

ONCOTERM¹: Sistema Bilingüe de Información y Recursos Oncológicos

0. Introducción

ONCOTERM es un proyecto de investigación interdisciplinar sobre terminología médica, que nuestro grupo está realizando conjuntamente con el Hospital Universitario Virgen de las Nieves. Financiado por el Ministerio de Educación y Cultura, nos encontramos a punto de completar el primer año de trabajo, habiendo conseguido ya resultados importantes. El objetivo global del proyecto es la elaboración de un sistema de información específico del subdominio biomédico de la ONCOLOGÍA. Al final del mismo se espera tener a punto una amplia variedad de recursos, ofrecidos en Internet, que serán de gran utilidad para profesionales de la salud, pacientes y familiares, así como para traductores y redactores de textos especializados. Los objetivos operativos van dirigidos a la creación y configuración de una amplia base de datos terminológica sobre el cáncer. Son los siguientes:

- (1) Crear un corpus de textos médicos tanto en español como en inglés, así como una tipología para su clasificación con enlaces de hipertexto a:
 - (a) una base de textos paralelos
 - (b) un asistente en pantalla para elaborar artículos biomédicos en inglés y en español
- (2) Especificar un lenguaje de definición terminográfica conciso, consistente y aplicable no sólo al subdominio de la oncología, sino también a otras especialidades médicas y a otras lenguas
- (3) Elaborar un inventario de relaciones conceptuales específicas del EVENTO MÉDICO en general y del EVENTO MÉDICO ONCOLÓGICO en particular
- (4) Configurar una base de datos terminológica articulada en torno a la estructura hallada en la definición de los términos

¹ El grupo de investigación *OncoTerm* (PB 98-1342) lo componen los siguientes miembros de la Universidad de Granada:..... de la Universidad de Málaga....., la Universidad Nacional de Educación a Distancia..... y el Hospital Universitario Virgen de las Nieves.

- (5) Ayudar a los profesionales de la salud a redactar distintos tipos de texto siguiendo las convenciones de la comunidad médica, tanto en inglés como en español
- (6) Facilitar la traducción de terminología médica especializada mediante enlaces entre términos contextualizados en diferentes lenguas por medio de hipervínculos

1. Representación de conocimiento

La parte fundamental de *ONCOTERM* es la representación de la estructura conceptual del dominio de ONCOLOGÍA MÉDICA, sus conceptos pertinentes, características, e interrelaciones. Esta estructura conceptual subyace a todos los demás componentes y constituye el vínculo entre términos en diferentes lenguas, ya que se trata de información no específica a ninguna lengua.

Es bien sabido que el terminólogo trabaja con conceptos. La gestión sistemática de terminología empieza con la adquisición de conocimiento especializado y la consiguiente elaboración de la estructura conceptual del dominio. Una vez delimitada la estructura conceptual del dominio de especialidad, el terminógrafo debe averiguar de qué manera los conceptos pertenecientes a esta estructura están representados por unidades del léxico especializado en una o varias lenguas. Precisamente por esta razón se dice que los términos son unidades de conocimiento especializado.

No obstante, aunque los resultados de este proceso se basan indudablemente en el conocimiento, no están sistemáticamente codificados como tal. Esto significa que gran parte del conocimiento adquirido por el traductor o terminógrafo no está representado de manera explícita en el producto terminológico final. Esta situación plantea dos problemas fundamentales:

- (a) Traductores, redactores técnicos e investigadores en la recuperación de conocimiento no pueden acceder a esta información que, sin duda, les beneficiaría en la codificación, descodificación y clasificación de documentos especializados
- (b) Este conocimiento no puede reutilizarse en otras aplicaciones informáticas de procesamiento de lenguaje natural

Nosotros pensamos que una posible solución a ambos problemas es la integración de técnicas de representación de conocimiento en la elaboración de bases terminológicas donde la información conceptual está explícitamente representada. Como se verá, ésta es una de las innovaciones más interesantes de nuestro proyecto de investigación.

2. La aplicación del MLF

En *ONCOTERM* se utiliza tanto la información extraída de diccionarios como la de textos especializados para la elaboración de una base de datos terminológica cuyos conceptos están integrados en una ontología de conceptos más extensa.

Creemos que puede utilizarse la información en definiciones lexicográficas, complementada por información extraída de corpora, para hallar la organización conceptual de un dominio o subdominio especializado, y así conseguir una representación verdaderamente multidimensional.

Establecer jerarquías mediante el análisis de definiciones lexicográficas no es nuevo. Amsler (1980) lo hizo en los años ochenta para derivar información hiponímica en relación a los sustantivos del lenguaje general. La premisa básica es que la información en los diccionarios constituye una red léxico-conceptual que necesariamente tendrá correspondencia con el conocimiento que se expresa. Estas cadenas léxicas revelan los parámetros de conocimiento especificados en las *differentiae* de los lexemas así relacionados (Meijs and Vossen 1992: 144-145). Un ejemplo de jerarquía léxica en WordNet sería la siguiente:

(1) {robin, redbreast}@ → {bird}@ → {animal, animate_being}

(Miller 1998: 25)

Esto es aplicable también a la terminología. La única diferencia entre los dos tipos de jerarquías es que las referidas a los sustantivos terminológicos son más largas, ya que empiezan a niveles mucho más específicos. Sin embargo, ambas cadenas acaban con genéricos pertenecientes a la lengua general.

(2) {small cell lung cancer}@→{lung cancer}@ →{cancer}@ →{disease}@
→{event}@→{ALL}

Dos premisas básicas de nuestro proyecto son que cualquier configuración de conocimiento debe hacerse con una base teórica y que es posible elaborar una estructuración conceptual basada en la información extraída de textos, que después debe ser validada por un especialista.

2.1. Microestructura: estructura de la definición

Se puede segmentar la información en entradas lexicográficas especializadas para extraer, clasificar y relacionar diferentes tipos de información. Aunque los términos en diccionarios especializados no deben ser, en teoría, polisémicos, a menudo en sus definiciones la información se presenta de manera caótica y poco conforme con las normas de la buena redacción lexicográfica. La divergencia en la cantidad y configuración de información sobre conceptos puede comprobarse en las siguientes definiciones de *adenocarcinoma*.

adenocarcinoma A malignant tumor with cells arranged in a glandlike pattern.
(*The HarperCollins Illustrated Medical Dictionary*)

adenocarcinoma a cancer made up of abnormal gland cells on the lining or inner surface of an organ. It can develop in virtually any part of the body. Adenocarcinomas may develop in the lung, pancreas, breast, prostate, esophagus, stomach, vagina, urethra, and small intestine, among others.
(*The Cancer Dictionary*)

adenocarcinoma A malignant growth of glandular tissue. This tissue is widespread throughout the body's organs and the tumours may occur, for example, in the stomach, ovaries, and uterus. Adenocarcinomas may be subdivided into those that arise from mucous or serous secreting glandular tissue. (*Black's Medical Dictionary*)

Estas son definiciones lexicográficas extraídas de diferentes diccionarios técnicos. La lengua en que están escritas no tiene importancia porque las definiciones van a utilizarse como fuente de conocimiento. Podemos segmentarlas y así extraer los siguientes tipos de información que forman el concepto:

	<i>The HarperCollins Illustrated Medical Dictionary</i>	<i>The Cancer Dictionary</i>	<i>Black's Medical Dictionary</i>
Categoría Conceptual	malignant tumor	cancer	malignant growth
		[type] abnormal gland cells	[type] glandular tissue
Forma	cell arrangement in glandlike pattern	_____	_____
Lugar	_____	on the lining or inner surface of any organ, e.g. lung, pancreas, breast, prostate, esophagus, stomach, vagina, urethra, and small intestine	where there is glandular tissue, e.g. stomach, ovaries, uterus
Subclasificación			(i) mucous secreting glandular tissue (ii) serous secreting glandular tissue

Estas categorías de información son a su vez reveladoras de estructuras de percepción:

- a) una etiqueta que indica la pertenencia a una categoría conceptual
- b) forma percibida <FORMA>
- c) lugar <LUGAR>
- d) subclasificación <TIPO>

Al hacer este tipo de análisis sistemático, se puede llegar a una estructuración lingüística que refleja la organización conceptual del subdominio especializado donde se ubica el término, *adenocarcinoma*, y los términos más relacionados. El contexto conceptual de *adenocarcinoma* sería el siguiente:

(4)

MALIGNANT NEOPLASMS	
DIMENSION: histologic derivation	
Epithelial tissue (tissue lining a surface/cavity) (italics show most common site)	
carcinoma malignant neoplasm of abnormal epithelial tissue. squamous cell carcinoma carcinoma of squamous epithelial cells. <i>Bladder, anus and vulva.</i> oat cell carcinoma very malignant, undifferentiated, small cell carcinoma. <i>Lung, esophagus.</i> basal cell carcinoma carcinoma arising in the basal layer of the skin/ structures derived from basal cells. It invades locally, frequently forming a rodent ulcer. <i>Skin, canthus of eye, tip of nose, chin, lip.</i> [...]	
Epithelial tissue (glandular epithelium)	
adenocarcinoma malignant neoplasm made up of abnormal gland cells on the lining or inner surface of an organ. <i>Breast, bronchi, digestive tract, pancreas, endocrine glands, prostate.</i> cystadenocarcinoma adenocarcinoma, most frequently occurring as a partially solid mass with a cystic pattern. <i>Salivary gland, ovaries, breast, thyroid.</i> serous cystadenocarcinoma rare cystadenocarcinoma with frequently bilateral loculations and cysts containing transudate. <i>Ovaries.</i> pseudomucinous cystadenocarcinoma cystadenocarcinoma with cell stratification and cysts containing viscid fluid. <i>Ovaries.</i> [....]	

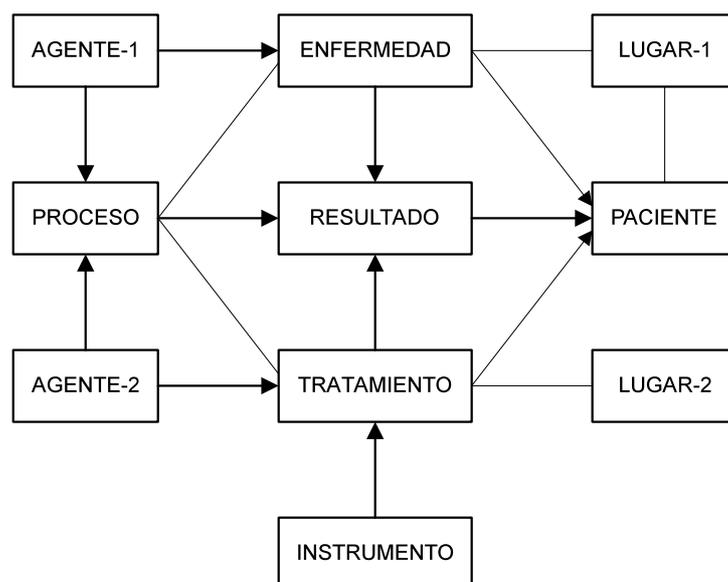
En cada definición la relación TIPO-DE está representada por el *genus* o etiqueta conceptual. Por ejemplo, la tabla indica que *serous cystadenocarcinoma* and *pseudomucinous cystadenocarcinoma* son ambos tipos de *cystadenocarcinoma*. *Cystadenocarcinoma*, a su vez un tipo de *adenocarcinoma*, que es un tipo de neoplasma maligno. Es cierto que en este ejemplo, la cadena terminológica es evidente dentro de la estructura del término, pero no es siempre éste el caso.

2.2. Categorías conceptuales

Hasta ahora, las macrocategorías conceptuales que hemos especificado en el dominio biomédico de la oncología son las siguientes:

- SÍNTOMA
- FÁRMACO
- EFECTO SECUNDARIO
- FACTOR DE RIESGO
- PROCEDIMIENTO DIAGNÓSTICO
- PARTE DEL CUERPO
- TUMOR
- TRATAMIENTO
- INSTRUMENTO
- ESPECIALISTA
- HOSPITAL

Sin embargo, también hemos tenido en cuenta estructuras más amplias de memoria, como marcos (*frames*) en el sentido de Fillmore. Dichas categorías están configuradas en el siguiente esquema relacional:



Como puede constatarse, los términos superordinados de cada categoría pertenecen a la lengua general, aunque sus niveles más específicos son propios de la terminología.

La estructura interna de cada categoría conceptual está representada por un conjunto de tipos de información. Dichos roles contienen información sobre las características de un concepto (su intensión) y también actúan como una descripción de entidades en el mundo real que pueden pertenecer a esta categoría (su extensión). Por lo tanto, cada concepto posee un esquema básico que actúa como modelo de la categoría conceptual en cuestión. Los ejemplos que siguen demuestran el esquema básico para la categoría, PROCEDIMIENTO DIAGNÓSTICO..

(5)

PROCEDIMIENTO DIAGNÓSTICO	(TIPO-DE) (FUNC) (LUGAR) (INSTR) (PROC)
----------------------------------	--

ENDOSCOPIA

TIPO-DE	procedimiento diagnóstico
FUNC	detección de una enfermedad
LUGAR	parte de cuerpo
INSTR	<i>endoscopio</i>
PROC	(1) examen con un instrumento flexible (2) resultados en imágenes del órgano /muestras de tejidos

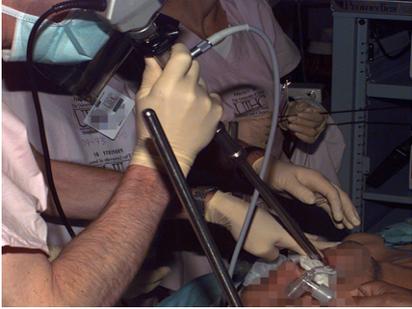
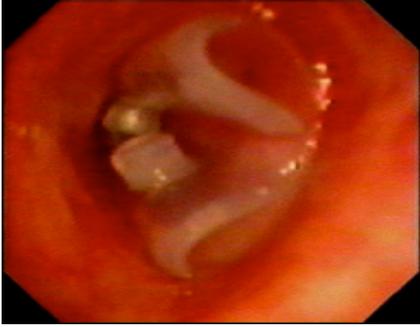
BRONCOSCOPIA

TIPO-DE	<i>endoscopia</i> de los <i>bronquios</i> y los <i>pulmones</i>
FUNC	detección de <i>cáncer de pulmón</i>
LUGAR	los <i>bronquios</i>
INSTR	<i>broncoscopio</i>
PROC	(1) <i>anestesia local</i> administrada (2) inserción del <i>broncoscopio</i> en la nariz / boca del paciente hacia el <i>pulmón</i> (3) examen visual por el médico (4) trozo de tejido de los bronquios obtenido para hacer una <i>biopsia</i> .

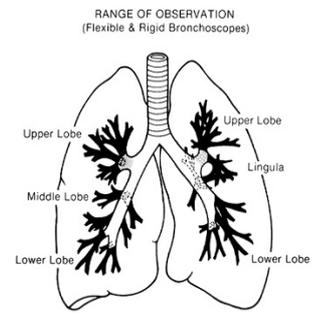
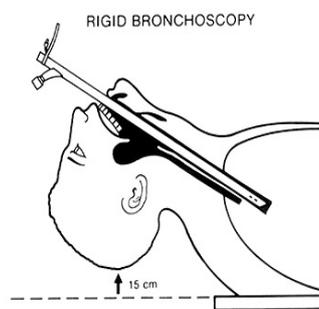
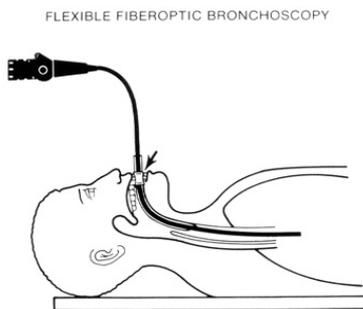
Esta información conceptual se complementaría con otras modalidades de información, como por ejemplo las siguientes:

Concepto: BRONCOSCOPIA

1. Tipo de aducto: Vídeo

<p>(1) Inserción</p>	
<p>(2) Visualización</p>	
<p>(3) Resultado</p>	

2. Tipo de aducto: imagen



3. Tipo de aducto: concordancias <contextos>

INGLÉS

1 second primary tumors are discovered. **Bronchoscopy** is not as productive
2 and pathologic (biopsy of lymph nodes, **bronchoscopy**, mediastinoscopy, or
3 gs but not visualized by imaging or **bronchoscopy** T0: No evidence of
4 en includes chest x-ray and selective **bronchoscopy** with close follow-up
5 Direct nasopharyngoscopy, laryngoscopy, **bronchoscopy**, and esophagoscopy,
6 at lead into the lungs. **bronchoscopy**: A procedure
7 the lung, as well as the lung. **Bronchoscopy** may be used to
8 ir passages that lead into the lungs. **bronchoscopy**: A procedure in
9 lead to the lung, as well as the lung. **Bronchoscopy** may be used to

ESPAÑOL

1 tivo de esputo (espontáneo o inducido), **broncoscopia** con lavado
2 uesto incorporar la panendoscopia (ORI, **broncoscopia** y esofagoscopia) en
3 les pero no ha sido visualizado por **broncoscopia** o imágenes. T0:
4 ye con frecuencia rayos X del tórax y **broncoscopia** selectiva con un
5 ocal una férula de metal extendible vía **broncoscopia**, lo que da como
6 dos tumores primarios subclínicos. La **broncoscopia** no es tan
7 undos tumores primarios subclínicos. La **broncoscopia** no es tan
8 Nasofaringoscopia, laringoscopia, **broncoscopia** y
9 tivo de esputo (espontáneo o inducido), **broncoscopia** con lavado

La inclusión de dichos archivos facilita la presentación del mismo concepto desde varias perspectivas. De esta manera se realza su naturaleza multidimensional.

Asimismo, como cada categoría tiene su propia sintaxis conceptual, las definiciones de los conceptos se formulan con la misma configuración de la información subyacente, y ganan así en consistencia y coherencia. Por ejemplo, las relaciones conceptuales *médula ósea / bone marrow* en (6) se plasmarían en las definiciones en (7):

(6)

PARTE DE CUERPO	(TIPO-DE)
	(LUGAR)
	(FUNC)
	(PERT)

MÉDULA ÓSEA

TIPO-DE	sustancia blanda y esponjosa
LUGAR	huesos (centro)
FUNC	producción de células sanguíneas
PERT	diagnóstica y tratamiento de cáncer (esp. <i>leucemia</i>)

(7)

<p>MÉDULA ÓSEA sustancia blanda y esponjosa [ETIQUETA CONCEPTUAL] situada en el centro de los huesos [LUGAR], cuya función es la producción de células sanguíneas, en particular de glóbulos rojos y plaquetas [FUNCIÓN]. En algunas variedades de cáncer, especialmente en la leucemia, la médula ósea produce células sanguíneas anormales, por lo que resulta un elemento importante en el diagnóstico y tratamiento del cáncer [PERTINENCIA].</p>
<p>BONE MARROW soft and spongy substance [ETIQUETA CONCEPTUAL] in the centre of bones [LUGAR] which is responsible for the production of blood cells, and particularly of red cells and platelets [FUNCIÓN]. In some varieties of cancer, especially leukemia, the bone marrow produces abnormal blood cells. It is, therefore, an important element for the diagnosis and treatment of cancer [PERTINENCIA].</p>

Estamos en el proceso de elaborar un *vocabulario controlado/restringido* para las definiciones. Dicho vocabulario no sería una lista de palabras, sino estructuras predicativas típicas de cada categoría. Los ejemplos que aparecen a continuación pertenecen a la macrocategoría de TRATAMIENTO:

Frase de definición	Relaciones conceptuales	Presuposición
<i>X consiste en/involves Y</i>	[X TENER-UN Y] [Y PARTE-DE X]	Un tratamiento tiene partes/ fases
<i>X es un TRATAMIENTO de Y</i> <i>X is a TREATMENT for Y</i>	[X AFECTAR _{pos} Y]	Un tratamiento positivamente afecta al paciente / el cuerpo del paciente/ enfermedad (condición física negativa).
<i>X recibe/receives Y</i>	[X BENEFICIARIO-DE Y]	Un paciente recibe un tratamiento.

resección abdominoperineal
abdominoperineal resection
<p><i>procedimiento quirúrgico que consiste en la extirpación del ano y parte del recto. Al mismo tiempo se crea un estoma para la eliminación del contenido intestinal (ver colostomía). Es un tratamiento de cáncer del recto.</i></p>

surgical procedure involving the removal of the *anus* and lower part of the *rectum* as well as the creation of a *stoma* for the elimination of body wastes (see *colostomy*). This is a treatment for *rectal cancer*.

trasplante autólogo de médula ósea

autologous bone marrow transplant

trasplante de médula ósea en que el paciente recibe su propia *médula ósea*, previamente extraída y limpiada de células cancerígenas.

bone marrow transplant in which the patient receives his/her own *bone marrow*, which has been previously removed and purged of all cancer cells.

3. ONTOTERM^{TM2}

Otra faceta importante de *ONCOTERM* es la incorporación de nuevas tecnologías. Su componente central es un sistema gestor de base de datos terminológica basado en el conocimiento, llamado ONTOTERMTM, aplicación informática que ha sido desarrollada e implementada por uno de los miembros de nuestro grupo, el Dr. Moreno Ortiz de la Universidad de Málaga. La necesidad de elaborar un nuevo recurso para el proyecto fue el resultado de las deficiencias percibidas en los sistemas gestores de bases de datos terminológicas existentes, como por ejemplo Trados MultitermTM, que aún careciendo de un sistema conceptual adecuado, gozan de una gran difusión entre los usuarios de herramientas de gestión terminológica.

Es cierto que hasta la fecha, muy pocos proyectos han empleado herramientas basadas en el conocimiento. Por esta razón, nuestro proyecto es innovador, ya que mediante ONTOTERM, se puede almacenar información terminológica en una auténtica base de conocimiento. La estructuración conceptual utilizada en esta aplicación informática depende de un recurso existente de representación de conocimiento, es decir, una ontología de conceptos que estamos en proceso de ampliar para organizar el conocimiento dentro de un dominio especializado.

Esta aplicación consta de dos módulos fundamentales. El primero, denominado *Ontology Editor*, es donde tiene lugar la representación conceptual propiamente. La

² OntoTerm se encuentra disponible en el sitio web <http://www.ontoterm.com/>

creación de esta aplicación se inspira en el software y ontología desarrollados en el Computing Research Laboratory (CRL) de la New Mexico State University bajo la dirección de S. Nirenburg (Viegas et al. 1999) para el proyecto de traducción automática basada en el conocimiento Mikrokosmos. El segundo módulo, llamado *TermBase Editor*, está enfocado a la gestión de información terminológica. Aquí el dominio conceptual modelado en el anterior adquiere su dimensión lingüística.

La función del editor de ontologías es capaz de gestionar bases de conocimientos jerarquizadas con relaciones complejas y se integra con otros módulos de gestión de información léxica y terminológica a los que suministra información. Aunque podemos crear bases de conocimiento nuevas, en nuestro proyecto hemos tenido la oportunidad de reutilizar un recurso de estas características existente previamente, la ontología desarrollada por el equipo del CRL para el proyecto anteriormente mencionado. Las ventajas de contar con un recurso de estas características son innumerables, ya que supone la disponibilidad de una fuente de conocimiento general, en la que hemos integrado el conocimiento especializado del ámbito de la oncología objeto de nuestro estudio. La metodología que hemos seguido para la integración de estos dos tipos de conocimiento se encuentra descrita en Pérez Hernández & Moreno Ortiz (2000).

La interfaz de OntoTerm ha sido diseñada para facilitar al usuario no experto la interactividad con las estructuras de datos de una forma lo más gráfica posible. Para conseguir esto, se utilizan controles estándar de Microsoft Windows y se hace uso extensivo de los menús de contexto. Todo, excepto la entrada de caracteres, se puede llevar a cabo empleando únicamente el ratón. También existen teclas de métodos abreviados para las tareas más comunes. No obstante, la interfaz no sirve únicamente como medio para facilitar las tareas de edición comunes, sino que también se encarga de aplicar un buen número de restricciones sobre los datos.

Así mismo, OntoTerm incluye una serie de herramientas destinadas a facilitar tareas de uso común, tales como la visualización gráfica de las jerarquías conceptuales, la gestión de usuarios múltiples mediante la implementación de subconjuntos de conceptos, o la revisión y publicación de información mediante un generador de informes, que permite publicar total o parcialmente la ontología en formato HTML, convirtiendo las relaciones entre conceptos en hiperenlaces.

En esta aplicación informática hemos comenzado creando una nueva base de datos terminológica, mediante la extensión de la ontología general en el programa para incluir niveles más específicos, ya que ningún término puede existir sin estar vinculado

a un concepto previamente definido. El editor de la base de datos complementa al editor de ontologías. En este módulo, se lleva a cabo la tarea de describir el término mediante diferentes tipos de categorías de datos. Aunque, por supuesto, cada término en la base tiene que haber sido asignado previamente a un concepto en el editor de ontologías. El sistema conceptual utilizado implementa las categorías del estándar ISO 12620, que sigue las indicaciones del marco CLS, resumidas en el esquema dado a continuación:

ISO 12620: GRUPOS DE CATEGORÍAS DE DATOS

Categorías de datos relacionados con el TÉRMINO

Subgrupo 1: categoría de datos que contiene un término u otra información tratada como tal (p. 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 y administrativa.

Aunque estos diez subgrupos engloban más de 150 categorías de datos, no se pretende que una base de datos incluya a todas, sino sólo a las que se consideran necesarias para representar la información pertinente a un determinado proyecto terminológico.

En nuestro caso, la información más interesante, sin duda, son las categorías de datos relacionados con el concepto (subgrupos 4-8). Con el grupo 4 se consigue asignar un concepto a un área temática. El subgrupo 7 enlaza los conceptos formando una red o estructura conceptual, establecida en el seno de un área de especialidad determinada

mediante relaciones jerárquicas. El grupo 9, por otra parte, enlaza los conceptos con los términos de un tesoro o lenguaje documentario, que suele cubrir un área de conocimiento más amplia que un sistema conceptual y que se ha construido para indexar documentos y para recuperar información en bases de datos documentales.

Las ventajas de usar el estándar englobado en la norma ISO 12620 son la versatilidad y combinación de categorías de datos que ofrece. Además, su formalismo permite expresar información terminológica de forma multidimensional, reutilizable e intercambiable.

4. Organigrama de actividades

El organigrama de actividades muestra tanto las tareas que se llevarán a cabo como la división del trabajo que suponen. La recopilación del corpus está ya totalmente terminada, así como el diseño del sistema. En este momento, estamos haciendo el vaciado, extracción y configuración de la información.

[SHOW ORGANIGRAMA?]

En el segundo año empezaremos una de las tareas más difíciles, pero también de las más interesantes: la elaboración de la arquitectura conceptual y la modelación de conocimiento. El tercer año se dedicará principalmente a la implementación total del sistema de información.

5. Conclusión

En conclusión, *ONCOTERM* es un proyecto innovador en el campo de la terminología por las técnicas de análisis, recuperación y representación de información utilizadas basadas en la semántica ontológica y la lingüística de corpus. Emplea un gestor de conocimiento terminológico, que almacena información en una auténtica base de conocimiento, ya que la estructuración conceptual utilizada se representa formalmente por medio de una

ontología de conceptos y sus relaciones. De este modo, pretende subsanar las carencias encontradas en otros recursos de terminología en Internet.

El conjunto de metodologías adoptadas en la ejecución de *ONCOTERM* nos permitirá la extensión del recurso terminológico a otros ámbitos especializados relacionados con un grado de reutilización muy alto. En concreto, la utilización de una ontología como repositorio central de conocimiento ha demostrado tener grandes ventajas y permitirá reutilizar nuestro trabajo en tareas de diversa índole.

Moreno Ortiz, A. & Pérez Hernández, C. (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.

Viegas, E., Mahesh, K., Nirenburg, S. & Beale, S. (1999). Semantics in Action. In P. Saint-Dizier, (Ed.) *Predicative Forms in Natural Language and Lexical Knowledge Bases* (pp. 171--204). Dordrecht, the Netherlands: Kluwer Academic Publishers.