



Universitat d'Alacant
Universidad de Alicante

OntoFIS: Tecnología ontológica en el
dominio farmacoterapéutico

María Teresa Romá Ferri

Tesis **Doctorales**

www.eltallerdigital.com

UNIVERSIDAD de ALICANTE



Universitat d'Alacant
Universidad de Alicante

OntoFIS: Tecnología ontológica en el dominio farmacoterapéutico

Tesis Doctoral

María Teresa Romá Ferri

Directores:

Dr. Manuel Palomar Sanz

Dr. Andrés Montoyo Guijarro

Departamento de Lenguajes y Sistemas Informáticos
Universidad de Alicante

Octubre 2009

Agradecimientos

Llegar hasta aquí ha supuesto recorrer un largo camino, durante el cual he tenido la suerte de encontrarme con muchos colegas y amigos. Sin sus aportaciones y colaboraciones, quizá, no hubiera logrado esta meta personal.

Al Departamento de Lenguajes y Sistemas Informáticos por el apoyo brindado en mi formación como investigadora. A los que fueron mis profesores, en los cursos de doctorado, por ayudarme a aprender.

A todos los colegas del Grupo de Investigación en Procesamiento del Lenguaje Natural y Sistemas de Información (GPLSI) por animarme a seguir. A todos los becarios, por hacer que siempre me haya sentido en casa. A Jesús y a Jorge por su colaboración y por nuestros debates, los cuales me estimularon a buscar y rebatir.

A Marisa, a Marisa y a Ángela por estar siempre y, sobre todo, en aquellos momentos menos gratos.

A Bea por mostrar siempre interés en lo que he ido haciendo pero, en especial, por sus lecturas y por obligarme a que le explicara. Ah!, y por sus comas y su boli rojo.

A Manolo y a Andrés, por creer y confiar en mí. Por su constante estímulo y respeto y, por brindarme la oportunidad de ser su doctoranda. Por haber hecho posible que un sueño sea una realidad.

A Tati por enseñarme lo que significaba @, por comprar nuestro primer ordenador y por renunciar, sin peros, a parte de nuestro tiempo.



Universitat d'Alacant

Esta Tesis ha sido parcialmente financiada por el proyecto 'Desarrollo de una ontología multilingüe para el dominio médico-farmacológico' (PI051438) del Fondo de Investigación Sanitaria (FIS), Instituto de Salud Carlos III.

ÍNDICE GENERAL

I. Introducción	1
1.1. Motivación	4
1.2. Objetivos	5
1.3. Organización de la Tesis	6
II. Estado de la Cuestión	11
2.1. Descripción del concepto ‘ontología’	13
2.2. Tipología de ontologías	16
2.2.1. La declaración de los componentes representados	18
2.2.2. La formalización del contenido	24
2.2.3. El ámbito de la conceptualización	25
2.3. Ontologías en Salud	27
2.3.1. Cobertura conceptual	28
2.3.2. Estructura organizativa y relaciones conceptuales	31
2.3.3. Granularidad conceptual	37
2.3.4. Grado de formalización	40
2.3.5. Usabilidad y funcionalidad de las terminologías como ontologías	41
2.4. Conclusiones	46
III. El Escenario Farmacoterapéutico	49
3.1. Características del sistema de información de la atención farmacoterapéutica	52
3.1.1. Características de la información relevante del dominio	55

3.2. Necesidades de información de los actores del escenario	57
3.3. Fuentes de información	58
3.3.1. Características del acceso y limitaciones	61
3.4. Prestaciones de los sistemas de información para la prescripción farmacológica	65
3.4.1. La prescripción electrónica	66
3.4.2. La prescripción farmacoterapéutica en la HCE	67
3.4.2.1. Aplicación de técnicas de PLN	68
3.5. Conclusiones y motivación	75
IV. OntoFIS: Desarrollo y Construcción de una Ontología Farmacoterapéutica	79
4.1. Análisis del modelo representado en las ontologías existentes	82
4.1.1. Metodologías de construcción de ontologías	85
4.2. Propuesta metodológica para el modelado de ontologías	88
4.2.1. UML (Lenguaje Unificado de Modelado)	90
4.2.1.1. Herramientas conceptuales de UML	90
4.2.2. Lenguajes de marcado Web	93
4.2.2.1. El lenguaje de metadatos RDF y RDFS	93
4.2.2.2. El lenguaje formal de ontologías OWL	97
4.3. Propuesta tecnológica: VisualWADE	102
4.3.1. Semantic Web Tools for VisualWADE	103
4.3.1.1. Limitaciones	110
4.4. OntoFIS	110
4.4.1. Diseño y construcción de OntoFIS	111
4.4.1.1. La etapa de captura del conocimiento aceptado del dominio	112
4.4.1.2. La etapa de descripción del conocimiento según el modelado conceptual	113
4.4.1.3. La etapa de formalización del conocimiento capturado	115
4.4.2. Captura del conocimiento del dominio farmacoterapéutico	117
4.4.3. Descripción del conocimiento con la técnica de modelado UML	119
4.4.3.1. El modelo de información de OntoFIS	126
4.4.4. Formalización del conocimiento farmacoterapéutico en OWL	127

4.4.5. Conclusiones	128
V. Evaluación de OntoFIS	131
5.1. Método de Evaluación	132
5.1.1. Indicadores de calidad	134
5.1.2. Procedimiento de evaluación	135
5.1.3. Herramientas para la evaluación	137
5.1.3.1. Protégé	137
5.1.3.2. FaCT++	139
5.1.3.3. DL Query	140
5.2. Evaluación de verificación	140
5.2.1. Causas de las limitaciones detectadas	145
5.2.2. Soluciones aplicadas	149
5.2.3. Verificación de las soluciones aplicadas	154
5.3. Evaluación de validación	156
5.3.1. Banco de cuestiones de competencia	157
5.3.2. Población de OntoFIS	158
5.3.3. Primer experimento: vocabulario completo	160
5.3.3.1. Selección de cuestiones	160
5.3.3.2. Transcripción de cuestiones en DL Query	162
5.3.3.3. Análisis de resultados	164
5.3.4. Segundo experimento: exactitud de la inferencia	166
5.3.4.1. Fragmentación de la base de conocimiento de OntoFIS	166
5.3.4.2. Medicamentos diana	168
5.3.4.3. La interrogación de la base de conocimiento	169
5.3.4.4. Análisis de resultados	170
5.3.5. Refinamiento	172
5.3.6. Experimento de comprobación de mejoras	172
5.4. Acreditación de OntoFIS	173

VI. Conclusiones y Trabajos Futuros.....	177
6.1. Aportaciones	178
6.2. Trabajos futuros	182
6.3. Producción científica	184
 Bibliografía	 187
 Anexos	 205
ANEXO I. Artículos y otras fuentes documentales seleccionadas	I
ANEXO II. Total de términos recopilados de las fuentes documentales	V
ANEXO III. Un fragmento del glosario creado	VII
ANEXO IV. Descripción informal de la ontología OntoFIS	IX
ANEXO V. Fragmentos del ‘diagrama de descripción de OntoFIS’	XXI
ANEXO VI. Grupos terapéuticos utilizados para la selección del conjunto de instancias de medicamentos de la base de conocimiento de OntoFIS	XXIII

ÍNDICE DE FIGURAS

2.1. Clasificación de las ontologías en función de los componentes declarados y la riqueza de su estructura interna	23
3.1. Fases del proceso farmacoterapéutico en relación con el proceso sanitario	50
3.2. Sistemas de Salud y de Atención Sanitaria	51
3.3. Componentes del SI y flujo de la información en el proceso de atención farmacoterapéutica ..	54
3.4. Representación de los módulos básicos de una HCE con información farmacoterapéutica ..	68
4.1. Identificación de las entidades en dos cuestiones de petición de información	80
4.2. Correspondencia entre las etapas de construcción de una ontología, de un esquema conceptual y el ciclo de vida de evolución del software	87
4.3. Colección de elementos del diagrama de clases UML para el modelado declarativo	91
4.4. Diagrama del procedimiento de transformación de <i>Semantic Web Tools for VisualWade</i> ...	104
4.5. Diagrama de clases de la ontología de referencia para las transformaciones	104
4.6. Interfaz para ampliar los campos de definición de los roles	109
4.7. Esquema de la propuesta de metodología	112
4.8. Proceso de retroalimentación para la toma de decisiones de la primera etapa	113
4.9. Proceso de verificación para la toma de decisiones de la segunda etapa	115
4.10. Proceso de verificación de la última etapa de desarrollo de una ontología	117
4.11. Sistemática de recogida de los términos relevantes de la atención terapéutica	118
4.12. Desarrollo inicial del modelado de OntoFIS en CASE VisualWADE	124
4.13. Creación de una asociación en la herramienta CASE VisualWADE	125
4.14. Fragmento del diagrama de clase UML de OntoFIS (representación no estándar)	125
4.15. Interfaz del plug-in para la inclusión de etiquetas terminológicas	130

5.1. Visualización de las clases de OntoFIS en la ventana principal de Protégé	142
5.2. Visualización de las propiedades de las clases de OntoFIS (<i>Datatype property</i>) en Protégé	143
5.3. Visualización de las relaciones entre clases de OntoFIS (<i>Object property</i>) en Protégé	144
5.4. El renombrado de las asociaciones con numerador secuencial	145
5.5. Fragmento del nuevo ‘diagrama de descripción’ de OntoFIS	150
5.6. Visualización en Protégé del fichero OWL DL del ‘diagrama para la transformación OntoFIS’	151
5.7. Jerarquía de relaciones semánticas incluidas como ‘ <i>Object Property</i> ’ en Protégé	153
5.8. Jerarquía diseñada y jerarquía inferida por FaCT++ en Protégé	155
5.9. Fragmento de la base de conocimiento OntoFIS para validar la adecuación del contenido ..	167
5.10. Resultado de la interrogación de la base de conocimiento OntoFIS en Protégé 4 beta	170
6.1. Interconexión de OntoFIS y otras ontologías, como propuesta de trabajos futuros	183

ÍNDICE DE TABLAS

2.1. Categorización de las ontologías	17
2.2. Uso práctico de las ontologías de salud en Gestión de Conocimiento	44
2.3. Uso práctico de las ontologías de salud en Interoperabilidad Semántica	45
2.4. Uso práctico de las ontologías en aplicaciones para el PLN	45
3.1. Variantes de denominación de un medicamento	56
3.2. Necesidades de información de los actores en la atención farmacoterapéutica	58
3.3. Tipo de contenido informativo de las fuentes de información sobre el medicamento.	61
3.4. Fragmento de la representación del conocimiento farmacoterapéutico en el MeSH	63
3.5. Fragmento del índice del campo título de la base de datos IME	63
3.6. Tiempo invertido para obtener respuestas a las cuestiones que surgen durante la atención ..	65
3.7. Características de cuatro aplicaciones comerciales de PLN en salud	71
4.1. Cobertura conceptual farmacoterapéutica en las ontologías de salud	83
4.2. Fuentes de conocimiento en español integradas en el Metathesaurus de UMLS	84
4.3. Etapas fundamentales de las metodologías para la construcción de ontologías	86
4.4. Las tres variantes de representación de los recursos y las relaciones en RDF	94
4.5. Vocabulario básico de RDFS	96
4.6. Variantes del vocabulario para la denominación de concepto, propiedad, relación e instancia de acuerdo al modelo de datos	114
4.7. Lista anidada de los conceptos seleccionados para diseñar OntoFIS	121
4.8. Relaciones asociativas entre conceptos, a partir de fragmentos de conocimiento capturado	122
4.9. Taxonomía de relaciones empleadas en OntoFIS	123
4.10. Transformaciones entre UML y OWL DL realizadas por <i>Semantic Web Tools for VisualWADE</i> en la formalización de OntoFIS	127
5.1. Categorías y número de cuestiones del banco de preguntas	158

5.2. Correspondencias entre las clases en OntoFIS y los campos de las tablas de ‘Digitalis’	159
5.3. Cuestiones seleccionadas del banco de preguntas para la evaluación de validación	161
5.4. Transcripción de las 6 primeras cuestiones en lenguaje DL Query	162
5.5. Transcripción de las cuestiones de la 7 a la 15 en lenguaje DL Query	163
5.6. Transcripción de las cuestiones de la 16 a la 23 en lenguaje DL Query	164
5.7. Resultados de la calidad de la inferencia de OntoFIS.	170
5.8. Cuestiones adicionales creadas en DL Query con la terminología de OntoFIS	173



Universitat d'Alacant
Universidad de Alicante

Capítulo I.

Introducción

Supongamos que un médico, una enfermera¹ o un farmacéutico, mientras atienden a un ciudadano, en una consulta, consideran necesario que éste comience a tomar un medicamento por un problema de salud. Para poder seleccionar el medicamento más idóneo y las alternativas para su uso adecuado, estos profesionales pueden necesitar de la consulta de fuentes de

¹ En el momento de elaboración de esta Tesis, en España, se está produciendo la modificación del apartado 1 del artículo 77 de la Ley 29/2006 de 26 de julio, de garantías y uso racional de los medicamentos y productos sanitarios. Esta modificación posibilita la prescripción de determinados medicamentos a otros profesionales sanitarios. Textualmente, el artículo señala: "...regulará la participación en la prescripción de determinados medicamentos por enfermeros y podólogos en el marco de los principios de atención integral de salud y para la continuidad asistencial, mediante la aplicación de protocolos institucionales de elaboración conjunta y en planes de cuidados estandarizados, autorizados por las autoridades sanitarias" [Sesión Plenaria 28 de abril de 2009. Diario de Sesiones del Congreso de los Diputados, nº79].

En el contexto europeo, países como Reino Unido, Holanda, Suecia e Irlanda, las enfermeras tienen legislado un procedimiento formal para la prescripción de medicamentos, al igual que ocurre en Estados Unidos, Canadá, Australia, Nueva Zelanda, Brasil o Argentina (Consejo Internacional de Enfermería).

información como prospectos, informes técnicos, resultados de ensayos clínicos o guías clínicas, por ejemplo. En otras palabras, necesitan de un acceso a una información de carácter textual especializada.

Para ello, sería muy útil una herramienta que, automáticamente, encontrase los fragmentos de información relevante dispersos en diferentes fuentes, sobre los medicamentos candidatos, en función del problema de salud a tratar, las precauciones necesarias a adoptar y los inconvenientes que pueden surgir, respecto a otros problemas de salud que el ciudadano pueda tener, o bien a otras medicaciones que pueda estar tomando. Esa misma herramienta, además, debería presentar la información organizada, a partir del medicamento o del problema de salud, para facilitar la toma de decisión del profesional en un reducido margen temporal. Desgraciadamente, esa herramienta no está disponible todavía.

En su lugar, los profesionales deben de utilizar distintas fuentes de información y deben realizar varias consultas hasta obtener un documento o un conjunto de documentos con la información potencialmente relevante. Posteriormente, deben leer, extraer y organizar la información de interés que determine la idoneidad y el uso del medicamento. Para ello, los profesionales deben invertir un tiempo, del que muchas veces no disponen, y demorar la atención directa al ciudadano.

Las tecnologías de información y comunicación (TIC) han penetrado, progresivamente, en el ámbito de la salud, y han cambiado la forma de comunicarse, diseminar y acceder a la información por parte de los profesionales. Pero, a pesar de ello, los sistemas de información (SI) capaces de interactuar e intercambiar la información, automáticamente, son insuficientes. Más escasos, aún, son los SI que pueden llegar a compartir el significado de la información contenida en cada uno de ellos (CE, 2008). En la actualidad, todavía se mantienen SI independientes para cada una de las instituciones, departamentos y servicios que componen el sistema de

atención sanitaria. El resultado es una información fragmenta, no integrada y, casi, imposible de compartir.

En el caso concreto de la atención farmacoterapéutica, la información viene determinada por la gran cantidad de medicamentos y de presentaciones comercializadas factibles de ser utilizadas. Por ejemplo, en España, en 2008, se comercializaban 17.204 presentaciones de medicamentos (CGCOF, 2009). Por otra parte, hay diversas fuentes de información referentes a la idoneidad de estos medicamentos y su uso adecuado. Estas fuentes difieren tanto en la estructura organizativa de la información como en el contenido informativo facilitado. Esta situación obliga a los profesionales, como hemos mencionado anteriormente, a tener que conocer el contenido y la forma de acceder a cada uno de los SI para obtener la respuesta a sus preguntas.

En la actualidad, los SI pueden conectarse e intercambiar su información, pero el reto está en que sean capaces de compartir la información entre ellos. En otras palabras, el problema se sitúa en proporcionar soluciones de interoperabilidad semántica. El concepto de interoperabilidad semántica hace referencia a una amplia variedad de cuestiones relacionadas con el contexto, en el que se crea y se utiliza la información, y aquellas relativas a la infraestructura, la estructura, la sintaxis y la semántica de la información (IDABC, 2004; RIDE, 2007). Los acuerdos respecto a la semántica de la información (significado de los datos), se centran en la forma de representar y descubrir una información en función de un contexto. Estos acuerdos deben asegurar que el significado de la información intercambiada es preservado e interpretado, adecuadamente, por el SI, tanto emisor como receptor, para poderla compartir (IDABC, 2004; EIF, 2004).

La tendencia actual, para facilitar soluciones a estos problemas de significación de los datos, es la incorporación de tecnologías semánticas y, en concreto, el empleo de ontologías. Las ontologías son un tipo de

representación del conocimiento, basado en conceptos, que facilita la comprensión del significado y del contexto de la información. Una de sus ventajas es que puede ser utilizada tanto por los humanos como por las aplicaciones informáticas (Gruber, 1993; CE, 2004; RIDE, 2007; Lewalle et al, 2008).

Las ontologías son clave para un SI al ser como un repositorio de información conceptualizada y relacionada entre sí, de forma consistente y no redundante. Por otro lado, son un recurso que facilita el razonamiento inductivo y la extracción o la recuperación de información a partir del conocimiento representado, explícitamente, en ellas.

En este sentido, el interés por las ontologías ha crecido, tanto en ámbitos generales como en específicos, conforme se ha centrado la atención en la necesidad de reutilizar y compartir el conocimiento para promover la integración de datos y de recursos, la gestión de conocimiento, el acceso a la información, los sistemas de apoyo a la toma de decisiones y el propio desarrollo de la Web Semántica, así como el procesamiento del lenguaje natural (PLN) (Guarino, 1998; Rodríguez, 2000; Bodenreider, 2008).

1.1. Motivación.

A pesar del creciente desarrollo de las ontologías y de la existencia de diversas fuentes de conocimiento terminológico, utilizadas en PLN, la representación de los medicamentos es deficitaria desde la perspectiva de la práctica asistencial.

Por otro lado, en este contexto, la información sobre la atención farmacoterapéutica se genera en la lengua nativa de los profesionales (en

nuestro ámbito geográfico el castellano), mientras que, en las fuentes de conocimiento del dominio de la salud, el predominio es la lengua inglesa.

En el ámbito internacional se están utilizando las técnicas de PLN para agrupar y clasificar (reconciliar) la información, sobre el tratamiento farmacológico, dispersa en los módulos que conforman la historia clínica electrónica (HCE). Reto que, en breve espacio de tiempo, también, se producirá de forma más extensiva en nuestro entorno.

La motivación para la realización de esta Tesis Doctoral se fundamenta, por ello, en la necesidad de aportar soluciones al ámbito de la terapéutica farmacológica en lengua castellana. Este trabajo incluye una aportación metodológica y tecnológica para el diseño y la construcción de una ontología para el dominio farmacoterapéutico, OntoFIS (Ontología Farmacoterapéutica e Información para el Seguimiento), con la finalidad de cubrir el vacío existente.

1.2. Objetivos.

El principal objetivo de esta Tesis es la definición de una propuesta metodológica y tecnológica, para la construcción de la Ontología Farmacoterapéutica e Información para el Seguimiento, OntoFIS. Este objetivo general se ha alcanzado a través de los siguientes objetivos específicos:

- Analizar el escenario farmacoterapéutico para establecer la cobertura y los vacíos de las fuentes de conocimiento que existen, así como el tipo de tecnologías que se emplean para representar el conocimiento en este ámbito.

- Valorar las metodologías existentes para la construcción de ontologías y establecer una propuesta metodológica y tecnológica para el diseño y la construcción de OntoFIS.
- Determinar el procedimiento de evaluación y los indicadores cualitativos y cuantitativos que acreditan la calidad de la ontología construida.
- Poblar OntoFIS con instancias de fármacos registrados y comercializados en España y crear un banco de cuestiones, basadas en las necesidades reales de los usuarios finales, para ejecutar los experimentos de evaluación de la adecuación del contenido de OntoFIS.
- Determinar la calidad del modelo y la adecuación del contenido representado en OntoFIS para su utilización como fuente de conocimiento.

1.3. Organización de la Tesis.

En este primer capítulo se ha presentado, brevemente, el problema de la información relevante en el dominio de la atención farmacoterapéutica, justificando que éste sea un área prioritaria de *eHealth*. Asimismo, se ha esbozado la propuesta de construcción de una ontología, como una tecnología para solucionar o facilitar la utilización de la información desde diversos perfiles de usuarios. También, se han destacado los objetivos que han orientado este trabajo. Como último apartado, de este capítulo, se

presentan los capítulos que estructuran esta Tesis y una breve descripción de los contenidos que aborda cada uno de ellos.

El segundo capítulo se estructura alrededor de tres núcleos. En primer lugar, se analiza, desde una perspectiva teórica, el concepto de ontología. En segundo lugar, se describen distintas tipologías de clasificación de las ontologías a partir de las definiciones de mayor relevancia. La descripción incluida se organiza atendiendo a tres criterios: el tipo de componentes declarados, el grado de formalización del contenido y el tipo de conceptualización del modelo de las ontologías existentes. Por último, se describen los artefactos clásicos utilizados para la representación de conocimiento, en el ámbito de la salud, desde una perspectiva ontológica (clasificaciones, taxonomías y nomenclaturas) y las funcionalidades desarrolladas a partir de estas fuentes de conocimiento.

En el tercer capítulo se establece el alcance global del escenario farmacoterapéutico. En sus distintos apartados se realiza un diagnóstico de la situación actual de los sistemas de información disponibles y de los requisitos necesarios, desde la perspectiva de los profesionales sanitarios, para el desarrollo del proceso de la atención farmacoterapéutica. El primer apartado contiene una descripción general del sistema de la atención farmacoterapéutica y de las características de la información relevante para este proceso. En el siguiente punto, se establecen las necesidades de información de los actores que intervienen en el escenario de la terapéutica farmacológica. El tercer apartado analiza la cobertura actual de las fuentes de información empleadas, con mayor frecuencia, en este dominio, así como sus características y limitaciones para acceder a sus contenidos informativos. Se continúa con un apartado en el que se identifican las prestaciones de los sistemas de información, centrados en la prescripción farmacológica, tanto en la atención ambulatoria como en la hospitalaria. En

este último ámbito, se reseña el uso de las técnicas de PLN para la identificación y la agrupación de la información sobre el tratamiento farmacológico disperso en la historia clínica electrónica (HCE).

En el cuarto capítulo se presenta la metodología de construcción de OntoFIS para la captura, la representación y la explicitación formal del conocimiento aceptado del dominio farmacoterapéutico, en el que se refleja una perspectiva compartida por médicos, enfermeras y farmacéuticos.

Este capítulo se estructura en cuatro apartados. En el primero de ellos, se analiza el modelado de las ontologías existentes y las metodologías de construcción de las mismas. El siguiente apartado se centra en la propuesta de modelado. En él, se describe la técnica de modelado conceptual UML y el lenguaje de representación de ontologías OWL DL. En el tercer apartado, se establece la propuesta tecnológica empleada y se especifica el funcionamiento del plug-in *Semantic Web Tools for VisualWADE*, utilizado para la transformación del modelado del diagrama de clases UML a OWL DL. El último apartado se dedica a la construcción de OntoFIS. En él, se detalla el proceso de captura del conocimiento del dominio farmacoterapéutico, el modelado conceptual de OntoFIS en UML y su formalización en OWL DL, de acuerdo con la metodología propuesta.

El proceso de evaluación de OntoFIS, con los experimentos llevados a cabo para asegurar la calidad de la formalización obtenida, se presenta en el quinto capítulo. Éste se ha organizado en cuatro apartados. El primer apartado tiene por objetivo presentar las distintas perspectivas existentes de evaluación de las ontologías y fundamentar el método seguido para la evaluación, de verificación de la estructura y de validación de los contenidos, del modelo de OntoFIS. El siguiente apartado presenta los resultados obtenidos en el proceso seguido para la verificación de OntoFIS, así como las causas de las limitaciones detectadas, inicialmente, y las

soluciones de mejora aplicadas. El tercer apartado detalla la evaluación de validación del contenido. En él, se incluye la descripción del banco de cuestiones de competencia, creado a partir de los patrones de preguntas que más frecuentemente se realizan los profesionales de la atención, y el proceso de población de OntoFIS. Asimismo, se presentan los resultados de los tres experimentos de evaluación realizados y el análisis de los mismos. En el último apartado se establece la acreditación de calidad del modelo representado en OntoFIS, a partir de los indicadores de calidad comprobados en el proceso de evaluación.

Para finalizar, en el sexto capítulo, se presentan las conclusiones más destacadas y las principales contribuciones aportadas con el desarrollo de esta Tesis. Seguidamente, se incluye la propuesta de trabajos en progreso y se finaliza con la presentación de la producción científica derivada de esta Tesis.

Capítulo II.

Estado de la Cuestión

La interoperabilidad semántica implica un acuerdo sobre la forma de representar, descubrir y dar un contexto a la información. Esto permite que las aplicaciones informáticas puedan llegar a compartir la información y no sólo a intercambiarla. El objetivo de la interoperabilidad semántica es obtener información interpretable automáticamente y, por consiguiente, reutilizable por aplicaciones que no participaron en su creación. Para lograr este objetivo se debe asegurar que el significado preciso de la información intercambiada sea preservado y entendido, adecuadamente, por el sistema de información receptor respecto al sistema de información emisor (IDABC, 2004; EIF, 2004). Máxime cuando los diferentes sistemas o aplicaciones pueden usar etiquetas idénticas pero con distinto significado; y,

alternativamente, el mismo significado puede ser expresado con diversas denominaciones.

Para conseguir la interoperabilidad entre sistemas de información se han propuesto, como una opción, las tecnologías semánticas y, entre ellas, las ontologías. Estas últimas se presentan como una forma de representar el conocimiento y están, comúnmente, asociadas a la idea de contenedores semánticos o de significado (Badía Cardús, 2003).

Desde la década de los noventa, las ontologías han sido un tema de interés para varias comunidades de investigadores como las de Inteligencia Artificial, Ingeniería del Conocimiento, Representación del Conocimiento, Diseño de Bases de Datos, Modelado de Información, Integración de Información, Gestión de Conocimiento y Organización, Diseños de Sistemas basados en Agentes o Procesamiento del Lenguaje Natural (PLN) (Guarino, 1998; Rodríguez, 2000). Más recientemente, este tema ha ido ganando peso en campos como la integración inteligente de información, la recuperación y la clasificación de información en la Web. Este interés es debido, fundamentalmente, a que las ontologías permiten una comprensión compartida y común de algún dominio, el cual puede ser comunicado entre individuos y aplicaciones (Studer, Benjamins y Fensel, 1998).

El objetivo de este capítulo es establecer una valoración de las definiciones de ontología aceptadas desde el ámbito de la Inteligencia Artificial, la Gestión de Conocimiento, la Ingeniería Ontológica, el Procesamiento del Lenguaje Natural (PLN) y la Web Semántica, diferenciando las distintas tipologías propuestas en estas áreas para los dominios, tanto abiertos como específicos, y, en concreto, para el dominio de la salud.

Este capítulo se organiza en cuatro apartados principales. En el primer apartado se delimita el significado descriptivo del término ‘ontología’, a partir del análisis de varias de las definiciones de mayor relevancia en el ámbito académico. En el segundo apartado se describen distintas

clasificaciones de las ontologías, de acuerdo con la declaración de los componentes, la formalización del contenido y el ámbito de la conceptualización, valorándose las funcionalidades de cada tipo para su integración en tareas de PLN. En el tercer apartado se realiza una descripción de los artefactos tecnológicos que, clásicamente, se han utilizado para la representación de conocimiento en el ámbito de la salud (terminologías, clasificaciones, taxonomías y nomenclaturas), desde una perspectiva ontológica y se presentan las funcionalidades desarrolladas en aplicaciones prácticas, a partir de estas fuentes de conocimiento. En el último de los apartados se incluyen las conclusiones obtenidas de la revisión del estado de la cuestión realizado.

2.1. Descripción del concepto ‘ontología’.

El término ‘ontología’ ha sido empleado, clásicamente, para referirse a la rama de la filosofía que trata del estudio del ser y de sus propiedades trascendentales (RAE, 2001). En el contexto de la representación de conocimiento, específicamente, este término empezó a emplearse para referirse a las teorías de modelado del mundo y a los componentes de un sistema de conocimiento, en la década de los ochenta. Desde esta perspectiva, una ontología se define como “una especificación explícita de una conceptualización [...], lo que ‘existe’ es lo que puede ser representado. Cuando el conocimiento de un dominio es representado en un formalismo declarativo, el conjunto de objetos que pueden ser representados es llamado el universo del discurso. Este conjunto de objetos y las relaciones descriptibles entre ellos, son reflejadas en el vocabulario representativo con el cual un programa basado en conocimiento representa el conocimiento” (Gruber, 1993).

Desde un punto de vista operacional, se podría definir como “un conjunto de términos estructurados jerárquicamente para describir un dominio, que puede ser utilizado como el esqueleto fundamental para una base de conocimiento” (Swartout et al, 1997). Aunque, a partir de estas dos definiciones, se puede observar el gradiente de complejidad y de ambigüedad que tiene, en la actualidad, el propio concepto de ‘ontología’. Estas diferencias en la descripción del concepto son la base para entender las semejanzas y las diferencias entre las ontologías existentes.

No obstante, la literatura muestra acuerdo al identificar que se habla de una ontología si el modelo de conocimiento representado, la conceptualización, es aceptado y compartido por una comunidad –entendida ésta tanto para diversas personas como para sistemas o agentes heterogéneos–, ya que dicha característica es esencial para que pueda ser reutilizada la representación y mantenida por diferentes sistemas de información (Uschold y Grüninger, 1996; Borst, 1997; Mizoguchi et al, 1997; Studer et al, 1998). Este hecho no implica un único y exclusivo punto de vista para representar lo que ‘existe’. Aunque, sí que establece un marco de unificación para resolver los problemas de interoperabilidad semántica (Uschold y Grüninger, 1996), al realizar una explicitación de la naturaleza y del contexto lógico de la información con una formalización que la hace legible, tanto por los humanos como por los ordenadores (Borst, 1997).

En una ontología, “las definiciones asocian los nombres de las entidades en el universo del discurso (clases, relaciones, funciones u otros objetos) con texto legible desde un punto de vista humano, describiendo lo que significan los nombres, y axiomas formales que restringen la interpretación y el uso adecuado de estos términos” (Gruber, 1993). En otras palabras, una ontología facilita la explicitación específica de los conceptos (clases), sus propiedades y sus restricciones, así como la organización parcial de los conceptos a partir de sus relaciones. Por ello, no puede interpretarse que los conceptos se organizan, exclusivamente, de forma jerárquica (por medio de las relaciones taxonómicas ‘*is-a*’ y ‘*part-of*’). En

función del dominio pueden existir otro tipo de relaciones semánticas que pueden determinar organizaciones de tipo grafo o redes (por ejemplo, relaciones del tipo '*afecta-a*', '*trata-a*', '*ingrediente-de*', '*reacciona-con*').

Una ontología es una tecnología conceptual que permite, por tanto, definir el vocabulario para quienes necesitan compartir la información dentro de un determinado dominio (Noy y McGuinness, 2001). Es decir, la ontología, necesariamente, incluirá un vocabulario de términos y una especificación de su significados para restringir las posibles interpretaciones (Uschold y Jasper, 1999). Aunque, no es el vocabulario, en sí, el que se califica como ontología, sino las conceptualizaciones que capturan los términos del vocabulario (Chandrasekaran et al, 1999). Por otra parte, los conceptos presentes en una ontología establecen un nexo de unión entre las distintas lenguas (idiomas), aunque de forma independiente a cada una de ellas (Badía Cardús, 2003). Todo ello explica el interés de las ontologías en el campo del PLN.

Como síntesis, indicar que una ontología es una definición operativa de un dominio en la que se realiza una declaración del conocimiento, de forma global o específica. Sus componentes son los conceptos (clases) y las relaciones empleadas en la declaración a partir del propio dominio. Estos componentes se caracterizan por medio de sus propiedades (atributos) y de las restricciones que se especifican, para concretar el significado y reducir la ambigüedad de los términos empleados. En este sentido, una ontología puede verse como un repositorio de información conceptualizada y relacionada entre sí, de forma consistente y no redundante que, bien organizada, es clave para un sistema de información. Un aspecto distintivo, que distingue el diseño de una ontología con respecto al de un sistema de información o una base de datos, es que la definición de sus componentes y la caracterización de éstos, no parte de una valoración personal o local, sino que parte del propio dominio a representar. El dominio es el que determina el diseño para que éste pueda ser compartido y reutilizado.

2.2. Tipología de Ontologías.

Cuando se valoran las ontologías existentes, o bien las características que deben considerarse para su diseño, se hace patente la gran disparidad de criterios entre diversos autores. Esta diversidad se debe a la utilización de distintas perspectivas de análisis para determinar la utilización y funcionalidad de las ontologías.

En la práctica, sin embargo, el límite entre los diversos tipos de ontologías es borroso, no está claramente determinado, y las situaciones mixtas o no excluyentes son las más frecuentes.

En los siguientes subapartados se recogen los diferentes tipos de clasificaciones propuestas para las ontologías. La presentación y la valoración de estas clasificaciones se ha organizado según: i) *la declaración de los componentes representados en el modelo del dominio*; ii) *la formalización del contenido* y iii) *el ámbito de la conceptualización del modelo*. Un esquema de la clasificación propuesta, en función de los criterios mencionados, se incluye en la tabla 2.1.

Tipos de Ontologías		Tipo de explicitación o descripción del significado.	Ligeras Pesadas
		Tipo de especificación o definición operacional del significado.	Lista de Términos Controlados Glosarios Tesauros Jerarquizaciones Informales Jerarquizaciones Formales Ontologías en Frames Ontologías incluyendo Restricciones
	Según la Formalización del Contenido.	Tipo de formalismo	Informales Semi-Informales Semi-Formales Formales
	Según el Ámbito de la Conceptualización del Modelo.	Tipo de cobertura conceptual.	Ontologías de Representación Ontologías Generales (Comunes o Genéricas) Ontologías de Alto Nivel Ontologías de Dominio Ontologías de Tarea Ontologías de Aplicación
Nota Aclaratoria: La tipología propuesta se basa en la definición de Gruber (1993) y en las aportaciones realizadas por Borst (1997) y por Studer y equipo (1998): "Una ontología es una especificación, explícita y formal, de una conceptualización compartida"			

Tabla 2.1: Categorización de las ontologías.

2.2.1. La declaración de los componentes representados.

Los constituyentes de una ontología pueden variar en función de las necesidades o el problema a resolver, de los diferentes tipos de aplicaciones en que van a ser utilizadas las ontologías y del lenguaje en el que se desarrolle (como se verá posteriormente). Los componentes comunes en una ontología deberían ser: conceptos, propiedades, relaciones, axiomas e instancias (Gruber, 1993; Farquhar, Fikes, y Rice, 1997; Sowa, 1999; Rodríguez, 2000; Noy y McGuinness, 2001; Gómez-Pérez et al, 2004).

- Los *Conceptos* (físicos o abstractos), en una ontología, son representados por un conjunto de clases relevantes para el dominio. Las clases se describen especificando las propiedades que deben satisfacer para pertenecer a ella. Las clases son, por tanto, la base de la descripción del conocimiento del dominio. Las clases se denominan por un término preferente, pero el término, en sí mismo, no es el concepto. En las ontologías lingüísticas se trata de conceptos relacionados con categorías lingüísticas mientras que, en las léxicas son términos simples o complejos, indistintamente.

Un ejemplo de clase podría ser '*Medicamento*'. Esta clase, idealmente, representa a todos los medicamentos existentes en el mundo y, de este modo, un medicamento concreto, '*Cardyl*', es un ejemplar o instancia de la clase '*Medicamento*'.

Una clase puede dividirse en subclases que representan conceptos más específicos y, a su vez, las subclases se agrupan en superclases más generales. De este modo la clase '*Medicamento*' podría dividirse en subclase de medicamentos del tipo: '*antilipemiente*', '*antihipertensivo*' y '*antidiabético*'. Al mismo tiempo, la clase '*Medicamento*' pasaría a ser considerada una superclase.

- Las *Propiedades* son el conjunto de las especificaciones, características o atributos, de una clase. Los valores de las especificaciones, de los rangos y de las restricciones (también se les denomina facetas) son los que describen a las clases en una ontología. De este modo podemos encontrar desde una simple etiqueta léxica hasta un conjunto, amplio y detallado, de atributos con los valores asociados a cada clase.

Ejemplos de propiedades podrían ser '*fórmula*', '*cantidad*', '*código nacional*' y '*nombre genérico*'.

En el caso de las superclases las especificaciones, los valores y las restricciones son heredadas por las subclases que, estructuralmente, dependen de ella.

- Las *Relaciones* representan la interacción existente entre los conceptos y conforman la estructura del dominio. Pueden ser tanto binarias como n-arias, aunque, en cada ontología, se suele permitir un número limitado y preciso de relaciones de acuerdo con el propio dominio representado. Cada relación puede estar dotada de propiedades que fijan su comportamiento. Las más frecuentes son las relaciones taxonómicas ('*es un tipo de*' o '*isa*') y las de meronimia ('*es parte de*' o '*part-of*', '*es ingrediente de*'...). Ambas establecen una organización jerárquica entre los conceptos que interaccionan a través de estas relaciones, del tipo super-subclase. Otro tipo de relaciones presentes, frecuentemente, son las relaciones semánticas de tipo: asociación ('*está asociado a*', '*depende de*', '*es responsable de*'...) , temporal ('*concurren con*', '*preceden a*'...) y topológica o espacial ('*a la derecha de*', '*está encima de*', '*está conectado a*'...).
- Los *Axiomas* contienen las reglas y las restricciones de los valores de las clases o de las propiedades de la relación. Los axiomas son las verdades

que se afirman en el dominio y suelen expresarse en lógica de primer orden. Pueden ser de dos tipos: estructurales y no estructurales. Los axiomas estructurales establecen las condiciones de organización de los conceptos por los atributos (*el concepto 'X' es una subclase de 'A'*). Los axiomas no estructurales determinan relaciones entre los atributos de un concepto.

Un ejemplo de axioma no estructural sería en la 'administración de una infusión endovenosa' la relación $got_min = (cant_total/t_total)*f_got$. Esta relación se debería de cumplir siempre entre los atributos *got_min* (*gotas minuto*), *cant_total* (*volumen solución*), *t_total* (*intervalo tiempo minutos*) y *f_got* (*factor gotas por ml*) para el concepto *RitmodeFlujo*.

- Las *Instancias* representan los elementos específicos de una clase (también, se les denomina individuos u objetos). Aunque, en un principio, no deberían de estar presentes en una ontología, ya que la combinación de la ontología con las instancias conforman la base de conocimiento. Sin embargo, dependiendo del punto de vista con el que se desarrolle la ontología, algunas instancias pueden ser consideradas clases si el nivel de granularidad de la representación es fino.

A partir de la presencia o ausencia de algunos de estos componentes se pueden distinguir entre ontologías pesadas (*heavyweight ontologies*) o ligeras (*lightweight ontologies*) (Gómez-Pérez et al, 2004). Se consideran *ontologías pesadas* a aquellas que contienen todos los componentes anteriormente descritos. Este tipo de ontologías tiene una especial significación para las tareas de PLN, debido a que incluyen la definición de los axiomas y las restricciones de las propiedades o de los valores, lo que reduce la ambigüedad y concreta el significado de los términos o unidades léxicas incluidas en la ontología. Mientras que las *ontologías ligeras* carecen de axiomas e incluso de restricciones. Este tipo de ontologías se

centran en la taxonomía de los conceptos incluidos y las relaciones entre éstos (organización jerárquica), así como las propiedades que describen a dichos conceptos.

La diferencia entre las ontologías pesadas y ligeras es el tipo de razonamiento de inferencia que se podrá desarrollar para el descubrimiento de nueva información (generalización o especificación), mucho más amplio si se dispone de axiomas y restricciones muy expresivas.

Por otra parte, si se consideran los componentes de una ontología ligera, se abre la puerta a la inclusión en esta categoría de otros artefactos, como son las terminologías y los glosarios, tanto de dominios generales como específicos. Estos artefactos pueden reutilizarse a pesar de las evidentes limitaciones para concretar el significado. Una de las limitaciones principales de estos artefactos es la ambigüedad del significado de los términos incluidos. Por ejemplo, términos como *‘enfermería’* ¿denominación de una profesión o lugar físico para la atención? o *‘proteinuria’* ¿síndrome o resultado de una prueba analítica de laboratorio?.

Una clasificación más abierta es la propuesta de Lassila y McGuinness (2001). Esta clasificación se fundamenta en los componentes y en la riqueza de la estructura interna de una ontología. Su tipología distingue un amplio espectro de artefactos:

- *Listas de términos fijados o controlados*, como los existentes en las listas alfabéticas, en los catálogos de biblioteca o de productos comercializados, a los que se les incluyen identificadores únicos para reducir la ambigüedad.
- *Glosarios*, donde los términos son especificados, por medio de definiciones o descripciones, en lenguaje natural para delimitar su significado, pero que no pueden ser interpretados automáticamente.
- *Tesauros*, que proporcionan información a partir de las relaciones de equivalencia (sinonimia) entre los términos y de jerarquía. Esta última

relación se puede deducir por la inclusión de términos genéricos (hiperónimos) y específicos (hipónimos). Aunque, no se distingue entre la relación ‘isa’ o ‘part-of’ por falta de explicitación.

- *Jerarquizaciones Informales de Términos*, que proporcionan una noción de generalización y especificación por la ordenación de los términos en la jerarquía, como ocurre, por ejemplo, en las categorías de los directorios de la Web. Aunque, se suelen caracterizar por una jerarquía no estricta respecto a la relación ‘isa’; es frecuente la existencia de subclases que no siempre lo son de las clases de niveles superiores.
- *Jerarquías Formales o Taxonomías* se constituyen por estrictas subclases a partir de la relación ‘isa’. Estas ontologías permiten explotar todos los requisitos de herencia entre las superclases y las subclases. Un subtipo de mayor complejidad son las taxonomías que incluyen, formalmente, instancias del dominio.
- *Ontologías diseñadas con Frames*, declaran, formalmente, las propiedades de las clases (atributos y valores), las cuales son heredadas por las subclases e instancias.
- *Ontologías que expresan Restricciones*, en las que se determinan los tipos o el rango de los valores que toma cada propiedad. Un nivel de expresividad mayor son la ontologías, que incluyen, en los valores, una fórmula de lógica de primer orden o matemática a partir de los valores declarados en otra propiedad.

Estas variaciones, en la riqueza de su estructura interna, condicionan la selección y la incorporación de las ontologías existentes (reutilización) en las aplicaciones que precisan asegurar la interoperabilidad semántica entre distintos SI o que desarrollan tareas de PLN (ver figura 2.1.). Considerando tan sólo las tareas de PLN, los cuatro primeros tipos de ontologías se podrían utilizar para la extracción de información y reconocimiento de entidades en dominios específicos. Pero, para la desambiguación de

significados dependientes de contextos, se debería optar por las ontologías del último tipo, que expresan las restricciones, y garantizan la completitud, la consistencia y la concisión del conocimiento representado. Respecto a las ontologías situadas en las posiciones intermedias de esta clasificación, taxonomías y frames, se podrían reutilizar para distintas finalidades: i) para enriquecer/extender las consultas de los usuarios en los sistemas de recuperación, ya que facilitan la reducción tanto de los ruidos como de los silencios de las respuestas, ii) para la extracción de terminologías vinculadas con la clasificación automática de documentos y, iii) para la búsqueda de respuestas.

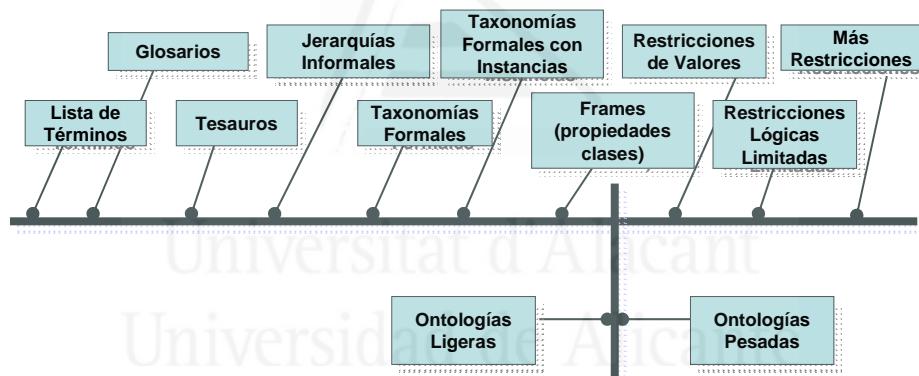


Figura 2.1: Clasificación de las ontologías en función de los componentes declarados y la riqueza de su estructura interna. Adaptación de la aportación de Lassila y McGuinness (2001) y Gómez-Pérez y equipo (2004).

2.2.2. La formalización del contenido.

Otra forma de clasificar las ontologías es considerando el tipo de formalización del contenido de la ontología. En la actualidad, existen diversos lenguajes de implementación capaces de contener el conocimiento representado. Dichos lenguajes determinan la expresividad del modelo o el desarrollo de una ontología. Desde esta perspectiva Uschold y Grüninger (1996) propusieron clasificar las ontologías en cuatro categorías:

- Las *ontologías rigurosamente formales*, en las que se definen los conceptos con semántica formal y que incluyen restricciones lógicas de primer orden entre los términos y relaciones más detalladas entre las clases (disyuntivas, inversas, simétricas, parte-todo...). En ellas, se pueden demostrar las propiedades de completitud, consistencia y solidez.
- Las *ontologías semi-formales*, que son expresadas en un lenguaje artificial con definición formal. Una de sus características es que puede expresar una herencia estricta, que asegura que una instancia de una subclase lo es de una superclase, también.
- Las *ontologías semi-informales*, que son expresadas de forma restringida y estructuradas en lenguaje natural. En ellas se reduce, sólo parcialmente, la ambigüedad. Aunque existe un esquema jerárquico, no cumplen siempre con los requisitos de herencia de las subclases o de disyunción entre subclases hermanas de una superclase.
- Las *ontologías altamente informales*, son las que se expresan en lenguaje natural. Entre ellas se encuentran los listados de términos o los glosarios, que incluyen los sinónimos o cuasi-sinónimos asociados al término representativo del concepto.

En este sentido, existe una convergencia entre el formalismo de expresividad del lenguaje y los componentes declarados (ver figura 2.1.) pero, a su vez, se manifiesta la discrepancia de que no todos estos

formalismos pueden ser legibles por los ordenadores. Para que, realmente, se pudiera interpretar y reutilizar el conocimiento representado, por aplicaciones informáticas que no participaron en su creación, de forma automática, se precisaría de ontologías formalizadas en lenguajes semiformales, como mínimo, aunque mucho mejor, serían las ontologías rigurosamente formalizadas.

2.2.3. El ámbito de la conceptualización.

Otro aspecto a considerar, para clasificar las ontologías, es el referido al tipo de fenómeno o de realidad representada, es decir, al ámbito de la conceptualización reflejada en cada modelo. Atendiendo a este criterio, nuevamente, se constata la disparidad en la denominación de las categorías para clasificar las ontologías, así como en las descripciones de éstas (Mizoguchi et al, 1997; Van Heijst, Schreiber y Wielinga, 1997; Guarino, 1998). Una recopilación y una ordenación de las propuestas existentes es la clasificación que facilitan Gómez-Pérez y su equipo (2004). Su propuesta establece un gradiente de generalización-especificación, respecto al fenómeno representado, identificando las denominaciones equivalentes a través de la concordancia de las descripciones analizadas. Dicha clasificación es la que se incluye:

- *Ontologías de Representación*: Proporcionan el vocabulario necesario para modelizar otras ontologías utilizando un determinado paradigma de representación del conocimiento (Van Heijst et al, 1997).
- *Ontologías Generales/Comunes o Ontologías Genéricas*: Contienen las descripciones generales sobre objetos, eventos, relaciones temporales, relaciones causales, modelos de comportamiento y funcionalidades. Proporcionan términos genéricos reutilizables en dominios diferentes (Mizoguchi et al, 1997; Van Heijst et al, 1997).

- *Ontologías de Alto Nivel*: También se les denomina ontologías de nivel superior. En ellas se describen los conceptos y las nociones generales bajo las cuales todos los términos raíz de las ontologías existentes deberían ser vinculados (Guarino, 1998).
- *Ontologías de Dominio*: Proporcionan los vocabularios sobre los conceptos asociados a un dominio y sus relaciones. También, sobre las actividades, las teorías o principios elementales que gobiernan un dominio particular. Usualmente, los conceptos y las relaciones de estas ontologías son definidos como especializaciones de conceptos existentes en ontologías de alto nivel o en las genéricas (Mizoguchi et al, 1997; Van Heijst et al, 1997; Guarino, 1998).
- *Ontologías de Tarea*: Establecen la forma en la que se puede usar el conocimiento proporcionado por las ontologías de alto nivel o por las ontologías de dominio. En ellas se describe el vocabulario especializado de los conceptos que se relacionan con tareas genéricas o actividades, por ejemplo para la búsqueda de respuestas –*Question answering*– (Mizoguchi et al, 1997; Guarino, 1998).
- *Ontologías de Aplicación*: Contienen las definiciones necesarias para modelizar el conocimiento requerido por una aplicación. En ellas se suelen reutilizar los conceptos de ontologías de dominio o de tareas. Se caracterizan por ser usables pero no reutilizables ya que, a diferencia de las ontologías de tarea, estas ontologías son dependientes de la aplicación que la utiliza.

Esta clasificación establece, teóricamente, una delimitación muy clara pero, en la práctica, puede ser más confusa. Las situaciones mixtas o no excluyentes son las más frecuentes en la realidad. De hecho, lo más habitual es valorar esta categoría de forma global y centrarse en la cobertura conceptual de una ontología a partir del análisis del vocabulario incluido.

2.3. Ontologías en Salud.

Las tecnologías empleadas en el ámbito sanitario, para la representación del conocimiento y el intercambio de información, han sido, clásicamente, las terminologías profesionales.

Aunque, se han intentado establecer distintas estandarizaciones en este dominio, no se ha logrado una unificación práctica que dé una respuesta completa a la diversidad de necesidades que tiene tanto el sistema sanitario como los profesionales y los usuarios de las prestaciones (los ciudadanos).

Estas tecnologías se han desarrollado y han evolucionado, conjuntamente, con las nuevas prestaciones de los sistemas de información (SI)². La gama de artefactos que se emplean, en la actualidad, es muy abundante y diversa. De hecho, la denominación de ‘terminología’ incluye desde los vocabularios controlados, los sistemas de codificación, las clasificaciones, los tesauros o las nomenclaturas hasta, más recientemente, las ontologías (Stead et al, 2000; Ciccicarese et al, 2005; Robinson, 2008).

En este sentido, las terminologías del contexto sanitario responden a la categorización realizada por Lassila y McGuinness (2001), expuesta anteriormente (pp. 21-22). Aunque, estos artefactos, además de diferenciarse por el tipo de componentes declarados y la riqueza de su estructura interna, se diferencian, fundamentalmente, por la cobertura conceptual, aún estando en un dominio restringido.

Las terminologías que, en los siguientes apartados, se analizan desde una perspectiva ontológica, son ejemplos de las variantes citadas y su empleo, en los actuales SI es altamente relevante. Cada una de ellas ha sido desarrollada para cubrir las demandas en diversos escenarios sanitarios,

² Más detalles sobre las funcionalidades de la primera, segunda y tercera generación de las terminologías, respecto a los sistemas de información, se encuentran en el trabajo de Stead y equipo, 2000.

tanto relativos al contexto clínico, epidemiológico y documental como de carácter administrativo-económico. Estas terminologías son:

- la Clasificación Internacional de Enfermedades, novena revisión, Modificación Clínica (CIE-9-MC),
- los *Medical Subject Headings* (MeSH),
- los Descriptores en Ciencias de la Salud (DeCs),
- la *Systematized Nomenclature of Medicine, Clinical Terms* (SNOMED CT),
- la *General Architecture for Languages Encyclopaedias and Nomenclatures in Medicine* (GALEN), y
- el *Unified Medical Language System* (UMLS).

En los siguientes subapartados se presenta la cobertura conceptual, la estructura organizativa y las relaciones conceptuales, la granularidad conceptual y, el grado de formalización de cada una de las terminologías indicadas. El objetivo es mostrar el grado de usabilidad de estos artefactos para las tareas de interoperabilidad semántica o de PLN. El análisis que se muestra forma parte del trabajo de Romá-Ferri y Palomar (2008).

2.3.1. Cobertura conceptual.

En este punto se analiza la finalidad, la utilidad y amplitud del conjunto de conceptos representados en cada terminología.

Terminología: Conceptualiza tanto los diagnósticos médicos, en función de la etiología y de las manifestaciones, como los procedimientos diagnósticos y terapéuticos. Contiene

CIE-9-MC

Componentes estudiados:	más de 11.000 rúbricas para especificar las entidades clínicas en castellano.
Lista Tabular e Índice Alfabético.	Su funcionalidad es la estandarización de los diagnósticos médicos para su codificación epidemiológica y administrativo-económica.
Terminología: MeSH	Conceptualiza el vocabulario incluido en las publicaciones científicas del ámbito de las Ciencias de la Salud. Incluye 24.767 encabezamientos de materia en lengua inglesa.
Componentes estudiados: Listado Alfabético Anotado y Lista Jerárquica.	Su funcionalidad es la de normalizar el vocabulario para la catalogación e indización ³ , así como para la posterior recuperación de los artículos indexados en la base de datos Medline ⁴ .
Terminología: DeCS	Su cobertura y finalidad es delimitar el conjunto finito de conceptos utilizados en el proceso de catalogación e indización para la base de datos LILACS ⁵ y para la recuperación de las publicaciones científicas indexadas en la base de datos Medline y LILACS.
Componentes estudiados: Listado Anotado y Lista Jerárquica.	Se compone de 29.590 descriptores en castellano, e igual número en portugués. El número de descriptores en inglés es igual que en el MeSH.

³ Desde el punto de vista de la documentación, “la indización es la actividad que conlleva examinar un documento y su contenido, identificar y elegir los principales conceptos y traducir los conceptos elegidos a un lenguaje elemental –en nuestro caso el lenguaje controlado o normalizado de referencia que aporta la nomenclatura, la taxonomía o la clasificación– y que, posteriormente, servirá para la recuperación del documento. El tipo de indización efectuado incide, directamente, en el tipo de consulta que el usuario podrá realizar posteriormente”. [Lancaster, FW. (2002). El control del vocabulario en la recuperación de información. (2ª ed.) (pp. 17-19). Valencia: Universitat de València].

⁴ *Medical Literature Analysis and Retrieval System Online*

⁵ Literatura Latinoamericana y del Caribe en Ciencias de la Salud

Terminología: Conceptualiza el vocabulario clínico asistencial en general y también incluye a las especialidades médicas y las terminologías de enfermería.

SNOMED CT

En la versión en castellano contiene 311.000 conceptos identificados por 724.000 términos (preferentes y sinónimos) y 920.000 relaciones especificadas (versión de Enero 2008).

Se orienta a diversas tareas que permiten el desarrollo de aplicaciones de HCE (historia clínica electrónica), a la captura de datos y a la recuperación de información, en el ámbito hospitalario y en atención primaria. Así como al análisis de resultados clínicos y el apoyo a la toma de decisiones.

Terminología: Contiene conceptos médicos sensibles, elementales y de alta especificidad, a partir de los cuales puede definirse cualquier otro tipo de concepto o de término.

GALEN

Componentes estudiados:

*Common
Referente
Model (CRM)
y GALEN
Representation
and
Integration
Language
(GRAIL).*

Su finalidad es la representación formal de terminologías médicas para los distintos usos clínicos y de codificación. Se desarrolló para facilitar la construcción de sistemas clínicos que pudieran representar y manipular la información de los pacientes, con un nivel de detalle suficiente, y que apoyasen la rutina del cuidado clínico cotidiano, tanto de médicos como de enfermería, al separar el modelo conceptual del modelo de uso o del idioma.

Terminología: Contiene la representación conceptual más amplia del dominio, debido a la combinación y alcance de los

UMLS

Componentes estudiados:

Metathesaurus

diversos vocabularios fuente que lo conforman.

y *Semantic Network* (Red Semántica).

En total contiene unos 140 vocabularios. Aunque, muchos de ellos son versiones, revisiones o traducciones de un mismo vocabulario. Esto facilita una amplia cobertura léxica, aunque desigual, en los 17 idiomas que integra. En la versión de 2008AA el Metathesaurus se componía de 1.553.638 conceptos y de 7.781.500 nombres no normalizados para ellos, e integrados en los 135 tipos de conceptos superiores que constituyen su Red Semántica. En la actualidad, incorpora 6 vocabularios fuente en castellano (el 4,14% del total de su contenido).

Su funcionalidad se orienta a tareas genéricas como la recuperación y la integración de información biomédica electrónica, desde o para sistemas de información dispares.

2.3.2. Estructura organizativa y relaciones conceptuales.

Los conceptos pueden ser organizados en diferentes estructuras denominadas monoaxiales, multiaxiales, composicionales o léxico-semánticas. Se entiende que una estructura es de tipo:

- *Monoaxial* cuando su estructura jerárquica está compuesta por una lista finita de categorías y subcategorías, elaboradas siguiendo un orden lógico, determinado por un criterio (cualidad o atributo) que aproxima los conceptos más específicos (hijos) y los agrupa dentro de los más generales (padres).
- *Multiaxial* cuando el dominio se estructura a través de varias jerarquías. Cada jerarquía o eje se desarrolla a partir de un atributo concreto que le da la naturaleza al eje. Este tipo de estructura organizativa permite que

un mismo concepto hijo este asociado a varios padres con la finalidad de distinguir su significado polisémico.

- *Composicional* cuando la estructura organizativa permite la creación de conceptos complejos a partir de la combinación de conceptos elementales y de modificadores. Un ejemplo sería el concepto elemental clínico '*dolor*' y su combinación con alguno o varios de los siguientes modificadores '*agudo*', '*izquierda*', '*proximal*'. En esta estructura no se precisa explicitar, por adelantado, todo el conjunto de conceptos de una especialidad o área. Aunque, para prevenir la creación de conceptos complejos sin significación, se incluyen reglas y restricciones para la combinación de los conceptos elementales y de los modificadores.
- *Léxico-semántica* cuando la estructura organiza el léxico a partir de su significado en tipos semánticos. En ella los conceptos incluidos y sus relaciones semánticas determinan el modelo de la interpretación del dominio.

Respecto a las relaciones conceptuales se deben distinguir entre los enlaces existentes y los que se declaran explícitamente. Los enlaces pueden ser de tipo jerárquico (relaciones de tipo *isa* y *part_of*) o incluir una gran variedad de otros tipos de enlaces semánticos (*es causado*, *se localiza*, *es tratado*...) que determinan o restringen las cualidades heredadas por los conceptos específicos (subconceptos) de los generales (superiores).

En este apartado se realiza la descripción de la organización conceptual, empleada para la representación del conocimiento del dominio, y del tipo de relaciones existentes entre los conceptos de cada terminología estudiada.

Terminología: Su organización es monoaxial. Está constituida por 17
CIE-9-MC subcategorías mayores (clases o conceptos principales)

Componentes estudiados: para los diagnósticos clínicos. Cada categoría es identificada por un código numérico. Los códigos numéricos de las categorías son consecutivos (desde 001 hasta 999.9). Aunque, algunas clases son representadas por cinco dígitos para facilitar una mayor especificidad. Los conceptos relativos a procedimientos se representan en un eje, independiente del anterior, constituido por 16 subcategorías principales definidas en función de su localización anatómica. Los códigos asignados, también, son consecutivos (del 00 hasta 99.99).

Las relaciones se derivan de la propia dependencia jerárquica en la que se estructura la clasificación. Cualquier categoría de nivel inferior responde a una relación de tipo '*isa*'. Aunque, en unos casos, cumple con ser 'un tipo de', mientras que, en otros, responde a 'una parte de' la clase superior, generalmente, cuando los conceptos tienen un contenido relacionado con una localización anatómica.

Terminología: Tiene una estructura multiaxial. Se organiza en 15 jerarquías que desarrollan clasificaciones independientes e identificadas por una letra distintiva. Cada eje equivale a una categoría principal. Las subcategorías de cada eje están asociadas a un código alfanumérico (de lo más genérico a lo más específico). Hay subcategorías que tienen más de un código. Cada código indica o la posición del término en varios de los ejes o las diferentes ubicación dentro del mismo eje.

Hay tres tipos de relaciones semánticas especificadas: *equivalencia*, *asociación* y *jerarquía*. Las relaciones de '*equivalencia*' son las que señalan los sinónimos o cuasi-

sinónimos por medio del verbo ‘*see*’ (referencia cruzada que encamina hacia el término aceptado para la representación del concepto). La relación de ‘*asociación*’ indica la existencia de vinculación temática, tanto en el mismo eje como en otro eje diferente, aunque sin ninguna relación de equivalencia, y se señala con ‘*see related*’. La relación de ‘*jerarquía*’ indica la existencia de una relación de orden, del tipo hiperonimia/hiponimia, respecto a un descriptor secundario o subordinado y se identifica con la expresión ‘*see under*’. La declaración realizada de la relación ‘*jerárquica*’ no distingue entre la relación de tipo ‘*isa*’ o ‘*part_of*’, tan sólo la especifica para la subordinación entre subcategorías y supercategorías.

Terminología:	Tiene una organización multiaxial conformada por 19
DeCS	ejes. De ellos, 15 son idénticos a los de MeSH
Componentes estudiados:	(mantienen la misma estructura y codificación). Los 4
Listado	ejes restantes son propios pero mantienen el sistema de
Anotado y	organización y codificación empleado por el MeSH.
Lista	
Jerárquica.	Las relaciones conceptuales son idénticas a las descritas en el MeSH.

Terminología	Tiene una estructura composicional. Los conceptos
SNOMED CT	representados se organizan en 18 jerarquías.

En 3 de estos ejes se organizan los tipos de relación entre los conceptos, pertenecientes a todos o alguno de los otros ejes, para facilitar o limitar la combinación entre ellos. En el eje ‘*Attribute*’ se localizan todos los conceptos que pueden ser usados para establecer un tipo de relación semántica (‘*está asociado*’). El eje ‘*Qualifier*

value’ contiene los posibles valores de una cualidad. Un valor puede ser usado en una definición de relación o como un calificador de un concepto (por ejemplo ‘agudo’, ‘remisión’).

Una particularidad es el eje denominado ‘*Special concept*’, que contiene los conceptos que comparten una característica común pero, que no forma parte de la actual jerarquía lógica de categorías semánticas. En este eje hay dos subcategorías principales ‘*concepto no actual*’ (ningún concepto o término descartado es nunca eliminado) y ‘*concepto de navegación*’.

Todos los ejes tienen una organización jerárquica básica establecida por la relación de tipo ‘*isa*’. Es decir, todos los conceptos representados se entienden como un subtipo del ‘concepto raíz’ o de un ‘Concepto de nivel superior’.

Terminología: GALEN Componentes estudiados: <i>Common Referente Model (CRM) y GALEN Representation and Integration Language (GRAIL).</i>	<p>La estructura jerárquica primaria de CRM es una red semántica composicional y generativa de lenguaje de representación de conocimiento. Se estructura a partir de las 3 jerarquías que determinan sus conceptos generales primarios y una cuarta jerarquía, en la que se explicitan todos los tipos de relaciones factibles.</p> <p>En cada jerarquía, las subcategorías sólo tienen una categoría superior. Aunque, existen relaciones entre categorías o subcategorías pertenecientes a otras jerarquías para facilitar la asociación entre los conceptos y la restricción en la composición de conceptos complejos.</p>
---	---

Cuenta con una rica jerarquía de relaciones semánticas,

especificadas por 592 tipos (internamente se denomina ‘atributos’), que se utilizan para expresar conceptos complejos. Destaca la diferenciación de las relaciones de tipo partitivo que permiten distinguir entre divisiones o partes de superficies, sólidos, capas y componentes estructurales, específicos o realizados específicamente.

Los conceptos de las subcategorías presentes, tanto en la estructura primaria como en la secundaria de CRM, mantienen una relación con las categorías superiores por medio del enlace jerárquico ‘isa’.

Los enlaces jerárquicos, como el resto de enlaces semánticos, han sido verificados para cumplir con la propiedad transitiva. Por lo tanto, siempre se puede admitir que las propiedades de los conceptos superiores son herederas por los conceptos subordinados inferiores, aunque sus propiedades no estén declaradas, explícitamente, por economía computacional.

Terminología:	En el Metathesaurus se mantiene la estructura jerárquica de cada uno de los vocabularios fuente incluidos. Todos
UMLS	los conceptos representados en el Metathesaurus son asignados a la categoría semántica más específica de la Red Semántica.
Componentes estudiados:	La Red Semántica de UMLS se estructura a partir 135 tipos de categorías semánticas. Su organización está determinada por el tipo de relaciones existente entre las categorías. La relación primaria, en la Red, es de tipo jerárquico ‘isa’. El resto de las relaciones semánticas declaradas se organizan en una taxonomía de 54 tipos. Las categorías principales de estas relaciones son:
Metathesaurus y <i>Semantic Network</i> (Red Semántica).	

'physically_related_to', *'spatially_related_to'*,
'temporally_related_to', *'functionally_related_to'* y
'conceptually_related_to'

Las relaciones semánticas son declaradas entre los tipos de categorías superiores de la Red, cuando es posible, y se heredan por todos sus tipos semánticos hijo por medio del enlace *'isa'*. Aunque, las relaciones semánticas declaradas pueden mantenerse o rescindirse entre cualquier par de conceptos, tanto de la Red como entre la Red y el Metathesaurus.

2.3.3. Granularidad conceptual.

En este punto se describe el grado de abstracción conceptual o la especificidad de las clases representadas en las terminologías (Romá-Ferri y Palomar, 2008).

Terminología: Una clase o concepto es explicitado por una rúbrica CIE-9-MC terminológica y, cada una de ellas, se identifica por un código numérico exclusivo. Las estructuras sintácticas que predominan en las rúbricas son los sintagmas nominales y los preposicionales. También, destacan las partículas de coordinación y negación (exclusión conceptual).

Componentes estudiados:

Lista Tabular e Índice Alfabético.

Las rúbricas que representan conceptos simples son casi anecdóticas. Hasta los enunciados realizados por un término simple, representan conceptos complejos (por ejemplo, el término *'apendicitis'* representa un concepto

de localización anatómica asociado con un concepto vinculado a proceso inflamatorio).

Terminología: MeSH Componentes estudiados: Listado Alfabético Anotado y Lista Jerárquica.	Los conceptos representados son tanto simples como complejos. Las etiquetas terminológicas, que los representan, tienen estructuras sintácticas constituidas por formas simples o unitérminos (no excluyen la representación de conceptos complejos) y etiquetas compuestas por más de un término (descriptores precoordinados). En este último caso, para lograr una mayor especificidad, las etiquetas pueden ser de tipo nominal como preposicional.
--	---

Terminología: DeCS Componentes estudiados: Listado Anotado y Lista Jerárquica.	Sus características de granularidad son coincidentes con el MeSH. Aunque, destaca por la inclusión de etiquetas terminológicas para representar los conceptos en inglés, español y portugués.
--	---

Terminología: SNOMED CT	Los conceptos representados tienen un único significado, que es referenciado por un término preferente. Otros posibles términos utilizados para su denominación, en la práctica asistencial o incluidos en otras terminologías, están agrupados en <i>clusters</i> (sinónimos).
-----------------------------------	---

Los términos preferentes pueden ser unitérminos o sintagmas nominales. Por lo general, los sintagmas incluidos son una precoordinación de unitérminos para reflejar una mayor especificidad de un concepto elemental.

Los términos de la versión en inglés se han traducido al

español y al alemán.

Terminología:	En el CRM de GALEN se representan los conceptos a través de expresiones formales de GRAIL. Estas expresiones están constituidas por dos conceptos primitivos o elementales y su relación semántica, con las que se definen nuevas categorías o conceptos complejos.
Componentes estudiados:	Las etiquetas para referenciar los conceptos son únicas y con una ambigüedad muy reducida. Un ejemplo de su semántica y sintaxis es: <i>(Process which hasSpecificOutcome Ulcer) name UlcerationProcess</i> .
<i>Common Referente Model (CRM) y GALEN Representation and Integration Language (GRAIL).</i>	La estructura subyacente interlingua del CRM le permite dar un servicio multilingüe –inglés, francés, italiano, holandés, alemán, finlandés y sueco–, sin precisar traducción.

Terminología:	El Metathesaurus es un repositorio de conceptos. En él se preservan los significados, los atributos, las conexiones jerárquicas, y otras relaciones entre los términos presentes en cada uno de los vocabularios fuente que se han incorporado (incluso si tienen una definición textual). Además, en el Metathesaurus se localiza la agrupación de los términos con equivalente significado (<i>clusters</i>). El <i>cluster</i> es el que define al concepto y es identificado por una etiqueta preferente para su denominación.
Componentes estudiados:	
Metathesaurus y <i>Semantic Network</i> (Red Semántica).	

En la Red Semántica el nivel de granularidad es variable. Hay zonas de la Red con gran número de subcategorías, tanto en profundidad (categorías hijo) como en amplitud (subcategorías disjuntas o hermanas). Pero, hay otras en que sólo aparecen categorías muy generales. Esta

organización tiene implicaciones para la interpretación del significado de los conceptos que le han sido asignados desde el Metathesarus. Cada concepto de la Red incluye una descripción o definición textual.

2.3.4. Grado de formalización.

En este punto se realiza una descripción del tipo de lenguaje utilizado para la formalización de la representación. La categorización empleada se basa en la aportación de Uschold y Grüninges (1996), que diferencian la expresividad de los lenguajes en: altamente informales, semi-informales, semi-formales y formales, como anteriormente se ha indicado (pp. 24).

Terminología: CIE-9-MC Componentes estudiados: Lista Tabular e Índice Alfabético.	Responde a un catálogo de términos normalizados y formalizados, exclusivamente, en lenguaje natural (altamente informal). Aunque, si se atiende a la información complementaria, de las notas de alcance para su aplicación y las situaciones de inclusión y exclusión, se calificaría como catálogo glosado.
Terminología: MeSH	Es un catálogo glosado expresado en lenguaje natural pero de forma estructurada. Es una ontología de tipo semi-informal por el grado de formalización empleado.
Terminología: DeCS	Catálogo glosado trilingüe de tipo semi-informal como el MeSH.
Terminología: SNOMED CT	El lenguaje utilizado para la representación de los conceptos es formal. Está basado en descripción lógica que permite hacer una explicitación de la semántica de

los conceptos.

Además, en la actualidad, está alineada con los estándares más relevantes, en el campo de las aplicaciones médicas, como: HL7, DICOM, XML, ANSI y ISO para el intercambio de datos.

Terminología:	Su grado de formalización es máximo. La base de GALEN
Componentes estudiados:	GRAIL se fundamenta en el lenguaje de representación de conocimiento KL-ONE.
<i>Common Referente Model (CRM) y GALEN Representation and Integration Language (GRAIL).</i>	Con GRAIL se ha construido el CRM ya que éste posee los modeladores que especifican el número de conceptos primitivos, la descripción de sus características, y las reglas para la creación de conceptos, con términos adicionales o con otros ya presentes en el modelo. También, permite diferenciar entre conceptos e instancias.

Terminología:	Es una ontología semi-informal expresada en formato UMLS
	RRF (<i>Rich Release Format</i>).

2.3.5. Usabilidad y funcionalidad de la terminologías como ontologías.

Las limitaciones del análisis presentado derivan de las restricciones propias de los estudios descriptivos. La limitación más importante es que el análisis está centrado en el conocimiento declarado y no en la exactitud teórica de la representación que proporciona cada una de las terminologías estudiadas. De hecho, la exactitud teórica, de acuerdo al propio conocimiento del dominio, es un aspecto muy discutido. Los mayores representantes de esta línea argumental son, en dominios abiertos, Guarino

(Guarino, 1995 y 1998) y, en el dominio de la salud, Smith (Smith, 2003 y Smith et al, 2006).

No obstante, con el análisis realizado, se puede concluir que tanto los MeSH, los DeCS como el UMLS tienen una formalización reducida, lo que permite controlar, parcialmente, la ambigüedad de los conceptos, aunque hay diferencias significativas entre ellas. Además, hay diferencias en la riqueza de su estructura, por ejemplo, ésta es mayor en la Red Semántica de UMLS por la variedad de relaciones declaradas.

Todas estas terminologías pueden reutilizarse para el desarrollo de ontologías ligeras (Gómez-Pérez et al, 2004). Una ontología ligera permite la integración de fuentes de información heterogéneas, porque trabaja con conceptos que pueden asociarse a diversas etiquetas terminológicas y responder a diferentes intereses simultáneamente –clínicos, investigación, epidemiológicos, económicos y formativos–.

Las nomenclaturas SNOMED CT y GALEN cuentan con una mayor expresividad para especificar los conceptos, las propiedades, las restricciones y el tipo de relaciones incluidas. En ambos casos, podrían reutilizarse para el diseño de ontologías pesadas debido a que su formalismo declarativo, basado en descripciones lógicas, les posibilita el descubrimiento por inferencia de nueva información (Gómez-Pérez et al, 2004). Las ontologías pesadas son idóneas para su inclusión en sistemas de ayuda a la toma de decisiones diagnósticas o de tratamientos (Tabeada et al, 2004; Sobrado, et al, 2004) y como recurso semántico para su incorporación en tareas de PLN.

La CIE-9-MC es la única de las terminologías analizadas cuya reutilización como ontología conllevaría el mismo esfuerzo e inversión que si se creara una ontología nueva, atendiendo a la definición de Gruber (1993), Borst (1997) y Studer (1998).

Recientemente, se ha comenzado a diferenciar entre las terminologías de interface y las terminologías de referencia. Las terminologías de interface

son las concebidas para el uso directo de los usuarios finales e incluyen, generalmente, los listados de términos controlados, las clasificaciones o los tesauros (los más frecuentes en el ámbito de la salud son los indicados anteriormente). Las terminologías de interface o se enlazan (*mapping*) o se extraen de la terminología de referencia que utiliza el SI fuente. Las terminologías de referencia son las que contienen el modelo lógico de los significados de las entidades sobre las cuales se expresa la información. En otras palabras, las terminologías de referencia son las verdaderas ontologías empleadas por los SI (RIDE, 2007). La terminología clínica que se está imponiendo, como terminología de referencia, para la solución de los problemas de interoperabilidad semántica en las HCE es SNOMED CT⁶.

Además de lo indicado, en términos generales, las ontologías están jugando un papel importante en la investigación en salud y en biomedicina. Como indica Bodenreider (2008), el número de referencias bibliográficas indizadas en la base de datos Medline con el término ‘ontología’, ‘vocabulario controlado’ o ambos ha crecido, aproximadamente, un 600%, entre 1998 y finales de 2007, pasando de 200 a cerca de 1200 artículos en ese último año.

Estos trabajos muestran las líneas de investigación prioritarias y los resultados obtenidos en distintos campos y tareas. El uso del conocimiento facilitado por estos artefactos se ha valorado, como funcionalmente óptimo, para la gestión de conocimiento –tanto para la anotación de datos y recursos de información como para el acceso a ésta– (tabla 2.2), la integración de datos, el intercambio de información e interoperabilidad semántica (tabla 2.3) y el empleo en aplicación para el PLN (tabla 2.4).

En las tablas que se adjuntan a continuación se incluyen las referencias de los trabajos más significativos, a partir de la revisión

⁶ El Ministerio de Sanidad y Política Social firmó (14 Julio 2009) el convenio de adhesión con *International Health Terminology Standards Development Organisation* (IHTSDO) para su empleo en el desarrollo de las historias clínicas electrónicas (HCE) y para facilitar la interoperabilidad semántica en el Sistema Nacional de Salud. El IHTSDO es el organismo que desarrolla y mantiene SNOMED CT. (http://www.ihtsdo.org/fileadmin/user_upload/Docs_01/Press_Releases/IHTSDO_SPMSPS_250609_SP_.pdf).

realizada por Bodenreider (2008). En ellas se destaca la tarea realizada, la fuente de conocimiento empleada en cada caso y se distingue entre los que han empleado como material de trabajo las historias clínicas o los recursos bibliográficos.

Trabajos Publicados	Tipo de Funcionalidad
<i>Autores, Año</i>	<i>Gestión de Conocimiento: Anotación de Datos y Recursos de Información.</i>
Aronson et al, 2004 Zhang et al, 2006	Asignación automática de vocabulario controlado MeSH a documentos de investigación para su indexación en bases de datos bibliográficas utilizando UMLS.
Richesson et al, 2006 Daniel-Le Bozec et al, 2007	Codificación de documentos clínicos electrónicos y como vocabulario para los documentos de investigación clínica empleando SNOMED-CT
Liu et al, 2005	Utilización del Metathesaurus de UMLS para apoyar la codificación de informes de anatomía patológica.
Friedman et al, 2004 Pakhomov et al, 2006	Técnicas automáticas, basadas en PLN, para la asignación de códigos diagnósticos en documentos clínicos (no identifican la fuente de información).
<i>Autores, Año</i>	<i>Gestión de Conocimiento: Acceso a Información.</i>
Srinivasan, 1996 Aronson y Rindflesch, 1997 Hersh et al, 2000	Expansión de las preguntas, para asegurar la recuperación de la información, usando como fuentes de conocimiento MeSH y UMLS
Lowe et al, 1999 Brandt y Nadkarni, 2001 Ruiz, 2006 Yu y Lee, 2006	Proporcionar listas de sinónimos, relaciones entre conceptos, categorización de nivel superior, información concurrente para recuperar información en Medline, libros en Internet e imágenes médicas, empleando UMLS.
Hersh y Donohoe, 1998 Liu et al, 2006	Recuperación multilingüe de información utilizando MeSH y UMLS.
Muller et al, 2004 Gaudinat et al 2006 Can y Baykal, 2007 Ide et al, 2007	Motores de búsqueda de información biomédica que utilizan MeSH y UMLS para acceso a la literatura como a recursos web con información para los ciudadanos.
Cole et al, 2004	Sistemas de ayuda para localizar médico especialistas empleando SNOMED-CT
Lee et al, 2006 Yamamoto y Takagi, 2007	Detección de los temas de los textos médicos y generación de clusters de documentos para su clasificación utilizando la Red Semántica de UMLS
Fuente: Bodenreider, 2008	

Tabla 2.2. Uso práctico de las ontologías de salud en Gestión de Conocimiento.

Trabajos Publicados	Tipo de Funcionalidad
<i>Autores, Año</i>	<i>Integración de Datos, Intercambio de Información e Interoperabilidad Semántica.</i>
Chabaliar et al, 2007	Integración de información de enfermedades y guías clínicas empleando SNOMED-CT
Bouhaddou et al, 2007	Estrategias de mediación entre sistemas de información clínico de diferentes organizaciones utilizando UMLS y SNOMED-CT
Fuente: Bodenreider, 2008	

Tabla 2.3. Uso práctico de las ontologías de salud en Interoperabilidad Semántica.

Trabajo Publicados	Tipo de Funcionalidad
<i>Autores, Año</i>	<i>Aplicaciones para el Procesamiento del Lenguaje Natural .</i>
Chapman et al, 2004 Meystre y Haug, 2006 Uzuner et al, 2008	Extracción de información identificando entidades relacionadas con hechos especializados en los informes médicos, como el tiempo de fumador o el tipo y la evolución de hallazgos respiratorios. La fuente de conocimiento empleada fue UMLS (el trabajo de Uzuner no la indica).
Friedman et al, 1994 Rindfleisch et al, 2005	Extracción de relaciones de documentos clínicos o de informes de investigación utilizando UMLS y el <i>Medical Entities Dictionary del Columbia Presbyterian Medical Center</i> (recurso <i>ad hoc</i>).
Jacquemart y Zweigenbaum, 2003 Wedgwood, 2005 Terol et al, 2007	Sistemas de búsqueda de respuesta. La fuente básica es UMLS, aunque Terol también emplea WordNet.
Fiszman et al, 2004 Whalen, 2005 Reeve et al, 2007	Resúmenes de documentos y de libros médicos. La fuente común es UMLS.
Fuente: Bodenreider, 2008	

Tabla 2.4. Uso práctico de las ontologías en aplicaciones para el PLN.

2.4. Conclusiones.

En este capítulo se ha delimitado el significado del término ‘ontología’ a partir del análisis de las definiciones de mayor relevancia en el ámbito académico. En síntesis, una ontología define, operativamente, un conjunto de conceptos y de relaciones por medio de los términos comunes de un dominio. En su definición establece las condiciones para la combinación de los términos por medio de las relaciones y restricciones, lo cual delimita o extiende la interpretación de los términos incluidos en la ontología. No obstante, en la revisión realizada, se ha mostrado una amplia variabilidad y diversidad de tipos de ontologías, con diferentes niveles de idoneidad para su reutilización, completa o parcial. Las diferencias en la declaración de los componentes, el grado de la formalización y el ámbito de la conceptualización de las ontología son aspectos fundamentales a considerar antes de incorporar una ontología para la resolución de problemas de interoperabilidad semántica y tareas de PLN.

Aunque, la carencia de criterios unánimes, hace difícil y laborioso determinar la idoneidad de una ontología para su reutilización global o parcial, y también, establecer el grado de vacío conceptual y justificar un desarrollo ‘*ad hoc*’ para un sistema de información. En este sentido, una demanda creciente, para optimizar la tarea de selección de las ontologías existentes, es que éstas estén acreditadas explícitamente. La acreditación permitiría, de forma práctica y objetiva, compararlas y clasificarlas respecto a las necesidades del sistema información que se desea construir (o mejorar). Aunque, tampoco existe un formato común para describir la información relevante que caracteriza a una ontología y que sirva a los usuarios potenciales (Tello, 2002). Una tipología básica de indicadores, para acreditar el contenido de una ontología con potencial usabilidad para las tareas de PLN, de acuerdo con Hovy (1997), sería:

- *la cobertura conceptual* o el universo representados por los términos,
- *el tamaño* de la ontología, relativo tanto al número de conceptos como de términos incluidos,
- *la especificidad* o grado de abstracción de los términos incluidos,
- *la organización* conceptual relativa a las relaciones declaradas,
- *la riqueza* organizativa caracterizada por el número de interconexiones entre términos,
- *los axiomas* declarados respecto a su número y el tipo de proposiciones declaradas,
- *las reglas de inferencia* existentes, y
- *las instancias* o elementos particulares incluidos.

Estos indicadores son más operativos y orientan, objetivamente, a los potenciales usuarios sobre el grado de declaración de los componentes representados en el modelo y del tipo de cobertura de la conceptualización. La capacidad de expresividad y de razonamiento de la ontología también debería ser indicado, añadiendo la información sobre la técnica de modelado y el lenguaje de implementación empleado.

Las dificultades para la selección de una ontología idónea se han observado, claramente, en el contexto de la salud. Este contexto se caracteriza por la coexistencia de una gran variedad de tecnologías empleadas para representar el conocimiento del dominio. Las terminologías, en general, han sido el recurso clásico empleado. Éstas han recorrido un largo camino de evolución, desde las simples listas o jerarquías de términos hasta las representaciones formales de conocimiento. Aunque, sus posibles prestaciones para controlar la ambigüedad de los significados y la inferencia de nuevo conocimiento son desiguales, como se ha mostrado en el análisis realizado.

La investigación sobre el uso de la representación de conocimiento y, específicamente, de ontologías ha tenido un incremento importante, en el ámbito de la salud, en los últimos años. Los escenarios de desarrollo han sido la atención asistencial, la epidemiología, la documentación y los aspectos económico-administrativos sanitarios. Estas investigaciones han estado vinculadas con la gestión de conocimiento, la integración de datos, de intercambio de información e interoperabilidad semántica y el procesamiento del lenguaje natural (PLN).



Universitat d'Alacant
Universidad de Alicante

Capítulo III.

El Escenario Farmacoterapéutico.

El medicamento es el producto terapéutico más utilizado por, prácticamente, todas las especialidades médicas y en situaciones muy diversas, con la finalidad de prevenir, resolver o paliar los problemas de salud. El medicamento es una pieza clave que une el proceso sanitario con el proceso farmacoterapéutico.

El diagnóstico clínico de un problema de salud determina la selección del medicamento adecuado y el inicio del proceso farmacoterapéutico. El proceso continúa con la prescripción, la dispensación y la administración, y finaliza con la valoración de la evolución del paciente, es decir, comprobando el logro alcanzado sobre el problema de salud que determinó el uso del medicamento (figura 3.1).

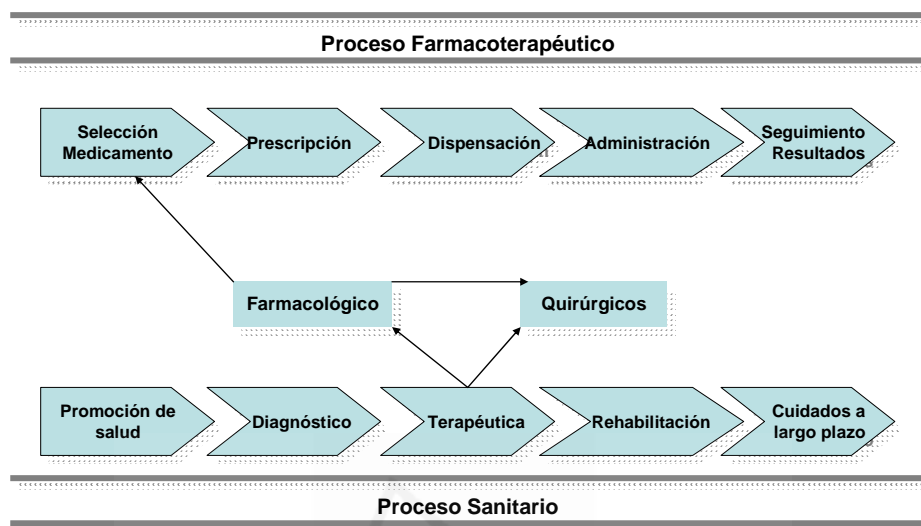


Figura 3.1. Fases del proceso farmacoterapéutico en relación con el proceso sanitario.

La terapéutica farmacológica brinda uno de los escenarios en el que mejor se puede concretar la complejidad y la interdependencia del *Sistema de Salud* y del *Sistema de Atención Sanitaria*. El *Sistema de Salud* abarca todas las organizaciones, las instituciones y los recursos de los que emanan iniciativas cuya principal finalidad es mejorar la salud global de los ciudadanos; por ejemplo, decidir la comercialización de un medicamento en el territorio nacional. En cambio, el *Sistema de Atención Sanitaria* lo conforman las instituciones, los recursos y los profesionales implicados en la prestación de servicios y de atención directa a los ciudadanos. Mientras que la atención sanitaria es el servicio que se recibe, indistintamente o de forma secuencial, en centros de salud, ambulatorios, centros de especialidades u hospitales, principalmente. Dicho servicio es realizado por uno o varios profesionales de la salud: médico, enfermera, odontólogo, matrona o farmacéutico, de acuerdo con su cualificación y de forma independiente o complementaria (Directiva, 2005). En la figura 3.2 se muestra, gráficamente, la infraestructura y los recursos humanos que

conforman tanto el Sistema de Salud como el Sistema de Atención Sanitaria relacionado con el propio proceso de atención sanitario (Robinson, 2007).

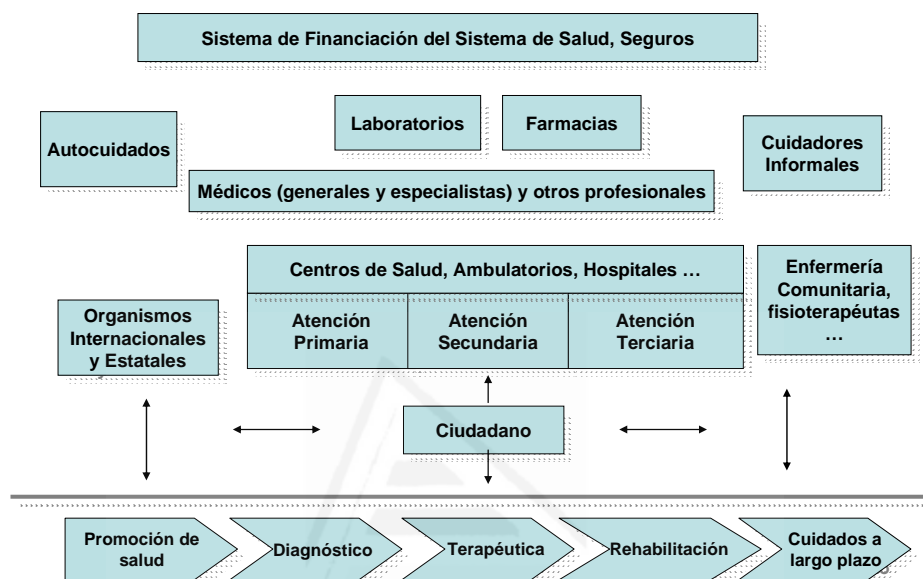


Figura 3.2. Sistemas de Salud y de Atención Sanitaria: instituciones, infraestructuras y recursos humanos. (Esquema adaptado de Robinson, 2007).

A partir de la identificación y el alcance global del escenario farmacoterapéutico, en los siguientes apartados se realiza un diagnóstico de la situación actual de los sistemas de información disponibles y de los requisitos necesarios para el desarrollo del proceso de la atención farmacoterapéutica.

El capítulo se ha organizado del siguiente modo: el primer apartado contiene una descripción general del sistema de información de la atención farmacoterapéutica y de las características de la información relevante para este proceso de atención. En el siguiente, se establecen las necesidades de información de los actores que intervienen en este escenario. El tercer apartado describe la cobertura actual de las fuentes de información, más comúnmente empleadas en este dominio, así como sus características y limitaciones para acceder a sus contenidos informativos. Se continúa con un

apartado en el que se identifican las prestaciones de los sistemas de información centrados en la prescripción farmacológica, tanto en la atención ambulatoria como en la hospitalaria. En este último ámbito, se reseña el uso de las técnicas de PLN para la identificación y la agrupación de la información sobre el tratamiento farmacológico disperso en la historia clínica electrónica (HCE). En el último apartado se indican las conclusiones alcanzadas.

3.1. Características del sistema de información de la atención farmacoterapéutica.

El sistema de atención sanitaria se puede entender, globalmente, como un '*sistema de información*' (SI) intensivo en conocimiento. De hecho, se le reconoce como uno de los sectores con mayor producción y uso de información (Cornella, 2000; Monteagudo, 2001). Como hemos indicado anteriormente, este sistema está constituido por un conjunto de profesionales y de recursos, combinados en una estructura diseñada para cumplir su misión y su finalidad social, la de prestar una atención de calidad para lograr el mejor estado de salud a cada ciudadano. Por consiguiente, es preciso comprender el papel de la información en el proceso de atención farmacoterapéutica y su vinculación en las relaciones profesionales que se producen en ella. También, es necesario entender cómo son las prestaciones actuales de las tecnologías de la información y comunicación (TIC), para que la información que se intercambia redunde en resultados óptimos para la propia organización en general y, en concreto, para el proceso farmacoterapéutico. Para ilustrar este contexto se puede utilizar el modelo de Itami (1987).

De acuerdo con este modelo, existen tres tipos de flujos de información, según de dónde proceda (información externa), cómo se constituya internamente (información interna) y a dónde se dirija (información corporativa). Los contenidos de estos flujos de información conforman el SI básico de una organización. Este modelo es muy simple pero muestra, claramente, la necesidad de intercambiar y de compartir la información que procede de diversos sistemas de información complementarios, tanto externos como internos, para fundamentar la toma de decisiones y facilitar la generación de nuevo conocimiento (Miralles, 2000).

La práctica sanitaria gira entorno a un núcleo de información vital de origen interno, conformada por los datos y la información sobre el estado de salud del ciudadano, así como por la información y el conocimiento verificado de los avances científicos, que constituye la información externa más destacada, aunque no la única.

Una ventaja del modelo de Itami es que permite definir una organización en su conjunto o en sus componentes elementales, a los que se denominan unidades funcionales (Cornella, 2000). En el escenario de la terapéutica farmacológica, se asume que una unidad funcional básica es la constituida por un conjunto de profesionales de la salud (médico, farmacéutico y enfermera) que se encuentran vinculados por el proceso de la atención farmacoterapéutica (selección, prescripción, dispensación, administración y seguimiento de resultados). Los componentes del SI y el flujo de información básico, en una unidad funcional prototipo de este escenario, se representan en la figura 3.3. Como puede observarse se ha especificado de forma independiente al nivel de atención o de la infraestructura donde se desarrolla la atención farmacoterapéutica.

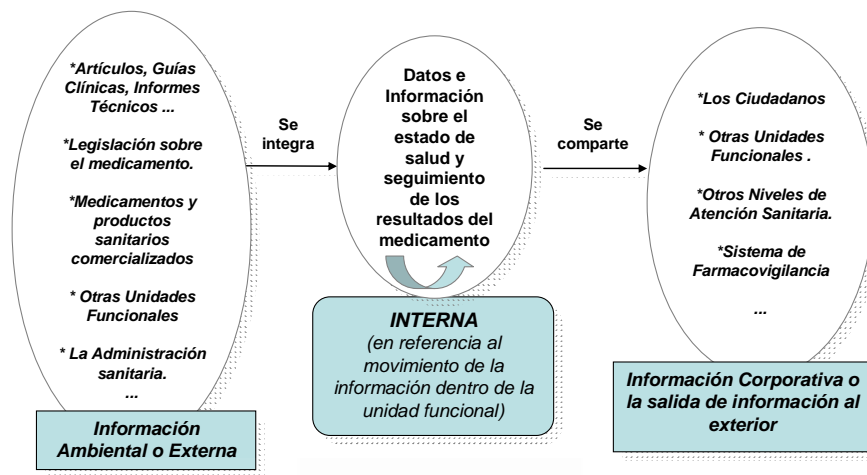


Figura 3.3. Componentes del SI y flujo de la información en el proceso de atención farmacoterapéutica de acuerdo con el modelo de Itami (1987).

Las tecnologías de información y comunicación (TIC), en el ámbito de la salud, han penetrado progresivamente y han cambiado la forma de comunicarse, diseminar y acceder a la información por parte de los profesionales de la salud. A pesar de esta creciente implantación, son escasos los SI que son capaces de interactuar e intercambiar la información automáticamente y, mucho menos, los SI que pueden llegar a compartir el significado de la información contenida en cada uno de ellos (CE, 2008). Todavía se mantienen SI independientes para cada una de las instituciones, departamentos y servicios que componen el sistema de atención sanitaria. Esta fragmentación de la información conlleva la existencia de información estanca, no integrada e imposible de compartir. Los profesionales son los que deben conocer el contenido y la forma de acceder a cada uno de los SI complementarios, para obtener la información necesaria que apoye su toma de decisiones, como se comprobará en los siguientes apartados.

3.1.1. Características de la información relevante del dominio.

La información relevante del SI farmacoterapéutico es toda aquella que facilita información sobre el medicamento más idóneo, de acuerdo a un problema de salud y las alternativas para su uso, respecto a las características individuales de cada sujeto. La información contenida en el SI debe ser fiable y actualizada para cubrir las necesidades de los diversos actores que intervienen en el proceso farmacoterapéutico: médicos, enfermeras, farmacéuticos (perfiles profesionales) y ciudadanos (perfil profano).

La pieza clave, tanto del proceso de atención como del SI, es el medicamento. Por medicamento se entiende toda preparación realizada con un fármaco con el fin de poderlo administrar y que interactúe con el organismo. Un medicamento puede contener uno o varios fármacos. Un fármaco es cualquier sustancia química capaz de producir un efecto sobre los seres vivos para prevenir, resolver o paliar un problema de salud. A los fármacos también se les denomina principio activo⁷ (de Ahumada, Santana y Serrano, 2003).

La extensión de la información relevante del dominio viene determinada por el número de medicamentos y por los formatos de comercialización factibles de ser utilizados, y a los que se les conoce como especialidades o presentaciones farmacéuticas. En España, a finales de 2008, se encontraban comercializadas 17.204 especialidades farmacéuticas. En ese mismo año se incorporaron 418 medicamentos con 30 principios activos nuevos y fueron dados de baja 923 presentaciones farmacéuticas. Asimismo, 45 medicamentos, ya comercializados, fueron autorizados para nuevas indicaciones terapéuticas (CGCOF, 2009).

Otra consideración a tener en cuenta, en el SI, es que no todas las especialidades farmacéuticas están disponibles en todos los países y, cuando

⁷ En la práctica asistencial se suelen utilizar los términos medicamento y fármaco como sinónimos y, a su vez, se diferencian de sus componentes, denominando a estos últimos como principios o sustancias activas.

lo están, no necesariamente se denominan igual. Incluso la denominación oficial del principio activo puede cambiar de un país a otro (ver un ejemplo en la tabla 3.1).

Nomenclatura	Denominación
<i>Internacional Nonproprietary Name</i> (INN) o Denominación Común Internacional (DCI) ⁸ .	Paracetamol
Denominación Oficial Española (DOE) del del fármaco o nombre genérico.	Paracetamol
<i>British Approved Name</i> (BAN).	Paracetamol
<i>United States Adopted Name</i> (USAN).	Acetaminophen (inglés) Acetominofen (castellano)
Otros nombre genéricos.	n-acetyl-p-aminophenol, p-acetamidophenol, APAP, acetamol...
IUPAC <i>name: system of naming chemical compounds</i> ⁹ .	N-(4-hydroxyphenyl)-acetamide (inglés) N-(4-hidroxifenil)-acetamida (castellano)
Denominación Comercial o nombre patentado.	Acecat, Adol, Apiretal, Atasol, Ben-u-ron, Calpol, Doliprane, Febrectal, Gelocatil, Panadol, Panamax, Paracetamol, Termalgin, Tylenol...
Formatos de Comercialización.	1 g polvo efervescente, 20 sobres. 500 mg 10 comprimidos. 500 mg comprimidos de dispersión bucal, 16 comprimidos. 100 mg/ml solución oral , 1 frasco de 30 ml. Pediátrico solución oral , 1 frasco de 90 ml.
Especialidad Farmacéutica.	PARACETAMOL ABEX Comp. 1 g APIRETAL Sol. oral gotas 100 mg/ml

Tabla 3.1. Variantes de denominación de un medicamento.

Asimismo, no hay que olvidar que la información sobre la atención sanitaria y farmacoterapéutica se crea en la lengua nativa. En la actual UE hay 23 idiomas y, en España, se cuenta con 4 lenguas cooficiales junto con el castellano. Pero, simultáneamente, gran parte de la información externa que se precisa para la atención se encuentra en lengua inglesa. Además, otro factor fundamental es la gran diferencia entre los lenguajes profesionales del

⁸ Este Sistema se inició por mandato de la Organización Mundial de la Salud (OMS) en 1950, <http://www.who.int/medicines/services/inn/en/index.html>

⁹ The International Union of Pure and Applied Chemistry (IUPAC), <http://www.iupac.org/>

ámbito de la salud (*hiperexia* o *hipertermia*) y el lenguaje coloquial (*calentura*) para referirse a los problemas de salud.

Estos aspectos, característicos de la información relevante para la atención farmacoterapéutica, justifican que este escenario sea un área prioritaria del espacio eHealth/eSalud. El objetivo de esta iniciativa es lograr compartir el significado entre las diversas nomenclaturas de denominaciones y las propiedades de los medicamentos comercializados en la UE, es decir, lograr la interoperabilidad semántica de los SI farmacoterapéuticos con el propósito de asegurar una completa libertad de tránsito a los ciudadanos y a los profesionales de la salud (Monteagudo, 2001; Robinson, 2008).

3.2. Necesidades de información de los actores del escenario.

Las necesidades de información vienen determinadas por el rol que se adopta en el proceso de atención farmacoterapéutica. Los diferentes roles son el de prescriptor, el de dispensador, el de administrador y el de evaluador de los resultados (Herrera, 2003).

Los estudios publicados sólo describen, globalmente, la necesidad informativa sobre el medicamento de los actores que intervienen en este escenario (Loza et al, 2000; Orueta, 2004; Baixauli, Salar y Barbero, 2004; Dee y Stanley, 2005; Muñoz-Izquierdo, 2008). Aunque, entre ellos, destacan los trabajos de Ely y equipo (1999, 2000, 2002) y de González-González y equipo (2005, 2006 y 2007), los cuales recogen las cuestiones que más frecuentemente se relacionan con la prescripción, la administración y el seguimiento de los resultados alcanzados con la prescripción, en la atención ambulatoria y en Sistemas de Salud diferentes (EEUU y España).

A partir del análisis de los trabajos citados, las necesidades de información de cada actor, de acuerdo con su rol en el proceso de atención farmacoterapéutica, se sintetizan en la tabla 3.2.

Rol	Prescriptor	Dispensador	Administrador	Evaluador
Profesional	Médico. Enfermera.	Farmacéutico.	Enfermera. ----- Ciudadano.	Médico. Enfermera. Farmacéutico.
Necesidad Informativa	Denominación, Principios activos o composición, Posología (dosis, frecuencia y duración), Eficacia, Farmacodinámica, Mecanismos de acción, Efectos adversos, Interacciones y Contraindicaciones.	Interacciones y Reacciones adversas, Indicación y Posología.	Composición, Indicación, Acción y Efecto, Modo de administrar, Precauciones, Dosis, frecuencia y duración, Efectos adversos, Interacciones y Contraindicaciones. ----- Indicación, Dosis, frecuencia y duración, Modo de administrar y Efectos adversos.	Efectividad del medicamento (efecto terapéutico). Efectos adversos (manifestaciones clínicas).

Tabla 3.2. Necesidades de información de los actores en la atención farmacoterapéutica.

3.3. Fuentes de Información.

Las fuentes de información, desde la perspectiva de la ciencia de la información, de forma amplia y genérica, son todos aquellos instrumentos y recursos donde se encuentra la información necesaria para obtener una respuesta pertinente y motivada (Villaseñor, 1998). Las fuentes de información sobre el medicamento son múltiples y diversas, tanto en estructura como en contenido. Hay fuentes de tipo narrativo como los manuales, los prospectos, los informes técnicos o los boletines, por ejemplo.

Además, están los recursos que proporcionan información estructurada o bases de datos.

En este apartado sólo nos vamos a referir a las fuentes de información que son citadas en la literatura, como más frecuentemente consultadas, en el ámbito español y que guardan relación con las internacionales (Ely et al, 1999; Loza et al, 2000; Orueta, 2004, Dee y Stanley, 2005, González-González et al, 2007). Dichas fuentes ofrecen una garantía de calidad sobre el contenido informativo recuperado (actualizado, fiable y válido) para la atención farmacoterapéutica (para otras fuentes de información disponibles en EEUU ver el trabajo de Sharp, Bodenreider y Wacholder, 2008).

Las bases de datos, por su contenido, se pueden agrupar en dos tipos: las relacionadas con los medicamentos comercializados y las relacionadas con la bibliografía científica. Dentro de las primeras se puede, a su vez, distinguir dos variantes. Las bases de datos cuyo contenido coincide con los prospectos autorizados para los ciudadanos, aunque dicha información, también, es relevante para los profesionales. En este grupo tenemos como ejemplos, el Vademécum Internacional¹⁰ o la Base de datos del Medicamento (BOT)^{11 12}. La segunda variante son las bases de datos de Organismos Oficiales cuyo contenido informativo es de tipo, exclusivamente, profesional. Estas bases de datos son, realmente, repositorios de los informes técnicos que han sido revisados, por los propios Organismos, para la autorización del uso comercial de un medicamento en un territorio geográfico (nacional o en todo el espacio de la UE). Destacan en este grupo la bases de datos de la AEMPS (Agencia Española de

¹⁰ Disponible en <http://www.vademecum.es/>

¹¹ Disponible en <http://www.portalfarma.com/pfarma/espacioges.nsf/wfAccesoEspacioGestion?OpenForm>

¹² La fuente de información equivalente en el Reino Unido es el *Electronic Medicines Compendium* (eMC). Disponible en <http://emc.medicines.org.uk/default.aspx>

Medicamentos y Productos Sanitarios)¹³ y la EMEA (Agencia Europea de Medicamentos)¹⁴.

El segundo tipo de bases de datos son las que incluyen bibliografía científica, cuyo contenido suele girar entorno a los resultados de eficacia-efectividad de un medicamento o estudios comparativos e indicaciones de uso. Entre los artículos de mayor interés están los ensayos clínicos, las revisiones sistemáticas y las guías clínicas para la prescripción y el seguimiento de la acción terapéutica. Los ejemplos más representativos de este grupo de bases de datos bibliográfica son Medline (Medical Literature Analysis and Retrieval System Online)¹⁵ e IME (Índice Médico Español)¹⁶, que, aunque no son exclusivas, sí son comunes para los actores profesionales de este escenario (Loza et al, 2000).

En la tabla 3.3 se ha sintetizado el contenido informativo de las fuentes que se relacionan, directamente, con el medicamento: Manuales de farmacología y prospectos de medicamentos en formato papel, así como informes técnicos de la AEMPS. También, se incluye el contenido que facilitan las bases de datos de Vademécum Internacional y BOT; pero, en ambos casos la valoración incluida es la obtenida por el acceso abierto, el cual facilita una información limitada. Otro aspecto a mencionar es que, aunque un contenido informativo pueda estar presente en todas las fuentes, la extensión y profundidad informativa difiere de una fuente a otra.

¹³ Centro de Información online de Medicamentos de la AEMPS, <https://sinaem4.agemed.es/consaem/fichasTecnicas.do?metodo=detalleForm>

¹⁴ *European Public Assessment Report*, <http://www.emea.europa.eu/htms/human/epar/eparintro.htm>

¹⁵ Acceso por el Sistema de Recuperación PubMed, <http://www.ncbi.nlm.nih.gov/pubmed/>

¹⁶ Disponible en http://bddoc.csic.es:8080/inicioBuscarSimple.do?tabla=docu&bd=IME&estado_formulario=show

	<div>Fuente</div> <div>Tipo Información</div>	Manuales Farmacología	Prospectos de Medicamento	Informes Técnicos	Vademécum Internacional	BOT
Prescripción	Nombre Comercial		X	X	X	X
	Concentración		X	X	X	X
	Forma	X	X	X	X	X
	Dosis		X	X	X	X
Química	Nombre químico	X	X	X		
	Nombre genérico (DOE)	X	X	X	X	X
	Composición		X	X		X
	Excipientes		X	X		X
Biología	Mecanismos de acción	X	X	X	X	
	Efecto biológico	X		X		
	Toxicidad	X		X		
Clínica	Clase terapéutica	X	X	X	X	X
	Indicación	X	X	X	X	X
	Interacción	X	X	X	X	X
	Contraindicación	X	X	X	X	
	Efecto adverso	X	X	X	X	
	Vía administración	X	X	X	X	X
	Precauciones	X	X	X	X	
	Intoxicación		X	X	X	
Nota Aclaratoria: Los criterios de valoración del contenido informativo son una adaptación de la valoración realizada por Sharp, Bodenreider y Wacholder, 2008. La valoración incluida de Vademécum Internacional y BOT se ha realizado a partir del contenido informativo obtenido por el acceso abierto de estas bases de datos, el cual facilita una información limitada.						

Tabla 3.3. Tipo de contenido informativo de las fuentes de información sobre el medicamento.

3.3.1. Características del acceso y limitaciones.

Las fuentes de información organizan su contenido informativo con su propio vocabulario. Por una parte están las fuentes de información que utilizan un vocabulario controlado, compuesto por descriptores o encabezamientos de materia, es decir, con términos normalizados como en

la base de datos Medline (ver el ejemplo de MeSH¹⁷ en la tabla 3.4). Por otro lado, están las fuentes que utilizan un vocabulario libre, constituido por las palabras clave o por la frecuencia de aparición de términos en algún campo de su base de datos, como en el caso de la base de datos IME (tabla 3.5), y en el que coexiste el término, tanto en singular como en plural, y sus posibles sinónimos. Los términos empleados por cada vocabulario, debido a su organización interna, establecen un modelo diferenciado del mismo dominio. Estas estructuras organizativas carecen de una explicitación formal de su significado, lo que dificulta tomar el conocimiento de una fuente y utilizarlo en otra para dar respuestas a las necesidades informativas de los usuarios.

En este sentido, para acceder a la información sobre el medicamento, en los ‘manuales’ y en los ‘prospectos’ se cuenta con la estructura narrativa de los apartados. Esta estructura permiten a los humanos identificar los bloques de interés y donde es probable encontrar la respuesta pertinente. Los ‘prospectos’ han normalizado la estructura de sus apartados. Recientemente, se ha incluido un apartado con los nombres que puede tener un medicamento en los diferentes países de la UE. Aunque, este último apartado sólo se incluye cuando el medicamento está autorizado mediante el procedimiento de reconocimiento mutuo (RD, 2007). En cambio, en los ‘manuales’, la estructuración de la información depende de la creatividad de sus autores; no obstante, mantienen un núcleo constante de información para los profesionales basado, normalmente, en las indicaciones terapéuticas de los fármacos pero, muy versátil.

¹⁷ El encabezamiento de materia utilizado en la base de datos bibliográfica Medline.

<div>All MeSH Categories</div> <div>Disciplines and Occupations Category</div> <div>Natural Science Disciplines</div> <div>Biological Science Disciplines</div> <div>Pharmacology</div> <div>Biopharmaceutics</div> <div>Chemistry, Pharmaceutical</div> <div>Drug Discovery +</div> <div>Ethnopharmacology</div> <div>Neuropharmacology</div> <div>Pharmacoepidemiology</div> <div>Pharmacogenetics</div> <div>Toxicogenetics</div> <div>Pharmacognosy</div> <div>Herbal Medicine</div> <div>Pharmacology, Clinical</div> <div>Psychopharmacology</div> <div>Toxicology</div> <div>Ecotoxicology</div> <div>Forensic Toxicology</div> <div>Toxicogenetics</div>	<div>All MeSH Categories</div> <div>Chemicals and Drugs Category</div> <div>Pharmaceutical Preparations</div> <div>Designer Drugs</div> <div>Dosage Forms</div> <div>Capsules</div> <div>Colloids +</div> <div>Delayed-Action Preparations +</div> <div>Drug Carriers +</div> <div>Liniments</div> <div>Micelles</div> <div>Ointments</div> <div>Pharmaceutical Solutions +</div> <div>Powders</div> <div>Suppositories</div> <div>Tablets +</div> <div>Vaginal Creams, Foams, and Jellies</div> <div>Drug Combinations</div> <div>Drugs, Chinese Herbal</div> <div>Drugs, Essential</div> <div>Drugs, Generic</div> <div>Drugs, Investigational</div> <div>Materia Medica</div> <div>Nonprescription Drugs</div> <div>Nostrums</div> <div>Pharmaceutic Aids</div>
--	--

Tabla 3.4. Fragmento de la representación del conocimiento fármacoterapéutico en el MeSH

<div>...</div> <div>FARMACO(124)</div> <div>FARMACO-ALIMENTARIAS(1)</div> <div>FARMACO-ANESTESICO(1)</div> <div>FARMACO-CLINICA(3)</div> <div>FARMACO-DIANA(1)</div> <div>FARMACO-EEG(1)</div> <div>FARMACO-NUTRICION(1)</div> <div>FARMACO-PSICODINAMICA(1)</div> <div>FARMACO-RADIOLOGICA(1)</div> <div>FARMACO-TERAPEUTICO(2)</div> <div>FARMACO-TERAPEUTICOS(1)</div> <div>FARMACO-TOXICOLOGIA(1)</div> <div>FARMACOACTIVOS(6)</div> <div>FARMACOBIOLOGIA(1)</div> <div>FARMACOCINETICA(233) ...</div>	<div>...</div> <div>MEDICAMENTO(60)</div> <div>MEDICAMENTO-ALIMENTO(1)</div> <div>MEDICAMENTO-MEDICAMENTO(2)</div> <div>MEDICAMENTOS(904)</div> <div>MEDICAMENTOSA(87)</div> <div>MEDICAMENTOSAS(83)</div> <div>MEDICAMENTOSSES(1)</div> <div>MEDICAMENTOSO(38)</div> <div>MEDICAMENTOSOS(9)</div> <div>MEDICAMENTS(4)</div> <div>MEDICARE(1)</div> <div>MEDICARUM(1)</div> <div>MEDICAS(278)</div> <div>MEDICATIONIS(1)</div> <div>MEDICATIONS(1)...</div>
--	---

Tabla 3.5. Fragmento del índice del campo título de la base de datos IME

La recuperación de los ‘informes técnicos’ se puede realizar a través de la base de datos de la Agencia Española de Medicamentos y Productos

Sanitarios (AEMPS). El formulario de búsqueda permite la recuperación a través de los campos: *principio activo*, *nombre de la presentación del medicamento* y *grupo terapéutico*. Estos mismos campos son los disponibles en el ‘Vademécum Internacional’ pero, en este caso, los denomina *nombre*, *principio activo* y *terapéutico*. En la ‘Base de Datos del Medicamento (BOT)’ se recupera por los campos *nombre de la especialidad* y *principio activo*. Las variantes de denominación de los campos puede llegar a ser un obstáculo para el acceso al contenido informativo. Asimismo, la información ofrecida varía, ya que, en el primer caso, se accede a un documento completo y organizado en apartados en los que hay que buscar la respuesta que se precisa. Mientras que, las dos últimas bases de datos, ofrecen información estructurada pero, por lo general, se precisa acceder en varias ocasiones al sistema de recuperación para lograr la respuesta requerida, sobre todo, cuando es necesario comparar las similitudes y las diferencias de más de dos medicamentos.

La queja generalizada de los profesionales sanitarios es la escasez de tiempo disponible para realizar la búsqueda de información. El problema, que sigue siendo un obstáculo en la práctica profesional actual, es recuperar la información que se precisa, durante la misma atención y en el menor tiempo (Ely et al, 1999; Dee y Stanley, 2005; Coumou y Meijman, 2006; González-González et al, 2006, 2007). Sólo se ha encontrado un estudio específico que cuantifique el tiempo requerido por los profesionales para obtener respuestas relacionadas con la terapéutica farmacológica (Ely et al, 1999). Otros trabajos valoran el tiempo total requerido para cualquier cuestión relacionada con la atención sanitaria, como el de González-González y equipo (2007), cuya población de estudio son los médicos españoles de atención primaria. Respecto al tiempo que precisa un ciudadano hay de nuevo un vacío. Aunque, se puede utilizar el trabajo de Baixauli, Salar y Barbero (2004) reinterpretando que el tiempo que el consumidor emplea para obtener la respuesta sobre las dudas de la prescripción es el tiempo que el farmacéutico dedica en responder. En la

tabla 3.6 mostramos el tiempo promedio que se precisa para encontrar la respuesta de una cuestión durante la atención sanitaria.

Estudio/Población estudiada	Tiempo para obtener la respuesta	Fuente de Información
González-González et al (2007). Médicos de Atención Primaria (familia y pediatría), España.	Media de 2,25 minutos (rango entre 1,23-3,27) para cualquier tipo de cuestión relacionada con la atención sanitaria.	El 9,6% de las cuestiones son relativas a la prescripción y a los aspectos clínicos del medicamento. La cuestión más frecuente fue la relativa a si un medicamento estaba indicado para un problema. La fuente de información principalmente empleada fue el vademécum.
Ely et al (1999). Médicos de Familia, EE.UU.	Consulta para la prescripción: entre 1,23 y 1,16 minutos (respuesta útil 85% de los casos). Utilizando aplicaciones Informáticas el tiempo fue de 6,58 minutos (respuestas útiles en el 20%).	El 19% de las cuestiones se refieren a la prescripción de un medicamento. La cuestión más buscada fue la dosis adecuada.
Baixauli, Salar y Barbero (2004): reinterpretado. Consumidor, España	Media de 4,4 minutos (rango entre 4 y 4,8)	El 49,5% de las consultas, atendidas por los farmacéuticos, están relacionadas con distintos aspectos de la prescripción del medicamento (dosis, frecuencia y uso).
Nota Aclaratoria: En España el tiempo de consulta en atención primaria está situado entre 5 y menos de 10 minutos por paciente. El tiempo de consulta indicado en la muestra de los médicos de familia de Iowa era entre 10 y 15 minutos.		

Tabla 3.6. Tiempo invertido para obtener respuestas a las cuestiones que surgen durante la atención.

3.4. Prestaciones de los sistemas de información para la prescripción farmacológica.

En el actual contexto del sistema de atención sanitaria, se diferencia entre los sistemas de información del ámbito de la atención primaria y los desarrollados para el ámbito de la atención hospitalaria. En el primer caso, se habla de los subsistemas de prescripción electrónica¹⁸ y, en el segundo, se

¹⁸ En la literatura anglosajona se la denomina *Computerized Physician Order Entry* (CPOE).

trataría de la histórica clínica electrónica (HCE¹⁹). Aunque, ambos casos, todavía se encuentran en una fase temprana de incorporación, principalmente, los del entorno hospitalario español.

A continuación se indican las prestaciones y las limitaciones actuales para la gestión y el procesamiento de la información farmacoterapéutica de estos sistemas desde el punto de vista del usuario.

3.4.1. La prescripción electrónica.

La prescripción electrónica es una de las tecnologías más extendidas, que ha facilitado un registro sistemático de la medicación de los pacientes en el medio ambulatorio, tanto para el control histórico como para el control de la medicación en uso actual. Esta tecnología es considerada como una herramienta de gran ayuda para la práctica profesional de médicos, enfermeras y farmacéuticos (Ferner, 2004). La funcionalidad principal es la de prevenir que se prescriba un mismo fármaco con distintas denominaciones y presentaciones, y la consecuencia de una intoxicación por sobredosificación (Ferner, 2004; Robinson, 2008).

En España, estas aplicaciones se auxilian de las tablas de la base de datos relacional '*Nomenclátor Digitalis*'²⁰, la cual contiene la actualización mensual de los medicamentos autorizados. Esta base de datos facilita el *nombre de la especialidad farmacéutica, los principios activos, la clasificación terapéutica, la forma farmacéutica, la dosis de referencia, la*

¹⁹ También se utiliza el nombre genérico de Historia de Salud Electrónica (*Electronic Health Record o EHR*). No hay un consenso sobre qué es y qué debe contener. Algunos sistemas incluyen, virtualmente, todos datos de los pacientes; otros, están limitados a ciertos tipos de datos. La mayor parte de las aplicaciones, de Historia de Salud Electrónica, son específicas del proveedor o del usuario final y sólo algunas permiten el intercambio y la comunicación entre sistemas de diferentes proveedores en una comunidad o una región. Una visión más detallada se encuentra en el 5º Informe de la Sociedad Española de Informática de la Salud (SEIS), "De la historia clínica a la historia de salud electrónica" (2003), disponible en: <http://www.conganat.org/SEIS/informes/2003/>.

²⁰ Base de datos mantenida por la Dirección General de Farmacia y Productos Sanitarios del Ministerio de Sanidad y Política Social.

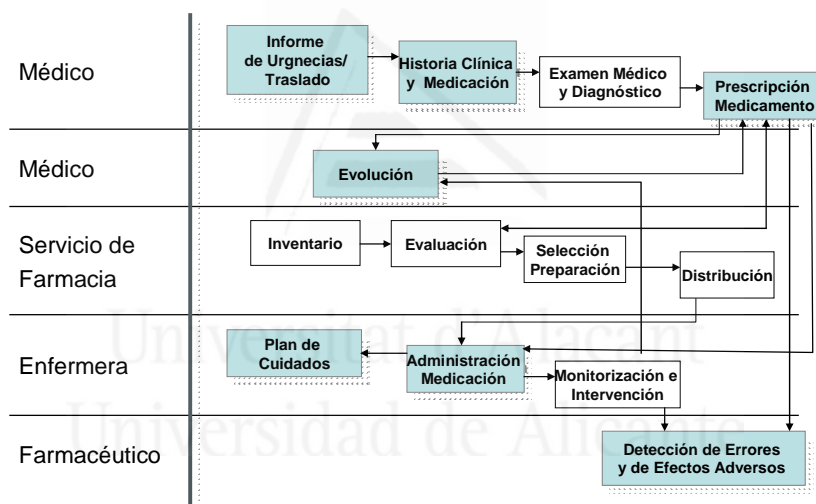
frecuencia y la vía de administración. Esta base de datos cuenta con, al menos, dos limitaciones. Por un lado, no cuenta con las posibles denominaciones de los medicamentos en otros países de la UE lo que, como manifiestan algunos profesionales, es un gran inconveniente para la atención de los ciudadanos europeos que residen, temporalmente, en nuestro país. Por otra parte, los campos como los valores incluidos en '*Nomenclátor Digitalis*' carecen de significación semántica y sólo pueden ser interpretados por un humano.

En los sistemas de prescripción ambulatoria más completos, se cuenta con sistemas de alerta para la detección anticipada de interacciones entre medicamentos y las relacionadas con las alergias (por ejemplo; en el Sistema de Salud Andaluz se está comenzando a desarrollar estos módulos complementarios). Aunque, en términos generales, se informa de deficiencias en el enlace entre la información clínica y la del medicamento para la detección de contraindicaciones relativas a la edad, los cambios fisiológicos (embarazo) y los problemas de salud colaterales (Fernando et al, 2004). Además, en ningún caso se hace referencia a la estructura y la semántica de los datos o al tipo de conocimiento que se utiliza para apoyar estos sistemas de alerta.

3.4.2. La prescripción farmacoterapéutica en la HCE.

En el ámbito de la atención hospitalaria se comienza a implantar la HCE. Su finalidad es recopilar, almacenar y facilitar el acceso a la información sobre el estado de salud del paciente, de acuerdo con el perfil profesional y la responsabilidad que se tenga en la prestación de atención. Existen variantes en la forma de diseñarla y en la rigidez o la flexibilidad respecto a la información, estructurada o abierta, que contienen los módulos que conforman la HCE.

El problema respecto a la terapéutica farmacológica se sitúa en la necesidad de reconciliar (detectar y agrupar) las distintas informaciones que, sobre la medicación, se encuentran dispersas por los distintos módulos de la HCE (figura 3.4). Por un lado, se encuentra la información incorporada a través de los registros no estructurados, como suelen ser el informe de ingreso por traslado o por urgencias, el informe preoperatorio, la hoja de evolución, las evaluaciones de las actividades de enfermería en los planes de cuidados. y el informe de alta; y por otra, la contenida en registros con mayor estructuración como son los de prescripción y los de administración.



Nota Aclaratoria: Los recuadros en color indican los módulos con información sobre el tratamiento farmacológico y sus efectos. Las flechas muestran el flujo teórico de intercambio de información entre los módulos y las actuaciones.

Figura 3.4 Representación de los módulos básicos de una HCE con información farmacoterapéutica.

3.4.2.1. Aplicación de técnicas de PLN.

Considerando el problema de la reconciliación de la información farmacológica, las técnicas basadas en procesamiento del lenguaje natural (PLN) se están introduciendo, en el ámbito internacional, para las tareas de reconocimiento y de extracción de los términos relativos a la medicación. El

propósito de su uso, en el ámbito hospitalario, es el de crear listados cronológicos de medicación que faciliten la actuación asistencial, evitando las interacciones y los efectos adversos prevenibles.

Las técnicas y las metodologías aplicadas para dar solución al problema de la reconciliación son diversas. Una de las estrategias es la que plantean Cimino, Bright y Li (2007), los cuales utilizan el sistema MedLEE²¹. Este sistema se auxilia de un vocabulario '*ad hoc*' del hospital y de UMLS (Metathesaurus y Red Semántica) para proporcionar otros términos y conceptos al sistema. Logran una clasificación de la medicación respecto a su indicación terapéutica. En su informe destacan la identificación del 95% de los términos, aunque indican que sólo obtienen una correspondencia (*mapping*) del 50% entre los conceptos de UMLS y el vocabulario '*ad hoc*' para asociar la indicación terapéutica con la medicación. Su recomendación final es la de mejorar el análisis realizado por MedLEE, con el objetivo de poder identificar las entidades temporales y poder diferenciar entre la medicación actual y la medicación suspendida (histórico); así como, la actualización de las correspondencias entre los recursos terminológicos auxiliares para incrementar la efectividad de los resultados.

Un aspecto que destaca del trabajo anterior es que, en él, no se hace referencia a la gran variabilidad de la escritura en los documentos asistenciales, donde la norma es, en muchas ocasiones, la carencia de normas. Estos documentos contienen un gran número de siglas (*IV: intravenoso*), abreviaturas (*sol o Sol: solución*), acrónimos (*SRAS: síndrome respiratorio agudo severo*), símbolos (*0-1-0: en la comida*) utilizando, en ocasiones, los referentes normalizados y, en otras, utilizando adaptaciones personales con cambios incluso en la ortografía. Además, pueden incluir

²¹ Es un sistema de PLN diseñado para extraer, estructurar y codificar información clínica a partir de los textos incluidos en los informes asistenciales. Está operativo desde 1995. MedLEE fue creado por Carol Friedman en colaboración con el Departamento de Biomedicina Informática y el Departamento de Radiología de la Universidad de Columbia. En la actualidad, está integrado en el Sistema de Información Clínico del Hospital Presbiteriano de Nueva York. Una demostración de MedLEE está accesible en <http://zellig.cpmc.columbia.edu/medlee/>

vocablos no recogidos en las terminologías controladas médicas (*AMC: amoxicilina clavulánico*) o incorporados recientemente al vocabulario profesional (*NIH1: virus gripe A*). En otros casos, los términos clínicos cambian de significado en función del contexto de especialización médica (*HTA: histerectomía total abdominal en ginecología o hipertensión arterial en el resto de especialidades médicas*). Y, por tanto, se ignora el problema que supone la detección y el reconocimiento, para lograr resultados óptimos, en la identificación de las entidades relacionadas con la medicación para su agrupación.

En este sentido se dirige el trabajo de Levin y su equipo (2007), los cuales plantean un mecanismo de normalización de los datos sobre medicación para facilitar su extracción. Este equipo utiliza un algoritmo de post-procesamiento genérico, Metaphone, diseñado inicialmente para la identificación de apellidos. Con él, analizan y normalizan la medicación en texto libre por medio de una clasificación de los sustantivos por agrupación de consonantes. La fuente de conocimiento externo que utilizan es RxNorm²², para verificar los nombres comerciales y establecer su equivalencia genérica o principio activo. Con el procedimiento descrito, indican obtener una sensibilidad del 92,3% y una especificidad del 95,7% en la identificación y en la extracción de entidades. Aunque proponen varias mejoras, tanto del método como del algoritmo, para reducir los falsos negativos. Entre las recomendaciones que consideran más efectivas está la de buscar, directamente, los *tokens* no emparejados en RxNorm y antes de utilizar Metaphone.

Otro problema práctico, en la reconciliación de la información sobre la medicación, es el de lograr el reconocimiento y la diferenciación de nombres de medicamentos equívocos por semejanza ortográfica o fonética. Para este área de investigación, también, se han incorporado técnicas de PLN, como la propuesta de Kondrak y Dorr (2006) que comparan la

²² RxNorm es una nomenclatura normalizada de la medicación clínica y de los dispositivos para su administración de EEUU. Este recurso ha sido desarrollado por la *National Library of Medicine* (NLM) (<http://www.nlm.nih.gov/research/umls/rxnorm/>)

eficiencia de dos algoritmos de similaridad ortográfica y fonética. Aunque, tampoco se pueden dejar de mencionar las aportaciones realizadas para el reconocimiento y el emparejamiento entre entidades representativas del par ‘enfermedad-medicamento’, a partir del texto de los documentos asistenciales (Chen et al, 2007), o entre las cuestiones de consultas de los ciudadanos y las respuestas facilitadas por los médicos en servicios de ayuda en la Web (Slaughter, Soergel y Rindflesch, 2006).

Un enfoque diferenciado es el realizado por Jagannathan y su equipo (2009), los cuales realizan una valoración de la capacidad actual de cuatro aplicaciones comerciales de procesamiento del lenguaje natural (tabla 3.7) para la extracción de información relacionada con la medicación en documentos clínicos.

Aplicación Comercial	Características
Language and Computing	Su aproximación cuenta con una ontología de conceptos clínicos. Su foco actual es el <i>mapping</i> de SNOMED CT y de CIE-9-MC para la codificación y facturación.
Coderyte	Su aproximación es estadística para la tareas de PLN. Su ámbito actual es la codificación de las notas clínicas ambulatorias y la asignación de códigos CIE-9-MC para facturación.
LingoLogics	Su desarrollo parte de los resultados de investigaciones realizados en la Clínica Mayo. Su aproximación cuenta con una ontología. Su ámbito de trabajo se centra en la codificación de la notas clínicas ambulatorios con CIE-9-MC.
Artificial Medical Intelligence (AMI)	Su aproximación es híbrida y combina terminologías con reglas. Su foco es apoyar la codificación y la facturación en ámbitos ambulatorios.

Tabla 3.7 Características de cuatro aplicaciones comerciales de PLN en salud.

El equipo de Jagannathan trabajó con 2.000 documentos (1.000 informes de alta y 1.000 notas clínicas de médicos de familia), previamente

des-identificados²³. Los mejores resultados se obtuvieron con los informes de alta y, globalmente, fue más efectivo el reconocimiento del nombre del medicamento (medida-F=93,2%), bajando su tasa a un 85,3% cuando la tarea era reconocer la concentración (cantidad y unidad de medida) y a un 80,3% para el reconocimiento de las entidades relacionadas con la vía de administración. Los peores resultados se obtuvieron en la tarea de reconocimiento de entidades relacionadas con la periodicidad (48,3%). Estos resultados se pueden explicar por la finalidad general que tienen todos estos sistemas comercializados, orientados al proceso de facturación, donde la mayoría de estas entidades no tienen relevancia. Estos sistemas analizados no son sistemas de PLN ideados para resolver problemas prácticos asistenciales, por lo que deberían ser adaptados.

Aunque, fuera de este problema concreto de la reconciliación de la información sobre el medicamento, los grupos de investigación en PLN españoles, también, han comenzado a mostrar interés por el dominio de la salud. Considerando el escenario de estudio de la farmacoterapéutica, se destaca, en primer lugar, el trabajo de Segura, Martínez y Samy (2008). Este grupo trabaja en la clasificación de los nombres genéricos de los fármacos reconocidos en los resúmenes de artículos en lengua inglesa. La fuente de conocimiento que emplean es UMLS y la clasificación de las denominaciones de USAN (*United States Adopted Name*). Respecto a esta última fuente, debemos comentar que la formación del término empleado para establecer la denominación USAN del medicamento genérico no siempre tiene correspondencia con la Denominación Oficial Española (DOE) (ver el ejemplo incluido en la tabla 3.1). Esta línea de trabajo tiene

²³ En el ámbito de la salud tiene una importancia vital la protección de los datos de privacidad y confidencialidad (Ley 16/2003 de Cohesión y Calidad del Sistema Nacional de Salud) y es casi imposible de separar de los datos personales (Ley 15/1999 de Protección de Datos de Carácter Personal). La des-identificación o anonimización es una tarea que elimina los datos de identificación, directa o indirecta, tanto de la persona atendida como de los profesionales que lo han tratado. La des-identificación está teniendo cada vez más protagonismo, en el ámbito sanitario, ante la necesidad de reutilizar los documentos asistenciales para la investigación y la docencia. Esta tarea se relaciona con el reconocimiento de las entidades nombradas. Los elementos clave de la des-identificación fueron regulados en EEUU por la *Health Information Portability and Accountability Act* (1996) y se han convertido en un referente internacional.

un gran valor para los investigadores y la industria farmacológica, pero le resta, inicialmente, efectividad para cubrir las necesidades de los profesionales asistenciales.

Otra línea significativa es la centrada en los sistemas de búsqueda de respuestas. En ella, se encuadran los trabajos realizados por Terol, Martínez-Barco y Palomar (2006 y 2007) y Terol (2009). Su propuesta tiene un interés destacable, ya que abordan tres de las cuestiones prototipo que con mayor frecuencia se producen en la atención farmacoterapéutica: i) '*What is the drug of choice for condition x?*', ii) '*What is the dose of drug x?*', y iii) '*Can drug x cause (adverse) finding y?*'. Su sistema tiene una estructura modular para abordar las tareas de generación de patrones, clasificación de preguntas y análisis de éstas. El conocimiento externo utilizado es el Metathesaurus de UMLS y WordNet. Aunque, su trabajo está centrado en lengua inglesa y, como ellos mismos indican, se precisa mejorar los índices de precisión (84-94%), sus resultados tienen un interés aplicativo frente a las dificultades que manifiestan los profesionales para el manejo de los sistemas de recuperación de información de las bases de datos (Ely et al, 2002).

Desde la perspectiva de los profesionales de la salud, y en un sentido aplicativo, es interesante mencionar la línea del equipo de Carrero (2007) para el acceso a información *crosslingüe*, en la cual proponen un sistema que conecta la documentación clínica en castellano²⁴ con información científica en inglés y en castellano. Los recursos léxico-semánticos con los que se apoya su sistema son SNOMED-CT, MeSH y UMLS. Sin embargo, no mencionan ningún problema con el reconocimiento de las entidades farmacológicas, la concentración y la forma farmacológica²⁵. Esto podría

²⁴ Tienen acceso a una colección de 9.413 notas de evaluación y 49 informes de alta des-identificados.

²⁵ Se deben prever algunas incompatibilidades o vacíos de información respecto a las entidades asociadas al concepto 'Compuesto Terapéutico Virtual' que incluye las 2 versiones de SNOMED-CT (inglés y castellano). Esta suposición se apoya en las características propias de la información farmacológica, que se han comentado en este capítulo, y en la existencia de 2 extensiones de SNOMED-CT para los productos medicinales comercializados en EEUU y en el Reino Unido [*The International Health Terminology Standards Development Organisation* (2008, Julio). *SNOMED Clinical Terms. User Guide*].

explicarse, al menos en parte, debido a que su contexto de trabajo es el ámbito hospitalario, donde no se suele utilizar, en la prescripción, el nombre de la especialidad farmacológica durante la hospitalización pero sí en las consultas externas para la atención ambulatoria de los pacientes.

Respecto al ámbito sanitario, en general, encontramos varias propuestas. El trabajo de Maña y su equipo (2006), para el acceso a información bilingüe, en el que se relacionan la documentación clínica y científica, con técnicas de categorización y generación automática de resúmenes y en el que emplean las fuentes de conocimiento MeSH y UMLS.

El trabajo de Perea y su equipo (2008), para la categorización de texto multietiqueta, los cuales utilizan un corpus de informes de radiología etiquetado con CIE-9-CM en lengua inglesa. En este estudio, los autores indican que la integración de conocimiento de UMLS mejora el sistema de expansión para la categorización²⁶.

Por último, habría que mencionar el trabajo de Iglesias y equipo (2008) relativo al sistema MOSTAS para el pre-procesamiento de los textos asistenciales. MOSTAS contiene cuatro módulos: analizador morfo-semántico, buscador de términos médicos, anonimizador y corrector ortográfico. Sus fuentes de conocimiento son un diccionario general español, un diccionario de siglas, abreviaturas y acrónimos biomédicos y un servidor terminológico específico, del que, únicamente, destacan a SNOMED-CT. Aunque este trabajo es de suma importancia, carece de resultados sobre su efectividad práctica.

²⁶ En un trabajo previo (Martín et al, 2007) integraba sólo el conocimiento de MeSH.

3.5. Conclusiones y motivación.

De todo lo anteriormente expuesto, y teniendo en cuenta que la información relevante del SI farmacoterapéutico es toda aquella que facilita información sobre el medicamento más idóneo, de acuerdo a un problema de salud y las alternativas para su uso, respecto a las características individuales de cada sujeto, se puede concluir que, en el contexto de la atención farmacoterapéutica, cuenta con las siguientes limitaciones:

1. No todas las especialidades farmacéuticas están disponibles en todos los países y, cuando lo están, no necesariamente se denominan igual. Incluso la denominación oficial del principio activo puede cambiar de un país a otro.
2. La información sobre la atención sanitaria y farmacoterapéutica se crea en la lengua nativa. Pero, simultáneamente, una gran parte de la información externa, que se precisa para la atención, se encuentra en lengua inglesa.
3. Las necesidades de información varían en función del rol que adoptan los profesionales (médico, farmacéutico y enfermera) durante el proceso de la atención farmacoterapéutica: prescriptor, dispensador, administrador y evaluador de los resultados (Herrera, 2003).
4. En el escenario farmacoterapéutico existe información fragmentada, no integrada e imposible de compartir en la actualidad por los SI. Todavía se mantienen abundantes SI independientes en el sistema de

atención sanitaria con muy reducida capacidad para interactuar entre ellos e intercambiar la información automáticamente. Sin olvidar que son, prácticamente, casi inexistentes los SI que llegan a compartir el significado de la información contenida en ellos (CE, 2008).

5. Los SI de información complementarios, a partir de su contenido informativo, son de dos tipos: los relacionados con los medicamentos comercializados y los que incluyen bibliografía científica. Los términos empleados por cada uno de estos SI, debido a su organización interna, establecen un modelo diferenciado del mismo dominio. Estas estructuras organizativas carecen de una explicitación formal de su significado, lo que dificulta tomar el conocimiento de un SI y utilizarlo en otro para dar respuestas a las necesidades informativas de los usuarios.
6. Los profesionales sanitarios son los que deben conocer el contenido y la forma de acceder a cada uno de los SI complementarios para obtener el contenido informativo que apoye su toma de decisión sobre la terapéutica farmacológica. El problema, que sigue siendo un obstáculo en la práctica profesional actual, es el de recuperar la información que se precisa, durante la misma atención y en el menor tiempo (Ely et al, 1999; Dee y Stanley, 2005; Coumou y Meijman, 2006; González-González et al, 2006, 2007).
7. Las prestaciones de los SI, para la prescripción ambulatoria, todavía muestran limitaciones en el enlace entre la información clínica y la del medicamento para la detección de contraindicaciones relativas a la edad, los cambios fisiológicos (por ejemplo, el embarazo) y los problemas de salud colaterales (Fernando et al, 2004). En la

actualidad, aún son nulas las referencias a la estructura y a la semántica de los datos utilizados en estos subsistemas de apoyo a la prescripción o al tipo de conocimiento que emplean para apoyar las alertas de contraindicaciones.

8. En el ámbito internacional, se están utilizando técnicas de PLN, para la reconciliación de la información sobre tratamiento farmacológico dispersa en los módulos que constituyen la HCE con la finalidad de agruparla y clasificarla.
9. La mayoría de los investigadores de PLN nacionales que trabajan en el contexto de la Salud, hasta el momento, focalizan su trabajo en documentos bien caracterizados, como los artículos o los resúmenes obtenidos, casi exclusivamente, de la base de datos Medline y con predominio del inglés.
10. Hay una escasez de recursos abiertos léxico-semánticos en lengua castellana para el dominio de la salud y del medicamento. Tampoco existen corpus abiertos de documentos asistenciales, des-identificados, en castellano. Esta última limitación dificulta la realización de chequeos de los métodos y de las aplicaciones sustentadas en PLN, orientadas a resolver los problemas asistenciales en general o específicos de la terapéutica farmacológica.

Capítulo IV.

OntoFIS: Diseño y Construcción de una Ontología Farmacoterapéutica.

Universitat d'Alacant
Universidad de Alicante

A partir de las limitaciones y de las carencias detectadas, surge la posibilidad de generar aplicaciones informáticas que puedan cubrir las necesidades de los actores del escenario farmacoterapéutico (médicos, farmacéuticos y enfermeras). En dichas aplicaciones, y con el menor número de interacciones por parte del usuario, se debería de poder solicitar información del tipo: “¿la atorvastatina ocasiona un incremento de glucemia?” o “si el paciente toma antidiabético oral y ahora hay que tratarlo de dolor neuropático periférico ¿está indicado un antiepiléptico?”. En el primer caso, la información solicitada se orienta al seguimiento terapéutico mientras que, en el segundo, se busca información para asegurar la instauración adecuada de un nuevo tratamiento farmacológico que está condicionada por un tratamiento previo. Las

aplicaciones informáticas, creadas para dar respuesta a estas peticiones de información, deberían de contar con una fuente de conocimiento que permitiera la identificación de las abstracciones incluidas en las distintas cuestiones.

Así, por ejemplo, en la cuestión de seguimiento terapéutico, la fuente de conocimiento debería identificar la abstracción ‘*atorvastatina*’ como una entidad del concepto *principio activo*, mientras que la abstracción ‘*incremento de glucemia*’ debería ser identificada como sinónimo de ‘*hiperglucemia*’ y, a su vez, por el verbo enunciado ‘*ocasiona*’ debería permitir deducir que esta abstracción es una entidad del concepto *reacción adversa* y distinguirla de la existente en el concepto *hallazgo clínico*. Las entidades del concepto *hallazgo clínico* serían las que justificarían el uso terapéutico de un fármaco mientras que, las entidades del concepto *reacción adversa* serían las que determinarían la suspensión o el cambio del medicamento (figura 4.1).

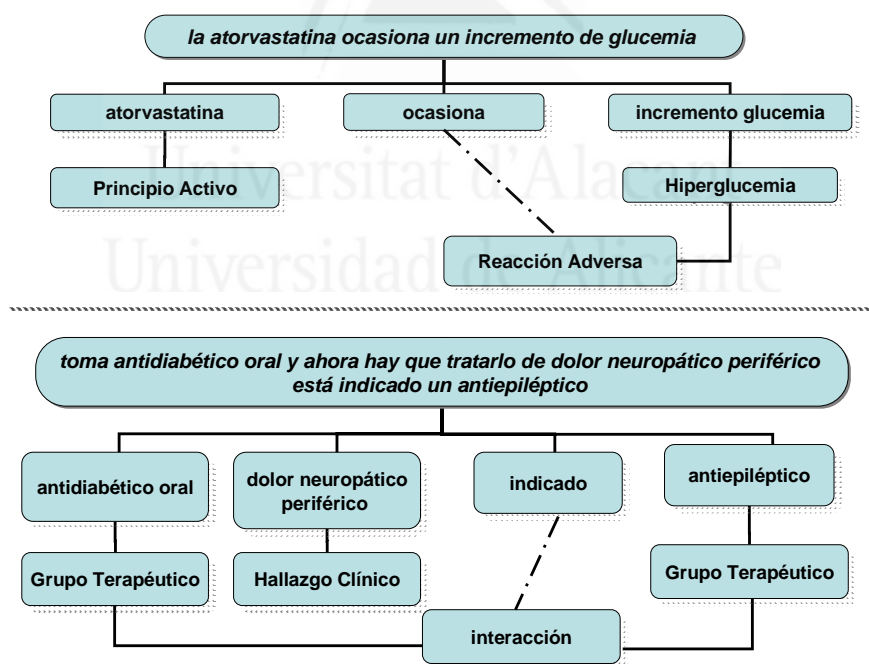


Figura 4.1. Identificación de las entidades en dos cuestiones de petición de información.

Respecto a la cuestión planteada sobre la instauración de nuevo tratamiento, se deberían identificar, además, otras abstracciones en la fuente de conocimiento. Estas entidades serían ‘*antidiabéticos orales*’ y ‘*antiepiléticos*’ relativas al concepto *grupo terapéutico*²⁷ y ‘*dolor neuropático periférico*’ como perteneciente al concepto *hallazgo clínico*. En esta cuestión la fuente de conocimiento debería, en primer lugar, confirmar que el hallazgo clínico (*dolor neuropático periférico*) podría ser tratado con el segundo grupo de fármacos (*antiepiléticos*) y, si la respuesta es afirmativa, debería seguir con la verificación de que no hay entre ambos grupos de fármacos (*antidiabéticos orales* y *antiepiléticos*) ninguna incompatibilidad o *interacción* (figura 4.1).

Para dar respuesta a estas demandas se precisa de una fuente de conocimiento que represente la información farmacoterapéutica con la finalidad de concretar el significado de las abstracciones y lograr las inferencias necesarias en este dominio; en otras palabras, se requiere una ontología. La ontología debería caracterizarse por: i) ser independiente, es decir, por ser una aplicación que, por ella misma, contuviera la información conceptualizada y relacionada entre sí, organizada de forma consistente y no redundante del dominio farmacoterapéutico; ii) ser interpretable o entendida, tanto por otras aplicaciones informáticas como por humanos; y, iii) ser reutilizable, para poder ser adaptada a condiciones de mayor especificidad o bien extendida por la incorporación de nuevo conocimiento. De acuerdo con estas directrices se ha determinado el procedimiento de diseño y de construcción de la Ontología Farmacoterapéutica e Información para el Seguimiento, OntoFIS.

Este capítulo se ha organizado en cinco apartados. En el primero de ellos se realiza un análisis del modelo representado en las ontologías existentes para determinar la cobertura conceptual que proporcionan. También, se describen las metodologías de construcción de ontologías y su correspondencia con un esquema conceptual. En el segundo apartado se presenta el modelo de datos como propuesta metodológica para el modelado de la ontología. El siguiente

²⁷ Agrupación de fármacos con una finalidad común.

apartado se centra en la propuesta tecnológica empleada para el diseño y la construcción de OntoFIS. El cuarto apartado se dedica a la descripción del proceso de captura del conocimiento del dominio farmacoterapéutico, el modelado conceptual de OntoFIS en UML y su formalización en OWL DL, de acuerdo con la propuesta metodológica y tecnológica establecida. En el quinto y último apartado, se reseñan las conclusiones obtenidas.

4.1 Análisis del modelo representado en las ontologías existentes.

La selección de una ontología para ser reutilizada, completa o parcialmente, se basa en la idoneidad de ésta para satisfacer necesidades del SI. A partir de la delimitación de las necesidades del SI para la atención farmacoterapéutica²⁸, el proceso de selección se debe iniciar garantizando la concordancia conceptual entre las demandas del SI y la cobertura de universo del discurso de la ontología.

La valoración de la conceptualización de una ontología se debe orientar hacia el tipo y la amplitud de la cobertura conceptual, relacionada tanto con el propósito del proceso terapéutico como con el propio medicamento. En el primer caso, se debe comprobar la representación de conceptos involucrados con los problemas de salud que pueden prevenirse, resolverse o paliarse. Estos se encuentran relacionados con los conceptos que representan a los *'diagnósticos'*, *'signos/síntomas'*, *'agentes etiológicos'* y *'efectos adversos'*. En el segundo caso, se debe de comprobar el conjunto de conceptos empleados para representar al medicamento, es decir, la existencia de conceptos relacionados con la *'denominación de la especialidad'*, *'nombre registrado del medicamento'*, *'principio activo'*, *'clasificación terapéutica'*, *'dosis'*, *'forma*

²⁸ Ver el apartado de conclusiones del capítulo anterior, pp 75-77.

farmacéutica, *‘vía de administración’* y *‘toxicidad’*. El análisis realizado se incluye en las tablas 4.1 y 4.2.

Cobertura Conceptual	CIE-9-MC (1)	MeSH (2)	DeCS (3)	UMLS (4)
Diagnósticos	X	X	X	X
Signos/síntomas	/	X	X	X
Agentes etiológicos	/	X	X	X
Denominación especialidad(*)	0	0	0	0
Nombre Registrado	0	/	/	/
Principio activo	/	X	X	X
Clasificación terapéutica	/	X	X	X
Dosis	0	/	/	0
Forma farmacéutica	0	X	X	X
Vía	/	X	X	X
Efectos adversos	/	/	/	X
Toxicidad	/	/	/	X
<p>Nota Aclaratoria: Las características generales, así como, la estructura organizativa y el grado de formalización de estas fuentes de conocimiento se describieron en el capítulo II.</p> <p>(*) La denominación de la especialidad es la utilizada para la prescripción ambulatoria. Se componen del nombre registrado, la cantidad y la forma farmacéutica. CARDYL 10 mg Comp. recub. es una especialidad diferente de CARDYL 20 mg Comp. recub.</p> <p>Legenda: (1)= Español; (2)= Inglés; (3)= Trilingüe; (4)= Multilingüe respecto al Metathesaurus (ver tabla 4.2.); X= amplia cobertura; /= cobertura parcial y 0= nula presencia.</p>				

Tabla 4.1. Cobertura conceptual farmacoterapéutica en las ontologías de salud.

Las conclusiones indican que el modelo, explicitado por las ontologías existentes, tiene una cobertura exhaustiva sobre los problemas de salud, pero una infrarepresentación de los conceptos que caracterizan el conocimiento sobre el medicamento. Destaca el vacío sobre la denominación de la especialidades farmacéuticas y la escasa representación de los nombres comerciales, necesarios en la atención farmacoterapéutica ambulatoria. Asimismo, se pone de manifiesto la reducida disponibilidad de fuentes de conocimiento léxicas y semánticas del dominio, con información en lengua española y en otras lenguas de la UE, a excepción del inglés.

Características de UMLS (2008AA)	Number of sources (distinct source families by language): 123 Number of sources contributing concept names: 145 Number of languages contributing concept names: 17																		
Representación de los idiomas en las fuentes y porcentaje que representan en el Metathesaurus	<table> <tr> <td>ENG → 96 → 66.21%</td><td>SWE → 2 → 1.38%</td></tr> <tr> <td>DUT → 6 → 4.14%</td><td>BAQ → 1 → 0.69%</td></tr> <tr> <td>GER → 6 → 4.14%</td><td>CZE → 1 → 0.69%</td></tr> <tr> <td>SPA → 6 → 4.14%</td><td>DAN → 1 → 0.69%</td></tr> <tr> <td>FRE → 5 → 3.45%</td><td>HEB → 1 → 0.69%</td></tr> <tr> <td>ITA → 4 → 2.76%</td><td>HUN → 1 → 0.69%</td></tr> <tr> <td>POR → 4 → 2.76%</td><td>NOR → 1 → 0.69%</td></tr> <tr> <td>FIN → 2 → 1.38%</td><td>RUS → 1 → 0.69%</td></tr> <tr> <td>JPN → 2 → 1.38%</td><td></td></tr> </table> <p>Fuentes que contribuyen a las relaciones y no están asociadas a ningún idioma → 5 → 3.45%</p>	ENG → 96 → 66.21%	SWE → 2 → 1.38%	DUT → 6 → 4.14%	BAQ → 1 → 0.69%	GER → 6 → 4.14%	CZE → 1 → 0.69%	SPA → 6 → 4.14%	DAN → 1 → 0.69%	FRE → 5 → 3.45%	HEB → 1 → 0.69%	ITA → 4 → 2.76%	HUN → 1 → 0.69%	POR → 4 → 2.76%	NOR → 1 → 0.69%	FIN → 2 → 1.38%	RUS → 1 → 0.69%	JPN → 2 → 1.38%	
ENG → 96 → 66.21%	SWE → 2 → 1.38%																		
DUT → 6 → 4.14%	BAQ → 1 → 0.69%																		
GER → 6 → 4.14%	CZE → 1 → 0.69%																		
SPA → 6 → 4.14%	DAN → 1 → 0.69%																		
FRE → 5 → 3.45%	HEB → 1 → 0.69%																		
ITA → 4 → 2.76%	HUN → 1 → 0.69%																		
POR → 4 → 2.76%	NOR → 1 → 0.69%																		
FIN → 2 → 1.38%	RUS → 1 → 0.69%																		
JPN → 2 → 1.38%																			
Fuentes en castellano y su aportación al Metathesaurus.	<ol style="list-style-type: none"> 1. Physicians' Current Procedural Terminology, Spanish Translation, 2001: Number of Strings: 7348 2. The International Classification of Primary Care (ICPC), Spanish Translation, 1993: Number of Strings: 723 3. Medical Dictionary for Regulatory Activities Terminology (MedDRA), Spanish Edition, 10.1 (Septiembre 2007): Number of Strings: 96161 4. Descriptores en Ciencias de la Salud (Spanish translation of the Medical Subject Headings), 2008: Number of Strings: 72293 5. SNOMED Terminos Clinicos (SNOMED CT), Edición en Español, Distribución Internacional, Octubre de 2007, 2007_10_31: Number of Strings: 1399646 6. WHO Adverse Drug Reaction Terminology (WHOART), Spanish Translation, 1997: Number of Strings: 3104 																		

Tabla 4.2. Fuentes de conocimiento en español integradas en el Metathesaurus® de UMLS (2008AA)
 [Disponible en http://www.nlm.nih.gov/research/umls/metab4.html#sb4_0 (recuperado 28/10/2008)].

Consecuentemente, para cubrir las necesidades informativas del SI para la atención farmacoterapéutica, se precisa desarrollar una ontología desde cero, aplicando técnicas manuales. Esta decisión se justifica, también, por la necesidad de asegurar la fiabilidad del razonamiento que se logre con la ontología. La validación de las inferencias implica un control exhaustivo de los resultados obtenidos en la recuperación de información. No hay que olvidar que los sistemas de información, derivados o apoyados por la ontología, serán utilizados para la gestión de problemas de salud y un error causado por el sistema puede resultar crítico. La estimación, realizada a partir del análisis del modelado de las ontologías existentes, indicaba que el esfuerzo y la dedicación

de tiempo serían equiparables empleando métodos semisupervisados o manuales para la creación de una ontología en el ámbito de la salud²⁹.

4.1.1. Metodologías de construcción de ontologías.

Desde el área de la ingeniería del conocimiento, se proponen varias metodologías para la construcción de nuevas ontologías sin reutilización de recursos. Las metodologías más significativas, por su repercusión en la literatura son: la ‘Metodología de elaboración Cyc’ (Lenat y Guha, 1990), la metodología de ‘Construcción de Ontologías de Uschold y King’ (1995), el método de ‘Construcción de Ontologías de Grüninger y Fox’ (1995), el procedimiento ‘Methontology’ (Fernández-López et al., 1999) y la guía de desarrollo de ontologías para sistemas declarativos basados en marcos, conocida como ‘Ontology Development 101’ (Noy y McGuinness, 2001).

Cuando se examina el conjunto de indicaciones y de tareas de cada una de estas metodologías (tabla 4.3), se detectan diferencias relativas a las denominaciones utilizadas, al número de tareas establecidas, al análisis que se propone en cada una de las etapas o en las tareas propuestas y asociadas con el uso de una herramienta diseñada ‘*ad-hoc*’.

No obstante, cuando se realiza un análisis más profundo de las indicaciones, se comprueba que hay coincidencias entre las actividades y las tareas propuestas. Las coincidencias se sintetizan en las garantías que debe proporcionar la metodología para realizar la representación del conocimiento aceptado y compartido. Los objetivos comunes son: i) realizar una captura sistemática del conocimiento aceptado del dominio con una formalización informal; ii) describir el conocimiento aceptado de acuerdo con una técnica de

²⁹ En Bodenreider, Rindflesch y Burgun (2002) se puede encontrar la valoración del funcionamiento y el rendimiento actual de los métodos no supervisados o semisupervisados para extender las terminologías en salud.

modelado conceptual; y iii) formalizar la descripción en un lenguaje de representación de conocimiento.

Metodología Cyc	Construcción de Ontologías de Uschold y King	Construcción de Ontologías de Grüninger y Fox	Methontology	Ontology Development 101
Extracción manual de conocimiento desde diferentes fuentes de información.	Identificación del propósito y alcance con los términos relevantes del dominio.	Identificación del escenario relacionándolo con las aplicaciones que usarán la ontología.	Especificación o delimitación de los objetivos de su creación y decidir el dominio de actuación de la Ontología.	Determinar el dominio y el alcance de la ontología.
Adquisición de nuevo conocimiento ayudándose de herramientas PLN y aprendizaje automático.	Capturar la ontología o la identificación de los conceptos y de las relaciones; con definiciones añadidas, que los relacionan con otros términos.	Establecer el conjunto de cuestiones de competencia o requerimientos específicos, en lenguaje natural, para determinar las opciones de respuesta.	Conceptualización o creación un glosario de términos que pertenecen al dominio, definirlos y crear una taxonomía.	Enumerar los términos importantes para la ontología.
Delegación de la gestión de extracción en las herramientas.	Selección de los términos representativos de los conceptos y escribir el código.	Extraer de las cuestiones de competencia los conceptos, sus propiedades, sus relaciones y axiomas.	Formalización o conversión del modelo anterior, en un modelo formal o semi-computable.	Definir las clases y su jerarquía, posteriormente, verificar su adecuación
Desarrollo de la representación y de la ontología de nivel superior.	Evaluación y documentación de la ontología.	Especificación formal en lógica de primer orden.	Implementación o conversión del modelo formalizado en un modelo computable con un lenguaje para construcción de ontologías.	Definir las propiedades de las clases y comprobar su funcionalidad.
		Definición de las condiciones para considerar las respuestas como completas	Mantenimiento o bien borrado, modificaciones o incorporaciones nuevas.	Definir las restricciones de los valores de las propiedades.
				Crear las instancias.
				Verificar si funciona adecuadamente

Tabla 4.3. Etapas fundamentales de las metodologías para la construcción de ontologías.

Estos objetivos comunes establecen un método secuencial para la construcción de una ontología y, por tanto, se pueden reinterpretar como las etapas propuestas para el diseño y la construcción de esta tecnología. Estas etapas, por otra parte, mantienen una correspondencia con el diseño de un sistema de información. La relación es más visible entre las primeras etapas propuestas y el proceso de obtención del ‘Esquema Conceptual’, en lo que atañe a las base de datos. Asimismo, hay una alta semejanza con el ciclo de vida clásico del software y, principalmente, con la fase de análisis (figura 4.2).

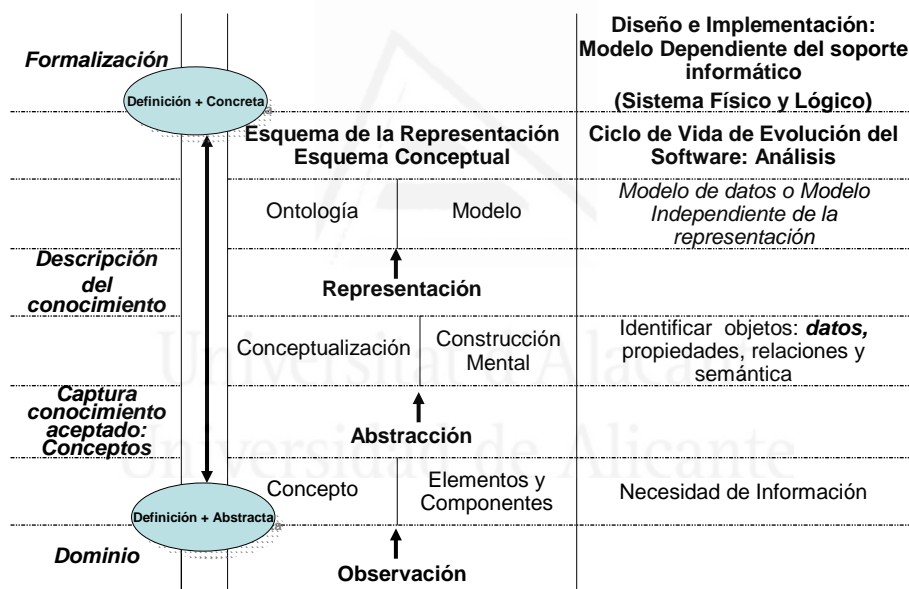


Figura 4.2. Correspondencia entre las etapas de construcción de una ontología, de un esquema conceptual y el ciclo de vida de evolución del software. (Adaptación a partir de Sánchez, Cavero y Marcos, 2005).

No hay que olvidar que un modelo es la descripción de una representación mental que contiene las características del conjunto de objetos y de sus relaciones, cuyo propósito es poder analizar, entender y expresar de forma más simplificada la realidad del mundo. Las ontologías son un tipo especializado de modelo que se concibe a partir de idénticos mecanismos mentales, pero con una

finalidad diferente, la de representar el conocimiento aceptado y compartido en el dominio y no, simplemente, una representación por consenso.

En ambos casos, para definir la representación de los conceptos relevantes y distintivos, se debe emplear un formalismo que evite el carácter subjetivo y la ambigüedad del lenguaje natural (Sánchez, Caverro y Marcos, 2005), aspecto que se trata en el siguiente apartado.

4.2. Propuesta metodológica para el modelado de ontologías.

El formalismo que permite estructurar los datos³⁰, de forma que se capte la semántica de los mismos, es el denominado ‘Modelo de Datos’. El *modelo de datos* es una colección de herramientas conceptuales que ofrecen un conjunto de conceptos y de reglas de composición para combinarlos. Estas herramientas conceptuales son capaces de representar, con mayor o menor fidelidad, el modelo de la abstracción de la realidad (el conjunto de datos interrelacionados del dominio y las restricciones que deberán cumplir en todo momento). El modelo de datos es, en definitiva, un lenguaje de tipo gráfico o lingüístico que puede utilizarse independientemente de que la representación se formalice o no en un ordenador.

El proceso de aplicar un *modelo de datos* se conoce por modelado y su resultado es *el modelo de datos conceptual*. En esencia el modelo de datos conceptual es un ‘*esquema de la representación*’ o un ‘*esquema conceptual*’; denominación que se emplea, respectivamente, para referirse a una representación de conocimiento o una representación de las propiedades de un sistema de información o de una base de datos (figura 4.2).

³⁰ Se utiliza la denominación general pero, en el caso de las ontologías, se refiere a los conceptos relevantes y distintivos del dominio.

Los mecanismos de abstracción generales que utiliza el *modelo de datos* para la construcción de las clases de objetos apropiadas son:

- La *clasificación*, para la definición de un concepto como una clase de objeto.
- La *agregación*, para la definición de una nueva clase de objetos a partir de otras clases.
- La *generalización*, para la definición de los subtipos de una clase.

Los tipos de restricciones aplicadas a las clases de objetos son:

- *Dominio*: definen el conjunto de valores agrupados en una clase de objeto.
- *Identificación*: no hay dos objetos dentro de una clase de objetos iguales.
- *Correspondencia entre clases*: por medio de restricciones de tipo cardinalidad, dependencia, dependencia de identificador, existencia y cobertura de las generalizaciones.

El modelo de datos sólo podrá considerarse adecuado para la construcción de una ontología si, con él, se realiza un modelo del conocimiento aceptado y compartido por una comunidad o un dominio (Gómez-Pérez et al, 2004). Existen varias alternativas para el modelado de los componentes del conocimiento, desde los formalismos empleados por la comunidad de Inteligencia Artificial como frames, descripción lógica y lógica de primer orden hasta los formalismo tipo UML (Lenguaje Unificado de Modelado) y E/R (Entidad/Relación) utilizados por la Ingeniería del Software y de Bases de Datos.

En los siguientes subapartados se describen las herramientas conceptuales de los formalismos UML y de los lenguajes de representación basados en marcas para la Web y, en concreto, de OWL DL, empleados en la propuesta metodológica para el modelado de la ontología OntoFIS.

4.2.1. UML (Lenguaje Unificado de Modelado).

El Lenguaje Unificado de Modelado (UML) es un lenguaje que proporciona un vocabulario y unas reglas que permiten la representación gráfica, del modelo de datos conceptual, de forma independiente al lenguaje de implementación. Los objetivos de UML se pueden sintetizar en sus capacidades para visualizar, especificar, construir y documentar:

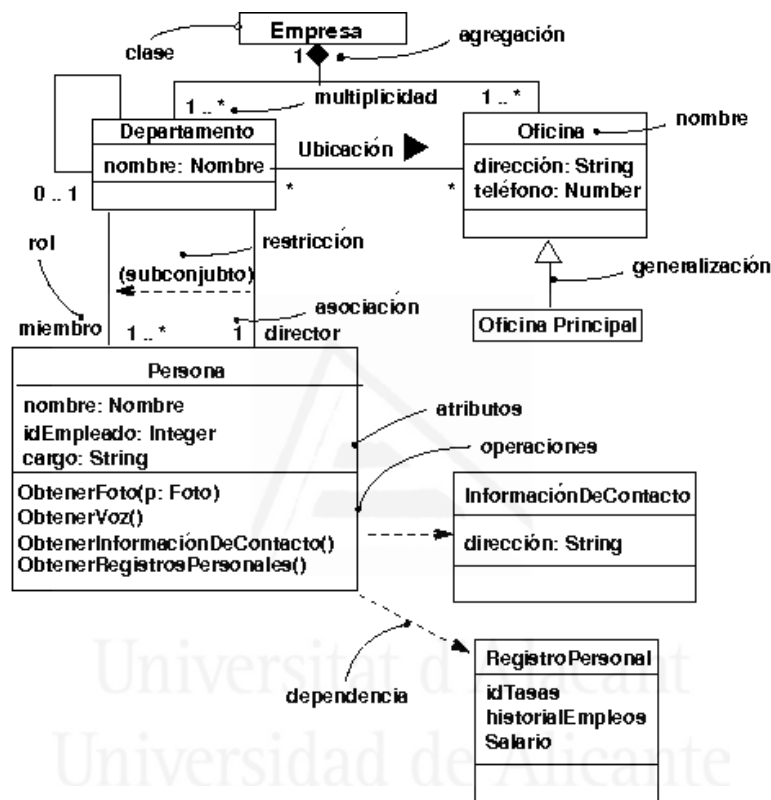
- UML permite expresar, de un forma gráfica, una representación de forma que otro usuario no participante lo pueda entender.
- UML permite especificar cuáles son las características del modelo de datos conceptual antes de su construcción.
- A partir del modelo especificado se pueden construir los sistemas diseñados.
- Los propios elementos gráficos sirven como documentación del sistema desarrollado.

4.2.1.1. Herramientas conceptuales de UML.

Un modelo UML está definido por tres tipos de bloques de construcción o anotaciones gráficas: i) los elementos para representar las abstracciones de la realidad, que se denominan clases; ii) las relaciones factibles entre los elementos; y, iii) los diagramas, que son la representación gráfica de un conjunto de elementos con su relaciones.

En el formalismo de modelado UML, un concepto es una versión simple de una clase. Una clase muestra distintiva y gráficamente: i) la denominación asignada para representar a la clase de forma precisa y no ambigua, para evitar la confusión con otra clase; ii) los atributos o propiedades identificadas, junto con su tipo de dato, longitud, valor inicial y visibilidad; y, iii) por último, los métodos que se definen en la clase con sus argumentos y el mecanismo mediante el cual otras clases o elementos del modelo se comunican con ella.

Aunque, esta zona distintiva no se utiliza para representar ontologías, ya que es la parte lógica asociada al objeto.



Nota Aclaratoria: Ejemplo tomado de <http://generacionesweb.com/paginas/6/7/uml-elementos.jsp>

Figura 4.3. Colección de elementos del diagrama de clases UML para el modelado declarativo: clases, tipos y sus contenidos y relaciones.

Las relaciones en UML permiten especificaciones del tipo ‘*asociación*’, ‘*agregación*’ y ‘*generalización*’. La ‘*asociación*’ es una relación de tipo estructural que modela una conexión binaria con significado único, determinada por el rol y la cardinalidad que se le asigna. La ‘*agregación*’, que es conocida también como una relación de composición, es una relación de tipo ‘*todo-parte*’ entre un agregado o clase que representa el todo y las clases que son sus partes constituyentes. La ‘*generalización*’ es una relación de tipo ‘*genérico-*

específico'. En otras palabras, es una relación de herencia donde la clase más específica es, completamente, consistente con la clase más general pero, contiene información adicional que la diferencia y la especifica.

El formalismo de UML ofrece una amplia variedad de diagramas para visualizar las distintas perspectivas o modelos de un sistema. El diagrama más empleado para el modelado conceptual es el 'diagrama de clases'. Éste muestra una colección de elementos de modelado declarativo (estático), tales como las clases, los tipos y sus contenidos, y las relaciones (figura 4.3).

Algunas de las razones que se citan para su uso son: i) la sencillez de su representación, ii) su estandarización para la representación gráfica de los modelos, iii) la disposición de múltiples herramientas CASE que apoyan la representación gráfica y iv) su versatilidad para hacer frente a diversos problemas de diseño, en los que se ha aplicado UML, para representar ontologías (Craniffeld y Purvis, 1999, 2000).

El diagrama de clase UML permite establecer el modelo que fundamentará el trabajo posterior de formalización y de razonamiento de la ontología (Kang et al, 2004). En el 'diagrama de clases' se establece, gráficamente, la declaración explícita del conocimiento del dominio (ontología), se especifican las propiedades de cada concepto, con las restricciones de sus valores y se organizan los conceptos a partir de las relaciones de tipo jerárquico (generalización o *'isa'* y *agregación* o *'part_of'*) y asociativo (*rol* o *relaciones semánticas del dominio*). El formalismo UML permite modelar una ontología ligera, especificando restricciones tanto en las propiedades identificadas de los conceptos como en las relaciones, aunque sin llegar a modelar una ontología pesada. Esto no significa que esta técnica de modelado conceptual no tenga limitaciones. Las principales desventajas son que: i) no hay diferencia entre clase y atributos de instancia, ii) no pueden ser representadas, directamente, relaciones de alta aridad y, iii) no hay un lenguaje estándar en todas las herramientas CASE para especificar las restricciones de objetos y formalizar los

axiomas. Aunque, parece ser más un problema de las herramientas a utilizar que del propio modelo que sustenta a UML.

4.2.2. Lenguajes de marcado Web.

Desde hace unos años, con el auge de la Web Semántica, los lenguajes desarrollados por el W3C (*World Wide Web Consortium*) han comenzado a tener un mayor impacto. Entre ellos, el OWL (*Web Ontology Language*³¹) que comienza a constituirse en un estándar para la representación de ontologías. Este lenguaje complementa y amplía a RDF (*Resource Description Framework*) y a los esquemas RDF (RDFS). En otras palabras, las posibilidades de expresividad, en relación a la semántica, se han incrementado desde los lenguajes de metadatos RDF y RDFS hasta llegar al lenguaje formal de ontologías OWL. A continuación, se describe el modelo de datos de cada uno de ellos para obtener una visión de la dependencia existente en su vocabulario y en las reglas de composición y combinación.

4.2.2.1. El lenguaje de metadatos RDF y RDFS.

Más que un lenguaje, RDF es un modelo de metadatos para describir recursos disponibles en Internet o de cualquier otro tipo. En este modelo la descripción se realiza en forma de tripletas del tipo: *sujeto-predicado-objeto*. Cada triplete corresponde con una sentencia RDF, donde el sujeto es el recurso a describir (a qué se refiere la sentencia), el predicado es una propiedad (identifica la característica del sujeto) y el objeto es el valor asignado a la propiedad o al recurso con el que se establece la relación. Cada constituyente de la triplete son recursos que no se corresponden, exclusivamente, con recursos presentes en la Web, sino con cualquier cosa con identidad. Los recursos, para

³¹ <http://www.w3.org/TR/owl-features/>

ser identificados, tienen un identificador URI (*Uniform Resource Identifier*); aunque, el URI no tiene significado por sí mismo y debe ser asignado por un humano.

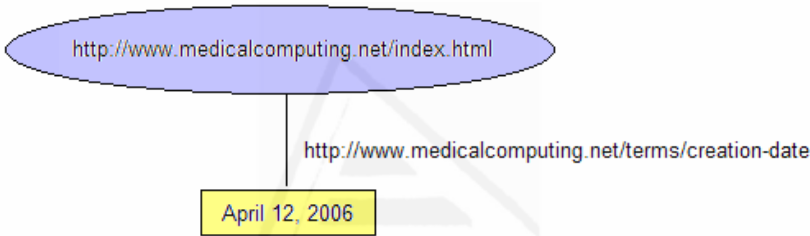
Un ejemplo de enunciado: <i>http://www.medicalcomputing.net/index.html has a creation-date whose value is April 12, 2006</i>
Componentes de enunciado para formar una tripleta en RDF: <ul style="list-style-type: none"> • El sujeto es el documento con URL (<i>Uniform Resource Locator</i>): 'http://www.medicalcomputing.net/index.html' • El predicado es la palabra: 'creation-date' • El objeto es valor literal 'April 12, 2006'
Representación gráfica de la tripleta: 
Representación en XML: <pre><?xml version="1.0"?> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:medcomterms="http://www.medicalcomputing.net/terms/"> <rdf:Description rdf:about="http://www.medicalcomputing.net/index.html"> <medcomterms:creation-date>April 12, 2006</medcomterms:creation-date> </rdf:Description> </rdf:RDF></pre>
Nota Aclaratoria: El ejemplo está tomado de Amies, A. (April, 2006). <i>Introduction to OWL Web Ontology Language for Medical and Biosciences Applications</i> . Disponible en: http://www.medicalcomputing.net/owl_printable.html .

Tabla 4.4. Las tres variantes de representación de los recursos y las relaciones en RDF.

La representación de los recursos y las relaciones entre los mismos, en RDF, puede expresarse por la tripleta (sujeto-predicado-objeto) y, además, por una representación gráfica (sujeto y objeto representados como nodos y unidos por una arco que es el predicado) o bien por una semántica muy sencilla codificada en XML³² (RDF/XML), como se muestra en la Tabla 4.4.

³² Uno de los lenguajes más extendidos es XML (*eXtensible Markup Language*). Éste es un metalenguaje para la definición de lenguajes que permiten el intercambio de datos estructurados, correctamente, con independencia de la plataforma en la que se ejecuten las aplicaciones. Se diseñó pensando en la estructuración, intercambio y publicación de

Sin embargo, una aplicación que extraiga datos RDF, se encuentra con el problema de no poder diferenciar si las propiedades asignadas a dos recursos son equivalentes o diferentes y, en este último caso, no podrá comprender (inferir) en qué son diferentes. Análogamente, si varios recursos son asignados a la misma URI podrá saber que corresponden con el mismo significado, pero sin comprender lo que significa dicho significado.

El vocabulario o la terminología utilizada, para ser asignada a los recursos RDF, puede ser definida en los RDFS. La terminología representada se define en función de las sentencias RDF que se deseen expresar. Esta versatilidad es la que le facilita especificar una terminología relativa a un posible dominio, con la especificación de la denominación de las etiquetas para los recursos, el orden y la combinación de éstas.

Además, RDFS incluye las propiedades y sus valores, así como las entidades en las que se pueden aplicar todas o parte de las propiedades incluidas. Esta declaración permite comprobar si las propiedades y los valores asignados a los recursos que componen un conjunto de tripletas resultan válidos respecto al propio esquema definido.

La aportación más significativa de RDFS es que especifica qué interpretación hay que dar a las sentencias de un modelo de datos RDF (significado), pero deja libre la representación sintáctica del modelo. El vocabulario que utiliza RDFS³³, para describir los metadatos que modelan el conocimiento, es de dos tipos: el vocabulario relativo a las clases (*Resource*, *Class*, *Literal*, *Datatype*, *Container*, *ContainerMembershipProperty*) y el

información con independencia del dispositivo, plataforma o aplicación de tratamiento. Facilita la interoperabilidad sintáctica, es decir, la capacidad del SI para leer los datos procedentes de otros SI y obtener una representación que pueden usar, pero no proporciona ningún mecanismo para garantizar la interoperabilidad semántica. XML no impone ninguna interpretación común de los datos del documento y, en él, no se pueden reconocer los objetos/conceptos y las relaciones del dominio de interés. Se complementa con los documentos DTD (*Document Type Definition*), de los cuales se puede extraer la semántica de sus etiquetas o de los comentarios incorporados, pero ello no conlleva que las denominaciones de las etiquetas sean terminologías aceptadas y compartidas del dominio de interés, ni que puedan ser procesadas por los ordenadores entendiendo su significado. En las DTD no hay nada que corresponda al concepto de herencia. Otro complemento pueden ser los esquemas XML, los cuales permiten incluir algunas expresiones semánticas: *X contiene a Y*; *X sigue a Y*; *si X, también Y*; *si X, no Y*. Pero, éstas no son suficientes para representar los conceptos de relación semántica. Estos lenguajes son lenguajes de modelado de datos y no de representación de conocimiento.

³³ Se define en <http://www.w3.org/TR/rdf-schema/>

vocabulario relativo a las propiedades (*domain*, *range*, *subPropertyOf*, *subClassOf*, *member*, *seeAlso*, *isDefinedBy*, *comment* y *label*). En la tabla 4.5 se indican las especificaciones del vocabulario RDFS más, frecuentemente, utilizado.

Vocabulario	Especificación
<code>rdfs:Resource</code>	Las entidades descritas por sentencias RDF son consideradas instancias de esta clase.
<code>rdfs:Class</code>	Esta etiqueta se asigna a cualquier recurso con identidad para indicar que es miembro una clase RDFS. El concepto de clase que se utiliza corresponde con el utilizado en la programación OO. Un recurso identificado en RDF por la propiedad <i>rdf:type</i> que se asocia al valor de una etiqueta <i>rdfs:Class</i> se dice que es una instancia de dicha clase.
<code>rdfs:subClassOf</code>	Es una propiedad que describe una clase como subclase de otra. Solo las instancias de <i>rdfs:Class</i> pueden tener esta propiedad.
<code>rdfs:Property</code>	Es una etiqueta que identifica al conjunto de recursos RDF que actúan como predicado. La jerarquía de propiedades se describe con <i>rdfs:subPropertyOf</i> . El dominio de estos recursos se describe con la etiqueta <i>rdfs:domain</i> y el rango con <i>rdfs:range</i> . Estas dos últimas etiquetas de propiedades de RDFS son instancias asociadas a la etiqueta <i>rdf:Property</i> , la cual se emplea para establecer que un recurso con cierto predicado (propiedad) es instancia de una clase o más.
<code>rdfs:comment</code>	Con esta etiqueta se pueden incluir definiciones o descripciones de la terminología que se declara en RDFS.
<code>rdfs:label</code>	Esta etiqueta es la que facilita la inclusión de la terminología que se asignará a los recursos RDF.
<code>rdfs:seeAlso</code>	Con ella, se especifican recursos que proporcionan información complementaria.
<code>rdfs:isDefinedBy</code>	Esta etiqueta de las propiedades de RDFS se usa para indicar cuál es el recurso donde se define el recurso principal. Es una subpropiedad, <i>rdfs:subPropertyOf</i> , de la propiedad <i>rdfs:seeAlso</i> .

Tabla 4.5. Vocabulario básico de RDFS.

En síntesis, el vocabulario de RDFS proporciona la identificación de las clases, sus propiedades, la organización jerárquica, tanto de las clases como de las propiedades, y las restricciones sobre los dominios y los rangos. RDFS es tan general que puede emplearse en diversos dominios y sirve de puente entre sus terminologías. Aunque, sus limitaciones son que: i) no puede declarar restricciones de rango para su asociación, en exclusiva, a algunas clases; ii) no

se pueden declarar algunas de las características lógicas de las propiedades como la transitividad, la simetría, la propiedad inversa o única; iii) no se puede indicar que un grupo de subclases son disjuntas para establecer que una instancia no puede pertenecer, simultáneamente, a todas ellas; iv) no puede expresar restricciones de cardinalidad o el número de valores que puede tomar una propiedad. En definitiva, RDFS tiene carencias de expresividad para establecer significados precisos y no ambiguos.

4.2.2.2. El lenguaje formal de ontologías OWL.

OWL (*Web Ontology Language*)³⁴ es un lenguaje descriptivo y basado en lógica. Este lenguaje ha sido especificado y estandarizado por W3C (febrero de 2004) y está diseñado para su uso en aplicaciones que requieran realizar funciones de procesamiento del contenido informativo y no limitarse, únicamente, a la presentación de una ontología para ser consultada por parte de los humanos.

OWL es una extensión de RDFS y emplea tripletas de RDF, pero proporciona un vocabulario adicional y una semántica formal que añade más posibilidades para describir las clases y las propiedades, así como las relaciones entre clases, la cardinalidad, las clases enumeradas o las características de las propiedades. OWL facilita una buena relación entre eficacia computacional y poder expresivo. Ofrece tres sublenguajes, con niveles de expresividad crecientes y de validez sintáctica y semántica, adaptados a las necesidades de uso de la ontología a formalizar:

- OWL Lite: está orientado a la elaboración de clasificaciones jerárquicas, tipo taxonomías y thesaurus. Facilita restricciones simples de cardinalidad, pero sólo permite los valores 0 ó 1.
- OWL DL: la denominación de este sublenguaje se debe a su base formal, la lógica de descripción (Description Logics). Se caracteriza por conservar la

³⁴ <http://www.w3.org/TR/owl-features/>

completitud computacional, es decir, porque garantiza que todas las conclusiones son computables y que todos los cálculos son resolubles en un tiempo finito. Este sublenguaje incluye todas las construcciones del lenguaje de OWL, pero sólo puede ser usado bajo ciertas restricciones; por ejemplo, mientras una clase puede ser una subclase de otras clases, una clase no puede ser una instancia de otra.

- OWL Full: este sublenguaje permite la máxima expresividad, por medio de la sintaxis libre de RDF, y posibilita que una ontología incremente el significado de un vocabulario predefinido (RDF). Aunque, es poco probable que un software de razonamiento sea capaz de obtener una inferencia completa para cada característica de OWL Full.

Cada uno de estos sublenguajes es una extensión más compleja de su predecesor, respecto a lo que puede ser expresado, legalmente, y a la validación de sus conclusiones. Su dependencia se mantiene sólo en una dirección:

- Cada ontología legal de OWL Lite es una ontología legal de OWL DL.
- Cada ontología legal de OWL DL es una ontología legal de OWL Full.
- Cada conclusión válida de OWL Lite es una conclusión válida de OWL DL.
- Cada conclusión válida de OWL DL es una conclusión válida de OWL Full.

OWL Full puede ser considerado como una extensión de RDF. Mientras que, OWL Lite y OWL DL pueden ser considerados como extensiones de una visión restringida de RDF. Cada documento OWL (Lite, DL, Full) es un documento RDF, y cada documento RDF es un documento de OWL Full pero, sólo algunos documentos RDF serán, legalmente, documentos OWL Lite u OWL DL.

Indistintamente del sublenguaje, los tres conceptos base de OWL son: clase, propiedad y restricción. A continuación se describe un subconjunto básico del vocabulario de OWL DL³⁵ que, más frecuentemente, se emplea:

- *owl:Class* define un grupo de individuos que pertenecen a la misma clase al compartir algunas propiedades. Representa a una clase de la ontología. Toda clase está identificada por un URI (atributo *rdf:ID*) y todas las clases que se definan, implícitamente, son subclases de la superclase *owl:Thing*. Los tipos de clase que pueden definirse son:
 - Clases aisladas o sin relación con otras clases.
 - Clases organizadas en una jerarquía de especialización, usando la etiqueta *rdfs:subClassOf*.
 - Clases definidas por la combinación booleana arbitraria de clases y de restricciones, por medio de las etiquetas *owl:unionOf* y *owl:intersecciónOf*.
 - Clases definidas por la enumeración de las instancias que contiene la propia clase, con la etiqueta *owl:oneOf*.
 - Clases sin instancias comunes dependientes de una clase, directamente común, que se definen con la etiqueta *owl:disjointWith*. Si se precisa distinguir entre dos instancias se utiliza *owl:differentFrom*.
 - Clases equivalentes o clases con las mismas instancias, se utiliza *owl:equivalentClass* y para indicar que dos instancias son equivalentes, aunque su denominación no coincida, se utiliza la etiqueta *owl:sameAs*.
 - Clases definidas como subclases de otra clase con restricciones en sus propiedades (ver seguidamente propiedades y restricciones).

³⁵ <http://www.w3.org/TR/2004/REC-owl-guide-20040210>

- *owl:DatatypeProperty* es una subclase de la clase de RDF *rdf:Property*. Esta propiedad relaciona una clase (dominio) con un tipo de valor o dato (rango) con las etiquetas *rdfs:domain* y *rdfs:rango*.
- *owl:ObjectProperty*, al igual que la anterior, es una subclase de *rdf:Property*. Esta propiedad tiene como dominio una clase y como rango otra clase y permite que, tanto la clase asignada a dominio como a rango, sea la misma (también emplea las etiquetas *rdfs:domain* y *rdfs:rango*). Su diferencia es que esta propiedad puede ser caracterizada para obtener una mayor precisión en el razonamiento por medio de restricciones.
- Las restricciones, sobre la forma en que las propiedades son utilizadas por las instancias de una clase, son de distintos tipos:
 - Restricciones de las características lógicas de las propiedades:
 - *owl:TransitiveProperty* permite especificar que una propiedad P es transitiva, si para cada instancia Y relacionada con X por $P(X \rightarrow Y)$ y si la instancia Z relacionada con Y por $R(Y \rightarrow Z)$ entonces implica que Z está relacionada con X por $R(X \rightarrow Z)$.
 - *owl:SymmetricProperty* permite especificar que una propiedad P es simétrica si, para cada instancia Y relacionada con X por $P(X \rightarrow Y)$ también X se relaciona con Y por $P(Y \rightarrow X)$.
 - *owl:inverseOf* es un tipo de propiedad simétrica. Una propiedad o relación puede ser inversa si, la propiedad R1 es la inversa de la propiedad R2 y X está relacionada con Y por R2, mientras que, Y está relacionada con X por R1. Sólo precisa su etiquetado en una de las instancia de la relación, bien la que especifica el dominio, bien la que constituye el rango para indicar el significado inverso de la propiedad.

- Restricciones sobre las instancias que podrán formar parte del rango de una propiedad y que, siempre, van asociadas a la etiqueta *owl:Restriction*:
 - *owl:AllValuesFrom* esta restricción especifica que todas las instancias de la clase rango estarán relacionadas por la propiedad con una instancia de la clase dominio.
 - *owl:SomeValuesFrom* establece que, al menos, una de las instancias de la clase rango, estará relacionada por la propiedad con una de las instancias de la clase dominio.
- Restricciones sobre las instancias que podrán formar parte del dominio de una propiedad o la cardinalidad. Siempre, van asociadas a la etiqueta *owl:Restriction*:
 - *owl:maxCardinality* establece el máximo de veces que una instancia de la clase dominio puede aparecer en una relación de la propiedad.
 - *owl:minCardinality* especifica el número mínimo en que una instancia de una clase dominio puede formar parte en una relación de la propiedad.
 - *owl:cardinality* especifica la cardinalidad exacta en que la instancia de una clase dominio podrá formar parte de una relación de la propiedad.
- Restricciones de cardinalidad global de la propiedad:
 - *owl:FunctionalProperty* permite establecer que los pares de la propiedad (X,Y) y (X,Z) tienen como valor de instancia Y el mismo que para Z, es decir, la instancia X sólo puede relacionarse con una sola instancia (Y=Z).
 - *owl:InverseFunctionalProperty* una propiedad es inversa funcional si dos pares de la propiedad (X,Y) y (Z,Y) tienen como valor de la instancia X el mismo valor para Y, es decir, la instancia Y sólo puede ser relacionada con una instancia (X=Y).

Además, hay dos etiquetas suplementarias empleadas para dar información complementaria a los usuarios humanos. La etiqueta *AnnotationProperty*, que permite introducir comentarios en lenguaje natural entre el vocabulario de OWL DL. La etiqueta *OntologyProperty*, que sirve para informar sobre las características de una ontología y está pensada para facilitar indicadores para su reutilización (acreditación).

En síntesis, OWL tiene una sintaxis abstracta e independiente de cualquier representación computacional, basada tanto en XML (define el esquemas de etiquetas a medida) como en RDF (flexibiliza la representación de los datos), con la que se puede describir, formalmente, el significado de la terminología empleada, además de una semántica superior a la que aporta RDFS.

4.3. Propuesta tecnológica: VisualWADE.

La propuesta tecnológica se respalda con la aplicación VisualWADE³⁶. Esta aplicación se basa en el método de modelado OO-H (*Object-Oriented-Hypermedia*). Este método facilita un entorno de producción de software automatizado Web (Gómez, Cachero y Pastor, 2000 y 2001). En la actualidad, es capaz de generar aplicaciones escritas en HTML (*HyperText Markup Language*), PHP (*Hypertext Pre-processor*), Java (*JSP/JavaServer Pages*) y Visual Basic Script (*ASP/Active Server Pages*) y, además, es capaz de generar el código necesario para emplear SGBD Oracle y MySQL.

Esta aplicación incluye una herramienta CASE (*Computer-Aided Software Engineering*) que permite desarrollar el esquema conceptual de un sistema de información utilizando el modelo de datos UML y, en concreto, el diagrama de clases. El diagrama de clases UML, que se diseña con CASE VisualWADE, es único para una aplicación software y, de él, se deriva el esquema de base de

³⁶ Más información sobre esta herramienta y su desarrollo en http://gplsi.dlsi.ua.es/iwad/ooh_project/index.htm

datos que empleará la propia aplicación, así como los diagramas de navegación (NAD, *Navigation Access Diagram*) y el diagrama de presentación (APD, *Abstract Presentation Diagram*)³⁷.

Los diagramas de clases generados con CASE VisualWADE presentan la misma limitación que tiene, en general, el modelo de datos UML. El conocimiento contenido en el diagrama es de tipo local y propio de la representación del diagrama, lo que hace difícil su reutilización, aún utilizando lenguajes diseñados con ese propósito (XMI, *XML Metadata Interchange*). Asimismo, las aplicaciones web generadas carecen del etiquetado semántico de su contenido, aspecto cada vez más imprescindible tanto para la organización como para la recuperación de la información. Esta desventaja de CASE VisualWADE se ha superado realizando una extensión de su funcionalidad, ampliando sus *plug-ins*. En concreto, con el desarrollo del *plug-in Semantic Web Tools for VisualWADE* (Hermida, 2007), cuya función es trasladar el esquema conceptual diseñado con la herramienta CASE VisualWADE al lenguaje de representación OWL DL.

En el siguiente apartado se describe el proceso de actuación del *plug-in* para trasladar los diferentes elementos del diagrama de clase UML al formalismo de OWL DL. Este proceso tiene una fase de transformación de contenido y una posterior de transformación de nombres.

4.3.1. Semantic Web Tools for VisualWADE.

Este *plug-in* es un método ‘*ad hoc*’ que define las transformaciones directas entre el diagrama de clases UML y el modelo de datos OWL DL. El procedimiento de transformación, para generar el esquema conceptual por

³⁷ El método de modelado OO-H proporciona otros dos diagramas: NAD (*Navigation Access Diagram*) y APD (*Abstract Presentation Diagram*). Éste último proporciona un editor de páginas web que permite al usuario ajustar el estilo de la aplicación. Estos diagramas no han sido descritos, al quedar fuera del alcance esta Tesis.

- Extraer la semántica del diagrama de clases.
- Generar una representación interna de la ontología.
- Generar la ontología en el lenguaje de representación OWL DL.

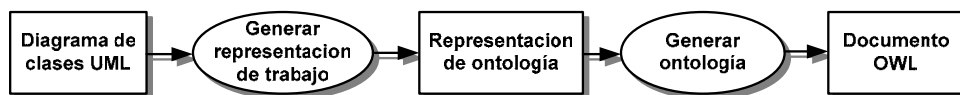


Figura 4.4. Diagrama del procedimiento de transformación del *plug-in Semantic Web Tools for VisualWade*.

Se generó una ontología de referencia para poder almacenar, de forma no persistente, cualquier ontología que se creara con la herramienta CASE VisualWADE. Esta ontología, para las transformaciones del plug-in *Semantic Web Tools for VisualWADE*, se describió a partir de los conceptos básicos del lenguaje de representación OWL (ver figura 4.5).

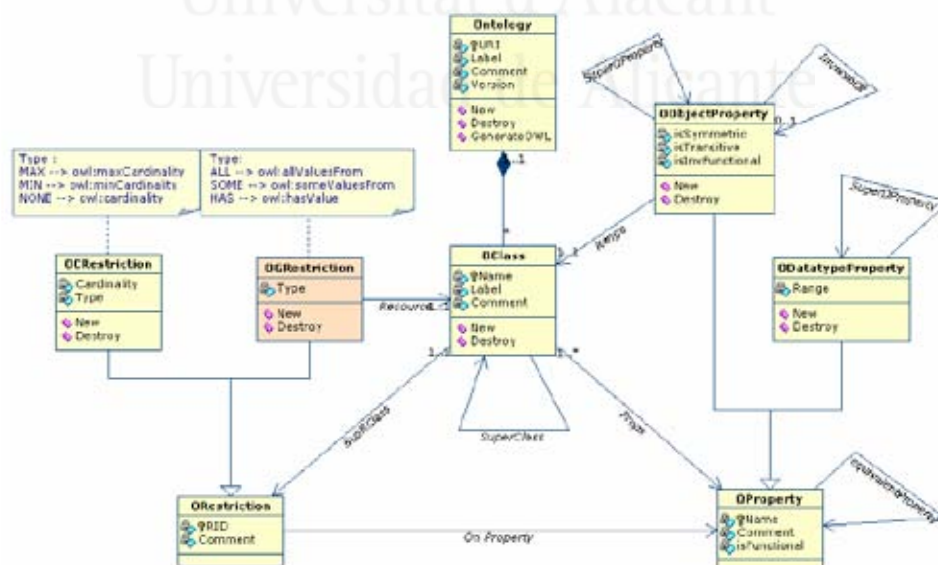


Figura 4.5. Diagrama de clases de la ontología de referencia para las transformaciones (Hermida, 2007).

El diagrama de clases, de la representación de la ontología de referencia, contiene los siguientes componentes: clases, propiedades, restricciones y relaciones. Dichos componentes han sido definidos del siguiente modo (Hermida, 2007):

Clases:

- *OClass* (Clase *Ontology Class*): Es la clase que representa los diferentes conceptos de un dominio. Esta clase se define como composición de varios elementos *OClass*. Para indicar que una clase es de tipo superclase se emplea la relación *Superclase*.

Propiedades:

- *OProperty* (Clase *Ontology Property*): Es una clase de tipo abstracto que referencia a las propiedades de forma global. Es una superclase de la que heredan las subclases *ODatatypeProperty* (Clase *Ontology Datatype Property*) y *OObjectProperty* (Clase *Ontology Object Property*).

Restricciones:

- *ORestriction* (Clase *Ontology Restriction*): Es la superclase de *OCRestriction* (Clase *Ontology Cardinality Restriction*) y *OGRestriction* (Clase *Ontology Grup Restriction*). Indican, respectivamente, las restricciones de tipo Cardinalidad y de Grupo. Esta última subclase de restricción permite representar, a su vez, las de tipo *someValuesFrom* y *allValuesFrom*.

Relaciones:

- *Prop*: Es la relación que se establece entre la clase *OClass* con aquellas propiedades que aparecen como dominio de dicha clase.
- *Range*: Es el tipo de relación que se crea entre el rango de la clase *OObjectProperty* para definirla en la clase *OClass*.
- *SubRClass*: Es la relación que se establece entre la clase *OClass* y la clase *ORestriction* para definir las restricciones de *OClass*.

- *OnProperty*: Es la relación que se crea entre la clases *ORestriction* y *OProperty* para indicar sobre qué propiedades se aplicará la restricción.
- *Resource*: Esta relación define las restricciones de grupo a las instancias de la clase *OClass*.

El proceso general de transformación del *plug-in* parte del diagrama de clase diseñado en UML, al que no se le incorpora ninguna de las restricciones OCL que contiene la herramienta CASE VisualWADE. Los pasos significativos del proceso, según el tipo de clases definidas en el diagrama, son:

- La ontología:

Globalmente, la ontología se identifica con un URI general. El resto de URI que identifiquen a los componentes de la ontología dependerán del URI general para su denominación.

- Clase individual con ciertos atributos:

Los conceptos clase en UML y de clase en una ontología OWL son similares, lo que permite traducirlos uno por uno de forma directa. El URI que identifica a una clase se constituye por el URI general y la agregación del nombre de la clase.

Los atributos de la clase son un concepto similar a la propiedad *owl:DatatypeProperty*. El URI que se le asigna a cada propiedad generada, es el URI de la ontología al que se agrega el nombre de la propiedad. Pero, la propiedad es renombrada, previamente, añadiendo información de la clase, pasando a denominarse de acuerdo con el patrón “NomClase..NomAtributo”.

- Clase con herencia:

En este caso la transformación crea la nueva clase OWL DL, como subclase de su respectiva superclase. Las propiedades distintivas de la subclase se definen como dominio de la superclase, por medio de la

propiedad *owl:DataTypeProperty*, y se denominan con el patrón “NomSubclase..NomAtributo”.

- Clases relacionadas mediante una relación asociativa:

En este caso no existe una correspondencia directa entre UML y OWL DL en la forma de expresar una relación de asociación entre dos clases. La transformación es por aproximación y se obtiene una *owl:ObjectProperty* por cada rol navegable que tenga la asociación entre dos clases.

El URI de cada *ObjectProperty* se compone del URI general agregándole el nombre del rol. Aunque, el rol es renombrado y se le añade el nombre de la asociación, de acuerdo con el patrón “NomAsociacion..NomRol”.

Si los dos roles entre dos clases en UML son navegables (rolA y rolB) a partir de una asociación, en la transformación, el segundo de los roles es identificado con la restricción lógica *owl:inverseOf*.

Un rol en UML con cardinalidad tiene una transformación por correspondencia directa con OWL. Aunque, pueden existir diversas variantes: i) Cardinalidad máxima O..1, en el diagrama se define con la restricción *owl:maxCardinality*; ii) Cardinalidad mínima de 1..*, se transforma en la restricción *owl:minCardinality*; y iii) Cardinalidad total 1..1, se define con *owl:cardinality*.

- Clases relacionadas mediante una composición:

En este caso, no existe una correspondencia directa entre UML y OWL, en la forma de expresar una relación de composición. La transformación es por aproximación. Para ello, se han definido, previamente, dos *ObjectProperty* para los roles parteDe y TieneParte como subclases de la Superclase de *owl:Thing*. La *ObjectProperty* TieneParte (rango) se ha definido con la restricción lógica *inverseOf* de la *ObjectProperty* ParteDe (dominio). Asimismo, *ObjectProperty* ParteDe (dominio) se ha definido con la característica de propiedad transitiva.

A estas propiedades de tipo general, definidas ‘*ad hoc*’, se le asocian como subpropiedades en la transformación, todas las relaciones de tipo ‘composición’ descritas en el diagrama de clases UML.

- Clases relacionadas por agregación:

En este caso, de nuevo, la semántica asociada es diferente entre UML y OWL DL. El proceso de transformación emplea el concepto de *owl:unionOf* para trasladar: i) una agregación como agrupación enumerada de clases; ii) una agregación como agrupación de clases con comportamiento similar a una generalización; y iii) una agregación como agrupación de clases siguiendo un criterio definido por el usuario.

Como se observa, por lo anteriormente expuesto, existen dos aspectos de las propiedades de OWL DL que no pueden especificarse empleando un diagrama de clases. En concreto, las características de las propiedades de tipo simétrica (*owl:SymmetricProperty*) y transitiva (*owl:TransitiveProperty*) y las restricciones de cardinalidad global, como la funcional (*owl:FunctionalProperty*) y la funcional inversa (*owl:InverseFunctionalProperty*). Para indicar qué características se definirán se han ampliado los campos de definición de los roles en la herramienta CASE VisualWADE (Figura 4.6). Esta información se incorpora, gráficamente, en el diagrama de clases y se visualizan entre corchetes las iniciales de cada propiedad (S, T, F e IF, respectivamente) y separadas por comas.

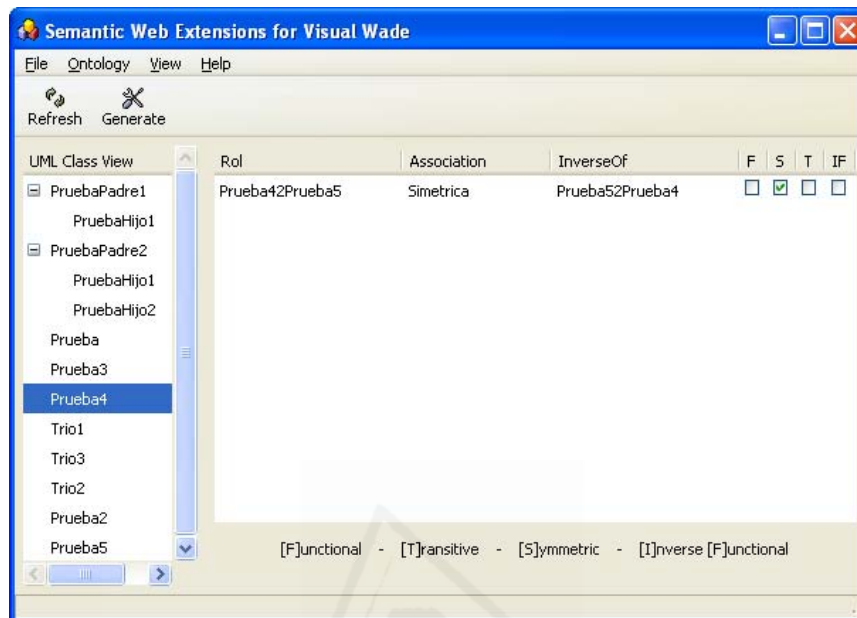


Figura 4.6. Interfaz para ampliar los campos de definición de los roles.

No obstante, existen restricciones para aplicar estas anotaciones a un rol en el diagrama UML:

- Un rol sólo puede ser marcado como funcional si es candidato a ser traducido como una restricción global de la propiedad de tipo *owl:FunctionalProperty*. Es decir, si tiene cardinalidad máxima o total de 1 y su rol inverso no tiene definida ninguna restricción sobre su cardinalidad. Si se marca un rol como funcional, automáticamente, tiene que ser marcado su rol inverso como funcional inverso.
- Un rol sólo puede ser marcado como funcional inverso si es candidato a ser traducido como una restricción global de tipo *owl:InverseFunctionalProperty*. Es decir, si no tiene restringida su cardinalidad y su rol inverso tiene una restricción de cardinalidad máxima o total de 1. Si se marca un rol como funcional inverso, automáticamente, tiene que ser marcado su rol inverso como funcional.

- La semántica de OWL DL impone que un rol marcado como transitivo no puede ser funcional ni funcional inverso.

4.3.1.1. Limitaciones.

La limitación de expresividad, más destacada, de este *plug-in* es la carencia para establecer que dos clases son equivalentes. Aunque, esta limitación no obstaculiza la funcionalidad de *Semantic Web Tools for VisualWADE*, teniendo en cuenta que, su diseño, no está pensado para la importación de otras ontologías y que su principal funcionalidad es crear el etiquetado semántico de las aplicaciones creadas localmente. Además, en el diseño de un diagrama de clases en UML, no se permite que dos clases se denominen de igual forma, ya que es considerado un error de diseño que dos clases representen un mismo concepto.

4.4. OntoFIS.

La ontología Farmacoterapéutica e Información para el Seguimiento, OntoFIS, contiene el conocimiento relacionado tanto con el proceso (terapéutica) como con el objeto (medicamento). OntoFIS especifica y explicita los conceptos relacionados con los problemas de salud clínicos (enfermedades, signos y síntomas, agentes etiológicos), los medicamentos (denominación, concentración, forma farmacéutica, principio activo, clasificación terapéutica, posología, conservación), la administración (vía, preparación) y la seguridad (efectividad, efectos adversos, toxicidad). La descripción responde al conocimiento compartido de los potenciales perfiles de usuarios (médico, farmacéutico, enfermera y consumidor) e incluye conocimiento multilingüe.

La finalidad de la ontología es disponer de un recurso, con conocimiento terminológico multilingüe, que facilite las tareas de: i) etiquetado semántico de las aplicaciones web autocreadas (VisualWADE), ii) clasificación automática de documentos, iii) recuperación de documentos y iv) extracción de información. Las tres últimas están relacionadas con tareas de PLN.

En los siguientes apartados se describen las etapas de diseño y de construcción de OntoFIS.

4.4.1. Diseño y construcción de OntoFIS.

En esta propuesta se entiende que el problema a tratar y el resultado esperado o, lo que es lo mismo, el propósito y el alcance de la ontología son los objetivos que determinan, tanto la conveniencia de utilizar una ontología, como el tipo de conocimiento a representar. A partir del enunciado del propósito y del alcance comienza la aplicación, en sí, de los procedimientos para lograr dichos objetivos. Los objetivos son inamovibles durante todo el proceso de construcción. Si se han de cambiar, entonces, se comienza de nuevo, pues el proyecto habrá cambiado. Esta consideración marca una diferencia con las metodologías analizadas anteriormente (tabla 4.3). Aunque, no establece ninguna contradicción ya que “no existe ni una sola forma o ni una sola metodología “correcta” para desarrollar ontologías” (Noy y McGuinness, 2001).

La propuesta de diseño y construcción de una ontología desde cero e independientemente de la herramienta cuenta con tres etapas: i) *la captura del conocimiento aceptado del dominio*, ii) *la descripción del conocimiento según la técnica de modelado conceptual* y iii) *la formalización del conocimiento capturado* (ver figura 4.7).

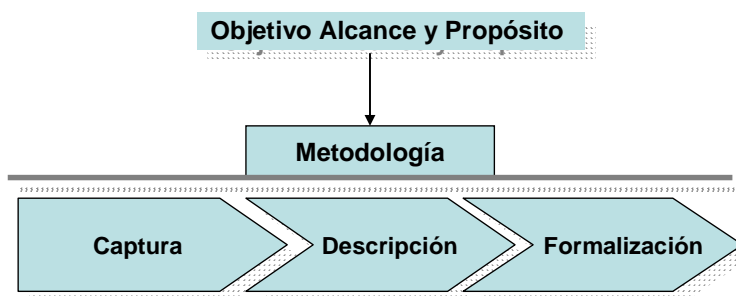


Figura 4.7. Esquema de la propuesta de metodología.

Una consideración general de esta propuesta, de acuerdo con las indicaciones de Noy y McGuinness (2001), es el enfoque iterativo que se mantiene durante todo el proceso de construcción de la ontología. Este proceso iterativo o *'feedback'* persigue controlar la descripción en relación al propósito y el alcance establecido previamente. La evolución y el refinamiento de la ontología, desde la etapa inicial de diseño hasta la versión final, se fundamentará en los resultados de las valoraciones realizadas con los expertos o en la aplicación de métodos de evaluación específicos.

4.4.1.1. La etapa de captura del conocimiento aceptado del dominio.

La sistemática para la captura del conocimiento tangible y tácito se apoya en la selección de las fuentes de información fiables del dominio diana. Hay diversas fuentes que contribuyen a delimitar la extensión y la profundidad del conocimiento utilizado en el ámbito de interés.

Entre las fuentes textuales se encuentran los diccionarios, las enciclopedias, los manuales, las monografías, las publicaciones seriadas o cualquier otro tipo de documentación del contexto a explorar. Entre las fuentes orales están la selección y la entrevista a expertos o a usuarios de la información del dominio. Asimismo, dependiendo del problema y de la finalidad por la que

se construya la ontología, se puede optar por la descripción de tareas utilizando técnicas de observación.

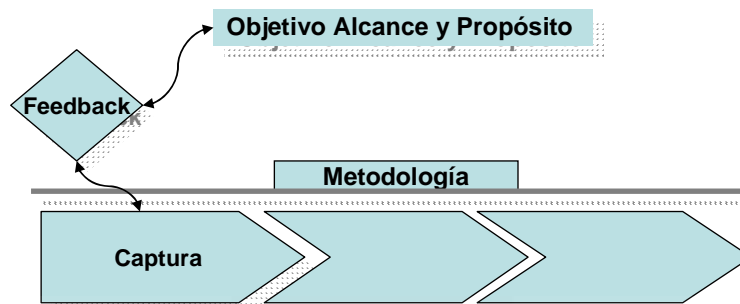


Figura 4.8. Proceso de retroalimentación para la toma de decisiones de la primera etapa.

Esta etapa cuenta con un proceso iterativo en el que el patrón de comparación está constituido por el alcance y el propósito establecidos (figura 4.8). Esta etapa termina cuando se puede responder, con suficiente detalle, a las siguientes cuestiones: ¿Está determinado el conjunto de términos que establece el alcance de la representación? ¿Se cuenta con conjuntos diferenciados de términos representativos (distintivos o esenciales, secundarios o complementarios y sinónimos) que capturan los conceptos característicos del dominio? ¿Está detallado, en un glosario, el significado de los términos de acuerdo con el ámbito de su uso?

4.4.1.2. La etapa de descripción del conocimiento según el modelado conceptual.

A partir de la información capturada en la etapa anterior, comienza el proceso de modelado. Esta etapa tiene dos tareas previas a la descripción: la determinación del nivel de detalle que se precisará y la identificación de las limitaciones que impondrá el modelo de datos.

Por una parte, el nivel de descripción informal se realizará de acuerdo con el alcance y el propósito previamente establecido y, asimismo, con el

conocimiento capturado en la etapa anterior. El objetivo de esta tarea es establecer el nivel de granularidad de los conceptos. La decisión se basará en el análisis de las ventajas y de los inconvenientes para cumplir las funcionalidades previstas. El análisis deberá valorar si es más funcional establecer conceptos de baja granularidad o si, realmente, lo que se precisará será poblar la ontología con instancias. Asimismo, para fundamentar una decisión de granularidad alta se realizará, previamente, un análisis comparativo para determinar las posibles pérdidas de información que pudieran resultar vitales respecto al propósito y el alcance, sobre todo para las posteriores inferencias, a partir del conocimiento representado.

Por otra parte, la descripción estará delimitada por el propio modelo de datos que se seleccione para construir la ontología. Como se ha indicado anteriormente, el tipo de modelo condicionará: la sintaxis para expresar los conceptos, las propiedades y las relaciones que contendrá la ontología (tabla 4.6), determinará la estructura organizativa y la especificación de las posibles restricciones o el tipo de herencia entre los conceptos descritos (Lassila y McGuinness, 2001).

Descripción Lógica	Frames	OO
Class, Concept	Frame, Schema	Class, Type
Role	Slot	Attribute, Instante Variable
Filler	Filler	Value
Instante, Individual	Frame, Instante, Individual	Instante

Tabla 4.6. Variantes del vocabulario para la denominación de concepto, propiedad, relación e instancias de acuerdo al modelo de datos (reproducido de Lassila y McGuinness, 2001).

De acuerdo con las conclusiones alcanzadas en estas tareas previas, se continua con la actividad de análisis de los términos utilizando el glosario creado. El resultado del análisis permite la identificación y la organización de los términos que representarán, bien a los conceptos bien a los atributos, y las relaciones de tipo jerárquico, u otro tipo de asociaciones semánticas de interés para el dominio. En definitiva, posibilita la descripción del conocimiento aceptado.

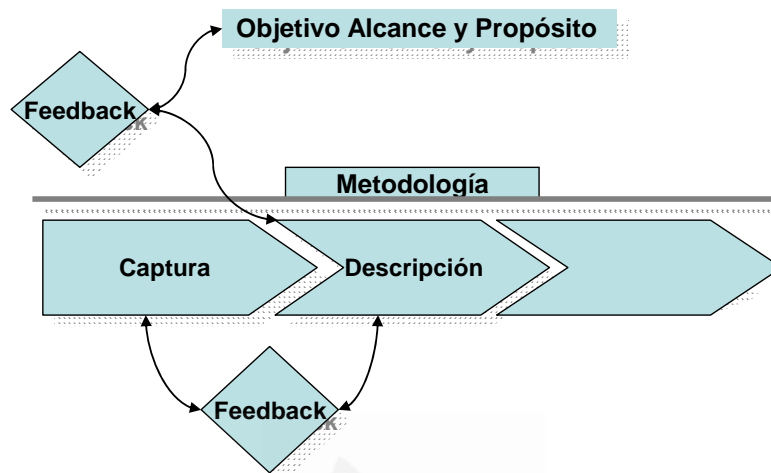


Figura 4.9. Proceso de verificación para la toma de decisiones de la segunda etapa.

La etapa se da por concluida cuando se obtienen respuestas precisas para, al menos, las siguientes cuestiones: ¿Se diferencian los conceptos recopilados en conceptos representativos, propiedades y relaciones? ¿Se determina el tipo de relación entre los conceptos? ¿Se establece una estructura de organización para representar los conceptos distintivos, sus propiedades, las posibles restricciones y su valores, de acuerdo con el conocimiento aceptado del dominio y las limitaciones impuestas por la técnica de modelado? Las respuestas de estas cuestiones se confirman por iteración respecto a la captura realizada y los objetivos previos al premodelado (figura 4.9).

4.4.1.3. La etapa de formalización del conocimiento capturado.

En esta última etapa, el lenguaje de representación terminará estableciendo el tipo de formalización del conocimiento capturado. Las opciones se encuentran delimitadas por los componentes del conocimiento capturado y por la técnica seleccionada para modelar la conceptualización de la representación (Gómez-Pérez et al, 2004). No obstante, como se comentó en el segundo capítulo de esta Tesis, el tipo de inferencia que, posteriormente, se

necesitará para llevar a cabo el propósito de la ontología será la que, en definitiva, estipulará la necesidad de un lenguaje rigurosamente formal o semiformal. En el primer caso, se conformaría una ontología pesada, mientras que, en el segundo caso, se constituiría una ontología ligera.

La gama de lenguajes es amplia y diversa. Están, desde los lenguajes enfocados al desarrollo de ontologías basados en técnicas de Inteligencia Artificial (KIF, CycL, OCML, Flogic, ...), hasta los lenguajes cuyo propósito es la identificación de los datos para el intercambio y su procesamiento automático. Entre estos últimos, se encuentran los lenguajes ontológicos desarrollados para la Web, como se ha comentado anteriormente. Todos estos lenguajes tienen ventajas y limitaciones pero, un análisis de todos ellos, queda fuera del propósito de esta Tesis (para más información consultar Uschold y Jasper, 1999; Cranefield y Purvis, 1999; Gómez-Pérez et al, 2004 y Nguyen, 2006).

El diseño y la construcción de la representación termina cuando la ontología es codificada en un lenguaje de representación de conocimiento, se comprueba la ausencia de errores de sintaxis y se confirma la adecuación del esquema de la representación, descartando, por ejemplo, la existencia de problemas de circularidad. La detección de deficiencias o de fallos desencadena un análisis, tanto de la descripción del modelo como una comparación con los objetivos previos. El alcance y el propósito de la ontología actúan, de nuevo, como una guía para el refinamiento (figura 4.10).

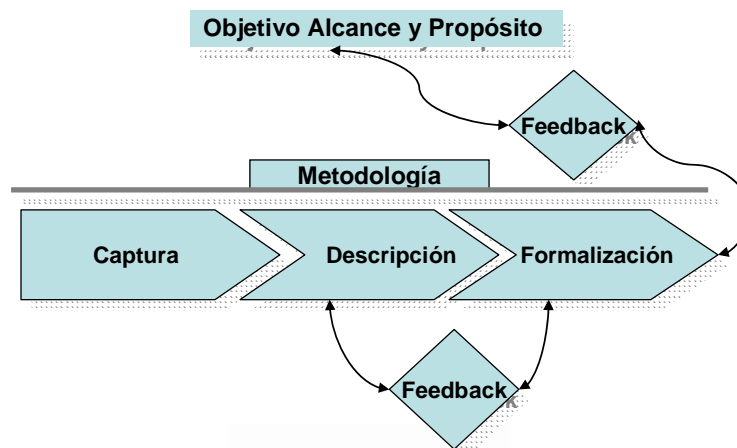


Figura 4.10. Proceso de verificación de la última etapa de desarrollo de una ontología.

4.4.2. Captura del conocimiento del dominio farmacoterapéutico.

La sistemática, aplicada para la captura del conocimiento explícito del dominio farmacoterapéutico y compartido por médicos, farmacéuticos y enfermeras, se orientó a las fuentes documentales. El criterio para su selección fue que éstas debían facilitar el conjunto de términos involucrados con el alcance de la representación: términos, sinónimos y traducciones en diversas lenguas de la UE. Se realizaron las siguientes tareas para la recuperación de documentos representativos (Anexo I: artículos y otras fuentes documentales seleccionadas):

- Búsqueda en la base de datos bibliográfica Medline y en la revista *Pharmaceutical Care* España. Los artículos seleccionados describían las necesidades de los usuarios para obtener información farmacológica, o bien las características de estos sistemas de información o de sistemas de ayuda en la Web basados en pregunta-respuesta y con contenido relacionado con tratamientos farmacológicos.
- Selección de otras fuentes documentales con información relacionada con el proceso farmacoterapéutico: manuales de medicina, de farmacología y de

procedimientos de enfermería para la administración de medicación, Vademécum, fichas técnicas y prospectos, entre otros.

Los términos se iban seleccionando conforme aparecían, explícitamente, en los documentos y siempre que tuvieran una vinculación con el proceso de la atención de la terapéutica farmacológica o con el medicamento o fármaco (figura 4.11). Se recopilaron 144 términos significativos para el dominio (Anexo II: total de términos recopilados de las fuentes documentales).

Conceptos \ Referencias	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Denominación	X	X	X	X	X	X	X	X	X
Principio Activo	X	X	X	X	X	X	X	X	X
Composición Química	X	X	X	X	X	X	X	X	X
Grupo terapéutico	X	X	X	X	X	X	X	X	X
Sirve para	X	X	X	X	X	X	X	X	X
Indicado [Nota: acción positiva]	X	X	X	X	X	X	X	X	X
Produce [Nota: acción negativa]	X	X	X	X	X	X	X	X	X
Signos/Síntomas	X	X	X	X	X	X	X	X	X
Enfermedades	X	X	X	X	X	X	X	X	X
Interacciones	X	X	X	X	X	X	X	X	X

Figura 4.11. Sistemática de recogida de los términos relevantes de la atención terapéutica.

Seguidamente, cada término fue definido o descrito, indistintamente, con la ayuda de los manuales y dos diccionarios, conformando el glosario de términos (ANEXO III: fragmento del glosario creado.). De acuerdo con las glosas se comprobó, posteriormente, que la colección de términos capturados contenía: i) todos los términos esenciales de acuerdo con el alcance y el propósito de OntoFIS, ii) algunos términos que contribuían a aumentar el conocimiento y, a su vez, especificar los términos esenciales y iii) un grupo de términos sinónimos o las traducciones de los términos en inglés, francés, italiano y alemán.

Esta etapa se dio por concluida al contar con un conjunto representativo de términos que capturaban los conceptos característicos del dominio definido, dando paso a la siguiente etapa³⁸.

4.4.3. Descripción del conocimiento con la técnica de modelado UML.

Esta etapa se inició abordando las tareas para la determinación del detalle que precisaba la descripción y la identificación de las limitaciones que impondría la técnica de modelado.

Respecto a la primera de estas tareas, se comenzó con la valoración de las ventajas e inconvenientes de una alta o de una baja granularidad de los conceptos a incluir en la representación. Un reto evidente del dominio es la amplitud del conocimiento que se precisa, tanto del ámbito de la salud en general (problemas de salud) como para el uso racional del medicamento (composición y forma farmacéutica), como ya se ha comentado. Se consideró la cobertura temática y la multilingüidad de otras fuentes de conocimiento existentes (tablas 4.1 y 4.2). El criterio final fue incluir conceptos de granularidad intermedia. La decisión era coherente con el alcance y el propósito, previamente establecido, y con la condición de cubrir el vacío detectado. Además, la decisión de describir el dominio con conceptos de granularidad intermedia planteaba la ventaja de permitir, posteriormente, ampliar y especificar, con mayor detalle, el dominio representado con otras ontologías³⁹ y evitar describir lo que ya está comprobado y que sí puede ser reutilizado.

³⁸ En este caso no se consultó con expertos externos, sólo se contó con el criterio del investigador.

³⁹ En la actualidad, hay aproximaciones que tratan el problema de la compatibilidad entre diferentes ontologías para permitir compartir la información entre ellas. No obstante, diferentes autores utilizan diversas expresiones para referirse a conceptos similares y, al mismo tiempo, diferentes conceptos son indicados con la misma denominación. Entre dichos conceptos se encuentran:

i) *Ontology Merging*, se refiere a la obtención de una ontología única que contiene todos los términos unidos de las ontologías originales, sin indicación de su origen anterior.

Respecto a la segunda tarea, la dirigida a seleccionar el tipo de modelado conceptual, hay que indicar que la propuesta de diseño y construcción puede ser adaptada a cualquier dominio y no preestablecer ningún tipo de dependencia sobre ningún tipo de modelado conceptual o herramienta concreta de implementación. No obstante, una característica propia de OntoFIS es que se diseña y se desarrolla aplicando una estrategia de reutilización del diagrama de clases UML, desarrollado con la herramienta CASE VisualWADE, para, a su vez, generar aplicaciones web de manera automática. La reutilización se orientó a la optimización de esfuerzo y de tiempo en el diseño.

Seguidamente, de acuerdo con el criterio adoptado de especificación, se analizaron los términos y el tipo de concepto que simbolizaban con la ayuda del glosario. También, se establecieron las reglas a emplear para la denominación de las clases y de las propiedades de acuerdo con el modelo de datos. Se seleccionaron y se ordenaron 23 conceptos de granularidad intermedia para ser representados en el diagrama de clases UML (tabla 4.7).

ii) *Ontology Integration*, se utiliza cuando se consigue una única ontología creada por recopilación, ampliación, especialización o adaptación de ontologías de diferentes área temáticas, pero cuando se integran las ontologías es posible identificar las relaciones hacia la ontología origen.

iii) *Ontology Alignment*, se indica el mantenimiento de las ontologías separadas aunque con enlaces entre ellas a partir de un conjunto de correspondencias expresadas (interconexiones o mapping) que permiten a las ontologías compartir los términos.

iv) *Ontology Mapping*, se refiere a una estructura formal que contiene expresiones que enlazan conceptos de un modelo a otro.

v) *Ontology Coordination*, es un término amplio que es aplicable siempre que el conocimiento de las dos ontologías (o más) puede ser utilizado al mismo tiempo de una manera significativa.

vi) *Ontology Transformation*, se refiere a cualquier proceso que conlleva una nueva ontología desde una ontología previa por la utilización de una función de transformación.

vii) *Ontology Translation*, indica la aplicación de una función de transformación para la traducción de una ontología escrita en algún lenguaje en otra ontología escrita en un lenguaje distinto.

viii) *Ontology Reconciliation*, es la denominación del proceso que armoniza el contenido de dos o más ontologías y que típicamente requiere cambios en una de las dos o incluso en ambas ontologías.

ix) *Meaning Negotiation*, se refiere al protocolo a través del cual dos agentes, humanos o artificiales, se ponen de acuerdo en los cambios requeridos para reconciliar sus ontologías.

En todos los casos, la interconexión (*mapping*) es una expresión formal que declara la relación semántica entre dos entidades pertenecientes a diferentes ontologías. Cuando la relación es orientada, ésta se corresponde con una función cuyo dominio es un conjunto que contiene un elemento simple.

Para más ampliación consultar: Bouquet, P. (Coord). (2004, Agosto). *Specification of a common framework for characterizing alignment. Technical Report: KWEB/2004/D2.2.1/v1.2. Proyecto Knowledge web: KWEB EU-IST-2004-507482*. Recuperado el 15 de Octubre de 2005, desde <http://eprints.biblio.unitn.it/archive/00000653/01/090.pdf>

Fármaco
Composición Química
Excipiente
Principio Activo
Acción
Efecto Alterado
Efecto Indeseable
Efecto Deseable
Indicación Terapéutica
Grupo Terapéutico
Forma Farmacéutica
Contraindicación
Posología
Vía de Administración
Sobredosis
Estabilidad
Proceso Clínico
Agente Tóxico
Agente Infeccioso
Población
Edad
Género
Unidad Medida

Tabla 4.7. Lista anidada de los conceptos seleccionados para diseñar OntoFIS.

Otros términos capturados se emplearon para establecer las propiedades intrínsecas de cada clase (con una media de 4 propiedades y un rango entre 3 y 6 propiedades) y las propiedades extrínsecas. Éstas últimas referidas a las etiquetas terminológicas que se utilizan para su denominación en español, inglés, alemán, francés e italiano (ANEXO IV: descripción informal de la ontología OntoFIS).

Para explicitar las relaciones entre conceptos se utilizó la taxonomía propuesta por Slaughter, Soergel y Rindflesch (2006), que incluye 59 tipos de relaciones asociativas. La asociación entre dos conceptos (domino-rango) fue establecida por aquellas relaciones de la taxonomía que reflejaban el conocimiento expresado, tanto en las fuentes de documentales consultadas como en el glosario creado (tabla 4.8).

Conocimiento Capturado		
<p>Medicament: Substància, simple o composta, emprada amb una finalitat terapèutica contra les manifestacions patològiques, tant si produeix el guariment suprimint la causa de la malaltia com si només n'atenua els símptomes.</p> <p>El medicament és presentat en forma medicamentosa adaptada a una determinada via d'administració, anomenada forma farmacèutica.</p> <p>L'efecte terapèutic del medicament s'acompanya, a vegades, d'efectes no desitjables o secundaris</p>		
Concepto Dominio	Relación	Concepto Rango
Farmaco	esta_contenido_en	Forma_farmaceutica
Fármaco	esta_formado_por	Composicion_quimica
Fármaco	esta_formado_por	Principio_activo
Fármaco	Provoca	Accion
Fármaco	Facilita	Efecto_deseable
Fármaco	Interrumpe	Proceso_clinico
Fármaco	Indicado	Proceso_clinico
Farmaco	Requiere	Posologia
Farmaco	Requiere	Via_administracion
Proceso_clinico	Relacionado_con	Farmaco
Efecto_indeseable	derivado_de	Farmaco
Efecto_indeseable	Afecta	Proceso_clinico
Conocimiento Capturado		
<p>Contraindicaciones son las circunstancias en las cuales no debe de administrarse el medicamento, y los antecedentes de hipersensibilidad al principio activo o a cualquiera de los componentes de la formulación.</p>		
Concepto Dominio	Relación	Concepto Rango
Contraindicacion	Impide	Farmaco
Contraindicacion	derivada_de	Farmaco
Contraindicacion	esta_relacionado	Proceso_clinico

Tabla 4.8. Relaciones asociativas entre conceptos, a partir de fragmentos de conocimiento capturado.

Se establecieron inicialmente 708 asociaciones. Tras los refinamientos realizados, posteriormente, se redujo el número a 647 relaciones totales. En el proceso se detectaron 61 relaciones inadecuadas, bien por duplicación o bien por estar asignadas incorrectamente.

Associated with:

1. full_related_to (ES DIR: integramente_relacionado_con)
 - 1.1. part_of (ES DIR: parte_de) (IN INV: has_part) (ES INV: tiene_parte)
 - 1.1.1. consists_of (IN DIR: is_ingredient_of, is_component_of) (ES DIR: forma, es_ingredient_de, compone) (IN INV: is_composed_of= has_ingredient) (ES INV: esta_formado_de/por, tiene_de_ingredient, esta_compuesto_por/de)
 - 1.2. related_to (ES DIR: relacionado_con) (IN INV: is_related_to) (ES INV: esta_relacionado)
 - 1.3. belongs (ES DIR: pertenece, integra) (IN INV: is_belonged, is_intgrated) (ES INV: esta_integerrado_por) SIMÉTRICA
2. functionally_related_to (ES DIR: funcionalmente_relacionado_con)
 - 2.1. affects (IN DIR: influences) (ES DIR: afecta, influye) (IN INV: is_affect, s_influenced) (ES INV: esta_afectado, esta_influido)
 - 2.1.1. delays = 3.3 (ES DIR: retrasa, retarda) (IN INV: is_delayed) (ES INV: es_retrasado, es_retardado)
 - 2.1.2. complicates (ES DIR: complica) (IN INV: has_complicated) (ES INV: esta_complicado)
 - 2.1.3. disrupts (IN DIR: interrupts) (ES DIR: interrumpe, perturba) (IN INV: is_interrupted, is_disrupted) (ES INV: es_interrumpido, es_perturbado)
 - 2.1.4. facilitates (ES DIR: facilita) (IN INV: is_facilitated) (ES INV: es_facilitado)
 - 2.1.5. increases (IN DIR: increments) (ES DIR: aumenta, incrementa) (IN INV: is_increased, is_incremented) (ES INV: es_aumentado, es_incrementado)
 - 2.1.6. decreases (IN DIR: falls) (ES DIR: disminuye, decrece) (IN INV: is_decreased, is_fallen) (ES INV: es_disminuido, es_decrecido)
 - 2.1.7. interacts_with (ES DIR: reacciona_con)
 - 2.1.8. manages (IN DIR: controls) (ES DIR: se_administra, se_controla) (IN INV: is_managed, is_controlled) (ES INV: es_administrado = es_controlado)
 - 2.1.9. prevents (IN DIR: avoids) (ES DIR: previene, evita, impide) (IN INV: is_prevented, is_avoided) (ES INV: es_prevenido, es_evitado, es_impedido)
 - 2.1.10. treats (ES DIR: trata) (IN INV: is_treated) (ES INV: es_tratado)
 - 2.2. brings_about (IN DIR: cuases, produces, results_in) (ES DIR: provoca, efectua_cambio, causa, ocasiona) (IN INV: is_produced, is_caused, is_resulted_in, is_brought_about) (ES INV: es_provocado, es_efectuado_cambio, es_causado, es_ocasionado)
 - 2.2.1. suitable_for (ES DIR: indicado_para, sirve_para)
 - 2.2.2. acts (ES DIR: actua_en)
 - 2.2.3. unsuitable_for (ES DIR: inadecuado_por)
 - 2.2.4. excludes (ES DIR: excluye) (IN INV: is_excluded) (ES INV: es_excluido)
 - 2.3. performs (IN DIR: works, satisfies) (ES DIR: realiza, cumple, funciona) (IN INV: is_performed, is_worked, is_satisfied) (ES INV: es_realizado, es_cumplido, es_funcionado)
 - 2.3.1. exhibits (IN DIR: exposes) (ES DIR: exhibe, expone) (IN INV: is_exhibited, is_exposed) (ES INV: exhibido, expuesto)
 - 2.4. occurs_in (IN DIR: happens) (ES DIR: ocurre, produce, sucede) (IN INV: has_occurrence, has_happened) (ES INV: ha_ocurrido, ha_producido, ha_sucedido)
 - 2.5. uses (ES DIR: utiliza, usa, emplea) (IN INV: is_used) (ES INV: es_utilizado, es_usado, es_employado)
3. temporally_related_to (ES DIR: relacionado_temporalmente_con)
 - 3.1. co-occurs_with (IN DIR: matches_with, coexits_with) (ES DIR: concurre_con, coincide_con, coexiste_con) (IN INV: has_co-occurred_with, has_matched_with, has_coexisted_with) (ES INV: ha_concurrido_con, ha_coincidido_con, ha_coexistido_con)
 - 3.2. precedes (IN DIR: anticipates) (ES DIR: precede, anticipa) (IN INV: is_preceded, is_anticipated) (ES INV: es_precedido, es_anticipado)
 - 3.3. delays = 2.1.1
 - 3.4. duration_of (ES DIR: tiene_de_duracion, tiene_duracion_de)
 - 3.5. time_position_of (ES DIR: asignación_de_frecuencia)
4. conceptually_related_to (ES DIR: relacionado_conceptualmente_con)
 - 4.1. assesses_effects_of (IN DIR: evaluates_effects_of) (ES DIR: valora_efectos_de, evalua_efectos_de) (IN INV: assessed_effects_of, evaluated_effects_of) (ES INV: valorado_efectos_de, evaluado_efectos_de)
 - 4.2. diagnoses (ES DIR: diagnostica) (IN INV: is_diagnosed) (ES INV: es_diagnosticado)
 - 4.3. property_of (ES DIR: es_propiedad_de)
 - 4.4. requires (IN DIR: needs) (ES DIR: requiere, necesita, exige) (IN INV: is_required, is_needed) (ES INV: es_requerido, es_necesitado, es_exigido)
 - 4.5. derivatives_of (ES DIR: deriva_de) (IN INV: is_derivated) (ES INV: es_derivado)
 - 4.6. method_of (ES DIR: tiene_metodo_de/para, tiene_modos_de/para)
5. is_a (ES DIR: es_un)

Tabla 4.9. Taxonomía de relaciones empleadas en OntoFIS (adaptación de la propuesta de Slaughter, Soergel y Rindflesch, 2006).

Tras los emparejamientos, la taxonomía de relaciones en OntoFIS está constituida por 36 tipos de asociaciones semánticas aplicadas. Aunque, en la estructura, se han mantenido las relaciones *full_related_to*, *functionally_related_to*, *temporally_related_to* y *conceptually_related_to* para conservar la organización taxonómica (tabla 4.9), éstas no se consideran relaciones efectivas al no haberse empleado para emparejar ningún concepto.

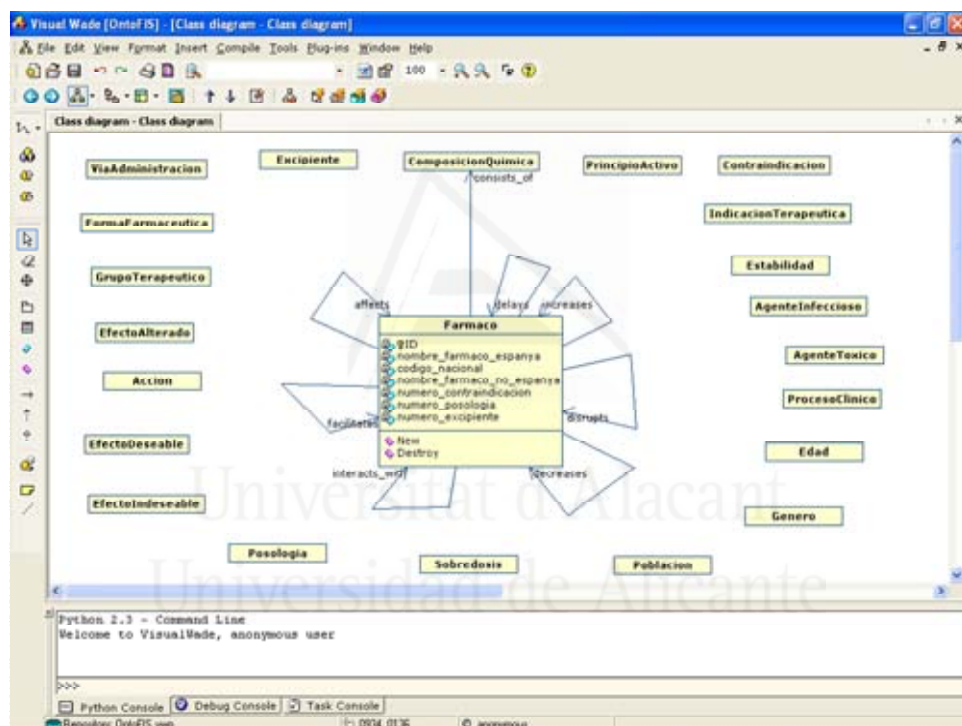
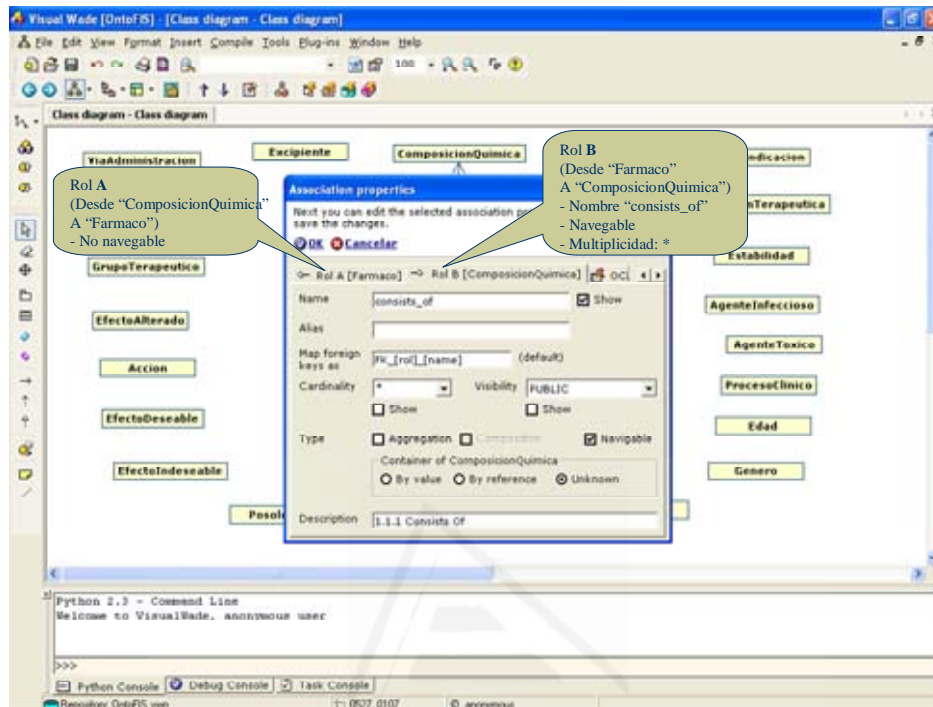


Figura 4.12. Desarrollo inicial del modelado de OntoFIS en CASE VisualWADE.

En la figura 4.12 se presenta una imagen del desarrollo inicial del diagrama de clases de OntoFIS. La figura 4.13 muestra la creación de una asociación entre dos conceptos en el diagrama de clases UML con la herramienta CASE VisualWADE. Mientras que, en la figura 4.14, se muestra un fragmento gráfico de las 4 clases principales de OntoFIS: 'Farmaco', 'FormaFarmaceutica', 'ViaAdministracion' y 'Contraindicacion', con un conjunto de relaciones declaradas entre ellas.



Nota Aclaratoria: Creación de la relación 'consists_of' entre la clase dominio Fármaco y la clase rango Composición Química.

Figura 4.13. Creación de una asociación en la herramienta CASE VisualWADE

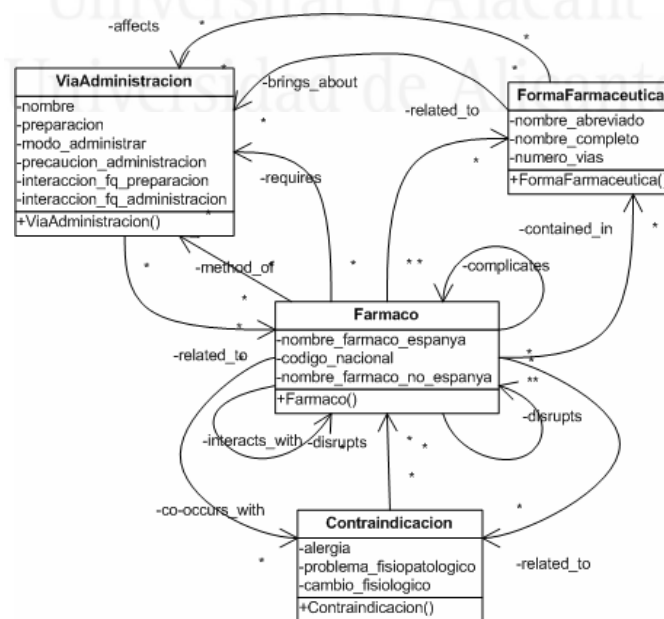


Figura 4.14. Fragmento del diagrama de clase de OntoFIS (representación no estándar).

4.4.3.1. El modelo de información de OntoFIS

El modelo de información de OntoFIS describe la realidad de los fármacos comercializados en España. Los fármacos comercializados son presentaciones que tienen un nombre particular y único, con él son prescritos y adquiridos. Su denominación se caracteriza por incluir su nombre de fantasía registrado, la concentración y la forma de presentación.

El esquema conceptual de OntoFIS describe un fármaco comercializado como un producto que contiene una composición química, definida por una concentración o cantidad de principio activo y de excipiente. Cada uno de ellos puede ser único o múltiple. El principio activo es la sustancia responsable de la acción de un fármaco. El resultado de la acción del principio activo produce tres tipos de efectos: deseable, alterado (interacciones) e indeseable (efectos adversos), pudiendo ser únicos o múltiples. El efecto deseable comprobado de un principio activo es el responsable de prevenir, curar o paliar uno o varios problemas de salud y constituye su indicación terapéutica. Los efectos indeseables y alterados son los que provocan o agravan otros problemas de salud. Los fármacos se ordenan en grupos terapéuticos, cuya acción deseable es común, pero con distintos efectos alterados e indeseables.

La forma de presentación de un fármaco comercializado determina la estabilidad y, también, las vías de administración.

La pauta de administración o posología (cantidad, frecuencia y duración) puede estar modificada por la edad, el sexo u otros problemas de salud. Una cantidad, frecuencia o duración excesiva provoca una sobredosis y un efecto alterado.

Un problema de salud es cualquier proceso clínico que pueda ser identificado como signo, síntoma, síndrome o cuadro patológico. También, puede ser causado por un agente infeccioso o tóxico.

4.4.4. Formalización del conocimiento farmacoterapéutico en OWL.

La formalización de la representación de OntoFIS, empleando el *plug-in Semantic Web Tools for VisualWADE*, se realiza automáticamente.

Transformación de Clases	
Elemento del diagrama de clase UML	Elemento de OWL DL
Clase Descripción de Clase Términos sinónimos y traducciones	owl: Class rdfs: comment rdfs: label
<pre> <owl: Class rdf: ID="http://www.dlsi.ua.es/ontofis.owl#Farmaco"> <!-- Disjoined with other classes --> <!-- Description of the concept in natural language --> <rdfs: comment xml:lang="es"> denominación comercial de sustancia química, simple o compuesta, empleada con una finalidad terapéutica... </rdfs: comment> <!-- Most representative terms and translations --> <rdfs: label xml:lang="es">especialidad farmacéutica</rdfs: label> <rdfs: label xml:lang="en">medicinal product</rdfs: label> </owl: Class> </pre>	
Transformación de Herencia	
Elemento del diagrama de clase UML	Elemento de OWL DL
Herencia	rdfs: subClassOf
<pre> <owl: Class rdf: ID="Accion"> <!-- Class Information --> </owl: Class> <owl: Class rdf: ID="EfectoDeseable"> <rdfs: subClassOf rdf: resource="#Accion"/> <!-- Class Information --> </owl: Class> </pre>	
Transformación de Atributo de Clase	
Elemento del diagrama de clase UML	Elemento de OWL DL
Atributo (dato simple)	owl: DatatypeProperty
<pre> <owl: DatatypeProperty rdf: ID="Farmaco..codigo_nacional"> <!-- Comments and translations --> <rdfs: Domain rdf: resource="#Farmaco"/> <rdfs: Range rdf: resource="&xsd:string"/> </owl: DatatypeProperty> </pre>	
Transformación del Rol de Asociación	
Elemento del diagrama de clase UML	Elemento de OWL DL
Rol de Asociación	owl: ObjectProperty
<pre> <owl: ObjectProperty rdf: ID="AS3..interacts_with"> <!-- Comments and translations --> <rdfs: Domain rdf: resource="#Farmaco"/> <rdfs: Range rdf: resource="#Farmaco"/> </owl: ObjectProperty> </pre>	

Tabla 4.10. Transformaciones entre UML y OWL DL realizadas por *Semantic Web Tools for VisualWADE* en la formalización de OntoFIS.

En la tabla 4.10 se muestran las transformaciones más significativas entre UML y OWL DL. El fragmento de código corresponde a las clases *Farmaco*,

Accion, *EfectoDeseable*, el atributo *codigo_nacional* de la clase *Farmaco* y la relación *interactss-with* entre las clases *Farmaco-Farmaco*. Es necesario recordar que las transformaciones se diseñaron y se aplicaron con un orden de incremento, comenzando con los mecanismos más sencillos y terminando con los de mayor complejidad.

En la propuesta de diseño y construcción de OntoFIS, la etapa de formalización del conocimiento capturado (pp 115-116) no termina hasta que se realiza una comprobación de la ausencia de errores en la codificación y se confirma la adecuación del contenido. Por la extensión, se ha preferido tratar todo el proceso de evaluación (verificación y validación) con sus consiguientes experimentos en el siguiente capítulo.

4.4.5. Conclusiones.

Como síntesis, en este capítulo se ha presentado: i) un análisis del modelo representado en las ontologías existentes y de las metodologías de construcción más frecuentemente aplicadas, estableciendo su correspondencia con la obtención de un modelo de datos conceptual; ii) se han propuesto los modelos de datos UML y OWL como metodologías para el modelado de OntoFIS; iii) se han descrito las funcionalidades de las aplicaciones tecnológicas empleadas para el diseño y la construcción de OntoFIS y iv) se ha determinado el proceso de captura, descripción y formalización de la Ontología Farmacoterapéutica e Información de Seguimiento hasta la construcción de OntoFIS en OWL DL.

Nuestra aportación reutiliza el esquema de representación, obtenido con la herramienta CASE VisualWADE, estableciendo una optimización del esfuerzo y tiempo de diseño. La extensión de las funcionalidades de los *plug-ins* de la herramienta permite reutilizar el esquema de la representación formalizado en el diagrama de clases UML. La incorporación del *plug-in Semantic Web Tools for*

VisualWADE facilita obtener, por un lado, una ontología de tarea para el enriquecimiento semántico de los recursos autogenerados y, por otro, una ontología de dominio independiente de la aplicación con la que se ha diseñado⁴⁰. El *plug-in* de transformación actúa como un centro de gestión de ontologías insertado en la herramienta CASE *VisualWADE*. Las funcionalidades más destacadas, que están desarrolladas en él, son:

- Multilingüidad y sinonimia: Permite introducir términos sinónimos y traducciones en todos los conceptos representados en el diagrama (figura 4.15).
- Descripciones de elementos: Facilita la incorporación de una descripción, en lenguaje natural, en cada uno de los elementos indicando el idioma en el que se expresa.
- Características de los roles de asociación: Incluye la posibilidad de indicar, en los roles de asociación característicos del diagrama de clases UML, las cuatro características más relevantes de las relaciones en OWL: funcionalidad, funcionalidad inversa, transitividad y simetría.
- Generación de un fichero con la ontología formalizada: Implementa un proceso de transformación automático, generando un fichero independiente en OWL DL, que facilita su reutilización para otras tareas (subapartado 4.3.1 y tabla 4.10).

⁴⁰ De acuerdo con la diferenciación de los tipos de ontología establecidos por Mizoguchi y equipo (1997), ver en pp. 26.

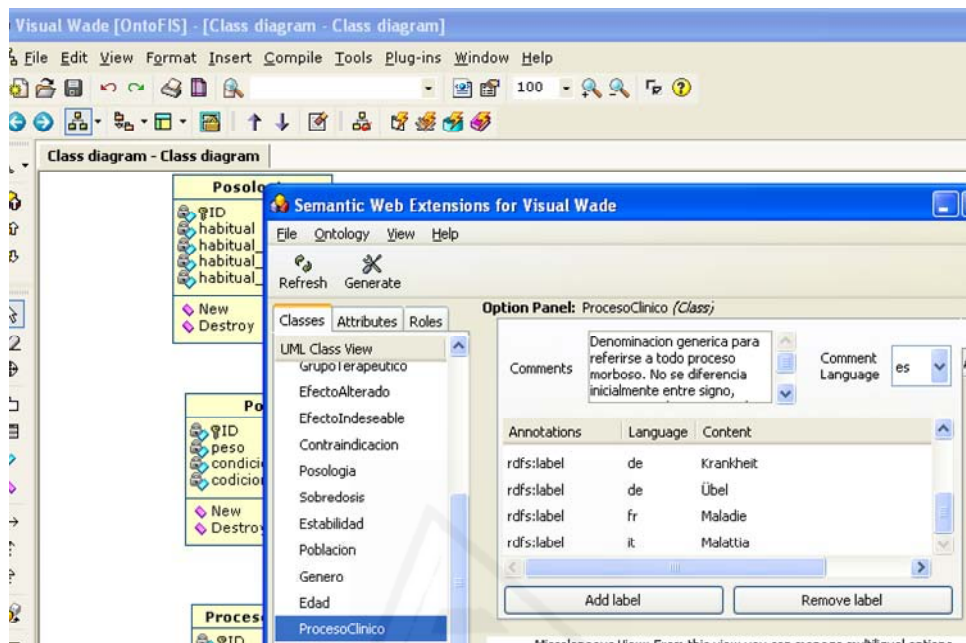


Figura 4.15. Interfaz del plug-in para la inclusión de etiquetas terminológicas

Otro aspecto importante en las transformaciones, ha sido el establecimiento de unas reglas de conversión de nombres, entre el modelado conceptual UML y la formalización OWL, para evitar los conflictos de los nombres idénticos. Mientras que OWL utiliza el URI para identificar cada uno de sus elementos de forma unívoca, en el diagrama de clases UML existen algunos elementos que pueden aparecer con el mismo nombre, como sería el caso de los atributos (el mismo nombre en clases diferentes) o los roles de asociación (el mismo nombre de rol en diferentes asociaciones). Como ejemplo, observar la conversión sufrida en el proceso de transformación por el atributo 'codigo_nacional' (*Farmaco..codigo_nacional*) y el rol de asociación 'interacts_with' (*AS3..interacts_with*) en la tabla 4.10. Todos estos aspectos serán tratados con más profundidad en el siguiente capítulo, debido a su relación con los resultados de la evaluación de OntoFIS.

Capítulo V.

Evaluación de OntoFIS

Una forma de realizar la evaluación de una ontología recién creada es usarla en la aplicación para la cual se diseñó (Noy y McGuinness, 2001). Pero, si ésta se lleva a cabo sin haber realizado una evaluación previa, sólo muestra la validez para su uso local y no acredita, objetivamente, el modelo representado para su utilización en otras aplicaciones o con otros propósitos.

El objetivo de este capítulo es evaluar el modelo conceptual de OntoFIS y comprobar que éste cumple con las funcionalidades previstas para una ontología, es decir: i) que proporciona una estructura de conocimiento del dominio, ii) que provee del conocimiento necesario para lograr inferencias significativas que concreten el significado de los términos contenidos y iii) que facilita la reutilización de la representación realizada.

Este capítulo se ha estructurado en cuatro apartados. En el primero se presenta la fundamentación del método de evaluación planificado, la definición de los indicadores utilizados para verificar y validar el modelo conceptual de OntoFIS y las fases del procedimiento aplicado. En el segundo, se describe el proceso de verificación desarrollado, los resultados obtenidos, las limitaciones detectadas y las soluciones adoptadas. En el tercer apartado, se expone el proceso de validación ejecutado y los experimentos realizados para determinar la adecuación del contenido de OntoFIS. En el último de los apartados, como conclusión del proceso de evaluación, se establece la acreditación de la ontología OntoFIS.

5.1. Método de evaluación.

En este trabajo, se entiende por acreditación el informe que establece la calidad del modelo representado de acuerdo a los indicadores de evaluación. Los indicadores son la forma particular en que se mide una condición considerada de calidad. En este sentido, existen distintas propuestas que tratan de establecer las condiciones a considerar y la forma de medirlas (Lozano, 2002; Gómez-Pérez et al, 2004; Tartir et al, 2005 y Obrst et al, 2007).

Lozano (2002) propone el método OntoMetric para establecer la calidad de ontologías candidatas para su incorporación en un sistema de información. Su método utiliza 160 indicadores de evaluación, agrupados en cinco dimensiones o condiciones, referentes a: i) el contenido y organización; ii) el lenguaje utilizado para la codificación; iii) la metodología de desarrollo, iv) la herramienta de software empleada para construir y editar la ontología y vi) el coste de construcción.

Gómez-Pérez y equipo (2004) establecen tres criterios para evaluar el modelado de conocimiento taxonómico. Estos criterios están asociados con los errores, más frecuentes, en la estructura de una taxonomía: i) la inconsistencia determinada por errores de circularidad, partición y semántica; ii) la clasificación incompleta (*incompletensse*) de conceptos por ausencia o errores de omisión de conocimiento en la partición; y, iii) la redundancia, tanto en las relaciones de subclases y de instancias como en la definición formal de clases y de instancias idénticas.

Tartir y equipo (2005) proponen un método de evaluación de la calidad de una ontología como fuente de información. Las dimensiones y los criterios que proponen son: i) el esquema conceptual: la riqueza de relaciones, atributos y herencia; ii) la base de conocimiento: la riqueza de clases, distribución media de instancias o de población y cohesión; y, iii) las clases: importancia, plenitud, riqueza de herencia y relaciones, conectividad y legibilidad.

Obrst y equipo (2007) centran su punto de vista en los requisitos, para la interoperabilidad semántica y la integración de ontologías, en el dominio biomédico. La tipología de su propuesta es: i) la representación del conocimiento en relación al modelo de datos, los conceptos y las instancias; ii) los requerimientos del dominio, de acuerdo a las cuestiones de competencia; iii) el acuerdo semántico inter-anotador para la clasificación de instancias y, iv) la similitud y la distancia semántica entre los conceptos de dos o más ontologías.

En conjunto, hay una gran variedad de indicadores, tanto cualitativos como cuantitativos, con una falta de acuerdo sobre qué elementos evaluar y qué tipo de indicador emplear para determinar, objetivamente, la calidad de la ontología. Todo ello se explica por las distintas perspectivas que adoptan los autores para abordar el concepto de calidad. Aunque, predominan dos aspectos comunes para medir la calidad:

- La formalización de la construcción, relativa al modelo de datos y la organización jerárquica.
- El dominio representado, relativo al vocabulario empleado y la aplicabilidad de las inferencias.

Estos aspectos comunes sobre la calidad coinciden con la finalidad de la evaluación de una ontología. La finalidad es comprobar el modelo en cuanto a su estructura (verificación) y la adecuación del contenido representado (validación) (Gómez-Pérez, 1996). El objetivo de la verificación es asegurar la libertad de errores o de defectos de la construcción, mientras que, el objetivo de la validación es comprobar que las definiciones de la ontología modelan, lo más exactamente posible, el dominio para el cual fueron creadas.

Tomando en consideración todo lo expuesto, en los siguientes subapartados, se indican los indicadores de calidad establecidos y el procedimiento desarrollado para determinar la calidad de OntoFIS.

5.1.1. Indicadores de calidad.

Los indicadores de calidad propuestos se agrupan de acuerdo al tipo de evaluación aplicado al modelo de OntoFIS:

1. La evaluación de verificación tiene por objetivo determinar la ausencia de errores en el modelo formalizado. Los indicadores son:
 - Respecto al lenguaje de codificación OWL DL tras la transformación del modelo conceptual en UML:
 - La ontología formalizada está libre de errores sintácticos y semánticos.

- La ontología formalizada contiene todos los componentes especificados en el modelo conceptual.
 - Respecto a la estructura organizativa:
 - La jerarquía taxonómica diseñada coincide con la jerarquía deducida y es consistente, al carecer de errores de circularidad y de errores de partición.
2. La evaluación de validación tiene por objetivo determinar el grado de adecuación del contenido para satisfacer el alcance y el propósito de la ontología.
- Respecto al vocabulario empleado:
 - Los conceptos explicitados son significativos y completos en relación al dominio.
 - Respecto a la aplicabilidad de la inferencia:
 - La representación formalizada logra inferencias correctas y completas respecto al dominio.

5.1.2. Procedimiento de evaluación.

La secuenciación de pasos que se llevaron a cabo para la evaluación fueron:

1. Verificación de la estandarización del código generado.

Para esta actividad se utilizó el editor de ontologías Protégé. Se exportó el fichero y se analizó la codificación, respecto a la ausencia de errores sintácticos y semánticos del código. También, se comprobó la presencia de todos los componentes modelados.

2. Verificación de la estructura organizativa.

Para esta actividad se empleó el razonador FaCT++ para confirmar la ausencia de fallos en la estructura jerárquica modelada.

3. Análisis de las limitaciones detectadas de acuerdo con los resultados obtenidos de las dos actividades anteriores.

4. Aplicación de soluciones.

5. Verificación de la soluciones.

6. Validación del contenido.

Para llevar a cabo esta actividad se pobló OntoFIS y se construyó un banco de cuestiones. El banco de cuestiones estaba constituido por los patrones de preguntas, más frecuentemente realizadas por los profesionales de la salud, en relación a la atención farmacoterapéutica. Se establecieron las hipótesis a confirmar en los experimentos.

7. Primer experimento.

Se estableció el procedimiento de ejecución, que incluía una muestra de preguntas seleccionadas del banco de cuestiones. La muestra de cuestiones se transformó, manualmente, en sintaxis DL Query empleando los conceptos de OntoFIS. La finalidad de este experimento era detectar los vacíos del vocabulario de OntoFIS respecto a las necesidades de los potenciales usuarios.

8. Segundo experimento.

El propósito de este experimento era comprobar la exactitud de las inferencias. Se usó el motor de consulta DL Query 1.0 para interrogar a la ontología poblada. Previamente, se estableció un grupo de medicamentos diana para el control de las respuestas correctas y completas y para determinar los índices de precisión y de exhaustividad (*recall*).

9. Análisis de las respuestas incorrectas.

Se realizó una comprobación manual en el modelo conceptual para localizar los fallos de inferencia. Se localizaron las relaciones de asociaciones semánticas que generaban el conflicto y se eliminaron los bucles no detectados anteriormente.

10. Tercer experimento.

Se comprobó la eficacia del refinamiento realizado, a partir de los índices de precisión y exhaustividad obtenidos.

11. Acreditación de OntoFIS.

Se estableció el informe de valoración de calidad de OntoFIS de acuerdo con los resultados de evaluación.

5.1.3. Herramientas para la evaluación.

Como se ha indicado en el procedimientos de evaluación, las herramientas utilizadas para la verificación de la ontología fueron el editor de ontologías Protégé y el razonador FaCT ++. Mientras que, para la validación se usó el motor de consulta DL Query. En los siguientes subapartados se realiza una breve presentación de las funcionalidades que motivaron su selección.

5.1.3.1. Protégé.

El editor de ontologías Protégé⁴¹ es una plataforma libre, de código abierto, desarrollada y mantenida por el *Center for Biomedical Informatics Research (BIMR)* de la Universidad de Stanford. Cuenta con una extensa comunidad de usuarios y con un conjunto de herramientas para construir

⁴¹ <http://protege.stanford.edu/>

modelos de dominio y aplicaciones basadas en conocimiento con ontologías.

Protégé implementa, en su núcleo, un conjunto de conocimientos de modelado de estructuras y acciones que apoyan la creación, la visualización y la manipulación de ontologías en distintos formatos de representación. Protégé se puede ampliar por medio de una arquitectura de plug-in y de la API (*Application Programming Interface*) de Java para la construcción de herramientas y aplicaciones basadas en conocimiento.

La plataforma Protégé soporta dos formas principales de modelado de ontologías:

- El editor de Frames de Protégé, que permite a los usuarios construir y poblar las ontologías, que están basadas en frames, de conformidad con el protocolo OKBC (*Open Knowledge Base Conectividad*).
- El editor OWL de Protégé que permite a los usuarios construir ontologías para la Web Semántica, en particular para la especificación OWL. La semántica formal de OWL especifica cómo derivar sus consecuencias lógicas, es decir, hechos no explícitos, literalmente, en la ontología, sino implícitos en la semántica. Estas implicaciones pueden estar basadas en un documento único o múltiples documentos distribuidos, y combinados mediante mecanismos OWL definidos.

El editor OWL de Protégé⁴² permite: i) cargar y guardar ontologías OWL y RDF; ii) editar y visualizar las clases, propiedades y reglas SWRL (*Semantic Web Rule Language*)⁴³; iii) definir las características lógicas de clase como expresiones OWL; iv) ejecutar razonamientos como la clasificación de descripción lógica; y, v) editar OWL individuales para el marcado de la Web Semántica.

⁴² Se utilizó, inicialmente, la versión 3.2.1 y, posteriormente, se pasó a la versión 3.3.1.

⁴³ <http://www.w3.org/Submission/SWRL/>

5.1.3.2. FaCT++.

El razonador FaCT++⁴⁴ es una herramienta imprescindible para extraer conocimiento útil de los conceptos definidos en una ontología e inferir nuevas relaciones o conceptos no explícitos en el modelo original. FaCT++ es una versión avanzada del clasificador de Lógica Descriptiva (DL) FaCT y está implementado en C++. Las ventajas de este razonador son su licencia *GNU General Public License* y, recientemente, su incremento en la capacidad de inferencia para dar soporte al nuevo estándar OWL1.1⁴⁵. Además, trabaja de forma eficiente con la T-Box o el conocimiento declarado de las propiedades de los conceptos y las relaciones. Entre sus funcionalidades están: i) el inferir las superclases de una clase (clasificación), ii) el determinar si una clase es o no consistente (una clase es inconsistente si no puede tener instancias), iii) el determinar si una clase es subclase de otra (subsumción) y iv) el construir la taxonomía (procesar tanto las relaciones de subclase como aquellas que no son explicitadas pero están implicadas en las definiciones dadas).

FaCT++⁴⁶ está incluido como un *plug-in* en el editor OWL de Protégé 3.3.1⁴⁷. Estas herramientas, como marco de prueba, facilitaban una verificación automática de la adecuación del código OWL-DL generado con el *plug-in Semantic Web Tools for VisualWADE* y la confirmación de la ausencia de fallos en la estructura desarrollada en OntoFIS.

⁴⁴ <http://owl.man.ac.uk/factplusplus/>

⁴⁵ Incluye tres nuevas características de las relaciones respecto a la versión OWL 1.0: la asimétrica, la reflexiva y la irreflexiva.

⁴⁶ El razonador debe ser lanzado antes de la comprobación, especificando el puerto donde Protégé estará esperándolo.

⁴⁷ Era la versión más estable y orientada al estándar OWL 1.0, al igual que el *plug-in* de transformación aplicado.

5.1.3.3. DL Query.

El motor de consulta DL Query 1.0 (build 3) está basado en lógica y orientado a la interrogación de ontologías. Este motor aporta una mayor facilidad y agilidad en la interrogación de ontologías pobladas, soportando un gran número de instancias. Además, se encuentra integrado en el editor Protégé versión 4 beta que, en el momento de proceder a la validación de contenido de OntoFIS, tenía sus funcionalidades estables y se adaptaba al proceso de evaluación previsto.

Su funcionalidad permitía comprobar que las inferencias realizadas, a partir del modelo declarado, facilitaban respuestas correctas y no habían pérdidas de información respecto al conocimiento del dominio farmacoterapéutico.

5.2. Evaluación de verificación.

El documento OWL DL, generado con el *plug-in Semantic Web Tools for VisualWADE*, fue exportado al editor Protégé. La carga de la ontología por este editor aseguraba que el código generado cumplía con el estándar. Por otra parte, se probaba la ausencia de errores de semántica y sintaxis, al menos inicialmente, ya que se podía verificar que la ontología estaba completa respecto al contenido modelizado, previamente, en la herramienta CASE VisualWADE. Es decir, que:

- contenía las 23 clases representadas en el diagrama de clase UML (figura 5.1),
- se recogía la descripción en lenguaje natural de todas clases,

- se conservaban las propiedades extrínsecas en cada clase y las relativas a las distintas etiquetas terminológicas, reseñadas en español, inglés, francés, italiano y alemán, que incluyen los sinónimos declarados,
- se identificaban las relaciones entre cada par de conceptos,
- se mantenían todas las propiedades intrínsecas en cada clase (figura 5.2), y
- se identificaban las asociaciones especificadas (figura 5.3).

No se pudo comprobar la consistencia de la estructura jerárquica. La estructura de OntoFIS visualizada en Protégé era un listado de clases (figura 5.1), al igual que ocurría con las asociaciones semánticas establecidas (figura 5.3).

La explicación de las causas de estos vacíos, se argumenta en el siguiente subapartado.

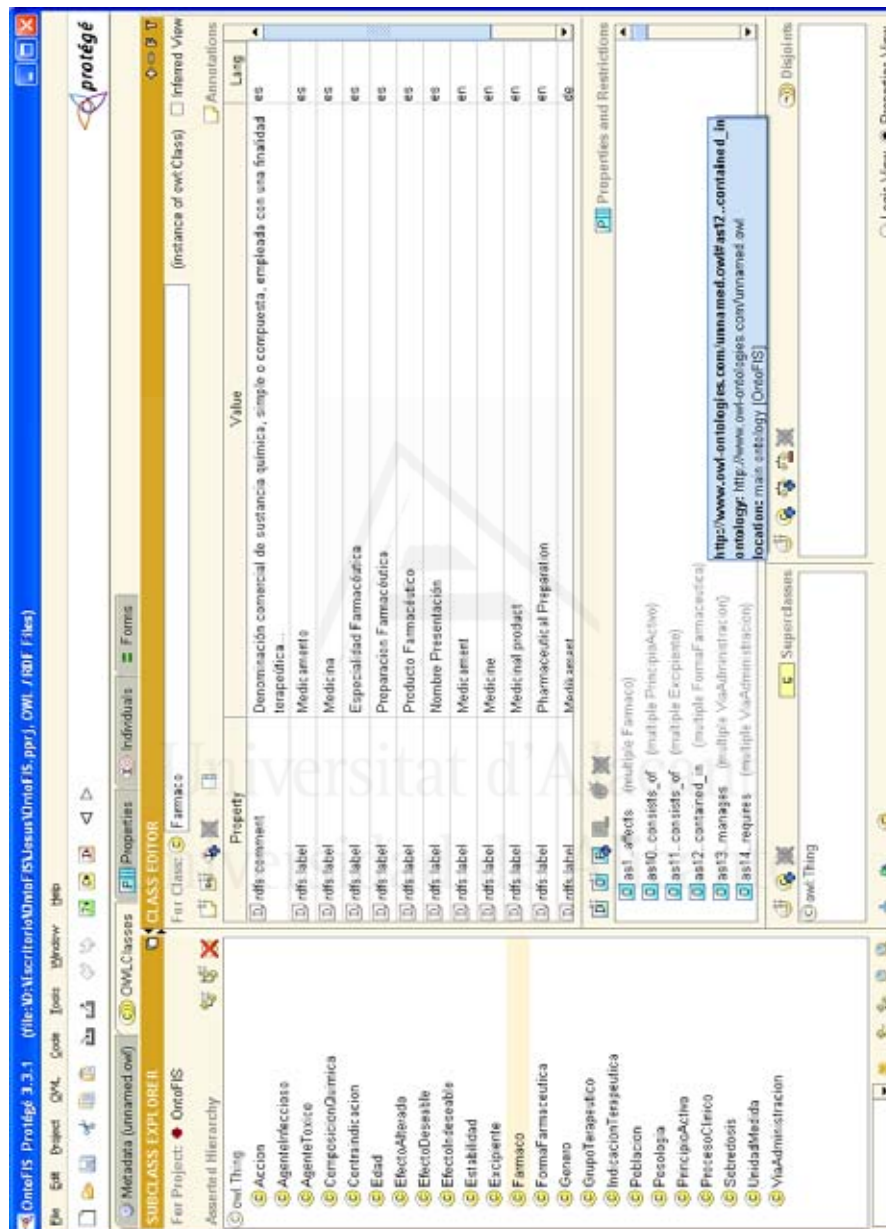


Figura 5.1. Visualización de las clases de OntoFIS en la ventana principal de Protégé.

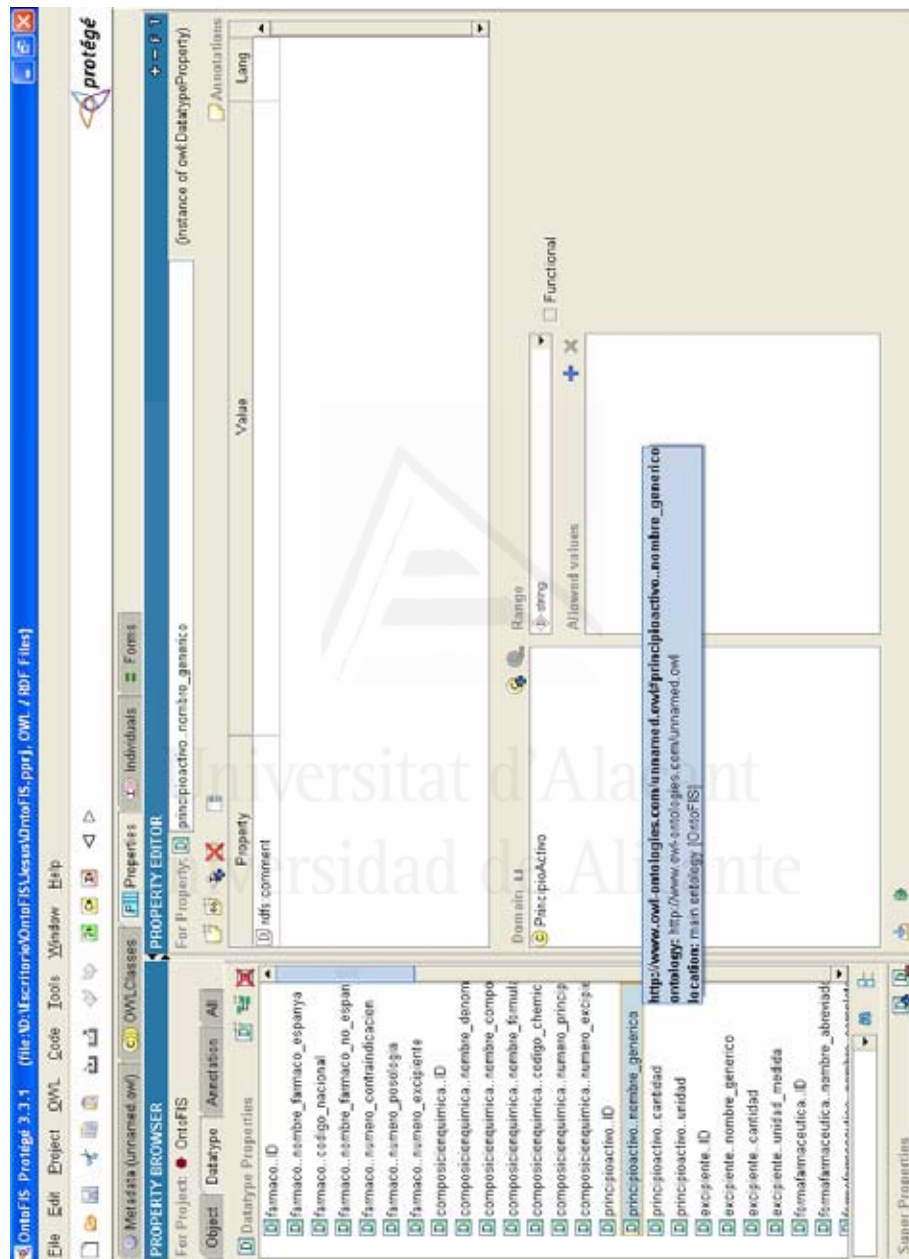


Figura 5.2. Visualización de las propiedades de las clases de OntoFIS (Datatype property) en Protégé.

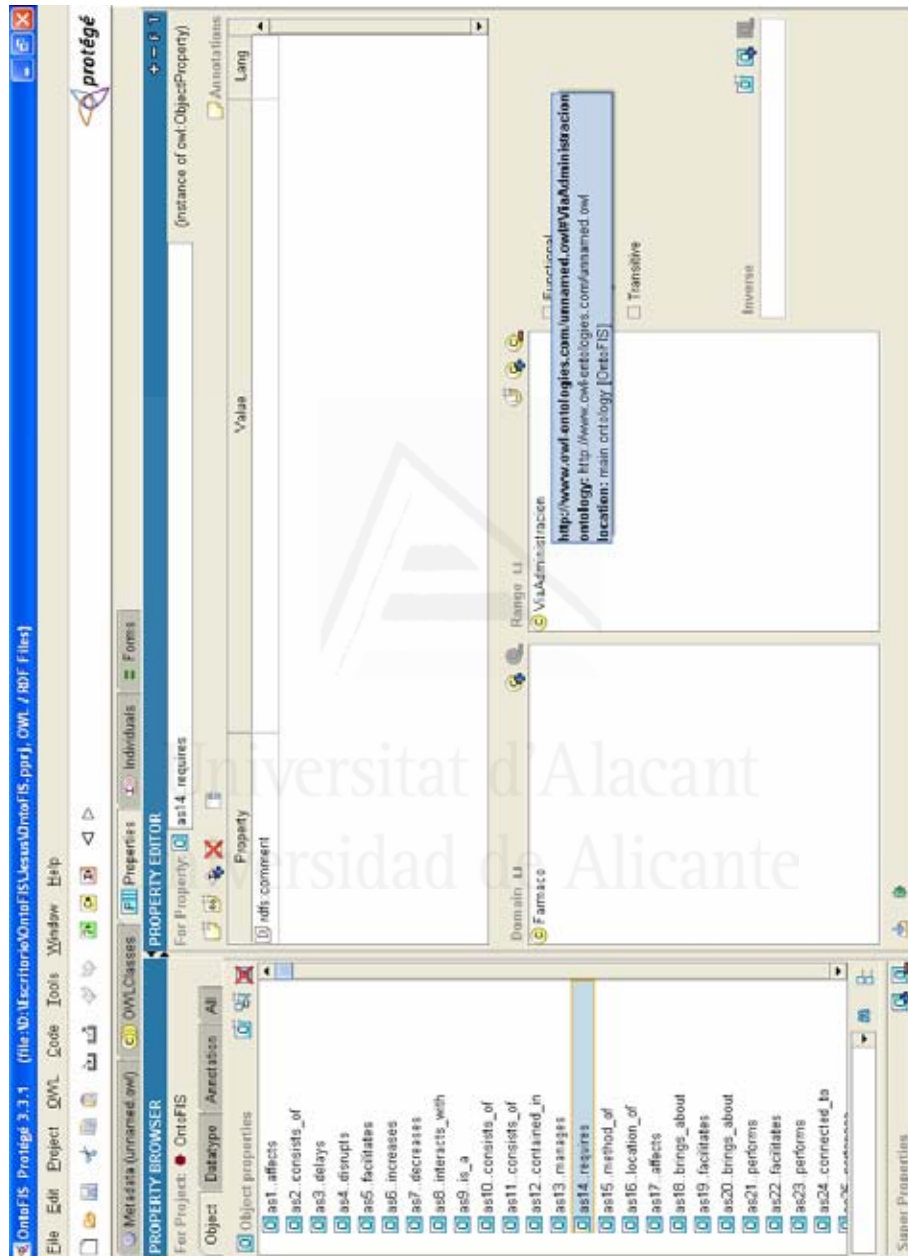


Figura 5.3. Visualización de las relaciones entre clases de OntoFIS (Object property) en Protégé.

5.2.1. Causas de las limitaciones detectadas.

Uno de los problemas se encuentra en el proceso de conversión de nombres entre el modelado conceptual UML y la formalización en OWL DL. El algoritmo anticipa que puedan aparecer algunos elementos con el mismo nombre en el diagrama de clases UML, lo que puede generar un conflicto de identificación. La regla establece utilizar un URI para identificar de forma unívoca cada elemento y evitar los nombres idénticos.

Cuando esta regla de transformación se aplica a los roles de asociación no se resuelve, satisfactoriamente, la identificación de éstos. El algoritmo va actuando 'rol' por 'rol'. Toma un 'rol' en el modelado de UML y lo renombra empleando el patrón 'NomAsociacion..NomRol'. El renombramiento incluye un numerador secuencial que hace que una misma asociación semántica se transforme en tantas como número de veces aparece. Por ejemplo, a partir del rol 'consists_of' en el modelado de UML se generan diferentes relaciones en OWL DL. El 'rol' es renombrado con un numerador secuencial que hace que se generen diferentes asociaciones como 'as10..consists_of' y 'as11..consists_of' (ver en el apartado 4.3.1, clases relacionadas mediante una relación asociativa, pp 107). El resultado conlleva una identificación unívoca pero, produce pérdidas de información para la inferencia, sobre todo, para la deducción de información a través de la propiedad transitiva (figura 5.4).

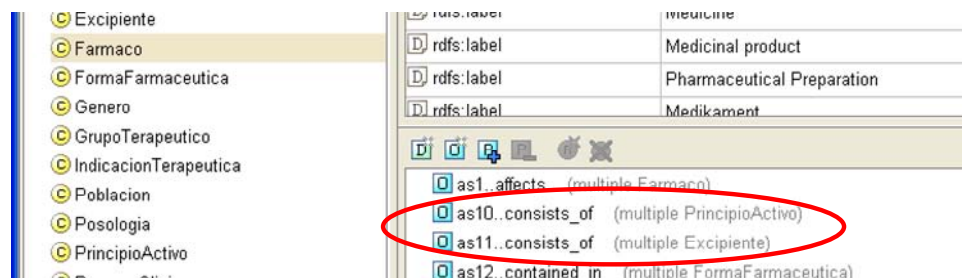


Figura 5.4. El renombrado de las asociaciones con numerador secuencial.

En el caso de OntoFIS, el documento OWL DL no contiene los 36 tipos de relaciones semánticas establecidos para definir las 647 asociaciones entre clases de la representación. El proceso de transformación hace que el documento OWL DL tenga 647 tipos diferentes de relaciones semánticas.

La solución propuesta sería rediseñar la regla para renombrar el nombre del rol en función al modelo de datos RDF. Una sentencia en RDF está constituida por una tripleta compuesta por tres constituyentes (sujeto-predicado-objeto), si uno de ellos cambia, el significado de la sentencia cambia. Por otra parte, varios recursos pueden ser identificados con la misma URI.

Así, el patrón de referencia para establecer la sentencia que define un ‘rol’ entre clases, en UML, sería “ClaseDominio-ObjectProperty(Rol)-ClaseRango” y que, respectivamente, serían “sujeto-predicado-objeto” para la sentencia en OWL. En esta propuesta, se mantiene la regla de denominación aplicada a las clases dominio y rango en la transformación entre UML y OWL DL. Estas transformaciones no plantean inconvenientes para la identificación unívoca.

Sin embargo, se debe cambiar la denominación del ObjectProperty a partir del nombre del ‘rol’. Hay dos alternativas para establecer la nueva denominación. Una alternativa es hacer un procesamiento posterior, una vez renombrados todos los roles. La otra opción es actuar, directamente, en la forma de renombrar.

En el primer caso, el procedimiento sería:

1. Una vez renombrados todos los roles, de acuerdo con la regla establecida y el patrón ‘NomAsociacion..NomRol’, se pasaría a comparar entre ellos:
 - 1.1. Todos los ObjectProperty cuya denominación es igual en la parte del patrón ‘NomRol’ se vuelven a renombran. La nueva denominación es, directamente, NomRol y se borra el NomAsociacion (el numerador secuencial).

Ejemplo: ‘as10..consists_of’ y ‘as11..consists_of’ pasarían a ser renombrados como ‘consists_of’.

En el segundo caso, el procedimiento sería:

1. La primera sentencia de referencia es la sentencia constituida por el primer ‘rol’ leído, con su correspondiente clase dominio y clase rango. El nombre de ObjectProperty es equivalente al nombre del ‘rol’.

Ejemplo: ‘Farmaco’-‘consists_of’-‘PrincipioActivo’

2. La siguiente sentencia se constituye a partir de:
 - 2.1. Si el ‘rol’ leído es distinto del nombre de ObjectProperty de una sentencia de referencia, entonces, se compone una nueva sentencia de referencia. El nombre de ObjectProperty es equivalente al nombre del ‘rol’ leído. La sentencia se almacena como sentencia de referencia.

Ejemplo: ‘Farmaco’-‘affects’-‘Farmaco’

- 2.2. Si el ‘rol’ leído es igual al nombre de ObjectProperty de una sentencia de referencia, entonces, se comparan los nombres de las clases dominio de las sentencias de referencia y la clase dominio candidata. Si la clase dominio candidata es distinta se compone una nueva sentencia de referencia. El nombre de ObjectProperty sigue siendo el nombre del ‘rol’.

Ejemplo: ‘ProcesoClinico’-affects’-‘Accion’

- 2.3. Si el ‘rol’ leído es igual al nombre de ObjectProperty y la clase dominio candidata es igual en una misma sentencia de referencia, entonces, se comparan las clases rango de las sentencias de referencia y la clase rango candidata. Si la clase rango candidata es distinta se compone una nueva sentencia de referencia. El nombre de ObjectProperty seguirá siendo el nombre del ‘rol’.

Ejemplo: ‘Farmaco’-‘consists_of’-‘Excipiente’

- 2.4. Si el ‘rol’ leído es igual al nombre de ObjectProperty y la clase dominio candidata como la clase rango candidata son idénticas en una misma sentencia de referencia, entonces, la sentencia ya existe y no se constituye una nueva sentencia de referencia.

La diferencia entre ambas alternativas es que, la segunda de ellas, permite verificar que no ha habido un error en el diseño del modelo conceptual. En modelos caracterizados por una amplia riqueza asociativa, la segunda alternativa puede servir para detectar duplicidades. Aunque, la primera de las alternativas supone una modificación menor del algoritmo actual.

Otra limitación encontrada en el algoritmo, del plug-in *Semantic Web Tools for VisualWADE*, es que las clases hermanas no están etiquetadas como disjuntas entre ellas. El inconveniente es la imposibilidad de detectar el error de haber declarado que un individuo puede ser instancia, simultáneamente, de dos clases hermanas. Esto supone un problema, tanto para el razonamiento derivado como para la instanciación como para la verificación de consistencia por partición. Una propuesta de solución es incluir la etiqueta *owl:disjointWith* en el proceso de transformación de las clases relacionadas mediante una composición o bien por agregación (apartado 4.3.1, pp 107-108).

Finalmente, se detectó un inconveniente en la relación taxonómica *is_a*, por su inclusión en el diagrama de clases como un rol. Este error en el modelado explica el problema de la deficiencia de la estructura organizativa y su visualización como un listado.

5.2.2. Soluciones aplicadas.

Como solución se reorientó la estrategia de obtención de la ontología, mediante la optimización de la dedicación, en cuanto al esfuerzo y tiempo. La solución aplicada mantenía la reutilización del diagrama de clases UML, pero sólo hasta el punto en que las descripciones impuestas por la herramienta CASE VisualWADE y las transformaciones realizadas por el *plug-in* no planteasen conflictos con el modelo de datos OWL DL. Las actividades desarrolladas fueron:

1. Se modificó el diagrama de clase UML eliminando todos los roles. Las relaciones de tipo *'isa'* se definieron como generalizaciones. Además, las relaciones de tipo *'part_of'* se redefinieron en el diagrama como herencia entre clases.

Esta decisión conllevaba mantener la utilidad del diagrama para generar las aplicaciones web con VisualWADE y un etiquetado semántico básico, basado en una organización taxonómica elemental.

La referencia dada a este diagrama de clase fue *'diagrama para la transformación OntoFIS'*.

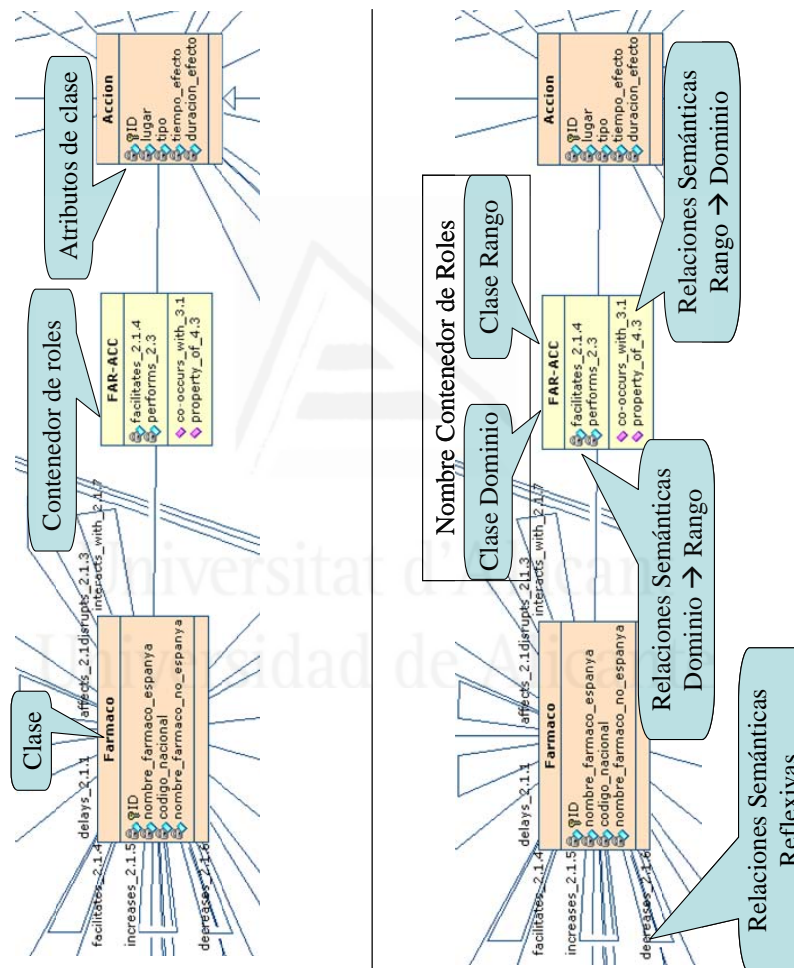
2. Se creó un nuevo diagrama de clases, con todas las asociaciones pero mejorando la visualización, para facilitar el seguimiento de las relaciones de asociación definidas entre las clases.

El diagrama de clases UML original era ininteligible, visualmente, por el número de roles establecidos (cerca de 600).

Para hacer el diagrama más usable se incluyó una identificación por colores, diferenciando entre las clases y un nuevo elemento propio de este diagrama, el *'contenedor de roles'*. Este último elemento se emplea para especificar las relaciones de asociación entre las clases (figura 5.5 y Anexo V).

Las relaciones de tipo reflexivo se mantuvieron en su representación original.

Este diagrama se denominó ‘*diagrama de descripción de OntoFIS*’ y contenía el modelo conceptual completo a formalizar.



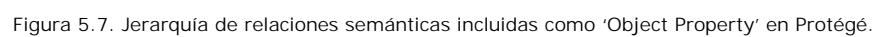
Nota Aclaratoria: Los ‘contenedores de roles’ se denominan con las tres primeras letras en mayúsculas de la clase dominio separadas, por un guión, de las tres primeras letras de la clase rango. Los roles de asociación entre la clase dominio y rango se sitúan en la parte superior y, en la parte inferior del contenedor, se sitúan los roles entre la clase rango y dominio.

Figura 5.5. Fragmento del nuevo ‘*diagrama de descripción*’ de OntoFIS.

‘PrincipioActivo’-‘Excipiente’ y entre las clases ‘EfectoAlterado’-‘EfectoIndeseable’-‘EfectoDeseable’. De acuerdo con el *‘diagrama de descripción de OntoFIS’*, las clases ‘PrincipioActivo’ y ‘GrupoTerapeutico’ no son disjuntas respecto a la clase ‘Farmaco’. Un medicamento debe poder ser clasificado, indistintamente, en cada uno de dichos conceptos por la relevancia que tiene para la práctica clínica, a pesar de la posible redundancia.

4. Las relaciones de asociación semántica fueron, nuevamente, creadas en Protégé 3.3.1, tal como se indicaba en el *‘diagrama de descripción de OntoFIS’*. En Protégé, las relaciones se crearon como ‘Object Property’, ya que este editor sí soporta el tratamiento de las relaciones como entidades y su estructuración jerárquica, así como la incorporación de etiquetas terminológicas en inglés y en castellano (figura 5.7).

Con estas soluciones se logra la finalidad de obtener un recurso de conocimiento terminológico del dominio farmacoterapéutico independiente, con dos tipos de utilidades directas. Por una parte, el recurso amplía el significado del etiquetado semántico de los recursos web, creados con VisualWADE. Por otra, se obtiene un recurso para posibles tareas relacionadas con PLN.



5.2.3. Verificación de las soluciones aplicadas.

A partir de las soluciones aplicadas, se realizó una verificación del modelo representado. Para ello se empleó el razonador FaCT++, el cual verificó la ausencia de inconsistencia de distintos tipos:

- errores de partición, es decir, se verificó que las clases y las subclases, por grupos disjuntos, no contenían ni denominación ni propiedades comunes,
- errores de circularidad en la estructura jerárquica, es decir, no había ninguna clase definida como generalización y especialización de sí misma,
- la redundancia mostrada con la clase ‘Farmaco’ respondía, como se ha indicado, a las características del propio conocimiento del dominio y,
- la estructura jerárquica inferida por el razonador coincidía con la estructura jerárquica exportada (figura 5.8).

Esta última prueba, realizada con FaCT++, confirmaba los siguientes requisitos de calidad OntoFIS: la ausencia de errores semánticos y sintácticos en el código OWL DL y la consistencia de la estructura jerárquica del modelo conceptual de la ontología.

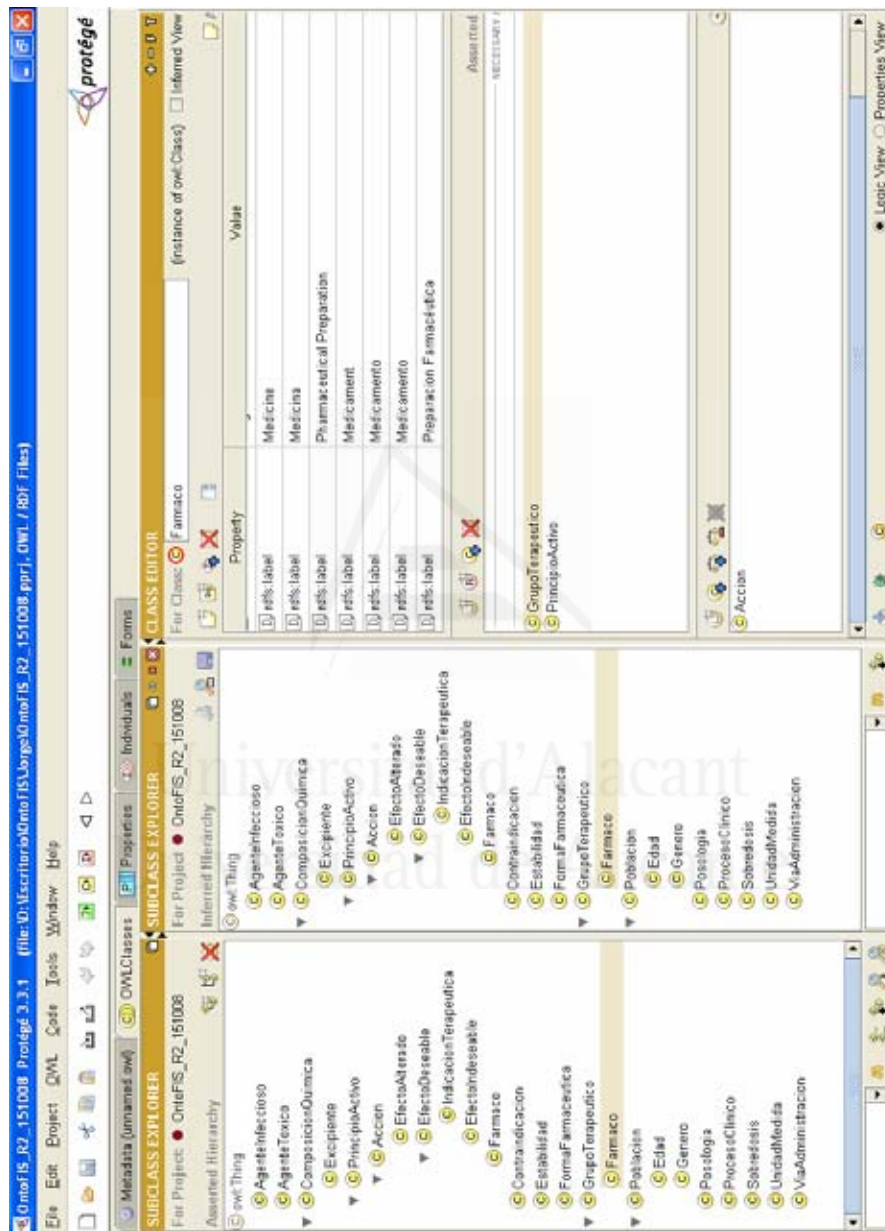


Figura 5.8. Jerarquía diseñada y jerarquía inferida por FaCT++ en Protégé.

5.3. Evaluación de Validación.

Como procedimiento para evaluar lo completo y lo consistente que es el vocabulario de la representación, respecto al conocimiento del dominio, se utilizó la propuesta de Grüninger y Fox (1995) de emplear ‘cuestiones de competencia’ (*competency questions*).

De acuerdo con estos autores, “las cuestiones de competencia se expresan en lenguaje natural y ellas determinan las respuestas que facilitará la ontología de acuerdo con el alcance especificado” (Grüninger y Fox, 1995). Sin embargo, la diferencia fundamental estriba en la forma de obtener estas cuestiones. Estos autores proponen elaborar, como segunda tarea de su metodología de construcción, un conjunto de cuestiones de competencia para determinar las opciones de respuesta y mantener estas cuestiones como un patrón de control del proceso. En cambio, en este trabajo, las ‘cuestiones de competencia’ son las cuestiones reales que se plantean los profesionales de la medicina durante la atención clínica, uno de los grupos potenciales de usuarios de la representación de conocimiento realizada.

Este criterio de obtención de las cuestiones de competencia asegura un control externo al diseñador para confirmar lo significativo, lo completo y lo consistente que es el vocabulario capturado y el modelado realizado.

Se establecieron dos hipótesis para asegurar la adecuación del contenido representado:

1. El modelo de conocimiento de OntoFIS se consideraría completo si los términos incluidos en las cuestiones de competencia tuvieran un concepto representativo en la ontología. La medida para determinar lo incompleto del modelo sería el número total de términos sin representación.

2. El modelo de conocimiento de OntoFIS se consideraría consistente si las respuestas inferidas fueran correctas y completas respecto al conocimiento del dominio farmacoterapéutico. La consistencia se mediría con los índices de precisión y exhaustividad.

En los siguientes apartados presentamos las características del banco de ‘cuestiones de competencia’ y de la base de conocimiento que se creó para comprobar estas hipótesis.

5.3.1. Banco de cuestiones de competencia.

Se construyó un banco de cuestiones a partir de los patrones de preguntas relacionadas con la atención farmacoterapéutica que, más frecuentemente, se plantean los profesionales médicos durante la actividad asistencial. Estos patrones de preguntas se obtuvieron de los trabajos de Ely y equipo (1999) y González-González y equipo (2005) que facilitaban cuestiones, cultural y sanitariamente, diferenciadas (España y EEUU).

El banco de cuestiones de competencia contiene 94 patrones de preguntas, tanto en inglés como en castellano. Aunque, las cuestiones sólo fueran traducciones, también se consideraron cada una de ellas como una expresión diferenciada. El banco de cuestiones se organizó en 12 categorías (tabla 5.1) equivalentes al nivel secundario y al terciario de la clasificación de González-González y equipo (2005) basada, a su vez, en la tipología de Ely y equipo (1999). Esta categorización recoge todos los aspectos involucrados en la atención farmacoterapéutica.

Prescripción (2.1):

Cómo prescribir (2.1.1)	= Número de cuestiones: 16
Eficacia/Indicaciones/Fármaco de elección (2.1.2)	= Número de cuestiones: 20
Efectos adversos (2.1.3)	= Número de cuestiones: 19
Interacciones (2.1.4)	= Número de cuestiones: 4
Nominación (2.1.5)	= Número de cuestiones: 2
Orientación/composición (2.1.6)	= Número de cuestiones: 6
Características físicas (2.1.7)	= Número de cuestiones: 8
Farmacodinamia o absorción (2.1.8)	= Número de cuestiones: 6
Mecanismo de acción (2.1.9)	= Número de cuestiones: 2
Coste (2.1.10)	= Número de cuestiones: 5
Niveles séricos (2.1.11)	= Número de cuestiones: 4
Disponibilidad (2.1.12)	= Número de cuestiones: 2

Nota Aclaratoria: el código, incluido entre paréntesis, responde a los niveles secundario y terciario de la clasificación de González-González et al (2005).

Tabla 5.1. Categorías y número de cuestiones del banco de preguntas.

5.3.2. Población de OntoFIS.

La inclusión de instancias para poblar OntoFIS se realizó en dos fases. En la primera, se utilizó un método semi-automático para extraer las instancias de la base de datos 'Digitalis'⁴⁸. Esta base de datos contiene todas las especialidades farmacéuticas utilizadas en el Sistema Nacional de Salud para el inventario de medicamentos en hospitales y farmacias.

Se estudió, previamente, la estructura de 'Digitalis' y se establecieron las correspondencia entre los campos de las tablas y los conceptos de OntoFIS (tabla 5.2).

Las instancias a incluir en el concepto 'Fármaco' tuvieron que ser filtradas previamente, ya que la base de datos 'Digitalis' incluye otros productos sanitarios cubiertos por el sistema de salud pero, sin la condición de medicamento (vendas, tobilleras, sondas, bolsas colectoras...).

⁴⁸ Ministerio de Sanidad y Consumo, versión de septiembre de 2008.

Las correspondencias facilitaban algo más de 50.000 instancias para los conceptos ‘Farmaco’, ‘ComposicionQuimica’, ‘PrincipioActivo’, ‘GrupoTerapeutico’, ‘FormaFarmaceutica’, ‘ViaAdministracion’ y ‘UnidadMedida’.

Fármaco	
nombre_farmaco_espanya	NESPECIFICO (DI0041)
codigo_nacional	CODNACIONAL (DI0041)
ComposicionQuimica	
grupo_quimicoOMS	DESCRIPCION (DI0034)
codigo_quimicoOMS	CODSUBGRUPO (DI0034)
PrincipioActivo	
nombre_generico	DESCRIPCION (DI0033)
cantidad	DOSISUNIDAD (DI0041)
unidad	DESCRIPCION (DI0037) a partir de CODUDOSI (DI0041)
GrupoTerapeutico	
codigo_OMS	CODSUBGRUP (DI0032)
uso_terapeutico_OMS	DESCRIPCION (DI0032)
FormaFarmaceutica	
nombre_completo	DESCRIPCION (DI0038)
caracteristicas	LFFAGRUP (DI0038)
ViaAdministracion	
nombre	DESCRIPCION (DI0039)
UnidadMedida	
unidad_cualitativa_abreviada	DESCRIPCION (DI0037)
unidad_cualitativa_completa	OBSERVACIONES (DI0037)

Tabla 5.2. Correspondencias entre las clases en OntoFIS y los campos de las tablas de ‘Digitalis’.

En la segunda fase, se incorporaron, manualmente, instancias para los conceptos: ‘Excipiente’, ‘Accion’, ‘EfectoAlterado’, ‘EfectoDeseable’,

‘EfectoIndeseable’, ‘Contraindicacion’, ‘Estabilidad’, ‘Edad’, ‘Posologia’, ‘ProcesoClinico’ y ‘Sobredosis’.

La base de conocimiento de OntoFIS contiene cerca de 55.000 instancias.

5.3.3. Primer experimento: vocabulario completo.

Este experimento inicial se realizó para comprobar si los conceptos explicitados eran significativos y completos en relación al dominio. Los pasos fueron:

- La selección de una muestra de cuestiones de competencia.
- La transcripción de las cuestiones seleccionadas al lenguaje DL query utilizando el vocabulario de OntoFIS.
- El análisis de resultados.

5.3.3.1. Selección de cuestiones.

La muestra de preguntas, para el experimento inicial, constó de 29 cuestiones de competencia. Estas preguntas se obtuvieron por medio de un muestreo aleatorio estratificado. Este proceso de selección aseguraba disponer de un grupo representativo, un tercio del total de las cuestiones del banco. Además, la muestra era equiparable, para cada una de las categorías de clasificación de las cuestiones, y se obtenía, al menos, una cuestión de aquellas categorías infrarepresentadas.

En la tabla 5.3. se muestran las cuestiones seleccionadas por cada categoría.

Categoría	Universo (N) Muestra (M)	Cuestiones seleccionadas
Cómo prescribir	N = 16 M = 5 (1,4,7,10,3)*	1. ¿Cómo se prescribe/administra el fármaco X (en la situación Y)? 4. ¿Debería cambiar la dosis del fármaco X (en la situación Y)? 7. ¿Cuándo debería comenzar el medicamento X (en términos de tiempo)? 10. ¿Cómo debería interrumpir el medicamento X (en términos de tiempo)? 3. What is the dose of drug X
Eficacia/ Indicación/ Fármaco Elección	N = 20 M = 6 (3,5,6,12,13,17)*	3. Is drug X indicated in situation Y? 5. ¿Está algún fármaco indicado en la situación Y? 6. Is any drug indicated in situation Y? 12. ¿Es el fármaco X1 mejor que el X2, X3..., Xn para la condición Y? 13. ¿Es el fármaco X1 tan efectivo como el fármaco X2 (en la situación Y)? 17. ¿Qué fármaco se debería prescribir para prevenir la condición Y?
Efectos Adversos	N = 19 M = 6 (2,3,7,13,15,18)*	2. Could finding Y be caused by drug X 3. ¿Produce el medicamento X el hallazgo Y? 7. ¿Cuál es la probabilidad de que el efecto adverso Y proceda del fármaco X? 13. ¿Cómo puede administrarse el fármaco X para minimizar los efectos adversos? 15. What dose of drug X would cause toxicity? 18. ¿Está contraindicado el fármaco X en la situación Y?
Interacciones	N = 4 M = 1 (2)*	2. Is it OK to use drug X with drug Y?
Nominación	N = 2 M = 1 (2)*	2. What is the name of that drug that I'm trying to think of?
Orientación/ Composición	N = 6 M = 2 (1,3)*	1. ¿Qué es el fármaco X? 3. ¿Qué tiene el medicamento X?
Características Físicas	N = 8 M = 2 (4,6)*	4. ¿Cuáles son las características físicas respecto al recipiente del fármaco X? 6. What are the physical characteristics tablet?
Farmacodinámica/ Absorción	N = 6 M = 2 (3,4)*	3. What are the pharmacodynamics characteristics of drug X? 4. What are the absorption characteristics of drug X?
Mecanismo de Acción	N = 2 M = 1 (1)*	1. ¿Cuál es el mecanismo de acción del fármaco X?
Coste	N = 5 M = 2 (2,5)*	2. What is the cost of drug X? 5. What drug can I use in situation Y, given third party rules Y?
Niveles Séricos	N = 4 M = 1 (4)*	4. how often to draw a drug level?
Disponibilidad	N = 2 M = 1 (1)*	1. ¿Está disponible ya el fármaco X?

Nota Aclaratoria: (*) obtenidos con el Generador de números aleatorios, con repetición http://nosetup.org/php_on_line/numero_aleatorio (número final = N y número de veces 1).

Tabla 5.3. Cuestiones seleccionadas del banco de preguntas para la evaluación de validación.

5.3.3.2. Transcripción de las cuestiones en DL Query.

Las 29 cuestiones de competencia seleccionadas fueron transformadas, manualmente, de acuerdo con el vocabulario de OntoFIS, en lenguaje DL Query.

En total se pudieron transcribir 23 cuestiones de competencia que se enumeraron correlativamente. Aunque, hubo dos cuestiones (la 4/5 y la 9/10) que resultaron idénticas porque eran traducciones en inglés y castellano. Esto no afectaba a la transcripción, por lo que se trataron como equivalentes. Sin embargo, varias de las cuestiones seleccionadas (la 9/10, la 14, la 15, la 16 y la 18) tenían 2 posibles transcripciones de acuerdo con el vocabulario de OntoFIS. En total se obtuvieron 26 enunciados en DL Query. En las tablas 5.4, 5.5 y 5.6 se muestra la transcripción de las cuestiones de la muestra y corresponden a las cuestiones de la 1 a la 6, de la 7 a la 15 y de la 16 a la 23, respectivamente.

<p>C1. ¿Cómo se prescribe el fármaco X (en la situación Y)? <i>C1T1: Posologia and (related_to_1.2 some ({Farmaco_X} and related_to_1.2 some {ProcesoClinico_Y}))</i></p> <p>C2. ¿Debería cambiar la dosis del fármaco X (en la situación Y)? <i>C2T1: Posologia and (related_to_1.2 some ({Farmaco_X} and related_to_1.2 some {ProcesoClinico_Y})) *</i></p> <p>C3. What is the dose of drug X <i>C3T1: Posologia and related_to_1.2 some {Farmaco_X}</i></p> <p>C4. ¿Está algún fármaco indicado en la situación Y? C5. Is any drug indicated in situation Y? <i>C4=5T1: Farmaco and suitable_for_2.2.1 some {ProcesoClinico_Y}</i></p> <p>C6. ¿Es el fármaco X1 mejor que el X2, X3..., Xn para la condición Y? <i>C6T1: (Contraindicacion or EfectoIndeseable) and (inv (related_to_1.2) some ({Farmaco_X} and related_to_1.2 some {ProcesoClinico_Y}))</i></p>
<p>Leyenda:</p> <ul style="list-style-type: none"> El código de identificación es el siguiente: C seguida de un número indica la cuestión de la muestra. C acompañada de número y seguida de T y un número indica la transcripción realizada respecto a la cuestión de la muestra. *El planteamiento es similar a la pregunta antecesora, en ésta se busca una variante centrada en sobredosis. <p>Nota Aclaratoria: Las cuestiones 4 y 5 son enunciados equivalentes expresados en inglés y castellano, lo cual no afectaba a la transcripción en DL-Query.</p>

Tabla 5.4. Transcripción de las 6 primeras cuestiones en lenguaje DL Query.

C7. ¿Es el fármaco X1 tan efectivo como el fármaco X2 (en la situación Y)?
C7T1: Accion and related_to_1.2 some ({Farmaco_X} and related_to_1.2 some {ProcesoClinico_Y})

C8. ¿Qué fármaco se debería prescribir para prevenir la condición Y?
C8T1: Farmaco and facilitates_2.1.4 some (IndicacionTerapeutica and related_to_1.2 some {ProcesoClinico_Y})

C9. Could finding Y be caused by drug X
C9=10T1: {EfectoIndeseable_Z} and related_to_1.2 some {Farmaco_X}
C9=10T2: {EfectoIndeseable_Z} and (inv (related_to_1.2) some {Farmaco_X})

C11. ¿Cuál es la probabilidad de que el efecto adverso Y proceda del fármaco X?
C11T1: EfectoIndeseable and (related_to_1.2 some {Farmaco_X} and (
(EfectoIndeseable..muy_raro value "Y"^^string) or
(EfectoIndeseable..raro value "Y"^^string) or
(EfectoIndeseable..general value "Y"^^string) or
(EfectoIndeseable..poco_frecuente value "Y"^^string) or
(EfectoIndeseable..frecuente value "Y"^^string) or
*(EfectoIndeseable..muy_frecuente value "Y"^^string)))***

C12. ¿Cómo puede administrarse el fármaco X para minimizar los efectos adversos?
C12T1: ViaAdministracion and related_to_1.2 some {Farmaco_X}

C13. What dose of drug X would cause toxicity?
C13T1: Sobredosis and related_to_1.2 some {Farmaco_X}

C14. ¿Está contraindicado el fármaco X en la situación Y?
C14T1: {Farmaco_X} and unsuitable_for_2.2.3 some {ProcesoClinico_Y}
C14T2: {Farmaco_X} and related_to_1.2 some (Contraindicacion and related_to_1.2 some {ProcesoClinico_Y})

C15. Is it OK to use drug X with drug Y?
*C15T1: {Farmaco_X} and related_to_1.2 some (EfectoIndeseable and co-occurs_with_3.1 some {Farmaco_Y})****
*C15T2: {Farmaco_X} and related_to_1.2 some (ViaAdministracion and related_to_1.2 some {Farmaco_Y} and (ViaAdministracion..interaccion_fisico_quimica_preparacion some string))*****

Leyenda:

- El código de identificación es el siguiente: C seguida de un número indica la cuestión de la muestra. C acompañada de número y seguida de T y un número indica la transcripción realizada respecto a la cuestión de la muestra.
- ** El campo en que se encuentra Y será su probabilidad.
- *** La transcripción se enfoca comprobar los riesgos por sobredosis
- **** Se enfoca a los riesgos en la preparación y administración de dos medicamentos conjuntamente (caso de la medicación parenteral).

Nota Aclaratoria: Las cuestiones 9 y 10 son enunciados equivalentes expresados en inglés y castellano, lo cual no afectaba a la transcripción en DL-Query.

Tabla 5.5. Transcripción de las cuestiones de la 7 a la 15 en lenguaje DL Query.

<p>C16. What is the name of that drug that I'm trying to think of? <i>C16T1: Farmaco and (inv (consists_of_1.1.1) some {PrincipioActivo_Y})</i> <i>C16T2: Farmaco and (inv (co-occurs_with_3.1) some {IndicacionTerapeutica_Y})</i></p> <p>C17. ¿Qué es el fármaco X? <i>C17T1: GrupoTerapeutico and (inv (belongs_1.3) some {Farmaco_X})</i></p> <p>C18. ¿Qué tiene el medicamento X? <i>C18T1: PrincipioActivo and consists_of_1.1.1 some {Farmaco_X}</i> <i>C18T2: ComposicionQuimica and consists_of_1.1.1 some {Farmaco_X}</i></p> <p>C19. What are the physical characteristics tablet? <i>C19T1: FormaFarmaceutica and property_of_4.3 some {Farmaco_X}</i></p> <p>C20. What are the pharmacodynamics characteristics of drug X? <i>C20T1: EfectoAlterado and related_to_1.2 some {Farmaco_X}</i></p> <p>C21. What are the absorption characteristics of drug X? <i>C21T1: ViaAdministracion and ViaAdministracion..nombre value "ORAL"^^string and related_to_1.2 some {Farmaco_X}</i></p> <p>C22. ¿Cuál es el mecanismo de acción del fármaco X? <i>C22T1: Accion and property_of_4.3 some {Farmaco_X}</i></p> <p>C23. ¿Está disponible ya el fármaco X? <i>C23T1: Farmaco and {Fármaco_X}</i></p>
<p>Leyenda:</p> <ul style="list-style-type: none"> • El código de identificación es el siguiente: C seguida de un número indica la cuestión de la muestra. C acompañada de número y seguida de T y un número indica la transcripción realizada respecto a la cuestión de la muestra. • *** La transcripción se enfoca comprobar los riesgos por sobredosis • ****Se enfoca a los riesgos por el uso de la misma vía de administración (medicación parenteral).

Tabla 5.6. Transcripción de las cuestiones de la 16 a la 23 en lenguaje DL Query.

5.3.3.3. Análisis de resultados.

Hubo 6 cuestiones de competencia que no pudieron ser transcritas y que reflejan necesidades informativas que no son afines con el alcance de la ontología modelada.

Las cuestiones ‘¿Cuándo se debería comenzar el medicamento X (en términos de tiempo)?’, ‘¿Cómo debería interrumpir el medicamento X (en términos de tiempo)?’ y ‘How often to draw a drug level?’ (controles analíticos) inciden en aspectos muy específicos del seguimiento

farmacológico y no son generalizables a la gran mayoría de los medicamentos.

En cuanto a la cuestión ‘¿Cuáles son las características físicas respecto al recipiente de fármaco X?’, en OntoFIS no se reflejan las características del envasado (*original package*). La representación de las características físicas se relaciona con las que se precisan para la prescripción, es decir, con la forma farmacéutica (comprimidos, cápsulas, inyectables...).

Por último, tampoco se trata el coste de los medicamentos, teniendo en cuenta la variabilidad del mercado y las discrepancias de cobertura sanitaria entre países. Esto explica que los conceptos incluidos en las cuestiones ‘What is the cost of drug X?’ y ‘What drug can I use in situation Y, given third party rules Y?’ no hayan sido tenidos en cuenta desde el inicio del modelado de OntoFIS.

Aunque, las preguntas de competencia detectan cuatro núcleos conceptuales ausentes en la representación (coste, tipo de envasado, condiciones temporales y controles analíticos), la valoración es que estos conceptos no determinan que la ontología sea incompleta, teniendo en cuenta que se refleja el interés común de tres perfiles profesionales diferenciados. Además, los conceptos no cubiertos corresponden con niveles de mayor especificidad, que no coinciden con el nivel de granularidad establecido en OntoFIS (ver el apartado 4.4.3, pp119). En este sentido, se podría indicar que la ontología precisará ampliarse a niveles de mayor granularidad. Los núcleos conceptuales candidatos a aumentar su detalle en OntoFIS son: el concepto ‘forma terapéutica’ para dar entrada al concepto ‘tipo de envasado’, el concepto ‘fármaco’ para ‘coste’ y el concepto ‘posología’ para dar entrada a los conceptos ‘condiciones temporales’ y ‘controles analíticos’.

5.3.4. Segundo experimento: exactitud de la inferencia.

El objetivo del segundo experimento fue determinar la exactitud de las inferencias para obtener respuestas correctas y completas.

Ante la sobrecarga computacional que suponía el tamaño de la base de conocimiento de OntoFIS, con cerca de 55.000 instancias, se optó por obtener una fragmentación de ella para la realización del segundo experimento. A partir de lo cual se ejecutaron los siguientes pasos:

- La fragmentación de la base de conocimiento de OntoFIS.
- La determinación del grupo diana de medicamentos para el control las de respuestas.
- La interrogación de la base de conocimiento, que consistió en:
 - La ordenación de las sentencias DL-Query por complejidad y tamaño.
 - La ejecución de las sentencias en Protégé 4 beta.
- La comparación de los resultados obtenidos con el patrón de respuestas.
- La determinación de los índices de precisión y exhaustividad.

5.3.4.1. Fragmentación de la base de conocimiento de OntoFIS.

La base de conocimiento empleada para el experimento fue fragmentada para reducir su tamaño. El criterio para la selección se basó en el código de grupo terapéutico eligiendo (Anexo VI):

- Todos los que comenzasen por el código B01.
- Todos con los estuvieran comprendidos entre el código B02A y B06.
- Todos los que comenzasen su código por la letra C, excepto los códigos C05A, C05B, C05D y C3D00.

En total, el fragmento de la base de conocimiento con la que se trabajó para verificar la adecuación y completitud de OntoFIS, quedó constituido por 25.099 instancias, distribuidas tal como muestra la figura 5.9. Este fragmento contiene 3.928 medicamentos comercializados que se emplean para problemas del sistema cardiovascular.



Nota Aclaratoria: Los principios activos en OntoFIS están declarados como componentes farmacológicos. Un componente farmacológico es un principio activo con una de las concentraciones con la que se comercializa. La combinación de los nombres de los principios activos con las concentraciones proporciona las 19.627 instancias incluidas.

Los medicamentos pertenecientes al fragmento de la base de conocimiento son 3.928. En la imagen se ve 3930 ya que se incorporaron manualmente 2 medicamentos que estaban fuera del grupo de selección.

Figura 5.9. Fragmento de la base de conocimiento OntoFIS para validar la adecuación del contenido.

Las instancias asociadas al concepto '*principio activo*' fueron las únicas que no se filtraron para fragmentar la base de conocimiento. A este respecto, es necesario indicar que el concepto '*principio activo*', en

OntoFIS, está declarado como ‘componente farmacológico’. Un componente farmacológico es un principio activo con una de las concentraciones con la que se comercializa. La base de datos Digitalis proporciona 4.456 denominaciones de principios activos. La combinación de los nombres de los principios activos con las concentraciones respectivas proporciona las 19.627 instancias incluidas en la base de conocimiento.

5.3.4.2. Medicamentos Diana.

Para el control de las respuestas se estableció un conjunto de medicamentos diana. Este conjunto está formado por 209 medicamentos. Los medicamentos que forman parte del fragmento de la base de conocimiento seleccionada son 207 y tienen como característica común su acción *antilipemiente*⁴⁹. Aunque, entre ellos se diferencian por el principio activo: ‘Atorvastatina’ (12), ‘Simvastatina’ (161) y ‘Atorvastatina calcio trihidrato asociado a amlodipino’ (34). Los otros 2 medicamentos incluidos en este conjunto de ‘medicamentos diana’, fueron incorporados, posteriormente, a la base de conocimiento. El propósito de su inclusión fue introducir una distorsión controlada en la homogeneidad de la base de conocimiento con la que se iba a trabajar. Estos medicamentos fueron: ‘aspirina masticable Comp. masticable 500 mg®’ con indicación ‘analgesica’ y ‘antipirética’, y cuyo principio activo es el ‘ácido acetilsalicílico’ y ‘Ciclochem uñas Sol. cutánea 80%®’ un ‘antifúngico’, cuyo componente es el ‘ciclopirox’.

⁴⁹ Descenso del colesterol.

5.3.4.3. La interrogación de la base de conocimiento.

Para la interrogación, los enunciados de las cuestiones en DL Query se especificaron con alguna de las siguientes alternativas:

- Cuando se tuvo que especificar el medicamento se utilizó: ‘ALCOSIN Comp. recub. con película 10 mg®’, ‘ALCOSIN Comp. recub. con película 20 mg®’, ‘ASTUCOR Comp. recub. 10 mg/10 mg®’, ‘SINVASTATINA TECNIGEN EFG Comp. recub. 20 mg ®’, ‘ASPIRINA MASTICABLE Comp. masticable 500 mg ®’ y ‘CICLOCHEM UÑAS Sol. cutánea 80% ®’.
- Cuando se precisó especificar la sustancia activa se emplearon: ‘Atorvastatina’ ‘Simvastatina’ y ‘Atorvastatina calcio trihidrato asociado a amlodipino’.
- Cuando se tuvo que indicar el proceso clínico se utilizaron: ‘hipercolesterinemia’, ‘aumento de CPK’, ‘dolor abdominal’, ‘estreñimiento’, ‘diarrea’, ‘anemia’ y ‘pancreatitis’.
- Cuando se tuvo que especificar las contraindicaciones se empleó: ‘hepatopatía activa’, ‘antecedentes hemorragia’, ‘perforación gástrica tras tratamiento con ácido acetilsalicílico’ y ‘perforación gástrica tras tratamiento con otros antiinflamatorios no esteroides’.
- Cuando se especificó el efecto adverso se empleó: ‘aumento transaminasas séricas’ y ‘flatulencia’.

Un ejemplo del proceso de interrogación se muestra a continuación:

1. Cuestión de competencia: ¿Cómo se prescribe el fármaco X?

Cuestión especificada: ¿Cómo se prescribe el fármaco “ALCOSIN 20MG 28 COMPRIMIDOS”?

2. Cuestión en DL Query: Posologia and (inv (related_to_1.2) some {Farmaco_326})

“ALCOSIN 20MG 28 COMPRIMIDOS” corresponde con la instancia de OntoFIS URI: “http://www.dlsi.ua.es/OntoFIS/Farmaco_326” (abreviado “Farmaco_326”).

3. Patrón de respuesta correcta seleccionada, manualmente:

Instancias de la clase Posología: Posologia_0, Posologia_1, Posologia_2, Posologia_3, Posologia_4, Posologia_5, Posologia_6 y Posologia_7

4. Resultados de la interrogación de la base de conocimiento (figura 5.10).

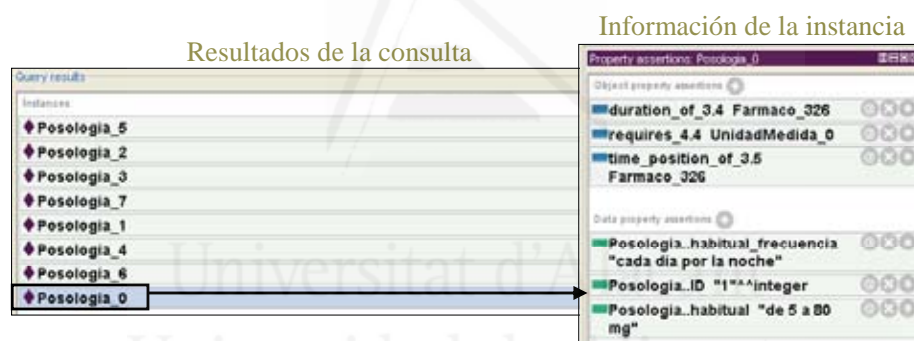


Figura 5.10. Resultado de la interrogación de la base de conocimiento OntoFIS en Protégé 4 beta.

5.3.4.4. Análisis de resultados.

Los resultados, para cada una de las cuestiones en DL Query, se recogen en la tabla 5.7. En la primera columna se incluye el código de identificación de las cuestiones (comprobar los enunciados con las tablas 5.4, 5.6 y 5.5). Las dos siguientes columnas se refieren a los resultados de precisión y exhaustividad obtenidos.

Las fórmulas de cálculo utilizadas fueron:

- *Precisión* = conjunto de respuesta relevantes recuperadas / conjunto de respuestas recuperadas.
- *Exhaustividad* = conjunto de respuestas relevantes recuperadas / conjunto de respuestas relevantes de la colección

Número de Orden	Precisión	Recall
C1T1	0,52	0,92
C2T1	0,51	0,83
C3T1	0,52	0,92
C4=5T1	1	1
C6T1	0,50	0,63
C7T1	0,54	1
C8T1	1	1
C9=10T1	0,50	1
C9=10T2	1	1
C11T1	1	1
C12T1	0,50	1
C13T1	0,51	1
C14T1	1	1
C14T2	1	1
C15T1	0,67	0,67
C15T2	0,67	0,67
C16T1	1	1
C16T2	1	1
C17T1	1	1
C18T1	1	1
C18T2	1	1
C19T1	1	1
C20T1	0,50	1
C21T1	0,50	1
C22T1	1	1
C22T2	1	1
C23T1	1	1

Tabla 5.7. Resultados de la calidad de la inferencia de OntoFIS.

En el 46,15% de las cuestiones se obtuvieron valores de precisión entre 50-67%, mientras que, en el 23% de las cuestiones, se obtuvieron resultados de exhaustividad comprendidos entre el 63-92%.

5.3.5. Refinamiento.

El proceso de inferencia de las 12 cuestiones, cuyos índices de precisión y exhaustividad eran inferiores a 1, se reprodujo, manualmente, en el *'diagrama de descripción de OntoFIS'*.

Los bajos índices obtenidos se explicaban por la existencia de ciclos, en las relaciones de asociación semánticas, que no habían sido detectados en las revisiones anteriores. El razonador FaCT++ sólo detectaba los errores de circularidad en la estructurada jerárquica, pero no ayudaba a detectar bucles por asociaciones semánticas organizadas como grafos; característica propia del modelo representado.

El problema se localizó en el 20% de las relaciones establecidas. Se detectaron 132 relaciones semánticas que precisaron cambios en la dirección de la relación para evitar la generación de ciclos, los cuales habían condicionado la falta de respuestas correctas y completas.

5.3.6. Experimento de comprobación de mejoras.

Tras la modificación de la dirección de las relaciones, se realizó un tercer experimento para determinar la calidad de la inferencia. Nuevamente, se interrogó al fragmento de la base de conocimiento de OntoFIS con el conjunto de preguntas en DL Query. Los resultados obtenidos indicaban que se habían resuelto las incongruencias de las respuestas al obtenerse, en todos los casos, una precisión y una exhaustividad del 100%.

No obstante, para verificar que no existía ninguna laguna, se crearon dos cuestiones complejas directamente en DL Query (tabla 5.8). Cada enunciado buscaba una respuesta inferida por la combinación de varios conceptos (clases) relacionados semánticamente. Estos nuevos enunciados tienen un valor determinante para la prescripción o el seguimiento de eficacia de un medicamento. En ambas preguntas, nuevamente, se lograron respuestas con contenidos correctos y completos (precisión y exhaustividad = 1). Lo cual determinó la finalización de la evaluación de validación de OntoFIS.

```
(Contraindicacion or EfectoIndeseable or EfectoDeseable or EfectoAlterado) and (inv
(related_to_1.2) some (Farmaco and (inv (consists_of_1.1.1) some (PrincipioActivo and
consists_of_1.1.1 some {Farmaco_X}))) and ((inv (property_of_4.3) some
(FormaFarmaceutica and property_of_4.3 some {Farmaco_X}))) or (inv (property_of_4.3)
some (FormaFarmaceutica and consists_of_1.1.1 some (PrincipioActivo and consists_of_1.1.1
some {Farmaco_X})))) and (related_to_1.2 some (ViaAdministracion and (inv
(related_to_1.2) some {Farmaco_X}))))))

(Contraindicacion or EfectoIndeseable or EfectoDeseable or EfectoAlterado) and (inv
(related_to_1.2) some (Farmaco and (related_to_1.2 some (GrupoTerapeutico (inv
(related_to_1.2) some {Farmaco_X}))) and (inv (consists_of_1.1.1) some (PrincipioActivo and
consists_of_1.1.1 some {Farmaco_X}))) and (related_to_1.2 some {ProcesoClinico_Y})))
```

Tabla 5.8. Cuestiones adicionales creadas en DL Query con la terminología de OntoFIS

5.4. Acreditación de OntoFIS.

La metodología de evaluación aplicada permite describir a OntoFIS con los siguiente indicadores de calidad:

- OntoFIS está formalizada en lenguaje OWL DL, su código cumple con el estándar y no incluye errores de sintaxis, ni contiene errores de partición ni errores de circularidad en su estructura jerárquica.

- Su cobertura conceptual es la farmacoterapéutica, con vocabulario común a médicos, enfermeras y farmacéuticos especificado por 23 conceptos con un grado de abstracción intermedio.
- El modelo representado se describe por medio de 647 interconexiones entre sus conceptos, a partir de una taxonomía de 36 tipos de relaciones semánticas.
- Incluye la descripción en lenguaje natural de cada concepto.
- Cada concepto incorpora las etiquetas terminológicas que se utilizan para su denominación en español, inglés, francés, italiano y alemán, así como sus posibles sinónimos en cada idioma.
- Se dispone de casi 55.000 instancias, etiquetadas semánticamente con OntoFIS. En las que destacan, por su valor terminológico, las correspondientes a :
 - La denominación de los medicamentos comercializados en España (17.204).
 - Los componentes farmacológicos que corresponden a los principios activos y sus respectivas concentraciones autorizadas (19.627).
 - El nombre genérico de principios activos (4.456).
 - La denominación de grupos químicos de acuerdo con la OMS⁵⁰ (3.200).
 - Las denominaciones de uso terapéutico (1.380).

⁵⁰ Organización Mundial de la Salud.

- Los tipos de formas terapéuticas (245).
 - Los tipos de vías de administración (34).
 - Las denominaciones de procesos clínicos (107).
 - Los tipos de efectos alterados (46).
 - Los tipos de efectos indeseables (44).
- El modelo representado responde con contenidos adecuados a cuestiones de la práctica clínica relacionadas con aspectos tales como:
 - La prescripción.
 - La elección de un fármaco de acuerdo con la eficacia y la indicación.
 - Los efectos adversos.
 - Las interacciones.
 - La nominación de los medicamentos y de los fármacos.
 - La composición de un medicamento.
 - Las características físicas relativas a la forma farmacéutica.
 - La farmacodinámica o absorción.
 - El mecanismo de acción.
 - La disponibilidad relativa a su comercialización en el territorio nacional.

En otras palabras, los resultados obtenidos indican que OntoFIS cumple con los objetivos, previamente establecidos a su premodelado, y contiene el conocimiento relacionado tanto con el proceso (terapéutica) como con el objeto (medicamento).

OntoFIS especifica los conceptos relacionados con los problemas de salud clínicos (enfermedades, signos y síntomas, agentes etiológicos), los medicamentos (denominación, concentración, forma farmacéutica, principio activo, clasificación terapéutica, posología, conservación), la administración (vía, preparación) y la seguridad (efectividad, efectos adversos, toxicidad). La descripción responde al conocimiento compartido de los potenciales perfiles de usuarios (médico, farmacéutico, enfermera y consumidor) e incluye conocimiento multilingüe. Proporciona el significado de los términos incluidos, con descripciones en lenguaje natural y descripciones objetivas, que permite hacer inferencias válidas y consistentes respecto al conocimiento del dominio.

No obstante, OntoFIS no puede responder a cuestiones relacionadas con las características del envasado (*original package*), aunque sí, con las características de la forma farmacéutica para la prescripción, que es uno de sus objetivos. Tampoco es un objetivo del modelo realizado representar el coste del medicamento. Asimismo, la representación de conocimiento realizada no incluye las consideraciones especiales de control de medicamentos específicos para su seguimiento. Todo ello se puede explicar por el nivel de granularidad (intermedio) con el que se ha trabajado.

Aunque, el modelado realizado facilita su ampliación hacia una mayor especificación y cobertura. Esta consideración abre la línea de trabajos futuros, orientados a la reutilización de otras ontologías con contenidos complementarios, como ya se mencionó en el capítulo de diseño y construcción de OntoFIS.

Capítulo VI.

Conclusiones y Trabajos Futuros.

En este capítulo final se presentan las conclusiones más destacadas, así como el conjunto de las principales contribuciones aportadas con el desarrollo de esta Tesis y, seguidamente, se incluye la propuesta de trabajos en progreso. Se finaliza con la producción científica derivada de la consecución de esta Tesis.

6.1. Aportaciones

La principal aportación de este trabajo es el desarrollo de una tecnología ontológica en el dominio farmacoterapéutico. La Ontología Farmacoterapéutica e Información para el Seguimiento, OntoFIS, contribuye a resolver el problema de la carencia de una fuente de conocimiento específica, en lengua castellana, que facilite la gestión avanzada de la información y permita compartirla en este dominio.

Las principales contribuciones de este trabajo son:

1. *El estudio del estado de la cuestión.*

Se ha realizado un estudio de la definición del concepto y de las características que se consideran para el diseño y la construcción de una ontología. Además, se han estudiado los tipos de ontologías existentes, distinguiendo entre la declaración de los componentes representados, la formalización del contenido y la conceptualización respecto al universo del discurso. Aunque, existe una gran variabilidad en este tipo de tecnología, se han valorado sus funcionalidades desde la perspectiva del PLN.

También, se han analizado varias de las tecnologías más características, en el ámbito de la salud, para la representación del conocimiento y el intercambio de información. Todas ellas responden a diferentes intereses sanitarios, lo que justifica su heterogeneidad terminológica y, al mismo tiempo, el solapamiento en algunos aspectos conceptuales. Aunque, el principal problema de estas tecnologías clásicas son sus diferencias para controlar la ambigüedad de los significados de los términos y la inferencia de nuevo conocimiento. Esta limitación se debe a las diferencias en la riqueza de sus estructuras y a la variabilidad en el grado de formalización de

su representación. Asimismo, hay núcleos conceptuales del dominio de la salud, ampliamente extendidos y con alta granularidad, mientras que otros núcleos conceptuales, todavía muestran representaciones escasas y con falta de especificaciones detalladas, como es el caso de los medicamentos.

2. *El estudio del escenario farmacoterapéutico.*

A partir de la identificación y el alcance global del escenario farmacoterapéutico, se ha realizado un diagnóstico de la situación actual de los sistemas de información disponibles. Se han descrito las características de la información relevante para el proceso de la atención terapéutica de tipo farmacológico. También, se han especificado las necesidades de información de los distintos actores que participan en este escenario y la cobertura que facilitan las fuentes de información sobre el medicamento.

Además, se han descrito las prestaciones de los sistemas de información (SI) para la prescripción farmacológica. En este aspecto se ha destacado la aplicación de técnicas basadas en PLN para la reconciliación de la información sobre la medicación que se encuentra en distintos módulos de la historia clínica electrónica (HCE).

Por último, se han subrayado las limitaciones existentes en este escenario. Todas ellas parten del hecho de la existencia de información fragmentada, no integrada e imposible de compartir en la actualidad por los SI. Los profesionales sanitarios son los que deben conocer el contenido y la forma de acceder, a cada uno de los SI que precisan, para apoyar su toma de decisión. Sin olvidar que la información sobre la atención se crea en la lengua nativa pero, una gran parte de la información complementaria, se encuentra en lengua inglesa.

Como solución se ha propuesto la construcción de una ontología para superar los inconvenientes de la gestión de información farmacoterapéutica.

3. *La descripción de una propuesta metodológica para el diseño y la construcción de una ontología farmacoterapéutica.*

Se han descrito las metodologías de construcción más relevantes, analizado su correspondencia con un esquema conceptual. Este análisis ha fundamentado la propuesta de la metodología aplicada para el diseño y la construcción de la ontología.

Se ha empleado el modelo de datos UML y OWL DL para el modelado de OntoFIS.

La propuesta realizada se ha aplicado en un dominio específico pero, ésta puede ser aplicada a cualquier otro escenario.

4. *La descripción de una propuesta tecnológica para la construcción de una ontología.*

Se ha utilizado el *plug-in Semantic Web Tools for VisualWADE*, como propuesta tecnológica para las transformaciones directas entre el diagrama de clases UML y el modelo de datos OWL DL. Esta propuesta tecnológica facilita la reutilización de los diseños de aplicaciones Web para la obtención de ontologías.

5. *La construcción de OntoFIS y su base de conocimiento.*

Se ha detallado el proceso de captura, descripción y formalización de la Ontología Farmacoterapéutica e Información de Seguimiento hasta la construcción de OntoFIS en OWL DL, de acuerdo con la propuesta metodológica y tecnológica establecida.

La ontología OntoFIS está constituida por 23 conceptos, con un grado de abstracción intermedio. El modelo farmacoterapéutico incluido está descrito por medio de 647 interconexiones, a partir de una taxonomía de 36 tipos de relaciones semánticas. Además, incluye conocimiento multilingüe para la denominación de los conceptos e incorpora descripciones en lenguaje natural.

La ontología OntoFIS está poblada con casi 55.000 instancias. Entre éstas destacan, por su número y su valor terminológico, las correspondientes a las denominaciones de los medicamentos comercializados en España (17.204), los componentes farmacológicos (19.627), los nombres genéricos de principios activos (4.456), las denominaciones de grupos químicos (3.200) y las denominaciones de uso terapéutico (1.380).

6. *La evaluación y la acreditación de la ontología construida.*

Se ha presentado la fundamentación del método de evaluación ejecutado. Se ha descrito el proceso de verificación y de validación del modelo conceptual de OntoFIS. Se han expuesto las limitaciones detectadas, las soluciones adoptadas y los resultados obtenidos.

Como consecuencia del proceso de evaluación, la ontología OntoFIS ha sido acreditada de acuerdo a los indicadores de calidad comprobados. Los resultados obtenidos indican que OntoFIS cumple con los objetivos, previamente establecidos a su premodelado. Los conceptos y relaciones especificadas, en OntoFIS, responden al conocimiento compartido de los potenciales usuarios (médico, farmacéutico, enfermera y consumidor) y permiten hacer inferencias válidas y consistentes respecto al conocimiento del dominio.

El modelo representado responde, con contenidos adecuados, a cuestiones de la práctica clínica relacionadas con la prescripción, la elección de un fármaco, los efectos adversos, las interacciones, la

denominación del medicamento y del fármaco, la composición de un medicamento, las características físicas relativas a la forma farmacéutica, la farmacodinámica o absorción, el mecanismo de acción y la disponibilidad, relativa a su comercialización, en el territorio nacional.

6.2. Trabajos futuros.

A partir de la aportación realizada, en este trabajo, existen diversas líneas abiertas de investigación que sería interesante abordar en futuros trabajos. Al menos, por una parte, está la inclusión de OntoFIS y su base de conocimiento en tareas de PLN y, por otra, la ampliación de la ontología para una mayor especificación y cobertura. Los trabajos futuros serían:

1. La elaboración de un sistema, para la clasificación de documentos y de búsqueda de respuestas, aplicado al dominio farmacoaterapéutico. El objetivo sería determinar si la fuente de conocimiento de OntoFIS mejora los resultados en estas tareas.
2. La ampliación de OntoFIS abre una línea de reutilización del conocimiento de otras ontologías, tanto del dominio de salud como de dominio abierto (figura 6.1), para cubrir, de forma más satisfactoria, las necesidades de los usuarios potenciales y la diversidad de información relevante para la atención farmacoterapéutica.

En este sentido, OntoFIS debería emplearse como ontología ‘puente’ y desarrollar, según las necesidades, la interconexión (*mapping*) entre un

modelo de dominio abierto (EuroWordNet) y un modelo de dominio específico (UMLS).

El objetivo sería determinar que tipo de interconexión semántica, entre estas fuentes de conocimiento, resulta más efectiva para ampliar el modelo de OntoFIS y ampliar su potencial multilingüe. Las alternativas iniciales podrían ser la aplicación de la metodología de negociación de significados (*meaning negotiation*) (Romá-Ferri y Palomar, 2005) y el algoritmo de similitud semántica.

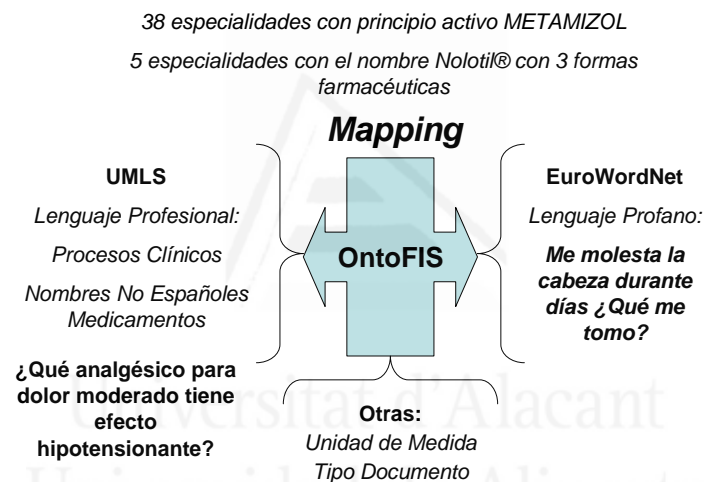


Figura 6.1. Interconexión de OntoFIS y otras ontologías, como propuesta de trabajos futuros.

Otra línea de trabajo futura podría ser mejorar el plug-in *Semantic Web Tools for VisualWADE*, como propuesta tecnológica para las transformaciones directas entre el diagrama de clases UML y el modelo de datos OWL DL. La mejora del *plug-in*, a partir de las propuestas realizadas en este trabajo, tiene como objetivo incrementar el potencial de reutilización de los diseños de aplicaciones Web para generar, de forma automática, ontologías de dominios específicos.

6.3. Producción científica.

Esta Tesis se ha desarrollado en el marco del proyecto ‘Desarrollo de una ontología multilingüe para el dominio médico-farmacológico’, financiado por el Fondo de Investigación Sanitaria (FIS) del Instituto de Salud Carlos III (PI051438), en la convocatoria de 2005 y cuya investigadora principal es M. Teresa Romá Ferri.

La producción científica derivada, como consecución de esta Tesis, ha sido:

Hermida, J., Romá-Ferri, M.T., Montoyo, A. y Palomar, M. (2009). Reusing UML Class Models to Generate OWL Ontologies. A Use Case in the Pharmacotherapeutic Domain. *Proceedings International Conference on Knowledge Engineering and Ontology Development, KEOD2009* (pp. 281-286). Funchal-Madeira, Portugal: Institute for Systems and Technologies of Information, Control and Communication (INSTICC).

Romá-Ferri, M.T., Cruanes, J. y Palomar, M. (2009). Quality indicators of the ‘OntoFIS’ pharmacotherapeutic ontology for semantic interoperability. *Proceedings IADIS International Conference e-Health 2009* (pp. 107-114). Algarve, Portugal: International Association for Development of the Information Society (IADIS).

Romá-Ferri, M.T. (2008). Ontologías en el ámbito sanitario. Conferencia presentada en el *II Encuentro Multidisciplinar de Bioingeniería e Informática Médica, EMBIM 2008*. Departamento de Tecnología Informática y Computación. Universidad de Alicante, 24 y 25 de Abril de 2008.

Romá-Ferri, M.T., Hermida, J.M., Montoyo, A. y Palomar, M. (2008). Representación del conocimiento farmacoterapéutico: diseño de una ontología. Actas XI Congreso Nacional de Informática de la Salud, INFORSALUD2008 (pp. 240-247). Madrid: Sociedad Española de Informática de la Salud (SEIS).

Romá-Ferri, M.T. y Palomar, M. (2008). Análisis de terminologías de salud para su utilización como ontologías computacionales en los sistemas de información clínicos. Gaceta Sanitaria, 22(5): 421-433.

Esta revista está indexada en Science Citation Index desde 2007. El valor del factor de impacto “no oficial” estimado para el año 2008 fue de 1,496. Este valor ha sido calculado por la editorial Elsevier España, y está pendiente de su publicación oficial. El primer valor de factor de impacto tiene siempre una demora de dos años.

Romá-Ferri, M.T. y Palomar, M. (2005). Interoperabilidad semántica de ontologías basada en técnicas de procesamiento del lenguaje natural (PLN). En: J. Gascón, F. Burguillos, A. Pons. La dimensión humana de la organización del conocimiento. 7º Congreso International Society for Knowledge Organization (ISKO)-España. (pp. 534-548). Barcelona: Publicaciones de la Universidad de Barcelona. Departamento de Biblioteconomía y Documentación.

Bibliografía.

- AEMPS, Agencia española de medicamentos y productos sanitarios. *Informes técnicos de medicamentos autorizados en España (uso humano)*. (s.f.). Versión 20081003. Ministerio de Sanidad y Política Social. Recuperado el 9 de Septiembre de 2007, desde <https://sinaem4.agemed.es/consaem/fichasTecnicas.do?metodo=detalleForm>
- Aronson, A. R. y Rindflesch, T. C. (1997). Query expansion using the UMLS Metathesaurus. *Proceedings American Medical Informatics Association (AMIA) Annual Fall Symposium*, 485-489.
- Aronson, A. R., Mork, J. G., Gay, C. W., Humphrey, S. M. y Rogers, W. J. (2004). The NLM indexing initiative's medical text indexer. *Studies in Health Technology and Informatics*, 107(Pt 1), 268-272.
- Badia Cardús, T. (2003). Técnicas de procesamiento del lenguaje. En M. A. Martín Antonin (Coord.), *Tecnologías del lenguaje* (pp. 239-244). Barcelona: Universitat Oberta de Catalunya.

- Baixauli, V. J., Salar, L. y Barbero, A. (2004) Demanda de información en la farmacia comunitaria. *Pharmaceutical Care España*, 6(3), 136-144.
- BDM, *Base de Datos del Medicamento*. (s.f.). Consejo General de Colegios Oficiales de Farmacéuticos (CGCOF). Recuperado el 9 de Septiembre de 2007, desde <http://www.portalfarma.com/pfarma/espacioges.nsf/wfAccesoEspacioGestion?OpenForm>.
- Bodenreider, O. y Burgun, A. (2005) Biomedical ontologies. En H. Chen, S. Fuller, W. R. Hersh y C. Friedman (Eds.), *Medical informatics: advances in knowledge management and data mining in biomedicina* (pp. 211-236). New York: Springer-Verlag.
- Bodenreider, O. (2008). Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearbook of Medical Informatics*, 67-79.
- Bodenreider, O., Rindflesch, T. C. y Burgun, A. (2002). Unsupervised, corpus-based method for extending a biomedical terminology. *Proceedings of the ACL'2002 Workshop Natural Language Processing in the Biomedical Domain* (pp. 53-60). Recuperado el 9 de Septiembre de 2007, desde <http://mor.nlm.nih.gov/pubs/pdf/2002-acl-nlpbd-ob.pdf>
- Borst, W. N. (1997). *Construction of engineering ontologies for knowledge sharing and reuse*. Ph. D-series No. 97-14. University of Twente, Enschede, The Netherlands. Recuperado el 8 de Febrero de 2009, desde <http://doc.utwente.nl/17864/1/t0000004.pdf>
- Bouquet, P. (Coord). (2004, Agosto). Specification of a common framework for characterizing alignment. Technical Report: KWEB/2004/D2.2.1/v1.2. Proyecto Knowledge web: KWEB EU-IST-2004-507482 . Recuperado el 15 de Octubre de 2005, desde <http://eprints.biblio.unitn.it/archive/00000653/01/090.pdf>
- Brandt, C. y Nadkarni, P. (2001). Web-based UMLS concept retrieval by automatic text scanning: a comparison of two methods. *Computer Methods and Programs in Biomedicine*, 64(1), 37-43.
- Burgun, A. y Bodenreider, O. (2001). Comparing terms, concepts and semantics classes in WordNet and the Unified Medical Language System. *Proceedings*

- of NAACL'2001 Workshop, WordNet and Other Lexical Resource: Appliactions, Extensions and Customizations* (pp. 77-82). Association for Computational Linguistics. Recuperado el 9 de Septiembre de 2007, desde <http://mor.nlm.nih.gov/pubs/pdf/2001-naacl-wordnet-ab.pdf>
- Can, A. B. y Baykal, N. (2007). MedicoPort: a medical search engine for all. *Computer Methods and Programs in Biomedicine*, 86(1), 73-86.
- Carrero, F., Gómez, J. M., de Buenaza, M., Mata, J. y Maña, M. (2007). Acceso a la información bilingüe utilizando ontologías específicas del dominio biomédico. *Procesamiento del Lenguaje Natural*, 38, 107-117.
- CE, Comisión Europea. (2008, 18 de Julio). *Recomendación de la Comisión sobre la interoperabilidad transfronteriza de los sistemas de historiales médicos electrónicos de 2 de Julio* [notificada con el número C(2008) 3282]. Diario Oficial de la Unión Europea, L 190, 0037-0043. Recuperado el 6 de Abril de 2009, desde <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:190:0037:01:ES:HTML>
- CGCOF, Consejo General de Colegios Oficiales de Farmacéuticos. (2009, Abril). Nuevos medicamentos 2008. *Farmacéuticos*, 345, 54-61.
- Chabalier, J., Mosser, J., y Burgun, A. (2007). Integrating biological pathways in disease ontologies. *Studies in Health Technology and Informatics*, 129(Pt 1), 791-795.
- Chandrasekaran, B., Josephson, J. R. y Benjamins, V. R. (1999, Enero-Febrero). What are ontologies, and why do we need them? *IEEE Intelligent Systems*, 14(1), 20-26.
- Chapman, W. W., Fiszman, M., Dowling, J. N., Chapman, B. E. y Rindflesch, T. C. (2004). Identifying respiratory findings in emergency department reports for biosurveillance using MetaMap. *Medinfo*, 11(Pt 1), 487-491.
- Chen, E. S., Stetson, P. D., Lussier, Y.A., Markatou, M., Hripcsak, G. y Friedman, C. (2007). Detection of practice pattern trends through natural language processing of clinical narratives and biomedical literature. *American Medical Informatics Association (AMIA) Annual Symposium Proceedings*, 120-124.

- Ciccarese, P., Caffi, E., Quaglini, S. y Stefanelli, M. (2005). Architectures and tools for innovative health information systems: the guide project. *International Journal of Medical Informatics*, 74, 553-62.
- Cimino, J. J., Bright, T. J. y Li, J. (2007). Medication reconciliation using natural language processing and controlled terminologies. *Studies in Health Technology and Informatics*, 129(Pt 1), 679-683.
- CIE-9-MC, *Clasificación Internacional de Enfermedades, 9ª Revisión, Modificación Clínica*. (6ª ed). (2008, Enero). Madrid: Ministerio de Sanidad y Consumo. Recuperado el 4 de Noviembre de 2008, desde <http://www.msc.es/ecie9mc-2008/html/index.htm>
- Cole, C. L., Kanter, A. S., Cummins, M., Vostinar, S. y Naeymi-Rad, F. (2004). Using a terminology server and consumer search phrases to help patients find physicians with particular expertise. *Studies in Health Technology and Informatics*, 107(Pt 1), 492-496.
- Cornella, A. (2000). *Organitzacions intenses en informació*. [CD-Rom]. Barcelona: Universitat Oberta de Catalunya.
- Coumou, H. C. y Meijman, F. J. (2006). How do primary care physicians seek answers to clinical questions? A literature review. *Journal of the Medical Library Association*, 94(1), 55-60.
- Cranefield, S y Purvis, M. (1999). UML as an ontology modelling language. *Proceedings of the Workshop on Intelligent Information Integration, 16th International Joint Conference on Artificial Intelligence (IJCAI-99)*, Germany. Recuperado el 12 de Enero de 2008, desde http://hcs.science.uva.nl/usr/richard/workshops/ijcai99/UML_Ontology_Modelling.pdf
- Cranefield, S. (2001). UML and the Semantic Web. *Proceeding. of the International Semantic Web Working Symposium*, Palo Alto. Recuperado el 12 de Enero de 2008, desde <http://www.semanticweb.org/SWWS/program/full/paper1.pdf>
- Daniel-Le Bozec, C., Steichen, O., Dart, T. y Jaulent, M. C. (2007). The role of local terminologies in electronic health records. The HEGP experience. *Studies in Health Technology and Informatics*, 129(Pt 1), 780-784.

- de Ahumada, J., Santana, M. L. y Serrano, J.S. (2003). *Farmacología práctica: para las diplomaturas en ciencias de la salud (enfermería, fisioterapia, podología) con autoevaluación*. Madrid: Ediciones Díaz de Santos.
- de Bruijn, J. y Fensel, D. (2005). Ontology definitions. En M.J. Bates, M. N. Maack y M. Drake, (Eds.). *Encyclopedia of Library and Information Science*. Marcel Dekker, inc. Recuperado el 10 de Enero de 2006, desde <http://www.debruijn.net/publications/ontology-definitions.pdf>
- Dee, C. y Stanley, E. E. (2005). Information-seeking behavior of nursing students and clinical nurses: implications for health sciences librarians. *Journal of the Medical Library Association*, 93(2), 213-222.
- DeCS, *Descriptores en Ciencias de la Salud*. (2008, Marzo). Biblioteca Virtual en Salud (bvs), Bireme. Recuperado el 16 de Febrero de 2007 y el 18 de Abril de 2009, desde <http://decs.bvs.br/E/decswebe2008.htm>
- Directiva 2005/36/CE del Parlamento Europeo y del Consejo, de 7 de septiembre de 2005, relativa al reconocimiento de cualificaciones profesionales. (2005, 30 Septiembre). Diario Oficial de la Unión Europea, L255, 22-142 Recuperado el 6 de Abril de 2009, desde <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2005:255:0022:0142:es:PDF>
- EIF, *European Interoperability Framework for Pan-European e-government services*. (2004). Luxembourg: Office for official publications of the European Communities. Recuperado el 13 de Abril de 2009, desde <http://ec.europa.eu/idabc/servlets/Doc?id=19529>
- Ely, J. W., Osherooff, J. A., Ebell, M. H., Bergus, G. R., Levy, B. T., Chambliss, M. L. y Evans, E. R. (1999). Analysis of questions asked by family doctors regarding patient care. *British Medical Journal*, 319(7206), 358-361.
- Ely, J. W., Osherooff, J. A., Gorman, P. N., Ebell, M. H., Chambliss, M.L., Pifer, E. A. y Stavri, P. Z. (2000). A taxonomy of generic clinical questions: classification study. *British Medical Journal*, 321(7258), 429-432.
- Ely, J. W., Osherooff, J. A., Ebell, M. H., Chambliss, M. L., Vinson, D. C., Stevermer, J. J. y Pifer, E. A. (2002). Obstacles to answering doctors' questions about patient care with evidence: qualitative study. *British Medical Journal*, 324(7339), 710-716.

- Farquhar, A., Fikes, R. y Rice, J. (1997) citado por Gómez-Pérez, A., Fernández-López, M. y Corcho, O. (2004). *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web*. (2ª ed., pp. 9, 47-52). London: Springer-Verlag.
- Fernández-López, M., Gómez-Pérez, A., Pazos-Sierra, A. y Pazos-Sierra, J. (1999) Building a chemical ontology using METHONTOLOGY and the ontology design environment. *IEEE Intelligent Systems & their applications*, 4(1), 37-46.
- Fernando, B., Savelyich, B. S., Avery, A. J., Sheikh, A., Bainbridge, M., Horsfield, P. y Teasdale, S. (2004). Prescribing safety features of general practice computer systems: evaluation using simulated test cases. *British Medical Journal*, 328(7449), 1171-1172.
- Ferner, R. E. (2004). Commentary: computer aided prescribing leaves holes in the safety net. *British Medical Journal*, 328(7449), 1172-1173.
- Fizman, M., Rindflesch, T. C. y Kilicoglu, H. (2004). Summarization of an online medical encyclopedia. *Studies in Health Technology and Informatics*, 107(Pt 1), 506-510.
- Friedman, C., Alderson, P. O., Austin, J. H., Cimino, J. J. y Johnson, S. B. (1994). A general natural-language text processor for clinical radiology. *Journal of the American Medical Informatics Association*, 1(2), 161-174.
- Friedman, C., Shagina, L., Lussier, Y. y Hripcsak, G. (2004). Automated encoding of clinical documents based on natural language processing. *Journal of the American Medical Informatics Association*, 11(5), 392-402.
- Gaudinat, A., Ruch, P., Joubert, M., Uziel, P., Strauss, A., Thonnet, M., Baud, R., Spahni, S., Weber, P., Bonal, J., Boyer, C., Fieschi, M. y Geissbuhler, A. (2006). Health search engine with e-document analysis for reliable search results. *International Journal of Medical Informatics*, 75(1), 73-85.
- Gómez, J., Cachero, C. y Pastor, O. (2000) Extending an object-oriented conceptual modelling approach to Web application design. advanced information systems engineering. *Proceedings. Lecture Notes in Computer Science*, 1789, 79-93.

- Gómez, J., Cachero, C. y Pastor, O. (2001) Conceptual modeling of device-independent Web applications. *IEEE Multimedia*, 8(2), 26-39.
- Gómez-Pérez, A., Fernández-López, M. y Corcho, O. (2004). *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web*. (2ª ed.). London: Springer-Verlag.
- González-González, A. I., Escortell Mayor, E., Hernández Fernández, T., Sánchez Mateos, J. F., Sanz Cuesta, T. y Riesgo Fuertes, R. (2005). Necesidades de información de los médicos de atención primaria: análisis de preguntas y su resolución. *Atención Primaria*, 35 (8), 419-431.
- González-González, A. I., Sánchez Mateos, J. F., Sanz Cuesta, T., Riesgo Fuertes, R., Escortell Mayor, E. y Hernández Fernández, T. (2006). Estudio de las necesidades de información generadas por los médicos de atención primaria (proyecto ENIGMA). *Atención Primaria*, 38(4), 219-224.
- González-González, A. I., Dawes, M., Sánchez-Mateos, J., Riesgo-Fuertes, R., Escortell-Mayor, E., Sanz-Cuesta, T. y Hernández-Fernández, T. (2007). Information needs and information-seeking behavior of primary care physicians. *Annals of Family Medicine*, 5(4), 345-352.
- Gruber, T. R. (1993). *Toward principles for the design of ontologies used for knowledge sharing*. Technical Report KSL-93-04, Stanford Knowledge Systems Laboratory Report. Recuperado el 10 de Febrero de 2004, desde <http://ksl-web.stanford.edu/knowledge-sharing/papers/onto-design.rtf>
- Grüninger, M. y Fox M. S. (1995) Methodology for the design and evaluation of ontologies. *Proceedings of the IJCAI Workshop on Basic Ontological Issues in Knowledge Sharing*. Montreal: AAAI Press.
- Guarino, N, (Ed.) (1998, Junio). Formal Ontology in information Systems. *Proceedings of Formal Ontology in Information Systems, FOIS'98* (pp. 3-15). Amsterdam: IOS Press. Recuperado el 10 de Febrero de 2004, desde <http://www.loa-cnr.it/Papers/FOIS98.pdf>
- Guarino, N. (1995) Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human-Computer Studies*, 43, 625-640.

- Hermida, J. (2007, Junio) *Desarrollo de un sistema de generación de páginas web etiquetadas semánticamente*. Proyecto final de carrera. Departamento de Lenguajes y Sistemas Informáticos, Escuela Politécnica Superior, Universidad de Alicante, Alicante, España.
- Herrera, J. (2003). *Manual de farmacia clínica y atención farmacéutica* (pp. 2). Madrid: Elsevier España.
- Hersh, W., Price, S. y Donohoe, L. (2000). Assessing thesaurus-based query expansion using the UMLS Metathesaurus. *American Medical Informatics Association (AMIA) Annual Symposium Proceedings*, 344-348.
- Hovy, E. (1997). Citado por Lozano Tello, A. (2002, Febrero). *Métrica de idoneidad de ontologías* (pp. 19-21). Tesis Doctoral. Departamento de Informática. Escuela Politécnica. Universidad de Extremadura, Cáceres, España.
- IDABC, *Decision of the European Parliament and of the Council on interoperable delivery of pan european services to public administrations, businesses and citizens*. (2004). Decision 2004/387/EC. Recuperado el 13 de Abril de 2009, desde <http://ec.europa.eu/idabc/en/document/2319/5644>
- Ide, N. C., Loane, R. F. y Demner-Fushman, D. (2007). Essie: a concept-based search engine for structured biomedical ext. *Journal of the American Medical Informatics Association*, 14(3), 253-263.
- Iglesias, A., Castro, E. Pérez, R., Castaño, L., Martínez, P., Gómez-Pérez, J. M., Kohler, S. y Melero, R. (2008). MOSTAS: Un etiquetador morfo-semántico, anonimizador y corrector de historiales clínicos. *Procesamiento del lenguaje Natural*, 41, 299-300.
- Itami, H. (1987). *Mobilizing invisible assets*. Cambridge: Harvard University Press.
- Jacquemart, P. y Zweigenbaum, P. (2003). Towards a medical question-answering system: a feasibility study. *Studies in Health Technology and Informatics*, 95, 463-468.
- Jagannathan, V., Mullett, C. J., Arbogast, J. G., Halbritter, K. A., Yellapragada, D., Regulapati, S. y Bandaru, P. (2009). Assessment of commercial NLP engines for medication information extraction from dictated clinical notes.

- International Journal of Medical Informatics*, (Epub 2008 Oct 5), 78(4), 284-291, DOI: 10.1016/j.ijmedinf.2008.08.006
- Kogut, P., Cranefield, S., Hart, L., Dutra, M., Baclawski, K., Kokar, M. y Smith, J. (2002). UML for ontology development. *The Knowledge Engineering Review*, 17 (1), 61-64.
- Kondrak, G. y Dorr, B. (2006). Automatic identification of confusable drug names. *Artificial Intelligence in Medicine*, 36(1), 29-42, DOI: 10.1016/j.artmed.2005.07.005
- Lancaster, F. W. (2002). El control del vocabulario en la recuperación de información. (2ª ed.) (pp. 17-19). Valencia: Universitat de València.
- Lassila, O. y McGuinness, D. (2001, Marzo). The role of frame-based representation on the semantic web. *Electronic Transactions in Artificial Intelligence*, 6(5). Recuperado el 10 de Enero de 2006, desde <http://www.ep.liu.se/ea/cis/2001/005/cis01005.pdf>
- Lee, M., Wang, W. y Yu, H. (2006). Exploring supervised and unsupervised methods to detect topics in biomedical text. *BMC Bioinformatics*, 7, 140. Recuperado el 14 de Abril de 2009, desde <http://www.biomedcentral.com/content/pdf/1471-2105-7-140.pdf>
- Lenat, D. B. y Guha, R. V. (1990). Citado por Gómez-Pérez, A. Fernández-López, M y Corcho, O. (2004). *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web* (2ª ed., pp. 113-115). London: Springer Verlag.
- Levin, M. A., Krol, M., Doshi, A. M. y Reich, D. L. (2007). Extraction and mapping of drug names from free text to a standardized nomenclature. *American Medical Informatics Association (AMIA) Annual Symposium Proceedings*, 438-442.
- Lewalle, P.; Rodrigues, J.M., Zanstra, P., et al. Ustun, B., Kalra, D., Surjan G., Rector, A., Stroetmenn, V. y Virtanen, M. (2008). A deployment and research roadmap for semantic interoperability: the EU SemanticHEALTH project. En S. K. Andersen, G. O. Klein, S. Schulz, J. Aarts y M. C. Mazzoleni (Eds.), *eHealth Beyond the Horizon – Get IT There* (pp. 635-640). Amsterdam: IOS Press.

- Liu, F., Ackerman, M. y Fontelo, P. (2006). BabelMeSH: development of a cross-language tool for MEDLINE/PubMed. *American Medical Informatics Association (AMIA) Annual Symposium Proceedings*, 1012.
- Liu, K., Mitchell, K. J., Chapman, W. W. y Crowley, R. S. (2005). Automating tissue bank annotation from pathology reports - comparison to a gold standard expert annotation set. *American Medical Informatics Association (AMIA) Annual Symposium Proceedings*, 460-464.
- Lowe, H. J., Antipov, I., Hersh, W., Smith, C. A. y Mailhot, M. (1999). Automated semantic indexing of imaging reports to support retrieval of medical images in the multimedia electronic medical record. *Methods of Information in Medicine*, 38(4-5), 303-307.
- Loza, M. I., Cordero, L., Fernández-Llimós, F., García, P., Cadavid, M. I., Sanz, F. y Calleja, J. M. (2000). Fuentes de Información sobre medicamentos utilizadas por los farmacéuticos comunitarios de Galicia. *Pharmaceutical Care España*, 2, 108-122.
- Lozano Tello, A. (2002, Febrero). *Métrica de idoneidad de ontologías*. Tesis Doctoral. Departamento de Informática. Escuela Politécnica. Universidad de Extremadura, Cáceres, España.
- Maña, M., Mata, J., Domínguez, J. L., Vaquero, A. y Álvarez, F (2006). Los proyectos SINAMED e ISIS: mejoras en el acceso a la información biomédica mediante la integración de generación de resúmenes, categorización automática de textos y ontologías. *Procesamiento del lenguaje Natural*, 37, 353-354.
- Martín, M. T., Montejo, A., Díaz, M. C. y Ureña, L. A. (2007). Integración de conocimiento en un dominio específico para categorización multietiqueta. *Procesamiento del lenguaje Natural*, 39, 63-70.
- MeSH, *Medical Subject Headings introduction* (2004, Diciembre). U.S. National Library of Medicine (NLM). Recuperado el 5 de Diciembre de 2004, desde <http://www.nlm.nih.gov/mesh/introduction2004.html>
- Meystre, S. y Haug, P. J. (2006). Natural language processing to extract medical problems from electronic clinical documents: performance evaluation. *Journal of Biomedical Informatics*, 39(6), 589-599.

- Miralles i Torner, F. (2000). *Introducció als sistemes d'informació*. [CD-Rom]. Barcelona: Universitat Oberta de Catalunya.
- Mizoguchi, R., Ikeda, M. y Sinita, K. (1997). Roles of share ontology AI-ED research, intelligence, conceptualization, standardization, and reusability. En B. de Boulay y R. Mizoguchi (Eds.), *Artificial Intelligence and Education, Proceedings of AIED*. (pp. 537-544). Amsterdam: IOS Press.
- Monteagudo Peña, J. L. (Coord.) (2001). *El marco de desarrollo de la e-Salud en España*. Madrid: Instituto de Salud Carlos III, Ministerio de Sanidad y Consumo. Recuperado el 28 de Febrero de 2006, desde http://bvs.isciii.es/mono/pdf/UCIS_01E.pdf
- Muller, H. M., Kenny, E. E. y Sternberg, P. W. (2004). Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS Biology*, 2(11), e309, DOI: 10.1371/journal.pbio.0020309
- Muñoz-Izquierdo, A., Puchades-Simó, A., Marco-Gisbert, A., Piquer-Bosca, C., Ferrer-Casares, E., Canela-Ferrer, A., Gómez-Muñoz, N. y Costa-Pastrana, M. C. (2008) Acceso a Internet de los profesionales de enfermería y tipo de consultas que realizan. *Enfermería Clínica*, 18(1), 16-23.
- Nguyen, B. (2006, 6 Junio). *Semantic Web best practices and deployment working group. Semantics Web tutorials*. World Wide Web Consortium (W3C). Recuperado el 12 de Enero de 2008, desde <http://www.w3.org/2001/sw/BestPractices/Tutorials>
- Digitalis, *Base de datos Nomenclátor Digitalis* (s.f.). Dirección General de Farmacia y Productos Sanitario. Ministerio de Sanidad y Política Social. Recuperado el 30 de Noviembre de 2008, desde <http://www.msc.es/profesionales/farmacia/nomenclator/DI.htm>
- Noy, N. F. y McGuinness, D. L. (2001, Marzo). *Ontology development 101: a guide to creating your first ontology*. Stanford Knowledge Systems Laboratory, Technical Report KSL-01-05 and Stanford Medical Informatics, Technical Report SMI-2001-0880. Recuperado el 12 de Noviembre de 2007, desde http://protege.stanford.edu/publications/ontology_development/ontology101.pdf

- Obrst, L., Clusters, W., Mani, I., Ray, S. y Smith, B. (2007). The evaluation of ontologies. Toward improved semantic interoperability. En C. J. O. Baker y K. Cheung (Eds.), *Semantic Web: revolutionizing knowledge discovery in the life sciences* (pp. 139-158). New York: Springer.
- Orueta, R. (2004). Utilidad de Internet en información de medicamentos para atención primaria. *Información Terapéutica del Sistema Nacional de Salud*, 28(3), 70-74.
- Pakhomov, S. V., Buntrock, J. D. y Chute, C. G. (2006). Automating the assignment of diagnosis codes to patient encounters using example-based and machine learning techniques. *Journal of the American Medical Informatics Association*, 13(5), 516-525.
- Perea, J. M., Martín, M. T., Montejo, A. y Díaz, M. C. (2008). Categorización de textos biomédicos usando UMLS. *Procesamiento del lenguaje Natural*, 40, 121-127.
- RAE, Real Academia Española. (2001). *Diccionario de la lengua española* [en línea]. Vigésimo segunda edición. Recuperado el 13 de Abril de 2009, desde http://buscon.rae.es/draeI/SrvltConsulta?TIPO_BUS=3&LEMA=ontología
- RD, *Real Decreto 1345/2007, de 11 de octubre, por el que se regula el procedimiento de autorización, registro y condiciones de dispensación de los medicamentos de uso humano fabricados industrialmente*. (2007, 7 de Noviembre). Boletín Oficial del Estado, 267, 45652-45698.
- Rector, A. (Coord). (2008, Junio). *Barriers, approaches and research priorities for integrating biomedical ontologies* (Deliverable D6.1). Semantic HEALTH. Recuperado el 8 de Febrero de 2009, desde http://www.semantichealth.org/DELIVERABLES/SemanticHEALTH_D6_1.pdf
- Reeve, L. H., Han, H. y Brooks, A. D. (2007). Biomedical text summarisation using concept chains. *International Journal of Data Mining and Bioinformatics*, 1(4), 389-407.
- Richesson, R. L., Andrews, J. E. y Krischer, J. P. (2006) Use of SNOMED CT to represent clinical research data: a semantic characterization of data items on case report forms in vasculitis research. *Journal of the American Medical Informatics Association*, 13(5), 536-546.

- RIDE (2007, Diciembre). *Standardization efforts for providing semantic interoperability in eHealth domain* (Deliverable D.2.2.1). Project A Roadmap for Interoperability of eHealth Systems in Support of COM 356 with Special Emphasis on Semantic Interoperability. Recuperado el 8 de Febrero de 2009, desde <http://www.srdc.metu.edu.tr/webpage/projects/ride/deliverables/RIDE-D2.2.1-standards-09.doc>
- Rindflesch, T. C., Fiszman, M. y Libbus, B. (2005). Semantic interpretation for the biomedical literature. En H. Chen, S. Fuller, W. R. Hersh y C. Friedman (Eds.), *Medical informatics: advances in knowledge management and data mining in biomedicine* (pp. 399-422). New York: Springer-Verlag.
- Robinson, S. (Coord.) (2007, Diciembre). *Conceptual framework for eHealth interoperability* (Deliverable D1.1). Semantic HEALTH. Recuperado el 8 de Febrero de 2009, desde http://www.semantichealth.org/DELIVERABLES/SemanticHEALTH_D1_1_finalC.pdf
- Robinson, S. (Coord.) (2008, Junio). *Draft semantic interoperability deployment and research roadmap* (Deliverable D7.1). Semantic HEALTH. Recuperado el 8 de Febrero de 2009, desde http://www.semantichealth.org/DELIVERABLES/SemanticHEALTH_D7_1_draft_21-05-08.pdf
- Rodríguez, H. (2000). Adquisición automática y uso de taxonomías de amplia cobertura. Tutorial. *Workshop on Automatic Acquisition of linguistic Knowledge*. San Millán. Recuperado el 20 de Octubre de 2003, desde <http://www.lsi.upc.es/~horacio/varios/sanmillan2000.zip>
- Romá-Ferri, M.T. y Palomar, M. (2005). Interoperabilidad semántica de ontologías basada en técnicas de procesamiento del lenguaje natural (PLN). En: J. Gascón, F. Burguillos, A. Pons. *La dimensión humana de la organización del conocimiento. 7º Congreso International Society for Knowledge Organization (ISKO)-España*. (pp. 534-548). Barcelona: Publicaciones de la Universidad de Barcelona. Departamento de Biblioteconomía y Documentación.
- Romá-Ferri, M. T. y Palomar, M. (2008). Análisis de terminologías de salud para su utilización como ontologías computacionales en los sistemas de información clínicos. *Gaceta Sanitaria*, 22(5), 421-433.

- Ruiz, M. E. (2006). Combining image features, case descriptions and UMLS concepts to improve retrieval of medical images. *American Medical Informatics Association (AMIA) Annual Symposium Proceedings*, 674-678.
- Sánchez, D.M., Cavero, J.M. y Marcos, E. (2005). Modelos y ontologías en informática. En J. M. Cavero, B. Vela y E. Marcos (Eds.), *Aspectos filosóficos, psicológicos y metodológicos de la informática* (pp. 97-110). Vol. 8 Ciencias Experimentales y Tecnológicas. Madrid: Dykinson.
- Segura, I., Martínez, P. y Samy, D. (2008). Detección de fármacos genéricos en textos biomédicos. *Procesamiento del lenguaje Natural*, 40, 27-34.
- Sharp, M., Bodenreider, O. y Wacholder, N. (2008). A framework for characterizing drug information sources. *American Medical Informatics Association (AMIA) Annual Symposium Proceedings*; 2008, 662-666.
- Slaughter, L. A., Soergel, D. y Rindflesch, T. C. (2006). Semantic representation of consumer questions and physician answers. *International Journal of Medical Informatics*, 75(7), 513-529, DOI: 10.1016/j.ijmedinf.2005.07.025.
- Smith, B. (2003). Ontology. En L. Floridi (Ed.), *Blackwell guide to the philosophy of computing and information* (pp. 155-166). Oxford: Blackwell. Recuperado el 18 de Abril de 2009, desde http://ontology.buffalo.edu/smith/articles/ontology_pic.pdf
- Smith, B., Kusnierczyk, W., Schober, D. y Ceusters, W. (2006). Towards a reference terminology for ontology research and development in the biomedical domain. En O. Bodenreider (Ed.), *Proceedings of KR-MED. Biomedical Ontology in Action* (pp. 57-66). November, 8. Baltimore MD (USA). Recuperado el 18 de Abril de 2009, desde http://ontology.buffalo.edu/bfo/Terminology_for_Ontologies.pdf
- SNOMED CT (Clinical Terms). (s. f.). *Publications*. International Health Terminology Standards Development Organisation (IHTSDO). Recuperado el 4 de Noviembre de 2008, desde <http://www.ihtsdo.org/snomed-ct/snomed-ct-publications/>
- Sobrado, F. J., Pikatza, J. M., Larburu, I. U., García, J. J. y López de Ipiña, D. (2004) Towards a clinical practice guideline implementation for asthma treatment. En R. Conejo, M. Urretavizcaya y J. L. Pérez de la Cruz (Eds.),

- Current topics in artificial intelligence, Lecture Notes in Artificial Intelligence* (pp. 587-596). Volumen 3040. Berlin: Springer-Verlag.
- Sowa, J. F. (1999) citado por Gómez-Pérez, A., Fernández-López, M. y Corcho, O. (2004). *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web*. (2ª ed., pp. 75-76). London: Springer-Verlag.
- Srinivasan, P. (1996). Retrieval feedback in MEDLINE. *Journal of the American Medical Informatics Association*, 3(2), 157-167.
- Stead, W. W., Miller, R. A., Musen, M. A. y Hersh, W. R. (2000). Integration and beyond: linking information from disparate source and into workflow. *Journal of the American Medical Informatics Association*, 7, 135-45.
- Studer, R., Bejamins, V. R. y Fensel D. (1998, Marzo). Knowledge engineering: principles and methods. *Data & Knowledge Engineering*, 25(1-2), 161-197. Recuperado el 8 de Febrero de 2009, desde [http://dx.doi.org/10.1016/S0169-023X\(97\)00056-6](http://dx.doi.org/10.1016/S0169-023X(97)00056-6)
- Swartout, B., Patil, R., Knight K. y Russ, T. (1997). Toward distributed use of large-scale ontologies. En A. Farquhar, M. Grüninger, A. Gómez-Pérez, M. Uschold y P. van der Vet (Eds.), *Proceedings of spring symposium on ontological engineering of AAAI*. (pp. 138-148), Stanford University (California). Recuperado el 10 de Febrero de 2004, desde http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html
- Swartout, W. y Tate, A. (1999, Enero-Febrero). Ontologies. *IEEE Intelligent Systems*, 18-19.
- Tabuada, M., González, J. L., Argüello, M., Des, J., Mira, J. y Martínez, D. (2004). Adquisición de conocimiento médico para aplicaciones en telemedicina. En F. González y C. Zamarrón (Eds.), *Telemedicina. Aplicaciones y nuevas tecnologías*. Santiago de Compostela: Sociedad Gallega de Telemedicina. Recuperado el 27 de Marzo de 2006, desde <http://aiff.usc.es/~elchus/personal/download/xornadastabuada.pdf>
- Tartir, S., Arpinar, I. B., Moore, M., Sheth, A. P. Aleman-Meza, B. (2005). OntoQA: metric-based ontology quality analysis. En D. Caragea, V. Honavar, I. Músla y R. Ramakrishnan (Eds.), *Knowledge acquisition from*

distributed, autonomous, semantically heterogeneous data and knowledge sources (KADASH). Proceedings of Workshop held in conjunction with 2005 IEEE International Conference on Data Mining. Houston, USA. Recuperado el 4 de Abril de 2009, desde <http://www.cild.iastate.edu/events/KADASH.pdf>

Terol, R. M., Martínez-Barco, P. y Palomar, M. (2006). Aplicación de técnicas basadas en PLN al tratamiento de preguntas médicas en búsqueda de respuestas. *Procesamiento del lenguaje Natural*, 36, 17-24.

Terol, R. M., Martínez-Barco, P. y Palomar, M. (2007). A knowledge based method for the medical question answering problem. *Computers in biology and medicine*, 37(10), 1511-1521.

Terol, R. M. (2009). *Representación del conocimiento textual mediante técnicas lógico-conceptuales en aplicaciones de tecnologías del lenguaje humano*. Tesis Doctoral. Departamento de Lenguajes y Sistemas Informáticos. Escuela Politécnica. Universidad de Alicante, Alicante, España.

GALEN, *The OpenGALEN Common Reference Model*. (s. f.) OpenGALEN. Recuperado el 16 de Febrero de 2007, desde <http://www.opengalen.org/themodel/structure.html>

UMLS, *Unified Medical Language System*. (2004, Marzo; 2008, Julio). U.S. National Library of Medicine (NLM). Recuperado el 5 de Diciembre de 2004 y el 18 de Abril de 2009, desde http://www.nlm.nih.gov/research/umls/about_umls.html

Uschold, M. y Grüninger, M. (1996, Febrero). Ontologies: principles, methods and applications. *Knowledge Engineering Review*, 11(2), 93-155. Recuperado el 8 de Febrero de 2009, desde <https://eprints.kfupm.edu.sa/55793/1/55793.pdf>

Uschold, M. y Jasper, R (1999). Citado por Gómez-Pérez, A. Fernández-López, M. y Corcho, O. (2004). *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web* (2ª ed., pp.8). London: Springer Verlag

Uschold, M. y King, M. (1995). Citado por Gómez-Pérez, A. Fernández-López, M. y Corcho, O. (2004). *Ontological engineering: with examples from the areas*

- of knowledge management, e-commerce and the semantic web* (2ª ed., pp.115-119). London: Springer Verlag.
- Uzuner, O., Goldstein, I., Luo, Y. y Kohane, I. (2008). Identifying patient smoking status from medical discharge records. *Journal of the American Medical Informatics Association*, 15(1), 14-24.
- Vademécum Internacional, información de medicamentos.* (s.f.). M Vademécum.es. CMP Medica España. Recuperado el 9 de Septiembre de 2007, desde <http://www.vademecum.es/index.cfm>
- Van Heijst, G., Schreiber, A. T., y Wielinga, B. J. (1997) citado por Gómez-Pérez, A., Fernández-López, M. y Corcho, O. (2004). *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web.* (2ª ed., pp. 29-34). London: Springer-Verlag.
- Villaseñor, J. (1998). Los instrumentos para la recuperación de la información: las fuentes. En I. Torres Ramírez (Coord.), *Las fuentes de información: estudios teóricos-prácticos*. Madrid: Síntesis.
- Wedgwood, J. (2005). MQAF: a medical question-answering framework. *American Medical Informatics Association (AMIA) Annual Symposium proceedings*, 1150.
- Whalen, G. (2005). Medical textbook summarization and guided navigation using statistical sentence extraction. *American Medical Informatics Association (AMIA) Annual Symposium proceedings*, 814-818.
- Yamamoto, Y. y Takagi, T. (2007). Biomedical knowledge navigation by literature clustering. *Journal of Biomedical Informatics*, 40(2), 114-130.
- Yu, H. y Lee, M. (2006). Accessing bioscience images from abstract sentences. *Bioinformatics*, 22(14), e547–e556, DOI: 10.1093/bioinformatics/btl261
- Zhang, D., Roderer, N. K., Huang, G. y Zhao, X. (2006). Developing a UMLS-based indexing tool for health science repository system. *American Medical Informatics Association (AMIA) Annual Symposium proceedings*, 1157.



Anexos.

Universitat d'Alacant
Universidad de Alicante

ANEXO I: Artículos y otras fuentes documentales seleccionadas

1. Selección de artículos: cuestiones planteadas para la obtención de información específica.

Profesionales de medicina:

Ely JW, Osheroff JA, Ebell MH, Bergus GR, Levy BT, Chambliss ML, Evans ER. Analysis of questions asked by family doctors regarding patient care. *BMJ*. 7 August 1999; 319:358-61.

Ely JW, Osheroff JA, Gorman PN, Ebell MH, Chambliss ML, Pifer EA, Stavri PZ. A taxonomy of generic clinical questions: classification study. *BMJ*. 12 August 2000; 321: 429-32.

Ely JW, Osheroff JA, Ebell MH, Chambliss ML, Vinson DC, Stevermer JJ, Pifer EA. Obstacles to answering doctors' questions about patient care with evidence: qualitative study. *BMJ*. 23 March 2002; 324:1-7

González-González AI, Escortell Mayor E, Hernández Fernández T, Sánchez Mateos JF, Sanz Cuesta T, Riesgo Fuertes R. Necesidades de información de los médicos de atención primaria: análisis de preguntas y su resolución. *Aten Primaria*. 2005; 35(8):419-31.

González-González AI, Sánchez Mateos JF, Sanz Cuesta T, Riesgo Fuertes R, Escortell Mayor E, Hernández Fernández T. Estudio de las necesidades de información generadas por los médicos de atención primaria (proyecto ENIGMA). *Aten Primaria*. 2006; 38(4):219-24.

Profesionales de medicina y de enfermería:

McKnight LK, Stetsob PD, Bakken S, Curran C, Cimino JJ. Perceived information needs and communication difficulties of inpatient physicians and nurses. *JAMIA*. Nov/Dec Suppl 2002; 9(6): 64-9.

Profesionales de farmacología

Baixauli VJ, Salar L, Barbero A. Demanda de información en la farmacia comunitaria. *Pharm Care Esp*. 2004; 6(3): 136-44

Espejo J, Fernández-Llimós, F, Machuca M, Faus MJ. Problemas relacionados con medicamentos: Definición y propuesta de inclusión en la Clasificación Internacional de Atención Primaria (CIAP) de la WONCA. *Pharm Care Esp*. 2002; 4(2): 122-7.

Loza MI, Cordero L, Fernández-Llimós F, García Corral P, Cadavid MI, Sanz F, Calleja JM. Fuentes de información sobre medicamentos usadas por farmacéuticos comunitarios en Galicia. *Pharm Care Esp*. 2000; 2:108-22.

Estudiantes de medicina:

Wildemuh BM, Bliet R, Friedman Ch P, Miya TS. Information-seeking behaviors of medical students: a classification of questions asked of librarians and physicians. *Bull Med Libr Assoc*. Jul 1994; 82(3):295-304.

Consumidores:

Slaughter LA, Soergel D, Rindfleisch, TC. Semantic representation of consumer questions and physician answers. *IJMI*. 2006; 75:513-29.

2. Selección de artículos: sistemas de información farmacológicos.

Duclos C, Cartolano GL, Ghez M, Venot A. Structured representation of the pharmacodynamics section of the summary of product characteristics for antibiotics: application for automated extraction and visualization of their antimicrobial activity spectra. *JAMIA*. 2004; 11(4): 285-93.

Duclos-Cartolano C, Venot A. Building and evaluation of a structured representation of pharmacokinetics information presented in SPCs: from existing conceptual views of pharmacokinetics associated with natural language processing to object-oriented design. *JAMIA*. 2003; 10(3):271-80.

Luna D, González F, Martínez M, Gambarte M, Morchón A, Costa G, López M. ¿Por qué se ignoran los alertas sobre interacciones farmacológicas? *Information and Communication Technologies in Healthcare Development. 3rd Virtual Congreso in Internet. Informedica 2004*. March 1st to 30th, 2004. Disponible en: http://www.informaticamedica.org/I04/papers/luna_62.pdf

3. Selección de artículos: sistemas de pregunta-respuestas con contenido farmacológico.

Demner-Fushman D, Lin J. Answer extraction, semantic clustering and extractive summarization for clinical questions answering. *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*. Sydney, July 2006. Association for Computational Linguistics: 841-48.

Demner-Fushman D, Lin J. Situated question answering in the clinical domain: selecting the best drug treatment for diseases. *In Proceedings of COLING/ACL 2006 Workshop on Task-Focused. Summarization and Question Answering*. July 2006, Sydney (Australia).

4. Selección de manuales de procedimientos de enfermería.

- Chocarro González, L. *Procedimientos y cuidados en enfermería médico-quirúrgica*. Madrid: Elsevier, 2006.
- Smeltzer, S.C. y Bare, B.G. Brunner y Suddarth, *Enfermería medicoquirúrgica* (10ªed). México: McGraw-Hill Interamericana, 2005.

5. Selección de manuales de farmacología, medicina y diccionarios específicos.

- Braunwald, E. (Ed). *Harrison, principios de medicina interna*. México: McGraw-Hill, 2006.
- Castells Molina, S. y Hernández Pérez, M. *Farmacología en enfermería* (2ª ed). Madrid: Elsevier, 2007.
- Diccionari enciclopèdic de medicina* (2ªed). Disponible en: <http://www.grec.net/home/cel/mdicc.htm>
- Diccionario Médico Roche*. Barcelona: Doyma, 1993.
- Fernández Alfonso, M.S. y Ruiz Gayo, M. (Coord). *Fundamentos de farmacología básica y clínica*. Editorial Ramón Areces, 2005.
- Gené Badia, J. y Font Pons, M. Utilización de los medicamentos. En: Martín Zurro, A. y Cano Pérez, J.F. *Atención Primaria. Conceptos, organización y práctica clínica* (3ªed). [CD-Rom]. Harcourt Brace y Doyma, 1997.
- Herrera Carranza, J. (Ed). *Manual de farmacia clínica y atención farmacéutica*. Elsevier España, 2003.
- Hitner, H. y Nagle, M.B *Introducción a la farmacología*. McGraw-Hill Interamericana, 2007.
- Manual Merck de información médica para el hogar*. Madrid: Merck Sharp & Dohme (MSD), 2005. Recuperado Agosto 2007, desde http://www.msd.es/publicaciones/mmerck_hogar/index.html

6. Selección de informes técnicos y prospectos:

- Base de Datos del Medicamento*. (s.f.). Consejo General de Colegios Oficiales de Farmacéuticos (CGCOF). Disponible en: <http://www.portalfarma.com/pfarma/espacioges.nsf/wfAccesoEspacioGestion?OpenForm>.
- Díez MV, Errecalde MF. Aclaraciones al concepto de genérico. *Información Terapéutica del SNS*. 1998; 22(3): 68-72 Recuperado Agosto 2007, desde <http://www.msc.es/biblioPublic/publicaciones/docs/generico.pdf>
- European Public Assessment Report (EPAR) for authorised medicinal products for human use*. European Medicines Agency (EMA). Recuperado Agosto 2007, desde <http://www.emea.europa.eu/htms/human/epar/a.htm>
- Ficha Técnica: Contenido e importancia de la prescripción. Notas de Interés. *Información Terapéutica del SNS*. 2006; 30(4):114-7 Recuperado Agosto

2007, desde http://www.msc.es/biblioPublic/publicaciones/docs/vol30_4NotasInteres.pdf

Medicamentos autorizados en España (uso humano). Agencia Española de Medicamentos y Productos Sanitarios. Ministerio de Sanidad y Consumo. Recuperado Agosto 2007, desde <https://sinaem4.agemed.es/consaem/fichasTecnicas.do?metodo=detalleForm>

Prospectos de medicamentos, con principio activo: *Ácido Acetilsalicílico, Metamizol Magnésico, Diazepam y Amoxicilina*.

Vademécum Internacional. (s.f.). El buscador de medicamentos, M Vademécum.es. CMP Medica España. Disponible en: <http://www.vademecum.es/index.cfm>



Universitat d'Alacant
Universidad de Alicante

ANEXO II: Total de términos recopilados de las fuentes documentales.

Acción negativa	Concentración en	Epidural	Nombre	Resistente
Aconsejar	sangre	Especialidad	genérico	Riesgo
Actúa en	Conservación	farmacéutica	Oftálmica	Ritmo gotero
Administrar	Contraindicaciones	Estabilidad	Oral	Ritmo perfusión
Adulto	Controles	Excipiente	Ótica	Seguridad
Advertencias	Coste	Farmacocinética	Parenteral	Signos
Afecta a	Cumplimiento	Farmacodinámica	Pauta	Síntomas
Agentes causales	Datos clínicos	Forma farmacéutica	tratamiento	Sirve para
Alergia	Denominación	Formas administrar	Período	Sobredosis
Anciano	común internacional	Fotosensible	tratamiento	Subcutánea
Antibiograma	Denominación	Frecuencia	Población diana	Sublingual
Asociados con	oficial española	Genérico	Posología	Susceptibilidad
Bacterias	Diluyente	Gestante	Precauciones	Sustancia
aerobias	Disolución	Gotas	Precio	Termolábil
Bacterias	Disponibilidad	Grupo	Preparación	Tiempo
anaerobias.	DOE	especialidades	fármaco	Tipo bacterias
Bacterias gram –	Dosis	Grupo terapéutico	Prescripción	Tomar
Bacterias gram +	Dosis diaria	Hongos	Prevenir	Tópica
Bactericidas	Dosis eficaz	Incompatibilidades	Principio activo	Toxicidad
Bacterioestático	Dosis máxima	Indicación	Procedimiento	Tóxico
Beneficioso	Dosis mínima	terapéutica	administración	Transdérmica
Cantidad	Dosis terapéutica	Indicado para	Procedimiento	Unidad de
Características	Dosis continua	Ingredientes	especial	medida
físicas	Dosis intermiten	Inhalatoria	Produce	Velocidad de
Características	Bolo	Interacciones	Profiláctico	flujo
preparado	Duración	Intranasal	Pronóstico	Via
Causadas por	tratamiento	Intravenosa	Propiedades	Vial
Clase terapéutica	Efectividad	Lactante	farmacológicas	Vida media
Composición	Efecto colateral	Lugar efecto	Pruebas	Virus
cualitativa	Efecto fármacos	Manipulación	laboratorio	
Composición	Efectos adversos	Medicamento	Quimioterapia	
cuantitativa	Efectos secundarios	Niño	Reacción	
Composición	Elección para	Niveles plasmáticos	adversa	
química	Endotraqueal	Nombre comercial	Recetar	
Concentración	Enfermedades	Nombre del fármaco	Recomendar	

ANEXO III: Un fragmento del glosario creado.

Administración: Acció d'aplicar a algú algun remei, sia per via oral o parenteral, amb una gran varietat de possibilitats, sia per moltes altres vies especials: subcutània, intramuscular, intravenosa, òtica, ocular, sublingual, lumbar, etc.

Administrar: Suministrar, proporcionar o distribuir algo. Aplicar, dar o hacer tomar un medicamento. Graduar o dosificar el uso de algo, para obtener mayor rendimiento de ello o para que produzca mejor efecto.

Advertencias y precauciones especiales de empleo: en este apartado se describen las contraindicaciones relativas, es decir, aquellas en las que debe de tenerse un cuidado especial si se administra el medicamento, así como las precauciones de uso consideradas importantes a tener en cuenta antes de utilizar el medicamento (circunstancias o grupos de pacientes en los que es esperable una reacción adversa o riesgos asociados al tratamiento) y las medidas que pueden tomarse en estos casos.

Composición cualitativa y cuantitativa: especifica el principio activo así como la cantidad que contiene, expresada de diferente manera según la forma farmacéutica del medicamento, pudiendo ser: por unidad de dosis, por unidad de volumen o por unidad de peso. También se hace una pequeña referencia a la composición cualitativa en excipientes.

Concentración: La cantidad de medicamento diluida en un determinado volumen de solución líquida.

Contraindicaciones: incluye las contraindicaciones absolutas, es decir, las circunstancias en las cuales no debe de administrarse el medicamento, y los antecedentes de hipersensibilidad, al principio activo o a cualquiera de los componentes de la formulación.

Datos Clínicos: indica las características clínicas de los pacientes que pueden recibir el medicamento en cuestión y aquellas circunstancias que han de tenerse en cuenta para optimizar el balance beneficio-riesgo del medicamento.

Datos farmacéuticos: detalla aquéllas características que desde el punto de vista químico-farmacéutico son relevantes para un correcto uso: Lista de

excipientes, Incompatibilidades, Período de validez, Precauciones especiales de conservación, Naturaleza y contenido del envase e Instrucciones de eliminación (antes denominado Instrucciones de uso y manipulación).

Diluyente: solución líquida necesaria para disolver una cantidad de soluto. Suele ser o bien agua (H₂O) o suero fisiológico (ClNa), aunque puede utilizarse otros diluyentes específicos.

Dosis: 1 quantitat d'una substància que entra en la composició d'un medicament. 2 quantitat d'un remei (droga, radiació, etc) que és administrada en una unitat de temps o durant tot el curs d'un tractament.

Dosis: toma de medicina que se da al enfermo cada vez.

Dosi fraccionada: cadascuna de les parts en què hom divideix la dosi total i que són administrades a intervals prefixats.

Efecto Colateral: efecte que acompanya l'acció principal d'un medicament. Quan és irreversible requereix un control mèdic estricte, que arribi fins i tot a suprimir la medicació. De vegades hom aprofita l'efecte col·lateral com a efecte principal.

Efecto Secundario: acció no desitjable, sovint de tipus nociu, produïda per un medicament. En són exemples l'acció ulcerativa de l'aspirina sobre els budells, l'acció virilitzant de certs anabolitzants i l'acció de les estreptomicines sobre el nervi acústic.

Enfermedad: nom genèric de tot procés morbós considerat des de la causa inicial fins a les darreres conseqüències. En l'estudi de tota malaltia (patologia) cal tenir en compte les causes (etiologia), la forma d'actuar de la causa morbosa sobre l'organisme (patogènia), les manifestacions clíniques que provoca (semiologia), les alteracions macroscòpiques i microscòpiques dels òrgans i dels teixits com a resultat de l'acció morbosa (anatomia patològica), la identificació del procés morbós (diagnosi), el judici anticipat sobre la seva probable fi (pronòstic) i el tractament profilàctic i curatiu (teràpèutica).

Especialidad Farmacéutica: medicament presentat en forma uniforme, precintat i amb un nom enregistrat, la fórmula medicamentosa del qual és també enregistrada i declarada en el prospecte adjunt.

ANEXO IV: Descripción informal de la ontología OntoFIS.

ID	Nombre Clase	Descripción	Ejemplo	Propiedades
1	Fármaco	Incluye la denominación del producto que llega al usuario final, como una unidad de comercialización. Su patrón se basa en un Nombre_Fantasia que se constituye con los siguientes elementos: Nombre Registrado + Cantidad + Forma Farmacéutica	<p>(1)</p> <p>-Augmentine 500/125mg Comprimidos</p> <p>-Ibuprofeno Davur 600mg comprimidos EFG</p> <p>-Nolotil cápsulas 575 mg</p> <p>-NOLOTIL Sol. iny. 2 g</p> <p>(2) Código adjudicado por el Ministerio que identifica al fármaco autorizado para su comercialización en España.</p> <p>(3) Este nombre puede cambiar de un país a otro.</p>	nombre_farmaco_españa (1)
				codigo_nacional (2)
				nombre_farmaco_no_españa (3)
				numero_contraindicacion
				numero_posologia
				numero_excipiente
				sinonimo_castellano_nombre_clase: Medicamento = Medicina = Especialidad Farmacéutica = Preparación Farmacéutica = Producto Farmacéutico = Nombre Presentación.
				sinonimo_ingles_nombre_clase: Medicament = Medicine = Medicinal product = Pharmaceutical preparation
				sinonimo_aleman_nombre_clase: Medikament = Arznei = Arzneimittels
				sinonimo_frances_nombre_clase: Médicament
				sinonimo_italiano_nombre_clase: Medicamento = Medicinale

ID	Nombre Clase	Descripción	Ejemplo	Propiedades
2	Composicion_ química	Se refiere a la denominación química de la sustancia que tiene acción terapéutica NOTA: El ejemplo se refiere al nolotil y su composición química sería la misma para el fármaco 'Nolotil cápsulas' y 'Nolotil sol iny 2g'	(1) Acido [(2,3-dihidro-1,5-dimetil-3-oxo-2-fenil-1H-pirazol-4-il) metilamino]metanosulfónico. (2) H 5.50% C 50.15% N 13.50% O 20.55% S 10.30% (3) C ₁₃ H ₁₇ N ₃ O ₄ S C ₁₆ H ₁₉ N ₃ O ₅ S + C ₈ H ₉ NO ₅ Las fórmulas empíricas son nolotil y amoxicilina+ac. Clavulánico (4) 68-89-3 (5) Fluroquilonas. Ver reseña en concepto Grupo_terapeutico (ID 10). (6) JO1MA02	nombre_denominacion (1)
				nombre_composicion_cuantitativa (2)
				nombre_formula_empirica (3)
				codigo_chemical (4)
				grupo_quimicoOMS (5)
				codigo_quimicoOMS(6)
				Sinonimo_castellano_nombre_clase: Composición
				Sinonimo_ingles_nombre_clase: Composition
				Sinonimo_aleman_nombre_clase: Zusammensetzung
3	Principio_activo	Denominación genérica de la sustancia responsable de la acción terapéutica	(1)- Metamizol -Amoxicilina + Acido Clavulánico (2) tiene como posible sinónimo el de concentración o dosis por unidad administrada.	nombre_generico (1)
				cantidad(2)
				unidad_medida
				sinonimo_castellano_nombre_clase: Sustancia = Sustancia Activa
				sinonimo_ingles_nombre_clase: active substance = drug
				sinonimo_aleman_nombre_clase
				sinonimo_frances_nombre_clase
				sinonimo_italiano_nombre_clase

ID	Nombre Clase	Descripción	Ejemplo	Propiedades
4	Excipiente	<p>Sustancia química inactiva terapéuticamente que es usada como vehículo o estabilizador de la sustancia activa.</p> <p>Suele haber varios pero, se destacan aquellos que tiene un factor de interacción o reacción en el organismo para algunos pacientes (alergias) y sus pruebas de laboratorio</p>	<p>(1) En general: Almidones, azúcares y celulosas como Hidroxipropil celulosa o lactosa. También se utilizan azúcares alcohólicas como Xilitol, sorbitol o maltitol.</p> <p>Nolotil: Estearato de magnesio.</p> <p>Componentes de la cápsula: indigotina (E132), eritrosina (E127), dióxido de titanio (E171) y gelatina.</p>	nombre_generico (1)
				Cantidad
				unidad_medida
				sinonimo_castellano_nombre_clase: Vehículo = Transportador = Diluyente (cuando es líquido) = Sustancia Auxiliar
				sinonimo_ingles_nombre_clase: Component = Excipient
				sinonimo_aleman_nombre_clase: Konstituens = Bestandteile
				Sinonimo_frances_nombre_clase: Excipient
5	Forma farmacéutica	<p>Estado físico en que se presenta un producto farmacéutico para facilitar su fraccionamiento, clasificación, administración o empleo</p> <p>NOTA: Toma valores de una lista extensa, pero finita.</p>	<p>(1) Cáps: cápsulas (2) Sol: solución</p> <p>Iny: inyectable...</p> <p>(3) Una sola forma puede tener más de una vía. Ejemplo: vial o inyectable puede tener como vía la IM, IV y Oral (caso Nolotil).</p>	nombre_abreviado_forma_f (1)
				nombre_completo_forma_f (2)
				Característica (3)
				sinonimo_castellano_nombre_clase
				sinonimo_ingles_nombre_clase: Pharmaceutical form
				sinonimo_aleman_nombre_clase: Darreichungsform
				sinonimo_frances_nombre_clase: Forme pharmaceutique
				sinonimo_italiano_nombre_clase: forma farmaceutica

ID	Nombre Clase	Descripción	Ejemplo	Propiedades
6	Via_ Administración	Método por la cual el fármaco se incorpora al organismo.	<p>(1) Oral IM: intramuscular IV: intravenosa...</p> <p>(2) mezclar antes, no agitar, esperar a desaparecer espuma.</p> <p>(3) Las cápsulas deben ingerirse enteras. IM profunda</p> <p>(4) Ingerir sin masticar y con líquido. No tome dosis dobles para compensar la posible dosis olvidada. No frotar zona posterior; abundante agua Ritmo no superior a 2ml/segundo;</p> <p>(5) no diluir con suero salino.</p> <p>(6) administrar con estómago vacío; fotosensible</p>	Nombre (1)
				preparacion (2)
				modo_administrar (3)
				precaucion_administra (4)
				interaccion_fisico_quimica_preparacion (5)
				interaccion_fisico_quimica_administracion (6)
				sinonimo_castellano_nombre_clase: método de administración = forma de administrar = procedimiento de administración = vías de uso,
				sinonimo_ingles_nombre_clase: method of administration
				sinonimo_aleman_nombre_clase: Art der anwendung
				sinonimo_frances_nombre_clase: mode d'administration
				sinonimo_italiano_nombre_clase: modo di somministrazione

ID	Nombre Clase	Descripción	Ejemplo	Propiedades
7	Accion	<p>La alteración inducida por los principios activos en las funciones fisiológicas.</p> <p>Las alteraciones se explican mediante el concepto de farmacodinámica El mecanismo de acción sobre el organismo que conlleva el efecto terapéutico o farmacológico y sus inconvenientes.</p> <p>En concreto, se refiere al mecanismo de acción intrínseco o cómo y dónde se realiza la acción (es una parte de farmacodinámica) y el tiempo para logra su efecto máximo y duración (es una parte de la farmacocinética)</p> <p>NOTA: Puede tener más de un mecanismo de acción.</p>	<p>(1) Lugar de la acción: enzima, pared celular...</p> <p>(2) Cómo: Impide la unión..., reduce..., retrasa..., potencia...</p> <p>(3) De treinta a 60 minutos después de la administración oral de Nolotil cápsulas se puede esperar una acción clara.</p> <p>(4) Entre 6 y 8 horas</p>	lugar_accion (1)
				tipo_accion (2)
				numero_efecto_deseable
				tiempo_efecto (3)
				duracion_efecto (4)
				numero_efecto_alterado
				numero_efecto_Indeseable
				sinonimo_castellano_nombre_clase: Mecanismo Acción = Acción del Medicamento.
				sinonimo_ingles_nombre_clase: Mechanism of action
				sinonimo_aleman_nombre_clase: Wirkungsmechanismus
				sinonimo_frances_nombre_clase: Mécanisme d'action
				sinonimo_italiano_nombre_clase: Meccanismo d'azione

ID	Nombre Clase	Descripción	Ejemplo	Propiedades
8	Efecto_deseable	Acción adecuada lograda por cómo actúa	Analgésico y Antipirético (1)	nombre_efecto_deseable (1)
				numero_indicacion_principal
				numero_indicacion_especifica
				sinonimo_castellano_nombre_clase: Acción Terapéutica = Propiedad Farmacológica = Función del Fármaco = Espectro de acción.
				sinonimo_ingles_nombre_clase: Activity
				sinonimo_aleman_nombre_clase: Wirkungsspektrum
				sinonimo_frances_nombre_clase: Spectre d'activité
				sinonimo_italiano_nombre_clase: Spettro di attività
9	Indicacion_Terapeutica	Referida a las indicaciones terapéuticas específicas autorizada y que son las únicas aprobadas para utilizar el fármaco.	(1) Los valores serían: tratamiento, preventivo primario o secundario y diagnóstico. (2) Dolor y Fiebre. (3) dolor agudo post-operatorio o post-traumático, dolor cólico, dolor oncológico	tipo_indicacion (1)
				nombre_indicacion_principal (2)
				nombre_indicaciones_especificas (3)
				sinonimo_castellano_nombre_clase: Indicaciones = Indicación Terapéutica Aprobada = indicación principal
				sinonimo_ingles_nombre_clase: Therapeutic indication
				sinonimo_aleman_nombre_clase: Anwendungsgebiete
				sinonimo_frances_nombre_clase: Indications thérapeutiques
				sinonimo_italiano_nombre_clase: Indicazioni terapeutiche

ID	Nombre Clase	Descripción	Ejemplo	Propiedades
10	Grupo_therapeutico	<p>Se refiere a la identificación por agrupación de los principios activos relacionándolos con la actuación de éstos en el organismo y de acuerdo con la clasificación anatomoterapéutica</p> <p>La clasificación fue desarrollada por la OMS y está adopta en Europa.</p>	<p>El código es alfanumerico y se interpreta a partir de los 5 niveles representados:</p> <p>Primer Nivel (anatómico): Órgano o sistema en el cual actúa el fármaco. Existen 14 grupos en total: de la A-D, G, H, J, L-N, P-S, V</p> <p>Segundo Nivel: Subgrupo terapéutico.</p> <p>Tercer Nivel: Subgrupo terapéutico o farmacológico</p> <p>Cuarto Nivel: Subgrupo terapéutico, farmacológico o químico</p> <p>Quinto Nivel: Mombre del principio activo o de la asociación farmacológica.</p> <p>El código y nombre a incorporar sería el del Tercer Nivel:</p> <p>Nolotil = N02BB02</p> <p>1º->N: Fármacos activos sobre el Sistema Nervioso.</p> <p>2º->N02: Analgésicos</p> <p>3º->N02B: otros analgésicos y antipiréticos.</p> <p>4º->N02BB: Pirazolonas</p> <p>5º->N02BB02: Metamizol Sódico</p>	codigo_OMS
				uso_therapeuticoOMS
				sinonimo_castellano_nombre_clase: Clasificación Anatomoterapéutica = Clasificación Farmacoterapéutica = Código ATC = Sistema de Clasificación Anatómica, Terapéutica, Química = Grupo farmacoterapéutico
				sinonimo_ingles_nombre_clase: Pharmacotherapeutic group = ATC = Anatomical, Therapeutic, Chemical classification system
				sinonimo_aleman_nombre_clase: Pharmakotherapeutische gruppe
				sinonimo_frances_nombre_clase: Classe pharmacothérapeutique
				sinonimo_italiano_nombre_clase: categoría farmacoterapeutica
11	Efecto_alterado	<p>Variaciones en el efecto como consecuencia de la administración conjunta de otros principios activos, alimento, agentes químicos (tabaco, alcohol) y el estado fisiopatológico concomitante. Dicha alteración se produce en el interior del organismo</p>	<p>(1)el resultado de la combinación da efectos MAYORES de los previstos aisladamente:</p> <p>Potencia el efecto anticoagulante del Acenocumarol y Warfarina</p> <p>Possible potenciación de la toxicidad del Metrexato</p> <p>Potencia su efecto con Alcohol y el de éste.</p> <p>(2)el resultado de la combinación da efectos MENORES de los previstos aisladamente:</p> <p>Reduce los niveles de la ciclosporina</p>	reaccion_sinergica (1)
				reaccion_antagonica (2)
				sinonimo_castellano_nombre_clase: Interacción = Interacción Medicamentosa
				sinonimo_ingles_nombre_clase: Interaction
				sinonimo_aleman_nombre_clase: Wechselwirkungen
				sinonimo_frances_nombre_clase: Interactions
				sinonimo_italiano_nombre_clase: Interazioni

ID	Nombre Clase	Descripción	Ejemplo	Propiedades
12	Efecto_ Indeseable	<p>Efecto no buscado. Son manifestaciones clínicas que aparecen cuando se administra un fármaco a dosis habitual.</p> <p>Los efectos indeseables se clasifican en muy raros, raros, poco frecuentes, frecuentes y muy frecuentes. A su vez, cada uno de ellos puede ser leve, moderado, grave o mortal. Se determinan por una probabilidad asociada.</p>	<p>(1) Probabilidad del ($<0,0001\%$). Reacciones alérgicas graves y poner en riesgo la vida; aunque se haya usado antes sin problemas y durante la primera hora. Reacciones cutáneas en las que aparece vesículas o ampolla.</p> <p>(2) Probabilidad del ($\geq 0,0001 - <0,001\%$). Reacciones alergias, aunque se haya usado antes sin problemas. Durante la primera hora. Las más leves se manifiestan en Piel y Mucosa: picor, quemazón, enrojecimiento, urticaria, hinchazón; .</p> <p>(3) Probabilidad del ($\geq 0,001 - <0,01\%$). Erupciones. Disminución presión arterial</p> <p>(4) Probabilidad del ($\geq 0,01 - <0,1\%$)</p> <p>(5) Probabilidad $\geq 0,1\%$</p>	e_indeseable_mraro(1)
				e_indeseable_raro(2)
				e_indeseable_poco_frecuente (3)
				e_indeseable_frecuente (4)
				e_indeseable_muy_frecuente(5)
				sinonimo_castellano_nombre_clase: Efecto Adverso = Reacción Adversa = Efecto Secundario = Efecto Colateral
				sinonimo_ingles_nombre_clase: Undesirable effects
				sinonimo_aleman_nombre_clase: Nebenwirkungen
13	Contraindicación	<p>Situación específica o factor por el cual NO se debe utilizar un principio activo ya que de antemano se sabe que puede ser dañino.</p>	<p>(1) Antecedentes de reacción alérgica al metamizol u otros derivados pirazolónicos o a cualquier otro componente (excipientes).</p> <p>(2) Agranulocitosis tras ingesta anterior de nolitil o Metamizol megnésico. Deficiencia congénita de glucosa-6-fosfato-deshidrogenasa.</p> <p>(3) Niños menores de 15 años. En primer y tercer trimestre de gestación; en el segundo valoración médica beneficio/riesgo. Se excreta en leche materna, evitar lactancia durante 48 horas después de ingesta.</p>	alergia(1)
				Problema_fisiopatologico (2)
				cambio_fisiologico(3)
				Sinonimo_castellano_nombre_clase: Incompatibilidad.
				Sinonimo_ingles_nombre_clase:: Contraindication
				Sinonimo_aleman_nombre_clase: Kontraindikation = Gegenanzeigen
				Sinonimo_frances_nombre_clase: Contre-indication
				Sinonimo_italiano_nombre_clase: Controindicazioni

ID	Nombre Clase	Descripción	Ejemplo	Propiedades
14	Posología	Pauta que indica la dosis estándar, referida a la cantidad diaria, su intervalo de tiempo entre dosis (frecuencia) y el período de tiempo total (duración). La dosis se tipifica para población adulta y las modificaciones para población infantil, a veces, para ancianos y para determinadas patologías.	(1) cápsula tres o cuatro veces por día durante 10 días. De 500-575mg/6-12h (2) No pauta hasta 15 años. (3) Debe evitarse la administración de dosis elevadas. (4) Función renal y hepática disminuida y pacientes debilitados debe evitarse la administración de dosis elevadas.	habitual(1)
				habitual_niño(2)
				habitual_ancianos(3)
				habitual_patología(4)
				sinonimo_castellano_nombre_clase: Dosificación = Prescripción
				sinonimo_ingles_nombre_clase: Posology = Dosage
				sinonimo_aleman_nombre_clase: Dosierung
				sinonimo_frances_nombre_clase: Posologie = Dosage
15	Sobredosis	Dosis excesiva del fármaco o incremento de la dosis recomendada que deriva en una intoxicación	(1) Náuseas, vómitos, dolor abdominal, deterioro función renal. Vértigo, somnolencia, coma, convulsiones y descenso tensión arterial (hipotensión). En dosis muy alta coloración roja de la orina. (2) No antídoto específico. Valorar si lavado gástrico y forzar el vómito. Considerar diuresis forzada o diálisis. Si hipersensibilidad grave, además aplicar otras medidas como posición lateral, vías respiratorias libres de obstrucción o administración oxígeno. Las medidas urgentes incluyen adrenalina, fluidoterapia y glucocorticoides. Monitorización funciones vitales.	manifestaciones
				actuación
				sinonimo_castellano_nombre_clase: Intoxicación medicamentosa
				sinonimo_ingles_nombre_clase: Overdose
				sinonimo_aleman_nombre_clase: Überdosierung
				sinonimo_frances_nombre_clase: Surdosage
				sinonimo_italiano_nombre_clase: Sovradosaggio

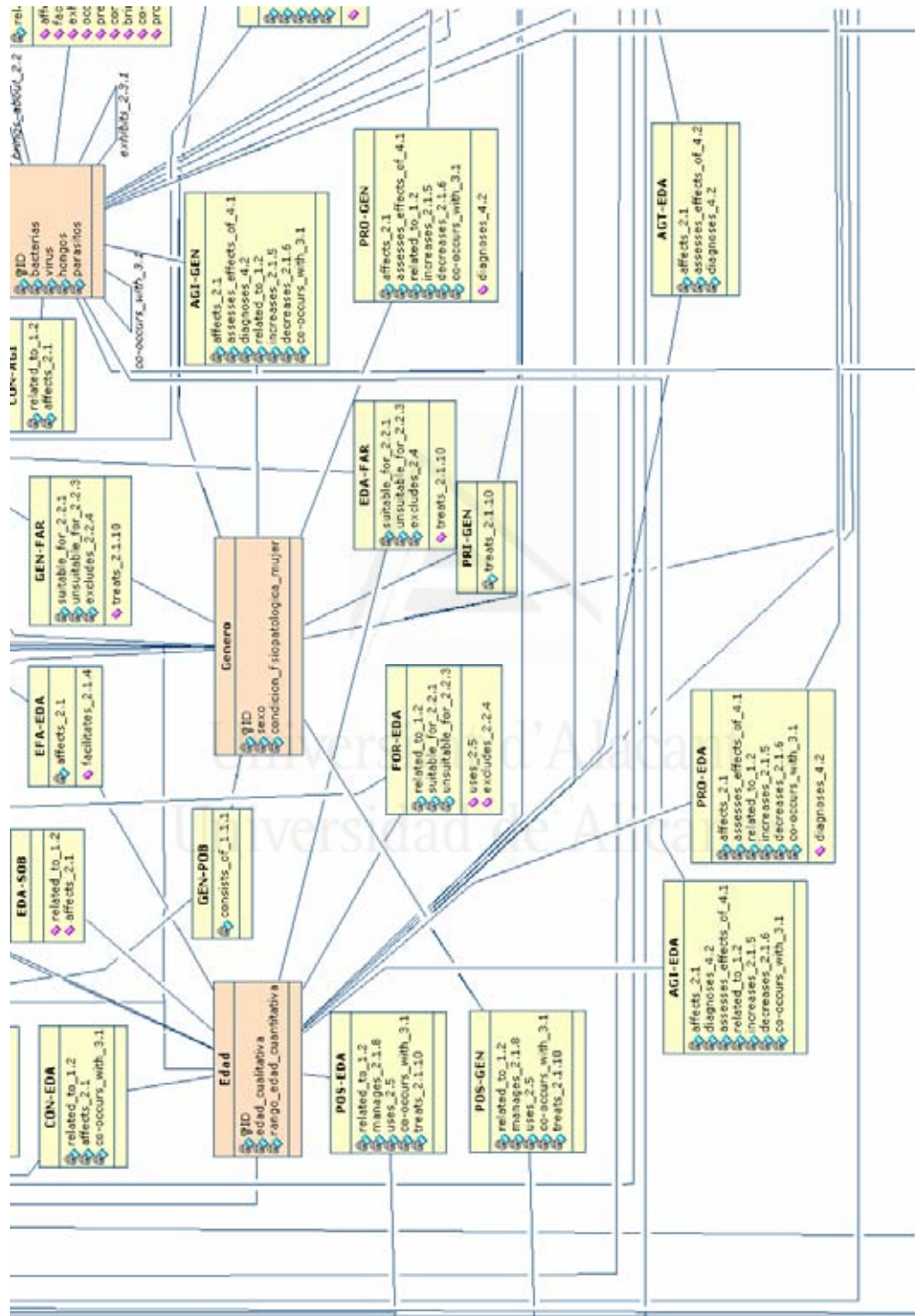
ID	Nombre Clase	Descripción	Ejemplo	Propiedades
16	Estabilidad	Condiciones necesarias para la conservación y permanencia de la funcionalidad del fármaco	(1) mantener a temperatura no superior de 15°C (2) proteger de luz directa.	termolabilidad (1)
				fotosensibilidad (2)
				sinonimo_castellano_nombre_clase: conservación
				sinonimo_ingles_nombre_clase: Special precautions for storage
				sinonimo_aleman_nombre_clase: Besondere vorsichtsmaßnahmen für die aufbewahrung
				sinonimo_frances_nombre_clase: Conservation = Précautions particulières de conservation
				sinonimo_italiano_nombre_clase: Conservazione = Precauzioni particolari per la conservazione
17	Poblacion	Referido al grupo de población para el cual es efectivo el fármaco y cuando presenta limitaciones por otras condiciones asociadas al sujeto.	(1) En ocasiones la condición para indicar la población diana se realiza a partir de tener o superar un peso; o bien la pauta cambia en función del peso. (2) Las condiciones exógenas que delimitan a la población diana: No conducir, no manipular maquinaria... (3) Problemas hepáticos y renales condicionan la prescripción o el ajuste de la dosis.	peso(1)
				condición ambiental(2)
				condicion_fisiopatologica (3)
				sinonimo_castellano_nombre_clase: Población diana
				sinonimo_ingles_nombre_clase: target population
				sinonimo_aleman_nombre_clase: Bevölkerung
				sinonimo_frances_nombre_clase: Population
				sinonimo_italiano_nombre_clase: Popolazione

ID	Nombre Clase	Descripción	Ejemplo	Propiedades
18	Genero	Condición orgánica, masculina o femenina, que diferencia a la población diana.	(1)Cada vez hay más medicación específica para un grupo diferenciado por sexo (Anticonceptivos, Viagra) o bien hay un cambio en la dosis. (2)Embarazo (en general o diferenciando según trimestre), lactancia, conlleva la prohibición o un ajuste de dosis.	sexo(1)
				condicion_fisiologica_mujer(2)
				sinonimo_castellano_nombre_clase: Sexo
				sinonimo_ingles_nombre_clase: Sex = Gender
				sinonimo_aleman_nombre_clase: Geschlecht
				sinonimo_frances_nombre_clase: Sexe
19	Edad	Cada uno de los períodos en que se considera dividida la vida humana. A veces se hace por negación, indicando la población que no es diana (los menores de 15 años no deben tomar este medicamento).	(1)Adulto, niño, lactante... (2)Niños mayores de 12 años; de 1 a 3 años.	edad_cualitativa(1)
				rango_edad_cuantitativa(2)
				sinonimo_castellano_nombre_clase
				sinonimo_ingles_nombre_clase: Age
				sinonimo_aleman_nombre_clase: Alter
				sinonimo_frances_nombre_clase: Âge
20	Proceso_clinico	Denominación genérica para referirse a todo proceso morboso. No se diferencia inicialmente entre signo, síntoma, síndrome, o cuadro patológico.		nombre_pclinica
				sinonimo_castellano_nombre_clase: Enfermedad = patología = manifestaciones clínicas
				sinonimo_ingles_nombre_clase: disease = illness
				sinonimo_aleman_nombre_clase: Krankheit = übel
				sinonimo_frances_nombre_clase: maladie
				sinonimo_italiano_nombre_clase: malattia

ID	Nombre Clase	Descripción	Ejemplo	Propiedades
21	Agente_infeccioso	Microorganismo causante de un proceso fisiopatológico	<p>(1)nominativamente tipo estafilococo o genéricamente: gram-, gram+, aerobias o anaerobias.</p> <p>(2)virus herpes simple o con siglas VHS: VHS1 y VHS2 ..</p> <p>(3)Alternaria, Aspergillus, Candida albicans, Cladosporium, Penicillium...</p> <p>(4)Externos: pulga...; Internos: tenia, triquina, Anisakis...; Protozoos: Toxoplasma gondii, Plasmodium.</p>	bacterias(1)
				virus(2)
				hongos(3)
				parasitos(4)
				sinonimo_castellano_nombre_clase: microorganismos
				sinonimo_ingles_nombre_clase: Microorganisms
				sinonimo_aleman_nombre_clase: Erreger
22	Agente_toxico	<p>Sustancia exógena que puede desencadenar en el organismo una reacción negativa, envenenamiento o deterioro orgánico.</p> <p>Pueden clasificarse por su origen (naturales y sintético), por sus efectos (corrosivos, irritantes, neurotóxicos...) y por su estado (sólidos, gases y líquidos). También pueden clasificarse por su naturaleza química, por su uso, por el órgano que afectan, por su mecanismo de acción, etc. Cada tóxico puede pertenecer simultáneamente a dos o más clasificaciones.</p>	(1) mercurio, plomo, alergénicos...	nombre_toxico (1)
				sinonimo_castellano_nombre_clase: Sustancia tóxica = tóxico
				sinonimo_ingles_nombre_clase: Toxic
				sinonimo_aleman_nombre_clase: Toxisch
				sinonimo_frances_nombre_clase: Toxique
23	Unidad_medida		<p>(1) mg, cc, min y otras variantes no normalizadas son: antes, después, durante, alternativamente</p> <p>(2) Masa densidad, volumen y tiempo.</p>	cantidad_cuantitativa
				unidades(1)
				magnitud(2)

The image displays a complex conceptual map with numerous nodes and connecting lines. The nodes are organized into several main categories, each represented by a colored box with a title and a list of related terms and their relationships.

- Sobredosis (Overdose):**
 - gID
 - manifestaciones
 - adición
 - dosis
- Contraindicación (Contraindication):**
 - gID
 - psico
 - problema fisiopatológico
 - combo fisiológico
- Población (Population):**
 - gID
 - psico
 - condición ambiental
 - condición fisiopatológica
- CON-PRO (Contraindication - Pro):**
 - related_to_1.2
 - affects_2.1
 - co-occurs_with_3.1
- PRI-AGI (Priority - Agitation):**
 - affects_2.1
 - delays_2.1.1
 - complicates_2.1.2
 - prevents_2.1.3
 - treats_2.1.9
 - brings_about_2.2
 - related_to_2.2
 - related_to_1.2
 - requires_4.4
- CON-AGI (Contraindication - Agitation):**
 - related_to_1.2
 - requires_4.4
- PRI-POB (Priority - Population):**
 - treats_2.1.10
- CON-POB (Contraindication - Population):**
 - related_to_1.2
 - affects_2.1
- EFA-POB (Efficacy - Population):**
 - affects_2.1
 - facilitates_2.1.4
- EFA-GEN (Efficacy - Genotype):**
 - affects_2.1
 - facilitates_2.1.4
- CON-GEN (Contraindication - Genotype):**
 - related_to_1.2
 - affects_2.1
 - co-occurs_with_3.1
- EFA-POB (Efficacy - Population):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- CON-PRO (Contraindication - Pro):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- PRI-POB (Priority - Population):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- CON-POB (Contraindication - Population):**
 - related_to_1.2
 - affects_2.1
 - co-occurs_with_3.1
- EFA-POB (Efficacy - Population):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- CON-GEN (Contraindication - Genotype):**
 - related_to_1.2
 - affects_2.1
 - co-occurs_with_3.1
- EFA-POB (Efficacy - Population):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- CON-PRO (Contraindication - Pro):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- PRI-POB (Priority - Population):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- CON-POB (Contraindication - Population):**
 - related_to_1.2
 - affects_2.1
 - co-occurs_with_3.1
- EFA-POB (Efficacy - Population):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- CON-GEN (Contraindication - Genotype):**
 - related_to_1.2
 - affects_2.1
 - co-occurs_with_3.1
- EFA-POB (Efficacy - Population):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- CON-PRO (Contraindication - Pro):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- PRI-POB (Priority - Population):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- CON-POB (Contraindication - Population):**
 - related_to_1.2
 - affects_2.1
 - co-occurs_with_3.1
- EFA-POB (Efficacy - Population):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- CON-GEN (Contraindication - Genotype):**
 - related_to_1.2
 - affects_2.1
 - co-occurs_with_3.1
- EFA-POB (Efficacy - Population):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- CON-PRO (Contraindication - Pro):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- PRI-POB (Priority - Population):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- CON-POB (Contraindication - Population):**
 - related_to_1.2
 - affects_2.1
 - co-occurs_with_3.1
- EFA-POB (Efficacy - Population):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- CON-GEN (Contraindication - Genotype):**
 - related_to_1.2
 - affects_2.1
 - co-occurs_with_3.1
- EFA-POB (Efficacy - Population):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- CON-PRO (Contraindication - Pro):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2
 - increases_2.1.5
 - decreases_2.1.6
 - co-occurs_with_3.1
 - facilitates_2.1.4
- PRI-POB (Priority - Population):**
 - affects_2.1
 - diagnoses_4.2
 - related_to_1.2
 - complicates_2.1.2



ANEXO VI: Grupos terapéuticos utilizados para la selección del conjunto de instancias de medicamentos de la base de conocimiento de OntoFIS.

B01 ANTITROMBOTICOS
B01A ANTITROMBOTICOS
B01AA ANTAGONISTAS DE LA VITAMINA K
B01AB HEPARINAS
B01AC INHIBIDORES DE LA AGREGACION PLAQUETARIA
B01AD ENZIMAS (ANTITROMBOTICOS)
B01AE INHIBIDORES DIRECTOS DE LA TROMBINA
B01AX OTROS ANTITROMBOTICOS
B02A ANTIFIBRINOLITICOS
B06 OTROS PREPARADOS HEMATOLOGICOS
C APARATO CARDIOVASCULAR
C01 TERAPIA DEL MIOCARDIO
C01A GLICOSIDOS CARDIACOS
C01AA GLICOSIDOS DE DIGITALIS (CARDIOTERAPIA)
C01AB GLICOSIDOS DE SCILLA (CARDIOTERAPIA)
C01AC GLICOSIDOS DE ESTROFANTO (CARDIOTERAPIA)
C01AX OTROS GLICOSIDOS CARDIACOS
C01B ANTIARRITMICOS
C01BA ANTIARRITMICOS CLASE IA
C01BB ANTIARRITMICOS CLASE IB
C01BC ANTIARRITMICOS CLASE IC
C01BD ANTIARRITMICOS CLASE III
C01C ESTIMULANTES CARDIACOS, EXCLUIDOS GLICOSIDOS CARD.
C01CA ADRENERGICOS Y DOPAMINERGICOS (ESTIMULANTES CARDIACOS)
C01CE INHIBIDORES DE LA FOSFODIESTERASA (ESTIMULANTES CARDIACOS)
C01CX OTROS ESTIMULANTES CARDIACOS
C01D VASODILATADORES USADOS EN CARDIOTERAPIA
C01DA NITRATOS ORGANICOS (CARDIOTERAPIA)
C01DB VASODILATADORES DRVDS. QUINOLONA
C01DX OTROS VASODILATADORES USADOS EN CARDIOTERAPIA
C01E OTROS PREPARADOS CARDIACOS
C01EA PROSTAGLANDINAS (CARDIOTERAPIA)
C01EB OTROS PREPARADOS CARDIACOS
C01EX ASOCIACIONES DE OTROS PREPARADOS CARDIACOS
C02 ANTIHIPERTENSIVOS
C02A ANTIADRENERGICOS DE ACCION CENTRAL (ANTIHIPERTENS)
C02AA ALCALOIDES DE LA RAUWOLFIA (ANTIHIPERTENSIVOS)
C02AB METILDOPA (ANTIHIPERTENSIVOS)
C02AC AGONISTAS RECEPTORES IMIDAZOLINA (ANTIHIPERTENSIVOS)

C02B ANTIADRENERGICOS BLOQ. GANGLIONARES (ANTIHIPERTEN)
C02BA DERIVADOS DEL SULFONIO (ANTIHIPERTENSIVOS)
C02C ANTIADRENERGICOS DE ACCION PERIFERICA (ANTIHIPERT)
C02CA BLOQUEANTES ALFA-ADRENERGICOS (ANTIHIPERTENSIVOS)
C02CB BLOQUEANTES ALFA-BETA ADRENERGICOS (ANTIHIPERTEN.)
C02CC DERIVADOS DE LA GUANIDINA (ANTIHIPERTENSIVOS)
C02D ANTIHIPERTENSIVOS ACCION MUSCULO LISO ARTERIOLAS
C02DA DERIVADOS DE LA TIAZIDA (ANTIHIPERTENSIVOS)
C02DB DERIVADOS DE LA HIDRAZINOFTALAZINA (ANTIHIPERTENSIVOS)
C02DC DERIVADOS DE LA PIRIMIDINA (ANTIHIPERTENSIVOS)
C02DD DERIVADOS DE LA NITROFERRICIANIDA (ANTIHIPERTENSIVOS)
C02DE BLOQUEADORES DE LOS CANALES DEL CALCIO (ANTIHIPERT)
C02DG DERIVADOS DE LA GUANIDINA (ANTIHIPERTENSIVOS)
C02K OTROS ANTIHIPERTENSIVOS
C02KX OTROS ANTIHIPERTENSIVOS
C02L ANTIHIPERTENSIVOS ASOCIADOS A DIURETICOS
C02LA ALCALOIDES DE RAUWOLFIA ASOCIADOS A DIURETICOS
C02LB METILDOPA ASOCIADA A DIURETICOS
C02LC AGONISTAS RECEPTOR IMIDAZOLICO EN ASOCIACION CON DIURETICO
C02N ASOCIACIONES DE ANTIHIPERTENSIVOS
C03 DIURETICOS
C03A DIURETICOS DE BAJO TECHO, TIAZIDAS
C03AA TIAZIDAS SOLAS (DIURETICOS)
C03AB TIAZIDAS ASOCIADAS A POTASIO (DIURETICOS)
C03AH TIAZIDAS ASOCIADAS A PSICOLEPTICOS (DIURETICOS)
C03AX TIAZIDAS ASOCIADOS A OTRAS SUSTANCIAS (DIURETICOS)
C03B DIURETICOS DE BAJO TECHO, EXCLUIDAS TIAZIDAS
C03BA SULFONAMIDAS SOLAS (DIURETICOS BAJO TECHO)
C03BB SULFONAMIDAS ASOCIADAS A POTASIO (DIURETICOS)
C03BC DIURETICOS MERCURIALES
C03BD XANTINAS (DIURETICOS)
C03C DIURETICOS DE ALTO TECHO
C03CA SULFONAMIDAS SOLAS (DIURETICOS ALTO TECHO)
C03CX OTROS DIURETICOS DE ALTO TECHO
C03E DIURETICOS AHORRADORES DE POTASIO CON OTROS DIURET
C03EA ASOCIACIONES DIURETICOS BAJO TECHO+AHORRADORES DE POTASIO
C03EB ASOCIACIONES DIURETICOS ALTO TECHO+AHORRADORES DE POTASIO
C04 VASODILADORES PERIFERICOS
C04A VASODILADORES PERIFERICOS
C04AA DERIVADOS DEL AMINOFENIL-ETANOL (VASODILADORES PERIFERICOS)
C04AB DERIVADOS IMIDAZOLICOS (VASODILADORES PERIFERICOS)
C04AC DERIVADOS DEL ACIDO NICOTINICO (VASODIL.PERIFERIC)
C04AD DERIVADOS DE LAS PURINAS (VASODILADORES PERIFERICOS)
C04AE ALCALOIDES ERGOTAMINICOS (VASODILADORES PERIFERICOS)
C04AX OTROS VASODILADORES PERIFERICOS