

Modelado conceptual para una terminología basada en el conocimiento

Chantal Pérez Hernández Antonio Moreno Ortiz
(mph@uma.es) (amo@uma.es)
(Universidad de Málaga)

1. Introducción

Tradicionalmente se considera que el trabajo terminológico parte de los conceptos o, mejor dicho, de una estructura conceptual que representa el conocimiento que una comunidad científica posee sobre su área de especialidad. Esta premisa determina el *modus operandi* del terminólogo, quien, en su análisis, estudia los conceptos relevantes en un dominio determinado, así como las relaciones entre ellos para, finalmente, determinar las unidades terminológicas que se usan para designar dichos conceptos.

Sin embargo, esa valiosa información conceptual asimilada por el terminólogo, que supone normalmente una gran inversión en tiempo, recursos humanos y dinero, no puede ser almacenada en las herramientas informáticas de gestión terminológica existentes. Es decir, esa información se queda en el conocimiento del terminólogo o se pierde en las notas previas a la compilación de la terminología.

La herramienta que utilizamos en el proyecto OncoTerm¹ presenta como principal característica la posibilidad de representar formalmente este conocimiento, haciendo uso de esquemas de representación típicos de la Inteligencia Artificial y presentando, al mismo tiempo, una gran facilidad de uso para el terminólogo. Esta herramienta, denominada OntoTerm™, no sólo permite, sino que impone, la creación de una ontología de conceptos, como recurso independiente de la lengua, en la que los conceptos se organizan jerárquicamente y se definen, además, mediante sus atributos y relaciones con los demás conceptos. En este trabajo exploramos las posibilidades

¹ Este trabajo se ha elaborado dentro del marco del proyecto *OncoTerm: Sistema bilingüe de información y recursos oncológicos* (PB 98/1342) subvencionado por el Ministerio de Educación y Ciencia.

que esta herramienta ofrece a los terminólogos para crear terminologías basadas realmente en un nivel conceptual, así como las dificultades que esto entraña para un trabajo terminológico sistemático.

2. 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. 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. En éste segundo significado de ontología, entendida como un cuerpo estructurado de conocimiento, es en el que centramos nuestra propuesta de terminología basada en ontologías.

El sinónimo más usual de ontología es *conceptualización*.² 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; *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):

² Guarino (1998) puntualiza que la denominación “conceptualización” se acerca más al entendimiento filosófico de una ontología como sistema de categorías que dan cuenta de una visión del mundo.

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.

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 permitirnos 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 permitirnos insertarlos en lugares diferentes de la ontología.³ 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: la ontología del proyecto Mikrokosmos (*μK*). Conocíamos ya con bastante detalle la estructura y composición de dicha ontología, puesto que había sido usada anteriormente por un miembro del grupo de investigación 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.

³ Por ejemplo, algunos tipos de quimioterapia pueden verse como un *tratamiento* para combatir el cáncer y como un *carcinógeno*, puesto que como efecto secundario pueden también provocar determinados tipos de cáncer.

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.

Estructura de mK. (all- event-object-properties).

3. Modelado conceptual en OntoTerm

Para describir de forma sucinta las funcionalidades de OntoTerm y las posibilidades que ofrece en el trabajo terminográfico, veamos cómo procedimos para incluir (y editar, borrar, modificar, etc.) la información referente a uno de los conceptos de la ontología: MYELOID-LEUKEMIA.⁴

Lo primero que tuvimos que hacer fue 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 nos venía 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:

⁴ La estructuración conceptual de los términos oncológicos y su inserción en la ontología ha sido fruto de un intenso trabajo de colaboración con otro miembro del grupo OncoTerm, M. García de Quesada, junto con los especialistas del Hospital Virgen de las Nieves de Granada integrados en el proyecto, a quienes queremos agradecer su entusiasta colaboración y su disposición a atender nuestras preguntas.

Add Concept

Concept Type

Object Attribute

Event Relation

Link Type

IS-A

Instance of

CONCEPT

MYELOID-LEUKEMIA

PARENT CONCEPT

MALIGNANT-NEOPLASM-OF-KIDNEY-AND-OTHER-AND-UNSPECIFIED
MALIGNANT-NEOPLASM-OF-LARYNX
MALIGNANT-NEOPLASM-OF-LIP
MALIGNANT-NEOPLASM-OF-LIP-ORAL-CAVITY-AND-PHARYNX
MALIGNANT-NEOPLASM-OF-LIVER-AND-INTRAHEPATIC-BILE-DUCT
MALIGNANT-NEOPLASM-OF-LYMPHATIC-AND-HEMATOPOIETIC
MALIGNANT-NEOPLASM-OF-MAJOR-SALIVARY-GLANDS

DEFINITION

form of leukemia characterized by an uncontrolled
proliferation of the myeloid lineage and their precursors in the

Add Close

Figura 1: Cuadro de diálogo para añadir conceptos.

Siguiendo el mismo procedimiento, incluimos a continuación los conceptos subordinados de MYELOID-LEUKEMIA (también obtenidos del ICD-9 CM). 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):

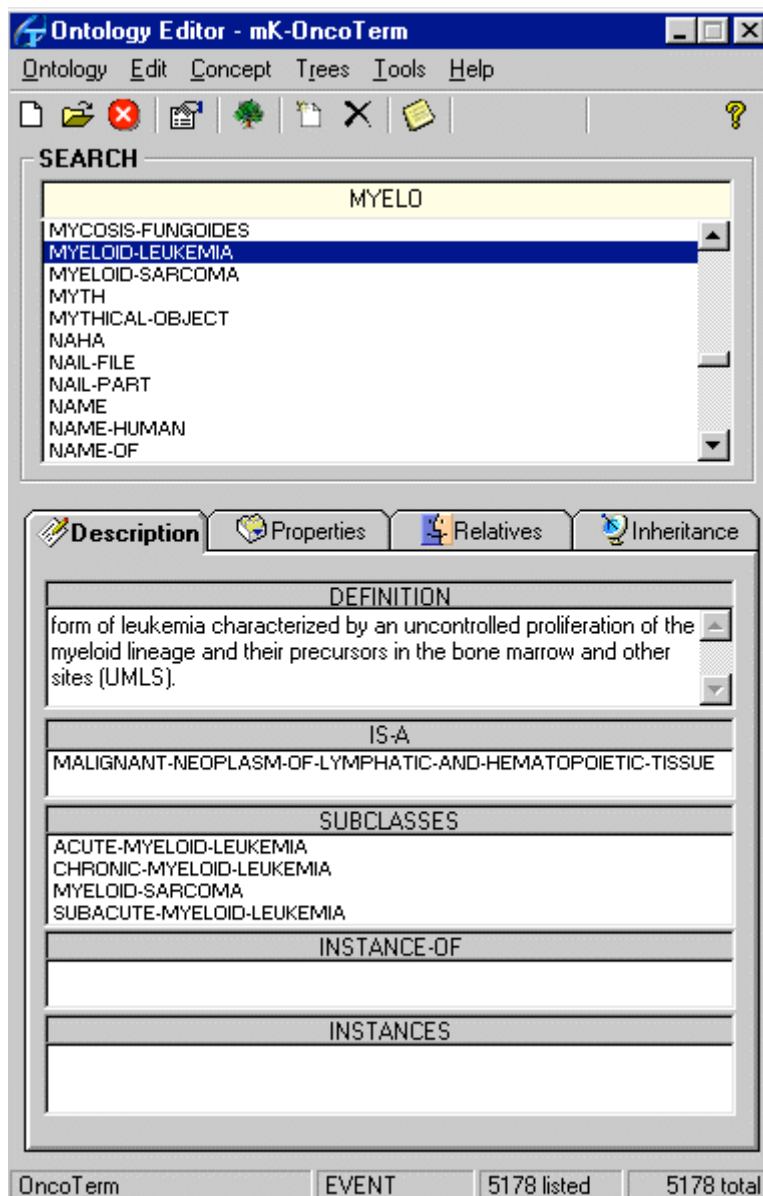


Figura 2: Pantalla de edición principal del Ontology Editor.

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 1), añadimos las RELACIONES y los ATRIBUTOS del concepto:

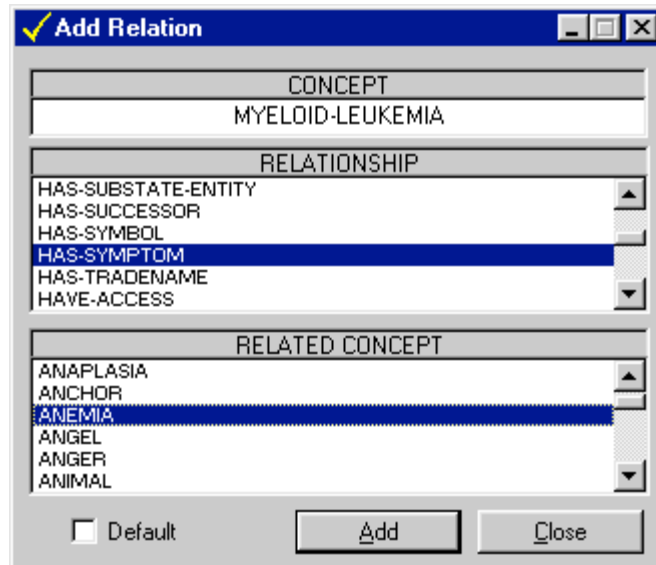


Figura 3: Cuadro de diálogo para añadir relaciones.

Al añadir esta relación, indicamos 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:

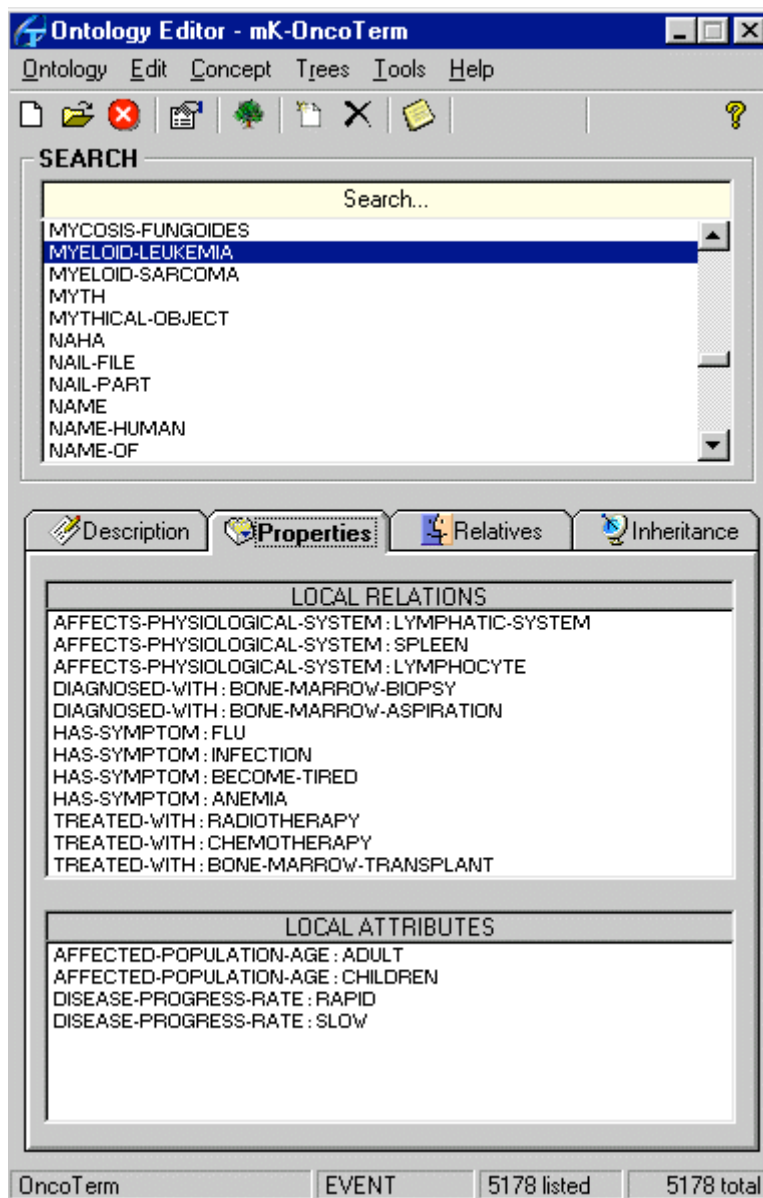


Figura 4: Relaciones y Atributos en el Ontology Editor.

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 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 4, vemos que, en nuestro ejemplo, hemos usado en la descripción de nuestro concepto las relaciones AFFECTS-PHYSIOLOGICAL-SYSTEM, DIAGNOSED-WITH, HAS-SYMPATOM y TREATED-WITH y los atributos AFFECTED-POPULATION-AGE y DISEASE-PROGRESS-

RATE. Para poder usarlas, debimos incluirlas previamente en sus lugares correspondientes de la ontología:

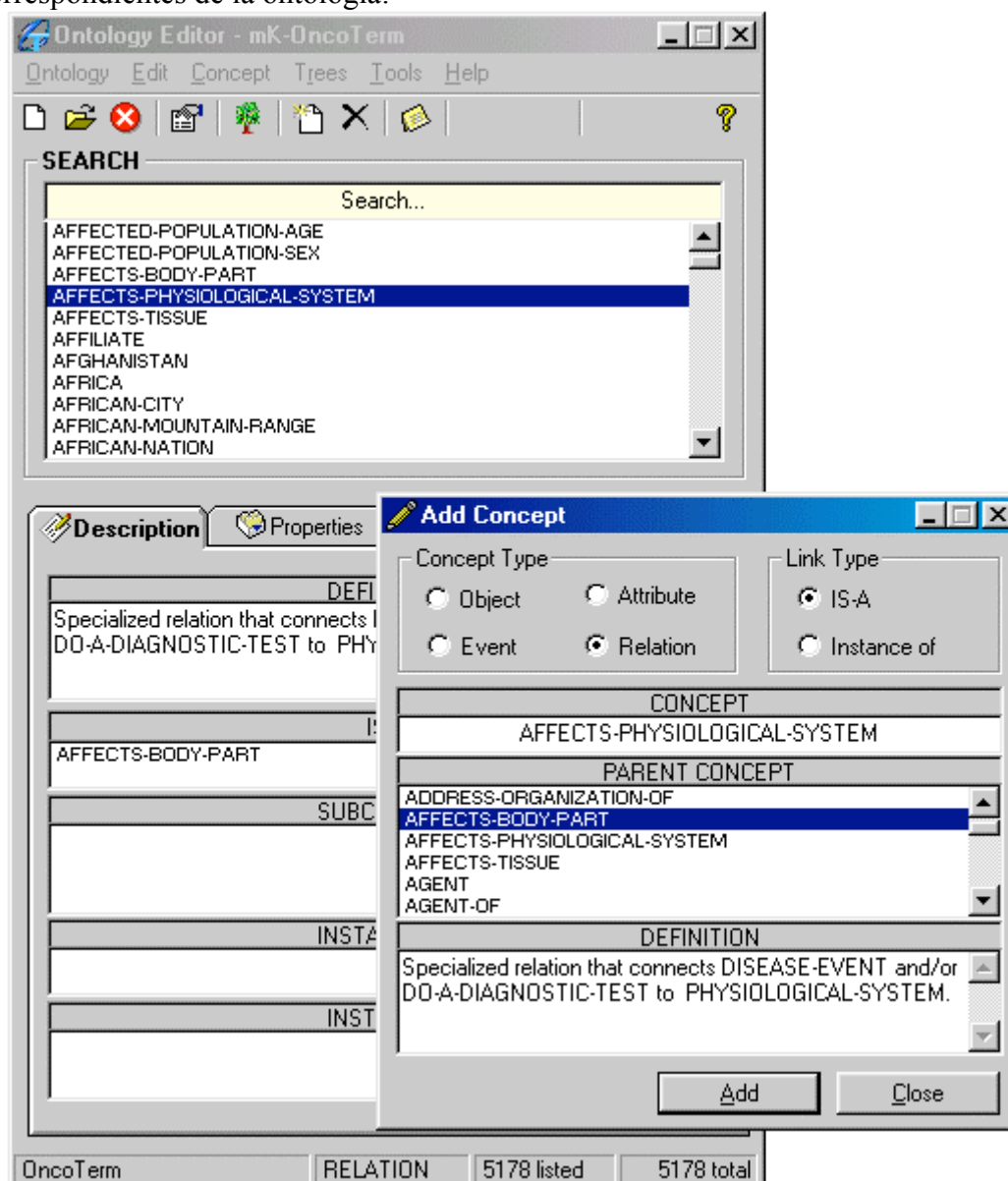


Figura 5: Relaciones en el Ontology Editor.

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 4, en la que mostrábamos 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.⁵

Hemos de recordar que la estructura conceptual de la ontología de Mikrokosmos obliga a que todo, incluyendo los elementos descriptivos (relaciones y atributos, dominios y rangos) y otras meta-entidades, sean, a su vez, conceptos en la ontología, por lo que nuestro siguiente paso fue la estructuración de los conceptos descriptivos (las propiedades). Usando la información extraída con las sondas de conocimiento siguiendo el procedimiento mostrado en la sección anterior, y ayudados por el asesoramiento de los especialistas médicos que integran el proyecto, hemos considerado necesario hasta ahora la inclusión de las siguientes relaciones, junto con sus inversos:⁶

DISEASE-EVENT-RELATION	INVERSE-DISEASE-EVENT-RELATION
DIAGNOSED-WITH	DIAGNOSTIC-TEST-FOR
HAS-RISK-FACTOR	RISK-FACTOR-FOR
HAS-SYMPTOM	SYMPTOM-OF
TREATED-WITH	TREATMENT-FOR
HAS-SIDE-EFFECT	SIDE-EFFECT-OF

DISEASE-EVENT-OBJECT-RELATION	INVERSE- DISEASE-EVENT-OBJECT-RELATION
AFFECTS-BODY-PART	BODY-PART-AFFECTED-BY
AFFECTS-PHYSIOLOGICAL-SYSTEM	PHYSIOLOGICAL-SYSTEM-AFFECTED-BY
AFFECTS-TISSUE	TISSUE-AFFECTED-BY
DIAGNOSED-WITH-THE-USE-OF	USED-IN-THE-DIAGNOSTIC-TEST-FOR
TREATED-WITH-THE-USE-OF	USED-IN-THE-TREATMENT-OF
SIDE-EFFECT-OF	HAS-SIDE-EFFECT

Las primeras (DISEASE-EVENT-RELATION e INVERSE-DISEASE-EVENT-RELATION) son relaciones que sirven para conectar unos eventos con otros, mientras que las segundas (DISEASE-EVENT-OBJECT-RELATION e INVERSE-DISEASE-EVENT-OBJECT-RELATION) conectan eventos con conceptos conceptualizados como objetos en la ontología. Además de estas relaciones, hemos incluido los siguientes atributos, cuyos valores pueden ser literales (LITERAL-DISEASE-EVENT-ATTRIBUTE) o escalares (SCALAR-DISEASE-EVENT-ATTRIBUTE):

LITERAL-DISEASE-EVENT-ATTRIBUTE	SCALAR-DISEASE-EVENT-ATTRIBUTE
CANCER-STAGING	FREQUENCY-RATE
DISEASE-PROGRESS-RATE	RECOVERY-RATE
PROGNOSIS	
RISK POPULATION	
AFFECTED-POPULATION-AGE	
AFFECTED-POPULATION-SEX	

⁵ Mostramos la posición de algunos de dichos conceptos en las capturas de pantalla que muestran los árboles de la ontología (Figuras 23 y 24).

⁶ Queremos recordar que, como dijimos en el capítulo 5, una ontología es un recurso artificial que se crea y, en el caso del proyecto OncoTerm, se expande y amplía según progresa el trabajo de nuestro equipo. En la actualidad, otro miembro del grupo, M. García de Quesada se encuentra trabajando en el uso de las relaciones conceptuales para la elaboración de definiciones terminográficas.

Con estas propiedades completamos la descripción de los conceptos de nuestro subdominio.

En muchas ocasiones, posicionar dichos conceptos en la ontología requería 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 **¡Error!No se encuentra el origen de la referencia.**, para llegar al nivel de dichos conceptos, partimos del concepto (original de la ontología de μ K) 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:

El editor de ontologías permite además muchas otras funcionalidades que sería demasiado extenso mostrar aquí en detalle;⁷ 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 en la Figura 6. 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*:

7

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

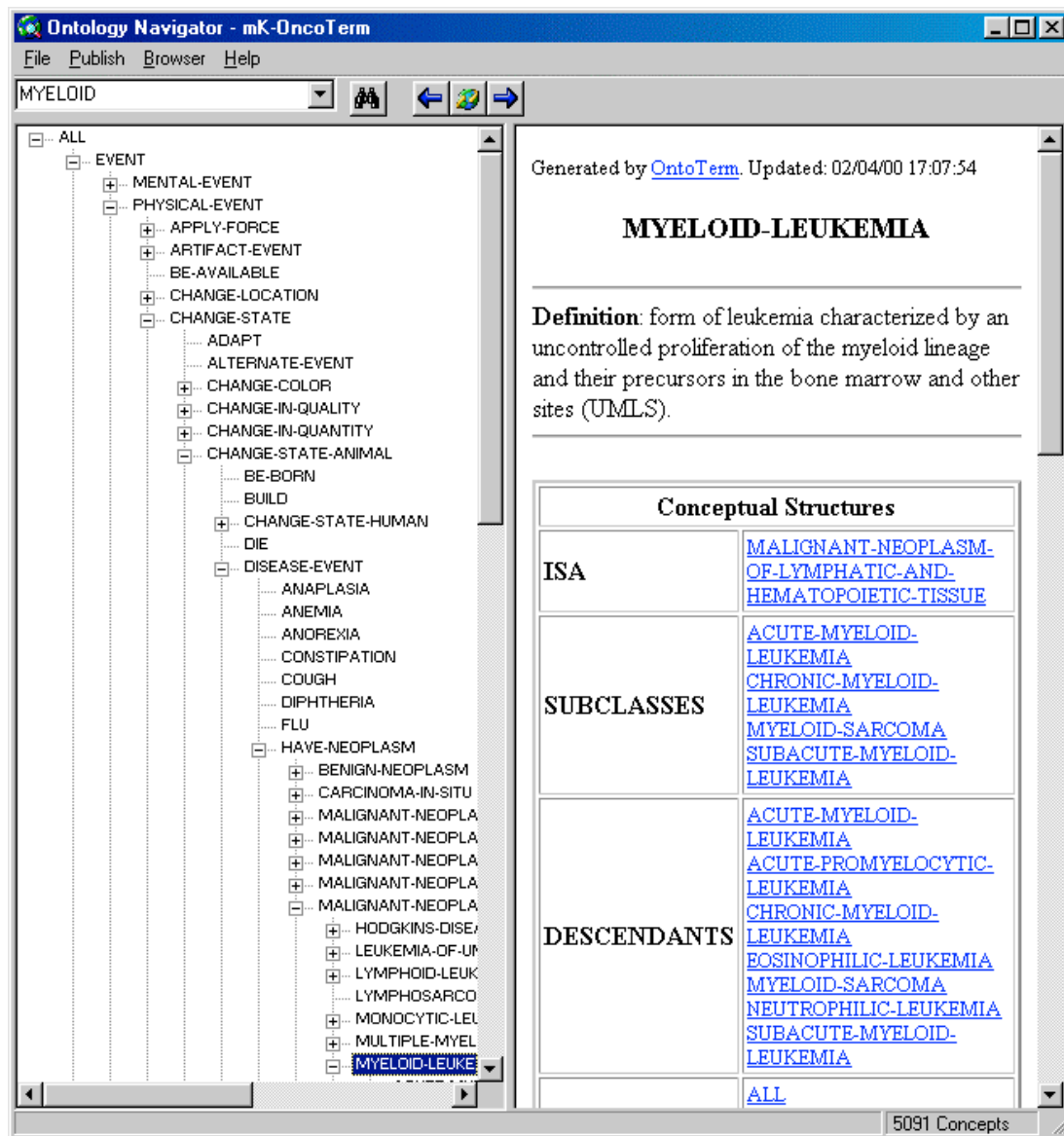


Figura 6: El "Ontology Navigator"

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.

4. Conclusión

Como puntos destacados de este sistema, deberíamos resaltar los siguientes:

- 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.
- 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.
- 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.
- Integridad: la aplicación, no el usuario, es la responsable de mantener en todo momento la integridad y coherencia interna de la información que contiene.
- 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.
- 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.
- 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.

- 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 podamos 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.