



UNIVERSIDAD
DE LA REPUBLICA
URUGUAY



UNIVERSIDAD DE LA REPÚBLICA
FACULTAD DE INGENIERÍA

Tesis para optar al Título de
Magíster en Ingeniería en Computación

**ESTUDIO DE METODOLOGIAS DE DISEÑO Y
DESARROLLO DE ONTOLOGIAS
APLICACION A UN CASO DE ESTUDIO DE
EVALUACION DE SITIOS WEB EN EL AREA
DE LA SALUD**

Autor: Edelweis Rohrer

Director de Tesis: Regina Motz

Montevideo, Uruguay

2009

ISSN 1510-7264

Edelweis Rohrer Errecarte

Tesis de Maestría en Ingeniería en Computación

Facultad de Ingeniería

Universidad de la República

Montevideo, Uruguay, 2009

Agradecimientos

A Regina, que además de su conocimiento y apoyo permanente, me transmitió mucha fuerza y energía positiva.

A Adriana y Lorena, por hacerme sentir tan bien en su lugar de trabajo y darme una mano cuando la necesité.

A los integrantes del proyecto CYTED-SALUS, que me dieron la posibilidad de realizar un aporte a su trabajo.

A mi familia.

Resumen

A principios de los años 90' comienza un proceso de análisis de buenas prácticas de construcción de ontologías, que tiene como resultado la publicación de metodologías de diseño con diferentes enfoques. El presente trabajo comienza con el estudio de seis metodologías de diseño de ontologías representativas de la evolución de este proceso, con un análisis comparativo que muestra que ninguna de las metodologías cumple, al mismo tiempo, con los siguientes tres requerimientos: (i) Especificar en detalle la actividad de adquisición de conocimiento (ii) Ser suficientemente flexible para dar soporte a la construcción inicial de una ontología y a su evolución posterior, tanto si se trata de una ontología independiente de una aplicación específica o creada especialmente para ser utilizada por una aplicación particular (iii) Proveer mecanismos de trabajo colaborativo y distribuido. En base a este análisis, se realiza una propuesta de metodología de diseño y desarrollo de ontologías, que se denomina Conceptual Ontology Design, que rescata las mejores prácticas identificadas en las metodologías estudiadas y define estrategias para contemplar estos tres requerimientos. A medida que fue elaborada, esta propuesta fue validada con un caso de estudio de evaluación de sitios web en el área de la salud. Esta realidad plantea el desafío de intentar emitir recomendaciones a los usuarios que consultan documentos de sitios web del dominio de la salud, de acuerdo a características de los contenidos de los documentos y a las características de los usuarios que

los consultan. Para ello es necesario clasificar estos documentos de acuerdo a diferentes factores de calidad como lo son confianza en la fuente y legibilidad, y tener en cuenta aspectos de los usuarios como nivel académico y rol que cumplen (usuario común, paciente, familiar del paciente, etc.). Cuando un usuario consulta un documento, en base a la evaluación del par <documento, usuario> debe emitirse una recomendación al usuario, que le sugiera en menor o mayor grado la lectura del documento. En este trabajo se construyó una ontología que a través de axiomas y reglas de inferencia hace posible emitir tal recomendación, a la vez que permitió llevar a cabo una validación primaria de la propuesta de metodología de diseño y desarrollo de ontologías.

Palabras clave

Ontología, diseño conceptual, adquisición de conocimiento.

Tabla de Figuras

Figura 1. Actividades propuestas por la metodología Methontology [GFC03].	13
Figura 2. Conceptualización de la evaluación de documentos y perfiles de usuario.	29
Figura 3. Pasos a seguir en la metodología SENSUS-based method [SPKR96].	38
Figura 4. Actividades propuestas en la metodología On-To-Knowledge [SSSS00].	45
Figura 5. Participación de los diferentes actores propuesta por la metodología DILIGENT [VPTS05].	52
Figura 6. Conceptualización del dominio específico Alzheimer.	58
Figura 7. Dinámica de definición de ontologías en la metodología HCOME [KV06].	61
Figura 8. Ciclo de vida y actividades propuestas por la metodología UPON.	66
Figura 9. Diagrama construido en base a un fragmento de Semantic Network, una de las partes que componen la ontología médica UMLS [WR1] [GFC03].	73
Figura 10. Modelo expresado a través de diagramas UML.	74
Figura 11. Fragmento de la estructura ontológica representado en UML y RDF(S).	77

Figura 12. Ciclo de vida y actividades propuestas por Conceptual Ontology Design.....	96
Figura 13. Lista de términos agrupados por área de trabajo.	105
Figura 14. Agrupación de conceptos principales en diferentes áreas. ...	108
Figura 15. Taxonomías de conceptos para el caso de estudio CYTED- SALUS.	113
Figura 16. Modelo con relaciones binarias del caso de estudio CYTED- SALUS.	116
Figura 17. Modelo del caso de estudio CYTED-SALUS con atributos de instancia.	119
Figura 18. Integración del modelo Calidad con el modelo Sitio.....	131
Figura 19. Ontología General de Calidad.....	148
Figura 20. Ontología General de Calidad actualizada con adaptaciones locales.	149
Figura 21. Primera conceptualización del área de trabajo Calidad.	152
Figura 22. Primera conceptualización del área de trabajo Salud.	152
Figura 23. Red de ontologías preliminar al modelo final.	154
Figura 24. Red de ontologías.	155
Figura 25. Ontología Dominio de Salud.	158
Figura 26. Ontología Dominio de Salud para una enfermedad específica.	160
Figura 27. Ontología Sitio Web.	163
Figura 28. Ontología Perfil Usuario.	164

Figura 29. Ontología Calidad.	169
Figura 30. Ontología Recomendación.	171
Figura 31. Mantenimiento de Métricas y Niveles de Calidad y Recomendación	173
Figura 32. Componentes de la aplicación de Mantenimiento de Métricas y Niveles de Calidad y Recomendación	174
Figura 33. Anotación y clasificación de contenidos Web de acuerdo a métricas de calidad	175
Figura 34. Componentes de la aplicación Anotación y Clasificación de Contenidos de Sitios Web.....	176
Figura 35. Recomendación de contenidos en un portal médico.	177
Figura 36. Componentes de la aplicación Recomendación de contenidos en un portal médico.	178

Tabla de contenido

1. Introducción.....	7
2. Análisis de metodologías de diseño de ontologías	11
2.1. Metodología Methontology	12
2.1.1. Actividades propuestas por la metodología	15
2.1.2. Conclusiones	34
2.2. Metodología SENSUS-based method	37
2.2.1. Actividades propuestas por la metodología	38
2.2.2. Conclusiones	40
2.3. Metodología On-To-Knowledge	44
2.3.1. Actividades propuestas por la metodología	45
2.3.2. Conclusiones	48
2.4. Metodología DILIGENT	51
2.4.1. Actividades propuestas por la metodología	53
2.4.2. Aplicación al caso de estudio	55
2.4.3. Conclusiones	56
2.5. Metodología HCOME	59
2.5.1. Actividades propuestas por la metodología	61
2.5.2. Conclusiones	62
2.6. Metodología UPON	65
2.6.1. Actividades propuestas por la metodología	68
2.6.2. Conclusiones	80
2.7. Conclusiones finales	83

2.7.1.	Criterio de evaluación de metodologías	83
2.7.2.	Resultados de la aplicación del criterio de evaluación	86
3.	Recomendación de metodología de diseño y desarrollo de ontologías: Conceptual Ontology Design	94
3.1.	Construcción de Ontología Inicial	97
3.1.1.	Especificación y técnicas de adquisición de conocimiento aplicadas.....	101
3.1.2.	Conceptualización apoyada en técnicas de adquisición de conocimiento.....	110
3.1.3.	Integración	121
3.1.4.	Evaluación.....	132
3.1.5.	Implementación.....	141
3.2.	Mantenimiento y evolución	142
3.2.1.	Adaptación de ontologías locales.....	143
3.2.2.	Actualización de ontología general.....	146
4.	Diseño de una ontología para la recomendación de sitios web en el área de la salud	150
4.1.	Evolución del diseño.....	151
4.2.	Descripción del modelo.....	155
4.2.1.	Ontología Dominio de Salud	157
4.2.2.	Ontología Sitio Web	161
4.2.3.	Ontología Perfil de Usuario	163
4.2.4.	Ontología Calidad.....	164

4.2.5. Ontología Recomendación	170
4.3. Utilización de la red de ontologías en una aplicación.....	172
4.3.1. Mantenimiento de métricas de calidad y recomendación ...	172
4.3.2. Anotación y clasificación de contenidos Web de acuerdo a métricas de calidad	174
4.3.3. Recomendación de contenidos en un portal médico	176
5. Conclusiones y trabajos futuros	179
Bibliografía	183

1. *Introducción*

Cualquier aplicación de software se basa en la construcción de un modelo de la realidad, tarea en la cual, la obtención de una conceptualización consensuada es uno de los principales desafíos del proceso de desarrollo de software. La representación de la realidad a través de modelos expresados en lenguajes formales, como son las ontologías, es una de las herramientas más adecuadas para llevar a cabo una conceptualización que asegure el consenso entre todos los usuarios involucrados. Sin embargo, existen algunos aspectos en el proceso de desarrollo de ontologías, que continúan siendo críticos para la obtención de un modelo que represente los aspectos verdaderamente relevantes de la realidad. La tarea de identificar qué debería ser modelado como un concepto y qué debería ser modelado como una relación, es uno de estos puntos críticos, que requiere de guías precisas para que los ingenieros de ontologías lleven a cabo la tarea de adquisición de conocimiento, especialmente en las actividades de especificación y conceptualización. Es necesario también que esta primera conceptualización de la realidad se encauce en un proceso de continua retroalimentación, de manera que contemple los cambios de esta realidad. Esto obliga a contar un mecanismo de evolución que permita a usuarios finales e ingenieros de ontologías trabajar en conjunto para evitar que la discontinuación de este proceso de cambio lleve a que el modelo se vuelva obsoleto. En muchos

casos la evolución natural de un modelo se da en contextos locales que por diversos factores tienen necesidades diferentes, lo que lleva a que la primera ontología diseñada desde una perspectiva general, deba ser adaptada de diferente manera de acuerdo al contexto local. Para responder a este fenómeno de evolución distribuida, debe ser provisto un mecanismo de trabajo para que cada grupo local pueda contar con una vista del modelo adaptada a sus necesidades, a la vez que el núcleo del modelo, común a todos los contextos, evoluciona también para reflejar cambios a nivel general. El alcance de la realidad a ser representada puede abarcar, desde la información relativa a un dominio de aplicación, como por ejemplo, medicina, química, arte, etc., que una vez construido el modelo puede ser explotado por diferentes aplicaciones que necesitan hacer uso de información del dominio, hasta información estrictamente restringida a una aplicación específica, para un proceso de negocio dado. Una propuesta flexible debería resultar útil, cualquiera sea la amplitud del alcance a modelar, sugiriendo diferentes mecanismos de trabajo si es necesario.

Puesto que los aspectos mencionados anteriormente se identifican como puntos relevantes en la ingeniería de ontologías, para llevar a cabo un abordaje de los mismos desde el punto de vista de la ingeniería de software tradicional, en este trabajo se comienza con un análisis y comparación de seis metodologías de diseño de ontologías. El resultado de esta comparación es el punto de partida para la elaboración de una

propuesta de metodología de diseño de ontologías, que además de incorporar las características más relevantes de cada una de las metodologías analizadas, intenta dar soporte a los aspectos mencionados, argumentados como relevantes. El enfoque de esta propuesta, denominada Conceptual Ontology Design, tiene entonces las siguientes características: (i) Especificar en detalle la actividad de adquisición de conocimiento (ii) Ser suficientemente flexible para manejar el desarrollo inicial como la evolución de la ontología, tanto si se trata de una ontología independiente de una aplicación específica o creada especialmente para ser utilizada por una aplicación particular (iii) Establecer mecanismos de trabajo colaborativo y distribuido.

La validación de la propuesta Conceptual Ontology Design fue realizada con la aplicación de la misma a un caso de estudio de diseño de una ontología que permite asistir en la recomendación de sitios web en el área de la salud a diferentes perfiles de usuario, en el marco del proyecto CYTED-SALUS. Para lograr este objetivo, el modelo debe permitir medir la calidad de los contenidos web de acuerdo a factores de calidad establecidos por expertos en el dominio. Para la elección de los factores de calidad, el enfoque de este caso de estudio es analizar la calidad que se desprende del valor de la información que el sitio provee y su adecuación a las expectativas del lector. Siguiendo este enfoque, se adquirió conocimiento a través de entrevistas con expertos para elegir los factores de calidad a ser medidos, resultando seleccionados los factores:

legibilidad, frescura, confianza en la fuente y completitud. Se construyó entonces una red de ontologías que representa los conceptos de la realidad desde cuatro principales puntos de vista: dominio de la salud, sitios web, factores de calidad y perfiles de usuario.

Lo que sigue de este documento está organizado de la siguiente manera: la Sección 2 presenta el análisis de la metodologías estudiadas, en la Sección 3 se detalla la propuesta Conceptual Ontology Design, la Sección 4 muestra el desarrollo completo del modelo de la ontología del caso de estudio y la Sección 5 presenta conclusiones y trabajos futuros.

2. Análisis de metodologías de diseño de ontologías

En los últimos años, se ha comenzado a trabajar en una cantidad significativa de proyectos de investigación acerca de metodologías para el diseño de ontologías. Como ejemplo pueden mencionarse OntoWeb [WR5], DOGMA [JM08], Neon [WR6], etc. Para realizar un estudio del estado del arte, se llevó a cabo un análisis comparativo de un conjunto de seis metodologías que se consideran representativas de la evolución de la disciplina Ingeniería de Ontologías.

Las metodologías seleccionadas para el análisis son: Methontology, SENSUS-based method, On-To-Knowledge, DILIGENT, HCOME y UPON. Methontology fue seleccionada por ser ampliamente citada en la gran mayoría de los documentos de esta área temática, pero la principal razón radica en que es una de las metodologías que tiene como actividades propuestas las principales actividades identificadas por el Estándar para Procesos de Ciclo de Vida de Desarrollo de Software de la IEEE [GFC03]. Es por este motivo que la misma es analizada en primer lugar, sirviendo como guía para el estudio de las demás metodologías elegidas. SENSUS-based method se eligió por ser una metodología de diseño de ontologías que obligatoriamente toma como base una ontología existente (en otras metodologías, la decisión de basarse en una estructura común es opcional). On-To-Knowledge, además de ser citada por otros autores, plantea una metodología orientada a la aplicación.

DILIGENT fue seleccionada por ser bastante más reciente que las anteriores y porque a diferencia de ellas pone énfasis en la evolución descentralizada de las ontologías, una vez construida la versión inicial. HCOME, al igual que DILIGENT se concentra en la construcción descentralizada de las ontologías, pero desde el inicio y haciendo más hincapié en la participación de los usuarios o expertos del dominio. UPON, uno de los enfoques más recientemente publicados, está basado en la metodología de desarrollo de software Unified Process, no tan reciente pero suficientemente conocida y adoptada por la comunidad de software, con la utilización del lenguaje de modelado UML.

A continuación se resumen las principales características de cada una de las metodologías, destacándose sus elementos más relevantes, como paso previo a la presentación de los resultados de la comparación.

2.1. Metodología Methontology

La metodología de diseño de ontologías Methontology, es el resultado del trabajo de Mariano Fernández, Asunción Gómez-Pérez y Natalia Juristo. En [FGJ97] y [GFC03], los autores describen una serie de actividades que deben llevarse a cabo durante el proceso de desarrollo de una ontología, así como cuál es el ciclo de vida que determina el orden y profundidad con que deben ser abordadas cada una de ellas. [FGJ97] hace referencia a estudios previos de los mismos autores y de otros como Uschold, Gruninger, Fox, De Vicente, Gangemi, Giacomelli y Speel. En este sentido

la metodología puede verse como una síntesis de diferentes iniciativas e investigaciones, así como prácticas aplicadas en diferentes dominios. El trabajo es una especie de selección de modalidades de trabajo cuya efectividad ya fue debidamente probada y argumentada. En la Figura 1 se muestran las actividades que propone la metodología, que como ya se mencionó forman parte del Estándar de Procesos de Desarrollo de Software de la IEEE. Como puede apreciarse dichas actividades se clasifican en: actividades de administración, desarrollo y soporte.

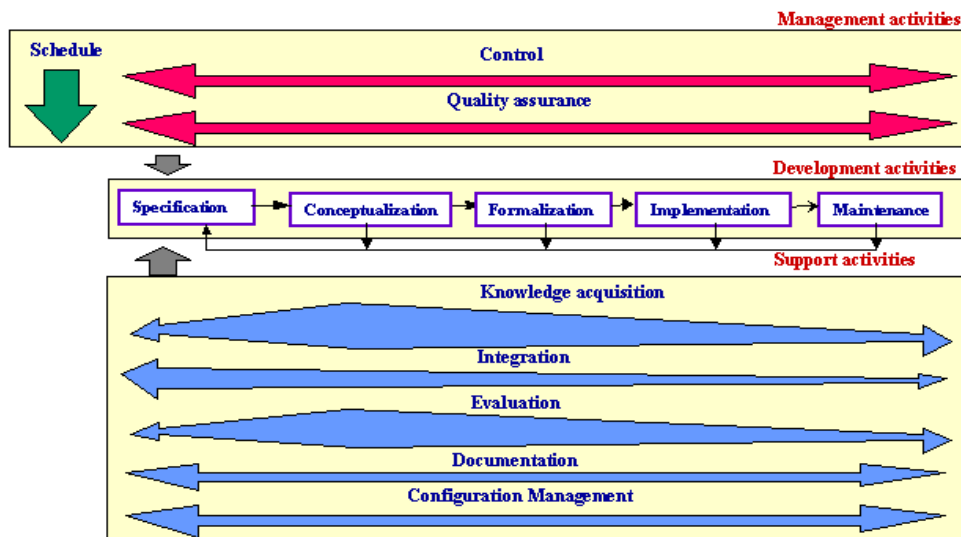


Figura 1. Actividades propuestas por la metodología Methontology [GFC03].

Las actividades de administración son el punto partida para iniciar la ejecución de cualquier proceso y tener un marco de referencia para medir su éxito, desvíos, etc. Las actividades de desarrollo constituyen la parte central del proceso, a través de ellas verdaderamente se “construye” la

ontología, a diferencia de las actividades de administración y soporte, que apuntan más que nada a la obtención de un producto de calidad, como herramientas de ayuda. Las actividades de soporte, al igual que las de administración, acompañan a las de desarrollo en todo el ciclo de vida, destacándose las actividades de adquisición de conocimiento y evaluación. La primera de ellas es importante ya que permite al ingeniero ontologista obtener una visión de la realidad a modelar, a partir de diversas fuentes. La actividad de evaluación, va a ser particularmente fundamental durante las instancias de conceptualización e implementación, para verificar la validez del diseño en base a propiedades que éste debe cumplir.

El ciclo de vida propuesto por la metodología es de Prototipos Evolutivos. Las actividades de desarrollo son ejecutadas en el orden propuesto para la obtención de una versión o prototipo, pudiendo introducirse modificaciones (en cualquier momento del desarrollo), en cuyo caso se vuelve a la actividad previa que sea necesario. Por cada nuevo prototipo se debe comenzar nuevamente desde la planificación y ejecutar todas las actividades de desarrollo a partir de la especificación, es decir cumplir un nuevo ciclo.

A continuación se realiza un análisis de algunas de las actividades más importantes de la metodología, como lo son especificación, conceptualización, evaluación, implementación e integración,

incluyéndose para ello ejemplos del caso de estudio del presente trabajo. Por último se emiten algunas conclusiones del análisis de la metodología. Para esta metodología y para todas las demás analizadas a continuación, en las conclusiones se indican ventajas y desventajas de las siguientes actividades: *especificación y adquisición de conocimiento, conceptualización y formalización, implementación, evaluación, y por último integración*. Se tomó la decisión de unir especificación con adquisición de conocimiento por considerarse actividades suficientemente ligadas como para ser tratadas en forma conjunta. Lo mismo ocurre con las actividades conceptualización y formalización.

2.1.1. Actividades propuestas por la metodología

Especificación

Esta actividad siempre está presente en cualquier proceso de desarrollo, cualquiera sea el ciclo de vida que lo guíe. Es fundamental para tener claro qué se quiere obtener del proceso, antes de encarar cómo se va a obtener. Como resultado de esta actividad se genera un entregable, el documento de especificación de la ontología. Dependiendo del nivel de formalidad con que se confeccione este documento, [FGJ97] lo clasifica en informal (en lenguaje natural), semi-formal (usando representaciones intermedias entre el lenguaje natural y otras más formales) y formal (usando construcciones que [FGJ97] considera más formales, como las preguntas de competencia). En definitiva, esta clasificación tiene que ver

con la rigurosidad con que se exprese el contenido del documento, que es sin duda lo más importante.

De acuerdo a [FGJ97], la especificación debe contener, con respecto a la ontología a desarrollar:

- Propósito
- Nivel de formalidad
- Alcance

Dentro del propósito, debe especificarse para qué se va a usar la ontología, cuáles van ser los escenarios de uso y los usuarios finales involucrados. Una correcta definición del propósito es parte fundamental del proceso de desarrollo, ya que todas las actividades posteriores, especialmente la conceptualización y la evaluación van a depender del propósito.

Concretamente, en el caso de estudio de este trabajo, se tiene:

Propósito

Evaluar y clasificar documentos del dominio médico de acuerdo a niveles de calidad previamente definidos, para determinar su adaptabilidad a diferentes perfiles de usuarios que los consultan.

En este caso de estudio, entonces, la ontología, debe brindar la semántica requerida por un sistema que ofrece adaptabilidad. La evaluación y clasificación de los documentos se realizará en base a criterios o niveles de calidad previamente definidos.

Entre los posibles escenarios de uso, en el caso de estudio se distinguen los siguientes:

Escenario 1

Un usuario realiza una consulta para obtener información relacionada con el dominio de la salud (consulta puntual de una afección, consulta más amplia acerca de un tema o mecanismo de prevención, etc.). De acuerdo al perfil del usuario que ingresa y la precisión requerida por la consulta, el sistema debe recuperar los documentos apropiados.

Escenario 2

Otro escenario de uso no tan explícito está dado por la ejecución de consultas de información de documentos médicos, realizadas por agentes de software para generar algún resultado intermedio, que no fueron solicitadas directamente por un usuario humano.

Escenario 3

También existirá una interfaz construida para el usuario que actúe como administrador de las ontologías, que será un experto encargado de

mantener sus instancias actualizadas, por ejemplo definiendo perfiles de usuario, clasificando documentos, etc. Este se considera también un escenario de uso en el que interviene un perfil de usuario muy específico.

Como quedó explícito con los escenarios de uso previamente expuestos, los usuarios finales en este caso de estudio van a ser personas de muy diferentes niveles de formación que consultan artículos médicos, así como aquellas que van a intervenir en calidad de administradores de las ontologías.

El nivel de formalidad de las ontologías de este caso de estudio, una vez implementadas, va a ser alto, ya que se va a utilizar una herramienta que permita expresar axiomas, realizar controles de consistencia del modelo, etc. En [FGJ97] se hace referencia a [UG96], donde se clasifica las ontologías en: altamente informal, semi-informal, semi-formal y rigurosamente formal, esta clasificación no merece mayor análisis dada la existencia de una buena cantidad de herramientas que permiten formalizar una ontología utilizando lógica descriptiva o lógica de primer orden.

El alcance se define en [FGJ97] como “el conjunto de términos a ser representados, sus características y granularidad”. La definición del alcance de una ontología va a depender del dominio (contexto de aplicación) y del punto de vista desde el que se quiere modelar la realidad. Para determinar el conjunto de términos que conforman el

alcance de una ontología, [FGJ97] propone diferentes técnicas extraídas de trabajos previos, que se apoyan fuertemente en la actividad de adquisición de conocimiento. Algunas de ellas se describen brevemente a continuación.

[UKMZ97] describe un método que tiene como punto de partida una tormenta de ideas para identificar la mayor cantidad de términos potencialmente importantes, luego éstos se agrupan en diferentes áreas de trabajo, de las que se seleccionan los conceptos que se consideran centrales para la realidad que se está modelando. Todos los demás conceptos que se definan se van a relacionar con los conceptos centrales, ya sea por generalización, especialización, etc., evitándose la ambigüedad de las definiciones. También se identifican conceptos no específicos del dominio, como por ejemplo “tiempo” y meta-conceptos, que se encuentran en un nivel de abstracción más alto a los identificados inicialmente. Por último se agregan otros términos relacionados a los centrales, como sinónimos y otros que se usen comúnmente, que tengan algún tipo de relación con éstos, que podrían pertenecer a otras ontologías.

Si se aplica el método descrito al caso de estudio, de la tormenta de ideas debe surgir un núcleo o conjunto de términos que contribuyen al propósito de la ontología. Es decir, ellos deben permitir expresar la

calidad de los documentos médicos en relación con los diferentes perfiles de usuario. Intuitivamente surgen como conceptos centrales posibles:

- Nivel de calidad

Los niveles que se definan van a estar basados en diferentes factores de calidad como: frescura del documento, confianza en la fuente, etc.

- Perfil de usuario

Ejemplos: usuario común, usuario experto, etc.

- Documento

Este es un concepto que desde el inicio se visualiza como un objeto complejo de la realidad, ya que en él deben estar representadas un conjunto de propiedades que permitan la evaluación de acuerdo a los criterios de calidad. A modo de ejemplo algunas de las propiedades relevantes son: área temática, origen del documento (sitio web, conferencia, etc.). Para la obtención del conjunto de valores posibles de las propiedades puede ser necesario utilizar taxonomías ya construidas, por ejemplo un tesoro que contenga los valores de la propiedad área temática. De acuerdo al valor de estas propiedades los documentos se van a clasificar de acuerdo a diferentes niveles de calidad. Por ejemplo, para el factor de calidad confianza en la fuente, los documentos podrían clasificarse en documentos de fuente confiable o no confiable.

Otra de las técnicas sugeridas en [FGJ97] para definir el alcance de una ontología, es la formulación de “competency questions”, que los autores toman de [GF95]. Básicamente, este método consiste en identificar los escenarios que motivan la ontología, con lo cual se visualizan los problemas de dicho escenario y se esboza una primera aproximación de los conceptos principales y sus relaciones, que dan solución a tales problemas. Para cada escenario surgen preguntas que definen los requerimientos de la ontología, que en principio se formulan en lenguaje natural y luego son traducidas a lógica de primer orden. La ontología que se va a diseñar debería ser capaz de responder a estas “competency questions”.

Al revisar el *Escenario 1* del caso de estudio, descrito anteriormente, se visualizan en primera instancia los conceptos de “documento” y “perfil”. Las preguntas de competencia que surgen de este escenario serían por ejemplo:

Pregunta 1

¿Qué información de un documento es relevante para determinar si éste se adapta a un determinado perfil?

Pregunta 2

¿Qué información es necesaria para distinguir un perfil de otro?

Pregunta 3

¿Qué conjunto de documentos tienen un nivel de recomendación adecuado a un perfil dado?

La respuesta a la primera de las preguntas puede motivar la asociación de diferentes atributos al concepto “documento”, a los efectos de clasificar los documentos de acuerdo al valor de tales atributos. La respuesta a la segunda pregunta podría llevar a definir una relación del concepto “perfil” con el concepto “documento” u otro concepto que represente las clasificaciones posibles de documentos.

[FGJ97] se basa en el trabajo [UG96] realizado por Mike Ushold y Michael Grüninger, para afirmar que el enfoque *middle-out* es el más conveniente en la identificación de los conceptos principales. Este enfoque consiste básicamente en definir los conceptos a un nivel de granularidad intermedio, que es en general la granularidad de los conceptos identificados aplicando las técnicas descritas anteriormente. Si es necesario, estos conceptos pueden generalizarse o especializarse. Los enfoques restantes, *top-down* y *bottom-up*, que por su nombre resultan claros, están detallados en [UG96].

En [FGJ97] se mencionan otras técnicas interesantes de adquisición de conocimiento, para definir el alcance de una ontología. Entre ellas se citan, entrevistas a expertos de dominio, formales e informales y análisis formal e informal de textos. Como forma de expresar los conceptos extraídos en estas instancias se sugiere también la utilización de

representaciones intermedias, que si bien se explotan en su totalidad en la actividad de conceptualización, algunas de ellas son adecuadas para expresar el alcance. La construcción de un glosario de términos o una versión primaria de árboles de clasificación de conceptos (que serán explicados en la actividad de conceptualización) pueden resultar de mucha utilidad, dando como resultado una conceptualización grosera que permite acotar el alcance del trabajo y realizar algunos chequeos como detectar términos que son sinónimos, etc.

Por último, [FGJ97] define tres cualidades que debe cumplir el documento de especificación:

- Conciso, esto significa que todos los conceptos abarcados deben ser relevantes.
- Parcialmente completo, con respecto a la cobertura de los términos definidos en el alcance. No es posible verificar la completitud absoluta ya que siempre es posible agregar nuevos términos, generalizar o especializar.
- Consistente, es decir, que los términos definidos con su correspondiente significado tengan sentido para el dominio de aplicación.

Estos tres criterios de evaluación de un documento de especificación son válidos para cualquier ontología de cualquier dominio, y en particular lo son para el caso de estudio del presente trabajo.

Conceptualización

La actividad de conceptualización en el desarrollo de una ontología es sin duda la más importante de todo el proceso. Es en ella que, a partir de una correcta especificación, debe realizarse una abstracción que “mire” la realidad desde el punto de vista que interesa a los efectos del propósito y alcance definidos en la misma. En [GFC03], [FGJ97] y [FGV96] se define un método para encarar esta actividad, en el cual se establecen una serie de pasos a seguir para construir el modelo conceptual, utilizando como herramienta representaciones intermedias (tablas y diagramas) que permiten realizar la conceptualización con un enfoque semi-formal. [FGPP99] es una ejemplificación muy clara del método, en base a un caso de estudio del dominio químico. A continuación se explican cada uno de los pasos propuestos en la bibliografía mencionada, a través de ejemplos que surgen de la idea, aún muy superficial, que se tiene de la ontología a desarrollar en el caso de estudio de este trabajo.

Tarea 1 – Glosario de términos

Se comienza por construir una tabla con los términos que se consideren relevantes. Si en la instancia de especificación se identificaron escenarios de uso y para cada uno de ellos preguntas de competencia (al menos en

lenguaje natural), éstos van a ser una de las fuentes para obtener este glosario. Entre otras fuentes, se utiliza tormenta de ideas o alguna otra forma de adquisición de conocimiento como entrevistas con expertos o estudio de documentos del dominio, como se mencionó en la actividad de especificación.

Alguno de los términos relevantes que se visualizan en el caso de estudio son: *documento, evaluación, clasificación de documentos, nivel de calidad, usuario, perfil de usuario.*

Para cada término, la metodología propone agregar una descripción, sinónimos, acrónimos y además su clasificación en concepto, relación, atributo, etc. *Clasificación de documentos y nivel de calidad* serían términos sinónimos en este ejemplo.

Tarea 2 – Taxonomías de conceptos

Esta tarea consiste en seleccionar y agrupar los términos obtenidos en la tarea anterior, a los efectos de identificar conceptos y organizarlos jerárquicamente. En [GFC03] se describen cuatro relaciones taxonómicas para construir la jerarquía, ellas son: *Subclass-Of, Disjoint-Decomposition, Exhaustive-Decomposition* y *Partition*.

Como ejemplo pueden citarse los conceptos *Documento, Nivel de Calidad, Perfil de Usuario y Usuario* y dos conceptos *Usuario Experto y Usuario Común*, que pueden verse como una *Partition* de *Usuario*.

Tarea 3 – Diagramas de relaciones binarias

Se identifican relaciones entre pares de conceptos. Entre los conceptos *Documento* y *Nivel de Calidad* posiblemente va a existir una relación *se clasifica*, para vincular cada documento a la clasificación que va a permitir su evaluación.

Tarea 4 – Diccionario de conceptos

Consiste en detallar los conceptos identificados en las taxonomías, colocándolos en una tabla y agregándoles la siguiente información: atributos de clase, atributos de instancia y relaciones. Para el concepto *Documento* se va a tener por ejemplo, el atributo de instancia *fecha de publicación* y la relación *se clasifica*.

Tarea 5 – Detalle de relaciones binarias

Con las relaciones identificadas en la Tarea 3 se construye una tabla cuyo objetivo es detallar para cada una: conceptos fuente y destino, cardinalidad, relación inversa y propiedades matemáticas. Para la relación *se clasifica* se tiene: concepto fuente *Documento*, concepto destino *Nivel de Calidad*, cardinalidad 1 si se asocia un único nivel de calidad a cada documento y si existe relación inversa sería *incluye*.

Tarea 6 – Detalle de atributos de instancia

Se construye una tabla con los atributos de instancia identificados en la Tarea 4, con los siguientes datos: concepto, tipo de valor, unidad de medida, precisión, rango de valores, valores por defecto, cardinalidades mínima y máxima, atributos de instancia, atributos de clase y constantes utilizados para inferir los valores del atributo, atributos que pueden ser inferidos utilizando el atributo, fórmulas o reglas utilizadas para inferir los valores del atributo y referencias usadas para definir el atributo. Un ejemplo es el atributo *fecha de publicación* del concepto *Documento*, con: tipo de valor fecha, cardinalidades mínima y máxima 1, si existiera un atributo *nivel de frescura* podría ser inferido a partir de este atributo, etc.

Tarea 7 – Detalle de atributos de clase

El detalle consiste en una tabla muy similar a la de los atributos de instancia, para los atributos fijos que pueden existir para un concepto. Si se justificara tener un atributo *tipo de usuario* en el concepto *usuario Común*, su valor sería siempre “Común”, a nivel de concepto.

Tarea 8 – Detalle de constantes

Para las constantes identificadas se tiene otra tabla donde se indica tipo de valor, valor y unidad de medida. En el dominio en cuestión una constante podría ser: *mínima fecha de publicación*, para establecer un límite de frescura aceptable para todos los documentos.

En la Figura 2 se muestra un primer modelo de alto nivel que conceptualiza la evaluación y clasificación de documentos para determinar su adaptación a diferentes perfiles de usuario, sin profundizar en el dominio médico, con detalles mínimos de la realidad del caso de estudio.

Una vez definidos los elementos principales del modelo conceptual, si la ontología a desarrollar es “pesada”, se sugiere establecer axiomas y reglas.

Tarea 9 – Definición formal de axiomas

Se establecen axiomas para especificar el modelo con más precisión. Volviendo a las “*competency questions*”, la *Pregunta 3* puede ser contestada formulando el siguiente axioma, de acuerdo al modelo de la Figura 2:

Forall(?P,?N,?D)

([se adapta](?P,?N) and [se clasifica](?D,?N) → [entiende](?P,?D))

siendo:

?P conjunto de perfiles

?D conjunto de documentos

?N conjunto de niveles de calidad

Tarea 10 – Definición formal de reglas

A los efectos de inferir conocimiento, se definen reglas con el formato:

If <conditions> then <consequent>

Tarea 11 – Definición de instancias

Esta tarea se justifica cuando existen instancias relevantes que se consideren importantes en la definición del modelo.

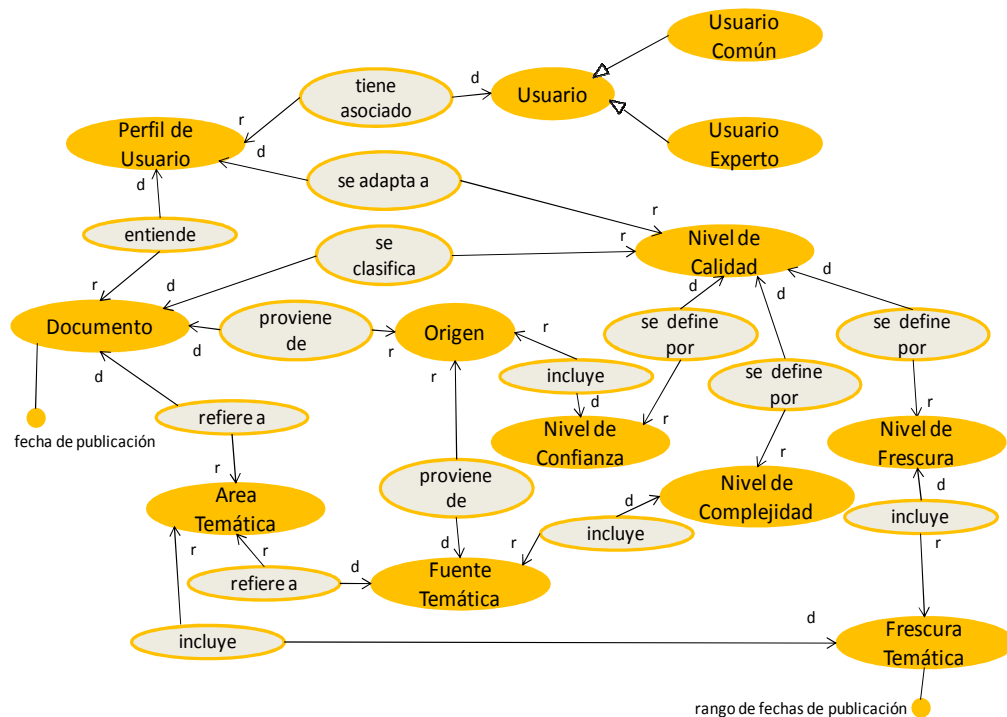


Figura 2. Conceptualización de la evaluación de documentos y perfiles de usuario.

Evaluación

Esta actividad puede realizarse durante la ejecución de cualquiera de las actividades de desarrollo. No obstante esto, es sumamente importante evaluar la ontología en la conceptualización, que es una instancia temprana del desarrollo, evitándose de esta forma la propagación de errores a actividades posteriores. La actividad de implementación es

también un buen momento para la evaluación, pues permite verificar la utilidad de la ontología desde un punto de vista más práctico.

Para cada uno de los pasos detallados anteriormente para la actividad de conceptualización, en [FGV96] se definen una serie de chequeos de completitud, consistencia, no repetición, etc. que permiten una evaluación temprana de la ontología. Por ejemplo, debe verificarse que si los términos *Documento*, *Nivel de Calidad*, *Perfil de Usuario*, *Usuario*, *Usuario Experto* y *Usuario Común* se incluyen en una taxonomía de conceptos (Tarea 2), deben existir en el glosario de términos (Tarea 1).

Según [GF95] la utilidad de las preguntas de competencia radica en que permiten evaluar la ontología, si la ontología es capaz de contestarlas se está validando el modelo en una etapa temprana del desarrollo. El axioma formulado a modo de ejemplo en la descripción de la Tarea 9 propuesta para la conceptualización, respondería formalmente a una de estas preguntas.

En [FGJ97] también se sugiere como marco de referencia para la evaluación, las tareas propuestas por [GJP95]. En este paper, Gómez-Pérez, Juristo y Pazos realizan un estudio meticuloso de todas las actividades de evaluación así como los actores que intervienen en la misma. Básicamente se definen diferentes tareas a realizar por tres equipos de desarrollo diferentes: de ontologías, de sistemas con bases de conocimiento y de agentes de software.

Los dos últimos representan al usuario final que utiliza las ontologías construidas por el primero de los actores, reusando y compartiendo el conocimiento que éstas proporcionan.

En [GJP95] las actividades de evaluación por parte del equipo de desarrollo de las ontologías son agrupadas en: actividades de verificación y actividades de validación. Las primeras aseguran más que nada la correctitud sintáctica, la consistencia, completitud y extensibilidad de las ontologías, entre otras propiedades, así como la correcta construcción del software asociado. El objetivo de la verificación apunta mayormente a asegurar la correctitud general de la ontología. Las actividades de validación controlan que el significado de las definiciones de la ontología expresen la realidad que se intenta modelar y que el software asociado sea capaz de dar una respuesta satisfactoria de acuerdo a los requerimientos. Es decir que la validación asegura que la representación de la realidad sea semánticamente correcta.

Al equipo de desarrollo de sistemas con bases de conocimiento le interesa fundamentalmente que la ontología sea extensible, usable, portable y con un grado de abstracción y granularidad que se adapte a sus requerimientos. Igualmente importante es contar con un ambiente de software que dé facilidades para encontrar las definiciones y permita una correcta integración de las mismas al sistema.

Por último, para el equipo de desarrollo de agentes de software es importante que la comunicación entre agentes y ontologías esté bien resuelta, siendo capaz de resolver ambigüedades, inconsistencias, omisiones, etc.

Implementación

La bibliografía consultada acerca de esta actividad no aborda el tema en profundidad, sólo menciona algunos aspectos a tener en cuenta.

[FGJ97] destaca la importancia de realizar una correcta selección del lenguaje en el que se implemente la ontología, y del ambiente de desarrollo a utilizar. Para la selección del lenguaje, se hace referencia a [SRVM95], que proporciona resultados acerca de tiempo y espacio consumidos en la carga y clasificación de bases de conocimiento de ontologías en diferentes lenguajes (LOOM, BACK, BACK++, CLASSIC). Actualmente sería de gran utilidad basarse en resultados similares para lenguajes como OWL.

La recomendación de Methontology [FGJ97] para los ambientes de desarrollo es que éstos cuenten con analizadores léxicos y sintácticos, traductores a otros lenguajes, editores ricos en funcionalidades, browsers, buscadores, módulos encargados de detectar incompletitud, inconsistencias y redundancia, y módulos de mantenimiento automático de definiciones.

Integración

Esta actividad debe realizarse cuando la ontología no se construye desde cero, es decir que de alguna manera se reutilizan ontologías ya construidas. La bibliografía consultada [GFC03] distingue varias modalidades de integración, algunas variaciones radican en que la ontología reutilizada se modifica o se mantiene intacta, también pueden fusionarse varias ontologías, etc.

Más allá del tipo de integración que se aplique, la misma debe realizarse durante la conceptualización, para que en el modelo conceptual ya se tengan en cuenta las interrelaciones con las ontologías a reutilizar, lo que simplifica las etapas posteriores del desarrollo. En la bibliografía se habla de ciclos de vida interrelacionados, expresando que las actividades realizadas en la ontología que se está diseñando pueden disparar actividades en las ontologías reutilizadas (por ejemplo adaptaciones, traducciones, etc.).

Es importante destacar que en el diseño de ontologías siempre debería analizarse la conveniencia de reutilizar trabajos anteriores, para no duplicar trabajo y poder concentrar energías en la creación de la porción del modelo que da respuesta a problemas aún no resueltos. Algunas de las tareas propuestas para la integración son: inspeccionar meta-ontologías que se puedan reusar y bibliotecas de ontologías que provean

definiciones coincidentes con los términos identificados en la conceptualización.

2.1.2. Conclusiones

Partiendo de la definición de *metodología* dada por la IEEE, como conjunto de técnicas o métodos para crear una teoría general de sistemas de cómo debe ser ejecutado todo proceso [GFC03], y comparándola con las actividades propuestas para Methontology en la bibliografía consultada, se puede ver que ésta cumple con dicha definición.

Como apreciación general de la metodología, para cada una de las actividades propuestas sus autores hacen referencia a trabajos previos, por lo que no existe uniformidad en el grado de detalle con que se aborda cada una de ellas.

A continuación se presentan las conclusiones para cada actividad propuesta.

Especificación y adquisición de conocimiento

En esta actividad se deja la puerta abierta a la aplicación de diferentes métodos o técnicas, que pueden diferir sustancialmente uno de otro, sin tomar posición por una forma de trabajo específica (tormenta de ideas, preguntas de competencia, entrevistas a expertos, análisis de textos). Esta característica permite una adaptación a los recursos disponibles para la adquisición de conocimiento, ya que por ejemplo, si existe gran

cantidad de documentación confiable, se realizará análisis de textos; si los expertos o usuarios especializados están disponibles para trabajar en el proyecto, se realizarán entrevistas y eventualmente se obtendrán preguntas de competencia, etc.

Si bien se proponen diversas alternativas de trabajo, ninguna de ellas se especifica con un nivel de detalle que permita al ingeniero de ontologías contar con una guía de trabajo concreta. En ninguno de los métodos sugeridos se define en forma precisa cómo surgen los conceptos principales o centrales de la ontología. Por ejemplo, no queda claro cuando se habla de entrevistas estructuradas cuál debe ser justamente la estructura de las mismas y cómo se extraen los términos principales a partir de la información provista por los expertos (por ejemplo si se extraen determinadas construcciones gramaticales). Lo mismo sucede con la obtención de preguntas de competencia a partir de escenarios de uso, no se especifica una técnica concreta para ello.

Conceptualización y formalización

Esta es la actividad a la que se da más importancia en la metodología, y como puede verse en la descripción realizada, se definen una serie de pasos que culminan con una definición del modelo utilizando construcciones semi-formales (representaciones) y lenguaje formal para la formulación de axiomas y reglas de inferencia.

Quizás por tratarse de una metodología concebida hace ya unos cuantos años, no se tiene en cuenta que la actividad de conceptualización debe dar soporte también a una forma de trabajo colaborativa y distribuida de construcción de las ontologías. En la actualidad esta modalidad es más común que el diseño centralizado que proponen las metodologías más antiguas. La conceptualización es la instancia en que se debe lograr un acuerdo consensuado de todos los actores involucrados acerca del modelo.

Implementación

Esta actividad es abordada en forma superficial, si bien es claro que es una actividad que depende en gran medida de la funcionalidad de las herramientas disponibles.

Evaluación

Los chequeos definidos a nivel de la actividad de conceptualización muestran que la metodología pone énfasis en esta actividad, ya desde las etapas tempranas del desarrollo. [GJP95] es una guía exhaustiva muy útil para no olvidar ningún punto importante a evaluar, si bien tampoco explica qué métodos (más o menos formales) aplicar para realizar cada uno de los chequeos.

Integración

Esta actividad es tratada con bastante profundidad. En [GFC03], además de detallarse las diferentes formas de integración, se ejemplifica esta actividad en base a un caso de estudio del dominio químico, en forma exhaustiva.

2.2. Metodología SENSUS-based method

SENSUS-based method, más que una metodología es un método propuesto por Bill Swartout, Ramesh Patil, Kevin Knight y Tom Russ, para construir el esqueleto de una ontología de dominio a partir de una ontología ya existente de amplia cobertura, SENSUS, desarrollada por el Natural Language Group at ISI. En [SPKR96], los autores describen una serie de procesos a seguir para identificar los términos que definen el alcance de la ontología. Además se especifican una serie de propiedades que las ontologías deben cumplir, algunas buenas prácticas de diseño y se propone una herramienta específica para la construcción de ontologías aplicando el método.

SENSUS es una gran estructura de conceptos asociados al lenguaje cotidiano, organizados jerárquicamente, que en los niveles superiores contiene conceptos más abstractos, agrupando a los más concretos de los niveles inferiores. Se diría más bien que es una ontología liviana o una taxonomía de términos enriquecida.

Básicamente el método consiste en diseñar una ontología tomando en principio toda la estructura de la ontología SENSUS, extendiéndola con

los conceptos más específicos del dominio que se quiere modelar y eliminando aquellos conceptos de SENSUS irrelevantes para el dominio. En la sección siguiente se detallan los pasos a seguir en la construcción de una ontología según [SPKR96], con una posible aplicación al caso de estudio de este trabajo. Estos pasos aparecen graficados en la Figura 3, para el dominio militar.

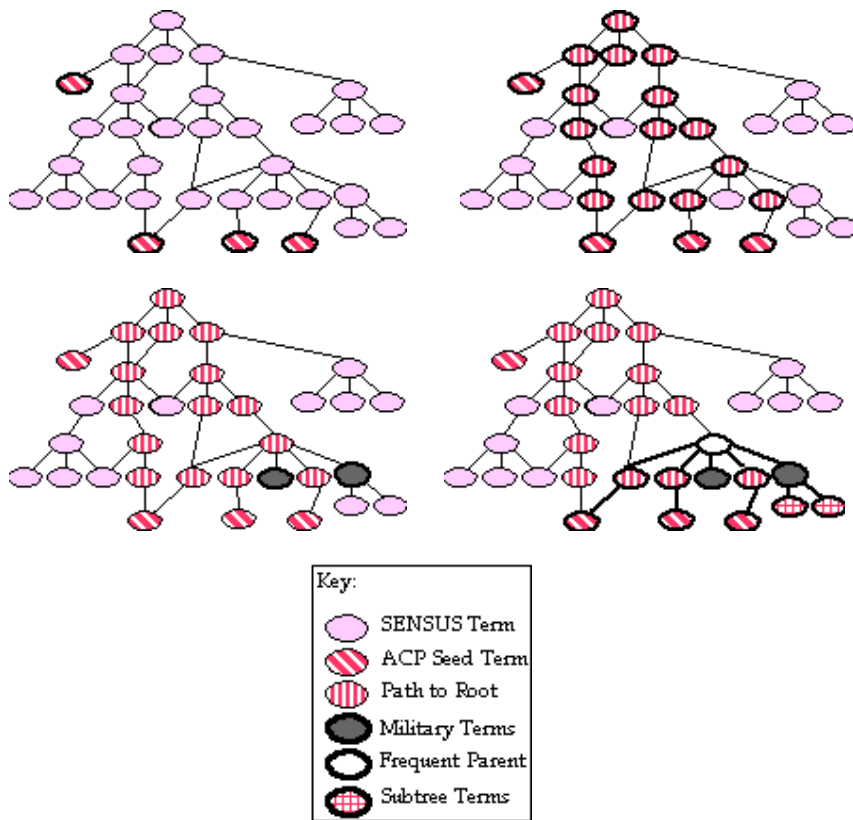


Figura 3. Pasos a seguir en la metodología SENSUS-based method [SPKR96].

2.2.1. Actividades propuestas por la metodología

Paso 1 – Identificación de términos “semilla”

En la descripción del método se sugiere que esta tarea se lleve a cabo con expertos en el dominio, encargados de identificar estos primeros términos, para obtener los principales conceptos relevantes. Aparte de esta modalidad, en [SPKR96] se sugiere la aplicación de “linguistics” o “conceptual clustering” para la identificación de los conceptos iniciales.

Si se considera el *Escenario 1* planteado anteriormente para el caso de estudio, *Documento y Perfil de Usuario* serían posibles términos semilla.

Paso 2 – Inserción manual de los términos “semilla” en la jerarquía SENSUS

Si los términos identificados en el paso anterior no existen en SENSUS, los mismos son agregados a la jerarquía como especializaciones de conceptos existentes.

En los pasos siguientes se realiza, por un lado una selección de los conceptos de SENSUS que van a permanecer en la nueva ontología, y por otro la inclusión de nuevos conceptos que no surgieron originalmente.

Paso 3 – Inclusión de todos los conceptos hasta la raíz de SENSUS

Siguiendo el camino desde cada término “semilla” hasta la raíz de SENSUS, se incluyen todos los conceptos del camino en la nueva ontología.

Paso 4 – Inclusión de otros términos relevantes del dominio

Se incluyen los términos relevantes para el dominio que no hayan surgido inicialmente, repitiéndose los pasos 2 y 3.

Por ejemplo el concepto *Nivel de Calidad* puede no ser de los primeros términos identificados, incluyéndose en esta instancia.

Paso 5 – Inclusión de subárboles completos

Considerando los conceptos incluidos en la nueva ontología a través de los pasos anteriores, si existen muchos caminos de niveles más bajos que llegan a un concepto, se incluye el subárbol completo cuyo nodo raíz es dicho concepto.

La argumentación de este paso es que si existen muchos conceptos de un subárbol que son relevantes, es muy probable que los nodos de todo el subárbol lo sean. De todas maneras lo que propone el método es que este paso no sea automático, se elige cuáles subárboles incluir, ya que para los niveles superiores aunque existan muchos caminos que lleguen a ellos, incluir esos nodos significaría incluir toda la estructura de SENSUS.

2.2.2. Conclusiones

Como comentario general acerca de esta metodología, puede decirse que no existe una referencia explícita acerca del ciclo de vida que guía los pasos descritos. Dado que cada vez que se quieren agregar nuevos conceptos relevantes se vuelven a ejecutar los pasos 2 y 3, estaría

implícito un ciclo de vida con iteraciones hasta obtener el conjunto de conceptos que abarca el alcance de la ontología. Además no se define la frecuencia con que se ejecutaría el paso 5 (cada vez que se agrega un nuevo concepto, o cuando el diseñador lo considere necesario, etc.).

Si se intenta realizar un paralelismo entre los pasos propuestos en SENSUS-based method con las actividades propuestas por la metodología Methontology (estándar desde una perspectiva de Ingeniería de Software), se puede ver que SENSUS-based method ataca especialmente las actividades: especificación, parte de la conceptualización, implementación e integración. A continuación se presentan las conclusiones para cada actividad propuesta.

Especificación y adquisición de conocimiento

Esta actividad estaría dada por las tareas que apuntan a la identificación de los términos que definen el alcance de la ontología. Una de las ventajas de partir de una estructura común es que en cierto modo se tiene parte de la tarea de especificación resuelta, ya que dicha estructura sugiere términos relevantes que pueden no haber sido identificados partiendo de cero. Este aspecto positivo está condicionado por la mayor o menor adaptación de esta estructura al dominio del problema particular.

Si bien en [SPKR96] no se toma partido por ninguna técnica particular para identificar los términos “semilla”, al menos se sugiere la aplicación de

procedimientos con cierto grado de formalidad, como “linguistics” o “conceptual clustering”.

Conceptualización y formalización

De la misma forma que en la especificación, en esta actividad una de las ventajas de partir de una estructura común es que se tiene parte de la tarea de modelado resuelta. Pero a pesar de esto, las vinculaciones jerárquicas entre los conceptos ya existentes en la estructura deberían coincidir con las relaciones que se quieren expresar de acuerdo a la perspectiva con la que se mira una realidad en particular. Se corre el riesgo de que en muchos casos la adaptación a la estructura de SENSUS lleve a forzar relaciones entre conceptos, que no aparecerían en una abstracción realizada únicamente en base al problema que se intenta resolver. Además, los conceptos de los niveles superiores (propios de la ontología SENSUS, incluidos en el camino hasta la raíz), por ser muy abstractos, pueden resultar irrelevantes para dominios muy específicos, aumentando innecesariamente la cantidad de conceptos de la solución que posteriormente se va a implementar. Este aspecto del método puede también restar claridad a dicha solución.

En SENSUS-based method, la red semántica está dada por relaciones de subclase entre conceptos (especialización dada por la estructura jerárquica). Aparte de ellas, en [SPKR96] no se especifica cómo expresar otro tipo de relaciones. Por ejemplo, mirando la Figura 2, si se quisiera

representar la relación *se refiere a* cuyo dominio es el concepto *Documento* y su rango el concepto *Area Temática*, en el método no está especificado ningún mecanismo para ello. Esta actividad estaría por lo tanto incompleta. La razón de esto es que ya desde el inicio esta metodología fue concebida para la construcción del esqueleto de una ontología de dominio, no de la ontología completa.

Implementación

Esta actividad forma parte del método ya que se proponen herramientas específicas para llevarla a cabo (como Ontosaurus), fundamentándose su utilización en base a las funcionalidades que brinda.

Evaluación

No se menciona explícitamente en [SPKR96], por lo que definitivamente es una carencia de la metodología.

Integración

La integración de dos o más ontologías construidas aplicando el método, surge naturalmente por estar todas basadas en un esqueleto común (SENSUS). Quizás esta sea la principal ventaja que provee este método, ya que la misma estructura de SENSUS establece la correspondencia entre los conceptos de diferentes contextos.

Este punto a favor en cuanto a la facilidad de integración, puede ir en detrimento de la actividad de conceptualización, como se explicó anteriormente, al tener que incluir forzosamente los conceptos de SENSUS más abstractos. Dependiendo del problema a resolver, el tamaño de la ontología (si ya tiene gran cantidad de conceptos específicos la adaptación a la estructura va a aumentar aún más su tamaño), las herramientas disponibles, etc., deberá obtenerse un equilibrio entre claridad (y cantidad acotada de términos) y facilidad de integración.

2.3. Metodología On-To-Knowledge

On-To-Knowledge es un proyecto cuyo objetivo es aplicar ontologías para mejorar la calidad de la gestión de información electrónicamente disponible en organizaciones grandes y distribuidas. En el marco de este proyecto, Steffen Staab, Hans-Peter Schnurr, Rudi Studer y York Sure proponen una metodología de diseño y desarrollo de ontologías, en la que se da particular importancia al uso final que se va a dar a las ontologías, es decir para qué aplicaciones concretas son desarrolladas. En [SSSS00] los autores proponen un conjunto de actividades, esquematizadas en la Figura 4 y detalladas a continuación.

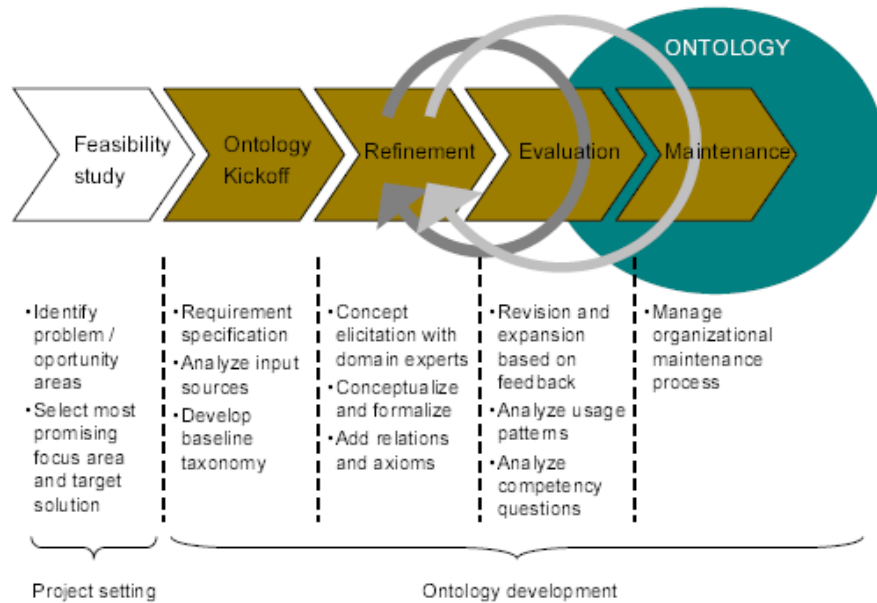


Figura 4. Actividades propuestas en la metodología On-To-Knowledge [SSSS00].

2.3.1. Actividades propuestas por la metodología

Estudio de factibilidad

Esta actividad, que no aparece en la descripción de otras metodologías, es importante en esta propuesta por el hecho de tratarse de una metodología orientada a la aplicación. Esto obliga a tener una fase previa donde se estudia la viabilidad del emprendimiento que va a utilizar la ontología.

En el caso de estudio de este trabajo, si la construcción de una ontología estuviera enfocada específicamente para ser usada por una herramienta que permita la evaluación de sitios Web que proveen información acerca de Alzheimer, debería estudiarse la factibilidad técnica y económica de

este emprendimiento, las necesidades reales de esta herramienta para los usuarios finales, etc.

Ontology Kickof

En esta actividad se determinan: propósito de la ontología, dominio y alcance, aplicaciones apoyadas por la ontología, fuentes de conocimiento, usuarios potenciales y escenarios de uso, preguntas de competencia y por último ontologías potencialmente reusables.

En el caso de estudio de este trabajo el propósito es el *Propósito* identificado en la sección 2.1.1.

La determinación del dominio y alcance, es decir, el conjunto de términos a tener en cuenta, surge, como se explica en un caso de estudio en [SSSS00], de entrevistas estructuradas con expertos de dominio.

En cuanto a los usuarios potenciales en el ejemplo se identifican: personas de muy diferentes niveles de formación que consultan artículos médicos, así como otras que intervienen en calidad de administradores de las ontologías. En cuanto a escenarios de uso se tendrían los *Escenarios 1, 2 y 3* detallados en sección 2.1.1.

La metodología sugiere la aplicación de la técnica *preguntas de competencia* para refinar los conceptos obtenidos al determinar el dominio y alcance. Estas preguntas también son formuladas por expertos de dominio, y como regla general, aunque con algunas excepciones, se

toman los sustantivos contenidos en estas preguntas para obtener los conceptos más específicos.

Refinamiento

En esta actividad se parte de los conceptos relevantes obtenidos en la actividad anterior, y nuevamente con los expertos del dominio, se realiza la conceptualización de la realidad, obteniéndose una “ontología semilla” que incluye además de conceptos y relaciones, la definición de axiomas que restringen el modelo. Además se implementa la ontología utilizando alguna herramienta que permita expresarla en un lenguaje de representación formal (seleccionado de acuerdo a los requerimientos de la aplicación que va a hacer uso de la ontología). También en esta instancia se podrían reusar ontologías existentes o al menos aprovechar decisiones de diseño.

Para esta actividad valen los ejemplos detallados en la actividad de conceptualización de Methontology.

Evaluación

La metodología propone como primera evaluación, verificar si la ontología diseñada satisface los requerimientos generales planteados en la actividad de Ontology Kickof, es decir: propósito, dominio, alcance y escenarios de uso para los usuarios identificados. Además se verifica que las preguntas de competencia puedan ser respondidas por el modelo.

Como ejemplo puede hacerse referencia a la *Pregunta 3* formulada en la sección 2.2.1, que es respondida a través de un axioma en la Tarea 9 de la actividad de conceptualización de Methontology.

También se considera una evaluación posterior en el ambiente de la aplicación, en base al feedback de los usuarios y realizando auditorías de las áreas más visitadas por éstos.

Estas evaluaciones pueden hacer que se vuelva a la actividad de refinamiento para realizar los ajustes necesarios, una y otra vez hasta obtener el nivel deseado para la ontología.

Mantenimiento

Una vez obtenida una primera versión de la ontología, en esta actividad se propone encarar los cambios que siempre sobrevienen, comenzando una nueva versión desde la primera actividad.

2.3.2. Conclusiones

El ciclo de vida propuesto es también de prototipos evolutivos. Después de cada evaluación, la metodología indica que es posible volver a la actividad de refinamiento. Como crítica a este mecanismo, se observa que en algunos casos puede ser necesario volver a la actividad *Ontology Kickof*, ya que si se modifica un requerimiento o surge una nueva pregunta de competencia, habría que esperar a completar la versión para

realizar este cambio. Este aspecto puede considerarse una rigidez de la metodología.

Más allá que la metodología se presenta en [SSSS00] como orientada a la aplicación, eliminando el estudio de factibilidad, las actividades planteadas se adaptan perfectamente al diseño de una ontología de dominio que pueda ser utilizada por diferentes aplicaciones.

Tomando como referencia Methontology, a continuación se evalúan las actividades que forman parte del estándar.

Especificación y adquisición de conocimiento

Corresponde a la actividad Ontology Kickof.

La utilización de entrevistas estructuradas a expertos para definir el dominio y alcance de la ontología, constituye una técnica con cierto grado de formalidad, aunque como aspecto negativo, en [SSSS00] no se detallan los criterios utilizados en la estructuración de las entrevistas.

Una vez más no se define en detalle cómo surgen las preguntas de competencia, si bien un punto a favor es que se define un método de identificación de conceptos (sustantivos), una vez formuladas dichas preguntas.

Conceptualización y formalización

Corresponden a parte de la actividad de refinamiento.

Estas actividades son abordadas con muy poca profundidad, a pesar de la importancia de las decisiones que se toman en esta instancia. Nuevamente no se propone ningún método, formal o informal, para la obtención de la “ontología semilla”.

Implementación

También es parte de la actividad de refinamiento.

El criterio de selección de las herramientas de acuerdo a la aplicación específica es correcto si la ontología se va a utilizar para un caso particular y no se tiene en mente su reutilización. De lo contrario debería cambiarse el criterio, es decir modificar esta actividad, teniendo en cuenta herramientas independientes de una aplicación, si la metodología se quiere adaptar a ontologías de dominio.

Evaluación

La metodología pone énfasis en esta actividad, ya desde las etapas tempranas del desarrollo, aunque definiendo los chequeos a realizar en forma poco precisa.

Integración

No se da mucha importancia a este aspecto, en parte por ser una metodología que se define como orientada a aplicaciones.

2.4. Metodología DILIGENT

DILIGENT (DIstributed, Loosely-controlled and evolVnG Engineering processes of oNTologies) es un enfoque presentado como metodología de diseño en [VPTS05] por Denny Vrandečić, Sofia Pinto, Christoph Tempich y York Sure, si bien en publicaciones anteriores ya se habla de DILIGENT como proceso de ingeniería de ontologías. Esta metodología difiere sustancialmente de las anteriormente estudiadas en que provee un mecanismo para facilitar la construcción colaborativa y distribuida de ontologías.

Una vez construido el núcleo central de una ontología que será utilizada y mantenida *descentralizadamente*, esta metodología describe cómo diferentes usuarios pueden modificar conceptualmente la ontología en forma local, así como sugerir cambios al núcleo central. La metodología DILIGENT está pensada para el caso en que existe un modelo central que es utilizado por diferentes grupos de usuarios, que a su vez extienden dicho modelo de acuerdo a sus necesidades. Como las necesidades varían a lo largo del tiempo, estas extensiones deben cambiar para satisfacerlas, lo que termina provocando también la necesidad de introducir cambios en la ontología central, ya que ésta tiene su razón de ser por la existencia de las aplicaciones de los diferentes grupos de trabajo. La gestión de todos estos cambios, es decir determinar cuáles se realizan centralmente y cuáles localmente, no es una actividad trivial, por

lo que en [VPTS05] esta actividad se divide en una serie de pasos que ayudan a trabajar más ordenadamente, facilitan la comunicación entre los diferentes actores y permiten que el núcleo continúe reflejando el conjunto de conceptos comunes a los diferentes grupos.

Los actores que intervienen en este proceso son: usuarios de la ontología, ingenieros de conocimiento, ingenieros de ontologías y expertos del dominio. La Figura 5 ilustra la participación de ellos en dicho proceso. Los pasos propuestos por esta metodología son: *construcción, adaptación local, análisis, revisión y actualización local*.

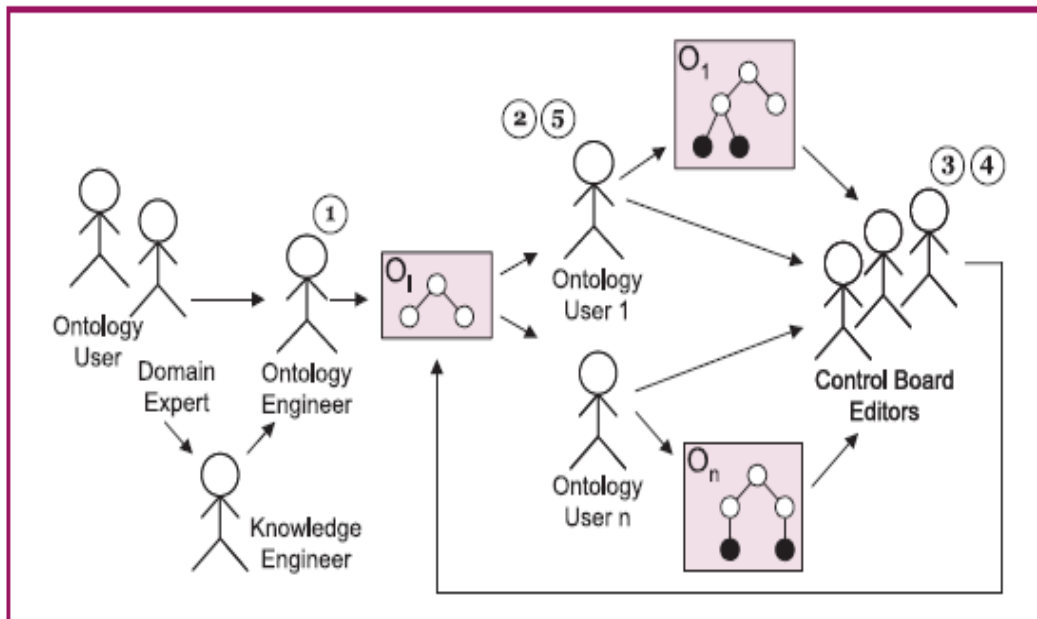


Figura 5. Participación de los diferentes actores propuesta por la metodología DILIGENT [VPTS05].

2.4.1. Actividades propuestas por la metodología

Construcción

Expertos del dominio, ingenieros de conocimiento e ingenieros de ontologías construyen una ontología inicial, que a diferencia de otras metodologías, no requiere completitud ya que se toma como una primera versión consensuada del núcleo, que será ampliado en etapas posteriores.

Adaptación local

Los usuarios de la ontología, cada uno con sus propios requerimientos de negocio, pueden adaptar localmente la ontología a sus necesidades, extendiendo e incluso modificando los conceptos compartidos.

Existe un ambiente central, el “control board”, donde se recogen los cambios realizados localmente a la ontología central como solicitudes de cambio y se auditan las adaptaciones locales.

Análisis

El “control board” realiza el análisis de las ontologías locales y los requerimientos de cambio al núcleo compartido e intenta identificar similitudes entre las ontologías y requerimientos locales. Se sugiere que en este ambiente intervengan todos los actores mencionados con sus diferentes niveles de conocimiento. En base al análisis mencionado se toma la decisión de qué cambios se introducen a la ontología central.

Dada la complejidad de esta tarea, [VPTS05] sugiere la utilización de un modelo de argumentación detallado en [TPSS05], que es una ontología construida para el manejo de argumentos, o sea las razones por las cuales se proponen los cambios al núcleo. Este modelo de argumentación está basado en el sistema IBIS (Issue-Based Information Systems), propuesto por Werner Kunz and Horst W. J. Rittel en 1970 [KR70], como un método de resolución de problemas en forma cooperativa. Básicamente consiste en que integrantes de un grupo proponen problemas a resolver, posibles soluciones y argumentos a favor o en contra de éstas, realizándose consultas concretas a expertos si es necesario. Se trata de un enfoque pensado para lograr una buena coordinación en los procesos de toma de decisiones.

Revisión

Periódicamente el “control board” debe revisar la ontología compartida, aún cuando no existan requerimientos de cambio, para asegurar que las ontologías locales no se desvíen demasiado del núcleo. Si esto ocurre la ontología compartida debe modificarse para que refleje las necesidades de los usuarios locales.

Es fundamental que en esta etapa participen todos los actores involucrados: usuarios para evaluar la usabilidad, expertos de dominio para evaluar los cambios desde el punto de vista del dominio, ingenieros de conocimiento para verificar cualidades técnicas como eficiencia y

mantenimiento de estándares, y obviamente ingenieros de ontologías para balancear técnicamente los argumentos y verificar la correctitud del diseño resultante.

Actualización local

Cada vez que una nueva versión de la ontología compartida es liberada, es bueno que los usuarios actualicen sus ontologías locales para una mejor explotación del conocimiento representado en la nueva versión compartida. Por ejemplo, si un término que había sido definido localmente fue incluido en el núcleo, es conveniente que localmente se utilice el concepto de la ontología central.

2.4.2. Aplicación al caso de estudio

En el caso de estudio del presente trabajo, podría aplicarse la metodología DILIGENT si se consideran dos niveles de ontologías bien diferenciados:

- Una ontología de alto nivel que modela conceptos generales relativos a documentos de diferentes áreas temáticas, criterios de evaluación de los mismos y perfiles de usuario. A modo de ejemplo, el modelo conceptual de la Figura 2 estaría representando la versión inicial de la ontología compartida, que será mantenida en forma centralizada.
- Varias ontologías de más bajo nivel, por ejemplo una por cada conjunto de documentos relativos a una determinada afección. En cada una de

estas ontologías los usuarios especializados en cada tema, junto con los ingenieros de ontologías se encargarían de refinar los conceptos de la ontología compartida y sugerir los cambios que consideren necesarios a dicha ontología.

En la Figura 6 se muestra un modelo que representa la especialización de algunos conceptos de la ontología compartida (*Area Temática, Usuario Experto y Usuario Común* de la Figura 2), para el caso específico de la enfermedad de Alzheimer. Los expertos en este dominio van a identificar estos conceptos en la instancia de Adaptación Local, mientras que expertos en otra afección van a agregar también localmente otro conjunto de conceptos, de los cuales algunos van a coincidir con los identificados por el grupo experto en Alzheimer y otros no. Por ejemplo, para las afecciones que no tienen efectos a nivel cerebral, no estaría identificado el concepto *Efecto Neuroquímico*, mientras que algunos conceptos como *Diagnóstico* son comunes a todas las enfermedades. Estos últimos serían los conceptos candidatos a formar parte de la ontología compartida, y una vez incluidos en ella por el “control board”, cada grupo local actualiza su modelo en la instancia de Actualización local, reutilizando los nuevos conceptos de la ontología compartida que se adaptan a su realidad.

2.4.3. Conclusiones

El ciclo de vida propuesto podría asimilarse a prototipos evolutivos, ya que por cada adaptación local, análisis, revisión o actualización local,

pueden surgir nuevos conceptos que son implementados para obtener una nueva versión. Como lo indica el propósito para el que fue creada la propia metodología, el ciclo de vida favorece la evolución de la ontología, lo que constituye un punto a favor.

Una vez más tomando como referencia Methontology, a continuación se evalúan las actividades que forman parte del estándar.

Especificación y adquisición de conocimiento

Están implícitas más que nada en las fases de construcción y adaptación local que son las etapas en las que surgen nuevas necesidades a ser conceptualizadas. En el análisis y revisión, en cambio, la creación de nuevos conceptos en la ontología compartida se alimenta de los conceptos sugeridos o ya creados localmente. De todas maneras [VPTS05] no profundiza ni hace referencia a ninguna técnica en particular para llevar adelante la especificación de requerimientos.

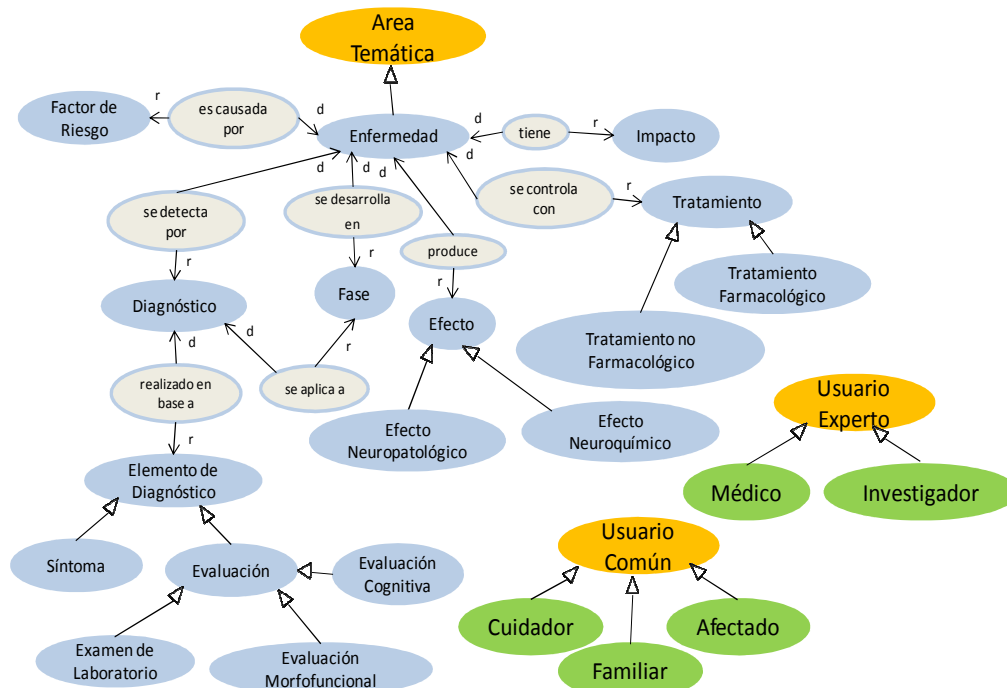


Figura 6. Conceptualización del dominio específico Alzheimer.

Conceptualización y formalización

Se puede decir que en todas las actividades propuestas por la metodología se está modificando el modelo, lo que permite la evolución de la conceptualización. La conceptualización es por lo tanto la actividad en la que más se profundiza, estableciéndose un cierto orden en la intervención de cada conjunto de actores. El intercambio de argumentos a través de la utilización de una ontología específica para ello, le imprime un carácter más formal a la conceptualización, ya que en [TPSS05] se hace referencia a un método bastante ordenado para tomar la decisión de incluir un nuevo concepto en la ontología compartida. Esta formalización del intercambio de argumentos no se ataca en ninguna de las

metodologías anteriormente estudiadas, por lo que se está en presencia de una metodología que da soporte a la necesidad de construcción colaborativa de ontologías, en especial a la toma de decisiones de modelado en la instancia de conceptualización.

Implementación

No se da importancia alguna a esta actividad en la metodología, salvo cuando se menciona que los ingenieros de conocimiento deben cuidar la eficiencia de las aplicaciones que utilicen las ontologías, cuidar que se mantengan los estándares, etc.

Evaluación

Tampoco se profundiza en esta actividad, únicamente se destaca la importancia de que los ingenieros de ontologías se encarguen de ella.

Integración

Está cubierta por la ontología compartida que integra las ontologías locales.

2.5. Metodología HCOME

Otro ejemplo de metodología orientada a la construcción colaborativa y descentralizada de ontologías es HCOME (Human-centered ontology engineering), propuesta por Konstantinos Kotis y George A. Vouros [KV06] poco después que DILIGENT, como una variación de ésta. Es por

este motivo que en este trabajo la misma será abordada con menor detalle que el resto de las metodologías, resaltando principalmente las diferencias con la metodología DILIGENT.

La razón de ser de este enfoque radica en la participación de los usuarios o expertos del dominio, apoyados por los ingenieros de ontologías, a lo largo de todo el proceso de definición de la ontología. Se propone además que este rol sea cumplido por los usuarios en sus actividades diarias dentro de la organización a la cual pertenecen, en cada momento que visualizan la necesidad de, ya sea representar nuevo conocimiento o modificar el ya existente, dado que la dinámica de cambios en las organizaciones aumenta su ritmo cada vez más.

Los pasos a seguir en este proceso son: *especificación*, *conceptualización* y *explotación*, los cuales se cumplen siguiendo un flujo que se muestra con claridad en la Figura 7. En ella se distinguen tres espacios de trabajo en los que se van modelando las ontologías:

- Espacios personales en los que los usuarios, y eventualmente ingenieros de conocimiento realizan conceptualizaciones localmente.
- Un espacio compartido o espacio de discusión como soporte al trabajo de un grupo.
- Un espacio consensuado que contiene los modelos que resultan de un acuerdo entre los actores involucrados.

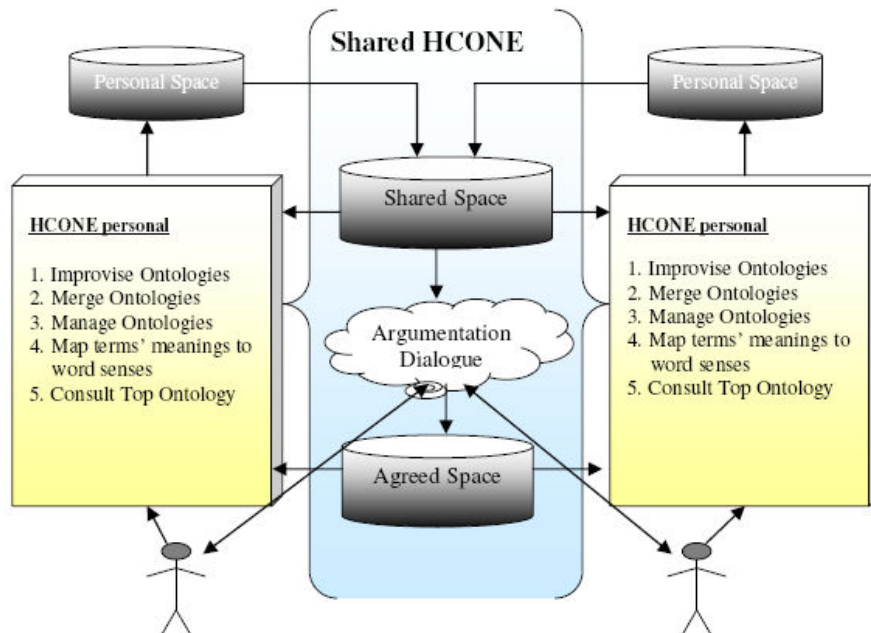


Figura 7. Dinámica de definición de ontologías en la metodología HCOME [KV06].

2.5.1. Actividades propuestas por la metodología

Especificación

En el espacio compartido los usuarios involucrados definen el propósito, alcance y requerimientos, ya sea para la construcción de una nueva ontología o la modificación de un modelo existente.

Conceptualización

En el espacio personal, cada usuario o ingeniero de ontologías, puede ingresar o modificar conceptos, y consultar y/o integrar ontologías existentes. Si lo requiere puede interactuar con expertos del dominio en el espacio compartido.

Explotación

En el espacio compartido se discuten las conceptualizaciones realizadas individualmente, proceso que está soportado por un modelo de argumentación basado en el modelo IBIS [KR70], ya mencionado en la exposición de DILIGENT. Las conceptualizaciones que alcanzan consenso pasan a formar parte del espacio consensuado.

Para llevar a cabo estos pasos en la forma descrita, la metodología propone la utilización de un ambiente de trabajo, HCONE (Human Centered ONtology Engineering Environment), que soporta este proceso de continuo intercambio, a la vez que brinda la funcionalidad de versionado.

2.5.2. Conclusiones

Por las mismas razones que las expuestas para la metodología DILIGENT, el ciclo de vida propuesto puede asimilarse a prototipos evolutivos, favoreciendo la evolución de la ontología.

Respecto al énfasis en el mantenimiento evolutivo y al enfoque colaborativo y distribuido, si bien en teoría el objetivo de la metodología HCOME es dar una participación activa a los usuarios en todas las instancias de la definición de una ontología, como parte de su actividad diaria, la aplicación práctica de la metodología tal cual es planteada en [KV2006], no queda clara, al menos cuando se trata de un número

importante de usuarios involucrados. Basta imaginar el hecho de que diariamente, además de cumplir con las actividades requeridas por la organización o empresa a la que pertenecen, los usuarios deban proponer nuevos requerimientos, conceptualizar en base a requerimientos acordados o participar de un debate acerca de una conceptualización, argumentando a favor o en contra. Si se trata de un grupo numeroso, cuyos integrantes tienen posiblemente diferentes percepciones de la realidad de acuerdo al rol que cumplen en la organización, surgen al menos dos interrogantes:

¿No se volverá demasiado tediosa la actividad diaria de los usuarios?

¿Aunque se tratara de usuarios sumamente competentes que son capaces de realizar todo lo que propone la metodología, se obtienen de esta forma buenos modelos?

Comparando una vez más con la metodología DILIGENT, si bien en ella se da participación a los usuarios permitiendo que éstos mantengan su modelo localmente y propongan cambios a la ontología central, las decisiones acerca del modelo compartido son tomadas por un grupo multidisciplinario dedicado específicamente a esa tarea, el “control board”. Si bien los integrantes de este grupo, que van a ser actores de diferente perfil (incluyo usuarios locales), intercambian argumentos para alcanzar una conceptualización consensuada, son en última instancia los

ingenieros de ontologías los responsables de la correctitud del modelo común.

A continuación se evalúan brevemente las actividades, en base al marco de referencia dado por Methontology.

Especificación y adquisición de conocimiento

Corresponde a la actividad de especificación, que a diferencia de DILIGENT, en HCOME se realiza siempre en forma descentralizada, sin partir de un modelo inicial. Tampoco se hace referencia a ninguna técnica en particular para llevar adelante esta tarea.

Conceptualización y formalización

Está presente en los pasos de conceptualización y explotación, manejándose también el intercambio de argumentos. Este intercambio, como se mencionó al inicio de las conclusiones, además de aparecer más desordenado que en la metodología DILIGENT, carece de un impulso inicial que guíe la argumentación de los usuarios, los cuales seguramente no son expertos en modelado. En DILIGENT, este impulso está dado por la actividad de construcción de una ontología inicial.

Implementación

Se propone una herramienta (HCONE) para llevar adelante esta actividad.

Cabe destacar que no se encontró en la bibliografía consultada ninguna aplicación concreta de la metodología, fuera de algunas pruebas realizadas en el propio ambiente académico en que ésta fue diseñada. Este es otro argumento para fundamentar el juicio emitido acerca de la eficiencia de la metodología en grupos de trabajo numerosos.

Evaluación

Tampoco se profundiza en esta actividad, aunque está implícito que los ingenieros de ontologías deben encargarse de ella.

Integración

Está cubierta por la ontología consensuada que integra las ontologías locales.

2.6. Metodología UPON

UPON (Unified Process for ONtology) es una metodología de diseño y desarrollo de ontologías presentada por Antonio De Nicola, Michele Missikoff y Roberto Navigli [NMN09] basada en el proceso de ingeniería de software Unified Process (UP), ampliamente conocido y utilizado.

Esta metodología propone el mismo ciclo de vida que UP, con las mismas actividades, adaptadas a las necesidades de diseño de ontologías. Esto significa que UPON está guiado por un ciclo de vida iterativo e incremental, basado en casos de uso [JBR99]. Puesto que la definición de

los casos de uso responde a necesidades de aplicaciones concretas, el enfoque UPON está concebido para la construcción de ontologías específicas de dominio e incluso para aquellas que dependen de la aplicación. En [NMN09] se propone también la utilización del lenguaje de modelado UML, ampliamente conocido y utilizado, que tiene la característica de ser un lenguaje extensible. Esta modalidad de representación, accesible a diversos tipos de usuarios, permite obtener fácilmente resultados intermedios a ser validados por éstos. La Figura 8 [NMN09] muestra el ciclo de vida y actividades de UPON, coincidentes con las definidas en UP.

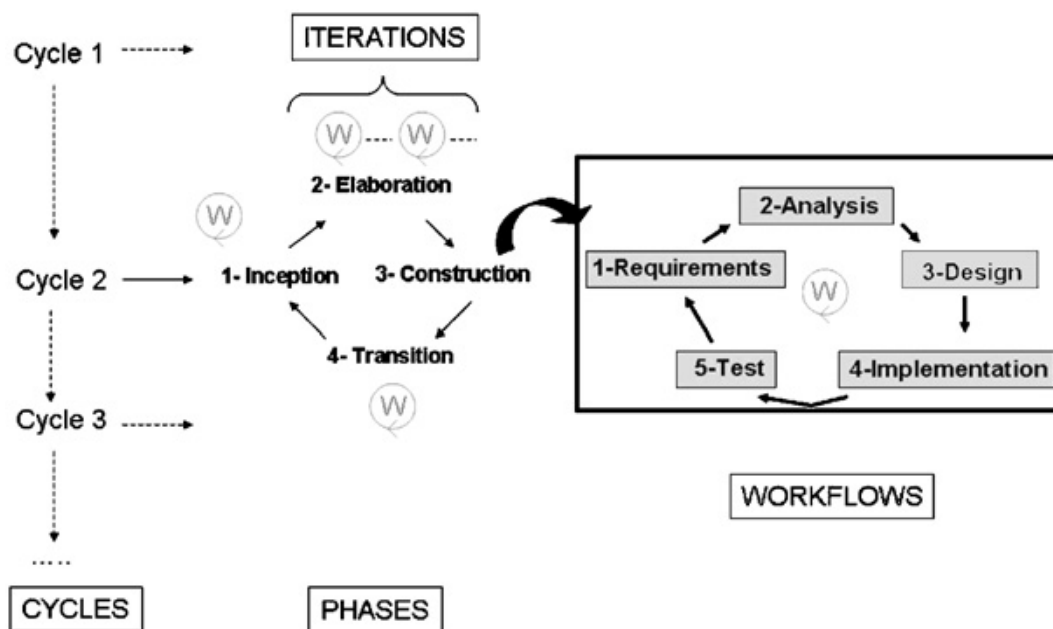


Figura 8. Ciclo de vida y actividades propuestas por la metodología UPON.

Esta metodología propone entonces para el diseño y desarrollo de ontologías, un ciclo de vida compuesto por ciclos, que a su vez están compuestos por fases, y en cada una de éstas ocurren una o más iteraciones. En cada iteración se van a ejecutar las actividades o “workflows”¹, que con la aplicación de técnicas específicas, constituyen las tareas a realizar para la obtención de los resultados esperados del proceso. En este enfoque, por cada ciclo se obtiene una nueva versión de la ontología. Al igual que en UP, si bien en cada fase se cumplen iteraciones en las que siempre se ejecutan las actividades de requerimientos, análisis, diseño, implementación y testeado, el peso de cada una de éstas difiere en cada fase. Por ejemplo, en la fase inception se realiza mayoritariamente análisis de requerimientos, si bien se analiza, diseña, implementa y prueba algún caso de uso puntual.

[NMN09] diferencia los roles a ser cumplidos por ingenieros de conocimiento y expertos de dominio en cada una de las actividades de la metodología, y además destaca que poco a poco está apareciendo un nuevo rol, el ingeniero de ontología, en el que se funden las habilidades de los dos primeros.

Además de describir cada una de las actividades propuestas incluyendo ejemplos del dominio de compras, [NMN09] presenta una formalización cada una de ellas, que no se incluye en este trabajo para no extender en

¹ En [NMN2009] se utiliza el término “workflow” para identificar cualquiera de las actividades que comúnmente se llevan a cabo en todo desarrollo de software (requerimientos, análisis, diseño, implementación, testeado).

exceso el espacio dedicado a cada una de las metodologías. En la siguiente sección se describen brevemente las actividades de *requerimientos, análisis, diseño, implementación y testeo*.

2.6.1. Actividades propuestas por la metodología

Requerimientos

El objetivo de esta actividad es determinar el dominio de interés y el alcance de la ontología, así como definir el propósito del negocio. Para obtener estos resultados, [NMN09] propone la aplicación de las siguientes técnicas:

- Escribir uno o más *escenarios de uso* o “storyboards”²
- Crear un *léxico de aplicación* (AL), consultando expertos del dominio y documentos específicos de la aplicación.
- Identificar *preguntas de competencia* (CQs).
- Identificar y priorizar los *casos de uso* relacionados.

Las tres primeras tareas pueden asimilarse a la identificación de los escenarios de uso, conjunto de términos relevantes y preguntas de competencia, ya definidos en la actividad de especificación de requerimientos de la metodología Methontology. En la sección 2.1.1 se

² [NMN2008] utiliza el término “storyboard” para referirse a lo que en Methontology fue identificado como escenario de uso.

explica en qué consiste cada una, además de proporcionarse ejemplos para el caso de estudio de este trabajo.

Con respecto al modelado de los casos de uso relacionados, según [NMN09], en el contexto de las ontologías los casos de uso corresponden a conocimiento que debe ser representado para ejecutar en forma exitosa las operaciones del negocio y responder a las preguntas de competencia. En la actividad de requerimientos únicamente se identifican y priorizan los casos de uso, para luego ser detallados en actividades posteriores. La priorización se realiza para determinar los casos de uso críticos que deben ser tratados en iteraciones tempranas, así como los menos relevantes que pueden ser pospuestos.

Por ejemplo, en el marco del *Escenario 1* identificado en la actividad de especificación descrita en la sección 2.1.1, se puede identificar:

Caso de Uso

Un médico (usuario experto) realiza una consulta con criterios específicos acerca del área temática Alzheimer.

De las tareas definidas para llevar a cabo la actividad de requerimientos, en las dos primeras hay mayor participación de los expertos de dominio, para las CQs se requiere una participación similar de expertos de dominio e ingenieros de conocimiento, y en la identificación y priorización de los casos de uso es fundamental el rol de los ingenieros de conocimiento.

Análisis

Siempre en base al trabajo de [NMN09], la actividad de análisis, que según propone UP es un análisis a nivel conceptual, puede resumirse como el “refinamiento y estructuración de los requerimientos de la ontología” identificados en la actividad anterior.

El léxico de aplicación generado en los requerimientos es enriquecido y transformado, para ir obteniendo paulatinamente una estructura que permite expresar el análisis conceptual de la ontología. A continuación se describen las tareas sugeridas por la metodología UPON para cumplir con el objetivo de la actividad de análisis.

- A partir de diferentes recursos del dominio, construir el léxico de dominio (DL).

Se extrae un conjunto de términos de diferentes fuentes del dominio de interés, en especial de documentos técnicos, tesauros, léxicos ya existentes, etc. Esta tarea puede ser realizada en forma automática aplicando técnicas como text mining.

Por ejemplo, tomando como base el *Caso de Uso* identificado en la especificación, pueden consultarse ontologías médicas como GALEN, UMLS, ON9 [GFC03], documentos Web de fuente confiable, etc.

- Construir el léxico de referencia (RL).

Se realiza un “merge” del léxico de aplicación y el léxico de dominio. Los términos coincidentes en ambos léxicos son incluidos en el léxico de referencia, y se seleccionan términos de uno y otro léxico que no forman parte de la intersección, con la aprobación de expertos de dominio y usuarios. De esta forma se extiende el conjunto intersección con términos de dominio que ayudan a la conceptualización de la aplicación y términos específicos de la aplicación que resulten relevantes aunque no tan generales para el dominio.

La Figura 9 muestra un fragmento de la ontología médica UMLS. Se puede considerar a modo de ejemplo que los conceptos representados en ella son parte del léxico de dominio y los conceptos específicos identificados en la Figura 6 para Alzheimer conforman el léxico de aplicación. Se tiene entonces:

Léxico de aplicación: Demencia, Factor de Riesgo, Síntoma, Evaluación, Examen de Laboratorio, Evaluación Morfofuncional, Evaluación Cognitiva, Fase, Efecto, Efecto Neuropatológico, Efecto Neuroquímico, Impacto, Tratamiento, Tratamiento Farmacológico, Tratamiento no Farmacológico, Médico, Investigador, Familiar, Afectado, Cuidador.

Léxico de dominio: Biologic Function, Pathologic Function, Disease or Syndrome, Mental or Behavioral Dysfunction, Pharmacologic Substance.

Si se aplica el “merge” que propone la metodología, los conceptos *Tratamiento Farmacológico* del léxico de aplicación y *Pharmacologic Substance* del léxico de dominio podrían, o bien considerarse coincidentes (aplicando alguna simplificación) transformándose en un único concepto para el léxico de referencia, o bien mantenerse separados, acordando que un tratamiento farmacológico va a estar compuesto por una o más sustancias farmacológicas. Por otra parte, los conceptos *Demencia* y *Mental or Behavioral Dysfunction* podrían considerarse coincidentes, al menos a los efectos del ejemplo. Teniendo en cuenta estas consideraciones:

Léxico de referencia: Demencia, Factor de Riesgo, Síntoma, Evaluación, Examen de Laboratorio, Evaluación Morfofuncional, Evaluación Cognitiva, Fase, Efecto, Efecto Neuropatológico, Efecto Neuroquímico, Impacto, Tratamiento, Tratamiento Farmacológico, Tratamiento no Farmacológico, Médico, Investigador, Familiar, Afectado, Cuidador, Biologic Function, Pathologic Function, Disease or Syndrome, Pharmacologic Substance.

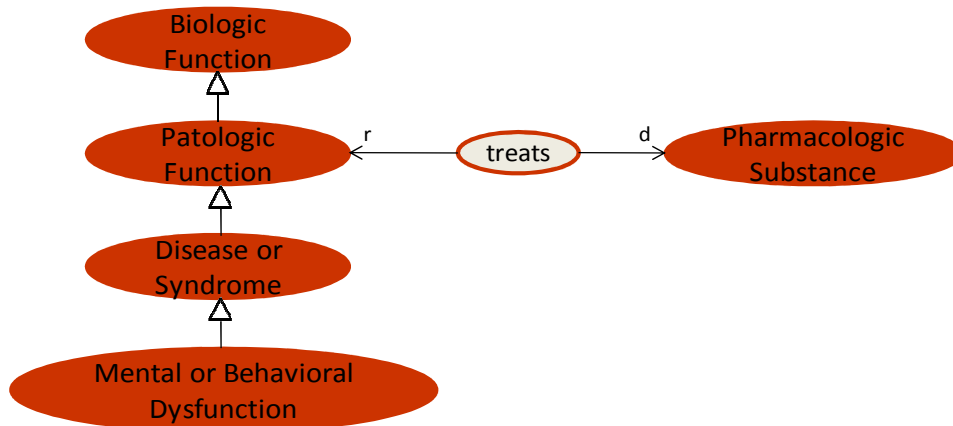


Figura 9. Diagrama construido en base a un fragmento de Semantic Network, una de las partes que componen la ontología médica UMLS [WR1] [GFC03].

- Modelar el escenario de la aplicación utilizando diagramas UML.

A los diagramas de casos de uso se agregan diagramas de actividad y de clases [JBR99], para construir un modelo de la aplicación, que va a ser utilizado en instancias posteriores para validar la ontología.

En la Figura 10 se muestra un ejemplo muy simple con diagramas UML que representan el *Caso de Uso* descrito anteriormente.

En la figura se observa que cada estereotipo o forma de representación caracteriza a cada elemento del modelo, como clase, actor o actividad, ubicándolo en una clasificación que coincide con las categorías OPAL (Object, Process, Actor modelling Language) [DMT07]. Esta categorización fue creada por Fulvio D'Antonio, Michele Missikoff y Francesco Taglino para guiar a los expertos de negocio que intervienen en la construcción de ontologías, proporcionando un marco en el cual

conceptualizar la realidad, siguiendo determinados patrones que cumplen una función asimilable a la aplicación de los patrones de diseño de software [GHJV94]. Se trata de un marco claramente orientado a aplicaciones de negocio.

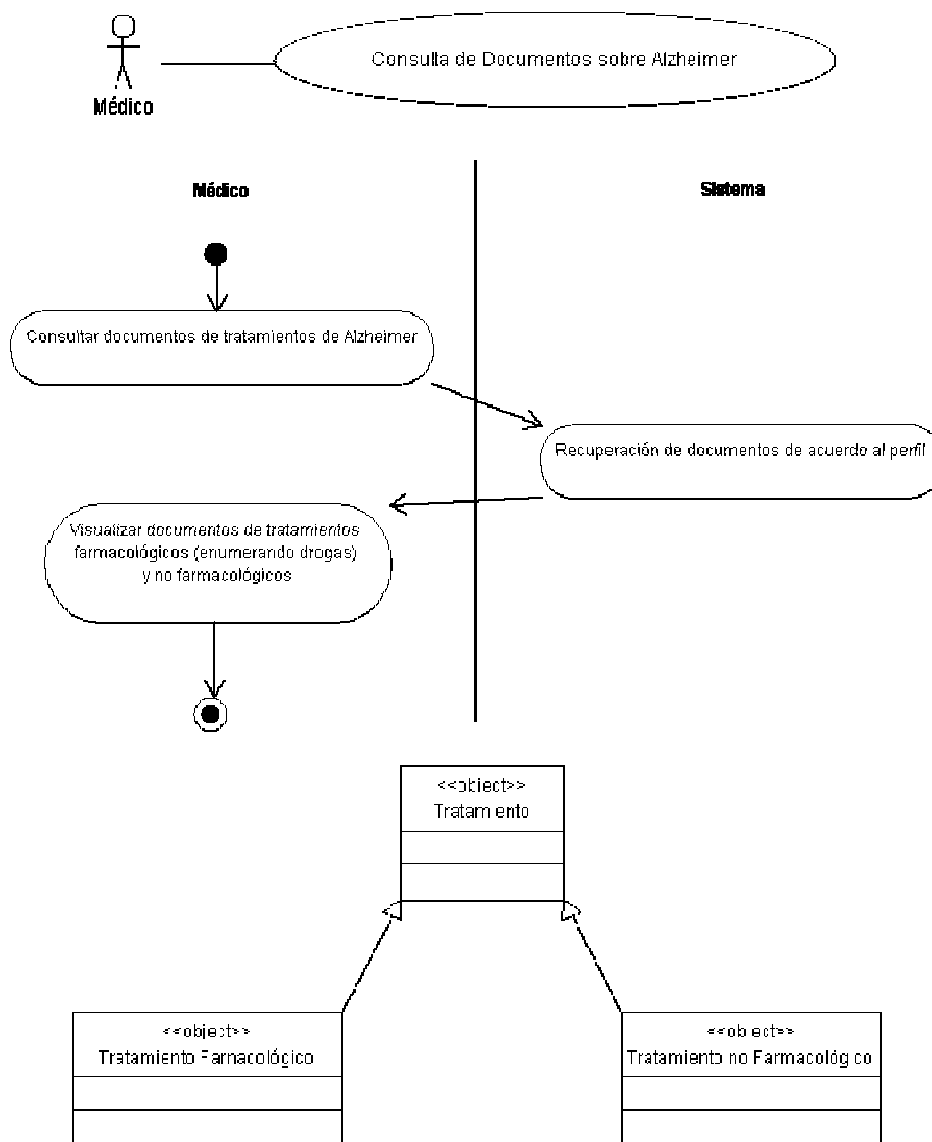


Figura 10. Modelo expresado a través de diagramas UML.

- Construir el glosario de referencia (RG).

A partir del léxico de referencia, se asocian a éste definiciones informales en lenguaje natural, extraídas de fuentes seleccionadas en acuerdo con expertos de dominio y usuarios.

A lo largo de toda la actividad de análisis los expertos de dominio participan aportando su conocimiento del dominio de interés y la aplicación, mientras que los ingenieros de conocimiento intervienen en el modelado conceptual con la utilización de UML.

Diseño

Tomando como entrada el glosario de referencia, esta actividad genera una estructura ontológica con el conjunto de términos del glosario, en base al modelo de la aplicación definido en la instancia de análisis. Como se expresa en [NMN09], se pasa de la dimensión lingüística a la dimensión conceptual. Para esta actividad se plantean las siguientes tareas:

- Modelar los conceptos

Cada concepto que surge del glosario de referencia es categorizado a través de la asignación de un *tipo*, de acuerdo a la clasificación OPAL ya mencionada. Para ello se realiza una correspondencia entre los elementos identificados en el modelo UML construido en la instancia de análisis y los conceptos del glosario de referencia. Se distinguen tres

tipos principales: actor de negocio, objeto de negocio y proceso de negocio, y dos complementarios: *mensaje* y *atributo*. En [DMT07] se puede encontrar una definición de cada una de estas categorías, que por cierto resultan bastante intuitivas.

A modo de ejemplo, si se parte de un glosario de referencia cuyos conceptos surgen de los términos identificados en el *Léxico de referencia* mencionado como ejemplo, al concepto *Tratamiento* se le asigna el tipo *objeto de negocio* y al concepto *Médico* el tipo *actor de negocio*, entre otros.

- Modelar jerarquías de conceptos y relaciones específicas de dominio.

Para construir la jerarquía de conceptos se identifican primero los conceptos principales, con un enfoque middle-out (detallado en la actividad de especificación de la metodología Methontology, sección 2.1.1), modelándose la estructura jerárquica a través de la generalización y especialización de los mismos. Además se agregan relaciones part-of y relaciones específicas de dominio. La estructura resultante con los tipos de relaciones mencionadas, se representa en UML como un diagrama de clases con las relaciones de generalización, agregación y asociación, respectivamente.

En la Figura 11 se puede ver un fragmento de la jerarquía construida con los conceptos mencionados en los ejemplos, representado en UML

como propone la metodología y en RDF(s), a los efectos de comparar la expresividad de ambas representaciones.

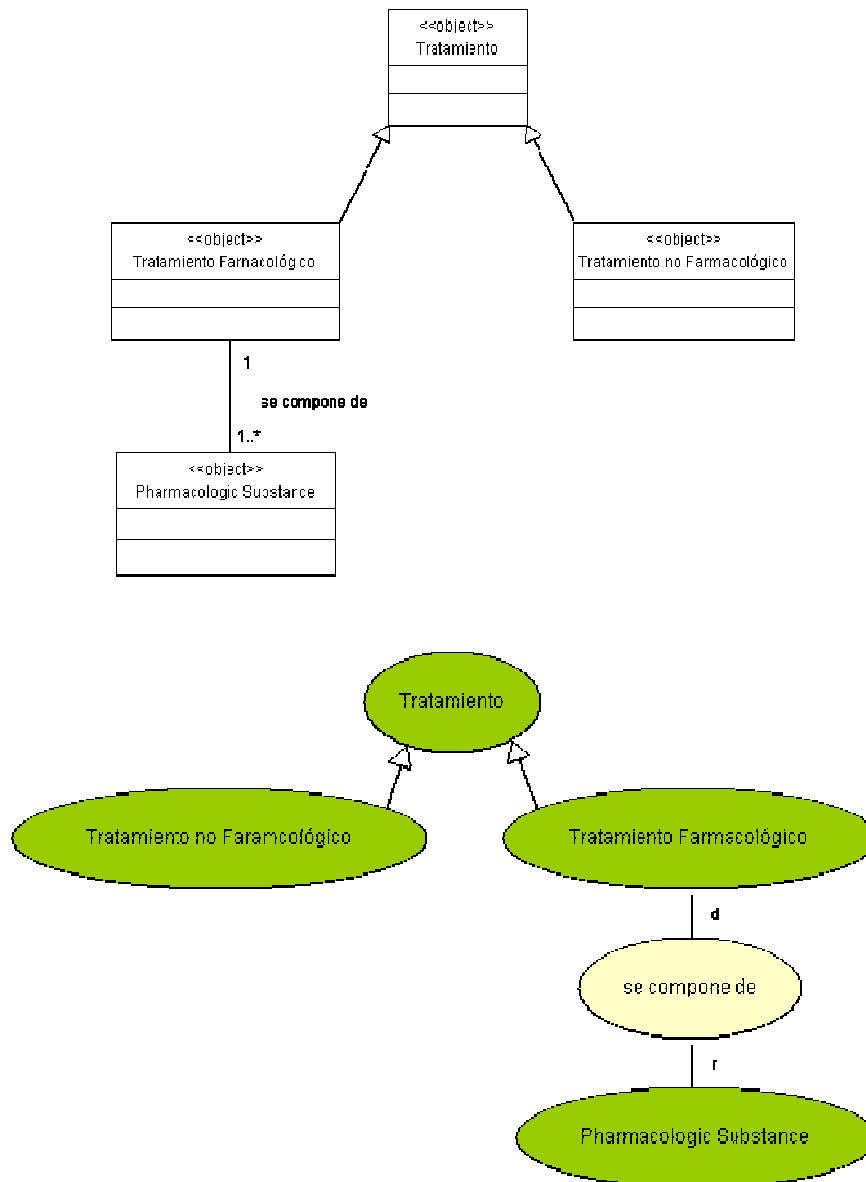


Figura 11. Fragmento de la estructura ontológica representado en UML y RDF(S).

La actividad de diseño concierne mayormente a los ingenieros de conocimiento, por su experiencia en modelado, mientras que los expertos de dominio intervienen en menor grado si es necesario algún aporte acerca del negocio.

Implementación

Esta actividad consiste en la codificación de la ontología en un lenguaje formal. [NMN09] da especial importancia a que el lenguaje seleccionado tenga suficiente nivel expresivo, cuente con un mecanismo de razonamiento efectivo y tenga aceptación en la comunidad. En esta instancia la metodología propone la definición de los axiomas que imponen las restricciones necesarias al modelo. [NMN09] propone OWL como lenguaje de implementación, por contar con los requerimientos mencionados [AH02].

Es claro que la actividad de implementación es realizada por los ingenieros de conocimiento.

Testeo

[NMN09] define cuatro características de una ontología que deben ser evaluadas: *calidad sintáctica*, *calidad semántica*, *calidad pragmática* y *calidad social*.

La calidad sintáctica es verificada en la instancia de implementación, ya que la ontología es representada en un lenguaje formal. La calidad social

puede ser evaluada recién después de la publicación de la ontología, al ser utilizada por expertos de dominio y otros usuarios.

En etapas tempranas es posible evaluar la calidad semántica y la calidad pragmática. Con la calidad semántica se verifica la consistencia de la ontología, es decir asegurar que no exista contradicción (existencia de ciclos en la jerarquía, clases disjuntas con instancias en común, etc.). Esta validación se realiza utilizando un razonador, siendo Pellet o Racer los que sugiere [NMN09]. Para verificar la calidad pragmática de una ontología deben tenerse en cuenta tres cualidades: *fidelidad*, *relevancia* y *completitud*. La primera significa que el conocimiento expresado e inferido por la ontología debe ser verdadero, lo que hace necesario verificar la veracidad de dicho conocimiento en las fuentes de información utilizadas para la especificación. Las cualidades de relevancia y completitud se validan en forma conjunta, constatándose la correcta implementación de los requerimientos de la ontología, para lo cual se verifica que el modelado de la aplicación representado por diagramas UML esté totalmente cubierto por el contenido de la ontología (conceptos, relaciones, etc.) y que el mismo responda a las preguntas de competencia.

La participación de ingenieros de conocimiento es clave para verificar la calidad semántica, por su conocimiento técnico de los razonadores, mientras que la calidad pragmática es verificada fundamentalmente por

expertos de dominio (consulta de las fuentes, asegurar la cobertura de requerimientos, etc.) con alguna intervención de ingenieros de conocimiento (por ejemplo en la interpretación de diagramas UML).

2.6.2. Conclusiones

El ciclo de vida iterativo e incremental, basado en casos de uso, en el que por cada ciclo se obtiene una nueva versión de la ontología, es también un ciclo de prototipos evolutivos.

Como apreciación general de la metodología hay tres aspectos a destacar:

- Es una metodología orientada al modelado de aplicaciones y más concretamente aplicaciones que involucran procesos. La existencia de procesos bien definidos, con actores también bien definidos, permite aprovechar al máximo el potencial de los artefactos UML. Por ejemplo, si se trata del diseño de una ontología de dominio, es posible definir casos de uso muy generales que modelen actividades que siempre tienen lugar en ese dominio [GFC03], pero va a ser el diagrama de clases el que más se adapte a la mayor parte de la conceptualización. En cambio, para una ontología orientada a una aplicación concreta con procesos definidos, el diagrama de actividad va a resultar más útil para la obtención del modelo.

- A pesar de ser una metodología propuesta recientemente, no tiene en cuenta la creciente tendencia al desarrollo distribuido y colaborativo de ontologías, no mencionando mecanismo alguno que ayude la coordinación de las necesidades de información de diferentes sectores de una organización.
- Se define claramente el rol de los ingenieros de conocimiento y expertos de dominio en cada una de las actividades.
- La metodología presenta una formalización cada una de las actividades propuestas, aspecto que no se da en ninguna de las restantes metodologías incluidas en el presente análisis.

A continuación se evalúan las actividades que forman parte del estándar.

Especificación y adquisición de conocimiento

Puesto que tres de las tareas propuestas (escenarios de uso, léxico de aplicación y preguntas de competencia) coinciden con las tareas propuestas en la metodología Methontology, se reiteran los mismos comentarios para la metodología UPON.

La identificación y priorización de casos de uso, resulta una herramienta efectiva para acotar el conjunto de términos del léxico de aplicación, ya que éstos van a surgir como relevantes en los procesos descritos por los casos de uso identificados. Como se explicó anteriormente, esto resulta más claro cuando se están diseñando ontologías orientadas a una

aplicación específica. Como la identificación de los casos de uso es una tarea que no da lugar a mucho margen de ambigüedad (en general no hay discusión en la ejecución de los pasos de un proceso), puede considerarse una técnica con una precisión aceptable.

Conceptualización y formalización

Abarca las actividades de análisis y diseño.

Una vez más haciendo referencia a Methontology, los diagramas UML estarían cumpliendo la misma función que las representaciones intermedias en Methontology, aunque con diferente enfoque. Ya se expresó que los diagramas UML proporcionan gran expresividad para la representación de procesos de negocio y en especial para indicar el orden de ejecución de los pasos en cada proceso. Si bien con las representaciones intermedias de Methontology, un paso de un proceso podría quizás ser modelado como una relación, no es posible indicar el orden de ejecución de los pasos de un proceso, ni mucho menos delimitar un proceso en una red semántica compleja.

Es en la actividad de conceptualización donde queda más al descubierto la falta de un mecanismo para el diseño colaborativo y distribuido.

Implementación

La recomendación de UPON en esta actividad se limita a sugerir a OWL como lenguaje de implementación, que permite restringir el modelo a través de axiomas y contar con un razonador como Pellet o Racer.

Evaluación

Se definen claramente chequeos a diferente nivel, que resultan ser una guía interesante a tener en cuenta en cada instancia de la metodología, aunque no se definen técnicas detalladas para llevar adelante cada una de las verificaciones propuestas.

Integración

Es una actividad que no está mencionada explícitamente por la metodología, quizás por su clara orientación a aplicaciones particulares. De todas maneras, la experiencia indica que siempre es bueno explorar si es posible diseñar algún fragmento de la ontología aprovechando modelos existentes.

2.7. Conclusiones finales

A continuación se presenta el mecanismo de comparación de las metodologías estudiadas y los resultados obtenidos.

2.7.1. Criterio de evaluación de metodologías

A los efectos de realizar un análisis medianamente objetivo acerca de las metodologías seleccionadas en este estudio, debe existir un marco de

referencia que permita realizar una evaluación consistente, analizando cada metodología de acuerdo a dicho criterio. Además de la bibliografía disponible de las seis metodologías estudiadas, se consultó literatura acerca de evaluación y comparación de metodologías, como [F99, GFC03]. Como resultado fue seleccionado un conjunto de aspectos o características de las metodologías de diseño de ontologías, que por su relevancia, merecen formar parte de este criterio de comparación.

La bibliografía consultada acerca de comparación de metodologías, [F99, GFC03], proporciona un resumen para un conjunto de metodologías, también a partir de una serie de aspectos seleccionados, en el que se indica para cada metodología, de qué forma está contemplado cada uno de estos aspectos. Además se ordenan las metodologías consideradas de acuerdo al grado de cumplimiento general de todos los aspectos considerados relevantes. La evaluación que se presenta a continuación difiere de las evaluaciones de la bibliografía consultada en los siguientes puntos:

- Se considera un espectro más amplio de metodologías con respecto al enfoque de las mismas. La presente comparación tiene en cuenta metodologías con enfoque centralizado y distribuido, así como metodologías orientadas a la construcción de ontologías de dominio y a ontologías que modelan procesos de negocio. Además se incluyen metodologías más recientes, como DILIGENT, HCOME y UPON.

Además, algunas metodologías, como Enterprise Ontology [UKMZ97] y TOVE [GF95], son consideradas en la presente evaluación como parte de Methontology (ya que esta metodología se apoya en aquéllas), mientras que en [F99, GFC03] se evalúan las tres metodologías en forma separada.

- Para cada aspecto identificado como relevante en el criterio de evaluación, en este trabajo se ordenan las metodologías de acuerdo al grado de cumplimiento de cada uno de ellos por separado, mientras que en la bibliografía consultada se establece un orden de cumplimiento general para todos los puntos evaluados, como ya se mencionó. Se adoptó esta modalidad de trabajo por considerarse que a la hora de proponer una recomendación de metodología más adecuada para el diseño y desarrollo de ontologías, surge casi directamente qué metodología(s) debe(n) tomarse como referencia en cada actividad propuesta, siendo de ayuda las justificaciones del orden asignado a cada una en cada aspecto relevante.

En cuanto a la forma de evaluación de cada uno de los aspectos del criterio de evaluación, se procedió de la siguiente manera:

- Para evaluar los aspectos más generales del criterio de evaluación, se realiza un análisis de alto nivel del enfoque general de cada metodología y de los elementos que ellas proponen. Por ejemplo, para determinar el ciclo de vida, se analiza si en la literatura disponible se

especifica algún orden en la ejecución de las actividades, si es posible retomar el proceso a partir de una actividad previa, etc. Para evaluar el cumplimiento del Estándar para Procesos de Ciclo de Vida de Desarrollo de Software de la IEEE, las actividades propuestas por cada metodología son comparadas con dicho marco de referencia.

- Los aspectos del criterio de evaluación que implican un análisis de más bajo nivel fueron evaluados a través de la aplicación de cada metodología a casos de uso identificados en el caso de estudio del proyecto SALUS-CYTED, que se utiliza en el presente trabajo.

2.7.2. Resultados de la aplicación del criterio de evaluación

A continuación se describe cada uno de los aspectos que componen el criterio de evaluación y se presentan los resultados de la comparación de las seis metodologías para cada uno de ellos. Las metodologías son ordenadas de acuerdo a la fortaleza que revisten en cada uno de estos aspectos, agregándose un comentario que argumenta el orden asignado, que se desprende de la evaluación individual de cada metodología, presentada anteriormente.

Aplicación del Estándar para Procesos de Ciclo de Vida de Desarrollo de Software de la IEEE.

Desde una perspectiva de Ingeniería de Software, el Estándar para Procesos de Ciclo de Vida de Desarrollo de Software de la IEEE propone

una serie de actividades a ser ejecutadas a lo largo del ciclo de vida de todo desarrollo. Como se explica en [F99], las ontologías, por ser productos de software, a nivel general deberían ser desarrolladas de acuerdo a dicho estándar, adaptado si es necesario, de acuerdo a las características especiales de las ontologías. El cumplimiento de este estándar es uno de los principales aspectos que componen el marco para la presente evaluación.

Las actividades propuestas por el estándar se agrupan en: actividades de administración, desarrollo y soporte. Las actividades de administración son pasos que deben cumplirse obligatoriamente para que todo proyecto llegue a buen fin (planificar, controlar, etc.), por lo que no se van a tener en cuenta en esta evaluación, ya que no tienen más peso en el desarrollo de ontologías que en cualquier otro producto de software. De las actividades de desarrollo y soporte, para realizar esta evaluación se han seleccionado algunas de ellas como de mayor relevancia para un proyecto de diseño y desarrollo de ontologías. Las actividades seleccionadas, que coinciden con las actividades tenidas en cuenta en la evaluación individual de cada metodología son: *adquisición de conocimiento, especificación, conceptualización, formalización, implementación, evaluación e integración*. Documentación y administración de configuración son actividades con el mismo grado de relevancia en todo proyecto de desarrollo de software, por lo que no se consideran en esta evaluación. La actividad de *mantenimiento* fue

evaluada individualmente en las metodologías que contemplan el mantenimiento evolutivo de las ontologías, siendo un aspecto importante que se tiene en cuenta en la comparación.

Para medir la aplicación del estándar en cada metodología, inicialmente se mide su fortaleza en cada una de las actividades seleccionadas. En base a estas evaluaciones individuales, se evalúa el grado de cobertura general del estándar. La Tabla 1 muestra el resultado de la evaluación de este aspecto.

Estrategia de construcción de la ontología.

Este aspecto de la evaluación de metodologías, tomado de [F99, GFC03], intenta responder si una metodología es más adecuada para la construcción de ontologías dependientes de una aplicación o independientes de una aplicación. Las metodologías más adecuadas para construir ontologías independientes de una aplicación se consideran más robustas, ya que también pueden aplicarse para construir ontologías dependientes de una aplicación, aunque no de forma óptima. La Tabla 2 muestra que ninguna de las metodologías es suficientemente flexible para responder de igual forma en la construcción de ontologías independientes y dependientes de una aplicación específica.

Tabla 1. Aplicación del Estándar para Procesos de Ciclo de Vida de Desarrollo de Software de la IEEE.

Actividad	Metodologías	Grado de desempeño en cada actividad
Especificación y adquisición de conocimiento	Methontology, UPON	Proponen varias alternativas, apoyándose en trabajos de otros autores, sin detallar en forma precisa.
	On-To-Knowledge	No se profundiza, pero sugiere algunas técnicas.
	SENSUS, DILIGENT, HCOME	Apenas se mencionan técnicas en SENSUS-based method, en DILIGENT y HCOME se ataca para la determinación consensuada del alcance.
Conceptualización y formalización	Methontology, UPON	Proporcionan una guía detallada de pasos para obtener el modelo, y en el caso de Methontology restricciones y reglas.
	DILIGENT	Se profundiza en el acuerdo de un modelo consensuado entre los actores involucrados, detallándose un método por el cual se toman las principales decisiones de diseño.
	HCOME	Define un mecanismo para obtener un modelo consensuado, que carece de control central.
	On-To-Knowledge, SENSUS	No se sugieren técnicas específicas.
Implementación	SENSUS, HCOME	Proponen una herramienta específica.
	On-To-Knowledge, Methontology, UPON	Apenas se dan algunas sugerencias.
	DILIGENT	Solamente se menciona.
Evaluación	Methontology	Se detallan gran cantidad de validaciones y verificaciones a realizar por diferentes actores.
	UPON	Define chequeos en diferentes momentos.
	On-To-Knowledge	Hace hincapié en la actividad, sin detallarla.
	DILIGENT, SENSUS, HCOME	Prácticamente no se trata.
Integración	Methontology	Especifica diferentes formas de integración y ejemplos de aplicación.
	SENSUS, DILIGENT, HCOME	Está implícita en la metodología (por la estructura común en SENSUS-based method y la ontología compartida en DILIGENT y HCOME).
	On-To-Knowledge, UPON	La actividad no es tratada.
Mantenimiento evolutivo	DILIGENT, HCOME	Definen ciclos para introducir cambios por parte de los usuarios, facilitándose la solicitud y/o ingreso de los mismos por los usuarios.
	Methontology, On-To-Knowledge, UPON	El ciclo de vida de prototipos evolutivos da soporte a esta actividad.
	SENSUS	No se menciona explícitamente la actividad.
Resultado general de la aplicación del estándar en base a las actividades seleccionadas	Methontology	Aborda todas las actividades del estándar, algunas con mucha profundidad.
	On-To-Knowledge, UPON	Cubren casi todas las actividades.
	DILIGENT, HCOME	Atacan algunas actividades importantes (especificación y conceptualización), desde el punto de vista de la construcción colaborativa.
	SENSUS	Ataca sólo algunas actividades, con poca profundidad.

Tabla 2. Estrategia de construcción de la ontología.

<i>Application-independent</i>	Methontology, DILIGENT, HCOME, SENSUS-based method	No están condicionadas por las aplicaciones.
<i>Application-semi-dependent</i>	On-To-Knowledge	Definida como dependiente, pero sus actividades aplican a ontologías no tan dependientes de la aplicación.
<i>Application-dependent</i>	UPON	La adaptación de la metodología UP al diseño de ontologías determina esta dependencia.

Existencia de un ciclo de vida, explícito o implícito.

En este aspecto se evalúa si la metodología propone algún ciclo de vida explícitamente, implícitamente (que se puede inferir de las actividades) o no está propuesto ningún ciclo de vida. Este aspecto es considerado también en [F99, GFC03].

Nivel de especificación de las actividades.

El grado de detalle con que se presenta cada actividad en la bibliografía consultada es un aspecto sugerido también en [F99]. En la presente evaluación también se tiene en cuenta que se propongan mecanismos no ambiguos con pasos bien definidos, para cada actividad.

Diseño distribuido y colaborativo.

Del la bibliografía consultada [VPTS05, TPSS05, KV06] se desprende que la modalidad de construcción de ontologías ha evolucionado en el tiempo, por lo que las metodologías más antiguas proponen un enfoque centralizado, mientras que las más actuales, un enfoque distribuido y

colaborativo. En el primer caso la construcción de ontologías es llevada a cabo por los ingenieros especializados en ontologías, obteniendo información y propuestas de los usuarios finales o expertos. En el segundo caso, si bien los ingenieros son los que en última instancia definen el modelo inicial y sus modificaciones, los usuarios que explotan la información estructurada por ontologías intervienen de forma directa realizando propuestas e incluso modificaciones al modelo dentro de ciertos límites y niveles de abstracción. Todo esto además puede realizarse en forma descentralizada, interviniendo equipos de trabajo de diferentes lugares. Dado que la tendencia actual se aproxima cada vez más a la segunda alternativa, en este estudio se considera relevante que las metodologías identifiquen métodos o procedimientos para hacer viable esta forma de trabajo. Si bien este aspecto se menciona en [F99], no se aplica a la evaluación de metodologías.

Nivel de aceptación de metodologías.

Se evalúa si se construyeron ontologías con cada una de las metodologías, y fueron utilizadas en aplicaciones [F99, GFC03, SSSS00, PST04, JB06].

La Tabla 3 muestra la evaluación de los últimos aspectos considerados (existencia de un ciclo de vida, nivel de especificación de las actividades, diseño distribuido y colaborativo, nivel de aceptación de la metodología). Es claro que Methontology y UPON son las metodologías con mayor nivel

de detalle en las actividades propuestas. Su punto más débil radica en no contemplar ambientes de trabajo distribuidos y colaborativos. DILIGENT y HCOME resultan mejor calificadas en este aspecto.

Tabla 3. Aspectos considerados para evaluar metodologías.

Característica	Metodologías	Grado de cumplimiento de cada característica
Existencia de un ciclo de vida	Methontology, UPON	Está definido explícitamente.
	On-To-Knowledge, DILIGENT, HCOME	El ciclo de vida no está explícito, pero se infiere de la definición de las metodologías.
	SENSUS-based method	No está propuesto y no puede inferirse.
Nivel de especificación de las actividades.	Methontology, UPON	Algunas actividades están muy detalladas y algunas presentan ambigüedades.
	DILIGENT, HCOME	Algunas actividades se definen con cierto grado de detalle.
	On-To-Knowledge, SENSUS-based method	Las actividades están definidas en forma superficial y ambigua.
Diseño distribuido y colaborativo.	DILIGENT	Sus actividades están orientadas a este objetivo.
	HCOME	Sus actividades están orientadas a este objetivo pero pueden resultar ineficientes para muchos actores, a la hora de alcanzar un consenso.
	Methontology, On-To-Knowledge, SENSUS-based method, UPON	No tienen en cuenta esta forma de trabajo.
Nivel de aceptación de la metodología.	Methontology	Es una de las metodologías más utilizadas en diversos dominios (química, medioambiente, etc.).
	On-To-Knowledge	Aplicada en portales, organizaciones virtuales, etc.
	DILIGENT, UPON y SENSUS-based method	Poco utilizadas, las dos primeras seguramente por ser más recientes.
	HCOME	No se encontraron referencias de utilización fuera del ambiente académico en que fue creada la metodología.

Una visualización general de este análisis comparativo de metodologías permite concluir que ninguna de ellas contempla todos los aspectos del criterio de evaluación, si bien Methontology es la que resulta mejor evaluada en la mayoría de los aspectos. Como factor común a todas se

observa que ninguna de ellas aborda en detalle las técnicas de adquisición de conocimiento a ser aplicadas en las actividades de especificación y conceptualización.

3. Recomendación de metodología de diseño y desarrollo de ontologías: Conceptual Ontology Design

A lo largo de la sección anterior se realizó el análisis y evaluación de seis metodologías de diseño y desarrollo de ontologías, que cubren una amplia gama de modalidades de trabajo, cuyas diferencias responden fundamentalmente al grado de madurez de la disciplina de Ingeniería de Ontologías del momento en que fueron publicadas. Con este estudio se trata visualizar de qué forma esta disciplina ha ido evolucionando, a la vez de rescatar de cada una de las metodologías, argumentación mediante, las mejores prácticas identificadas en cada una de ellas.

En esta sección se dan los primeros pasos hacia una nueva metodología de diseño de ontologías que aborda con mayor profundidad la actividad de adquisición de conocimiento, con un ciclo de vida que propone además un mecanismo de gestión del mantenimiento evolutivo con un enfoque distribuido. Las directrices que se proponen en este trabajo combinan todos los aspectos identificados como relevantes en el marco de referencia de comparación de metodologías e incluyen también aportes de diversas publicaciones, en especial de [SAAHSVW00, SE08, ED08]. [SE08] y [ED08] presentan un estudio de cómo aspectos psicológicos y cognitivos del experto consultado tienen incidencia en la conceptualización de una ontología, mientras que [SAAHSVW00] profundiza en el análisis de las técnicas a aplicar para llevar a cabo la

actividad de adquisición de conocimiento, de acuerdo al perfil del experto que provee su conocimiento de la realidad.

La presente propuesta de metodología de diseño de ontologías, que se ha denominado Conceptual Ontology Design, está basada en un ciclo de vida de Prototipos Evolutivos. Es un ciclo de vida tradicional de ingeniería de software, que sugiere la ejecución de las actividades de desarrollo: Especificación, Conceptualización e Implementación de forma iterativa, y de las actividades de soporte: Adquisición de Conocimiento, Integración y Evaluación, en la medida que lo requieren las actividades de desarrollo, con diferente intensidad a lo largo del proceso.

La Figura 12 describe las actividades de desarrollo con la posible interacción entre ellas, y las actividades de soporte como círculos concéntricos envolviendo a las primeras, con diferente intensidad. La metodología propone dos fases bien diferenciadas para la ejecución de las actividades de desarrollo: *Construcción de Ontología Inicial* y *Mantenimiento y Evolución*. La fase de Construcción de Ontología Inicial se ilustra en la figura a través de un círculo ubicado a la izquierda, a partir del “Mundo real”. En dicho círculo están representadas las actividades de esta fase, llevadas adelante por ingenieros de ontologías y expertos de dominio, para la construcción de una primera versión de una ontología que representa un primer modelo consensuado entre todos los involucrados, es decir que es una ontología de carácter general. Esta

primera versión es la Ontología General₀, que en la figura es la Ontología General_i con $i = 0$.

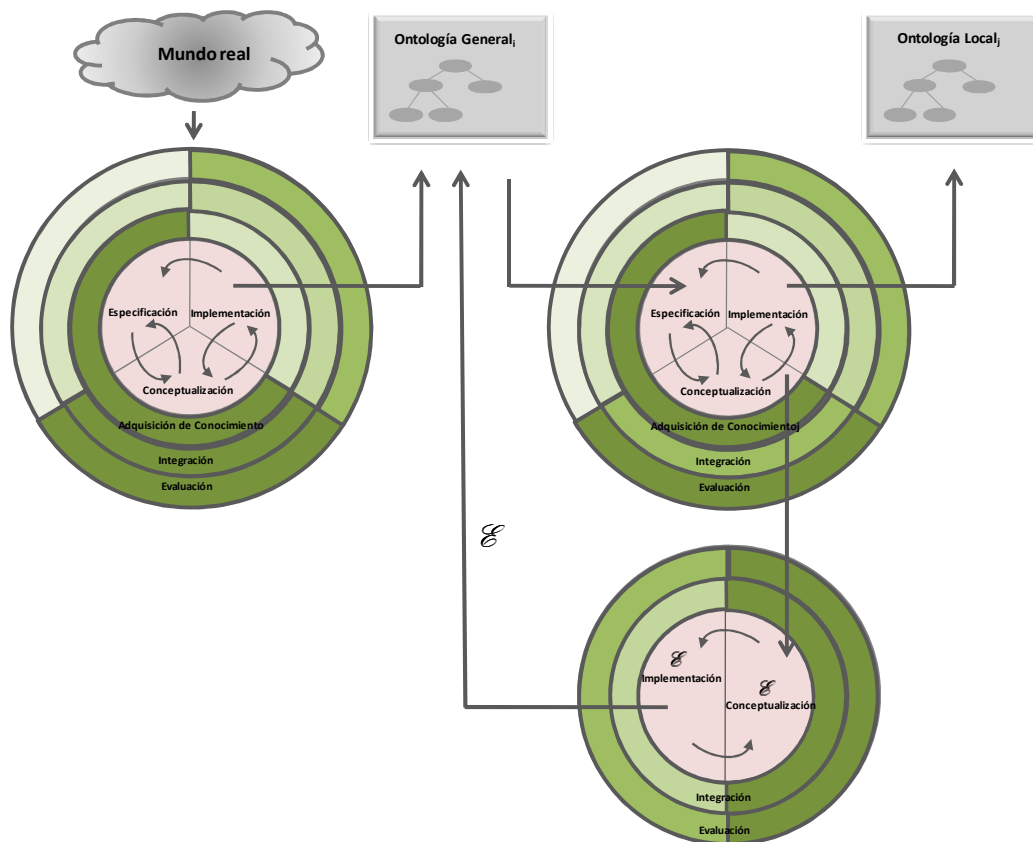


Figura 12. Ciclo de vida y actividades propuestas por Conceptual Ontology Design.

En la fase de Mantenimiento y Evolución, ingenieros de ontologías y expertos de dominio trabajan para adaptar ontologías locales, de acuerdo a necesidades que surgen de la realidad de su contexto. Las ontologías locales tienen como núcleo conceptual la ontología general, siendo modificaciones y extensiones de ese modelo núcleo. Las actividades para llevar adelante esta adaptación de las ontologías locales están

representadas en la Figura 12 a través del círculo ubicado a la derecha en la parte superior, partiendo de la Ontología General_i, lo que ilustra que el modelado de estas adaptaciones se basa en dicho núcleo conceptual. Periódicamente, el grupo de ingenieros de ontologías y expertos de dominio que trabajan en forma centralizada, debe actualizar la ontología general de acuerdo a las adaptaciones locales. El círculo ubicado a la derecha y abajo representa esta acción, partiendo del círculo correspondiente a la adaptación de ontologías locales. El resultado de estas actividades de actualización es una nueva versión de la ontología general, es decir Ontología General_j en la figura 12.

En resumen, la metodología propuesta es un proceso iterativo a través del cual se gestiona la evolución de la ontología general, respondiendo a cambios en la realidad.

3.1. Construcción de Ontología Inicial

La Construcción de Ontología Inicial tiene como objetivo principal obtener una ontología general inicial, que contemple todos los escenarios posibles, sin profundizar en requerimientos particulares a contextos locales. Esta ontología es tomada como punto de partida para la siguiente fase, Mantenimiento y Evolución. Es por ello que las actividades propuestas para esta fase están inspiradas, en parte, en las metodologías Methontology [FGJ97] y UPON [NMN09], ambas con un enfoque centralizado, que por otra parte resultaron mejor calificadas en el análisis

comparativo de la Sección 2. Aunque UPON no es tan general como Methontology, provee técnicas interesantes a ser aplicadas en el diseño de ontologías orientadas a una aplicación.

Como ya se mencionó, en esta propuesta y en particular en la fase de Construcción de Ontología Inicial, se da particular importancia a la actividad de adquisición de conocimiento, por ser ésta la base para una correcta abstracción de los principales conceptos de un modelo. Existen múltiples aportes respecto a las técnicas de elicitación de conocimiento disponibles, que tienen en cuenta cómo los expertos estructuran su conocimiento, pero no se ha encontrado una vinculación explícita de éstos con las actividades de especificación y conceptualización en la literatura disponible acerca de ingeniería de ontologías. Si bien ni el alcance ni el tiempo disponible para el presente trabajo permitieron profundizar en la gran variedad de investigaciones acerca de estas técnicas, este trabajo intenta al menos marcar un comienzo para que aspectos de la elicitación del conocimiento que tienen que ver con la forma de razonamiento y abstracción del experto sean tenidos en cuenta en la ingeniería de ontologías.

Por lo tanto, sin hacer una exposición de todas las técnicas de elicitación de conocimiento disponibles, este trabajo presenta una guía que orienta en la aplicación de las técnicas más adecuadas. De acuerdo a [ED08], las representaciones conceptuales humanas son flexibles y dinámicas,

mientras que las ontologías representan construcciones relativamente fijas. Esta flexibilidad está dada en parte por el contexto en el cual la actividad de adquisición de conocimiento es llevada a cabo, ya que como [SE08] expresa, las representaciones internas de los conceptos son el producto de procesos constructivos, dependientes del contexto. Otro factor importante en la adquisición de conocimiento es el perfil del experto de dominio, de acuerdo a su nivel de conocimiento (académico, experiencia en el dominio, etc.) [SAAHSWOO], ya que el nivel de conocimiento del dominio contribuye significativamente en la determinación del contenido, alcance y estructura de las ontologías de dominio [SE08]. Por lo tanto, en lo que tiene que ver con la actividad de adquisición de conocimiento, este trabajo tiene en cuenta fundamentalmente dos aspectos: el perfil del experto de dominio y el contexto en el que la ontología va a ser explotada.

En relación con el contexto, es importante tener en cuenta él o los usos finales que se va a dar a la ontología. Por ejemplo, si se construye una ontología para ser explotada por una aplicación con procesos de negocio bien definidos, seguramente el alcance de la ontología va a estar acotado concretamente por los términos relacionados con dichos procesos, por lo que para adquirir el conocimiento necesario probablemente sea suficiente con consultar a los usuarios clave en la ejecución de dichos procesos. Por el contrario, el alcance de una ontología de dominio que va a ser utilizada por diferentes aplicaciones, no va a resultar tan claro desde un principio y

es posible que sea necesario consultar expertos, fuentes escritas, etc. Como puede apreciarse, el contraste entre las ontologías construidas en cada caso está dado por el *nivel de generalidad* de la ontología, quedando claro que la ontología de dominio es más general que la ontología orientada a una aplicación específica. Existen diversas clasificaciones de ontologías de acuerdo a su nivel de generalidad, como la mencionada por Nicola Guarino en [G98], que define como categorías *top-level*, *domain*, *task* y *application* o los niveles *model-type*, *field*, *domain* y *case*, definidos por Ameen Abu-Hanna y Wouter Jansweijer en [AJ94], entre otras. A los efectos de mantener un equilibrio entre un enfoque tan simple que aplique las mismas técnicas de adquisición de conocimiento en la construcción de ontologías de cualquier nivel de generalidad y otro enfoque opuesto que discurra en una casuística engorrosa de acuerdo a diferentes clasificaciones que puedan encontrarse en las publicaciones existentes, a los efectos del presente trabajo se ha optado por categorizar las ontologías en: *orientadas a una aplicación* e *independientes de una aplicación*. Las técnicas de adquisición de conocimiento sugeridas en la metodología propuesta pueden aplicarse a cualquiera de estas dos categorías, especificándose algunas variantes que facilitan la construcción de ontologías orientadas a una aplicación, inspiradas en la metodología UPON [NMN09]. De esta forma se pretende evitar que la metodología propuesta caiga en alguna de las categorías: *application-dependent*, *application-semi-dependent* o

application-independent, definidas en la Sección 2.7.2, ya que la idea es producir una recomendación lo más general posible, pero que a su vez sea lo suficientemente flexible para optar por el camino más conveniente en cada caso.

En lo que sigue se detallan las actividades propuestas para esta primera fase.

3.1.1. Especificación y técnicas de adquisición de conocimiento aplicadas

La actividad de especificación se basa principalmente en el conocimiento provisto por el experto de dominio, mientras que el rol del ingeniero de ontologías radica en la extracción del conocimiento que es relevante desde la perspectiva del propósito de la ontología a construir. Su principal objetivo es la definición del propósito y el alcance de la ontología. El propósito se define como “el uso que se va a dar a la ontología, los escenarios de uso y los usuarios finales involucrados” [FGJ97]. El alcance es “el conjunto de términos a ser representados, sus características y granularidad” [FGJ97].

Considerando el caso de estudio del proyecto CYTED-SALUS, se presenta el primero de una serie de ejemplos que se numeran correlativamente a lo largo de la presente propuesta.

Ejemplo 1 - Propósito

“Evaluar y clasificar documentos del dominio médico de acuerdo a niveles de calidad previamente definidos, para emitir una recomendación de lectura a diferentes perfiles de usuarios”.

Para obtener el alcance, la actividad de especificación se apoya fuertemente en la actividad de adquisición de conocimiento, como se ilustra en la Figura 12. Como ya se mencionó, desde el comienzo es importante situar a los expertos de dominio en el contexto adecuado de acuerdo a los escenarios de motivación dados por el propósito de la ontología [ED08]. En caso que la ontología se construya para ser reutilizada en una amplia variedad de tareas y contextos, el ingeniero de ontologías debería llevar a cabo la actividad de adquisición de conocimiento en un ambiente neutral, en la medida de lo posible, aunque habrá que tener en cuenta que el experto siempre va a estar inmerso en su propio contexto particular.

A continuación se detallan los pasos a seguir en la especificación para obtener el alcance, especificando cómo se aplican las técnicas de adquisición de conocimiento.

Primera lista de términos del dominio y definición de áreas de trabajo

La tarea de obtener una primera lista de términos puede ser conducida a través de entrevistas no estructuradas y/o tormentas de ideas. Estas técnicas proveen un mecanismo adecuado al comienzo de la actividad de especificación, ya que son técnicas que no imponen restricciones al

experto, para abordar en su propio lenguaje los aspectos que considera más importantes, permitiendo al ingeniero adquirir una visión amplia del dominio [SAAHSWOO]. La elección de una u otra técnica va a depender del número de expertos a ser consultados. Si hay muchos expertos que aportan elementos al alcance de la ontología, la tormenta de ideas va a ser más provechosa [UKMZ97]. Para un grupo reducido de expertos de dominio, la entrevista directa resulta un método más práctico. Es claro que la información obtenida a través de la aplicación de estas técnicas puede no ser suficientemente precisa, pudiendo obviarse aspectos del dominio que el experto no considera importantes, pero que desde el punto de vista del modelo futuro lo son, etc. De todas formas, estas carencias son atacadas en los siguientes pasos propuestos por la metodología.

Dependiendo del perfil de los expertos consultados, puede ser necesario realizar una *consulta de fuentes documentales* para ampliar la lista de términos obtenida. Los expertos con perfil académico transmiten en general una visión del dominio que tiene una estructura lógicamente organizada. Ellos van a proporcionar información acerca de términos más generales del dominio, comunicando una visión más amplia. Tienden a creer que los problemas siempre pueden resolverse con la apropiada aplicación de la teoría, pero en algunos casos pueden estar lejos de la resolución de problemas del día a día. Los expertos no académicos, que adquieren el conocimiento del dominio a través de su experiencia diaria, brindan información más acotada. Su visión tiende a estar limitada a las

restricciones de los recursos con los que trabaja [SAAHSVW00]. Es en este último caso que puede resultar necesario consultar fuentes escritas, a los efectos de enmarcar el conocimiento adquirido dentro de principios más generales. Esta técnica de adquisición de conocimiento puede ser llevada a cabo en forma manual o en forma automática, utilizando por ejemplo técnicas de text mining, como se sugiere en [NMN09].

Una vez obtenidos, se propone asociar a cada término una descripción en lenguaje natural, para obtener un primer consenso con respecto a su significado e indicar qué términos son sinónimos. También se propone agrupar los términos obtenidos en diferentes *áreas de trabajo*. En cada grupo deberían incluirse términos con significado similar [UKMZ97] o relacionado, de manera que cada área de trabajo o grupo de términos tenga sentido independientemente de la existencia de otras áreas.

Ejemplo 2 - Lista de términos del dominio agrupados en áreas de trabajo

Términos identificados: Documento, evaluación de calidad de documentos, clasificación de documentos, enfermedad, síntoma, examen de laboratorio, sitio web, conferencia, revista, tratamiento, paciente, médico.

La Figura 13 muestra una primera agrupación de los términos en dos áreas de trabajo.

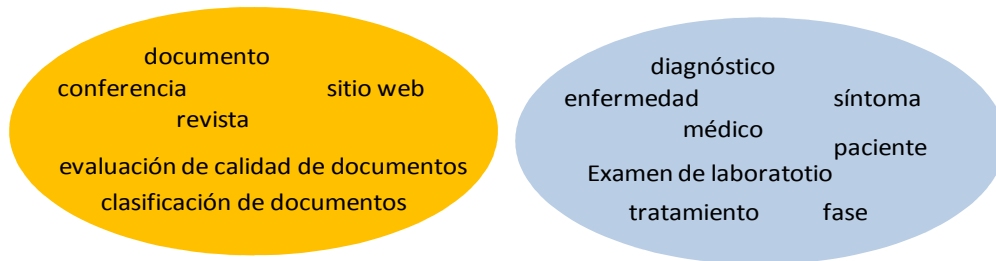


Figura 13. Lista de términos agrupados por área de trabajo.

Área Calidad: documento, sitio web, conferencia, revista, evaluación de calidad de documentos, clasificación de documentos.

Área Salud: enfermedad, diagnóstico, síntoma, examen de laboratorio, fase, tratamiento, paciente, médico.

Para ontologías *orientadas a una aplicación*, además de lo especificado anteriormente, se recomienda realizar la identificación de los casos de uso del negocio. Esta técnica va a facilitar la tarea de obtención de una lista de términos a los expertos de dominio o usuarios del negocio, ya que este conjunto de términos va a estar más acotado al contexto específico del proceso de negocio. Ellos deben entonces identificar los casos de uso correspondientes a los escenarios de motivación del negocio y realizar los diagramas de casos de uso (notación UML). Una vez identificados, los casos de uso pueden priorizarse [NMN09], para luego ir atacando los más críticos en primer lugar.

Identificación de conceptos principales

En este paso la actividad de especificación se mueve de la dimensión lingüística a la dimensión conceptual [NMN09].

De la lista de términos construida en el paso anterior, por cada área de trabajo identificada deben obtenerse los conceptos principales del modelo. Para ello se sugiere la aplicación de un enfoque middle-out, que como se explica en [UG96], implica identificar los conceptos representados por los términos obtenidos a un nivel de granularidad intermedio. En [ED08], un trabajo más reciente ya mencionado, se explica que las personas cuando hablan o piensan acerca del mundo tienden a hacerlo a un nivel básico de granularidad, buscando un equilibrio entre “distinción” e “información”. Esto significa que los conceptos principales deben poder expresar la máxima cantidad de información siempre que los mismos resulten fácilmente distinguibles unos de otros. Por ejemplo, en el caso de estudio relativo a la enfermedad de Alzheimer, el término “tratamiento farmacológico” resulta más distinguible (mayor distinción) que los términos “tacrina” o “acetilcolina”, que son tratamientos farmacológicos específicos. Sin embargo estos últimos expresan mayor cantidad de información que el primero (mayor información). Por lo tanto, la obtención de los conceptos principales o centrales, que deben expresar un equilibrio entre distinción e información, se realiza generalizando (se gana en distinción y se pierde en información) y especializando (se gana en información y se pierde en distinción), hasta alcanzar el equilibrio.

Dado que en este paso el proceso de especificación se sitúa en una dimensión conceptual, se sugiere refinar la agrupación en áreas de trabajo a nivel de conceptos, ya que en esta etapa se tiene una idea más clara del significado de los términos definidos en el paso anterior.

Con la identificación de los conceptos principales, se obtienen las “unidades básicas” de conocimiento, de acuerdo a un estudio de las estructuras de conocimiento de los expertos realizado por [GW06]. Posteriormente en la conceptualización se obtiene el “principio de organización” de estas estructuras básicas (relaciones entre conceptos) y el uso inferencial del conocimiento (axiomas y reglas).

Ejemplo 1 – Conceptos principales agrupados en áreas de trabajo

Se identifican los siguientes conceptos a partir de los términos del ejemplo anterior, con un refinamiento de las áreas de trabajo que se visualiza en la Figura 14.

Area Sitio: Documento, Sitio, Conferencia, Revista.

Area Calidad: Métrica (surge de los términos evaluación de calidad y clasificación de documentos), Nivel de Calidad (surge para representar el resultado de aplicar una métrica de calidad).

Area Salud: Enfermedad, Diagnóstico, Elemento de Diagnóstico (se agrupan síntoma y examen de laboratorio), Fase, Tratamiento.

Area Usuario: Paciente, Médico.

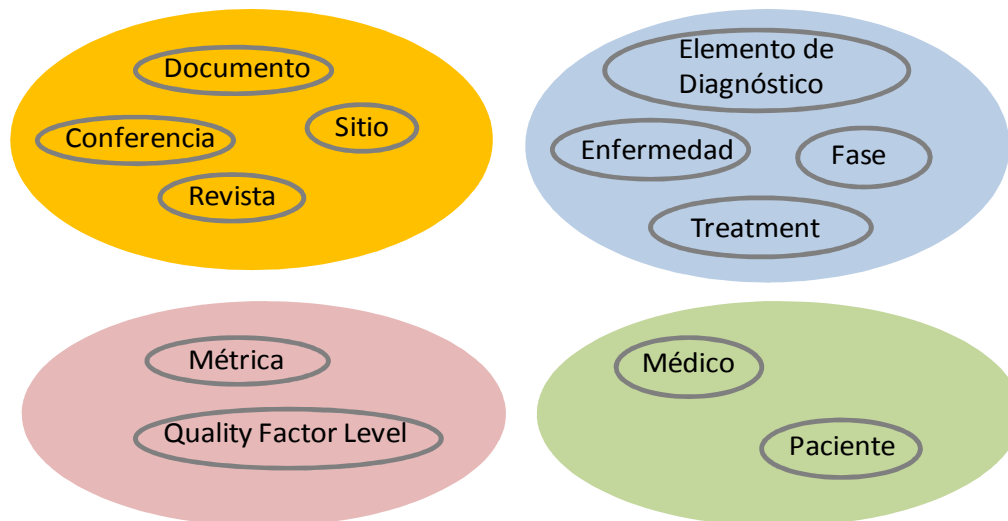


Figura 14. Agrupación de conceptos principales en diferentes áreas.

Para ontologías *orientadas a una aplicación* se recomienda también la construcción de diagramas UML y la clasificación de los términos de acuerdo a la clasificación OPAL. Previo a la aplicación del enfoque middle-out, se sugiere refinar el modelo de negocio para cada caso de uso, construyendo los diagramas de actividad y clases. La notación UML caracteriza cada elemento (a través de un estereotipo) como clase, actor o actividad. Esta clasificación coincide con las categorías OPAL [DMT07]. OPAL es una categorización creada para guiar a los expertos de negocio que intervienen en la construcción de ontologías, que proporciona un marco en el cual conceptualizar la realidad, claramente orientado a aplicaciones de negocio. Los términos obtenidos en el paso anterior

deben ser clasificados, teniendo en cuenta el modelo expresado en UML, de acuerdo a las categorías OPAL, de la siguiente manera:

- Los sustantivos o verbos que representan acciones, correspondientes a elementos del diagrama de actividad, se caracterizan como procesos de negocio.
- Los sustantivos que denotan agentes activos que ejecutan acciones, en general actores de los casos de uso, representan *actores de negocio*.
- Los sustantivos que no caen en ninguna de las categorías anteriores, sobre los que los actores ejecutan acciones, representan objetos de negocio. Estos términos se relacionan con elementos del diagrama de clases.

Con los términos ya categorizados, para cada área de trabajo, se obtienen los conceptos principales aplicando el enfoque middle-out, como se explicó. En general, para ontologías orientadas a una aplicación, los conceptos principales se obtienen de los términos relacionados con los diagramas de clases, que estructuran la información jerárquicamente. Debe tenerse en cuenta también que para este tipo de ontologías, la mayoría de los términos obtenidos en el paso anterior se convierten en conceptos principales, por estar involucrados directamente en el proceso de negocio representado a través de la notación UML. La agrupación de los términos de acuerdo a las categorías OPAL va a facilitar la

identificación de los elementos del modelo en la actividad de conceptualización.

3.1.2. Conceptualización apoyada en técnicas de adquisición de conocimiento

La actividad de conceptualización consiste en la estructuración y organización de los términos obtenidos en la actividad de especificación, partiendo de los conceptos principales identificados, para la obtención de un modelo que abstraer los aspectos que son relevantes desde el punto de vista del propósito de la ontología. El actor principal en la actividad de conceptualización es el ingeniero de ontologías, si bien no deja de intervenir el experto de dominio aportando el conocimiento necesario para que la definición del modelo responda a la realidad que se quiere representar.

Las tareas propuestas para la actividad de conceptualización coinciden con las especificadas por la metodología Methontology, que tiene en cuenta todos los elementos importantes para obtener un modelo que incluye axiomas y reglas de inferencia [FGV96, GFC03]. Es importante destacar [FGV96] que si bien el orden de las tareas propuesto es el que resulta más natural, la metodología permite que se retome el proceso a partir de cualquier tarea previa si es necesario, por lo que la conceptualización es un proceso iterativo en sí mismo. En cuanto a la forma de representación del modelo que se obtiene en la actividad de conceptualización, la presente propuesta admite, tanto las

representaciones intermedias propuestas en [FGV96, GFC03], como cualquier otra notación independiente de cualquier lenguaje de implementación, que brinde la expresividad requerida por el modelo. Al igual que la actividad de especificación, la conceptualización se apoya fuertemente en la actividad de adquisición de conocimiento, como se indica en la Figura 12. Las tareas propuestas para la actividad de conceptualización, tomadas de la metodologías Methontology, se detallan a continuación. También se agregan algunas contribuciones de otros trabajos con respecto a las técnicas de adquisición de conocimiento y algunos ejemplos del caso de estudio.

Glosario de términos

A partir de los conceptos principales y términos identificados en la actividad de especificación, si es necesario se amplía su descripción en lenguaje natural, se asocia a cada uno de ellos información adicional como identificación de acrónimos, detección de nuevos sinónimos y se comienza a categorizar los términos para determinar si corresponden a relaciones, conceptos, atributos, instancias, etc. Cada vez que se identifique algún nuevo elemento del modelo en alguno de los pasos posteriores en la conceptualización, éste debe ser agregado al glosario.

Taxonomías de conceptos

Esta tarea se inicia como una estructuración grosera de las relaciones jerárquicas del modelo, partiendo de los conceptos principales

identificados en la actividad de especificación. De acuerdo a [ED08], el conocimiento conceptual en los humanos no está organizado taxonómicamente, y los conceptos surgen como representaciones sensitivas al contexto, las que se articulan dinámicamente en la medida de lo necesario. En base a esta afirmación se argumenta que las representaciones conceptuales en la mente se estructuran más bien como jerarquías de pocos niveles o “llanas”, no como jerarquías profundas. Esto significa que en las primeras iteraciones de la conceptualización, la actividad de adquisición de conocimiento está orientada a la obtención de jerarquías básicas (de pocos niveles) que surgen de lo que en primera instancia el experto de dominio es capaz de transmitir. En iteraciones posteriores se profundiza en la especialización de los conceptos. A medida que se especializan o generalizan conceptos, se propone también que el ingeniero de ontologías y el experto de dominio determinen en qué se diferencian los conceptos hijos de un concepto padre dado, a los efectos de encontrar un atributo en este último con un peso tal que justifique su especialización. Este mecanismo se desprende de la técnica de adquisición de conocimiento “Repertory grids” especificada en [SAAHSVW00]. De esta forma se identifican atributos importantes de los conceptos principales.

Dependiendo del perfil del experto, en general cuando se trata de un experto no académico, puede existir conocimiento implícito al cual el experto no tiene acceso introspectivo directo de forma consciente

[SE08]. En estos casos, se sugieren algunas técnicas de elicitación de conocimiento tales como “repertory grids” y “laddering”, que por el alcance del presente trabajo no se detallan pero pueden ser consultadas en [SAAHSWOO].

Ejemplo 2 – Taxonomías de conceptos.

La Figura 15 muestra la generalización de los conceptos Sitio, Conferencia y Revista en el concepto Origen. También ilustra la especialización del concepto Elemento de Diagnóstico en los subconceptos Síntoma y Examen de Laboratorio, estructura que evoluciona en la identificación de los conceptos Evaluación Morfofuncional y Evaluación Cognitiva, que junto con Examen de Laboratorio se generalizan bajo el concepto Evaluación.

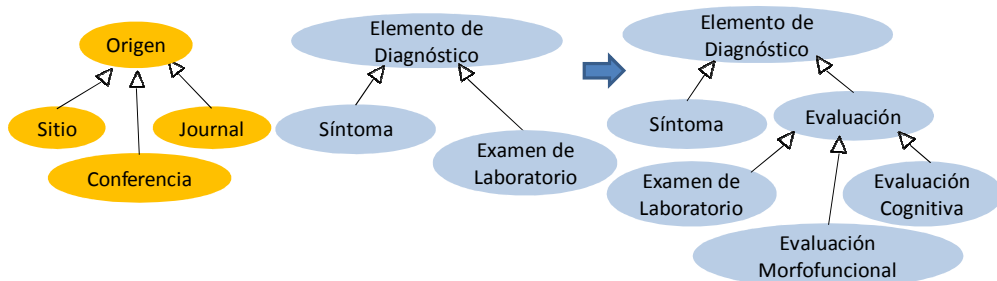


Figura 15. Taxonomías de conceptos para el caso de estudio CYTED-SALUS.

Diagramas de relaciones binarias

En esta tarea el ingeniero de ontologías debe guiar al experto de dominio en la obtención de relaciones muy claras y fuertes, no jerárquicas, que

aporten al propósito de la ontología a construir. Una vez más, el perfil del experto aparece como un factor importante, ya que según ha sido estudiado [SE08], los expertos académicos visualizan relaciones entre conceptos que los expertos no académicos no identifican en forma inmediata. Para facilitar la abstracción de estas relaciones, la técnica de adquisición de conocimiento recomendada es la entrevista estructurada, en especial para expertos no académicos. La presente propuesta admite diferentes formatos de entrevistas para identificar estas relaciones. A continuación se presenta una posible alternativa de entrevista estructurada.

Por cada par de conceptos identificados como relevantes en pasos anteriores, se realiza al experto de dominio el siguiente cuestionario [SAAHSVWOO]:

- Dados los conceptos A y B, ¿existe alguna relación entre ellos?
- Dado un concepto A que se relaciona con un concepto B a través de una relación r , ¿en qué forma A y B se relacionan a través de r ?

Esta pregunta apunta a caracterizar la relación desde el punto de vista semántico, a los efectos de dar un nombre y significado a la relación.

- Por cada instancia i_a del concepto A, ¿existe siempre al menos una instancia i_b del concepto B vinculada con i_a a través de r ?

Esta pregunta tiene como objetivo distinguir las relaciones obligatorias de aquellas que no lo son. [ED08] expresa que el experto de dominio a menudo tiende a caracterizar como obligatorias algunas relaciones que ocurren en la mayoría de los casos, pero no en todos. Si esto ocurre, podrían ser erróneamente definidos cuantificadores universales en axiomas de la ontología. Con la formulación de esta pregunta se requiere explícitamente que el experto de dominio se detenga a pensar si existe algún caso en el cual la relación no existe. También es posible identificar la cardinalidad de la relación, si la pregunta se extiende además para identificar si i_a está vinculada a más de una instancia i_b , a cuántas, etc.

- Si existen instancias i_a que no se relacionan con ninguna instancia i_b a través de r , ¿se vinculan alternativamente con instancias de otro concepto?

Con esta última pregunta, se descubren relaciones alternativas, no tan frecuentes, que no surgen desde el principio.

Ejemplo 3 – Diagrama de relaciones binarias

La Figura 16 muestra el modelo correspondiente al área de trabajo definida por los conceptos centrales Sitio, Conferencia, Revista y Documento, con la identificación de nuevos conceptos y relaciones. La relación “tiene Origen” se define como obligatoria ya que todo documento

va a tener un origen. La relación “tiene certificación” no es obligatoria, dado que un sitio puede o no estar certificado con un sello de calidad.

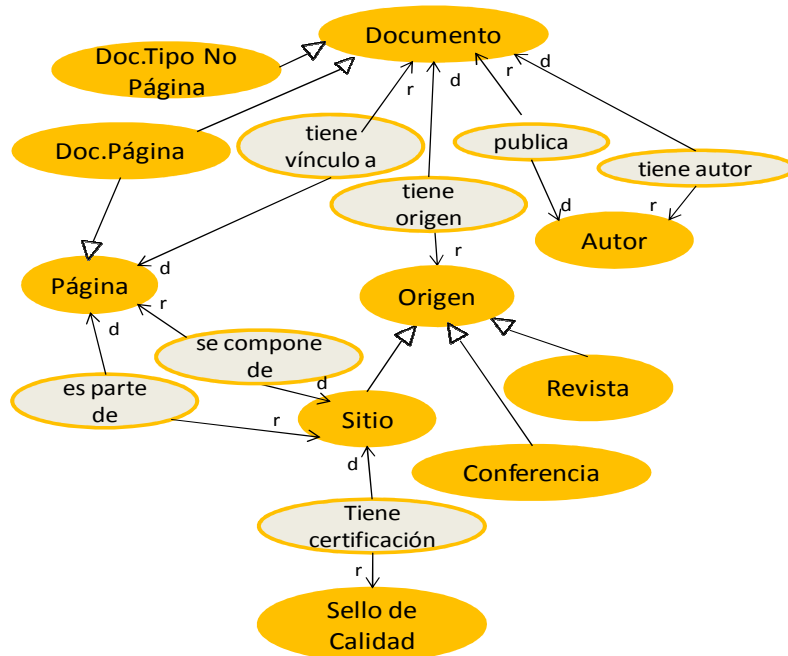


Figura 16. Modelo con relaciones binarias del caso de estudio CYTED-SALUS.

Para *ontologías orientadas a una aplicación*, se sugiere además tener en cuenta que algunas de las relaciones importantes que se identifiquen corresponden a conceptos caracterizados como procesos de negocio, donde un actor de negocio actúa sobre un objeto de negocio. En general, este proceso de negocio se debe modelar como una relación no jerárquica entre los conceptos identificados como actor y objeto de negocio. Pueden surgir además otras relaciones entre conceptos, que corresponden a los términos identificados a la documentación consultada. Ellas representan relaciones más generales en el dominio que el usuario

no visualiza por estar demasiado involucrado en la aplicación de negocio particular.

A partir de esta tarea se obtiene una estructura semántica o “principio de organización” [GW06] con las principales relaciones jerárquicas y no jerárquicas, la cual se refina paulatinamente en los pasos siguientes, agregándose atributos, axiomas, reglas, etc. como se detalla en [FGV96, GFC03]. A continuación se presenta una breve mención de algunas de las restantes tareas de la actividad de conceptualización, incluyéndose algunos ejemplos del caso de estudio.

Diccionario de conceptos

Ejemplo 4 – Entrada en diccionario de conceptos

Para el concepto Documento, la información asociada en el diccionario de conceptos [FGV96, GFC03] contiene:

Atributo de instancia: “tipo de documento”, es un atributo que se identifica para la especialización del concepto Documento, indicando si es una página o no.

Relaciones: “tiene origen” y “tiene Autor”.

Detalle de relaciones binarias

Ejemplo 5 – Entradas de detalle de relaciones binarias

Para la relación “tiene origen” que se muestra en la Figura 16, la entrada en el detalle de relaciones binarias contiene:

Concepto fuente: Documento

Cardinalidad fuente: 1

Concepto Destino: origen

Propiedad matemática: funcional

Para la relación “tiene Autor”:

Concepto fuente: Documento

Cardinalidad fuente: N (un documento puede tener más de un autor)

Concepto Destino: Autor

Propiedad matemática: no funcional

Detalle de atributos de instancia

Ejemplo 6 – Entradas de detalle de atributos de instancia.

La Figura 17 muestra el modelo de la Figura 16 con la identificación de algunos atributos de instancia. Para el atributo “tipo de documento”, la entrada en el detalle de atributos de instancia contiene:

Concepto: Documento

Tipo de valor: String

Cardinalidades mínima y máxima: (1, 1)

Para el atributo “largo promedio de oraciones”:

Concepto: Documento

Tipo de valor: Integer

Cardinalidades mínima y máxima: (1, 1)

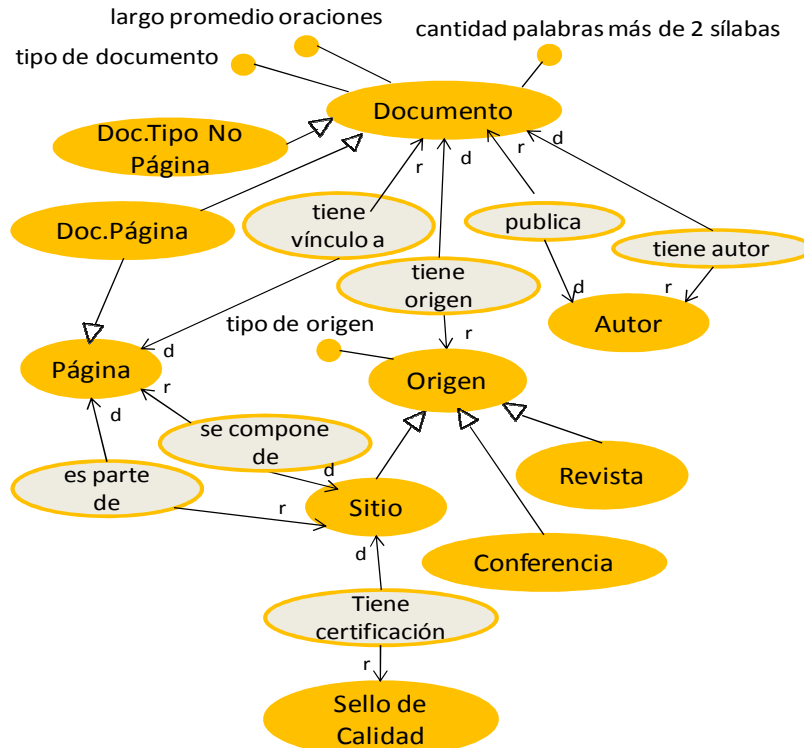


Figura 17. Modelo del caso de estudio CYTED-SALUS con atributos de instancia.

Definición formal de axiomas

Ejemplo 7 – Definición formal de axiomas.

Los axiomas son especificados en lógica de primer orden [GFC03]. De la Figura 17, el axioma que establece la condición necesaria y suficiente que deben cumplir las instancias del concepto Doc. Página, expresa que van a

ser las instancias del concepto Documento cuyo atributo tipo de documento tiene el valor "Página".

forall(?X)

([Documento](?X) and [tipo de documento](?X, "Página") →
[Doc.Página](?X))

Definición formal de reglas

Ejemplo 8 – Definición formal de reglas

La siguiente regla, aplicada al modelo de la Figura 17 expresa que si el origen de un documento de tipo "DocPagina" es un sitio, entonces el documento es de tipo "Pagina" y además forma parte de dicho sitio.

$\text{DocPagina}(?x) \wedge \text{Sitio}(?y) \wedge \text{tieneOrigen}(?x, ?y) \rightarrow \text{Pagina}(?x) \wedge \text{formaParteDe}(?x, ?y)$

Pueden aparecer dificultades en la identificación de axiomas y reglas cuando el conocimiento se adquiere de expertos con un perfil orientado a tareas de administración, que en general no dan información acerca de predicciones o hipótesis, es decir el "uso inferencial" [GW06]. En estos casos es posible agregar la siguiente pregunta a la entrevista estructurada sugerida para la identificación de relaciones:

- Dado un concepto A que se relaciona con un concepto B a través de una relación r, ¿en qué condiciones una instancia i_a de A se relaciona con una instancia i_b de B a través de r?

3.1.3. Integración

La actividad de integración consiste en la reutilización del conocimiento representado en una o más ontologías existentes, que resulte útil en la definición de un nuevo modelo. Como se muestra en la Figura 12, la actividad de integración es fundamentalmente una actividad de apoyo a la actividad de conceptualización, ya que es en el transcurso de esta última que puede tomarse la decisión de integrar con otros modelos. El ingeniero de ontologías es el actor más involucrado con la actividad de integración, ya que se requiere conocimiento técnico para llevar a cabo la elección de ontologías a integrar y la implementación de la solución integrada. De [BEFMSW06] y [GFC03] se toman las definiciones de los dos tipos de integración más comúnmente utilizados, que son *merge* y *mapping*. *Merge* es la creación de una nueva ontología a partir de una o más ontologías originales, capturando todo el conocimiento representado en ellas. *Mapping* (o *alignment* según [GFC03]) consiste en establecer correspondencias entre dos ontologías, preservando las ontologías originales.

Para argumentar la importancia de la integración durante la conceptualización vale la pena destacar que desde que se inicia la actividad de conceptualización y también en parte de la especificación, explícita o implícitamente existe una forma de integración en la que el ingeniero de ontologías reusa los modelos de una o más teorías de representación de conocimiento u ontologías genéricas. La consulta de

bibliografía publicada al respecto, como [GCG94], [GSG96], [GPS98] y [GWG06], permite afirmar que la elaboración de todo modelo debe estar basada en la teoría de representación de conocimiento general que más se adapta a la realidad que se quiere representar, reusándose los conceptos de las ontologías genéricas que conceptualizan los principios básicos de dichas teorías. Tales ontologías genéricas son denominadas ontologías fundacionales en [GWG06], destacándose la importancia de que las metodologías utilicen los principales “patrones de diseño” de estas ontologías como guía en la adquisición del conocimiento de los expertos de dominio. De esta forma el modelo resultante queda alineado con el diseño principal de la ontología fundacional [GWG06]. Para ejemplificar la argumentación anterior, se sugiere analizar los pasos propuestos para llevar a cabo la actividad de especificación. Al obtener los conceptos principales en el caso de ontologías orientadas a una aplicación, implícitamente existe una forma de integración. La propuesta sugiere que los términos identificados en los casos de uso se incluyan en alguna de las categorías OPAL: objeto de negocio, proceso de negocio o actor de negocio, que son patrones de diseño correspondientes a la teoría de participación. Las ontologías independientes de la aplicación, dependiendo del dominio, se enmarcarían en las teorías de ontología formal o meta-ontología.

Partiendo entonces de una o más teorías subyacentes como base de la ontología a construir, y teniendo claro el alcance de la misma (definido en

la actividad de especificación), es necesario evaluar la conveniencia de encarar el diseño conceptual reutilizando modelos de ontologías de dominio existentes, ya sea para construir todo o parte del modelo de la nueva ontología. Al describir la actividad de especificación se sugiere la agrupación de los conceptos principales en diferentes áreas de trabajo. La identificación de estas áreas es un punto de partida para realizar un estudio de ontologías existentes que puedan ser reutilizadas, si se adaptan a las necesidades de cada área.

De [GSG96, PM01, BEFMSW06, GFC03] se ha obtenido un amplio espectro de los métodos de integración existentes. La presente recomendación propone que, paralelamente a la ejecución de las tareas de conceptualización sugeridas, y especialmente al encarar las tareas: Taxonomías de conceptos, Diagramas de relaciones binarias y Detalle de relaciones binarias (que es cuando se definen los aspectos estructurales del modelo conceptual), se lleve a cabo la actividad de integración, guiada por los pasos que se detallan a continuación. Ellos pueden ser consultados en la literatura mencionada, para más detalle.

Identificación de ontologías a ser reutilizadas con definición de adaptaciones individuales y modalidad de integración al modelo.

El objetivo de este paso es la exploración de ontologías existentes, para seleccionar las que efectivamente van a ser integradas a la ontología a construir. De este estudio podría decidirse no aplicar integración alguna y

diseñar la ontología desde cero. En caso de seleccionarse una o más ontologías a integrar, debe definirse también si cada una de ellas por separado requiere de alguna adaptación previa, y en qué forma cada una va a ser integrada al modelo final. Lo más importante de todo este proceso es que se realiza *a nivel conceptual*, es decir que se integran modelos, no es a nivel de implementación. La importancia de integrar modelos radica en que se logra conceptualizar el modelo de la ontología final con una mejor comprensión de los modelos de las ontologías a reusar, que incluso podrían ser adaptados, antes de la implementación. Por el contrario, la integración a partir de una implementación, dificulta esta conceptualización por no estar tan explícitos los modelos de las ontologías a reusar. Un detalle pormenorizado las tareas de este paso puede encontrarse en [PM01]. A continuación se describen brevemente, profundizando sólo en aquellos puntos en los que este trabajo aporta contenido o se incluyen técnicas tomadas de otras publicaciones. También se agregan algunos ejemplos.

- *Identificación de ontologías candidatas*

Consiste en seleccionar posibles ontologías, exigiendo que las mismas cumplan algunas condiciones mínimas. Entre estas condiciones se exige que las ontologías candidatas representen al menos parte del dominio correspondiente al alcance de la ontología a construir y que los principales acuerdos ontológicos sean consistentes con los de ésta, así

como los conceptos principales. Obviamente la ontología a integrar debe estar disponible y su paradigma de representación de conocimiento (o teoría subyacente) debe coincidir también con el de la futura ontología. Son además aspectos importantes aunque no excluyentes, si las ontologías candidatas están disponibles a nivel conceptual, qué documentación puede encontrarse acerca de las ontologías, etc.

Ejemplo 9 – Ontologías candidatas a ser integradas en el caso de estudio.

Para construir una ontología genérica que contemple los requerimientos del área de trabajo Salud, existen ontologías ya implementadas como UMLS y GALEN. Para construir una ontología que contemple la especificación del módulo Calidad, cuyo propósito es evaluar la calidad de documentos, la conceptualización del área Sitio ilustrada en la Figura 17 es un posible modelo a ser reutilizado.

- *Obtener la información necesaria de las ontologías candidatas*

Si únicamente se tiene disponible la implementación de la ontología a reusar, es necesario realizar una reingeniería para obtener el modelo conceptual. En [GFC03] se propone el método Ontological re-engineering, para llevar a cabo esta tarea. En algunos sólo se cuenta con la documentación, en cuyo caso se evalúa si ella es suficiente para construir la ontología.

Ejemplo 10 – Información disponible acerca de UMLS.

UMLS provee información completa acerca del modelo, en forma gráfica y en lenguaje natural.

- *Evaluación técnica de las ontologías candidatas*

Entre los principales aspectos a evaluar es importante analizar si en las ontologías candidatas está representado el conocimiento requerido con una granularidad que corresponda al alcance definido en la especificación, detectándose posibles carencias o la existencia de conocimiento que es innecesario para el propósito de la ontología a construir. En esta tarea es importante evaluar el grado de cobertura de las ontologías candidatas respecto al alcance definido para cada área de trabajo. Otro aspecto a tener en cuenta es la estructura general de las ontologías, el criterio y clasificación de los conceptos y la calidad de las definiciones, como principales características.

Ejemplo 11 – Evaluación de la ontologías UMLS

La red semántica que organiza la terminología provista por UMLS brinda la granularidad requerida en el caso de estudio, y aún más. Por ejemplo, si se toma el concepto principal identificado Elemento de Diagnóstico, por una especialización se llega al concepto Examen de Laboratorio. UMLS proporciona, a partir del tipo semántico Laboratory Procedure, una mayor especialización de este concepto, por ejemplo

en: Anatomic pathology procedure e Histopathology test, con diferentes niveles de granuralidad.

- *Selección de las ontologías a integrar y la modalidad de integración.*

En base a la evaluación técnica se ponderan las ventajas y desventajas de las ontologías candidatas, para cubrir las necesidades de cada una de las áreas de trabajo identificadas en la actividad de especificación. Los aspectos más relevantes a tener en cuenta son: el grado de madurez de las ontologías (si está en desarrollo, terminada, evaluada, etc.), las características del desarrollo de la ontología (confiabilidad de las fuentes de conocimiento, políticas de mantenimiento, paradigma formal, mecanismos de inferencia, lenguaje), su contenido (completitud de los conceptos representados, grado de detalle, definiciones correctas, estructura adecuada, que permita agregar fácilmente conocimiento no representado). Este estudio va a tener como resultado la definición de adaptaciones a realizar a algunas ontologías, que aunque han resultado seleccionadas, necesitan cambios para ser integradas. Las tareas referentes a la evaluación técnica de ontologías candidatas se detallan al abordar específicamente la actividad de evaluación, en la sección siguiente.

Una vez definidas las posibles adaptaciones a cada una de las ontologías seleccionadas por separado, en este punto se cuenta, para cada área de trabajo, con una o más ontologías identificadas como

más adecuadas para la representación del modelo definido por el alcance de cada área. Como estas áreas de trabajo deben estar relacionadas entre sí, ya que forman parte de una conceptualización consensuada, el siguiente aspecto a considerar es que las ontologías seleccionadas sean compatibles en cuanto a las definiciones de conceptos comunes. En caso de existir incompatibilidades entre dos o más ontologías, deben establecerse mecanismos para resolverlas (unificar denominaciones de conceptos con la misma definición, unificar definiciones de conceptos semánticamente equivalentes, etc.), siempre que sea posible.

Para determinar la modalidad de integración a ser aplicada (*merge* o *mapping*), se sugiere aplicar el siguiente criterio:

- Si para alguna de las áreas de trabajo definidas en la actividad de especificación, existen varias ontologías seleccionadas cuyo contenido corresponde a los conceptos que deben ser representados en dicha área, de forma que el conocimiento representado en ellas es complementario y estructuralmente no puede mantenerse separado, es adecuado integrar aplicando *merge*.
- La aplicación de técnicas de *mapping* en cambio, va a resultar más adecuada para establecer las vinculaciones que existen entre las áreas de trabajo. Habrá porciones de conocimiento (conceptos, relaciones) en los modelos de las ontologías seleccionadas para

cada una de las áreas, cuyo significado va a coincidir, o al menos va a estar relacionado. Para obtener una conceptualización unificada de la ontología a construir, es necesario vincular de alguna manera estos modelos. El mapping es precisamente la definición de esas correspondencias. También puede aplicarse mapping entre dos ontologías para representar el modelo de una determinada área, cuando, sin necesidad de realizar un merge, cada ontología representa una porción bien diferenciada del modelo.

Si a partir del análisis anterior se decide aplicar integración, es necesario también determinar si existen partes del modelo final que deben construirse desde cero, y en este caso debe definirse claramente cuáles son estas porciones del modelo.

Ejemplo 12 – Selección de ontología a integrar.

Al conceptualizar un modelo para el área Calidad, se toma la decisión de aplicar mapping con el modelo que representa el área de trabajo Sitio (Figura 17), ya que es posible mantener los modelos totalmente independientes. La vinculación se define a través del concepto Documento.

Aplicación de adaptaciones a las ontologías a integrar y posible extensión.

De la evaluación técnica realizada en el paso anterior, surgen una serie de modificaciones a realizar a las ontologías. Entre ellas, puede ser necesario extender las ontologías con conocimiento no representado en ellas, por ejemplo a través de la especialización de conceptos, definición de nuevos conceptos y relaciones con conceptos existentes, etc.

Aplicación de las operaciones de integración.

Para realizar efectivamente las operaciones de integración, en caso de aplicarse técnicas de merge una de las metodologías existentes es ONIONS [GSG96]. Esta metodología hace hincapié en enmarcar las definiciones de las ontologías locales en teorías generales, a través de un método que busca la organización de diversas terminologías en un modelo integrado. Otros enfoques más recientes son FCA-Merge y PROMPT [GFC03]. El primero tiene la particularidad de tener en cuenta también las instancias de las ontologías. PROMPT propone la integración de las ontologías a un modelo de conocimiento común, basado en frames. Tiene la ventaja de ser un método dinámico en el que se van resolviendo conflictos paso a paso. Respecto al mapping entre ontologías, si bien [PM01] propone una metodología de integración a nivel general, está más orientado al mapping, es decir a mantener los modelos conceptuales separados. Aunque este trabajo maneja múltiples alternativas, sugiere como método generar la implementación de la ontología resultante a partir de las representaciones conceptuales. Otro enfoque menos completo

pero más reciente acerca de integración aplicando mapping es [BEFMSW06]. En este enfoque, si bien se requiere conocer los modelos conceptuales de las ontologías para establecer las correspondencias entre ellos, las ontologías se mantienen intactas y se define una ontología para representar las correspondencias, es decir una ontología de mapping. Existen algunos frameworks como MAFRA, RDFT y C-OWL para construir estas ontologías, en los cuales se utiliza el término “puente” para hacer referencia a estas correspondencias.

Ejemplo 13 – Integración aplicando mapping.

Como se ilustra en la Figura 18, al construir el modelo del área de trabajo Calidad, se reutilizan el concepto Documento, para asociarle un determinado nivel de calidad.

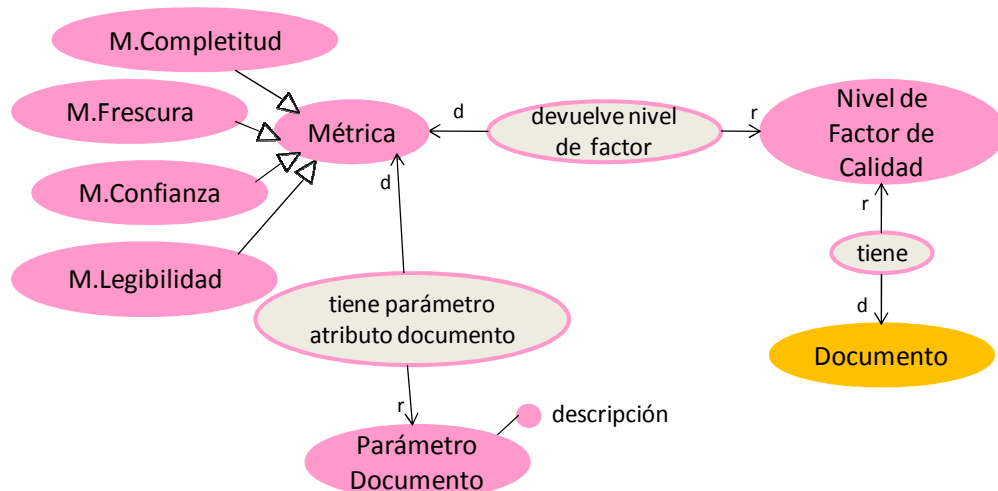


Figura 18. Integración del modelo Calidad con el modelo Sitio.

Es importante destacar [PM01], que si bien se da un orden en la ejecución de los pasos para llevar a cabo la integración, siempre es posible retomar el proceso a partir de un paso anterior. Esto significa que el proceso de integración en sí está guiado también por un ciclo de vida de prototipos evolutivos, dentro del ciclo de vida más general de la metodología propuesta en esta recomendación. También es fundamental contar con herramientas que faciliten la tarea de conceptualización conjuntamente con la integración.

3.1.4. Evaluación

La actividad de evaluación puede definirse como: un juicio técnico acerca del contenido de una ontología, con respecto a un marco de referencia, cuyo objetivo es asegurar que sus definiciones implementan correctamente la especificación de requerimientos y responden efectivamente a las necesidades del mundo real [GFC03, G99].

Esta definición sugiere dos elementos clave a ser considerados: en qué momentos realizar la evaluación y qué aspectos integran el marco de referencia o criterio de evaluación de ontologías. Respecto a cuándo debe llevarse a cabo la actividad de evaluación, en la bibliografía consultada, en general hay acuerdo en que debe aplicarse a lo largo de todo el ciclo de vida de la ontología [GFC03]. También existen publicaciones que definen momentos específicos para la evaluación, como [HSGMCSS05], que sugiere realizar esta actividad en las siguientes instancias: pre-

modelado, modelado y una vez liberada la ontología para su utilización. Aún cuando se reconoce la necesidad de llevar a cabo la actividad de evaluación a través de todo el ciclo de vida de la ontología, a los efectos de que esta recomendación adquiriera un carácter de guía de trabajo concreta, se definen dos momentos en los que la actividad de evaluación debe cobrar mayor importancia: *durante la conceptualización*, como se visualiza en la Figura 12, y *durante la utilización de la ontología a través de una aplicación*. Con relación al marco de referencia utilizado para la evaluación de una ontología, pueden encontrarse criterios muy variados en la bibliografía publicada. [G93] establece principios muy generales a tener en cuenta al tomar decisiones de diseño, es decir, al evaluar la ontología durante la conceptualización. [GJP95] agrupa el conjunto de tareas acerca de la construcción de ontologías (definiciones, estructura, software asociado, documentación) como actividades de verificación. A la comprobación de que la ontología represente efectivamente la realidad, la denomina validación, y define el término “assessment”, para referirse a la evaluación que debe realizarse al momento de reusar y compartir ontologías. Otro análisis bastante general pero más reciente, es realizado por [BGM05], que clasifica las técnicas de evaluación en diferentes “niveles”: léxico, jerarquía, otras relaciones semánticas, contexto, sintaxis y estructura. [LG04] presenta un método que abarca lo que para [GJP95] sería la verificación de ontologías, aunque mucho más exhaustivo. Más que una recomendación es una guía de trabajo que agrupa los aspectos a

tener en cuenta en “dimensiones” (contenido, lenguaje, metodología, herramientas y costos). [GCCL05] establece tres tipos de medidas para la evaluación de ontologías: medidas estructurales, medidas funcionales y medidas de usabilidad, en base a una meta-ontología que caracteriza las ontologías y a una ontología que modela los aspectos a tener en cuenta en la evaluación de ontologías. Es un enfoque general, que aborda la actividad desde diferentes perspectivas. Existen además estudios menos generales que encaran con más profundidad algunos de los aspectos mencionados, como [G99] y [GFC03], que profundizan en la detección de problemas de inconsistencia, incompletitud y redundancia en las relaciones taxonómicas. [BD05] y [FA08] proponen extensiones al trabajo anterior, agregando otras posibles fuentes de error en las clasificaciones taxonómicas, así como la identificación de anomalías en las estructuras conceptuales, tanto para relaciones taxonómicas como para otras relaciones entre conceptos. Un enfoque diferente es [WG01], cuyo análisis de “buena formación” de taxonomías, es realizado desde el punto de vista filosófico, en relación con las ontologías fundacionales o meta-ontologías.

A continuación se indican los aspectos a evaluar en cada uno de los dos momentos definidos. Estos elementos han sido tomados de la bibliografía consultada, aprovechando al máximo los conceptos vertidos por sus autores y combinándolos en un intento de presentar una guía de evaluación completa.

Evaluación durante la conceptualización

A lo largo de la actividad de conceptualización es fundamental asegurar, en primer lugar, que la ontología sea construida de forma correcta, y que además represente un modelo del mundo real para el cual fue creado. Estos conceptos corresponden a las definiciones de *verificación* y *validación*, respectivamente [GJP95, GFC03], y se adoptan en la presente recomendación. El actor principal en esta instancia es el ingeniero de ontologías, principalmente en las actividades de verificación, con la colaboración del experto de dominio que aporta su conocimiento de la realidad en las tareas de validación.

Para evaluar la correctitud de la ontología (verificación), y parte de la validación, se toma como guía el método Ontometric [LG04], que a su vez se basa en el trabajo de tesis doctoral [L02]. Este método agrupa los aspectos relevantes a tener en cuenta en la evaluación, en cinco dimensiones: *contenido*, *lenguaje*, *metodología*, *herramientas* y *costos*. Cada una de estas dimensiones está organizada jerárquicamente de la siguiente manera: cada dimensión está compuesta por factores (elementos fundamentales a ser analizados en las ontologías) y cada factor se desglosa en características (aspectos particulares enumerados con mayor detalle). Esta estructura jerárquica puede ser modificada al momento de realizar la evaluación, extendiéndola agregando nuevas características o eliminando aquellas que resulten innecesarias, de

acuerdo al propósito de la ontología. [L02] brinda una descripción detallada de las características a tener en cuenta en cada dimensión, las cuales son completamente reutilizadas como guía en la presente recomendación.

La dimensión *contenido* tiene que ver con el universo de discurso de la ontología, es decir los términos que incluye y su organización. En base a la consulta de diferentes trabajos presentada anteriormente, para esta propuesta se han identificado múltiples aportes de otras fuentes, que proponen chequeos, los cuales claramente deberían ser incluidos en esta dimensión. Aprovechando el mecanismo de extensión de características definido por Ontometric, la presente recomendación propone agregar estos chequeos, relevantes para la actividad de evaluación, a las características identificadas por Ontometric [LG04]. La Tabla 4 muestra las características de la dimensión contenido, unión de las ya existentes y las incluidas de otras fuentes, para cada uno de los factores que se definen en [LG04]: conceptos, relaciones, taxonomía de conceptos y axiomas. Las que aparecen con fondo claro, son las incluidas en este trabajo, que además presentan la correspondiente referencia a la fuente. Por ejemplo, en el factor conceptos se agrega la característica “Lazy Concepts”, para verificar si existe en la ontología algún concepto que nunca es utilizado. La decisión de borrar dicho concepto puede depender de que la ontología sea reutilizada por otra aplicación que podría hacer uso del concepto. De todas maneras debe realizarse este análisis. En el

factor taxonomía de conceptos se agrega el chequeo de circularidad, que resulta básico para evitar los ciclos en las taxonomías.

Respecto a los restantes factores que tiene en cuenta Ontometric (lenguaje, metodologías, herramientas y costos), se da una breve descripción de lo que se chequea en cada uno.

El *lenguaje* en el que se implementa una ontología, es otro de los aspectos a tener en cuenta en la actividad de evaluación. Fundamentalmente lo que debe medirse es el grado de expresividad del lenguaje, de modo que permita reflejar la granularidad requerida para el modelo e inferir nuevo conocimiento. Los factores que define Ontometric en esta dimensión son: representación del conocimiento y mecanismos de inferencia.

La *metodología* de diseño de ontologías que se aplique es un elemento muy importante a ser evaluado, puesto que si no se han definido en forma precisa pasos a seguir durante el ciclo de vida completo de la ontología, además de verse afectado el diseño inicial, van a surgir dificultades en el mantenimiento evolutivo de la ontología. Los factores identificados para esta dimensión son: precisión, usabilidad y madurez.

Tabla 4. Dimensión Contenido (Ontometric) y elementos de otras fuentes.

Dimensión	Contenido
Factor	Conceptos
Características	Los conceptos esenciales están en la ontología. Para ontologías dependientes de una aplicación, validar de que todos los casos de uso están cubiertos por conceptos esenciales. [NMN09]
	Los conceptos esenciales están en los niveles superiores de la ontología
	Los conceptos están descriptos convenientemente en lenguaje natural
	La especificación formal de los conceptos coincide con su descripción en lenguaje natural
	Los atributos describen de forma precisa a los conceptos
	El número de conceptos representados en la ontología
	Anomalía de diseño: "Lazy concepts" [BD05]
	Anomalía de diseño: "Property clumps" [BD05]
Factor	Relaciones
Características	Las relaciones esenciales están en la ontología
	Las relaciones relacionan los conceptos apropiados
	Las relaciones están descriptas convenientemente en lenguaje natural
	La aridad de las relaciones es la apropiada
	Las relaciones tienen especificadas sus propiedades formales
	El número de relaciones representadas en la ontología
	Extensión en errores de incompletitud: "Functional Property Omission for single valued property" [FA08]
	Extensión en errores de incompletitud: "Inverse-Functional Property Omission for a unique valued property" [FA08]
Factor	Taxonomía de conceptos
Característica	Errores de inconsistencia
Subcaracterísticas	Errores de circularidad [G99] [GFC03]
	Errores de partición: incluye "Errores de partición" [G99] [GFC03] y anomalía de diseño: "Lonely Disjoints" [BD05]
	Errores semánticos: incluye "Errores semánticos" [G99] [GFC03] y extensiones a los errores taxonómicos: "Weaker domain specified by subclass error" y "Domain breach specified by subclass error" [FA08]
Característica	Errores de incompletitud
Subcaracterísticas	Clasificación de conceptos incompleta [G99] [GFC03]
	Errores de partición (partición disjunta y exhaustiva) [G99] [GFC2003]
Característica	Errores de redundancia
Subcaracterísticas	Redundancia gramatical [G99] [GFC03]
	Definición formal idéntica de clases ([G99] [GFC03]
	Definición formal idéntica de instancias ([G99] [GFC03]
	Anomalía de diseño: "Chains of Inheritance" [BD05]
	Extensión a los errores taxonómicos: "Redundancy of Disjoint Relation error" [FA08]
Características	Errores lógicos: violaciones a las propiedades de OntoClean [WG01]
	Los conceptos son clasificados desde varias perspectivas
	La relación <i>no_subclase_de</i> es usada convenientemente
	La relación de partición disjunta es usada convenientemente
	La relación de partición exhaustiva es usada convenientemente
	La profundidad máxima en la jerarquía de conceptos
	La media de hijos por concepto
Factor	Axiomas
Características	Los axiomas se utilizan para hacer deducciones respondiendo a consultas. En esta característica tener en cuenta la extensión en errores de incompletitud: "Sufficient knowledge Omission Error" [FA08]
	Los axiomas se usan para completar valores en los atributos de instancia
	Los axiomas pueden utilizarse para verificar la consistencia
	Los axiomas se definen como elementos independientes en la ontología
	El número de axiomas de la ontología

En cuanto a las *herramientas*, las facilidades que brinda el entorno de desarrollo son sin duda un aspecto a tener en cuenta, siendo particularmente importantes las funcionalidades de visualización, modificación, chequeos de consistencia, integración con otras ontologías, etc. Los factores que componen esta dimensión son: prestación, visualización, edición, interacción, aspectos metodológicos, aspectos cooperativos, traducción e integración.

Finalmente, la dimensión *Costos* cobra importancia en caso que se utilice un software de índole comercial en el desarrollo de una ontología. Si se está evaluando una ontología existente para ser reutilizada, deben tenerse en cuenta los costos de utilización de la misma. Los factores definidos para esta dimensión son: adquisición de la ontología, recursos de uso, adquisición de módulos de acceso y adquisición de software de adaptación.

Con la especificación de un método basado en Ontometric y otros trabajos, se ha definido cómo llevar adelante la evaluación de ontologías, en todo lo que refiere a la *verificación*. Este método cubre también parte de lo que [GFC03] y [GJP95] definen como *validación* de ontologías. En la dimensión contenido, por ejemplo, algunas de las características definidas, como “La relación de partición disjunta es usada convenientemente para las necesidades del sistema”, correspondiente al factor “taxonomía”, se basan en aspectos del contenido en relación con

las necesidades del mundo real, que validan el significado de las definiciones de la ontología.

Para completar la *validación* de las ontologías, deben tenerse en cuenta los siguientes aspectos [GJP95]:

- El comportamiento del ambiente de software en el que la ontología se implementa debe ser adecuado para ejecutar las tareas definidas en la especificación de los requerimientos.
- El significado de la documentación de la ontología en lenguaje natural debe ser consistente con el significado de las definiciones de la ontología y con la documentación del ambiente de software.

Evaluación durante el uso de la ontología

Una vez implementada la ontología e integrada a una aplicación, es importante la obtención de una retroalimentación desde el punto de vista del usuario. Esta instancia de evaluación es referenciada como “assessment” en [GJP95]. Los principales actores que intervienen en esta evaluación son los ingenieros encargados del desarrollo de aplicaciones que utilizan ontologías. De [GJP95] se desprenden los siguientes aspectos a ser evaluados:

- Las definiciones y axiomas de la ontología deben cumplir con las siguientes propiedades: claridad, usabilidad, generalidad, granularidad,

calidad, portabilidad, extensibilidad, mantenibilidad y uniformidad, a los efectos de ser integradas fácilmente a la aplicación.

- El ambiente de software para integrar ontologías debe cumplir con las siguientes propiedades: robustez, precisión, portabilidad, extensibilidad y eficiencia computacional.
- La documentación en lenguaje natural acerca de cómo integrar las definiciones de la ontología a las aplicaciones debe ser clara.

[BGM05] aporta otro elemento a esta instancia de evaluación al definir uno de sus “niveles”, el “contexto”:

- Realizar un análisis de cómo los resultados de la aplicación son afectados por el uso de la ontología.

3.1.5. Implementación

En esta actividad, el modelo conceptual de la ontología expresado en un lenguaje formal, queda disponible para ser procesado en forma automática. En general, la implementación se lleva a cabo en forma conjunta con la conceptualización e integración. El actor principal en esta actividad es el ingeniero de ontologías, si bien el experto de dominio puede llevarla a cabo si se dispone de alguna herramienta que limite la porción del modelo a ser implementado por éste. La herramienta utilizada para la implementación debe cumplir con los siguientes requerimientos:

- Permitir representar el modelo de la ontología con la expresividad del lenguaje de representación de conocimiento elegido.
- A la vez de brindar la expresividad requerida, es deseable que la complejidad del lenguaje resulte transparente para el usuario.
- Debe contar con un mecanismo de inferencia de nuevo conocimiento.

3.2. Mantenimiento y evolución

Una vez obtenida la primera versión de la ontología, la fase de mantenimiento y evolución da soporte al mantenimiento evolutivo de la ontología, en un ambiente distribuido. Se define un mecanismo para introducir cambios a la ontología general desde contextos locales, inspirado en la metodología DILIGENT [VPTS05]. Como se observa en la Figura 12, a partir de cambios que se introducen en diferentes contextos locales, cada uno adaptando su versión Ontología Local_i, después de un análisis la ontología general absorbe los cambios que ameritan ser realizados a nivel general, de manera que en cada evolución \mathcal{E} se obtiene una nueva versión Ontología General_i. Este enfoque permite que los usuarios finales de la ontología, que en alguna medida explotan la semántica que ésta brinda y los datos que la pueblan, estén involucrados en el proceso de diseño, con ciertas restricciones. Esta fase se compone de dos etapas: *adaptación de ontologías locales* y *actualización de ontología central*. En caso que no existan ontologías locales, se simplifica

el mecanismo propuesto para manejar la evolución de una única ontología.

3.2.1. Adaptación de ontologías locales

En esta etapa intervienen principalmente ingenieros de ontologías y expertos de dominio, si bien los expertos van a ser los responsables de promover los cambios, por ser quienes perciben las transformaciones de la realidad de su contexto, para luego adaptar el modelo, con la intervención de los ingenieros. Como “expertos de dominio”, además de expertos en el tema a nivel académico, en este trabajo se abarca también a los usuarios que como ya se mencionó utilizan la ontología en su trabajo, independientemente de su nivel de conocimiento técnico. La idea es que intervengan todos los involucrados, es decir expertos con diferentes habilidades, que aportan conocimiento que se complementa, sin que ello signifique que todos ellos van a introducir cambios a la ontología, lo que importa es que promuevan cambios. También debe tenerse en cuenta la intervención de los ingenieros encargados de construir aplicaciones que hacen uso de las ontologías.

En la actividad de *especificación*, se va a modificar el alcance de la ontología local, que en la primera iteración de esta fase coincide con la ontología central y se van a cumplir los pasos del procedimiento detallado en la Sección 3.1.1, aplicando las técnicas sugeridas para la actividad de

adquisición de conocimiento. Al aplicar estas técnicas es importante que intervengan todos los actores mencionados.

La actividad de *conceptualización* se va a ejecutar también en base a los pasos ya detallados en la Sección 3.1.2, para extender y/o modificar una versión local de la ontología que en cada iteración se modifica de acuerdo a las necesidades locales. Esta actividad se intercala con la actividad de *implementación*. Además de la actividad de *adquisición de conocimiento*, en la conceptualización es fundamental para los ingenieros de ontologías apoyarse en la actividad de *Evaluación*, a los efectos de asegurar la correctitud del modelo local. En menor medida, podría ser necesario *integrar* con otros modelos, aunque es más factible que este tipo de decisiones sean tomadas en la actualización de la ontología general.

Tal como propone la metodología DILIGENT [VPTS05], para que diferentes grupos de trabajo puedan llevar adelante la adaptación de ontologías basadas en un mismo modelo general, contribuyendo además a la evolución de este último, resulta imprescindible contar con una herramienta que facilite esta modalidad de trabajo colaborativo y distribuido. Existen algunas herramientas, como HCONE, propuesta por la metodología HCOME y otra herramienta del proyecto SEKT. Desde la perspectiva de trabajo local, en la presente recomendación se propone que la herramienta de apoyo tenga las siguientes características:

- Permitir extender el modelo de la ontología general y modificar la porción del modelo construido a nivel local. Es importante además proveer un espacio para la inclusión de argumentos que justifiquen los cambios realizados. Este punto coincide totalmente con el enfoque de la metodología DILIGENT, que además sugiere la utilización de un modelo de argumentación [TPSS05], basado en una ontología que conceptualiza el intercambio de argumentos, que debería estar integrado a la herramienta.

- Si en alguno de los contextos locales surge la necesidad de modificar una porción del modelo que forma parte de la ontología general, a diferencia de la recomendación de la metodología DILIGENT, esta propuesta sugiere que la herramienta restrinja esta funcionalidad, permitiendo únicamente argumentar el cambio, que posteriormente debería ser estudiado a nivel general. Existen dos razones por las que se recomienda introducir esta restricción:
 - Si cada grupo modifica su visión del modelo general en forma independiente, a lo largo del tiempo puede ocurrir que cada ontología local se desvíe significativamente del modelo que en principio debía ser el núcleo conceptual de toda la red de ontologías.

 - Puesto que la ontología general fue concebida desde un principio para representar los aspectos generales y comunes a los diferentes contextos locales, resulta poco probable que a nivel local surjan

cambios importantes a la conceptualización consensuada desde un inicio. Por lo tanto, un mecanismo de trabajo consistente no debería violar la regla de que siempre exista un modelo común que hace posible la integración de los diferentes contextos. Si existe alguna porción del modelo general que necesita ser modificado para algún contexto local, en la etapa de actualización de la ontología general, debería ser manejado este requerimiento, adaptándose el modelo general para que siga contemplando todos los contextos.

3.2.2. Actualización de ontología general

En esta etapa, ingenieros de ontologías y expertos de dominio trabajan juntos, aunque lo hacen mayoritariamente los primeros, por ser los responsables de adaptar la ontología general. Respecto a los expertos de dominio, como lo propone DILIGENT, se sugiere contar en este grupo de trabajo centralizado con representantes de los grupos locales.

La actividad de *conceptualización* en esta etapa se alimenta de la conceptualización local, debe estar apoyada por la actividad de *Evaluación* y si es necesario también por la actividad de *Integración*. Durante la actualización de la ontología general, la actividad de conceptualización tiene especial importancia, ya que los ingenieros deben analizar las adaptaciones realizadas a nivel local para decidir, teniendo en cuenta las argumentaciones de los expertos que trabajan en contextos locales, qué cambios van a ser incluidos en la ontología general.

La actividad de conceptualización en la actualización de la ontología general, se vuelve una tarea crítica, en especial cuando aumenta la cantidad de grupos de trabajo locales. Por lo tanto, en un entorno de trabajo distribuido altamente escalable es imperiosa la necesidad de contar con una herramienta para facilitar la conciliación entre modelos. Desde una perspectiva central, entonces, la herramienta requerida debería tener las siguientes características:

- Permitir visualizar fácilmente los cambios de cada ontología local aún no analizados a nivel general, con sus correspondientes argumentos.
- Comparar cambios realizados en diferentes contextos locales y cambios requeridos al modelo general, identificando el grado de similitud entre ellos [ES04].
- Cada vez que se modifique la ontología general a raíz de la identificación de extensiones comunes en los contextos locales, automáticamente deberían actualizarse las ontologías locales para que reusen el modelo general, evitando redundancia. Aunque debería ser menos probable, si ocurre que una porción del modelo general debe ser eliminada por no satisfacer todos los contextos locales, automáticamente debería ser incluida en las ontologías locales que sí la necesiten. Respecto a esta característica, la metodología DILIGENT en cambio, sugiere que los grupos de trabajo locales deben encargarse de reusar las actualizaciones realizadas a nivel general. Esto puede

Llevar a que los modelos locales sufran un importante desvío respecto al modelo general, en caso que no se reutilice el modelo general.

Ejemplo 14 – Actualización de ontología general

En la Figura 18 se ilustra un posible modelo de la ontología general para área de trabajo Calidad. La Figura 19 muestra dos fragmentos de especializaciones de esta ontología para los factores de calidad Confianza y Legibilidad, respectivamente.

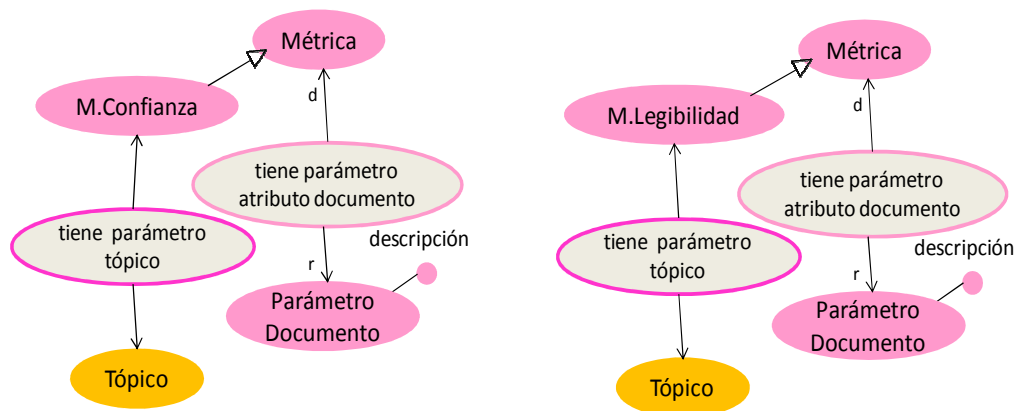


Figura 19. Ontología General de Calidad

Si las ontologías locales, para cada uno de los factores de calidad, tienen para sus métricas particulares una relación “tiene parámetro tópico”, cuyo dominio es el concepto correspondiente a la métrica específica y cuyo rango es el concepto Tópico, esta relación debe ser adoptada a nivel general, incluyéndose dicha relación en la ontología general, como se ilustra en la Figura 20.

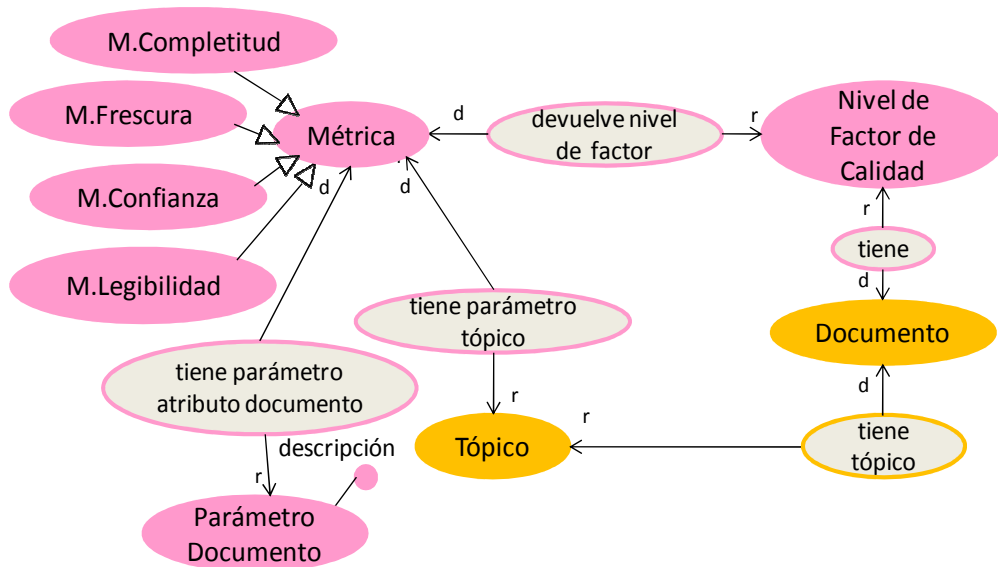


Figura 20. Ontología General de Calidad actualizada con adaptaciones locales.

4. Diseño de una ontología para la recomendación de sitios web en el área de la salud

Una vez completado el análisis de seis metodologías representativas del estado del arte en lo que refiere a ingeniería de ontologías, el proceso de elaboración de una guía de trabajo para el diseño de ontologías evoluciona hasta la obtención de los primeros pasos de una propuesta metodológica, Conceptual Ontology Design, presentada en la sección anterior, que si bien está sujeta a futuras mejoras, alcanza cierto grado de completitud. La evolución de la propuesta se nutre de la selección de los aspectos más relevantes de estas metodologías y la incorporación de otros aportes de la literatura acerca de técnicas específicas, y desde el punto de vista práctico, la propuesta se afirma a medida que es aplicada a un caso de estudio de evaluación de sitios web del área de la salud, en el marco del proyecto “Red Temática SALUS” del programa CYTED.

En esta sección se presenta el diseño de una ontología que tiene como objetivo principal emitir una recomendación de sitios web del área de la salud a diferentes perfiles de usuario. Para lograr este objetivo, es necesario contar con una estructura semántica cuyo núcleo conceptual permita asociar niveles de calidad a contenidos web, de acuerdo a criterios previamente establecidos, permitiendo emitir una recomendación de lectura de los contenidos a diferentes perfiles de usuario. Esto lleva a la definición de una ontología para abstraer los conceptos más relevantes

de la realidad, que permite visualizar los contenidos web de acuerdo a cuatro perspectivas diferentes: dominio de la salud, sitios web, factores de calidad y perfiles de usuario. La ontología que se construye tiene una estructura compuesta de un conjunto de sub-ontologías interrelacionadas, que surge entonces como resultado de un proceso de diseño en el que se aplica la propuesta Conceptual Ontology Design. El punto de partida de este proceso es el conocimiento adquirido del sitio [WR1] y entrevistas con algunos expertos. A continuación se da una breve explicación de la evolución del modelo y un detalle de la estructura final obtenida.

4.1. Evolución del diseño

El proceso de construcción de una ontología general aplicando la propuesta Conceptual Ontology Design, comienza con la ejecución de múltiples iteraciones de las actividades de especificación y conceptualización hasta la obtención del modelo general, que puede ser extendido en contextos locales.

Siguiendo los pasos detallados en la metodología propuesta para las actividades de especificación y conceptualización, se obtiene una primera lista de términos que en primera instancia se agrupan en dos áreas de trabajo: Calidad y Salud. Estas áreas de trabajo evolucionan identificándose conceptos principales, que se especializan, generalizan y vinculan obteniéndose las primeras taxonomías y otras relaciones binarias, para dar lugar a dos incipientes modelos cuya estructura se

ilustra en las Figuras Figura 21 y Figura 22. En este punto se tiene entonces un modelo compuesto únicamente por dos ontologías.

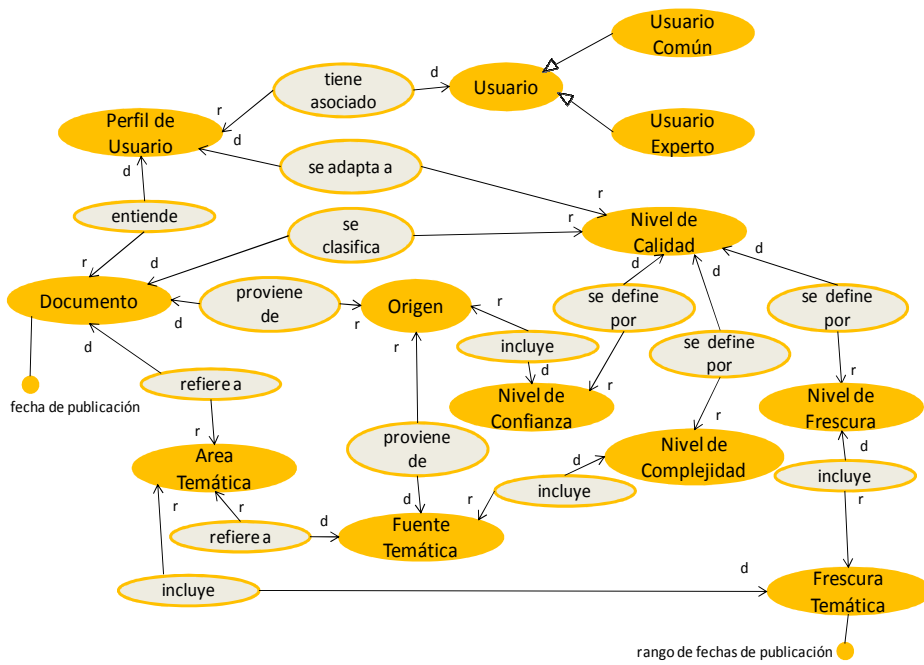


Figura 21. Primera conceptualización del área de trabajo Calidad.

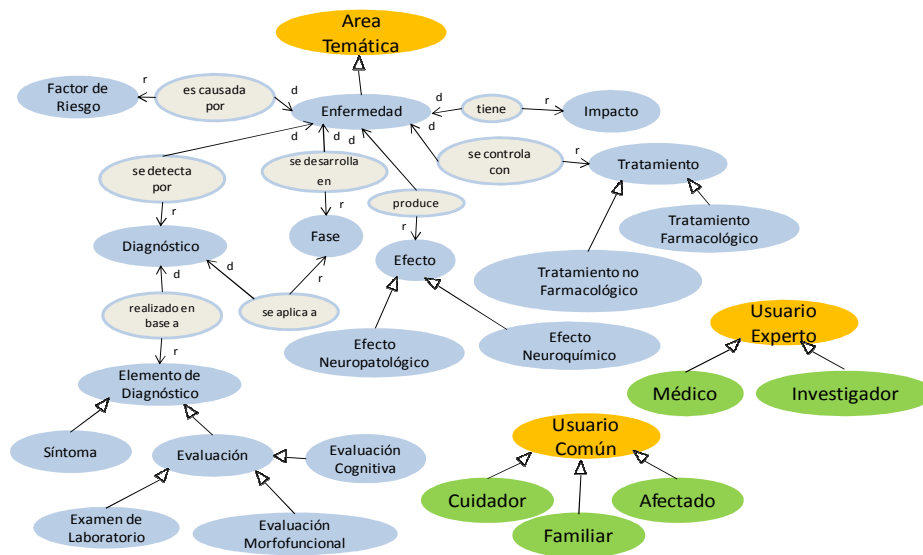


Figura 22. Primera conceptualización del área de trabajo Salud.

En subsecuentes iteraciones de las actividades de especificación y conceptualización aparecen nuevos términos, que dan lugar a conceptos principales y paulatinamente algunos de estos conceptos se separan para crear nuevas áreas de trabajo. Usuario es un área compuesta por conceptos que en principio aparecían formando parte del área Calidad, pero son identificados como independientes del resto, ya que es posible construir con ellos un modelo que puede ser reutilizado en otra aplicación diferente del caso de estudio de evaluación de calidad de sitios web. En esta etapa del proceso se fue manejando la posibilidad de construir ontologías especializando las correspondientes a las áreas de trabajo Calidad y Salud, que son generales. El modelo de Calidad puede especializarse por factor de calidad y enfermedad, y el modelo de Salud, por enfermedad. Se obtiene así una red de ontologías compuesta por las ontologías generales de Salud, Calidad y Usuario, y ontologías específicas de Salud y Calidad, que se modelan para la enfermedad Alzheimer y los factores de calidad completitud y confianza. Se identifica también el área de trabajo Sitio, cuyos conceptos también forman un modelo separado, independientemente de que la aplicación final tenga como objetivo medir la calidad de los contenidos. A la red de ontologías se agrega entonces la sub-ontología Sitio. Cada vez que se separa una nueva área se ajustan las vinculaciones entre los modelos. Se tiene entonces una red de ontologías del estilo de la que se muestra en la Figura 23. En ella se aprecia una vinculación del modelo Perfil Usuario

hacia el modelo Calidad, ya que para emitir una recomendación de un contenido web a un usuario debe asociarse de alguna manera la calidad del documento consultado a las características que definen el perfil del usuario que consulta, para así determinar si el documento es adecuado a ese perfil. Pero esta vinculación quita independencia al modelo Perfil Usuario, por quedar encapsulada la conceptualización de la recomendación misma en este modelo. Por este motivo se define una nueva ontología, Recomendación, que vincula los modelos Calidad y Perfil Usuario representando los conceptos y relaciones que resuelven la recomendación de contenidos a usuarios, manteniéndose independientes las ontologías Calidad y Perfil Usuario. Se obtiene así el modelo final que se explica en detalle en la Sección siguiente.

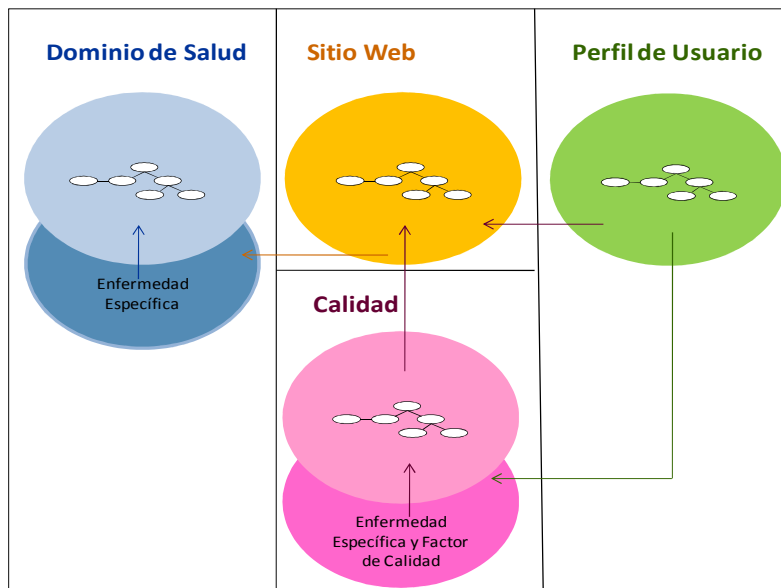


Figura 23. Red de ontologías preliminar al modelo final.

4.2. Descripción del modelo

La Figura 24 muestra la red de sub-ontologías que componen el modelo de la ontología del caso de estudio. En este modelo, las sub-ontologías que aparecen más a la izquierda tienen mayor grado de independencia que las que se ubican a la derecha. De esta forma, la integración se resuelve de manera que cada una reutiliza las sub-ontologías ubicadas a su izquierda.

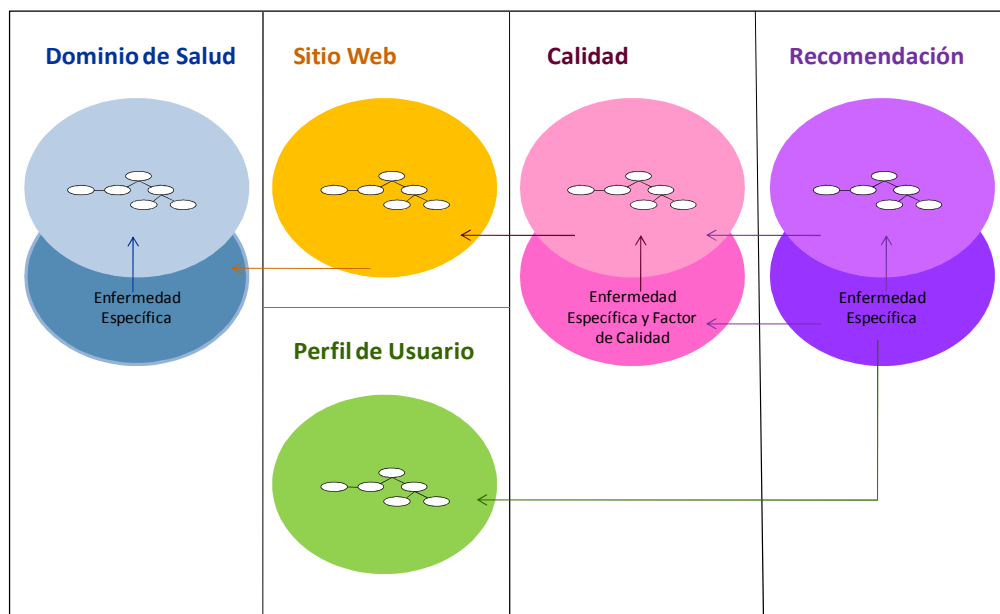


Figura 24. Red de ontologías.

Las ontologías Dominio de Salud y Perfil de usuario son totalmente independientes. Dominio de Salud representa el dominio de la aplicación, es conceptualizada como una ontología de carácter general, de manera que debe ser especializada para cada enfermedad específica, con el conocimiento provisto por el experto. La ontología Perfil de usuario

representa el sujeto de la recomendación, es decir los usuarios. Con ella se modelan diferentes características de los usuarios que influyen en la recomendación de contenidos web. La ontología Sitio Web se crea para modelar el objeto de la recomendación, es decir los documentos consultados por los usuarios. Esta ontología reutiliza la ontología Dominio de Salud para una enfermedad específica (la especialización), que representa el tema a que refieren los documentos publicados en sitios web. El modelo de la ontología Calidad se define para representar los criterios de evaluación de diferentes factores de calidad. Ella reutiliza la ontología Sitio Web (y por lo tanto la ontología Dominio de Salud) ya que el objetivo de la ontología Calidad es clasificar documentos de sitios web de acuerdo a criterios establecidos. La ontología Calidad también debe ser extendida para diferentes enfermedades y factores de calidad, de manera que los expertos del área médica y del área de calidad definan las métricas de evaluación de contenidos para cada factor de calidad. Finalmente, la ontología Recomendación es la estructura que explota el resto de las ontologías definidas de acuerdo a diferentes perspectivas, para representar la recomendación de un contenido a un usuario dado. Las ontologías utilizadas por el modelo de Recomendación proveen los parámetros que resultan determinantes en el proceso de obtención de un nivel de recomendación. La ontología Recomendación también debe especializarse dependiendo de la enfermedad, permitiendo a los expertos de cada afección definir las métricas a ser aplicadas. Esta ontología

reutiliza las ontologías Perfil de Usuario y Calidad para una enfermedad y factor de calidad específicos (y por lo tanto las ontologías Sitio Web y Dominio de Salud).

Si bien el presente diseño fue realizado para un caso de estudio de dominio salud, resulta adecuado para ser aplicado a la evaluación de contenidos de diferentes dominios, como e-learning o leyes, por lo que se trata de un diseño más general. A continuación se detalla cada una de las sub-ontologías de la Figura 24.

4.2.1. Ontología Dominio de Salud

La estructura de esta sub-ontología se ilustra en la Figura 25. Los conceptos identificados en este modelo surgen del análisis del sitio [WR1]. Si bien esta estructura se construye en base a información contenida en un sitio acerca de una enfermedad específica (Alzheimer), la consulta a sitios de otras afecciones visualizando los conceptos más relevantes lleva a concluir que el modelo resultante puede ser considerado como una ontología general para cualquier enfermedad. Al analizar el dominio de una enfermedad específica con más profundidad, estudiando la terminología utilizada, van a surgir nuevos conceptos que pueden ser introducidos.

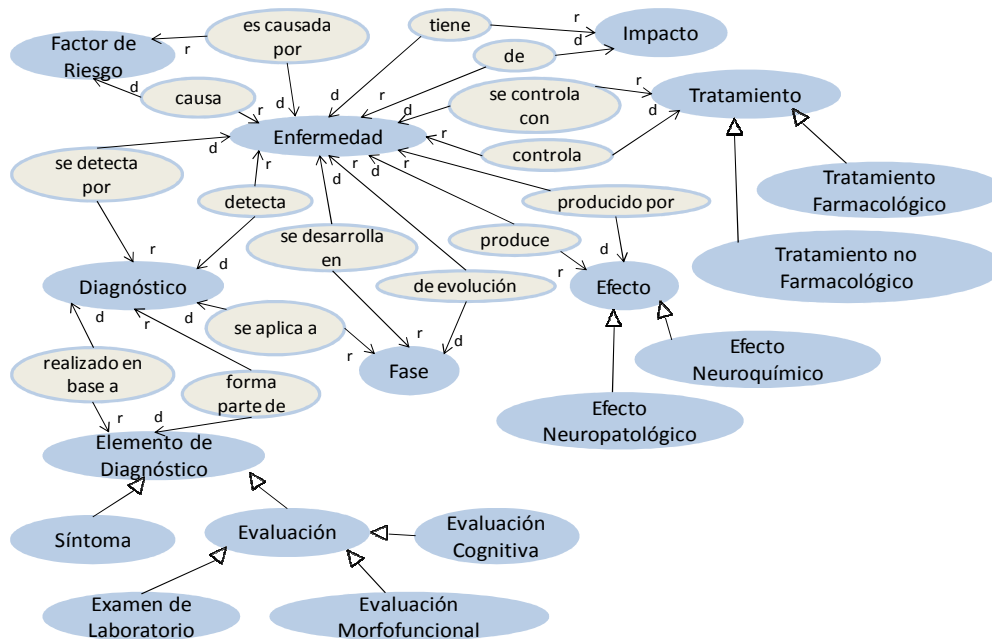


Figura 25. Ontología Dominio de Salud.

Como se muestra en la Figura 25, por cada relación que vincula una instancia del concepto Enfermedad a instancias de otros conceptos, se define una relación inversa. Esta decisión es tomada ya que por cada instancia de cualquier concepto relacionado con el concepto Enfermedad, tal como Diagnóstico o Tratamiento, es necesario obtener la correspondiente instancia del concepto Enfermedad. Por ejemplo, es necesario saber a qué enfermedad corresponde un tratamiento dado. De esta forma, se agrega la relación “controla”, con dominio Tratamiento y rango Enfermedad. Esta decisión de diseño permite deducir, para cualquier instancia de cualquier concepto, a qué enfermedad corresponde.

Como ya se mencionó, la ontología Dominio de Salud para Enfermedad Específica es una especialización de la ontología Dominio de Salud, que representa información acerca de una enfermedad particular, provista por el experto. La Figura 26 muestra un modelo que corresponde a la enfermedad Alzheimer. En esta extensión se definen subclases cuyas instancias corresponden a la enfermedad específica. Por ejemplo, la subclase Diagnóstico Alzheimer se define como el conjunto de todos los individuos de la clase Diagnóstico que se vinculan a la instancia “Alzheimer” de la clase Enfermedad, a través de la relación “detecta”. A continuación se muestra la definición de la subclase Diagnóstico Alzheimer en lenguaje OWL.

```

<owl:Class rdf:ID="DiagnosticoAlzheimer">
  <owl:equivalentClass>
    <owl:Class>
      <owl:intersectionOf rdf:parseType="Collection">
        <owl:Restriction>
          <owl:onProperty rdf:resource="#p1;detecta"/>
          <owl:hasValue rdf:resource="#Alzheimer"/>
        </owl:Restriction>
        <owl:Class rdf:about="#p1;Diagnostico"/>
      </owl:intersectionOf>
    </owl:Class>
  </owl:equivalentClass>
</owl:Class>

```

Esto significa que los axiomas que establecen las condiciones necesarias y suficientes a ser satisfechas por las instancias de las subclases para una enfermedad específica utilizan las relaciones inversas introducidas en la ontología Dominio Salud.

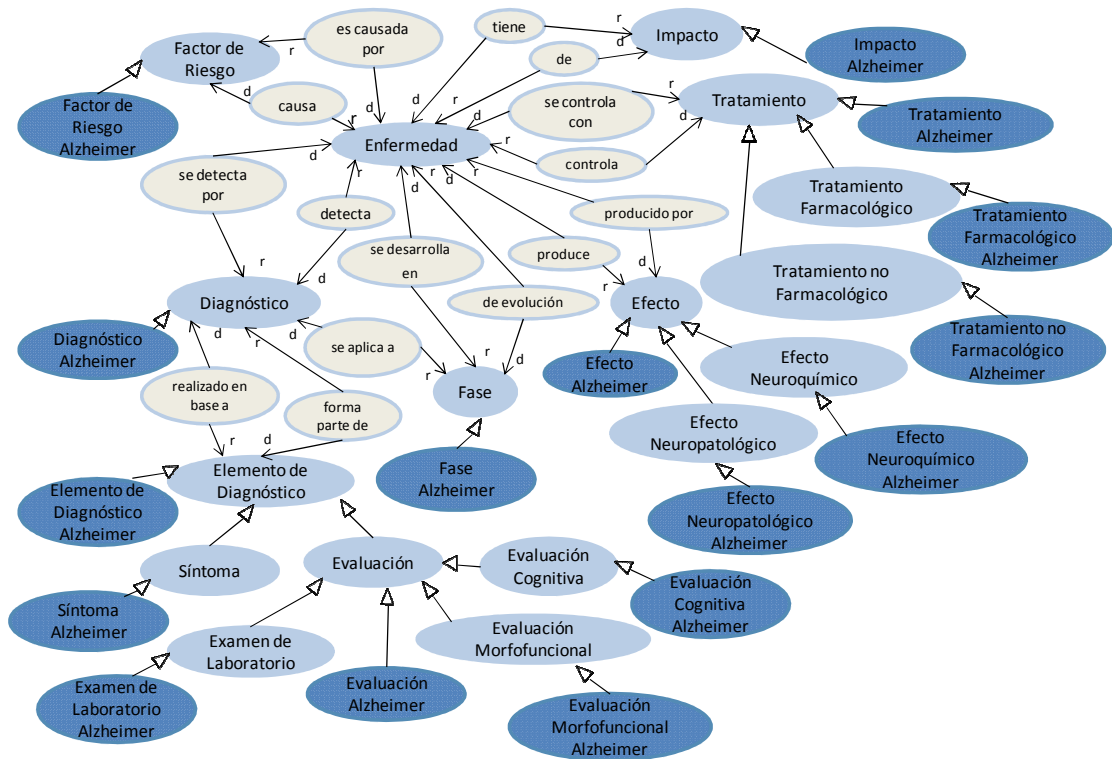


Figura 26. Ontología Dominio de Salud para una enfermedad específica.

Por cada enfermedad es posible especificar una especialización diferente, con la extensión de la ontología Dominio de Salud agregando estructuras más complejas, con nuevos conceptos y relaciones específicos, de acuerdo a la expresividad requerida. Para asegurar que la ontología Dominio de Salud incluya toda la terminología médica sin omitir ningún concepto, se plantea la posibilidad de integrar estos primeros conceptos identificados con alguna ontología general o estructura taxonómica médica. Tal estructura debería proveer toda la terminología médica y si está disponible, una estructura semántica que organice esta terminología. GALEN, UMLS y ON9 son algunas de las bases de datos de información biomédica que están disponibles. De estas, UMLS ha sido analizada con

más profundidad, por integrar diferentes bases de datos terminológicas en una estructura semántica común denominada Semantic Network. Por lo tanto, UMLS es la ontología candidata a ser integrada. Para ello debería definirse un mecanismo de integración, de acuerdo a la propuesta metodológica de la Sección 3.1.3. Si se aplica merge, una posible solución sería extraer la porción de la estructura Semantic Network de UMLS que tiene que ver con el modelo Dominio de Salud de la Figura 25 y combinarlo con este modelo para obtener una nueva ontología. El mapping, por el contrario, es un proceso de integración por el cual deben establecerse correspondencias entre la estructura de la Figura 25 y los conceptos adecuados de la estructura Semantic Network de UMLS, para obtener de ésta la información requerida (estructura y terminología), preservando los modelos originales.

4.2.2. Ontología Sitio Web

Esta ontología representa el contenido de los sitios web, es decir el conjunto de documentos publicados que deben ser clasificados de acuerdo a criterios de calidad, asociando a ellos niveles de calidad que hagan posible la recomendación de estos contenidos a diferentes usuarios. La Figura 27 muestra el diseño de esta sub-ontología. En ella existe un concepto central Documento, con diferentes atributos y relaciones que lo vinculan a otros conceptos tales como Autor, Origen, Tópico, que representa el tema del documento y la unión de conceptos de la sub-ontología Dominio de Salud para vincularlo a cada elemento de

contenido del dominio médico. Con respecto al origen de los documentos, si es un sitio web, el modelo expresa si tiene certificación de calidad o no.

Dos subclases del concepto Documento son definidas: Doc. Página y Doc. No Página. Esta especialización es necesaria a los efectos de aislar los documentos que son páginas web, ya que éstos tienen características que los distinguen del resto, como el hecho de que sus instancias pertenecen también al concepto Página. Por lo tanto, el concepto Doc. Página debe ser una subclase del concepto Página, que incluye todas las instancias que representan páginas web (sean éstas documentos o no). Pero la implementación con Protégé 3.4 no permite la representación de la relación de herencia múltiple desde el concepto Doc. Página hacia los conceptos Documento y Página. Por lo tanto debe escribirse una regla para implementar la relación “Doc. Página is-a Página”.

`DocPage (?x) → Page (?x)`

Además, se escribe también una regla para controlar que toda instancia del concepto Doc. Página que tiene como origen un sitio web (una instancia del concepto Sitio) es una instancia del concepto Página y está vinculada al concepto Sitio a través de la relación “es parte de”.

`DocPagina (?x) ∧ Sitio (?y) ∧ tieneOrigen (?x, ?y) → Pagina (?x) ∧ esParteDe (?x, ?y)`

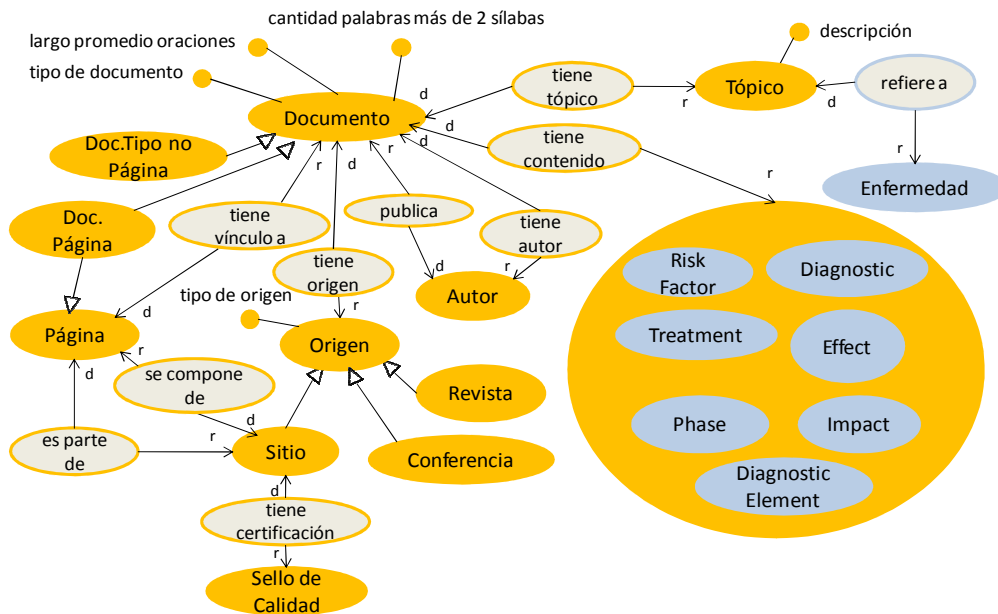


Figura 27. Ontología Sitio Web.

4.2.3. Ontología Perfil de Usuario

El análisis exhaustivo de los elementos de este modelo no está en el alcance de este trabajo, por lo que la sub-ontología definida, refleja los conceptos más básicos identificados acerca del perfil de usuario. Como se ilustra en la Figura 28, en la estructura se tienen en cuenta tres diferentes aspectos de los usuarios: nivel académico, el rol que juega el usuario (médico, investigador, paciente, cuidador, familiar, etc.) y la franja etárea (niño, adolescente, adulto joven, adulto mayor, etc.). Estas son algunas de las características de los usuarios que influyen en la comprensión e interés de éstos con respecto a documentos médicos, contribuyendo a definir su perfil.

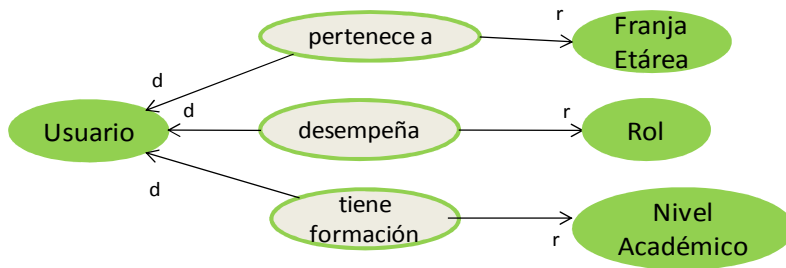


Figura 28. Ontología Perfil Usuario.

4.2.4. Ontología Calidad

Tradicionalmente, la calidad de los sitios web ha sido evaluada teniendo en cuenta factores de calidad como: navegación, aspectos de interface de usuario, legibilidad (en lo que tiene que ver con tamaño de letra, colores, imágenes, etc.), aspectos de desempeño (tiempo que toma el acceso al contenido de un sitio), etc. Sin embargo, el enfoque de este trabajo apunta al análisis de la calidad que surge del valor de la información que el sitio provee, su adaptación al uso que el lector desea darle y en qué grado la información satisface sus expectativas. Desde este punto de vista, la calidad del sitio depende del contexto de uso y del consumidor final.

Siguiendo este enfoque, el conocimiento adquirido con expertos permitió seleccionar algunos factores de calidad como más factibles de ser medidos. Algunas consideraciones acerca de estos factores se describen brevemente a continuación.

Confianza

De [PLW02] se extraen dos definiciones:

“Believability: the extent to which data is regarded as true and credible”.

“Reputation: the extent to which data is highly regarded in terms of its source or content”.

La primera es una definición general que expresa que el significado de los datos es creíble, mientras que la segunda menciona determinadas propiedades del documento, como el origen, que son considerados a la hora de evaluar si el documento es creíble o no.

Un aspecto importante a tener en cuenta con respecto a este factor de calidad, es la existencia de sitios certificados con sellos de calidad tales como HON [WR2], WIS [WR3] and WMA [WR4]. Estas certificaciones constituyen una herramienta de evaluación, ya que serían lógico pensar que los documentos apuntados por sitios certificados sean evaluados con un nivel de calidad más alto que los contenidos de sitios sin certificación.

Frescura

En [PLW02] puede encontrarse la siguiente definición:

“Timeliness: the extent to which data is sufficiently up-to-date for the task at hand”

Con respecto a este factor, lo que realmente importa es medir la frescura de los datos publicados más que su fecha de publicación.

Legibilidad

[D04] es una investigación de las diferentes métricas de legibilidad que han sido creadas para diferentes dominios y perfiles de usuario. De este trabajo se extrae la siguiente definición:

“Readability is what makes some texts easier to read than others”.

El mismo trabajo menciona la definición de G. Harry McLaughlin (1969), creador de la fórmula de legibilidad SMOG:

“The degree to which a given class of people find certain reading matter compelling and comprehensible.”

Existe gran cantidad de formulas de legibilidad creadas por diferentes autores, como los niveles FOG y SMOG, que al ser testeados alcanzaron resultados aceptables.

FOG grade level = $0.4 (\text{average sentence length} + \text{hard words})$

SMOG grade level = $3 + \sqrt{\text{polysyllable count}}$

Compleitud

De [PLW02] se toma la siguiente definición:

“Completeness: the extent to which data is not missing and is of sufficient breadth and depth for the task at hand”.

En las primeras iteraciones de la especificación y conceptualización, se intentó modelar este factor con alto nivel de detalle. La experiencia transmitida por expertos llevó a redefinir la representación de las métricas de evaluación de este factor, descartando la idea de modelar métricas con condiciones complejas, ya que se estaría evaluando la correctitud y no la completitud de los contenidos.

Del presente análisis de factores de calidad, la principal conclusión que surge es que la posibilidad real de medir la calidad de los documentos y la profundidad o máximo nivel de detalle que puede ser alcanzado en la evaluación, deben ser establecidos por el experto a través de la definición de métricas.

El modelado de todos los aspectos relativos a la evaluación de calidad de los documentos, resulta en una ontología general que es un modelo núcleo, la ontología de Calidad. Esta ontología tiene como objetivo proveer una base para la implementación de sistemas adaptativos, cualquiera sea el factor de calidad y dominio de aplicación. En esta ontología no se definen métricas específicas a ser aplicadas, ya que esta tarea debe ser responsabilidad del experto en el dominio y factor de calidad. Esto lleva a la especialización de la ontología general, de acuerdo a la enfermedad y el factor de calidad a ser evaluados. Cada extensión debería ser construida por el ingeniero de ontologías y el experto, trabajando en conjunto. De esta forma se obtienen una o más ontologías

que extienden el modelo núcleo, a través de un proceso en el cual el experto provee las métricas de calidad, generalmente expresadas utilizando reglas.

La Figura 29 muestra la ontología de Calidad. Este modelo tiene dos conceptos relevantes, Factor de Calidad y Métrica. Con respecto al concepto Métrica, se definen subclases para clasificar las métricas que miden el mismo factor de calidad. Se expresa también que las métricas pueden depender del tema del documento y de diferentes propiedades del documento representadas con el concepto Parámetro Documento. El concepto Nivel de Factor de Calidad se especializa también por cada factor de calidad y se asocia al concepto Documento.

Durante el proceso de conceptualización se analiza la idea de asociar un nivel de calidad general a los documentos, combinando los niveles de calidad de todos los factores evaluados. Esta idea es finalmente descartada, porque las posibles instancias de ese nivel de calidad general representarían múltiples combinaciones de las instancias del concepto Nivel de Factor de Calidad, por cada factor de calidad considerado. Por ejemplo, no es fácil determinar cuál sería el significado de un nivel de calidad general que exprese la combinación de un nivel de frescura bajo, un nivel de confianza alto y un nivel de legibilidad medio. Parece más útil evaluar los factores de calidad individualmente para cada documento,

considerando el propósito de dar una recomendación a un cierto perfil de usuario.

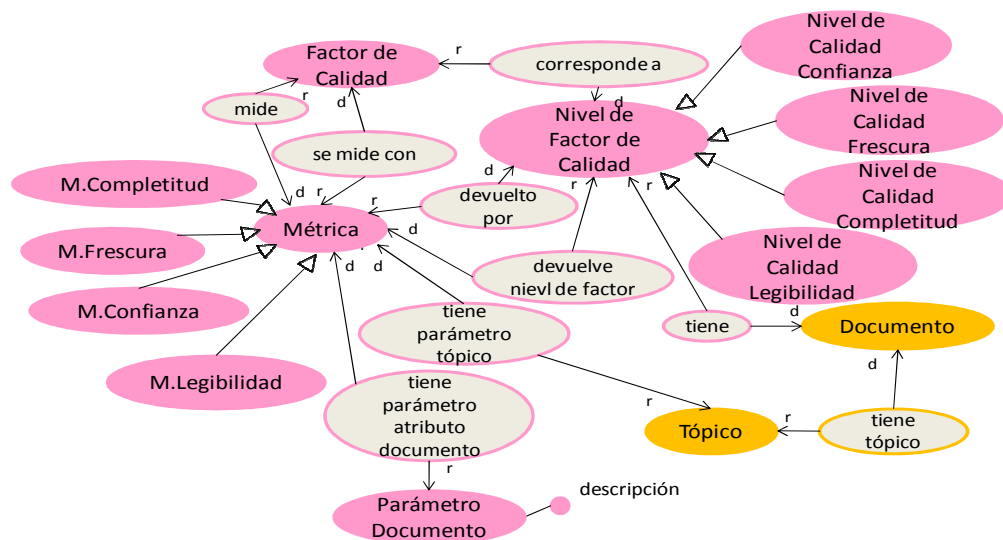


Figura 29. Ontología Calidad.

La ontología Calidad debe entonces especializarse por enfermedad y factor de calidad a ser evaluado. En cada especialización el experto define las métricas para medir la calidad de los documentos, por lo que en principio la extensión a la ontología general consiste en la especificación de las métricas. Para representarlas se propone el uso de reglas SWRL, teniendo en cuenta los siguientes argumentos:

- Permiten expresar claramente en el modelo la función que cumplen las métricas en el proceso de asignación de niveles de calidad de cada documento. A través de ellas puede indicarse en forma explícita que a partir de los parámetros dados por determinadas propiedades del documento (expresión a la izquierda del símbolo \rightarrow) se infiere el nivel

de calidad (a la derecha de \rightarrow). A continuación se muestra un ejemplo de regla SWRL para la evaluación del factor de calidad confianza.

```
Documento(?d) ^ Sitio(?s) ^ SelloDeCalidad(?l) ^ tieneCertificacion(?s, ?l)
^ tieneOrigen(?d, ?s) -> tieneNivelDeCalidadFactor(?d, Confiable)
```

- Puesto que la definición de métricas es responsabilidad de los expertos, la sintaxis de escritura que contiene el antecedente y el consecuente es un mecanismo intuitivo y ordenado para usuarios que no son especialistas en diseño de ontologías. De esta forma no tienen que explorar en el modelo para decidir dónde definir axiomas, etc.

4.2.5. Ontología Recomendación

La Figura 30 muestra la estructura de esta ontología, de carácter general, siendo la base para la construcción de ontologías de recomendación para una enfermedad específica.

Uno de los conceptos más relevantes en este modelo es Lectura de Contenido, que establece una relación entre dos conceptos principales: Documento y Usuario. Lectura de Contenido expresa la acción que ejecuta un usuario al leer un documento, de manera que el concepto Usuario es el sujeto y el concepto Documento, clasificado de acuerdo a los diferentes factores de calidad, es el objeto de la recomendación. Se reutilizan por lo tanto, los modelos Perfil Usuario y Calidad.

Otro concepto central es Métrica de Recomendación, que representa si se recomienda o no y en qué grado, que un usuario lea un documento. Los

parámetros de entrada (aspectos que influyen en la recomendación) son las características de los usuarios, representados por el concepto Parámetro Usuario, los niveles de calidad del documento, representados por el concepto Parámetro Nivel de Calidad y el tópico del documento. El resultado de aplicar una métrica de recomendación permite la categorización de las instancias del concepto Lectura de Contenido, asociándoles un nivel de recomendación. El método de cálculo del nivel no está expresado en esta ontología general, esas fórmulas se definen utilizando reglas SWRL en la especialización de la ontología, para una enfermedad específica. Esta tarea es responsabilidad del experto.

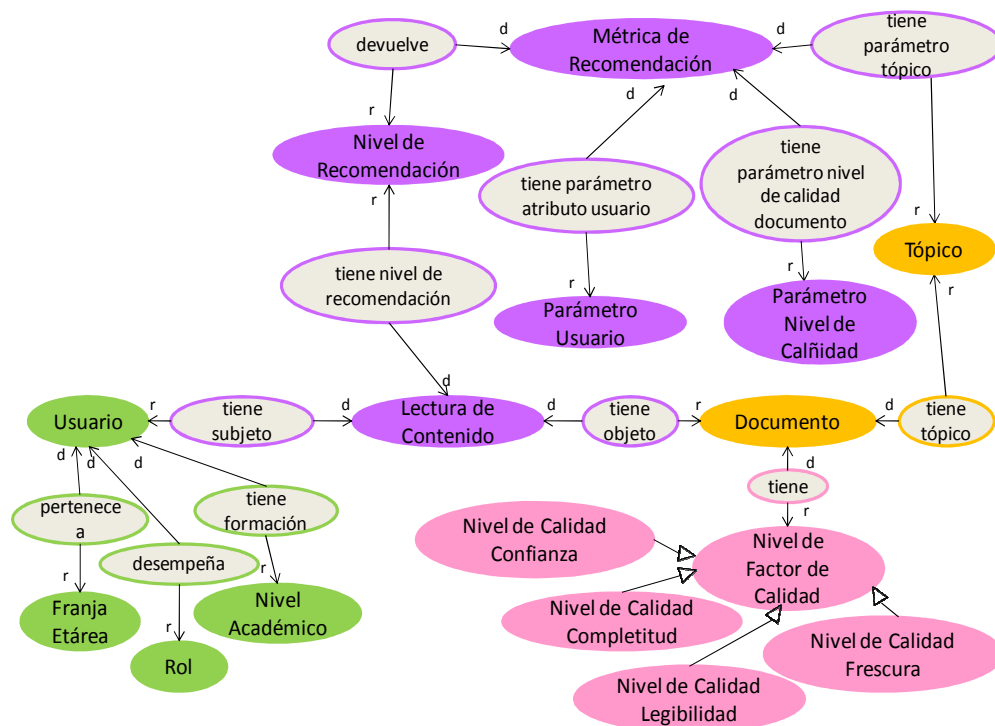


Figura 30. Ontología Recomendación.

4.3. Utilización de la red de ontologías en una aplicación

La construcción de una red de ontologías de acuerdo al modelo descripto tiene como objetivo más inmediato su explotación por tres aplicaciones a ser implementadas en el futuro. Estas aplicaciones van a formar parte de una plataforma que permita a usuarios con diferente perfil consultar contenidos médicos clasificados de acuerdo a criterios de calidad, obteniendo una recomendación de acuerdo a su perfil y al nivel de calidad del documento consultado. Las aplicaciones a ser implementadas son las siguientes:

- Mantenimiento de métricas de calidad y recomendación por parte de un usuario experto.
- Anotación y clasificación de contenidos web de acuerdo a métricas de calidad por parte de un usuario administrador.
- Recomendación de contenidos a perfiles de usuario en un portal médico.

A continuación se describen brevemente cada una de ellas.

4.3.1. Mantenimiento de métricas de calidad y recomendación

La Figura 31 muestra una visión general de la aplicación, con una interfaz que permite al experto en una enfermedad específica y/o factor de calidad ingresar las métricas de evaluación de calidad de documentos y

recomendación a diferentes perfiles de usuario. La capa de negocio se va a encargar de realizar las invocaciones necesarias a la capa de datos para generar las instancias correspondientes a métricas, niveles de calidad y reglas de cálculo de los niveles de calidad, en las ontologías de Calidad y Recomendación. En la capa inferior se encuentran las ontologías Sitio Web, Perfil de Usuario, Calidad y Recomendación.

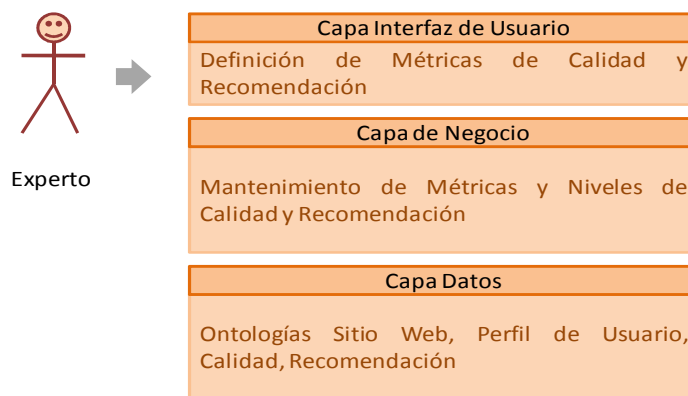


Figura 31. Mantenimiento de Métricas y Niveles de Calidad y Recomendación

La capa de negocio se desarrolla en la Figura 32, con un módulo de Mantenimiento de Métricas de Calidad que en base a la estructura de la ontología Calidad (integrada a la ontología Sitio Web), una vez validado el ingreso de datos del experto, invoca los métodos que resuelven la persistencia de las instancias que representan las métricas. De forma similar existe un módulo de Mantenimiento de Métricas de Recomendación que de acuerdo al modelo de la ontología Recomendación (integrada a toda la red de ontologías) se encarga de

poblar esta ontología con las instancias correspondientes a las métricas definidas por el experto.

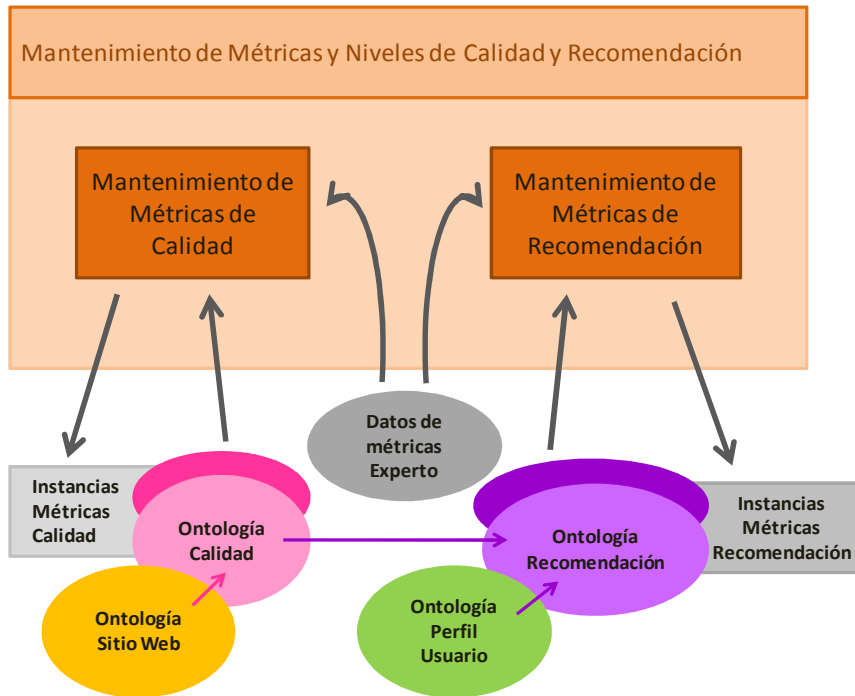


Figura 32. Componentes de la aplicación de Mantenimiento de Métricas y Niveles de Calidad y Recomendación

4.3.2. Anotación y clasificación de contenidos Web de acuerdo a métricas de calidad

La Figura 33 muestra una visión general de la aplicación, con una interfaz que permite al administrador del portal médico anotar y clasificar documentos del dominio salud de acuerdo a las métricas de calidad definidas por el experto. Por cada documento vinculado al sitio, la capa de negocio va a invocar a un módulo de extracción de metadatos y a un módulo que clasifica el documento de acuerdo a las métricas de calidad.

En la capa inferior se encuentran las ontologías Dominio Salud, Sitio Web y Calidad.

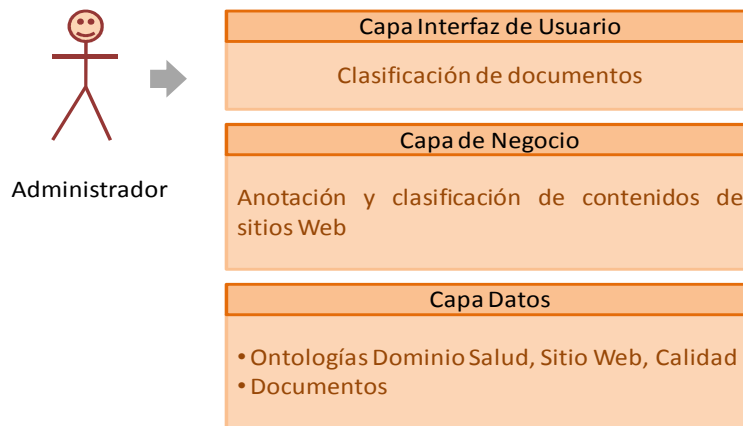


Figura 33. Anotación y clasificación de contenidos Web de acuerdo a métricas de calidad

La Figura 34 muestra más detalle acerca de la capa de negocio. Guiado por la ontología Sitio Web (que está integrada a la ontología Dominio Salud), el módulo Extractor recupera información de determinados atributos de los documentos, que es necesaria para el cálculo de los niveles de calidad. El módulo Clasificador de Documentos utiliza la semántica provista por la ontología Calidad para anotar los documentos con los niveles de calidad correspondientes a los factores de calidad seleccionados por el experto. Esto se realiza ejecutando las reglas SWRL que definen las métricas de calidad.

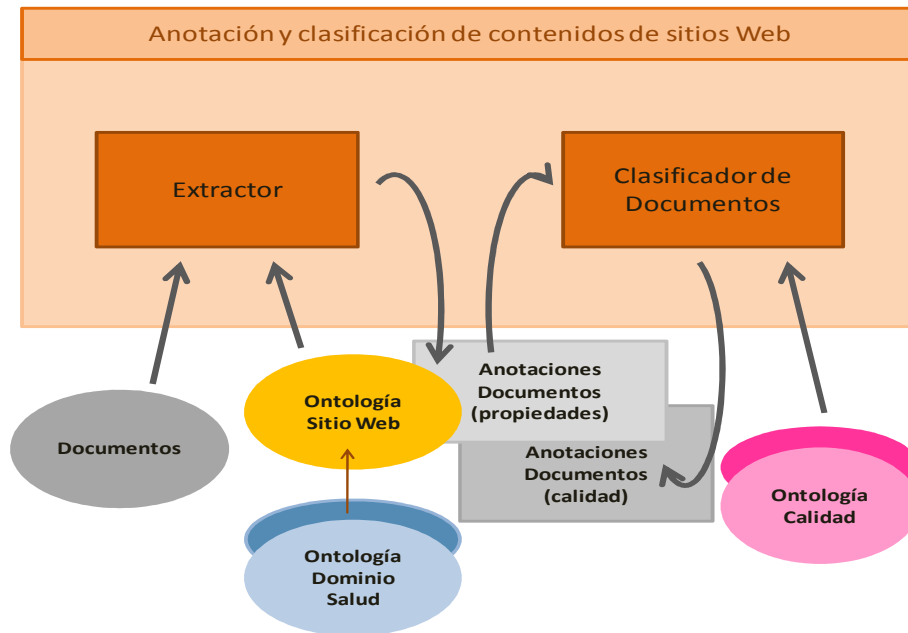


Figura 34. Componentes de la aplicación Anotación y Clasificación de Contenidos de Sitios Web

4.3.3. Recomendación de contenidos en un portal médico

En esta aplicación, según se ilustra en la Figura 35, un usuario común consulta un sitio con contenido del dominio salud, aportando además datos de su perfil (por ejemplo edad, nivel académico, etc.). La capa de negocio recoge esta información, que combina con la calidad de los documentos ya clasificados, accesibles a través del portal, para recomendar al usuario que consulta, los contenidos más adecuados a su perfil. En la capa de datos se encuentran las ontologías Sitio Web, Perfil de Usuario, Calidad y Recomendación.

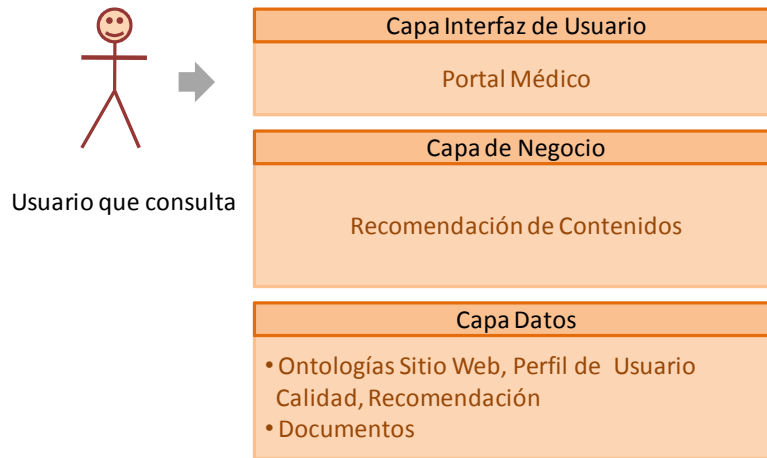


Figura 35. Recomendación de contenidos en un portal médico.

La aplicación de Recomendación de Contenidos, como se observa en la Figura 36, consta de un módulo de Mantenimiento de Perfiles de Usuario que captura los datos del usuario que ingresa, para poblar la ontología Perfil de Usuario. A continuación, el módulo de Cálculo de Niveles de Recomendación, en base a las métricas de la ontología Recomendación, para cada documento consultado por el usuario, crea la correspondiente instancia del concepto Lectura de Contenido (que vincula el documento con el usuario) y ejecutando las reglas SWRL asociadas a las métricas, calcula el nivel de recomendación para dicha instancia.

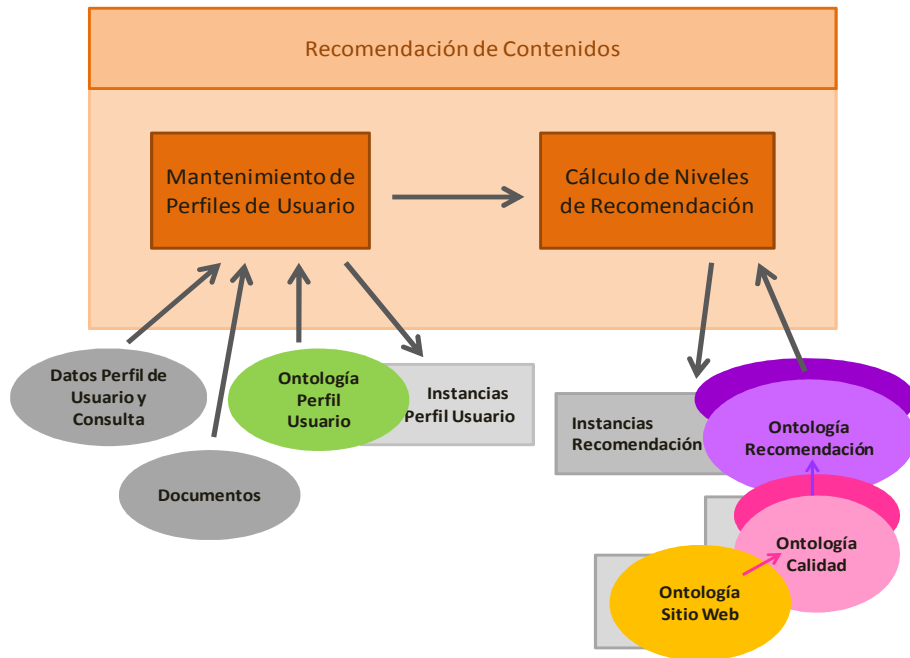


Figura 36. Componentes de la aplicación Recomendación de contenidos en un portal médico.

5. Conclusiones y trabajos futuros

La razón de ser de este trabajo es la presentación de los primeros pasos hacia la elaboración de una metodología de diseño y desarrollo de ontologías, *Conceptual Ontology Design*. La propuesta presentada incluye guías y recomendaciones extraídas de publicaciones de metodologías existentes, algunas muy ampliamente usadas. En muchas de ellas pueden encontrarse importantes aportes en lo que tiene que ver con las actividades de conceptualización, integración y evaluación, que se han incluido en este trabajo. Respecto a la actividad adquisición de conocimiento, de fundamental importancia para llevar a cabo las actividades de especificación y conceptualización, no se encontraron recomendaciones mínimamente detalladas que formen parte de alguna de las metodologías estudiadas. Como la correcta abstracción de los principales conceptos de un modelo es un aspecto fundamental, tanto para la obtención de la primera versión de la ontología, como para lograr una correcta gestión de su evolución posterior, en este trabajo se dio especial énfasis a la actividad de adquisición de conocimiento. Fue necesario entonces, realizar una exploración de diversas publicaciones en esta disciplina, incluso acerca de aspectos psicológicos que influyen en el proceso de abstracción conceptual de las personas. A partir de este estudio se logra elaborar una propuesta detallada de cómo las técnicas de adquisición de conocimiento deben ser aplicadas para dar soporte a las actividades de especificación y conceptualización. Otro aporte, con el que se intenta distinguir esta propuesta de las que se conocen hasta el momento, es

la recomendación de técnicas alternativas para la construcción de ontologías orientadas a aplicaciones específicas, que además contemplen ontologías independientes de una aplicación. UPON es una metodología cuyo enfoque está orientado específicamente a la construcción de este tipo de ontologías, por lo que el presente trabajo integra las técnicas sugeridas por esta metodología como pasos alternativos para facilitar el diseño cuando se trata de ontologías orientadas a una aplicación. Este aspecto le da cierta flexibilidad a la propuesta, que de esta forma se adapta al nivel de generalidad de la ontología a construir. Por último, al detalle de las técnicas sugeridas para llevar a cabo las principales actividades de diseño, se agrega la propuesta de un mecanismo para gestionar la evolución distribuida del proceso de desarrollo, inspirado en la metodología DILIGENT.

La metodología propuesta fue además evaluada a través de su aplicación al caso de estudio de evaluación de calidad de sitios web en el área de la salud, lo que permitió ir ajustando la guía de pasos especificados en cada actividad de la metodología. En particular, el caso de estudio sugiere una forma de trabajo descentralizada en la que cada grupo de trabajo define métricas para diferentes factores de calidad y enfermedades. En el modelo obtenido se definen ontologías de carácter general que no dependen de ninguna métrica de calidad ni temática particular y ontologías locales que extienden las centrales, en las que se especifican las métricas particulares. Se obtiene así un modelo flexible que admite adaptaciones de acuerdo a diferentes métricas de calidad.

Dado que la metodología propuesta fue validada en forma individual con la construcción de la ontología para el caso de estudio CYTED-SALUS, el próximo desafío de este trabajo consiste en propiciar la utilización de la presente propuesta por algún diseñador, para la creación de una ontología totalmente independiente al caso de estudio del presente trabajo. De esta forma sería posible realizar un monitoreo de la aplicación de los pasos propuestos por la metodología, que permitiría realizar ajustes de acuerdo a los resultados obtenidos. A los efectos de que la metodología propuesta adquiriera un carácter más estricto, otra tarea a realizar en el futuro consiste en escribir los pasos de la presente recomendación en lenguaje formal. En el corto plazo se tiene previsto también realizar una evaluación y/o propuesta de herramientas que faciliten la aplicación de una metodología colaborativa y distribuida de diseño de ontologías. La herramienta ideal debería permitir comparar modelos e intercambiar argumentos que ayuden al especialista a tomar las decisiones de diseño más adecuadas, al momento de adaptar la ontología general. Una de las metodologías estudiadas en la evaluación de metodologías existentes, HCOME, propone la utilización del ambiente de trabajo HCONE (Human Centered ONtology Engineering Environment). Se trata de una herramienta que soporta este proceso de continuo intercambio, a la vez que brinda la funcionalidad de versionado. Otro objetivo igualmente importante es llevar a cabo la integración del modelo obtenido con ontologías existentes que refuercen las áreas de

trabajo del caso de estudio que no fueron abordadas con la profundidad necesaria, como Salud y Perfil de Usuario. En el primer caso, se cuenta con una ontología que estructura la terminología médica existente, UMLS. Para enriquecer el modelo Perfil de Usuario se tiene previsto integrar con un modelo que está siendo elaborado por otro grupo de trabajo del proyecto CYTED-SALUS.

Bibliografía

[AH02] Grigoris Antoniou, Frank van Harmelen. Web Ontology Language: OWL. In Handbook on Ontologies in Information Systems, pages 67–92, 2003.

[AJ94] Ameen Abu-Hanna, Wouter Jansweijer. Modeling Domain Knowledge Using Explicit Conceptualization. IEEE Expert: Intelligent Systems and Their Applications, Volume 9, Issue 5 (October 1994), Pages: 53 – 64, Year of Publication: 1994, ISSN:0885-9000.

[BD05] Joachim Baumeister and Dietmar Seipel. Smelly Owls – Design Anomalies in Ontologies. 18th Intl. Florida Artificial Intelligence Research Society Conference (FLAIRS), pp 251-220, 2005.

[BEFMSW06] Jos de Bruijn, Marc Ehrig, Cristina Feier, Francisco Martín-Recuerda, François Scharffe, Moritz Weiten. Ontology mediation, merging and aligning. Semantic Web Technologies. Published Online: 3 Jul 2006. Editor(s): John Davies, Rudi Studer, Paul Warren. Print ISBN: 9780470025963 Online ISBN: 9780470030332 DOI: 10.1002/047003033X Copyright © 2006 John Wiley & Sons, Ltd.

[BGM05] Janez Brank, Marko Grobelnik, Dunja Mladenić. A survey of Ontology Evaluation Techniques. In Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2005), Ljubljana, Slovenia, 2005.

[D04] William H. DuBay. The Principles of Readability. Costa Mesa, CA: Impact Information, 2004.

[DMT07] Fulvio D'Antonio, Michele Missikoff, Francesco Taglino. Formalizing the OPAL eBusiness ontology design patterns with OWL. Third International Conference on Interoperability for Enterprise Applications and Software, I-ESA (2007)

[ED08] Paula C. Engelbrecht, Itiel E. Dror. How Psychology and Cognition can inform Creation of Ontologies in Semantic Technologies. 18th European-Japanese Conference on Information Modelling and Knowledge Bases (EJC2008), Tsukuba, Japan, 2008.

[ES04] Mark Ehrig, York Sure. Ontology mapping: an integrated approach. In Bussler, C., Davis, J., Fensel, D. and Studer, R. (Eds), Proceedings of the First European Semantic Web Symposium, Volume 3053 of Lecture Notes in Computer Science, Springer Verlag, Heraklion, Greece, pp. 76-91, 2004.

[FA08] Muhammad Fahad, Muhammad Abdul Qadir. A Framework for Ontology Evaluation. Supplementary Proceedings of the 16th International Conference on Conceptual Structures, ICCS 2008, Toulouse, France, July 7-11, 2008.

[FGJ97] Mariano Fernández, Asunción Gómez-Pérez, Natalia Juristo. Methontology: From Ontological Art Towards Ontological Engineering. In Proceedings of the AAAI97 Spring Symposium Series on Ontological Engineering, 1997.

[FGPP99] Mariano Fernández López, Asunción Gómez-Pérez, Juan Pazos Sierra, Alejandro Pazos Sierra. Building a Chemical Ontology Using Methontology and the Ontology Design Environment. IEEE Intelligent Systems, Vol.14, No.1, Special Issue on Ontologies, January/February 1999, pp. 37-46.

[FGV96] Mariano Fernández, Asunción Gómez-Pérez, Antonio J. de Vicente. Towards a Method to Conceptualize Domain Ontologies. Proceedings of Workshop on Ontological Engineering/ECAI; Budapest, Hungary, August 1996.

[F99] Mariano Fernández López. Overview Of Methodologies For Building Ontologies. In: Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods, Stockholm, Sweden, August 2, 1999.

[G93] Thomas R. Gruber. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. In Formal Ontology in Conceptual Analysis and Knowledge Representation, edited by Nicola Guarino and Roberto Poli, Kluwer Academic Publishers, in press. Substantial revision of paper presented at the International Workshop on Formal Ontology, March, 1993, Padova, Italy. Available as Technical Report KSL 93-04, Knowledge Systems Laboratory, Stanford University.

[G98] Nicola Guarino. Formal Ontology and Information Systems. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15.

[G99] Asunción Gómez-Pérez. Evaluation of taxonomic knowledge in ontologies and knowledge bases. In Proceedings of the 12th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop, Banff, Alberta, Canada, 1999.

[GCCL05] Aldo Gangemi, Carola Catenacci, Massimiliano Ciaramita, Jos Lehmann. A theoretical framework for ontology evaluation and validation.

In Semantic Web Applications and Perspectives (SWAP) – 2nd Italian Semantic Web Workshop – 2005.

[GCG94] Nicola Guarino, Massimiliano Carrara, Pierdaniele Giaretta. Formalizing Ontological Commitments. In Proceedings of National Conference on Artificial Intelligence (AAAI-94). Seattle, Morgan Kaufmann, 1994.

[GF95] Michael Grüninger, Mark S. Fox. Methodology for the Design and Evaluation of Ontologies. In J. Brown and D. O’Sullivan, editors, Reengineering the Enterprise, pages 83-98. Chapman and Hall, 1995.

[GFC03] Asunción Gómez-Pérez, Mariano Fernández-López, Oscar Corcho. Ontological Engineering. Springer Verlag, 2002/2003.

[GHJV94] Erich Gamma, Richard Helm, Ralph Johnson, John Vlissides. Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley, Reading, Massachusetts, 1995.

[GJP95] Gómez-Pérez A, Juristo N, Pazos J. Evaluation and assessment of knowledge sharing technology. In: Mars N (ed) Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing (KBKS’95). University of Twente, Enschede, The Netherlands. IOS Press, Amsterdam, The Netherlands, pp 289-296, 1995.

[GPS98] Aldo Gangemi, Domenico M. Pisanelli, Geri Steve. Ontology Integration: Experiences with Medical Terminologies. Proceedings of the First International Conference (FOIS’98), June 6-8, Trento, Italy. Escrito por N. Guarino. Publicado por IOS Press, 1998 ISBN 9051993994, 9789051993998.

[GSG96] Aldo Gangemi, Geri Steve, Fabrizio Giacomelli. ONIONS: An Ontological Methodology for Taxonomic Knowledge Integration. Proc. ECAI96 Workshop on Ontological Engineering, 1996. Reparto Informatica Medica, Istituto Tecnologie Biomediche, CNR, Roma, Italy.

[GW06] Erwin W. Van Geenen, Cilia M. N. Witteman. How experts reason: the acquisition of expert’s knowledge structures. The Knowledge Engineering Review, Vol. 21:4, 335-344. ©2006 Cambridge University Press, doi:10.1017/S0269888906000968 Printed in the United Kingdom.

[GWG06] Andreas Gruber, Rupert Westenthaler, Eva Gahleitner. Supporting domain Experts in creating formal knowledge models (ontologies). In 6th International Conference on knowledge management, Tochtermann Klaus und Maurer Hermann, pages 252–260, Graz, Austria, 2006.

[HSGMCSS05] Jens Hartmann, Peter Spyns, Alain Giboin, Diana Maynard, Roberta Cuel, Mari Carmen Suárez-Figueroa, York Sure. D1.2.3 Methods for ontology evaluation. EU-IST Network of Excellence (NoE) IST-2004-507482 KWEB Deliverable D1.2.3 (WP 1.2). This document is part of a research project funded by the IST Programme of the Commission of the European Communities as project number IST-2004-507482.

[JB06] Ernesto Jimenez-Ruiz, Rafael Berlanga. A View-Based methodology for Collaborative Ontology Engineering: An Approach for Complex Applications (VIMethCOE). Proceedings of the 15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises. Pages 376-381. Year of Publication: 2006. ISBN ~ ISSN: 1524-4547, 0-7695-2623-3.

[JBR99] I. Jacobson, G. Booch, J. Rumbaugh. The Unified Software Development Process. Addison Wesley, 1999.

[JM08] Mustafa Jarrar, Robert Meersman. Ontology Engineering – The DOGMA approach. Book Chapter (Chapter 3). In Advances in Web Semantics I. Volume LNCS 4891, Springer. 2008.

[KR70] Werner Kunz, Horst W. J. Rittel. Issues as Elements of Information Systems. Working Paper 131 Center for Planning and Development Research. University of California, USA. (1970).

[KV06] Konstantinos Kotis, George A. Vouros. Human centered ontology engineering: The HCOME methodology. Knowledge and Information Systems 10 109-131 – Springer, 2006.

[L02] Adolfo Lozano Tello. Métrica de Idoneidad de Ontologías. Ph.D. thesis. Departamento de Informática. Universidad de Extremadura, 2002.

[LG04] Adolfo Lozano-Tello, Asunción Gómez-Pérez. ONTOMETRIC: a method to choose the appropriate ontology. Journal of Database Management. Special Issue on Ontological analysis, Evaluation, and Engineering of Business Systems Analysis Methods. Volumen: 15(2). Abril-June 2004.

[NMN09] Antonio De Nicola, Michele Missikoff, Roberto Navigli. A software engineering approach to ontology building. Information Systems, 34(2), Elsevier, 2009, pp 258-275.

[PM01] Helena Sofia Pinto, João P. Martins. A Methodology for Ontology Integration. Proceedings of the 1st international conference on Knowledge Capture K-CAP01 – 2001.

[PLW02] L. Pipino, Y. Lee, R. Wang. Data quality assessment. Communications of the ACM 4 (2002) 211-218.

[PST04] Helena Sofia Pinto, Steffen Staab, Christoph Tempich. Diligent: towards a fine-grained methodology for Distributed, Loosely-controlled and evolving engineering of ontologies. In European conference on Artificial Intelligence: 2004. Valencia, Spain; 2004: 393-397.

[SAAHSWOO] Guus Schreiber, Hand Akkermans, Anjo Anjewierden, Robert de Hoog, Nigel Shadbolt, Walter van den Velde, Bob Wielinga. Knowledge Engineering and Management: The CommonKADS Methodology. Publicado por MIT Press, 2000 ISBN 0262193000, 780262193009.

[SE08] Paul R. Smart, Paula C. Engelbrecht. An Analysis of the Origins of Ontology Mismatches on the Semantic Web. In: 16th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2008), 29th September-3rd October 2008, Acitrezza, Catania, Italy. In: 16th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2008), 29th September-3rd October 2008, Acitrezza, Catania, Italy.

[SPKR96] Bill Swartout, Ramesh Patil, Kevin Knight, Tom Russ. Toward Distributed Use of Large-Scale Ontologies. In Proc. the 10th International Workshop on Knowledge Acquisition, Modeling and Mangement (KAW'96), Banff, Canada, November 9-14, 1996.

[SRVM95] Speel H., Raalte F., Vet P., Mars N. Scalability of the Performance of Knowledge Representation Systems. In: Mars N (ed) Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing (KBKS'95). University of Twente, Enschede, The Netherlands. IOS Press, Amsterdam, The Netherlands, pp 173-183.

[SSSS00] Steffen Staab, Hans-Peter Schnurr, Rudi Studer, York Sure. Knowledge Processes and Ontologies. IEEE Intelligent Systems 16(1):26-34, 2000.

[TPSS05] Christoph Tempich, H. Sofia Pinto, York Sure, Steffen Staab. An Argumentation Ontology for Distributed, Loosely-controlled and evolving Engineering processes of oNTologies (DILIGENT). In Bussler, C., Davies, J., Fensel, D., and Studer, R., editors, *Second European Semantic Web*

Conference, ESWC 2005, LNCS, Heraklion, Crete, Greece. Springer. to appear.

[UG96] Mike Ushold, Michael Grüninger. Ontologies: Principles, Methods y Applications. *Knowledge Engineering Review* 11(2):93-155, 1996.

[UKMZ97] Mike Ushold, Martin King, Stuart Moralee, Yannis Zorgios. The Enterprise Ontology. *The Knowledge Engineering Review* 13(1):31-89, 1997.

[VPTS05] Denny Vrandečić, Sofia Pinto, Christoph Tempich, York Sure. The DILIGENT knowledge processes. *Journal of Knowledge Management*, 9(5):85–96, 2005.

[WG01] Christopher Welty, Nicola Guarino. Supporting ontological analysis of taxonomic relationships. *Data & KnowledgeEngineering* 39(1):51–74, 2001.

Referencias Web

[WR1] <http://www.alzheimermed.com.br/>

Fecha última visita: 02/11/2009

[WR2] <http://www.hon.ch/>

Fecha última visita: 15/10/2009

[WR3] http://www.portalesmedicos.com/web_interes_sanitario/index.htm

Fecha última visita: 15/10/2009

[WR4] <http://wma.comb.es/>

Fecha última visita: 15/10/2009

[WR5] <http://www.aifb.kit.edu/web/OntoWeb/en>

Fecha última visita: 29/11/2009

[WR6] <http://www.neon-project.org/web-content/>

Fecha última visita: 29/11/2009