra definición de concepto del apartado 2; *formal* indica que la especificación debe representarse por medio de un lenguaje de representación formalizado y *compartida* refleja que una ontología debe, en el mejor de los casos, dar cuenta de conocimiento aceptado (como mínimo, por el grupo de personas que deben usarla).

Una definición de ontología más concreta la ofrece Weigand (1997):

An ontology is a database describing the concepts in the world or some domain, some of their properties and how the concepts relate to each other.

Por tanto, aunque en filosofía una ontología es una explicación sistemática de la Existencia, en los sistemas basados en el conocimiento, lo que existe es exactamente lo que se puede representar, y lo que se representa, mediante un formalismo declarativo se conoce con el nombre de *Universo de Discurso* (UoD). El UoD de una ontología es el conjunto de objetos que están representados en ella y sobre los cuales se puede hablar y razonar.

Cuando hablamos de ontologías como “sistemas de representación de conocimiento” debemos especificar a qué tipo de sistemas nos referimos. En realidad, las ontologías se están empleando en todo tipo de aplicaciones informáticas en las que sea necesario definir concretamente el conjunto de entidades relevantes en el campo de aplicación determinado, así como las interacciones entre las mismas. Algunas ontologías se crean con el mero objetivo de alcanzar una comprensión del UoD pertinente, ya que su creación impone una especificación muy detallada. Otras ontologías han sido creadas con un propósito general, como por ejemplo el proyecto Cyc (Guha & Lenat 1990), que está orientado a la construcción de una base de conocimiento que contenga el conocimiento humano necesario para hacer inferencias.

La ontología que hemos reutilizado en el proyecto de investigación OncoTerm ha sido la creada de MikroKosmos. Esta es una ontología a la que sus autores denominan “situada” (Mahesh 1996: 6), y con ello se refieren al hecho de que la creación de dicha ontología ha estado enfocada a la realización de una tarea concreta, dentro de un ámbito bien definido y diseñada para trabajar de forma conjunta con unas herramientas determinadas. Sin embargo, el hecho de que sea una ontología situada no impide que la cobertura de conceptos generales sea extensiva (aunque no pueda considerarse como una ontología genérica). Por otra parte, integra su propia *meta-ontología*, ya que todos los conceptos incluidos (excepto el nodo primario) parten de conceptos previamente definidos en la ontología. Por ejemplo, la relación fundamental IS-A, que estructura la jerarquía, es un concepto de la ontología, concretamente es una PROPERTY de tipo ONTOLOGY-SLOT. En nuestra opinión, estas dos características la hacen muy valiosa y reutilizable para nuestros propósitos.

3.2 La ontología de MikroKosmos y su reutilización para el subdominio del cáncer

Cuando nos propusimos usar una ontología de conceptos para la representación de la estructura conceptual del dominio del cáncer, la primera decisión a tomar atañía al tipo de ontología que queríamos crear. Según nuestros planteamientos, para que esta ontología fuera de utilidad en la terminografía, debía permitimos representar de modo explícito y formalizado, no sólo las características definitorias de los conceptos incluidos, sino también priorizar algunas de éstas características, para dar cuenta de las diferentes perspectivas desde las que puede verse un mismo concepto y, por tanto, debía permitimos insertarlos en lugares diferentes de la ontología.

Dada la transdisciplinariedad del ámbito de investigación en el que nos movemos, el cáncer y su tratamiento, la ontología debía reflejar una estructuración conceptual muy rica, en la que estuvieran incluidos conceptos bastante genéricos y de ramas del saber diferentes (como por ejemplo, las partes del cuerpo humano o las relaciones causa-efecto), puesto que, si nos ceñíamos sólo a los conceptos específicos del cáncer, muchas de las propiedades y relaciones de éstos conceptos no iban a poder representarse, al hacer referencia a conceptos de otros ámbitos de especialidad o más genéricos.

La tarea de construir una ontología de estas características, partiendo desde cero, es enorme, por lo que, después de analizar algunas iniciativas de construcción de ontologías con fines médicos ya existentes, decidimos adaptar a nuestras necesidades una ontología (no

especializada) ya existente, a la que se ofrece libre acceso para fines académicos:² la ontología del proyecto MikroKosmos (μK). Conocíamos ya con bastante detalle la estructura y composición de dicha ontología, puesto que ya la habíamos usado anteriormente en el desarrollo de un lexicón computacional para la traducción automática basada en el conocimiento (véase Moreno Ortiz 1997). Los magníficos resultados obtenidos en dicho proyecto hacían patente la utilidad de los niveles superiores y medios de la ontología de MikroKosmos para otras aplicaciones de procesamiento del lenguaje natural.

El proyecto MikroKosmos (μK), del que la ontología es el componente central, es un sistema de KBMT interlingüe desarrollado por el Computing Research Laboratory (CRL) de la New Mexico State University (NMSU), EE.UU., financiado por el Ministerio de Defensa de este país (Beale, Nirenburg & Mahesh 1995, Mahesh & Nirenburg 1995ab). A diferencia de otros proyectos de KBMT anteriores, de dimensiones más reducidas, μK es un sistema práctico a gran escala, enfocado en principio a traducir entre los idiomas inglés y español, y que actualmente está siendo expandido para dar cabida a otros idiomas.

μK comenzó siendo un proyecto derivado de Pangloss; de hecho, se inició con el objetivo de superar las deficiencias mostradas por el motor de KBMT de este sistema (Nirenburg et al. 1995). En la actualidad, μK traduce artículos periodísticos españoles e ingleses sobre adquisiciones y fusiones empresariales sin restricciones de *input*. Para ello utiliza una serie de lexicones específicos para cada lengua y una ontología de conceptos independiente de cualquiera de las lenguas. Estos dos recursos son utilizados para generar las denominadas TMRs (*Text Meaning Representations*) que constituyen las representaciones interlingües propiamente. Además del analizador sintáctico, que es el mismo que el utilizado en el proyecto Pangloss, en μK se están desarrollando una serie de algoritmos específicos para problemas concretos, denominados *microteorías*; se contemplan, por ejemplo, teorías específicas para el aspecto, el tiempo, el género, etc. A pesar de que μK fuera diseñado para traducir un tipo específico de textos (sobre fusiones y adquisiciones empresariales), el hecho de que no se impusiera restricción alguna en el contenido de los textos de entrada obligó a los creadores de μK a construir una ontología en la que se incluyera un gran número de conceptos generales, y a

² Hay que aclarar que lo que el proyecto μK pone a disposición de la comunidad académica es la ontología en formato Lisp, pero no los programas que facilitan la consulta, edición y modificación de la ontología. Información sobre el proyecto μK y su ontología puede obtenerse en su sede web en <http://crl.mnsu.edu/users/mahesh/onto-intro-page.html>.

representar muy variados objetos del mundo, las propiedades de dichos objetos y los eventos en los que dichos objetos se relacionan entre sí o con sus propiedades, es decir, una ontología genérica. La ontología original de μK cuenta con unos 4.700 conceptos, los cuales están conectados con una media de otros 14 conceptos, a través de la asignación de atributos y relaciones.

Para dar una idea de su estructura, la Figura 1 muestra gráficamente la estructuración algunas de las ramas de los niveles superiores de la jerarquía:

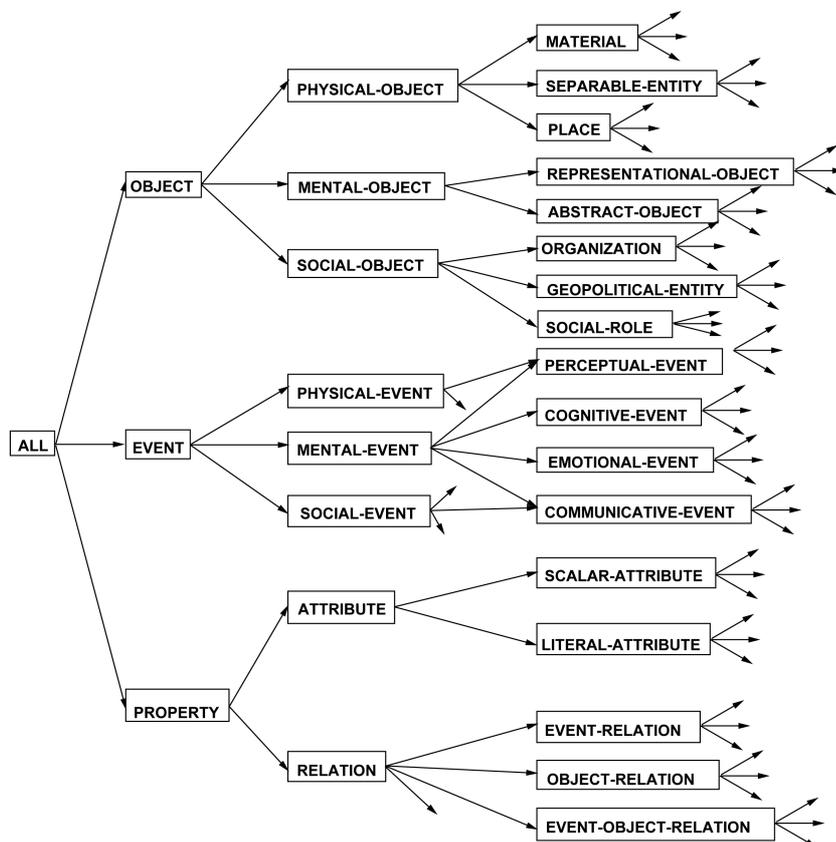


Figura 1: Niveles superiores de la ontología de MikroKosmos.

Como se puede observar existen tres entidades superiores, OBJECT, EVENT y PROPERTY, a partir de las cuales se desarrollan todas las demás. La jerarquía es una red semántica de marcos; cada uno de estos marcos posee una rica estructura interna que le permite una gran expresividad. La ontología puede ser considerada como una entidad autónoma en el sentido de que se define a

sí misma. Por ejemplo, todas las propiedades adscritas a los objetos o eventos están a su vez definidas en algún punto de la rama PROPERTY (a su vez, las propiedades están divididas en ATTRIBUTES y RELATIONS).

Por otra parte, las propiedades de los objetos y eventos se heredan a lo largo de los sucesivos niveles de la jerarquía. Por ejemplo, todos los EVENTS heredan por omisión la propiedad de tomar la propiedad de requerir un AGENT. Si especificamos esta característica al nivel EVENT, todos los conceptos hijos de éste heredarán esta propiedad. Además, la herencia puede ser no monotónica, también llamada herencia negativa, es decir, ha de ser posible especificar que algún elemento no herede alguna propiedad. Por ejemplo, los PASSIVE-COGNITIVE-EVENTS y los INVOLUNTARY-PERCEPTUAL-EVENTS no requieren un agente. En cuanto a la profundidad (número de conceptos hijo sucesivos) de las ramas de las diferentes jerarquías, la ontología llega a poseer diez niveles o más en un buen número de ramas.

Sin embargo, necesitábamos un sistema gestor de ontologías que nos permitiera aprovechar las características de la estructura y composición de la ontología de μK . Como ya hemos mencionado, la ontología de μK se ofrece al público en código LISP por lo que, sin contar con un programa específicamente diseñado para tal fin, era imposible realizar ni la más simple de las consultas en la ontología, por lo que nos decidimos a crear una aplicación informática que gestionara la ontología y nos propusimos crear una interfaz gráfica de usuario que permitiera a los demás miembros del proyecto de investigación posible realizar, de forma muy sencilla, las tareas de consulta, modificación y entrada de conceptos en la ontología.

Para que la aplicación informática desarrollada pudiera usarse en todas las fases de la gestión terminológica y no sólo en la estructuración del conocimiento pertinente al dominio de especialidad, diseñamos e implementamos un sistema gestor de base de datos terminológica basado en el conocimiento, del que el gestor de ontologías es uno de sus componentes. La aplicación en su conjunto, denominada OntoTerm®, hace posible relacionar de forma directa conceptos de la ontología con términos contenidos en una o varias bases de datos terminológicas. Las ventajas de este tipo de estructuración de la información terminológica, así como la filosofía de diseño de dicha base de datos serán expuestas en el apartado 4.2 y siguientes; pasamos ahora a describir la filosofía del diseño e implementación del sistema gestor de bases de datos en el apartado siguiente y a describir el funcionamiento del primer componente de OntoTerm®, el gestor de ontologías, en el apartado 4.1

4 El sistema gestor de representación de conocimiento: OntoTerm®

Teniendo en cuenta todo lo expuesto hasta ahora en este artículo, podríamos resumir las características que exigimos en un sistema de representación terminológica basada en conceptos en las siguientes:

- Gestión conceptual real basada en ontologías.
- Adopción de un esquema de representación estándar, preferiblemente basado en marcos, o compatible con este esquema.
- Interfaz gráfico que facilite las operaciones de gestión comunes como agregar, editar y eliminar conceptos.
- Modularidad: el gestor conceptual debería operar de modo independiente al gestor terminológico propiamente dicho.
- Integración: el gestor terminológico debería ser capaz de aprovechar las ventajas derivadas del uso de un gestor conceptual avanzado, tales como la herencia por defecto o la integridad de la jerarquía conceptual.

Evidentemente, el primer problema con el que nos encontramos a la hora de poner en práctica nuestras ideas es la disponibilidad de un sistema de estas características. Por un lado, las aplicaciones de gestión terminológica disponibles en el mercado no ofrecen este tipo de características avanzadas, pues se reducen a la organización del conjunto de términos en torno a conceptos, cuyo estatus dentro del sistema no está claro. Por otro, las aplicaciones y sistemas que sí permiten la construcción de un sistema conceptual u ontología, no están preparadas para la gestión terminológica.

En el primero de los grupos se enmarcan las aplicaciones comerciales que, a pesar de carecer de un sistema conceptual adecuado, gozan de una gran difusión entre los usuarios de herramientas de gestión terminológica. De entre estos destaca, sin duda, TRADOS MultiTerm. Resulta interesante resaltar lo que para los creadores de este sistema significa “terminología orientada al concepto”:³

An entry always corresponds to one concept. This is why terminologists call MultiTerm “concept-oriented.” For the English homonym “plane”, you would create at least two entries: one for airplane, and one for the tool used to create a smooth surface on wood.

³ Tomado de la ayuda de TRADOS MultiTerm '95 Plus!, © 1992-96, TRADOS GmbH.

Dicho de otro modo: un gestor terminológico merece el calificativo de “orientado al concepto” si permite almacenar palabras homónimas. Sin duda el terminólogo moderno necesita herramientas que se tomen su trabajo mucho más en serio; de poco sirve postular la necesidad de realizar terminología basada u orientada al concepto si el enorme trabajo que supone la estructuración conceptual pormenorizada en un dominio no va a quedar plasmado en modo alguno en el resultado final.

Sin salir del ámbito de la terminología, el único sistema que permite una representación del sistema conceptual subyacente a una terminología es CODE4 (Skuce & Lethbridge 95). Este sistema se utilizó en la ejecución del proyecto Cogniterm desde 1991.⁴ CODE4 evolucionó hacia la herramienta llamada IKARUS (1996-97), la cual aportaba como novedad fundamental el estar escrita en Java, de modo que, en teoría podía ejecutarse a través de Internet desde cualquier navegador con una Máquina Virtual Java integrada. En nuestra experiencia, sin embargo, jamás conseguimos hacerla funcionar de este modo, dando señales de ser una aplicación extremadamente inestable. Tal vez debido a esto, IKARUS dejó de existir pronto para transformarse en la herramienta que este grupo canadiense está desarrollando en el presente: DOCKMAN, la cual, aunque también está escrita en Java, no se ofrece para su uso a través de Internet.

Aunque no tenemos experiencia directa con este sistema o sus antecesores, algo que se desprende de las publicaciones en las que sus autores lo describen (Lethbridge & Skuce 1992; Skuce & Lethbridge 1994, 1995; Meyer, Eck & Skuce 1997), es el hecho es que nunca se ha empleado para la creación de una ontología a gran escala tal y como la que nosotros nos planteamos utilizar, por lo que pensamos que quizá este sistema está pensado para trabajar con pequeños dominios conceptuales en los que las relaciones entre los distintos conceptos puedan ser fácilmente controladas. Tampoco encontramos una descripción clara del tratamiento concreto que se ofrece para la gestión de información terminológica en lo que respecta a cuestiones fundamentales como son las categorías de datos o el intercambio de información. En realidad, toda la información lingüística asignada al término parece ser un único valor de un atributo, algo que dista mucho de lo que consideramos idóneo.

Incluso así éste es, sin duda, el sistema que más se acerca en cuanto a filosofía de diseño a OntoTerm®, el sistema gestor de conocimiento que hemos desarrollado para la representación

⁴ Todo lo referente a este proyecto y sus posteriores evoluciones se puede encontrar en <http://aix1.uottawa.ca/~imeyer/research.htm>

de información terminográfica y su estructuración conceptual. Como puntos destacados de este sistema, deberíamos resaltar los siguientes:

1. Representación de conocimiento: el gestor de ontologías (*Ontology Editor*) de OntoTerm® permite la creación, gestión y navegación de ontologías, es decir, de cuerpos de conocimiento altamente estructurados susceptibles de ser utilizados en diversas tareas.
2. Escalabilidad: al contrario que otras aplicaciones, en las que se fomenta la creación de prototipos a pequeña escala o sistemas conceptuales “de juguete”, OntoTerm® está enfocado a la creación de ontologías a gran escala, en las que se pueda trabajar con un número muy elevado de nodos y relaciones entre nodos. Tanto es así que los límites los impone el sistema informático del usuario, no la propia aplicación.
3. Flexibilidad: el usuario es quien decide el nivel de complejidad que quiere otorgar a su sistema conceptual, por lo que la aplicación impone un número lo más reducido posible de restricciones.
4. Integridad: la aplicación, no el usuario, es la responsable de mantener en todo momento la integridad y coherencia interna de la información que contiene.
5. Compatibilidad: tanto el esquema de representación utilizado como el formato de almacenamiento deben ser lo más estándar posible. Además, el sistema debería permitir la exportación de información a sistemas conceptuales extendidos. OntoTerm® puede exportar sus bases de conocimiento a los sistemas gestores de bases de conocimiento Loom y Clips. Por otra parte, OntoTerm® almacena sus datos internamente en una base de datos relacional, sin duda el formato de almacenamiento de datos más estándar.
6. Modularidad: el sistema de representación conceptual es independiente del sistema de gestión terminológica, con el que interactúa de diversas maneras, pero cuya existencia no presupone. Esto implica una separación y delimitación de la información puramente conceptual, y por tanto independiente de la lengua, de la información léxica y administrativa, que se asigna al término en una base de datos independiente.
7. Facilidad de manejo: el usuario no tiene por qué ser un especialista en computación. La interfaz de usuario debe permitir realizar todas las operaciones de una forma fácil y amigable en la medida de lo posible.

8. Facilidad de acceso: el sistema ofrece diversos mecanismos encaminados a una navegación eficaz. Esta característica es fundamental en un sistema que permite el trabajo con ontologías a gran escala.

Estas características son las que han hecho posible que en el proyecto OncoTerm se haya podido manipular un cuerpo estructurado de conocimiento tan complejo como es la ontología de MikroKosmos de forma relativamente sencilla, facilitando la integración de los conceptos específicos a nuestro ámbito de especialidad.

4.1 El gestor de ontologías de OntoTerm®

Para describir de forma sucinta las funcionalidades del gestor de ontologías de OntoTerm® y las posibilidades que ofrece en el trabajo terminográfico, veamos cómo procedería un terminógrafo para incluir (y editar, borrar, modificar, etc.) la información referente a uno de los conceptos de la ontología: MYELOID-LEUKEMIA.

Lo primero que debe hacer es localizar la posición del concepto en la ontología, es decir, el concepto superordinado de nuestro concepto. En este caso, la elección del concepto padre de MYELOID-LEUKEMIA le vendría dada, puesto que la clasificación de neoplasias incluidas en la ontología está tomada del ICD-9 CM, en la que la que MYELOID-LEUKEMIA se clasifica como un tipo de MALIGNANT-NEOPLASM-OF-LYMPHATIC-AND-HEMATOPOIETIC-TISSUE.

La siguiente captura de pantalla muestra el cuadro de diálogo en el que se añaden conceptos nuevos en la ontología. Como puede apreciarse en dicho cuadro de diálogo se especifica el tipo de concepto que se añade (EVENT, OBJECT, ATTRIBUTE o RELATION), se elige también el concepto superordinado, el tipo de relación que les une (INSTANCE-OF o IS-A) y puede, además, añadirse una definición, tomada en este caso del UMLS Metathesaurus:

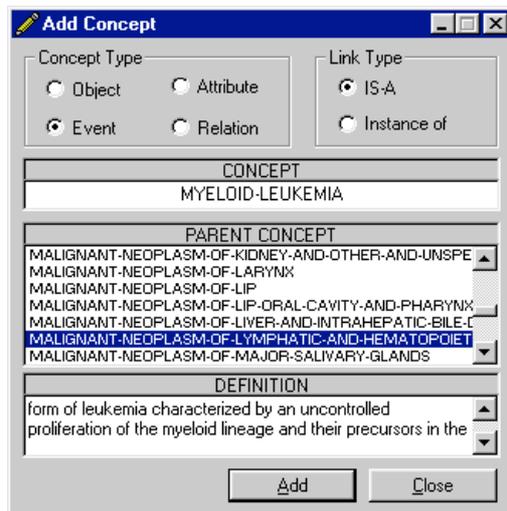


Figura 2: Cuadro de diálogo para añadir conceptos.

Siguiendo el mismo procedimiento, incluimos a continuación los conceptos subordinados de MYELOID-LEUKEMIA. Una vez añadida esta información, podemos comprobar el concepto (o conceptos) superordinados y subordinados de MYELOID-LEUKEMIA en la pantalla de edición principal del editor de ontologías. Como muestra la siguiente captura, esta pantalla de edición nos permite en su barra de herramientas tener acceso a las funciones básicas del gestor de ontologías. A la hora de visualizar un concepto, la ficha de esta pantalla denominada description nos ofrece la información que hemos añadido sobre su conceptos super- y subordinados, su definición e identifica, en la parte inferior del editor, a MYELOID-LEUKEMIA como un concepto (en este caso un EVENT), añadido por un miembro de OncoTerm (para diferenciarlos de los conceptos ya existentes en la ontología de MikroKosmos):

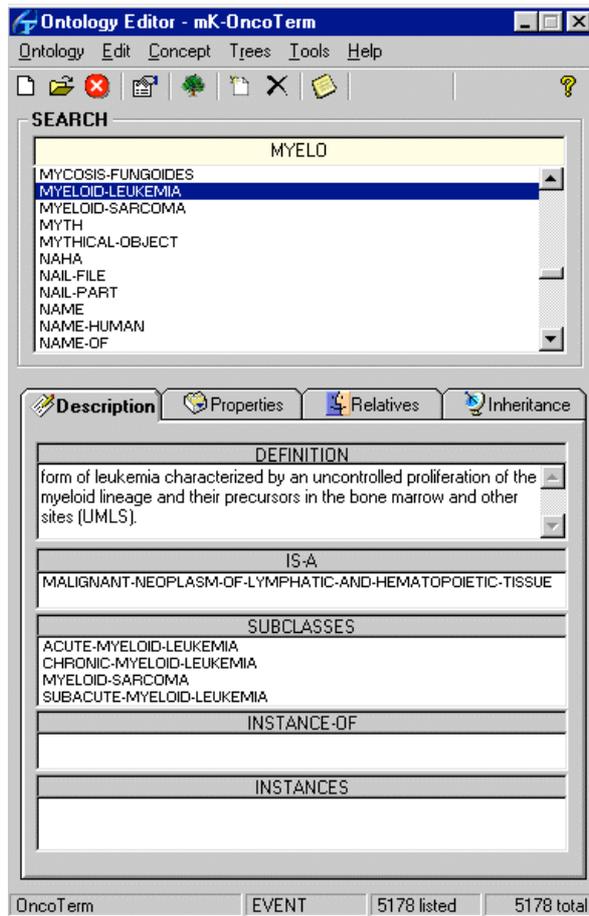


Figura 3: Pantalla de edición principal del Ontology Editor.

Además de la descripción básica del concepto, esta pantalla nos permite especificar las propiedades del concepto, el tipo de información que, sin lugar a dudas, más enriquece la ontología. Con ventanas de edición de estructura similar a la que nos permite añadir los conceptos (véase Figura 2), añadimos las RELACIONES y los ATRIBUTOS del concepto:

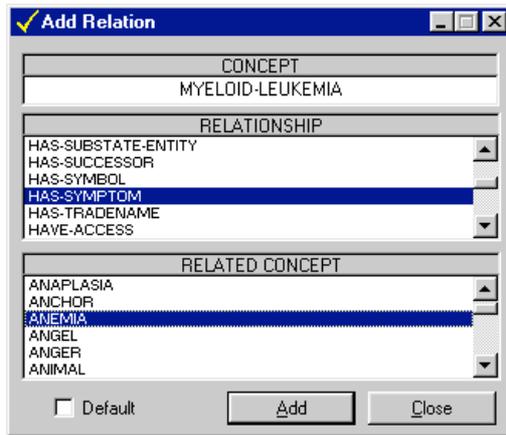


Figura 4: Cuadro de diálogo para añadir relaciones.

Al añadir esta relación, el gestor de ontologías nos permite indicar de forma explícita que la anemia es uno de los síntomas de MYELOID-LEUKEMIA. Una vez (o a la vez) que añadimos las relaciones de éste concepto con otros conceptos de la ontología y los atributos pertinentes, ambos tipos de información pueden verse resumidos en la pantalla de edición principal, en la ficha correspondiente a PROPERTIES:

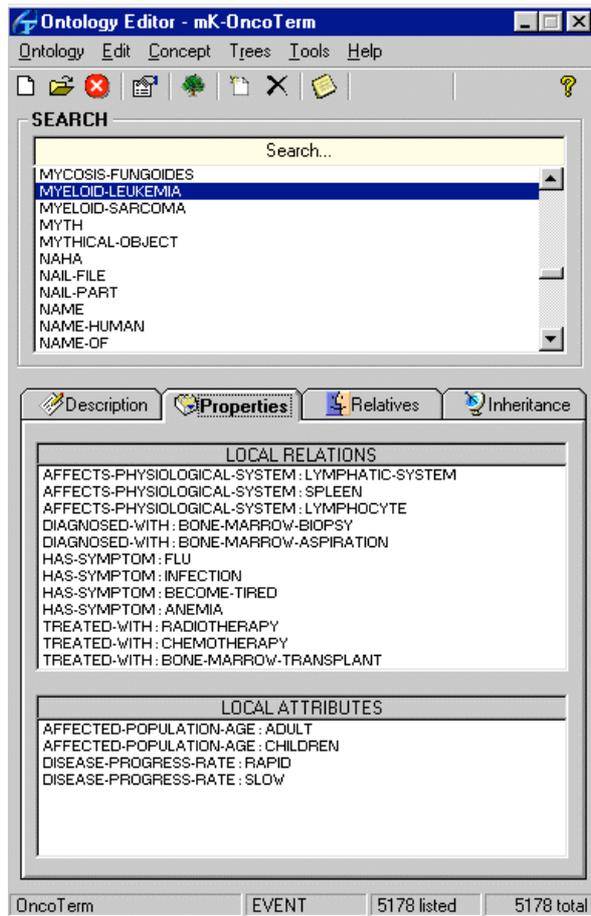


Figura 5: Relaciones y Atributos en el Ontology Editor.

En el proceso de entrada de información que acabamos de mostrar, hay un hecho que no debemos dejar que pase desapercibido, ya que muestra la complejidad de la estructura de la ontología y la inherente consistencia que el gestor de ontologías impone en el proceso de trabajo. Ya mencionamos anteriormente que en la estructura de la ontología original de MikroKosmos, tanto relaciones como atributos se consideran conceptos de la ontología, con respectivas ramas en la jerarquía, sus definiciones y su explicitación por medio de la asignación de DOMAINS y RANGES. Este hecho implica que, para poder asignar una relación o un atributo a un concepto de terminado, dicha relación o atributo se haya tenido que insertar previamente en su lugar correspondiente de la ontología. El gestor de ontologías, además, obliga al usuario a crear a la vez una relación y su relación inversa, actualizando después de forma automática la

información de la ontología. Si volvemos a la captura de pantalla de la Figura 5, vemos que, en nuestro ejemplo, hemos usado en la descripción de nuestro concepto las relaciones AFFECTS-PHYSIOLOGICAL-SYSTEM, DIAGNOSED-WITH, HAS-SYMPTOM y TREATED-WITH y los atributos AFFECTED-POPULATION-AGE y DISEASE-PROGRESS-RATE. Para poder usarlas, se deben incluir previamente en sus lugares correspondientes de la ontología:

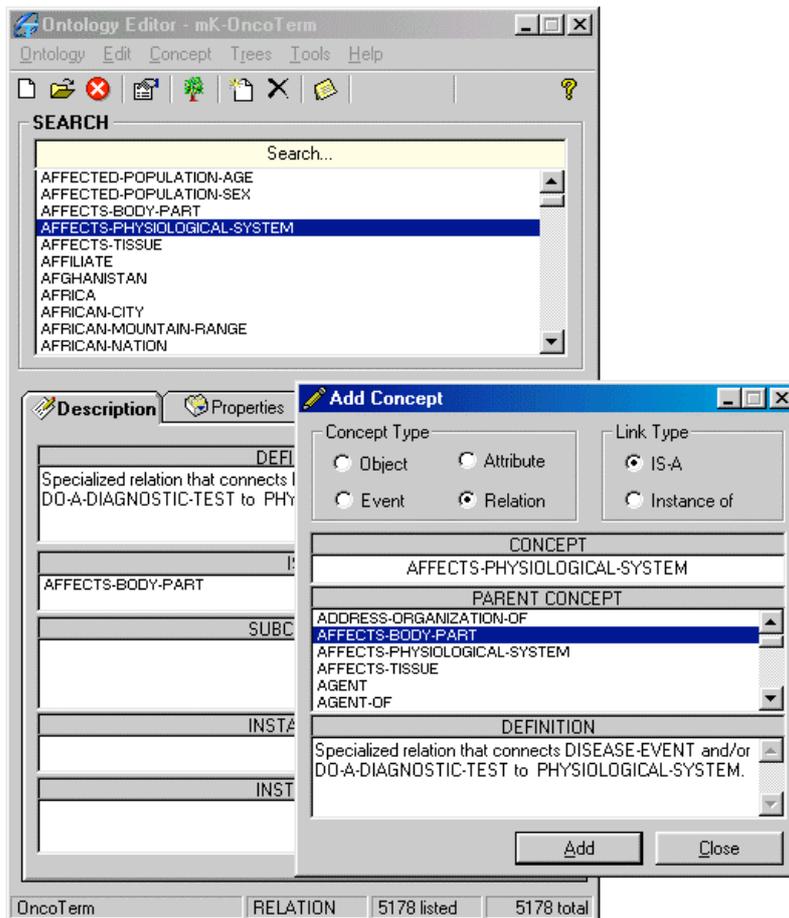


Figura 6: Relaciones en el Ontology Editor.

Como se puede observar en las capturas de pantalla, la relación AFFECTS-PHYSIOLOGICAL-SYSTEM es un concepto subordinado de la relación AFFECTS-BODY-PART. Además, hemos añadido la relación inversa PHYSIOLOGICAL-SYSTEM-AFFECTED-BY en su lugar correspondiente de la jerarquía (INVERSE-DISEASE-EVENT-OBJECT-RELATION).

Otro aspecto importante de la estructuración de la ontología que se hace patente en la Figura 5, en la que hemos mostrado la pantalla principal de las relaciones y los atributos, es el hecho de que los conceptos relacionados con MYELOID-LEUKEMIA también son conceptos de la ontología. Esto, lógicamente, hace que fuera necesario insertarlos previamente en su lugar correspondiente de la jerarquía y, si se considera apropiado, definirlos y explicitar sus características propias.

En muchas ocasiones, posicionar dichos conceptos en la ontología requiere construir o completar una rama completa, sobre todo cuando se trata de un concepto muy específico, como por ejemplo LYMPHOCYTE o NEUTROPHIL. Como puede observarse en la captura de pantalla de la Figura 8, para llegar al nivel de dichos conceptos, partimos del concepto (original de la ontología de μK) INTERNAL-ANIMAL-PART y debimos completarlo con los conceptos hijos sucesivos PHYSIOLOGICAL-SYSTEM \rightarrow IMMUNE-SYSTEM \rightarrow LEUKOCYTE \rightarrow GRANULOCYTE \rightarrow NEUTROPHIL.

Para facilitar la visualización de los conceptos insertados en la jerarquía, el editor de ontologías nos permite ver la ontología completa en forma de árbol, o bien árboles parciales para EVENT, OBJECT, ATTRIBUTE y RELATION. La siguiente captura de pantalla muestra el árbol parcial de RELATION, donde se pueden ver las relaciones específicas a la oncología incluidas en la ontología (que parten de DISEASE-EVENT-RELATION y DISEASE-EVENT-OBJECT-RELATION), así como sus relaciones inversas:

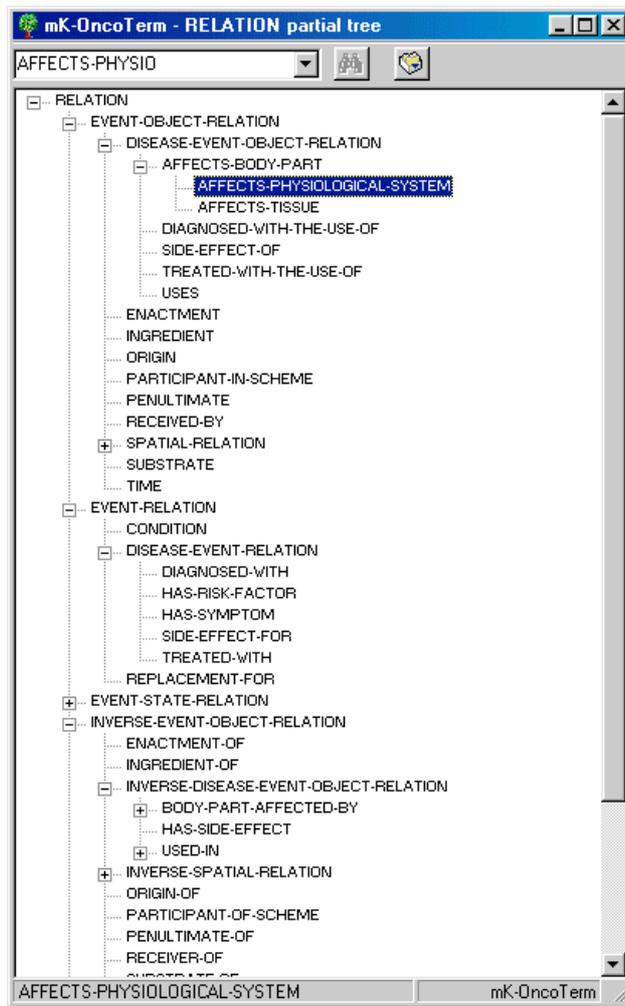


Figura 7: Árbol parcial de RELATION.

Desde el punto de vista del terminógrafo que compila la información, estos árboles son de gran utilidad para ver de forma completa el desarrollo de la estructuración conceptual. En estos árboles puede, además, verse resumida la información que la ontología contiene sobre cada concepto o movernos a la pantalla de edición principal de un concepto:

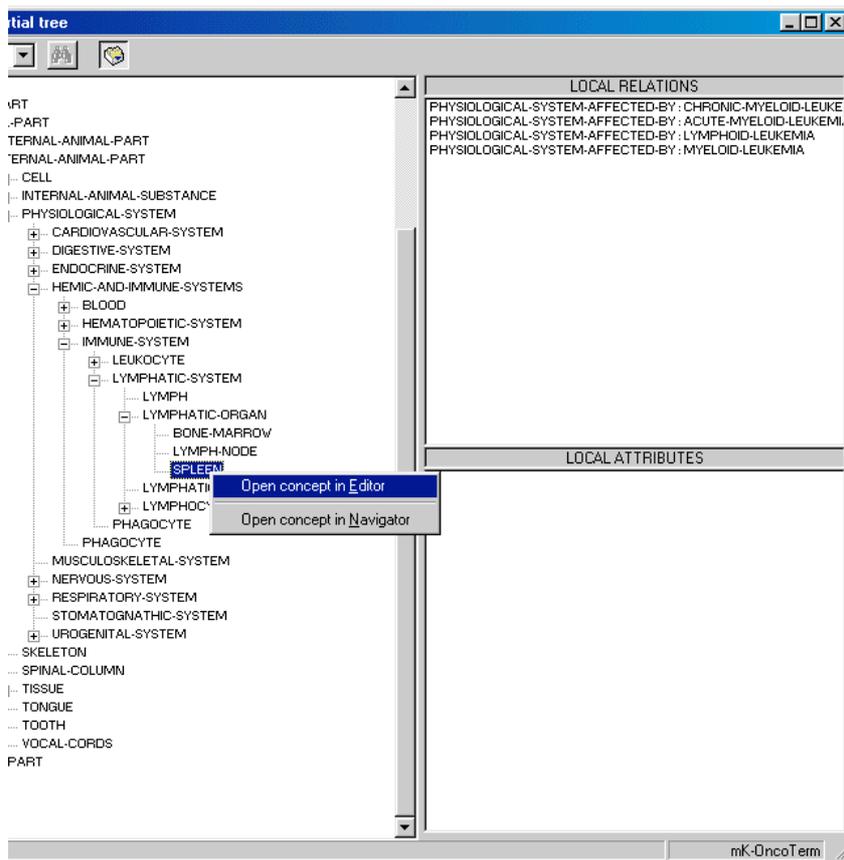


Figura 8: Árbol Parcial de OBJECT (concepto SPLEEN).

El editor de ontologías permite además muchas otras funcionalidades que sería demasiado extenso mostrar aquí en detalle;⁵ permite, por ejemplo, modificar la información ya incluida en la ontología, asignar más de un concepto superordinado (herencia múltiple), ver todos los conceptos superordinados o subordinados de un concepto determinado y los valores que un concepto hereda de sus subordinados, de qué concepto hereda cada valor y la distancia que separa ambos conceptos (superordinado y subordinado).

El editor de ontologías ofrece, además, dos herramientas para la navegación y publicación de nuestra ontología. El primero es el llamado *Ontology Navigator*, que mostramos

⁵ Véase Moreno Ortiz (2000a/b) y Moreno Ortiz & Pérez Hernández (2000).

en la Figura 9. Se trata de una herramienta que combina los árboles ya mostrados en el panel izquierdo con un navegador web integrado en el panel derecho, en donde se nos mostrará de una forma mucho más legible que en el editor toda la información referente a cada uno de los conceptos que conforman nuestra ontología. Sirviéndose de esta herramienta, OntoTerm® genera un documento HTML para cada uno de los conceptos introducidos. Lo único que debemos hacer para visualizar dichos documentos es hacer clic en cualquiera de los nodos del árbol del panel izquierdo. Los documentos se almacenan en un directorio específico que OntoTerm® crea durante su instalación. Al hacer clic la aplicación busca en este directorio la existencia de la página web en cuestión y, si no la encuentra, la crea y la visualiza. Si queremos que se refleje información actualizada tras una primera visualización deberemos pulsar el botón “Refresh” en la barra de herramientas. La siguiente figura muestra parte de un documento HTML generado por el *Ontology Navigator*:

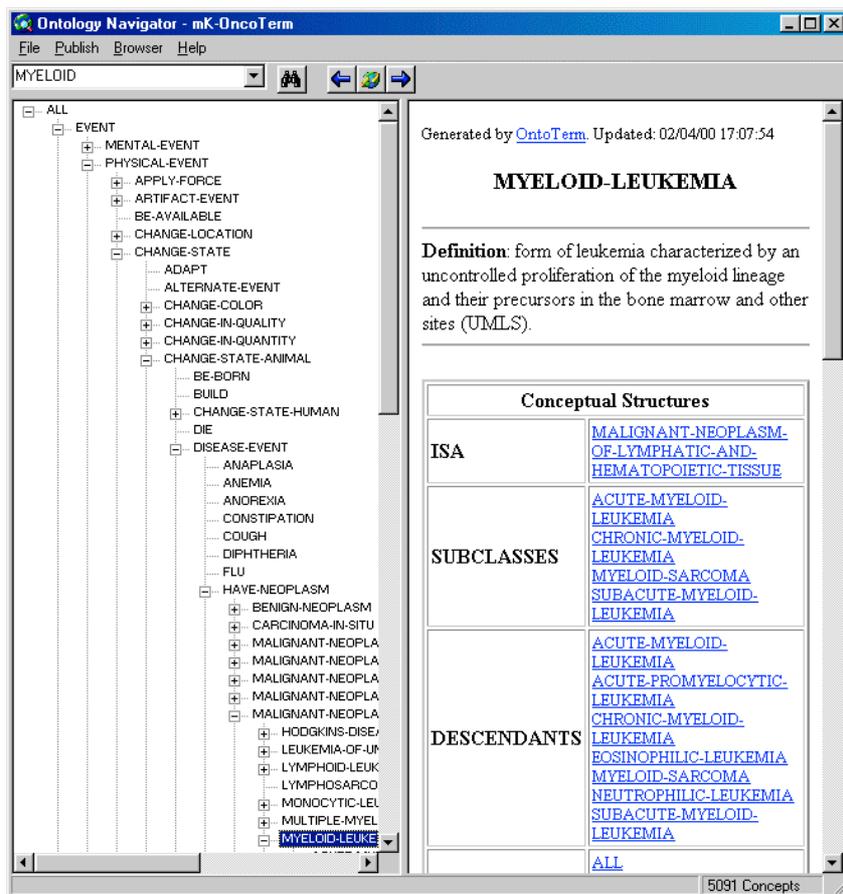


Figura 9: El "Ontology Navigator"

Como se puede apreciar en esta captura de pantalla, OntoTerm® incluye un encabezado en el que se nos informa de la fecha y hora en la que esa página fue creada. Desde esta herramienta también se nos ofrece la posibilidad de publicar la ontología en conjunto. Si seleccionamos esta opción, OntoTerm® generará una página web para cada uno de los conceptos incluidos en la ontología, convirtiendo las referencias a conceptos en hipervínculos y generando una página índice, con marcos en HTML. Esto nos permite publicar todo el contenido de nuestra ontología, bien localmente, bien en la red.

La segunda herramienta de visualización e informe nos permite elegir para qué conceptos de nuestra ontología deseamos generar páginas, la información a incluir en cada una de ellas y la localización de los archivos resultantes.

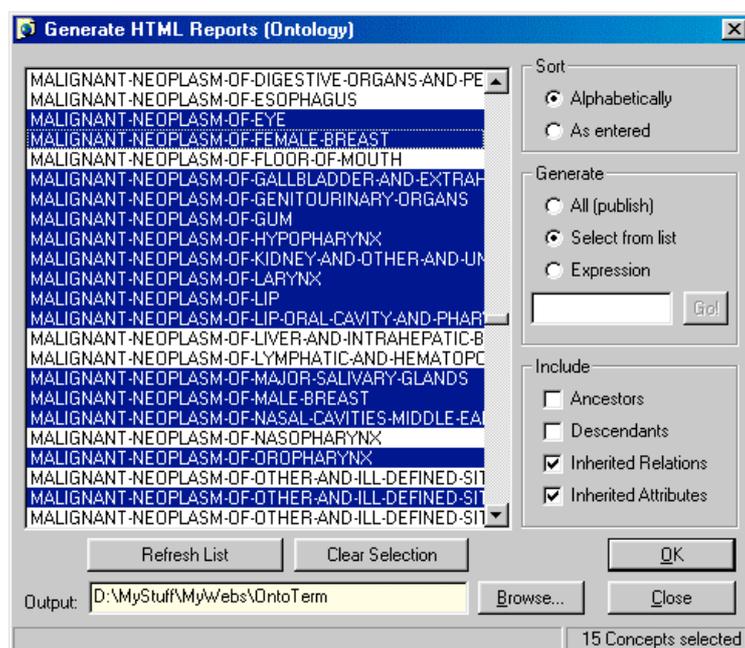


Figura 10: El generador de informes de OntoTerm®

Como muestra la Figura 10, esta herramienta nos permite, además seleccionar qué secciones queremos incluir en nuestros informes, pudiendo omitir aquellas en las que no estemos interesados.

Como ya hemos mencionado, éste es el primero de los dos módulos principales de los que consta OntoTerm®; el segundo, el dedicado a la gestión de información terminológica propiamente dicha (lingüística y administrativa) es el llamado *Termbase Editor*, que describimos en el apartado 4.3. Antes, sin embargo, es importante que conozcamos las categorías de datos con las que trabaja este módulo de la aplicación.

4.2 Categorías de datos en las bases de datos terminológicas y su estructuración

Antonio Moreno Ortiz 62//00 9:41 PM
Con formato: Numeración y viñetas

Aunque durante décadas traductores, lexicógrafos y terminólogos han almacenado la información terminológica en fichas de papel (las denominadas “fichas de vaciado” y “fichas terminológicas”), hoy por hoy, la forma más frecuente (y completa) de almacenamiento de la información terminológica se lleva a cabo en formato electrónico, ya sea por medio de un procesador de texto, un sistema gestor de base de datos genérico o un programa específico de gestión de información terminológica.

Las desventajas que acarrear el uso de las dos primeras formas de almacenamiento (el papel o su correspondiente electrónico en un procesador de textos) son bastante obvias:⁶ falta de rigurosidad, inconsistencia en el formato de los datos que contienen, dificultad de consulta y extensión o modificación de las entradas y los repositorios, etc.

De los otros dos medios de almacenamiento, los sistemas gestores de bases de datos genéricos resultan bastante apropiados para estructurar, guardar y gestionar la información terminológica, aunque también pueden plantear algunos problemas, ya que puede resultar muy laborioso configurar una base de datos de forma que se adapte a las necesidades de un terminólogo y pueden presentar también problemas de compatibilidad con otras herramientas computacionales de uso común en la terminología o la traducción, como procesadores de texto o memorias de traducción (Wright 2001a).

En cuanto a los sistemas gestores de terminología específicos que se encuentran en el mercado o que han sido desarrollados para bancos terminológicos específicos, estos varían considerablemente en sus aspectos técnicos, comerciales o en lo que se refiere a las

⁶ Los problemas y ventajas de las diferentes formas de almacenamiento de información terminológica han sido tratados extensamente en Fischer et al. (1993), Wright (2001a) y Melby, Schmitz & Wright (2001).

posibilidades de modelado y presentación final de la información que ofrecen al usuario.⁷ Algunos sistemas ofrecen al usuario un repertorio fijo de categorías de datos predefinidas, o pueden incluso limitar la extensión de los campos de datos. En estos casos, los usuarios pueden sentirse frustrados si necesitan añadir alguna categoría no contemplada por el sistema o si intentan introducir información que sobrepasa la extensión asignada al campo de datos. Otros sistemas, sin embargo, ofrecen al usuario una mayor flexibilidad en lo que se refiere a la creación y estructuración de las categorías de datos que conforman una entrada terminológica.

Los sistemas gestores de bases de datos terminológicas difieren, además, en otros aspectos: algunos mantienen en gran medida la mentalidad de las tradicionales fichas terminológicas impresas, por lo que la información referente a los sinónimos o los equivalentes de un término han de ser tratados en registros separados. En otros sistemas es posible, al menos en teoría, que una entrada terminológica incluya toda la información concerniente a un *concepto*, aunque esto en muchas ocasiones se realiza enlazando varios registros, de modo que forman una especie de “entrada virtual”, la que representaría el concepto. Esta solución, sin embargo, puede presentar serios problemas en lo que se refiere a la gestión de la base de datos, ya que puede dificultar la realización cambios o modificaciones de forma automática en el caso de que un término aparezca como sinónimo en una entrada virtual y a la vez posea una entrada propia, y hace muy difícil asignar información conjunta a, por ejemplo, a los términos que designan un concepto en un mismo idioma.

Otro aspecto en el que los sistemas gestores de bases de datos difieren considerablemente es en el tipo de modelado de datos que ofrecen, sobre todo en lo que se refiere a las categorías de datos que identifican los términos de entrada en una base de datos. En algunos sistemas, existen diferentes nombres para la categoría **término de entrada** (*término principal, sinónimo, forma abreviada*), en la que se identifica el tipo de término de entrada. El ejemplo que sigue está tomado de Wright (en 2001a: 592):

term:	serializer
synonym:	parallel-serial converter
synonym:	dynamicizer

⁷ El proyecto *POINTER* realizó un estudio comparativo de siete sistemas gestores de bases de datos terminológicas (*POINTER Project* 1998: capítulo 5; Ver nota al pie **Error! Marcador no definido.**), basándose en una serie de criterios técnicos, comerciales, de modelado de datos y de facilidad de intercambio y distribución de la información.

En otros sistemas se usa un único nombre para el campo que identifica al término de entrada (ej. *término*, *término de entrada*, etc.) y añaden la información sobre el tipo de término en otra categoría de datos (que puede denominarse, por ejemplo, *tipo de término* (*term type*):

term:	serializer
term type:	main entry term
term:	parallel-serial converter
term type:	synonym
term:	dynamicizer
term type:	synonym

Con este tipo de modelado de datos que, como veremos en la sección siguiente, es el que se recomienda para el intercambio de información terminológica, todos los términos se consideran iguales (de modo que se pueden asociar al mismo nivel al concepto que designan) y se diferencian entre sí por el valor asignado en otra categoría de datos (*term type*).

Existen otros aspectos muy relevantes a tener en cuenta en el modelado de los datos, como son la repetición (*repeatability*) y combinación (*combinability*) de categorías de datos, así como los posibles enlaces entre entradas o elementos de una entrada. Si el modelado de datos permite la repetición de categorías de datos, la misma categoría puede usarse más de una vez en una misma entrada terminológica. Algunas categorías deben aparecer sólo una vez (ej. número de identificación de la entrada), mientras que otras (ej. la categoría *term* en el ejemplo anterior o categorías para incluir contextos o ejemplos) deberían poder aparecer más de una vez.

El principio de repetición, además, permite que en cada campo de datos se incluya un solo elemento de información, ya que cada categoría puede repetirse tantas veces como elementos de información sean requeridos. El hecho de que unos campos de datos puedan enlazarse con otros permitirá, además, que cada campo de datos consigne un único *tipo* de información. Por ejemplo, una definición o un contexto de uso debe contener un *enlace* a una referencia bibliográfica, pero no debe contener la referencia bibliográfica en sí.

La correcta implementación de estas funcionalidades en una base de datos terminológica parte de una base fundamental: la correcta especificación de las categorías de datos y el establecimiento de un protocolo de utilización (en el sentido de Cabré 1993: 283) de dichas categorías de forma clara e inequívoca, de modo que el terminólogo pueda estar seguro de que los datos que incluye se ajustan (en forma, número y lugar en la entrada) al tipo de datos permitidos en esa categoría.

Una de las iniciativas llevadas a cabo por la comunidad internacional con vistas al desarrollo y difusión de estándares de este tipo es el trabajo desarrollado por el comité técnico 37 de la ISO (sub-comité 3, grupo de trabajo 1) quien, después de varios años de documentación y consultas extensivas, ha elaborado un amplio listado de las categorías de datos usadas que con mayor frecuencia se usan en las bases de datos terminológicas. Estas categorías de datos forman la base de la norma ISO 12620: 1999, *Computer Applications in Terminology - Data Categories*.

Las categorías descritas en esta norma no deben considerarse exhaustivas ni, según las describe uno de los miembros del comité técnico encargado de su elaboración (Wright 2001b: 567), es probable que un proyecto de creación de base de datos terminológica necesite incluir todas y cada una de las categorías descritas. El trabajo llevado a cabo por el comité técnico 37 de la ISO para definir y explicitar las categorías de datos se complementa con las propuesta llevada a cabo por el grupo de trabajo denominado *CLS Framework*, en la que se realiza una propuesta de estructuración de dichas categorías muy completa, y es la que hemos usado para la estructuración de la base de datos terminológica del proyecto OncoTerm.

4.2.1 Estructura de una base de datos en el *CLS Framework*

Con independencia del formato que se elija para representar la información terminológica, (una base de datos relacional, un texto estructurado con códigos SGML, un programa de gestión terminológica, etc.), es de vital importancia establecer de forma explícita un modelo de datos para todos los tipos de información terminológica (Higgins & Ahmad 1996: 215-224). La norma ISO 12620 ofrece un inventario completo de *categorías de datos*. Según esta norma, una entrada terminológico-conceptual (denominadas en inglés *terminological concept entry* o *term entry*, en forma abreviada) está compuesta por una serie de campos de datos y cada uno de estos elementos o campos conforma una categoría de datos. Lo que la norma ISO 12620 **no** especifica es la **estructura de la entrada terminológico-conceptual**, es decir, no indica de forma explícita las relaciones que existen (o deben existir) entre las distintas categorías de datos dentro de una entrada.

Precisamente ésta es una de las propuestas que el *CLS Framework* está desarrollando en la actualidad. En ella, se propone una forma de estructurar los elementos que se integran en una entrada terminológica de modo que dicha estructuración esté en consonancia con los fundamentos generales de la teoría terminológica (se parte del concepto para llegar a la denominación). Las siglas elegidas para denominar al grupo de trabajo hacen referencia explícita al enfoque propuesto para la estructuración de una base de datos terminológica:

Antonio Moreno Ortiz 62//00 9:41 PM
Con formato: Numeración y viñetas

Concept-oriented with Links and Shared references -una estructuración basada en los conceptos, con enlaces y referencias compartidas. Por tanto, según el *CLS Framework* una base de datos terminológica se estructura en tres partes fundamentales:

1. información global sobre la base de datos,
2. entradas terminológico-conceptuales,
3. un grupo de referencias que pueden compartir varias entradas o partes de una entrada.

La *información global* puede incluir detalles sobre el nombre de la base de datos, versión, derechos de autor y creadores de la base de datos. Puede además incluir información sobre las lenguas y los alfabetos usados en la base de datos, el grupo de usuarios potencial de la información contenida u otra información administrativa o de gestión.

El segundo de los grupos, *las entradas terminológico-conceptuales*, también denominado *cuerpo (body)*, está compuesto por los conceptos en un campo de especialidad y los términos que les son asignados como designación específica en una o varias lenguas. A cada uno de estos términos se le asigna, a su vez, una serie de informaciones específicas (descriptivas, administrativas, lingüísticas, de uso, etc.).

En consonancia con los principios de la teoría terminológica tradicional, la situación ideal en este tipo de estructuración de base de datos (y de cualquier otra, deberíamos añadir) se da cuando todos los términos (en una o varias lenguas) contenidos en una entrada terminológico-conceptual designan el mismo concepto. De este modo, en cada entrada, los términos de una misma lengua se agrupan y forman los denominados *langsets*, por lo que determinadas categorías de datos pueden también asignarse al *langset* completo (autor de las entradas, fecha, fuentes bibliográficas, notas, etc.).

Sin embargo, es posible que el terminólogo tenga que hacer frente a situaciones en las que la equivalencia interlingüística (equivalencia entre términos de diferentes *langsets*) o intralingüística (sinonimia entre términos de un mismo *langset*) no sean perfectas. En estos casos, puede añadirse un comentario sobre transferencia (*transfer comment*), es decir, una observación que indique el grado de equivalencia de los términos. En otros casos, el grado de no-equivalencia puede ser mayor, por lo que el terminólogo deberá decidir si es necesario definir dos conceptos diferentes y crear por tanto dos entradas terminológicas independientes.

La tercera parte de la base de datos se compone de una serie de enlaces (*links*) que relacionan unas entradas (o elementos de una entrada) con otras entradas o elementos de otras entradas, o bien con un conjunto de referencias que se consideran compartidas, ya que pueden

asignarse a varios elementos (entradas o partes de ellas) de la base de datos (*SharedRefs*). Una entrada bibliográfica completa, información sobre el autor de la entrada o la fecha de redacción se considera información compartida, ya que es probable que se asignen a varias entradas o elementos de una entrada. Otro tipo de información compartida la constituyen los gráficos, tablas, imágenes y archivos de sonido, etc.

El lugar donde estas referencias compartidas se almacenan reciben el nombre de *back matter*. En el *back matter* también se incluyen otros tipos de referencias que pueden compartir varias entradas o partes de una entrada: enlaces (*links*) a elementos externos a la base de datos, como puede ser la dirección que identifica un documento o una página personal en la web.

La utilidad de este tipo de estructuración de base de datos es clara: evita que incluya información redundante, un principio fundamental en la gestión de bases de datos. Si, por ejemplo, una referencia bibliográfica o cualquier otro tipo de información compartida se incluye sólo una vez en la base de datos y varias entradas la incluyen, no físicamente, sino como un enlace, no sólo se evita la redundancia, sino que además supone un importante ahorro en cuanto a la cantidad de información que se guarda. Otra ventaja de este tipo de estructuración es que facilita la consistencia y la coherencia de la información, por lo que evita errores o variaciones posibles al incluir información repetida y hace mucho más fácil corregir o actualizar la información que se repite en una base de datos, asegurando que la corrección se hace en todos los lugares en los que esa información aparece.

Las categorías de datos contempladas por la norma ISO 12620 que se agrupan en estos tres grandes bloques (información global, cuerpo terminológico y referencias compartidas) poseen, además, otras dos características implícitas: el tipo de datos que puede incluirse en un campo determinado (numérico, alfanumérico, textual) y la lengua en la que debe consignarse, en caso de ser un campo textual. En otros casos, el valor de una categoría de datos puede estar restringido a uno o más de una lista de opciones permitidas (*pick lists*): por ejemplo, en la categoría *género gramatical* las opciones permitidas pueden ser cuatro: *masculino*, *femenino*, *neutro* y *otro*.

Esta estructuración de base de datos puede representarse gráficamente (de forma muy simplificada) con el siguiente esquema:

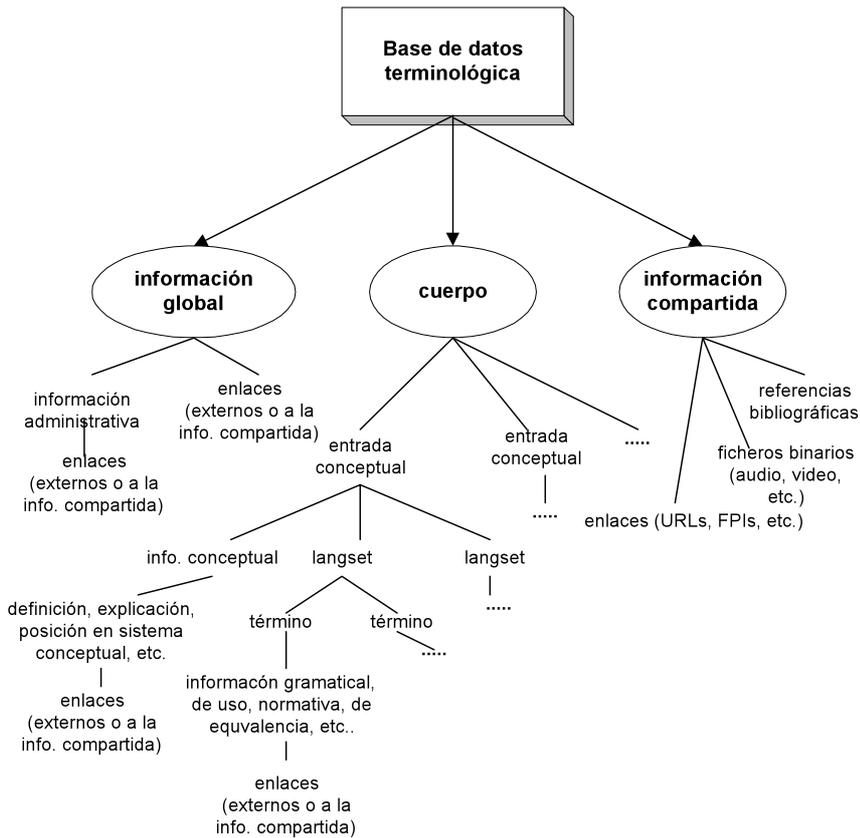


Figura 11: Resumen de la estructura de base de datos en el CLS Framework

Hay que tener en cuenta que este esquema es una simplificación de la estructura de base de datos propuesta por el *CLS Framework*, ya que es imposible representar gráficamente las posibles combinaciones de tipos y categorías de datos que pueden asignarse a otras categorías.

4.2.2 Categorías de datos de la norma ISO 12620 y su aplicación

Las categorías de datos propuestas por el comité técnico 37 de la ISO tomadas por el *CLS Framework* han sido agrupadas en diez subcategorías que se complementan con dos grupos de categorías de la norma ISO 12200. Estas categorías se resumen en el cuadro siguiente adaptado de Wright (en 2001b: 574):

Antonio Moreno Ortiz 62//00 9:41 PM
Con formato: Numeración y viñetas

ISO 12620: GRUPOS DE CATEGORÍAS DE DATOS

Categorías de datos relacionadas con el TÉRMINO

Subgrupo 1: categoría de datos *término* y contiene un término u otra información tratada como tal (ej. una unidad fraseológica o un texto estándar).

Subgrupo 2: categorías de datos que contienen información sobre el término.

Subgrupo 3: categorías de datos sobre la equivalencia entre términos asignados al mismo concepto o a conceptos similares.

Categorías de datos descriptivas, relacionadas con el CONCEPTO

Subgrupo 4: categorías de datos que clasifican los conceptos en áreas o subáreas temáticas.

Subgrupo 5: categorías de datos para la descripción del concepto, por ejemplo, diferentes tipos de definiciones, explicaciones o material contextual usado con el fin de definir el concepto o para determinar el área temática y el concepto al que se asigna un término.

Subgrupo 6: categorías de datos que indican relaciones entre dos conceptos.

Subgrupo 7: categorías de datos que se usan para indicar la posición de los conceptos en un sistema conceptual.

Subgrupo 8: categoría de datos *nota*. Esta categoría aparece aislada ya que puede asignarse a cualquier otra categoría y por tanto no puede subordinarse a ningún grupo específico.

Categorías de datos ADMINISTRATIVAS (gestión)

Subgrupo 9: categorías de datos de los lenguajes documentales y tesauros.

Subgrupo 10: categorías de datos de la información de gestión y administrativa.

ISO 12200: CATEGORÍAS DE DATOS SUPLEMENTARIAS

Subgrupo 11: especifica los códigos especiales usados en el estándar MARTIF.

Subgrupo 12: especifica las categorías de datos para la información bibliográfica.

Figura 12: Grupos de categorías de datos de la norma ISO 12620.

Estos diez subgrupos engloban un total de más de 150 categorías de datos que, a juicio de los miembros del comité técnico de la ISO, no pretenden ser exhaustivas. Tampoco se pretende en el *CLS Framework* que una base de datos incluya todas y cada una de estas categorías, sino sólo las que se consideren necesarias para representar la información pertinente a un proyecto terminológico determinado.

4.3 El editor de base de datos terminológica de OntoTerm®

El editor de base de datos terminológica (*Termbase Editor*) que hemos diseñado en la aplicación OntoTerm® es el módulo que complementa al ya expuesto *Ontology Editor* (§ 4.1). En el

Termbase Editor el terminólogo lleva cabo la tarea de describir las categorías lingüísticas, de uso y administrativas de los términos.

Puesto que un sistema basado en conceptos, no es posible introducir un término en la base de datos sin asignarlo previamente a un concepto que con anterioridad ha tenido que ser definido conceptualmente en el editor de ontologías. Por tanto, este sistema impone una metodología de trabajo acorde con la orientación conceptual del trabajo terminológico y al mismo tiempo permite una división del trabajo clara en equipos de terminógrafos.

Al igual que en el *Ontology Editor*, una nueva base de datos terminológica se crea simplemente dando un nombre de fichero y un nombre lógico. En el momento de su creación, la ontología que esté abierta en el *Ontology Editor* será considerada la “ontología de trabajo” de la nueva base de datos y los conceptos serán copiados a la nueva base de datos. A partir de este momento, todos los cambios realizados en la ontología de trabajo quedarán automáticamente reflejados en la lista espejo que contienen todas las bases de datos que la tengan como tal. No existe ningún límite en cuanto al número de bases de datos terminológicas que pueden estar asignadas a una ontología en concreto. La ontología contiene un listado de las mismas y, en caso de faltar alguna de ellas en el momento de cualquier actualización (porque el usuario la ha eliminado), se mostrará un mensaje de error y se eliminará este enlace. Del mismo modo, una base de datos terminológica previamente creada no se podrá abrir si su ontología de trabajo no es la que se encuentra abierta en ese momento.

Esto implica que el *Termbase Editor* está subordinado al *Ontology Editor*, que en todo momento deberá permanecer activado (con la ontología de trabajo en cuestión) durante el proceso de edición o navegación una de sus bases de datos terminológicas. Esto implica que no se puede abrir más de una base de datos terminológica aunque estén asignadas a la misma ontología de trabajo. Sin embargo, es posible ejecutar múltiples instancias del *Ontology Editor*, abriendo la misma ontología, es decir, *OntoTerm®* permite el acceso concurrente a sus bases de datos, de modo que esto no es realmente una limitación.

El esquema conceptual de las bases de datos que genera el *Termbase Editor* está basado en la especificación *Reltef™* (Hardman 2000), el cual ha sido adaptado para conseguir mayor velocidad de acceso mediante el uso de claves indizadas numéricas y algunas otras mejoras. Como ya hemos mencionado, implementa las categorías de datos detalladas en el apartado anterior.

La implementación de estos estándares se ha enfocado a facilitar su utilización por parte del terminógrafo, a quien se pretende aislar de la complejidad interna del modelo de datos que se

está empleando. Esto lo consigue OntoTerm® mediante una interfaz de usuario innovadora, que poco tiene que ver con las fichas terminológicas que suelen emplear los sistemas de bases de datos a los que estamos acostumbrados. Todas las categorías de datos se encuentran disponibles, pero es el usuario quien decide cuándo utilizarlas, siempre y cuando el sistema se lo permita (según sean combinaciones de categorías o tipos de datos permitidas o no en el *CLS Framework*). Al igual que en el *Ontology Editor*, el usuario es el que decide la complejidad de sus entradas terminológicas, la aplicación se limita a ofrecer las posibilidades y bloquear las incoherencias o instancias no permitidas.

En el diseño de la aplicación se ha intentado en todo momento que la interfaz gráfica de usuario sea lo más didáctica posible, y que no sea necesario poseer más que un conocimiento superficial de las categorías de datos para empezar a trabajar. Puesto que la aplicación impedirá cometer errores en cuanto al uso de las mismas, se da al terminógrafo la facilidad de experimentar con las posibilidades que se nos ofrecen, de modo que poco a poco se hace evidente la enorme flexibilidad que el *CLS Framework* permite.

Para conseguir esto, la interfaz del *Termbase Editor* se basa en dos árboles y un marco de edición. Prácticamente todas las operaciones de edición y navegación se realizan en esta misma ventana, que mostramos en la Figura 13.

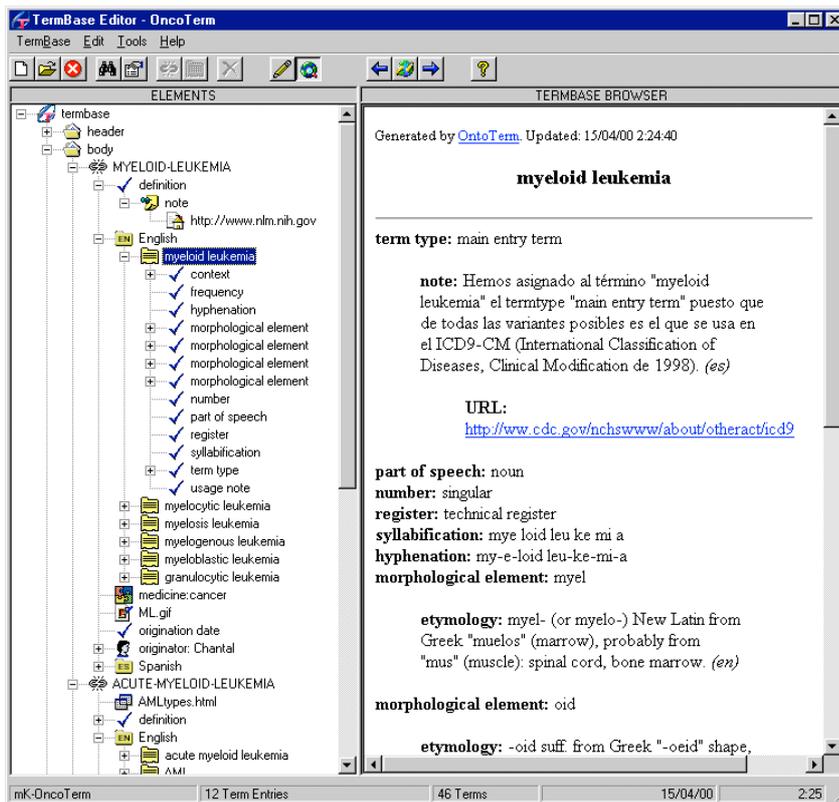


Figura 13: El Termbase Editor de OncoTerm®

El árbol de la izquierda refleja de forma gráfica la estructura jerárquica que subyace a toda base de datos terminológica que siga el *CLS Framework*. Según vimos en el apartado 4.2.1, una base de datos consta de un *encabezado*, un *cuerpo* y la información compartida. El grueso de la información terminológica se encuentra en el cuerpo de la base de datos, que engloba las denominadas *entradas terminológicas*. A éstas se les asignan los términos, que se agrupan en los denominados *langsets*, así como otros elementos de información. Lo realmente interesante de este sistema es que absolutamente todos estos elementos, al igual que cada uno de los atributos que se asignan a cualquier elemento gozan del mismo estatus por lo que se refiere al modelo de datos. Internamente, una base de datos basada en el modelo propuesto por el *CLS Framework* almacena del mismo modo todos estos elementos y se les asigna un tipo que es el que realmente les caracteriza.

La razón para ello es que de este modo podemos hacer que una categoría de datos determinada se pueda caracterizar mediante un subconjunto de categorías de datos determinado,

que incluso puede contener un atributo igual (por ejemplo, una nota se puede describir mediante otra nota). En este principio se basa el funcionamiento básico del gestor de base de datos de OntoTerm®: validar o bloquear la asignación de determinadas categorías a otras. Esto lo hace de una forma totalmente gráfica: al hacer clic sobre cualquiera de los elementos que conforman el árbol de la base de datos (el izquierdo), el árbol derecho muestra las categorías de datos válidas para el elemento seleccionado mediante el icono con la flecha azul y bloquea las no válidas mediante una cruz roja.

De esta manera podemos asignar atributos (es decir, describir) no sólo a conceptos y términos, sino también a los propios atributos. Por ejemplo, podemos asignar a la definición de un término la fuente bibliográfica de donde se ha extraído y, a su vez, describir ésta mediante un hipervínculo a una sede web determinada a la que, a su vez, le podemos asignar un código de fiabilidad bajo, indicando que no estamos seguros de esta información. En definitiva, la herramienta nos permite incluir información tan explícita o genérica como deseemos para un elemento en cuestión, sin necesidad de asignar las mismas categorías a todos los elementos de un mismo tipo.

Hemos mencionado que el editor se compone de estos dos árboles y el marco de edición. Sin embargo, como se puede observar en la Figura 13, éste último no se puede apreciar. En realidad, éste sólo aparece cuando hacemos clic en una de las categorías de datos (válida) del árbol derecho. En ese momento la aplicación procesa que deseamos añadir o editar un valor para el elemento que hemos seleccionado, y OntoTerm® muestra el marco de edición en la parte inferior del árbol izquierdo con un control apropiado según el tipo de datos de que se trate, ya sea una caja de texto para la inclusión de texto libre, una lista de la que escoger valores, un deslizador para valores en un rango numérico o un cuadro de texto con máscara para fechas (ver Figura 14).

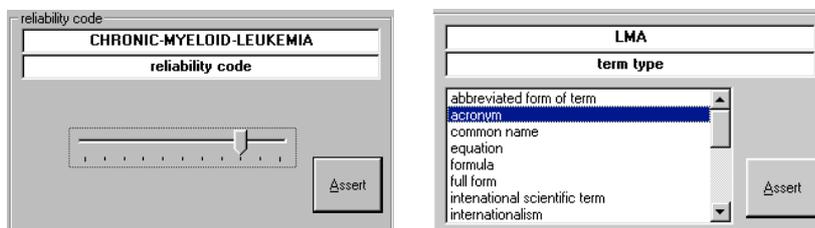


Figura 14: Ejemplos de marcos de edición en OntoTerm®

El mismo marco sirve tanto para añadir un atributo determinado como para editarlo una vez ha sido añadido, la única diferencia es que en el segundo caso, el botón de confirmación mostrará “OK” en lugar de “Assert”.

En lugar de ofrecer un listado de campos, OntoTerm® muestra las categorías de datos de forma gráfica, facilitando la localización de los mismo mediante iconos descriptivos del tipo de categoría. La inserción de remisiones o referencias cruzadas se realiza también de forma totalmente gráfica: en primer lugar seleccionamos el elemento que contendrá la referencia, después hacemos clic en la categoría de datos que implique la remisión, que puede ser bien una referencia genérica a la que deberemos dar una etiqueta (por ejemplo, “véase”) o bien una categoría de datos que requiere otro elemento de la base de datos (por ejemplo un homógrafo o un *false friend*). En ese momento OntoTerm® nos informa de que debemos seleccionar el elemento al que se remite, solicita una etiqueta en caso de ser necesaria, cambia el puntero del ratón a una flecha con interrogación y espera a que pulsemos sobre el elemento al que queremos hacer referencia. Las remisiones son mostradas mediante un icono con dos anillos y la etiqueta que los describe. Para saber a qué elemento se hace referencia deberemos usar el menú contextual (ver Figura 15).

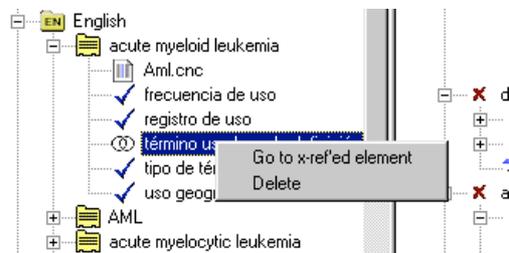


Figura 15: Gestión de remisiones en OntoTerm®

Hasta aquí hemos mostrado el modo en que creamos y editamos una base de datos terminológica en OntoTerm®. Sin embargo, las características de navegación no se limitan al árbol de la base de datos, sino que las entradas terminológicas también pueden ser visualizadas en modo de navegación. Al pasar al modo de navegación, el panel derecho se convierte en un navegador web en el cual OntoTerm® visualiza un documento HTML que genera sobre la marcha en caso de no haber sido generado previamente, para el concepto sobre el que se haga clic en el árbol de la base de datos. Esto nos permite una lectura y visualización rápida de toda la información que hayamos asignado a la entrada terminológica en cuestión. Las imágenes u otros

elementos multimedia (sonido, vídeo, etc.) también se visualizarán en este navegador (véase Figura 16).

Como se puede observar en esta captura de pantalla, OntoTerm® convierte las referencias a otras entradas terminológicas en hipervínculos, lo que nos permite, por ejemplo navegar libremente entre las mismas, por ejemplo, siguiendo remisiones.

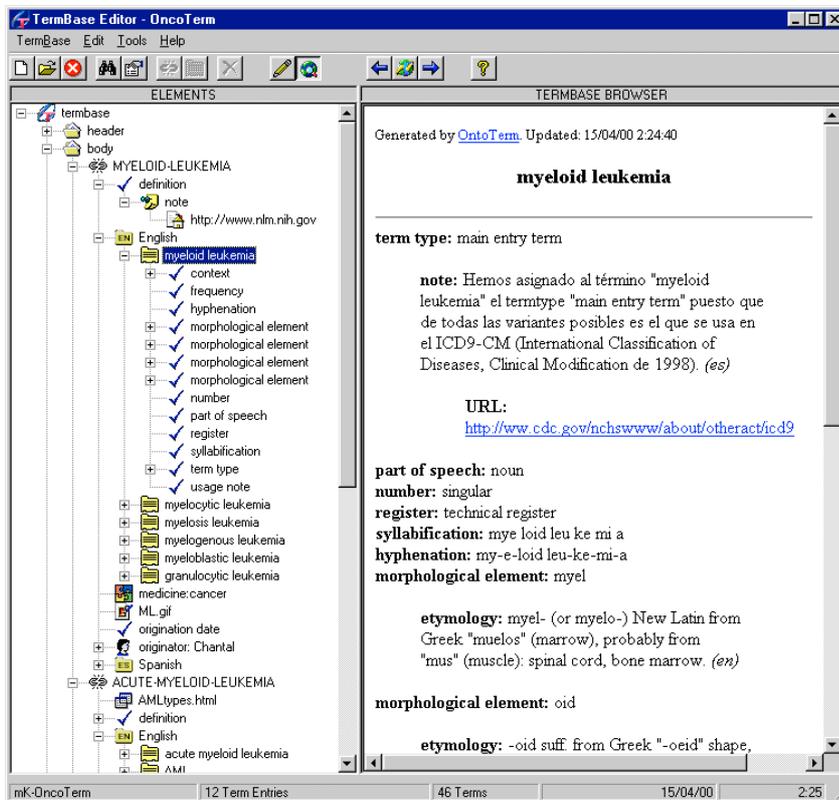


Figura 16: Ejemplo de visualización en el navegador de OntoTerm®

El sistema gestor guarda todos estos documentos HTML en un directorio especial que la aplicación genera durante el proceso de instalación, de modo que podemos utilizar cualquier navegador web para visualizarlos, modificarlos, etc. En realidad, no es necesario generar las páginas una a una, ya que, al igual que el *Ontology Editor*, el *Termbase Editor* tiene un generador de informes (véase sección 4.1, Figura 10) que permite generar masivamente todas las entradas terminológicas de nuestra base de datos, o bien una parte de ellas, en una localización determinada. Los informes generados de esta forma pueden contener toda la información

incluida en la base de datos terminológica para un concepto determinado (por ejemplo todos los términos que designan dicho concepto en las lenguas de trabajo) o, si se prefiere, se puede generar sólo la información pertinente a un término. En estas páginas, además, puede incluirse también la información conceptual del *Ontology Editor*, de forma que en una sola página puedan verse ambos tipos de información y, por tanto, navegar en modo de hipertexto por las páginas generadas correspondientes a los conceptos o los términos. En los apéndices incluimos ejemplos de documentos HTML generados por el navegador para algunos conceptos y/o términos de la base de datos terminológica.

5 Conclusiones

En este artículo hemos mostrado la filosofía y consideraciones metodológicas que han guiado el diseño e implementación de OntoTerm®, el sistema gestor de bases de datos terminológicas basado en el conocimiento usado para representar la información terminológica en el proyecto OncoTerm. Hemos discutido las implicaciones que la denominación “basado en el conocimiento” posee en el trabajo terminográfico y hemos mostrado la forma en la que el gestor de ontologías de OntoTerm® permite representar de forma consistente y completa la estructura de conocimiento del dominio de especialidad, y ha facilitado la labor de integrar dicho conocimiento en una ontología de gran amplitud y escala, como es la de Mikrokosmos. Después hemos detallado la estructura de base de datos que permite crear el sistema gestor, en el que se implementa el estándar de representación de información terminológica ISO 12620.

En estas páginas hemos detallado algunas de las ventajas que supone para un proyecto terminográfico contar con una herramienta de trabajo como OntoTerm®, entre las que destaca (i) el contar con un sistema de representación estándar, que garantiza la reutilización de los datos y la compatibilidad; (ii) el diseño de un interfaz gráfico de usuario que facilita tanto las labores de compilación de las entradas como las administrativas y la publicación de los datos; (iii) la modularidad, ya que el sistema de representación conceptual es independiente del sistema de gestión terminológica, con el que interactúa de diversas maneras, pero cuya existencia no presupone; (iv) la escalabilidad, ya que OntoTerm® está enfocado a la creación de bases de datos terminológicas a gran escala; (v) la flexibilidad, ya que es el usuario quien decide el nivel de complejidad que quiere otorgar a su sistema conceptual y a las entradas terminológicas, por lo que la aplicación impone un número lo más reducido posible de restricciones, y (vi) la integridad de los datos, derivada de las capacidades de la aplicación de bloquear las operaciones inconsistentes.

6 Bibliografía

- Beale, S., S. Nirenburg & K. Mahesh** (1995). "Semantic Analysis in the MikroKosmos Machine Translation Project", en *Proceedings of the Second Symposium on Natural Language Processing (SNLP-95)*, Agosto 2-4. Kaser Sart University, Bangkok, Tailandia.
- Gruber, T. R.** (1993). "A Translation Approach to Portable Ontologies". *Knowledge Acquisition*, 5(2): 199-220.
- Guarino, N.** (1995). "Formal Ontology, Concept Analysis and Knowledge Representation". *International Journal of Human and Computer Studies*. 43 (5/6): 625-640.
- Guarino, N.** (1998a). "Formal Ontologies and Information Systems", en N. Guarino (ed.): 3-15.
- Guarino, N. (ed.)** (1998b). *Formal Ontologies and Information Systems*. Amsterdam: IOS Press.
- Guha, R. V. & D. B. Lenat** (1990). "CyC: a Midterm Report", *Artificial Intelligence*, 11: 32-59.
- Hardman, D.** (2000). *The Reltef™ Specification*. Documento disponible en la red en: <http://www.ttt.org/clsframe/reltef/>
- Lethbridge, T.C. & D. Skuce** (1992). "Integrating Techniques for Conceptual Modeling", *Proc. 7th Knowledge Acquisition for Knowledge-based Systems Workshop*, Banff, Alberta: 15.1-15.20
- Mahesh, K. & S. Nirenburg** (1995a). "Semantic Classification for Practical Natural Language Processing", en *Proceedings of the Sixth ASIS SIG/CR Classification Research Workshop: An Interdisciplinary Meeting*. October 1995. Chicago, Illinois.
- Mahesh, K. & S. Nirenburg** (1995b). "A Situated Ontology for Practical NLP", en *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing. International Joint Conference on Artificial Intelligence (IJCAI-95)*, August 1995. Montreal, Canada.
- Mahesh, K.** (1996). *Ontology Development for Machine Translation: Ideology and Methodology*. NMSU. Computing Research Laboratory. Technical Report MCCS-96-292. New Mexico.
- Mahesh, K. (ed.)** (1997). *Ontologies and Multilingual NLP. Proceedings from the 15th International Joint Conference on Artificial Intelligence (IJCAI)*, August 23-29, 1997. Nagoya, Japan.
- Melby, A. K., K. D. Schmitz & S. E. Wright** (1997). "The Machine Readable Terminology Interchange Format (MARTIF): Putting Complexity in Perspective". *TermNet News* 54/55 (1996).
- Melby, A. K., K. D. Schmitz & S. E. Wright** (2001). "Terminology Interchange" en Wright, S. E. & G. Budin (eds.).
- Meyer, I., K. Eck & D. Skuce** (1997). "Systematic Concept Analysis within a Knowledge-Based Approach to Terminology", en S. E. Wright & G. Budin (eds.): 98-118.
- Moreno Ortiz & C. Pérez Hernández** (2000). "Reusing the MikroKosmos Ontology for Concept-Based Multilingual Terminology Databases". *Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC 2000 Athens, Greece. 31 May -2 June)*: 1061-1067.
- Moreno Ortiz, A.** (1997). *Diseño e Implementación de un Lexicón Computacional para Lexicografía y Traducción Automática*. Tesis doctoral, Universidad de Córdoba.

- Moreno Ortiz, A.** (2000a) “Managing Conceptual and Terminological Information in a User-friendly Environment”. Proceedings of OntoLex 2000. Workshop on Ontologies and Lexical Knowledge Bases.
- Moreno Ortiz, A.** (2000b) “OntoTerm: un sistema abierto de representación conceptual”. Actas del XVI Congreso de la SEPLN (Sociedad Española para el Procesamiento del Lenguaje Natural).
- Nirenburg, S.** (ed.) (1995). *The Pangloss Machine Translation System*. Joint Technical Report, Computing Research Laboratory (New Mexico State University), Center for Machine Translation (Carnegie Mellon University), Information Sciences Institute (University of Southern California).
- Nirenburg, S., D. Attardo, R. Frederking & D.** (1995). “Lexical Transfer MT Engine”, en S. Nirenburg (ed.).
- Nirenburg, S., I. Monarch, T. Kaufmann, I. Nirenburg, & J. Carbonell** (1988). *Acquisition of Very Large Knowledge Bases: Methodology, Tools and Applications*. Pittsburgh: Carnegie Mellon University. (Technical Report CMU-CMT-88-108).
- POINTER Project** (1998). *Proposals for an Operational Infrastructure for Terminology in Europe*. Documento disponible en la red. <http://www.mcs.surrey.ac.uk/AI/pointer/report>.
- Sager, J.** (1990). *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins.
- Skuce, D. & T.C. Lethbridge** (1994). “CODE4: A Multifunctional Knowledge Management System”, *Proc. 8th Knowledge Acquisition for Knowledge-Based Systems Workshop*, Banff, Alberta: 12.1-12.21
- Steve, G. A. Gangemi, D. Pisanelli** (1998a). “Integrating Medical Terminologies with ONIONS Methodology”. Documento disponible en la red. <http://saussure.irmkant.rm.cnr.it>.
- Steve, G. A. Gangemi, D. Pisanelli** (1998b). “Ontology Integration: Experiences with Medical Ontologies”, en N. Guarino (ed.): 163-178.
- Weigand, H.** (1997). “Multilingual Ontology-Based Lexicon for News Filtering –The TREVI Project”, en K. Mahesh (1997): 138-159.
- Wright, S. E. & G. Budin (eds.)** (1997). *Handbook of Terminology Management*. Vol. 1 Amsterdam/Philadelphia: John Benjamins.
- Wright, S. E. & G. Budin (eds.)** (2001). *Handbook of Terminology Management*. Vol. II. Amsterdam/Philadelphia: John Benjamins.
- Wright, S. E.** (2001a). “Terminology Management Entry Structures”, en Wright, S. E. & G. Budin (eds.).
- Wright, S. E.** (2001b). “Data Categories for Terminology Management”, en Wright, S. E. & G. Budin (eds.).
- Wüster, E.** (1968). *The Machine Tool, An Interlingual Dictionary of Basic Concepts*. London: Technical Press.