

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

FACULTAD DE CIENCIAS E INGENIERÍA



PONTIFICIA
**UNIVERSIDAD
CATÓLICA**
DEL PERÚ

USO DE INFERENCIA BASADA EN ONTOLOGÍAS PARA DAR SOPORTE
AL DIAGNÓSTICO VETERINARIO.

TESIS PARA OPTAR POR EL TÍTULO DE INGENIERO INFORMÁTICO, QUE PRESENTAN LOS
BACHILLERES:

DIEGO ANDRÉS SALAS GUILLÉN

JACKLIN DEL ROCIO GONZÁLES MACEDA

ASESOR: DR. ANDRÉS MELGAR SASIETA

Lima, noviembre de 2014

RESUMEN

La salud a nivel mental y psicológico que nos brindan los animales domésticos es una necesidad poco valorada. La medicina veterinaria aplica conocimientos científicos al mejoramiento de la salud animal y al control enfermedades transmisibles al hombre. Las clínicas veterinarias, en el contexto peruano, no cuentan con suficiente soporte tecnológico para el proceso de diagnóstico. De la revisión del estado del arte se sabe que son escasos los proyectos de investigación realizados en este dominio específico. Existen herramientas de diagnóstico en el dominio de la medicina humana, varios ejemplos hacen uso de la representación del conocimiento por medio de ontologías y de sistemas de inferencia que trabajan con esta representación.

En vista de los problemas presentados se plantea la interrogante. ¿Cómo dar soporte al diagnóstico de enfermedades en el dominio de la veterinaria en el contexto peruano, específicamente en Lima metropolitana?. Este es un proyecto de investigación aplicada en el área de la inteligencia artificial motivado por un interés personal en la representación de conocimiento, la simulación de procesos mentales racionales y la medicina veterinaria. Se plantea aplicar conceptos de ingeniería del conocimiento para contribuir a alcanzar el estado ideal, trata del uso de inferencia basada en ontologías para el diagnóstico en enfermedades de animales.

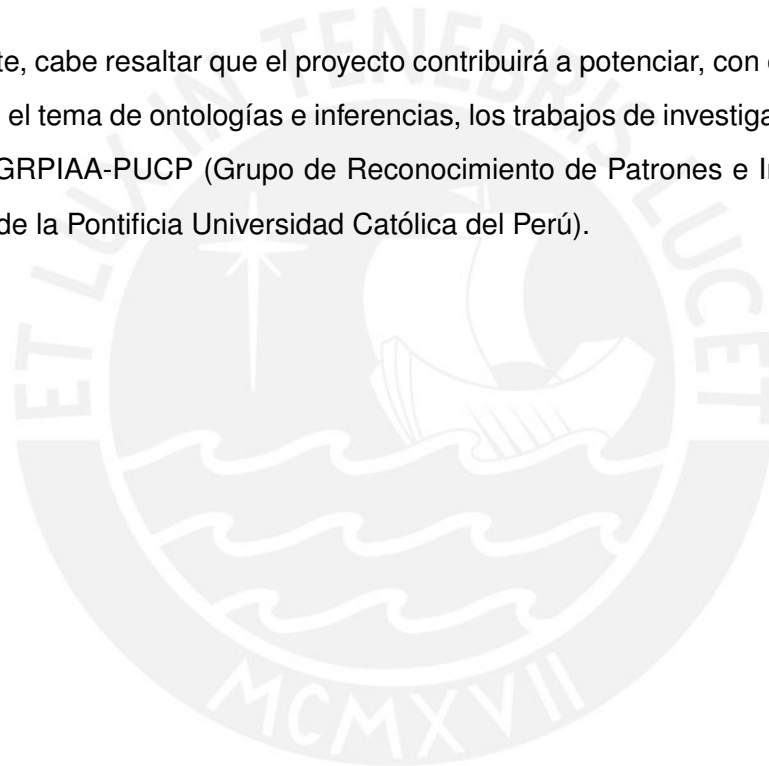
Una ontología encapsula y uniformiza el conocimiento de un dominio de interés y se puede utilizar para resolver problemas en este dominio. Las ontologías permiten un alto grado de expresividad, estas permiten representar conocimiento complejo. El conocimiento, experiencia y criterio de un veterinario son habilidades complejas de replicar, por lo que no es posible hacerlo con exactitud. Una ventaja de las ontologías es que permiten representar supuestos, metas, hipótesis y predicciones sobre un dominio específico.

El principal objetivo de la investigación es aplicar inferencia basada en ontologías para dar soporte al proceso de diagnóstico encapsulando el conocimiento del especialista en una base de información que de diagnóstico consultable. Mediante herramientas de inferencia el experto podrá extraer información filtrada en base a los síntomas observados.

De esta manera obtendrá un abanico de posibilidades que reducirán significativamente el esfuerzo del veterinario al momento de dar un diagnóstico. Se trata de abstraer el conocimiento del veterinario, organizarlo y acceder a este por medio de herramientas que simulen el razonamiento.

Esto reducirá el tiempo invertido por el veterinario en revisión de bibliografía, remembranza y análisis de posibilidades. Es como ofrecer, para un determinado problema, una limitada posibilidad de opciones de solución entre las cuales se encuentra respuesta acertada con una probabilidad bastante alta. Se perfila mucho más sencilla y rápida la labor de diagnóstico frente a esta alternativa.

Finalmente, cabe resaltar que el proyecto contribuirá a potenciar, con el aporte de investigación en el tema de ontologías e inferencias, los trabajos de investigación realizados por el grupo GRPIAA-PUCP (Grupo de Reconocimiento de Patrones e Inteligencia Artificial Aplicada de la Pontificia Universidad Católica del Perú).



*Dedicado a nuestras familias
por el apoyo brindado a lo largo de la carrera.
Por confiar en nosotros y hacer un gran esfuerzo
para darnos la oportunidad salir adelante.*



Agradecemos al Doctor Andrés Melgar Sasieta por su asesoría y dirección en el trabajo de investigación. Por la oportunidad de realizar investigación en temas innovadores de la rama de inteligencia artificial. Y a los doctores Josmel Pacheco y Milagros Montesinos por su colaboración en el proyecto.



Índice General

1. Capítulo 1	1
1.1. Definición del Problema	1
1.1.1. Problemática	1
1.1.1.1. Introducción	1
1.1.1.2. Situación Actual	2
1.1.1.3. Situación Ideal	3
1.1.2. Objetivo General	4
1.1.3. Objetivos Específicos	4
1.1.3.1. Objetivo Específico 1	4
1.1.3.2. Objetivo Específico 2	4
1.1.3.3. Objetivo Específico 3	4
1.1.3.4. Objetivo Específico 4	4
1.1.4. Resultados Esperados	4
1.1.4.1. Resultado 1 para el objetivo 1	4
1.1.4.2. Resultado 2 para el objetivo 1	5
1.1.4.3. Resultado 3 para el objetivo 2	5
1.1.4.4. Resultado 4 para el objetivo 2	5
1.1.4.5. Resultado 5 para el objetivo 2	5
1.1.4.6. Resultado 6 para el objetivo 3	5
1.1.4.7. Resultado 7 para el objetivo 3	6
1.1.4.8. Resultado 8 para el objetivo 3	6
1.1.4.9. Resultado 9 para el objetivo 4	6
1.1.4.10. Resultado 10 para el objetivo 4	6
1.2. Herramientas, Métodos, Metodologías y Procedimientos	6
1.2.1. Introducción	6
1.2.2. Herramientas	8
1.2.3. Métodos y Procedimientos	10
1.2.4. Metodologías	10
1.3. Alcance	11
1.3.1. Limitaciones	11
1.3.2. Riesgos	12

1.4. Justificativa y viabilidad del proyecto	13
1.4.1. Justificativa	13
1.4.2. Viabilidad	14
2. Capítulo 2	15
2.1. Marco Conceptual	15
2.1.1. Objetivo del Marco Conceptual	15
2.1.2. Ontología	16
2.1.2.1. Definición	16
2.1.2.2. Tipos de Ontologías	19
2.1.2.3. Componentes de las Ontologías	20
2.1.2.4. Importancia de las ontologías	21
2.1.2.5. Aplicaciones	22
2.1.3. Raciocinio	24
2.1.3.1. Definición	24
2.1.3.2. Ejemplos de Raciocinio Usando Ontologías	25
2.1.4. Inferencia	26
2.1.4.1. Definición	26
2.1.4.2. Uso de las Ontologías en las inferencias	26
2.1.5. Conclusión	27
2.2. Estado del arte	28
2.2.1. Objetivos del estado del arte	28
2.2.2. Método usado en la revisión del estado del arte	28
2.2.3. Resultados de la revisión sistemática	29
2.2.4. Conclusiones sobre el estado del arte	31
3. Capítulo 3	33
3.1. Introducción	33
3.2. Resultado Esperado 1: Informe de Herramientas Aplicables - Ontologías	33
3.2.1. OBO	34
3.2.2. OWL	34
3.2.3. Ontologías Aplicables	36
3.2.4. Conclusión	37
3.3. Resultado Esperado 2: Informe de Herramientas Aplicables - Inferencias	38

3.3.1. Inferencias sobre OWL	38
3.3.2. Jena	39
3.3.3. Pellet, Racer y FaCT	40
3.3.4. BaseVISor	41
3.3.5. Conclusión	41
4. Capítulo 4	42
4.1. Introducción	42
4.2. Resultado Esperado 3: Análisis de la Base de Conocimiento	42
4.2.1. Proceso de Diagnóstico	43
4.2.2. Catálogo de Enfermedades	43
4.2.3. Conclusión	44
4.3. Resultado Esperado 4: Diseño de la Base de Conocimiento	44
4.3.1. Esquema del Dominio	45
4.3.2. Estructura de Inferencias	47
4.3.3. Conclusión	48
4.4. Resultado Esperado 5: Construcción de la Base de Conocimiento	49
4.4.1. Sobre la herramienta Protégé	49
4.4.2. Definición de las Clases	50
4.4.3. Relaciones entre Clases	50
4.4.4. Inserción de Individuos	51
4.4.5. Relaciones entre Individuos	52
4.4.6. Conclusión	53
5. Capítulo 5	53
5.1. Introducción	53
5.2. Resultado Esperado 6: Análisis del Razonamiento para el Diagnóstico	54
5.2.1. Procesos Mentales	54
5.2.2. Diagnóstico Veterinario	55
5.2.3. Tarea de Diagnóstico Segun CommonKADS	56
5.2.4. Conclusión	56
5.3. Resultado Esperado 7: Diseño del Motor de Inferencia	56
5.3.1. Plantilla de Diagnóstico CommonKADS	57
5.3.2. Diagrama de Diseño	59

5.3.3. Implementación de cada Función de Inferencia	61
5.3.3.1. Cubrir	61
5.3.3.2. Seleccionar	61
5.3.3.3. Especificar	62
5.3.3.4. Verificar	62
5.3.4. Conclusión	63
5.4. Resultado Esperado 8: Construcción del Motor de Inferencia	63
5.4.1. Librería Jena	63
5.4.2. Acceso a las Clases, Individuos y Propiedades de la Ontología	64
5.4.3. Implementación de las Inferencias y Resultados	65
5.4.4. Conclusión	67
6. Capítulo 6	67
6.1. Introducción	67
6.2. Resultado Esperado 9: Construcción del Prototipo	67
6.2.1. Funcionalidades	67
6.2.2. Diseño	68
6.2.3. Conclusión	69
6.3. Resultado Esperado 10: Interfaz Gráfica	69
6.3.1. Diseño	69
6.3.1.1. Ventana de Inicio	69
6.3.1.2. Ventana de Inicio	70
6.3.1.3. Ventana de Evaluación de Síntomas	71
6.3.1.4. Ventana de Resultados	71
6.3.1.5. Mensajes	72
6.3.2. Flujo	72
6.3.3. Resultados	73
6.3.4. Conclusión	74
7. Capítulo 7	74
7.1. Discusión de Resultados	74
7.1.1. Introducción	74
7.1.2. Objetivo Específico 1:	75
7.1.3. Objetivo Específico 2	75

7.1.4. Objetivo Específico 3	76
7.1.5. Objetivo Específico 4	76
8. Capítulo 8	77
8.1. Conclusión	77



Índice de Figuras

1.	Ejemplo de taxonomía: <i>Yahoo Categories</i>	18
2.	Ejemplo de catálogo: <i>Amazon product catalog</i>	19
3.	Vista de UMLS tab, herramienta de acceso a la información de UMLS obtenido de http://protegewiki.stanford.edu/wiki/UMLS_Tab	19
4.	Arquitectura de Ontobroker [Decker et al., 1999].	27
5.	Estructura de inferencia para el método de diagnóstico [Schreiber, 2000].	37
6.	Imagen obtenida de la documentación de Jena [Foundation, 2014]	40
7.	Esquema del dominio de la veterinaria para el diagnóstico	46
8.	Imagen propia, adaptada del modelo de diagnóstico de CommonKADS [Schreiber, 2000]	47
9.	Definición de las clases <i>Disease</i> y <i>Symptom</i> en Protégé	50
10.	Definición de la relación <i>evidenceOf</i> en Protégé	51
11.	Creación de los individuos de la clase <i>Disease</i> en Protégé	51
12.	Creación de los individuos de la clase <i>Symptom</i> en Protégé	52
13.	Asignación de la relación <i>evidenceOf</i> entre el síntoma <i>Alopecia</i> y la enfermedad <i>Folliculitis</i> en Protégé	52
14.	Relación de los síntomas asociados a la enfermedad <i>Folliculitis</i> en Protégé	53
15.	Traducido de la especificación de la tarea de diagnóstico de la metodología CommonKADS [Schreiber, 2000]	57
16.	Diagrama de diseño construido por los autores.	60
17.	Pseudocódigo para la función cubrir.	61
18.	Pseudocódigo para la función seleccionar.	61
19.	Pseudocódigo para la función especificar.	62
20.	Pseudocódigo para la función verificar.	62
21.	Resultado de la función diagnóstico para síntoma <i>Lesions</i> y 1 enfermedad.	66
22.	Resultado de la función diagnóstico para síntoma <i>Lesions</i> y 3 enfermedades.	66
23.	Resultado de la función diagnóstico para síntoma <i>Lesions</i> y 5 enfermedades.	66
24.	Diagrama de diseño del prototipo elaborado por los autores.	68
25.	Ventana inicial de prototipo de la herramienta de diagnóstico	70
26.	Ventana de configuración de la herramienta de diagnóstico	70
27.	Ventana de aceptación y rechazo de síntomas	71

28. Ventana de resultados	71
29. Diversos mensajes de la herramienta	72
30. Diagrama de estados funcional de la herramienta de diagnóstico.	73

Índice de Tablas

1. Herramientas y Resultados	7
2. Riesgos y contingencia	12
3. Resultados de la revisión sistemática relacionados a la medicina humana. .	29



1. Capítulo 1

1.1. Definición del Problema

1.1.1. Problemática

1.1.1.1. Introducción

La salud a nivel mental y psicológico que nos brindan los animales domésticos es una necesidad poco valorada. Mantener saludables a las mascotas no solo es importante para evitar problemas con nuestra salud física sino también para nuestro bienestar psicológico. No deja de ser sorprendente como estos seres pueden ayudar a mantener la salud tanto física como psicológica, por lo que es importante agradecer su compañía con cada acto de vida. [Gómez, 2007].

La medicina veterinaria aplica conocimientos científicos y técnicos al mejoramiento de la salud animal y a la prevención y control de aquellas enfermedades de los animales transmisibles al hombre. Esta contribuye al bienestar físico y psicológico de las muchas personas que conviven con los animales. Los veterinarios poseen conocimiento de las técnicas para establecer diagnósticos acertados, instaurar tratamientos eficaces y de la misma manera criterio para tomar las medidas necesarias en la prevención de las enfermedades de animales que se transmiten al humano [Serrano, 2008].

El cuidado de mascotas es una preocupación principal en la mayoría de personas. Entre 1998 y el 2002, el mercado de alimentos y productos para el cuidado de mascotas en América Latina creció 5% [Vera Ramirez, 2009]. En el Perú, las veterinarias han diversificado los servicios que ofrecen, que van desde servicios de salud y alimentación hasta de peluquería, y son más las personas que consideran a sus mascotas como sus propios hijos, por lo que en la actualidad existen cientos de tiendas especializadas en Lima y en las principales ciudades del país [Vera Ramirez, 2009].

En el Perú, donde existen aproximadamente 4 millones de mascotas, una clínica veterinaria puede atender hasta 80 consultas diarias y se atienden aproximadamente a 100 mascotas [Vera Ramirez, 2009]. Las principales especies domésticas en el país son el gato y el perro [Vera Ramirez, 2009].

1.1.1.2. Situación Actual

En su mayoría, las clínicas veterinarias, en el contexto peruano, no cuentan con suficiente soporte tecnológico, en cuanto a tecnologías de información se refiere, a nivel de procesos de diagnóstico. De la revisión del estado del arte, las herramientas encontradas son escasas y altamente especializadas.

Esto se debe en parte a la escasez de herramientas informáticas orientadas al soporte de diagnóstico en veterinaria. Se puede ver en la revisión del estado del arte que son escasos los proyectos de investigación realizados en este dominio específico. Existen herramientas de diagnóstico en el dominio de la medicina humana, estas hacen uso de la representación del conocimiento por medio de ontologías y de sistemas de inferencia que trabajan con esta representación. Una ontología encapsula y uniformiza el conocimiento de un dominio de interés y se puede utilizar para resolver problemas en este dominio [Uschold et al., 1996].

Además, resulta bastante complicado analizar cada especie. Con un espectro tan amplio de enfermedades, el dominio de investigación se extiende y se hace complejo [Lian et al., 2012]. Las ontologías permiten un alto grado de expresividad, estas permiten representar conocimiento complejo [Studer et al., 2000]. Esta limitación es más reducida en el dominio de la medicina humana al tratarse de una sola especie.

El conocimiento, experiencia y criterio de un veterinario son habilidades complejas de replicar, por lo que no es posible hacerlo con exactitud. Existe un amplio espectro de técnicas de diagnóstico veterinario que varían de acuerdo al tipo de enfermedad. Una ventaja de las ontologías es que permiten representar supuestos, metas, hipótesis y predicciones sobre un dominio específico [Chandrasekaran et al., 1999].

El diagnóstico de las enfermedades suele ser bastante ambiguo porque los síntomas varían, se repiten en más de una enfermedad y dependen del progreso de esta [Houe et al., 2011]. Nuevamente las ontologías arrojan luz sobre posibles maneras de afrontar esta ambigüedad, estas permiten reducir la ambigüedad mediante leguajes formales basados en lógica descriptiva [Guarino, 1998].

1.1.1.3. Situación Ideal

En vista de los problemas presentados se plantea la interrogante. ¿Cómo dar soporte al diagnóstico de enfermedades en el de la veterinaria en el contexto peruano, específicamente en Lima metropolitana?. La informática ofrece distintas alternativas para afrontar el problema, esta investigación aplicará conceptos de ingeniería del conocimiento para contribuir a alcanzar el estado ideal.

Al intentar solucionar este problema se obtendrán los siguientes beneficios:

Brindar soporte al diagnóstico de enfermedades, lo que contribuirá a reducir el tiempo que le toma al veterinario concluir el tratamiento de un paciente. Esto se logrará realizando inferencias sobre una base de conocimiento basada en ontologías, simulando el proceso de razonamiento que sigue el médico veterinario durante el diagnóstico. Es decir, se expresará el conocimiento del veterinario mediante una ontología y se realizarán consultas a este conocimiento siguiendo una estructura que permita llegar a un resultado. Esta estructura dependerá del análisis que se realice del proceso racional detrás de la tarea de diagnóstico. El conocimiento y experiencia del médico se encapsulará en una base de conocimiento y un motor de inferencia permitirá simular el trabajo de asociación, distinción y deducción necesario para ofrecer un abanico de posibles enfermedades que permitan al veterinario afinar su diagnóstico.

De esta manera, se daría soporte al diagnóstico reduciendo la probabilidad de cometer errores y acelerando el proceso de atención. Esto además contribuirá a disminuir el periodo de malestar de la mascota y, por consiguiente, del dueño.

Finalmente, se deja abierta la posibilidad de utilizar los resultados de la investigación para explorar la alternativa de ofrecer servicios de diagnóstico en línea para reducir el costo de atención de enfermedades bastante comunes para los dueños de las mascotas. Es conveniente recalcar que este servicio médico no remplazará al médico veterinario, siempre será necesario realizar chequeos periódicos que garanticen la salud de las mascotas. Se puede entender de manera análoga a la consulta que realizamos al farmacéutico ante un resfriado común u otro malestar conocido, si el dueño tiene conocimiento de

enfermedades recurrentes de su mascota el servicio web podría servirle también de guía.

1.1.2. Objetivo General

Aplicar métodos y técnicas de inferencia basada en ontologías para dar soporte al diagnóstico en medicina veterinaria.

1.1.3. Objetivos Específicos

1.1.3.1. Objetivo Específico 1

Revisar los métodos y técnicas de inferencia basada en ontologías existentes para soportar el diagnóstico en el dominio de la medicina veterinaria.

1.1.3.2. Objetivo Específico 2

Encapsular el conocimiento de un médico veterinario, necesario para el diagnóstico de enfermedades de la piel en perros, en una base de conocimiento basada en ontologías.

1.1.3.3. Objetivo Específico 3

Utilizar técnicas y métodos de inferencia para simular el razonamiento de un médico veterinario sobre el conocimiento encapsulado en una base de conocimiento basada en ontologías.

1.1.3.4. Objetivo Específico 4

Reducir la ambigüedad inherente al diagnóstico en medicina veterinaria utilizando ontologías formales y herramientas de inferencia precisas, buscando un balance entre la expresividad de la ontología y su complejidad computacional.

1.1.4. Resultados Esperados

1.1.4.1. Resultado 1 para el objetivo 1

Informe de ontologías existentes y formatos de representación que se pueden adaptar al proyecto.

Se tendrá en cuenta dos opciones de formatos OWL y OBO-Format. Se revisarán ontologías de una base de conocimiento existente.

1.1.4.2. Resultado 2 para el objetivo 1

Informe de métodos de inferencia sobre ontologías aplicables al dominio y al problema planteado.

Se considerarán dos motores de inferencia que trabajan sobre ontologías, JENA y BaseVisor. Se evaluará un algoritmo propuesto por CommonKADS

1.1.4.3. Resultado 3 para el objetivo 2

Análisis de la información necesaria para el desarrollo de la base de datos de conocimiento que permita realizar el diagnóstico de enfermedades en las mascotas. Este incluye información sobre síntomas, enfermedades de la piel y gastrointestinales, y las relaciones entre estos conceptos.

1.1.4.4. Resultado 4 para el objetivo 2

Diseño de una base de datos incluyendo estructura, relaciones, campos, clases y relaciones. Esto basado en el análisis realizado para el resultado 3.

1.1.4.5. Resultado 5 para el objetivo 2

Construcción de una base de datos de conocimiento basada en una ontología existente, seleccionada en el resultado 1, que permita encapsular el conocimiento necesario para el diagnóstico de enfermedades de la piel seleccionadas. Esta se construirá sobre las especificaciones desarrolladas en los resultados 3 y 4.

1.1.4.6. Resultado 6 para el objetivo 3

Análisis de los procesos de razonamiento que sigue un médico veterinario para llegar al diagnóstico de una enfermedad. Esto incluye una especificación formal de las hipótesis,

planteamientos, descarte de ideas, filtros y relaciones entre síntomas y enfermedades.

1.1.4.7. Resultado 7 para el objetivo 3

Diseño del motor de inferencia incluyendo la mecánica a seguir, la definición de entradas y salidas, casos de prueba.

1.1.4.8. Resultado 8 para el objetivo 3

Construcción del motor de inferencia basado en los métodos de inferencia seleccionados en el resultado 2 que permita realizar inferencias sobre una base de conocimiento basada en ontologías. Este se construirá sobre las especificaciones desarrolladas en los resultados 6 y 7.

1.1.4.9. Resultado 9 para el objetivo 4

Herramienta piloto que integre la base de datos y el motor de inferencia desarrolladas en los resultados 5 y 8. La herramienta debe realizar sugerencias de enfermedades en base de síntomas especificados con un grado de certeza de por lo menos 70%. Este grado se determinará comparando los resultados en diagnóstico de médico veterinario y las enfermedades sugeridas como salida de la herramienta para un conjunto determinado de síntomas.

1.1.4.10. Resultado 10 para el objetivo 4

Interfaz gráfica simple que permita el ingreso y salida de datos para la prueba del piloto.

1.2. Herramientas, Métodos, Metodologías y Procedimientos

1.2.1. Introducción

El presente proyecto de investigación aplicada se desarrolla en el área de ciencias de la computación, particularmente en la subárea de inteligencia artificial en lo respectivo a representación del conocimiento. Las principales herramientas, que serán el núcleo de la investigación, son las ontologías y los métodos de inferencia sobre ontologías. Es-

tas comparten la visión de mundo de la web semántica. A continuación se detallarán, además de las principales herramientas a utilizar, metodologías y métodos auxiliares que permitirán alcanzar los resultados esperados.

Cuadro 1: Herramientas y Resultados

Resultado Esperado	Herramientas a usarse
RE1:Informe de ontologías.	Métodos cualitativos de obtención de información. Revisión de bibliografía. Revisión de ontologías existentes.
RE2:Informe de métodos de inferencia.	Métodos cualitativos de obtención de información. Revisión de bibliografía.
RE3:Análisis para el desarrollo de la base de conocimiento.	Métodos cualitativos de obtención de información. Metodología CommonKads. Clasificación de enfermedades de la piel en perros existentes de la bibliografía recomendada.
RE4:Diseño de la base de conocimiento.	Metodología CommonKads
RE5:Construcción de la base de conocimiento.	Ontologías(OWL,OBO). Protégé. Metodología CommonKads.
RE6:Análisis de los procesos de razonamiento.	Metodología CommonKads.
RE7:Diseño del motor de inferencia.	Inferencias basadas en reglas, API Jena
RE8:Construcción del motor de inferencia.	API Jena. Razonador Pellet. Lenguaje de programación Java.
RE9:Construcción del prototipo.	Lenguaje de programación Java. Metodología CommonKads.
RE10:Interfaz gráfica.	Lenguaje de Programación Java

1.2.2. Herramientas

Como se mencionó, el proyecto de investigación está centrado en dos herramientas principales, las ontologías y los métodos de inferencia sobre ontologías. Como parte de los resultados esperados se considera dos informes de comparación de herramientas. Estos permitieron seleccionar la herramienta más adecuada entre posibles candidatas. Estas se preseleccionaron durante la revisión del estado del arte.

La base de conocimiento se puede construir utilizando diferentes métodos de representación de conocimiento como lo son las reglas, representación semántica o una base de datos. El rol de las ontologías en el proceso de la ingeniería del conocimiento es facilitar la construcción de un modelo de dominio [Studer et al., 1998].

Se decidió hacerlo utilizando ontologías pues así se tiene una visión más ordenada, se puede agrupar los hechos en una estructura que permite tener una visión global de los elementos y sus interacciones. Otra ventaja es que son portables y facilitan el agregar conocimiento nuevo contribuyendo a la escalabilidad del sistema.

Además se cuenta con plantillas de desarrollo e información reutilizables que permiten ahorrar tiempo y esfuerzo en investigación, la metodología CommonKads ofrece plantillas para desarrollar sistemas para procesos intensivos en conocimiento. En el artículo «What are ontologies, and why do we need them?» mencionan que al tener compartida una representación de conocimiento con otros con necesidades similares se evita tener que replicar el proceso de análisis de conocimiento [Chandrasekaran et al., 1999]. De esta manera, las ontologías permiten un mejor orden y reducen considerablemente el tiempo requerido para el desarrollo de sistemas basados en conocimiento.

Para el caso de las ontologías se contaba con dos posibilidades para realizar la adaptación a las necesidades del proyecto.

La primera, el lenguaje OWL (*Ontology Web Language*). Este es un lenguaje diseñado para ser usado por aplicaciones que necesitan procesar el contenido de la información en vez de solo presentarla de manera comprensible [Van Harmelen and McGuinness, 2004].

OWL utiliza ontologías para representar de manera explícita el significado de términos en vocabularios y las relaciones entre estos. [Van Harmelen and McGuinness, 2004]. Esta herramienta ha sido utilizada en dos de los artículos relacionados con el proyecto que se rescataron en el cuadro 1. La herramienta MEDBOLI utiliza la herramienta Jena, una plataforma de código abierto para el procesamiento de ontologías que soporta varios formatos, entre estos OWL [Rodríguez et al., 2009].

La segunda opción es OBO (*Open Biological and Biomedical Ontologies*). La OBO-Foundry ofrece una serie de ontologías relacionadas a la medicina y la representación de organismos [Smith et al., 2007]. OBO aplica los principios base de GO (*Gene Ontology*), una ontología debe ser abierta, ortogonal e instanciada en una sintaxis bien especificada y diseñada para compartir un espacio común de interés [Smith et al., 2007]. Al tratarse de ontologías que giran entorno a la medicina, anatomía y procesos biológicos, podría encontrarse ontologías útiles para el proyecto. En el artículo PsyDis se menciona esta base de ontologías como una opción para la representación de información médica [Casado-Lumbreras et al., 2012].

Para el caso de las inferencias se trabajará con inferencias basadas en reglas. De acuerdo al artículo MEDBOLI (2009), para que el sistema permita el diagnóstico, el motor de inferencia deberá acceder a la base de conocimiento que contenga reglas expresadas en función de síntomas, enfermedades y relaciones entre estos [Rodríguez et al., 2009]. Esta es una posibilidad interesante, por otro lado, la herramienta Jena ofrece herramientas de inferencia sobre ontologías [Rodríguez et al., 2009]. Otra herramienta evaluada que trabaja inferencias basadas en reglas con soporte para OWL es baseVisor [Matheus et al., 2006]. La estructura de inferencias seguirá los lineamientos que la metodología CommonKADS propone para la tarea de diagnóstico.

Adicionalmente se utilizará el lenguaje de programación Java para el desarrollo de una interfaz gráfica simple.

1.2.3. Métodos y Procedimientos

Para la obtención de la información necesaria para el proyecto se utilizarán métodos cualitativos de recolección de datos como entrevistas, trabajo de campo e información histórica.

Se medirán los resultados utilizando dos procedimientos. Primero, se contrastará los resultados obtenidos por el prototipo contra información histórica de diagnóstico para determinar un porcentaje de efectividad. Segundo, se realizarán pruebas en paralelo comparando el diagnóstico sugerido por el prototipo con el de un veterinario. Esto permitirá determinar en que grado se cumplió los objetivos.

1.2.4. Metodologías

Para el proyecto, se empleará la metodología CommonKADS, la cual pertenece a la visión de mundo de la gestión de conocimiento y ofrece una metodología para la construcción de sistemas basados en conocimiento [Schreiber, 2000].

Las secciones a usar de la metodología seleccionada, debido a su importancia en el desarrollo de los resultados planteados, son las siguientes:

- Sección5: *Knowledge Model Components* -Como base conceptual para el desarrollo y la comprensión de los modelos de conocimiento.
- Sección6: *Template Knowledge Models* - Como base conceptual para el desarrollo y la comprensión de los modelos de conocimiento.
- Sección7: *Knowledge Model Construction* - Se seguirá las especificaciones de esta sección para la construcción de los modelos de conocimiento.
- Sección8: *Knowledge-Elicitation Techniques* - Se obtendrá una guía para la captura de conocimiento de esta sección.
- Sección14: *UML Notations Used in CommonKADS* - Se utilizará la notación UML sugerida por CommonKADS para los diagramas necesarios.

1.3. Alcance

A continuación se detallará el alcance del proyecto, justificando las limitaciones que existen para el dominio en cuestión y los riesgos a los que está sujeto.

El proyecto a desarrollar trata del uso de ontologías para representar el conocimiento y métodos de inferencia sobre ontologías para acceder al conocimiento con el fin de brindar soporte al diagnóstico en enfermedades de animales. Se centra en el desarrollo de una base de datos del conocimiento veterinario y el de un motor de inferencia que permita dar soporte al diagnóstico.

Para la realización de este proyecto, se ha identificado las limitaciones y alcance para que sea viable su desarrollo en el transcurso del periodo académico definido por la universidad.

El animal de estudio en el proyecto es el perro, animal doméstico más común en los hogares limeños, enfocándonos en las características de las enfermedades dérmicas.

El alcance está sujeto a las siguientes decisiones:

- Reducir el número de enfermedades a tratar limitandolas a un tipo.
- Empleo de fuentes modernas, de no más de 5 años de antigüedad.

1.3.1. Limitaciones

Los obstáculos y restricciones que limitaron el alcance son los siguientes:

- El número de enfermedades y variantes es demasiado amplio.
- Dificultad de identificación de los síntomas de los tipos de enfermedades.
- Falta de ejemplos de adaptación de herramientas, ontologías y métodos de inferencia, orientados al diagnóstico veterinario.
- Déficit en la comunicación del malestar del animal.
- Capacidad de procesamiento necesario para soportar el prototipo de sistema.

1.3.2. Riesgos

En la siguiente tabla se presentan los riesgos identificados para el proyecto, así como su probabilidad de ocurrencia, impacto en el proyecto y las medidas de contingencia que se tomarán en caso ocurran.

Cuadro 2: Riesgos y contingencia

Id	Riesgo	Probabilidad (0..1)	Impacto (0..1)	Contingencia
1	Las herramientas no alcanzan el grado de precisión esperado.	0.3	0.7	Tener disponibles herramientas opcionales. Realizar pruebas previas de resultados, tomando como referencia el conocimiento de un veterinario o de una base de datos de una veterinaria.
2	Pérdida de la información obtenida.	0.1	0.8	Manterner la información versionada y respaldada en backups periódicos.
3	Dificultades en la integración de la base de datos del conocimiento y el motor de inferencia.	0.4	0.5	Buscar opciones de interconexión intermedias.
4	No obtener algún resultado esperado en el tiempo planificado, retrasando el avance de todo el proyecto.	0.6	0.3	Realizar la planificación de las actividades considerando la duración aproximada de cada actividad en relación a la disponibilidad de tiempo para la fecha prevista. Tener un respaldo temporal para amortiguar retrasos.

Id	Riesgo	Probabilidad (0..1)	Impacto (0..1)	Contingencia
5	Reestructuración del análisis o diseño del prototipo por errores en estas etapas.	0.3	0.8	Tratar de cubrir todos los vacíos posibles durante la etapa de análisis y diseño. Adelantar pruebas de funcionalidad para detectar posibles cambios. Destinar el tiempo de respaldo a la reestructuración del prototipo.

1.4. Justificativa y viabilidad del proyecto

1.4.1. Justificativa

A continuación se presentarán los motivos de selección del tema de investigación, la motivación del proyecto y los beneficios que nos impulsan a culminar con éxito esta iniciativa.

Razones que motivan el proyecto de investigación aplicada en la sub-área de representación de conocimiento, en el área de la inteligencia artificial:

- Interés personal en la representación de conocimiento basada en ontologías y la simulación de procesos mentales racionales.
- Interés personal en la medicina veterinaria y el bienestar que brinda tanto a animales domésticos como a sus dueños.
- Beneficios psicológicos y fisiológicos de mantener a las mascotas saludables.

La labor del médico veterinario en el tratamiento de un paciente comienza con la revisión de los síntomas para asociarlos con problemas de salud generales (por ejemplo enfermedades de la piel o estomacales), revisar la documentación en búsqueda de posibles enfermedades, filtrar posibilidades, realizar exámenes para obtener mayor información y volver a contrastar con las posibilidades. Finalmente dar un diagnóstico y determinar el

tratamiento.

El principal objetivo de la investigación es aplicar las tecnologías en cuestión para dar soporte a este proceso de diagnóstico, encapsulando el conocimiento del especialista en una base de información que de diagnóstico consultable. Mediante herramientas de inferencia, el experto podrá extraer información filtrada en base a los síntomas observados. De esta manera, se obtendrá un abanico de posibilidades que reducirán significativamente el esfuerzo del veterinario al momento de dar un diagnóstico. Se trata de abstraer el conocimiento del veterinario, organizarlo y acceder a este por medio de herramientas que simulen el razonamiento.

Esto reducirá el tiempo invertido por el veterinario en revisión de bibliografía, remembranza y análisis de posibilidades. Es como ofrecer, para un determinado problema, una limitada posibilidad de opciones de solución entre las cuales se encuentra respuesta acertada con una probabilidad bastante alta. Se perfila mucho más sencilla y rápida la labor de diagnóstico frente a esta alternativa.

Adicionalmente se contribuye a la construcción de una base de conocimiento reutilizable en el dominio de la medicina veterinaria. Es parte del encanto de las ontologías el ser reutilizables y fácilmente adaptables. Esto permite que el conocimiento encapsulado pase a formar parte de un legado de conocimiento disponible en un mismo formato.

Finalmente, cabe resaltar que el proyecto contribuirá a potenciar, con el aporte de investigación en el tema de ontologías e inferencias, los trabajos de investigación realizados por el grupo GRPIAA-PUCP (Grupo de Reconocimiento de Patrones e Inteligencia Artificial Aplicada de la Pontificia Universidad Católica del Perú).

1.4.2. Viabilidad

A continuación se analizará la viabilidad del proyecto basada en tres aspectos generales: Recursos, Tiempo y Herramientas.

Respecto a los recursos disponibles para el desarrollo del proyecto se cuenta con ac-

ceso a la información necesaria para la construcción de la base de conocimiento, así como con el apoyo de un médico veterinario que facilitará la información y los medios necesarios por el lado de la medicina veterinaria. Se cuenta además con recursos materiales y humanos que provee la universidad, y principalmente GRPIAA, como acceso a información, equipos de cómputo, asesores e investigadores dispuestos a brindar su apoyo, ambientes de trabajo y acceso a impresoras. Por parte del equipo del proyecto se cuenta con dos computadoras personales y una de escritorio, todas con un desempeño satisfactorio y con disponibilidad económica para costear los gastos a lo largo del desarrollo del proyecto.

El proyecto se desarrollará en seis meses, duración del semestre académico, tiempo que se considera razonable para cumplir con los objetivos dado el alcance determinado para el proyecto. La disponibilidad de recursos humanos es alta, aproximadamente 40 horas-hombre semanales. En los anexos se muestra la programación de entrega de cada resultado.

Finalmente, las herramientas a utilizarse para el desarrollo del proyecto son de código abierto (OpenSource) por lo que no limitarán económicamente el desarrollo del proyecto. Por el lado del manejo de las mismas, los conceptos aprendidos a lo largo de la carrera universitaria facilitan en gran medida el aprendizaje de herramientas nuevas y, teniendo en cuenta la sólida base conceptual, se considera que la curva de aprendizaje no se extenderá mucho en el tiempo. Se espera que la motivación personal de los gestores y desarrolladores del proyecto impulse significativamente el progreso del mismo.

2. Capítulo 2

2.1. Marco Conceptual

2.1.1. Objetivo del Marco Conceptual

El presente proyecto plantea utilizar un motor de inferencias para obtener información útil para el diagnóstico de enfermedades de una base de datos de conocimiento basada en ontologías en el dominio de la medicina veterinaria. A continuación se presentan definiciones útiles para la comprensión del proyecto. Se abordarán los conceptos de

Ontología, Raciocinio e Inferencia para luego establecer relaciones entre estos. Esto se realizará mediante una revisión sistemática de estudios relacionados a los conceptos. Se intentará dar respuesta a preguntas como ¿Qué es una ontología?, ¿Qué es raciocinio?, ¿Qué es inferencia?, ¿Cuál es la importancia de estos conceptos?, ¿Cómo se relacionan entre sí?. Sobre la información recopilada, revisada y delimitada se realizará un escaneo en busca de la respuesta a estas preguntas. De esta manera se irá enlazando la información encontrada de manera que se pueda estructurar un marco de referencia que facilite la comprensión de los conceptos relevantes para el proyecto.

2.1.2. Ontología

2.1.2.1. Definición

Dentro de la diversa bibliografía que gira en torno al estudio y aplicación de las ontologías en el área de ingeniería del conocimiento e inteligencia artificial, el concepto de ontología ha sido definido desde varias perspectivas por distintos autores. A continuación se presentarán diversos puntos de vista con el objetivo de identificar los conceptos comunes.

Una ontología es una especificación de una conceptualización [Gruber, 1995]. Una conceptualización es una vista abstracta y simplificada del mundo que queremos representar con algún propósito. [Gruber, 1995]. Para el conocimiento de un dominio específico, el conjunto de objetos que pueden ser representados se denomina universo de discurso; estos objetos, y las relaciones entre los mismos, se reflejan en el vocabulario representacional con el que un programa basado en conocimiento representa el conocimiento [Gruber, 1995].

Una ontología está referida al entendimiento compartido de ciertos dominios de interés [Uschold et al., 1996]. Comprende el concepto de conceptualización, es decir, una visión global respecto a un dominio, así como un conjunto de conceptos, definiciones y las relaciones entre estos. [Uschold et al., 1996].

Desde el punto de vista de la filosofía, la ontología es el estudio de las cosas que existen. [Chandrasekaran et al., 1999]. Para la inteligencia artificial, se refiere a un vocabulario

especializado para cierto dominio del conocimiento, en esencia, las conceptualizaciones que este vocabulario define [Chandrasekaran et al., 1999]. Se podría cambiar de idioma a la ontología sin afectar la conceptualización [Chandrasekaran et al., 1999]. Identificar el vocabulario y las conceptualizaciones requiere un análisis exhaustivo de los tipos de objetos y relaciones de un dominio [Chandrasekaran et al., 1999].

Una ontología se refiere a un artefacto de ingeniería que consiste en un vocabulario usado para describir cierta realidad acompañado de presunciones respecto a su significado [Guarino, 1998]. Este conjunto de suposiciones tiene, por lo general, la forma de una teoría lógica de primer orden (lógica de predicados) [Guarino, 1998].

Otra definición resaltante de ontología es la siguiente:

“Una ontología es una especificación formal y explícita de una conceptualización compartida” [Studer et al., 1998].

Una conceptualización se refiere a un modelo abstracto de cierto fenómeno perteneciente al mundo habiendo identificado los conceptos relevantes de ese fenómeno [Studer et al., 1998]. Explícita quiere decir que se define los conceptos y restricciones de manera declarativa [Studer et al., 1998]. Formal se refiere a que la ontología esta hecha para ser entendida por un computador [Studer et al., 1998]. Compartida quiere decir que hay un consenso en el conocimiento que esta captura [Studer et al., 1998].

Habiendo revisado las distintas formas de definir el término ontología, se ha rescatado conceptos comunes entre los autores los cuales nos han permitido construir una definición general.

Los autores concuerdan en que parte de la definición de una ontología tiene que ver con el concepto de conceptualización. La conceptualización es la representación abstracta de un mundo, al que se le conoce como dominio. Esta representación constituye un modelo lógico de los elementos y relaciones entre estos para un dominio respectivo. Es decir, es una estructura organizada del conocimiento. Una ontología consiste en una especificación explícita de una conceptualización. Esta define un vocabulario de términos

para expresar una conceptualización de manera formal. Este vocabulario encapsula la conceptualización y debe reflejarla independientemente de la lengua en el que este se exprese. El vocabulario en la ontología puede estar expresado con distintos grados de formalidad, pero se suele utilizar una teoría lógica de primer orden.

Una ontología comprende una conceptualización que implica un consenso entre los interesados en un dominio respectivo. Esta nace del análisis exhaustivo de los elementos y relaciones existentes en un dominio y de conjeturas consensuadas respecto al mismo. De este modo, la ontología actúa como un lenguaje común entre los agentes que hacen uso de esta.

Noy presenta como ejemplos de ontologías:

- Taxonomías en la Web: *Yahoo! categories*. [Noy and McGuinness, 2001]

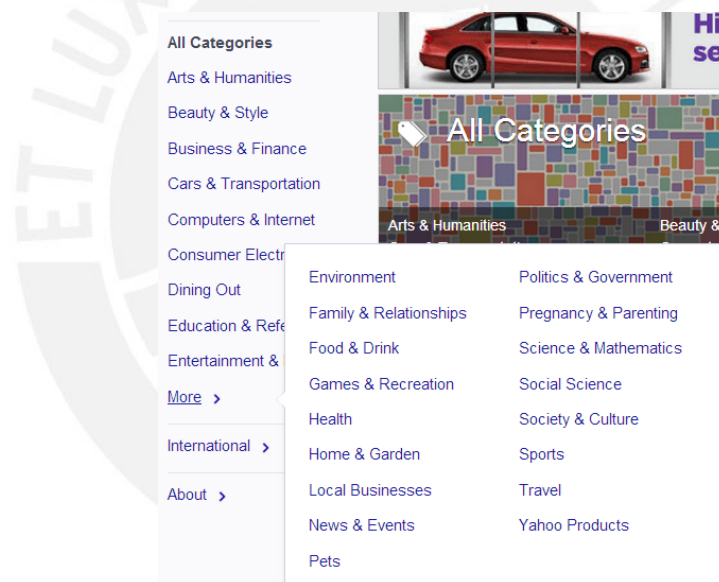


Imagen 1: Ejemplo de taxonomía: *Yahoo Categories*

- Catálogos de venta en línea: *Amazon product catalog*. [Noy and McGuinness, 2001]

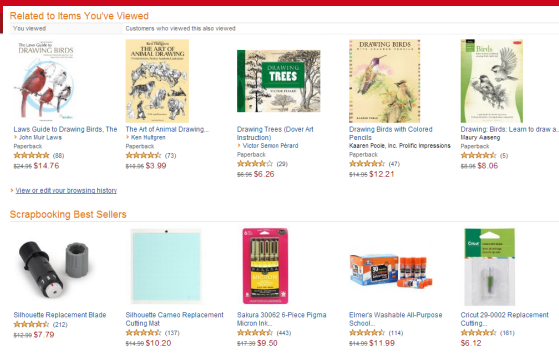


Imagen 2: Ejemplo de catálogo: *Amazon product catalog*

- Dominio específico de la terminología estándar: Sistema Unificado de Lenguaje Médico (UMLS), UNSPSC-terminología de los productos y servicios. [Noy and McGuinness, 2001]

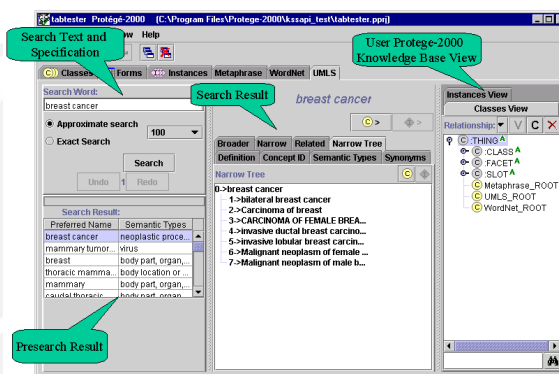


Imagen 3: Vista de UMLS tab, herramienta de acceso a la información de UMLS obtenido de http://protegewiki.stanford.edu/wiki/UMLS_Tab

2.1.2.2. Tipos de Ontologías

Dependiendo del nivel de generalidad, se pueden identificar diferentes tipos de ontologías que cumplen determinados roles en la construcción de un sistema basado en el conocimiento [Studer et al., 2000].

El primero, las ontologías de dominio, se refiere a las ontologías que reflejan el conocimiento de un dominio específico [Studer et al., 2000]. Por ejemplo, una ontología que contenga información sobre vinos, uvas, productores de vinos y las relaciones correspondientes.

El segundo, ontologías genéricas o de sentido común, tratan de replicar conocimiento sobre las nociones y conceptos básicos del mundo en general [Studer et al., 2000]. Por ejemplo una ontología que modele las relaciones entre objetos en el tiempo y el espacio.

El tercer tipo, ontologías de representación no están ligadas a ningún dominio en particular, representan entidades sin una utilidad específica, pero adaptables para cubrir distintas necesidades, por ejemplo, una conceptualización basada en marcos y ranuras [Studer et al., 2000].

2.1.2.3. Componentes de las Ontologías

El conocimiento de un dominio consiste en la información estática principal de los objetos del conocimiento dentro del dominio [Corcho and Gómez-Pérez, 2000]. Se puede especificar una ontología utilizando cinco componentes: conceptos, relaciones, funciones, axiomas e instancias [Corcho and Gómez-Pérez, 2000].

Conceptos: Son todo aquello sobre que pueda ser descrito, y cuya descripción sea relevante. Se les conoce también como clases. Pueden ser abstractos o concretos, elementales o compuestos, reales o ficticios [Corcho and Gómez-Pérez, 2000].

Relaciones: Interacción entre los conceptos de un dominio [Corcho and Gómez-Pérez, 2000].

Funciones: Las funciones son un tipo especial de relaciones en las que el último parámetro es único para cada posible conjunto de valores del resto de parámetros [Corcho and Gómez-Pérez, 2000].

Axiomas: Modelan enunciados que son siempre verdaderos. Su objetivo es restringir la información y permitir que se deduzca nueva información, también se utilizan para verificar si una ontología es correcta [Corcho and Gómez-Pérez, 2000].

Instancias: Representan los elementos de un concepto respectivo. Adicionalmente se

define como un hecho una relación entre elementos. Tanto a los hechos como a las instancias se les conoce como “*individuals*” [Corcho and Gómez-Pérez, 2000].

Tomando el ejemplo del dominio de la vinería de Noy:

- Conceptos: Cada vino tiene características como color, contenido de azúcar, sabor, etc. [Noy and McGuinness, 2001]
- Relaciones: Un vino puede estar relacionado con un productor y con un tipo de uva. [Noy and McGuinness, 2001]
- Funciones: Un vino puede tener un precio, el cual depende del tipo de vino y su productor.
- Axiomas: Se puede afirmar que: “Todo vino tiene como ingrediente principal a la uva”.
- Instancias: El vino “Chateau Lafite”. [Noy and McGuinness, 2001]

2.1.2.4. Importancia de las ontologías

Existen buenas razones para desarrollar ontologías. Las ontologías permiten:

Compartir un entendimiento común sobre la estructura de la información, entre personas y entre los agentes de software [Noy and McGuinness, 2001].

La reutilización del dominio del conocimiento, evitando reinventar todo y así introducir estándares [Noy and McGuinness, 2001].

Hacer suposiciones explícitas, haciendo más fácil cambiar las suposiciones del dominio (considerar una base genérica de conocimiento), así como también hace más fácil el entender y actualizar la información [Noy and McGuinness, 2001].

Partiendo del hecho de que una ontología permite representar una conceptualización proporcionando un vocabulario sin el cual no sería posible representar el conocimiento, afirman que la ontología conforma el núcleo de todo sistema basado en conocimiento [Chandrasekaran et al., 1999]. Realizar un análisis ontológico esclarece la estructura

del conocimiento para cualquier dominio, un buen análisis permite a las ontologías trabajar de manera coherente y cohesiva [Chandrasekaran et al., 1999].

El conocimiento representado por una ontología puede ser compartido [Chandrasekaran et al., 1999]. La ontología captura la estructura conceptual de un dominio en un lenguaje que permita representar esta estructura [Chandrasekaran et al., 1999]. Este lenguaje puede ser compartido con otros que tengan necesidades similares en el mismo dominio, eliminando, de esta manera, la necesidad de un análisis exhaustivo [Chandrasekaran et al., 1999].

2.1.2.5. Aplicaciones

Distintos autores rescatan aspectos destacables sobre su utilidad y aplicaciones. A continuación se presentan ejemplos de aplicaciones que algunos autores reconocen en distintas áreas de la informática.

Las ontologías son usadas para describir compromisos entre agentes, de modo que si un agente está comprometido con una ontología sus acciones son consistentes con la definición de esta [Gruber, 1995].

También, se puede usar las ontologías para crear una red de relaciones, de forma que se puede explorar esta red y rastrear los elementos conectados [Uschold et al., 1996]. Una ontología provee de definiciones libres de ambigüedad para definiciones de términos empleados en un sistema [Uschold et al., 1996].

Las ontologías son de utilidad para comunicar diferentes herramientas de software. Los sistemas de información cuentan con modelos empresariales compartidos por todas las herramientas de software. [Uschold et al., 1996].

Entre los diversos usos de las ontologías también se encuentran la recuperación de información, librerías digitales, integración de fuentes de información heterogéneas y motores de búsqueda en la web [Chandrasekaran et al., 1999].

Otras dos áreas importantes de aplicación de las ontologías son la comprensión de lenguaje natural y en la resolución de problemas basada en conocimiento [Chandrasekaran et al., 1999]. En el primer caso, la ontología actúa como un diccionario de conceptos que permite identificar las categorías semánticas involucradas en la comprensión de un discurso [Chandrasekaran et al., 1999]. En el segundo caso, estas se utilizan para representar conocimiento, con el que se pretende simular el razonamiento basado en el sentido común para llegar a la solución de un problema [Chandrasekaran et al., 1999].

La reutilización y adaptación de elementos es otra de las aplicaciones en las que las ontologías resultan particularmente útiles. La arquitectura UPML describe un estándar en el que se distinguen componentes intercambiables entre sistemas basados en conocimiento [Studer et al., 2000].

Además existen nuevas áreas de aplicación de las ontologías:

Servicios de información inteligentes: Servicios que combinan técnicas de obtención y extracción de información para integrar diferentes fuentes de información facilitando al usuario el acceso a información relevante desde su perspectiva [Studer et al., 2000].

Servicios de razonamiento inteligentes: Servicios que permiten a usuarios comunes o empresas acceder a servicios de razonamiento que ofrezcan soluciones para determinados dominios [Studer et al., 2000]. El usuario accede a una librería de razonamiento que puede configurar para adaptarla a sus necesidades [Studer et al., 2000]. Este servicio se puede brindar a usuarios comunes mediante el pago por uso o suscripción a empresas ofreciendo soluciones básicas que pueden refinarse [Studer et al., 2000].

Web semántica: El crecimiento desmedido de la información de la web motivó el desarrollo de herramientas que ayuden al humano a procesar y filtrar la masiva cantidad de información disponible [Studer et al., 2000]. Las ontologías se perfilan como una posible solución enfocada en la creación de datos formalmente definidos e interconectados en la web [Studer et al., 2000]. La W3C creó el marco de descripción de recursos (RDF por sus siglas en inglés) con este fin [Studer et al., 2000].

2.1.3. Raciocinio

2.1.3.1. Definición

El raciocinio sobre cualquier dominio de la realidad siempre requiere simplificación. El solo preparar el conocimiento para soportar el raciocinio requiere dejar muchos hechos sin explicar o resumidos muy vagamente [Pearl, 1988]. Al declarar cualquier hecho se realizan ciertas asunciones, por ejemplo, al afirmar como regla “los pájaros vuelan”, se descartan ciertas excepciones y se aceptan las condiciones bajo las que sea aplica la regla a pesar de que podrían ser ambiguas o difíciles de definir [Pearl, 1988]. El raciocinio, con excepciones, es como atravesar un campo minado. La mayoría de pasos son seguros pero un paso en falso puede ser devastador [Pearl, 1988]. Se podría marcar la localización de cada mina pero intentarlo es como marcar en un mapa del tamaño de una tarjeta postal, lo que resulta bastante impreciso [Pearl, 1988]. Una alternativa a este problema es enumerar las excepciones y resumirlas, como si se marcara las áreas de mayor riesgo en el mapa de minas [Pearl, 1988]. Este proceder es indispensable para balancear seguridad y velocidad en el movimiento [Pearl, 1988].

La habilidad de usar tanto información predictiva como de diagnóstico es un componente importante del raciocinio plausible, un mal uso de esta información conduce a resultados extraños [Pearl, 1988].

Frente al problema de la falta de certeza existen diferentes posturas que pueden ser clasificadas en dos categorías: Aproximación extensional y aproximación intencional [Pearl, 1988].

La aproximación extensional maneja la falta de certeza con un valor de verdad asociado a las fórmulas y calcula la incertidumbre de una fórmula como una función de las incertidumbres de sus subfórmulas [Pearl, 1988]. Esta aproximación descuida la parte semántica pero son convenientes computacionalmente [Pearl, 1988].

La aproximación intencional relaciona la incertidumbre al estado de los hechos, un subconjunto de posibles “mundos”. Esta aproximación es más clara en el aspecto semántico pero entorpece la labor computacional [Pearl, 1988].

2.1.3.2. Ejemplos de Raciocinio Usando Ontologías

Un claro ejemplo del uso del raciocinio basado en ontologías se da en la búsqueda de imágenes basada en inferencias. Si se quiere encontrar decoraciones de fiestas se puede recurrir a imágenes con anotación semántica correctamente etiquetadas como decoraciones [Guha et al., 1998]. Sin embargo una fiesta infantil tiene decoraciones y puede estar anotada como fiesta infantil [Guha et al., 1998]. La búsqueda basada en la inferencia puede basarse en un conjunto de reglas sobre el mundo que le permite inferir que una fotografía sobre fiestas probablemente tenga decoraciones [Guha et al., 1998].

Un motor de inferencia trabaja, visto de manera muy superficial, mediante un conjunto de reglas del tipo “Si a y b entonces c” aplicadas a un grupo de hechos atómicos [Guha et al., 1998]. Cuando se hace una consulta o se obtienen nuevos hechos, en base a los anteriores, se aplican las reglas para obtener conclusiones [Guha et al., 1998].

Los motores de inferencia por lo general usan modelos lógicos, los cuales son abstracciones generalmente representadas usando teoría de conjuntos [Guha et al., 1998]. Estos definen una entidad abstracta y se formalizan en una o más sintaxis concretas, modelos físicos. [Guha et al., 1998] Es decir, un modelo físico es una representación lógica codificada o traducida a pseudocódigo bajo cierta sintaxis. Una manipulación concreta en particular se suele dar en un modelo físico. [Guha et al., 1998]. Sin embargo, se suele usar modelos lógicos para el trabajo de inferencia por las siguientes razones:

- Un modelo lógico puede tener uno o muchos modelos físicos asociados, distintas maneras de representarlo. Cada modelo físico se compromete con una sintaxis, con una semántica limitando el modelo y dificultando el trabajo de los motores de inferencia y de consultas de alto nivel. [Guha et al., 1998].
- Los modelos lógicos permiten realizar algunas operaciones que en un modelo físico serían difíciles o imposibles de aplicar. [Guha et al., 1998].
- La mayoría de sistemas de inferencia tienen conclusiones de deducción infinitas, el conjunto de sentencias utilizado puede tener infinitas representaciones concretas. Consultar un resultado de este tipo requiere consultas en los términos de un modelo lógico. [Guha et al., 1998].

2.1.4. Inferencia

2.1.4.1. Definición

Respecto a la información representada sobre un dominio específico se puede definir:

“Inferencia: La información que puede obtenerse de la información representada” [Corcho and Gómez-Pérez, 2000].

La habilidad de usar información tanto predictiva como de diagnóstico son componentes de un razonamiento plausible [Pearl, 1988]. El uso inadecuado de esta información conduce a resultados inesperados [Pearl, 1988]. Un patrón común del discurso es el razonamiento abductivo, si una proposición A implica B, darle un valor de verdadero a B aumenta la credibilidad de A [Pearl, 1988]. Se define como inferencia bidireccional a esta capacidad de ir de la evidencia a la hipótesis y viceversa. [Pearl, 1988].

2.1.4.2. Uso de las Ontologías en las inferencias

Una ontología puede ser usada para formular consultas semánticas y ofrecer exactamente la información en la que estamos interesados [Decker et al., 1999]. Además, los axiomas provistos por las ontologías sirven como medios para derivar información que ha sido especificada de manera implícita [Decker et al., 1999]. El sistema Ontobroker, utiliza las ontologías formales para extraer, razonar y generar metadatos en el WWW [Decker et al., 1999]. Este sistema tiene los siguientes importantes componentes:

La parte central son las ontologías, usadas en la mayoría de componentes del sistema [Decker et al., 1999]. Estas proveen de una estructura formal del contenido de las páginas, permitiendo así información integrada y fácil de manejar [Decker et al., 1999].

El motor de inferencia aprovecha esta semántica formal de la representación del lenguaje y permite un raciocinio automático bien definido [Decker et al., 1999]. Un ejemplo de este es el motor de inferencias de Ontobroker, el cual está conformado por dos elementos. El primero traduce el lenguaje especializado en modelado a uno más restringido que es utilizado por el segundo elemento, que realiza la evaluación de expresiones en

este lenguaje [Decker et al., 1999].

La interfaz de consulta permite la formulación interactiva de consultas durante la navegación por la ontología y la selección de los términos que integran la consulta [Decker et al., 1999].

A continuación se muestra un diagrama de la arquitectura de Ontobroker:

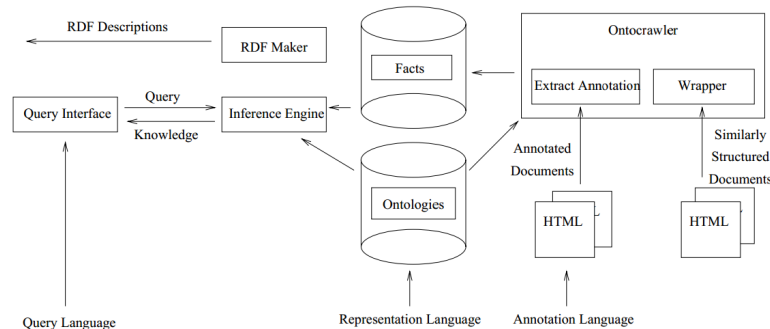


Imagen 4: Arquitectura de Ontobroker [Decker et al., 1999].

Hay una fuerte relación entre las estructuras usadas para representar el conocimiento y la las bases del proceso de raciocinio [Corcho and Gómez-Pérez, 2000]. Mientras más expresivo es un lenguaje menores son sus capacidades de inferencia [Corcho and Gómez-Pérez, 2000]. El motor de inferencia razona con el conocimiento representado [Corcho and Gómez-Pérez, 2000]. Este realiza clasificaciones automáticas, maneja excepciones, revisa restricciones de los axiomas y encadena conceptos hacia adelante o hacia atrás al razonar en base a reglas [Corcho and Gómez-Pérez, 2000].

2.1.5. Conclusión

Se ha definido el concepto de ontología primero porque sobre este concepto se estructura el resto del proyecto. Se parte de la abstracción del conocimiento de un dominio específico, encapsulada en un vocabulario formal, para construir una base de conocimiento. Esta base de conocimiento estructurada dará soporte para realizar inferencias sobre el dominio seleccionado. Esto se realizará simulando el raciocinio necesario para la obtención de información de diagnóstico mediante un mecanismo que realice inferencias. Las inferencias se obtendrán partiendo del conocimiento formalizado en la onto-

logía, utilizando un conjunto de reglas para hacer deducciones válidas para el dominio seleccionado. De esta manera se explotará el potencial de las ontologías en cuanto a expresividad al representar conocimiento y los mecanismos de inferencia que aprovechan la formalidad lógica de esta representación para realizar diagnósticos. Trabajando sobre estos conceptos se intentará buscar una solución a la problemática definida.

2.2. Estado del arte

2.2.1. Objetivos del estado del arte

Con el fin de conocer las herramientas más adecuadas para hacer frente al problema planteado se revisará el reciente desenvolvimiento de las tecnologías de interés. En los últimos años se ha hecho uso de las ontologías para realizar inferencias con distintos fines. Se ha listado distintos ejemplos de sistemas que ofrecen propuestas de soluciones a problemas específicos utilizando estas tecnologías.

Los objetivos de la revisión del estado del arte son:

- Identificar ejemplos exitosos de sistemas que realicen inferencias basadas en ontologías.
- Identificar ontologías aplicables al problema planteado.
- Identificar técnicas de inferencia aplicables al problema planteado.
- Identificar herramientas que propongan soluciones a problemas similares.

2.2.2. Método usado en la revisión del estado del arte

El método a utilizar en la revisión del estado del arte es la revisión sistemática. Se realizará una búsqueda de fuentes relevantes siguiendo una pregunta que oriente la búsqueda. Se planteó la siguiente pregunta para direccionar la búsqueda: ¿Cómo han sido utilizadas las ontologías en el proceso de inferencia de conocimiento?

Se utilizó los buscadores Scopus¹, ACM², IEEE³ y ScienceDirect⁴.

¹<http://www.scopus.com/>

²<http://dl.acm.org/>

³<http://ieeexplore.ieee.org/Xplore/home.jsp>

⁴<http://www.sciencedirect.com/>

Los criterios de búsqueda fueron todos los estudios primarios con las palabras claves «ontología» e «inferencia» en el título de un artículo y se excluyó todo estudio primario en donde de forma explícita no se usen las ontologías en el proceso de inferencia.

2.2.3. Resultados de la revisión sistemática

A continuación se presentan los resultados de la revisión sistemática.

En la siguiente tabla se presentan las fuentes relacionadas con la medicina humana:

Cuadro 3: Resultados de la revisión sistemática relacionados a la medicina humana.

Artículo	Propósito de la ontología	Métodos, procedimientos, herramientas usadas
MEDBOLI: Medical Diagnosis Based on Ontologies and Logical Inference [Rodríguez et al., 2009].	Diagnóstico Diferencial. Se utiliza la ontología para estructurar el amplio número de enfermedades e identificarlas usando diagnóstico diferencial (basado en síntomas representativos de la enfermedad).	Aplica el uso de ontologías combinadas con inferencia lógica y computo de probabilidades.

Artículo	Propósito de la ontología	Métodos, procedimientos, herramientas usadas
<p>Gene Ontology-driven inference of protein-protein interactions using inducers [Maetschke et al., 2012].</p>	<p>Representación de Proteínas y sus interacciones. El proyecto Gene Ontology ofrece una ontología de términos que representan productos genéticos y sus propiedades. La ontología cubre tres dominios: componentes celulares, funciones moleculares y procesos biológicos.</p>	<p>Inferencia basada en interacción de anotaciones GO.</p>
<p>Experience of using OWL ontologies for automated inference of routine pre-operative screening tests [Bouamrane et al., 2010].</p>	<p>Inferencia lógica en el dominio de conocimiento, de acuerdo al contexto médico del paciente. Manejo eficiente de un vasto repositorio de conocimiento de seguimiento pre-operatorio, incluyendo clasificación de cirujías, enfermedades y guías para exámenes pre-operatorios.</p>	<p>Desarrollo del sistema utilizando tecnología de la web semántica, incluyendo módulos de ontologías desarrollados en OWL y una interfaz de programación para OWL diseñada en JAVA que utiliza un razonador lógico.</p>

Artículo	Propósito de la ontología	Métodos, procedimientos, herramientas usadas
<p>ODDIN: Ontology-driven differential diagnosis based on logical inference and probabilistic refinements [García-Crespo et al., 2010].</p>	<p>Diagnóstico Médico Diferencial. Estimación de múltiples parámetros para determinar el diagnóstico más probable.</p>	<p>Construir un sistema de diagnóstico médico diferencial inteligente implica usar tecnologías basadas en conocimiento que permitan evitar ambigüedades. Para esto se utilizan ontologías que representen información estructurada y estrategias como cómputo de probabilidades de varios factores e inferencia lógica.</p>
<p>Discovering relationship associations in life sciences using ontology and inference [Guo and Kraines, 2009].</p>	<p>Dar soporte a los investigadores para realizar descubrimientos en base al conocimiento publicado en la actualidad.</p>	<p>Usar tecnologías de la web semántica para medir la relevancia entre relaciones con tipos específicos que relacionen a alguna entidad. Estas relaciones pueden ayudar a los investigadores a generar hipótesis científicas o a crear descriptores semánticos para sus artículos.</p>

2.2.4. Conclusiones sobre el estado del arte

Se listó la información relevante sobre los artículos encontrados y se logró obtener, respecto a los objetivos de la revisión, información de utilidad.

Como se puede apreciar en la tabla anterior, existen gran cantidad de papers que tratan acerca de la aplicación de las ontologías e inferencias en diversos dominios. En estos, las ontologías son usadas con el propósito de esquematizar el conocimiento de un área determinada para que, con esta estructura, se puede trabajar de manera más eficientes las inferencias lógicas; de forma que, las máquinas puedan hacer deducciones sin la intervención humana.

Se han encontrado papers acerca de la aplicación de ontologías e inferencias para el dominio de la medicina humana. Siendo esta una ciencia afín a la medicina veterinaria, se puede desprender que la utilización aplicada en dicha área también puede ser aplicada en la medicina veterinaria.

A manera de ejemplo, podemos enlistar una serie de aplicaciones y usos, como lo son: Ontologías basada en el diagnóstico médico, uso para la gestión de repositorios de las historias clínicas combinada con los procedimientos quirúrgicos previstos, uso para la anotación de genes, uso en la estimación de múltiples parámetros distintos con el fin de determinar el diagnóstico más probable, así como en la reconstrucción de árboles filogenéticos y analizar las relaciones evolutivas entre las especies y otros.

De manera similar se encontraron herramientas recurrentes para el caso del sistema de inferencias. Entre estas se encuentra el uso de lógica difusa, relaciones entre clases y sub-clases, aprendizaje en una web semántica, cómputo de probabilidades, inferencia lógica y asociación entre relaciones de elementos.

El caso específico de “ODDIN: Ontology-driven differential diagnosis based on logical inference and probabilistic refinements” se asemeja mucho al problema propuesto en la investigación actual. Las herramientas que se encontraron factibles son las ontologías como representación de información estructurada y el cómputo de probabilidades junto a la inferencia lógica.

3. Capítulo 3

3.1. Introducción

En este capítulo se abordarán los resultados obtenidos para el primer objetivo específico. Se fijó como primer objetivo específico documentar una revisión de los métodos y técnicas de inferencia basada en ontologías existentes para soportar el diagnóstico en el dominio de la medicina veterinaria. Se cuenta con dos resultados, el primero es un informe de ontologías aplicables y el segundo un informe de herramientas de inferencia que puedan trabajar con las ontologías seleccionadas.

3.2. Resultado Esperado 1: Informe de Herramientas Aplicables - Ontologías

El proyecto apunta a modelar el conocimiento necesario para realizar un diagnóstico en el dominio de la veterinaria haciendo uso de ontologías para dar soporte al médico veterinario. Realizando inferencias sobre este conocimiento modelado se obtendrá un producto, información relevante y útil para el diagnóstico.

Analizando la problemática se determinó como una causa principal la falta de herramientas orientadas al diagnóstico veterinario. El presente documento tiene como objetivo presentar la evaluación de dos herramientas candidatas para el modelamiento de la ontología a utilizar, como parte del primer resultado esperado para el objetivo específico 1. Estas consisten en dos formatos, OWL, *Web Ontology Language*, y OBO, *Open Biological and Biomedical Ontologies*.

Además se listará ontologías que den luces acerca de la mejor manera de modelar el problema en cuestión. Se revisarán ontologías de la base de datos de OBO y ontologías utilizadas en otros proyectos con un enfoque aproximado en búsqueda de opciones. Esto con el fin de encontrar una ontología adecuada para realizar la adaptación al caso planteado.

3.2.1. OBO

OBO es una comunidad abierta que los autores de ontologías se comprometen a mantener y mejorar trabajando juntos en pro del avance de la ciencia [Smith et al., 2007]. Basados en ciertos principios, tratan de aplicar el método científico a la tarea de desarrollar ontologías, este se basa en la crítica constante y la asunción de que todo recurso puede ser perfeccionado [Smith et al., 2007].

Su meta es construir una colección de ontologías de dominio público que compartan diseño y revisión y que faciliten el compartir información respecto a las ciencias biológicas y de la salud [Smith et al., 2007]. Con un conjunto de principios, tratan de dar ejemplo de buenas prácticas en el desarrollo y la aplicación de las ontologías [Smith et al., 2007].

OBO cuenta con su propio editor de ontologías programado en Java, este sigue el formato OBO-Format especificado en el enlace siguiente: http://www.geneontology.org/G0.format.obo-1_2.shtml.

Existen diferentes sistemas de interconversión entre OBO y OWL pero todos son incompletos y tienen serias desventajas, estos formatos aun no son compatibles ni cuentan con herramientas comunes [Smith et al., 2007]. La fundación OBO necesita una conversión estándar para garantizar la reutilización de ontologías sin importar el formato, esto implica trabajar con las dos herramientas existentes OBO-Edit y Protégé [Smith et al., 2007]. Actualmente OBO acepta ontologías en formatos OBO, OWL y OWL2.

3.2.2. OWL

El Lenguaje de Ontologías de la Web Semántica (OWL) está diseñado para ser usado en aplicaciones que necesitan procesar el contenido de la información en lugar de únicamente representar información para los humanos [McGuinness et al., 2004]. OWL puede ser usado para representar explícitamente el significado de términos en vocabularios y las relaciones entre esos términos, denominada ontología [McGuinness et al., 2004]. OWL facilita un mejor mecanismo de interpretabilidad de contenido Web que los mecanismos admitidos por XML, RDF, y esquema RDF (RDF-S) proporcionando vocabulario adicional junto con una semántica formal [McGuinness et al., 2004]. OWL va más

allá de estos lenguajes en su capacidad para representar contenido interpretable por un ordenador en la Web [McGuinness et al., 2004]. OWL es una revisión del Lenguaje de Ontologías Web DAML+OIL incorporando lecciones aprendidas a partir del diseño y aplicación de DAML+OIL [McGuinness et al., 2004]. Posee tres sublenguajes, con un nivel de expresividad creciente: OWL Lite, OWL DL, y OWL Full [McGuinness et al., 2004].

OWL proporciona tres lenguajes, cada uno con nivel de expresividad mayor que el anterior, diseñados para ser usados por comunidades específicas de desarrolladores y usuarios [McGuinness et al., 2004]

OWL Lite está diseñado para aquellos usuarios que necesitan principalmente una clasificación jerárquica y restricciones simples. Por ejemplo, a la vez que admite restricciones de cardinalidad, sólo permite establecer valores cardinales de 0 o 1 [McGuinness et al., 2004].

OWL DL está diseñado para aquellos usuarios que quieren la máxima expresividad conservando completitud computacional (se garantiza que todas las conclusiones sean computables), y resolubilidad (todos los cálculos se resolverán en un tiempo finito) [McGuinness et al., 2004]. OWL DL incluye todas las construcciones del lenguaje de OWL, pero sólo pueden ser usados bajo ciertas restricciones (por ejemplo, mientras una clase puede ser una subclase de otras muchas clases, una clase no puede ser una instancia de otra) [McGuinness et al., 2004].

OWL Full está dirigido a usuarios que quieren máxima expresividad y libertad sintáctica de RDF sin garantías computacionales [McGuinness et al., 2004]. Por ejemplo, en OWL Full una clase puede ser considerada simultáneamente como una colección de clases individuales y como una clase individual propiamente dicha [McGuinness et al., 2004]. OWL Full permite una ontología para aumentar el significado del vocabulario preestablecido (RDF o OWL) [McGuinness et al., 2004]. Es poco probable que cualquier software de razonamiento sea capaz de obtener un razonamiento completo para cada característica de OWL Full [McGuinness et al., 2004].

La ventaja que representa usar OWL respecto a otros lenguajes en ontologías web es

que este lenguaje ha sido definido para ser compatible con la arquitectura World Wide Web y la web semántica en particular a diferencia de los lenguajes anteriores que se desarrollaron para comunidades de usuarios específicas [Hendler, 2004].

OWL ratifica esto proporcionando un lenguaje que utiliza la conexión proporcionada por RDF para agregar las siguientes capacidades a las ontologías:

- Capacidad para ser distribuidos a través de muchos sistemas.
- Escalable a las necesidades de la Web.
- Compatible con los estándares web para la accesibilidad y la internacionalización.
- Abierta y extensible [Hendler, 2004].

Además la nueva versión de OWL, OWL2, ha permitido la ampliación del sistema Protégé con las construcciones adicionales proporcionadas por OWL2 [?].

3.2.3. Ontologías Aplicables

Para encontrar la ontología que más se ajusta a las necesidades del proyecto se revisarán diversas opciones obtenidas tanto de la librería de ontologías de OBO como de otras fuentes.

Una de las candidatas es una ontología obtenida de la librería de OBO sobre enfermedades humanas (<http://www.berkeleybop.org/ontologies/doid.owl>). Esta muestra una clasificación de enfermedades y relaciones de tipo “tiene síntoma”, “deriva de” o “transmitido por” y nos dará una idea de como modelar las enfermedades del animal.

CommonKADS propone varias plantillas para el modelamiento de conocimiento aplicables a diversos casos. Busca reutilizar elementos de estos modelos en la medida de lo posible para reducir esfuerzo [Schreiber, 2000]. Dentro de estas plantillas se presenta una orientada al diagnóstico de errores en un sistema. Esta ofrece terminología, métodos, entradas y salidas útiles para el modelamiento del conocimiento asociado a la tarea de diagnóstico [Schreiber, 2000]. Se utilizará esta plantilla como guía para la construcción de la ontología.

A continuación se presenta la estructura planteada por la metodología CommonKADS para el modelo correspondiente al diagnóstico⁵. Las entidades están representadas como cuadros rectangulares y las relaciones dentro de elipses. Las flechas indican la dirección de efecto de la relación.

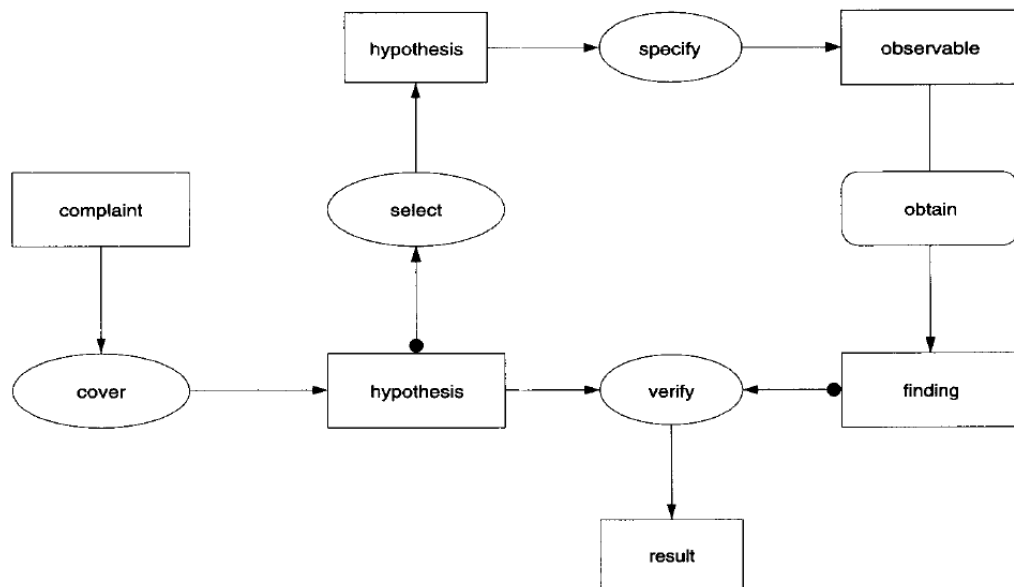


Imagen 5: Estructura de inferencia para el método de diagnóstico [Schreiber, 2000].

3.2.4. Conclusión

De acuerdo a lo presentado previamente se ha decidido utilizar el formato OWL para la construcción de la ontología por ser el estándar al que tiende el desarrollo de ontologías. Particularmente se utilizará el lenguaje OWL DL, descartando el Lite por su simpleza y el Full porque no es computable. Además se utilizará la herramienta Protégé para el modelado de la ontología.

Se dió preferencia a este formato sobre OBO porque es un lenguaje más estable, mejor documentado y con mayor compatibilidad de herramientas. Lo más atractivo de las ontologías OBO es el contenido y revisando la página web de OBO se puede confirmar que

⁵Algunos diagramas se han mantenido en inglés con el fin de reutilizarse en la redacción de un artículo sobre el presente proyecto.

están en un proceso bastante avanzado de integración de sus ontologías al formato OWL.

Se tomará en cuenta los diseños de las ontologías revisadas así como el análisis y diseño de un sistema de diagnóstico básico presentado en la metodología CommonKads. Con estas bases se construirá una ontología simple abarcando los principales aspectos que permitan realizar inferencias orientadas al diagnóstico.

Se plantea construir una ontología estándar que pueda ser reutilizada sin depender del dominio. Sin embargo, estará orientada al diagnóstico veterinario por lo que otros trabajos en este dominio permitirán una mejor adaptación de la ontología.

3.3. Resultado Esperado 2: Informe de Herramientas Aplicables - Inferencias

Seleccionadas las herramientas a utilizar para modelar el conocimiento, continuaremos definiendo cómo este se procesará para llegar a un resultado satisfactorio. El objetivo es obtener un diagnóstico que refleje el de un médico veterinario. Esto se realizará recorriendo el contenido de la base de conocimiento y relacionando conceptos.

En esta sección se presentarán ejemplos de cómo OWL permite representar el conocimiento de forma que se pueda procesar para realizar inferencias. Además, se compararán dos herramientas de inferencia que trabajan sobre OWL, Jena y Base Visor.

3.3.1. Inferencias sobre OWL

La documentación de OWL contiene ejemplos sobre distintas maneras de realizar inferencias sobre ontologías construidas en este lenguaje. Para esto utilizan como base una ontología de personas.

Un primer ejemplo que presentan son las inferencias de clases. Para el caso de conductores de autobús se tiene:

```
Class(a:bus_driver_complete_intersectionOf(a:person  
restriction(a:drives_someValuesFrom_(a:bus))))
```

```
Class(a:driver_ complete_intersectionOf(a:person
restriction(a:drives_someValuesFrom_(a:vehicle))))
```

```
Class(a:bus_partial_a:vehicle)
```

Se puede revisar que la relación tiene `_mascota` tiene su dominio en la clase persona y su rango en la clase animal. La relación es `_mascota_de` se invierte el dominio y el rango. Por lo tanto si Spike está asociado a Pete por la relación es `_mascota_de` Pete debe pertenecer a la clase persona y Spike a la clase animal.

Realizar este tipo de inferencias se hace posible por la capacidad de representar por medio del lenguaje OWL reglas de cuantificación existencial propias de la lógica proposicional [Bechhofer, 2003]. OWL permite expresar expresiones como $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ [Bechhofer, 2003].

Además incluye los conceptos de existencia (existe al menos un...) y universalidad (para todo...) de la teoría de conjuntos [Bechhofer, 2003]. Estas son herramientas poderosas para realizar inferencias entre clases e individuos [Bechhofer, 2003].

3.3.2. Jena

Jena es un framework open source, en Java, para construir aplicaciones en el contexto de la web semántica. Esta cuenta con una API que permite trabajar con Ontologías.

La API de Jena reconoce los lenguajes RDFS, OWL Lite y OWL DL, mencionados en orden de expresividad de menor a mayor [Foundation, 2014]. Para trabajar con OWL DL utiliza razonadores intermedios que permiten realizar inferencias sobre estructuras más complejas [Foundation, 2014].

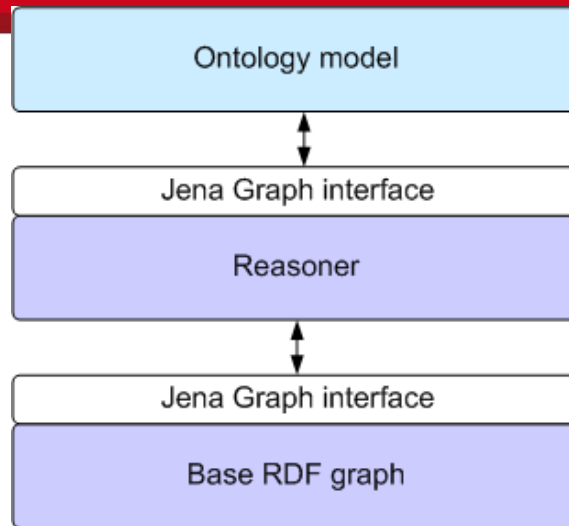


Imagen 6: Imagen obtenida de la documentación de Jena [Foundation, 2014]

En la imagen se puede ver la comunicación entre la ontología y la API, utilizando un razonador de intermediario [Foundation, 2014]. Jena provee una adaptación de su clase base para trabajar con RDF, Model, para trabajar con ontologías, esta es OntModel [Foundation, 2014]. Esta se construye asociándola a un modelo de inferencia infModel que contiene un razonador [Foundation, 2014]. Jena contiene tres razonadores que trabajan sobre OWL DL, Pellet, Racer y FaCT.

3.3.3. Pellet, Racer y FaCT

Según el artículo Pellet: An OWL DL Reasoner; de los tres razonadores más conocidos para trabajar con OWL DL (Pellet, Racer y FaCT), los dos últimos son eficientes pero no cumplen con ciertos requerimientos importantes [Parsia and Sirin, 2004]. Un razonador en el contexto de la web semántica debería manejar inferencia sobre individuos, no debería asumir nombres únicos, debería soportar verificaciones de vinculamiento y debería trabajar con esquemas XML. Afirma que Pellet ha sido desarrollado para cumplir estas expectativas y que, a pesar de que no alcanza el rango de desempeño de Racer o FaCT es una buena opción para soluciones poco pesadas como el prototipo que se desea construir.

3.3.4. BaseVISor

BaseVISor es un herramienta de software libre de inferencia que utiliza encadenamiento progresivo basado en una red Rete optimizada para el procesamiento de RDF triples con soporte para OWL 2 RL y XML Schema Datatypes [Matheus et al., 2006]. BaseVISor 2.0 cuenta con un motor de inferencia de encadenamiento hacia adelante muy eficiente, optimizada para el razonamiento ontológico y basado en normas [Matheus et al., 2006]. Las reglas y los hechos se compilan en una red Rete y procesados con respecto a las ontologías especificadas a fin de generar todos los nuevos hechos inferibles [Matheus et al., 2006]. BaseVISor 2.0 puede ser embebido en aplicaciones existentes y ampliado para satisfacer los requisitos específicos de la aplicación [Matheus et al., 2006].

Utiliza una estructura de datos simple para sus hechos (es decir, triples), más que arbitrarias estructuras de lista. BaseVISor está escrito en Java, e incluye una API para añadir fácilmente los archivos adjuntos de procedimiento definidos por el usuario [Matheus et al., 2006]. Su API facilita la integración dentro de otra aplicación Java [Matheus et al., 2006].

La forma general de funcionamiento de BaseVISor consiste en escribir un archivo XML que contiene los hechos (es decir, triples primas), una base de normas y, posiblemente, una o varias consultas que se someten a continuación al procesador estándar [Matheus et al., 2006]. También es posible escribir las declaraciones para incluir otros archivos en el procesamiento por lotes [Matheus et al., 2006]. En particular, se puede incluir múltiples bases de normas mediante el elemento de inclusión y proporcionar la dirección URL del documento base de reglas [Matheus et al., 2006]. También se puede utilizar para importar un documento RDF / OWL especificando el atributo lang como RDF [Matheus et al., 2006].

3.3.5. Conclusión

Se cuentan con herramientas que permiten trabajar inferencias sobre una base de conocimiento construida en OWL. Estas ofrecen distintos tipos de inferencias basadas en las expresiones que provee el lenguaje OWL. Se revisó la herramienta Jena y la herramienta BaseVisor, ambas son software libre y trabajan sobre una base de RDF. Jena

ofrece una interfaz para la extracción de información y razonadores que permiten construir nueva información en base a lo especificado en la ontología. El API BaseVISor proporciona todos los métodos necesarios para la lectura de las reglas y archivos de ontologías, la consulta de la base de hechos y la aplicación de servicios web. Se decidió utilizar Jena, con el razonador Pellet, por ofrecer un entorno más amigable, mejor documentación y mayor integración con OWL. Si bien ambas cuentan con métodos para acceder a ontologías, a diferencia de BaseVISor, Jena cuenta con un conjunto de clases especializadas para el tratamiento de la ontología para tareas de lectura, construcción, modificación y razonamiento. Además permite hacer uso de razonadores externos construidos específicamente para trabajar sobre OWL, los mismos que suelen usarse en paralelo con la herramienta Protégé para construir y probar la congruencia de la ontología.

4. Capítulo 4

4.1. Introducción

En este capítulo se presentará el análisis y diseño de la base de conocimiento junto a los resultados del proceso de construcción de la misma. Se evaluará el conocimiento necesario para realizar el diagnóstico y se modelará teniendo en cuenta la estructura propuesta por la metodología CommonKADS para la tarea de diagnóstico. También se presentará la clasificación de enfermedades y síntomas a utilizar. Finalmente, se explicará el proceso de construcción de la ontología.

4.2. Resultado Esperado 3: Análisis de la Base de Conocimiento

En esta sección se presentará el análisis realizado del problema planteado respecto al conocimiento a representar. Se presentarán esquemas que muestren las tareas que realiza el médico para llegar a un diagnóstico. Además se estructurarán los resultados de una clasificación de enfermedades de la piel en perros y los síntomas respectivos. De esta manera se busca dar forma al contenido de la base de conocimiento y obtener entradas útiles para proceder con el diseño de la misma.

4.2.1. Proceso de Diagnóstico

Según lo revisado en el libro “Manual of skin Diseases of the Dog and Cat” de Sue Paterson, para poder diagnosticar a un paciente se deben realizar básicamente 2 pasos: el primero es identificar el área(s) del cuerpo en donde se está manifestando el síntoma y el segundo paso es identificar los síntomas relacionados al área(s) del cuerpo que se haya indicado. Al poder identificar los síntomas de las diversas áreas y al relacionarlos se puede llegar a un diagnóstico posible [Paterson, 2009].

Sin embargo, existen síntomas que no están relacionados directamente con alguna parte del cuerpo, como la fiebre o la presión, pero que si se consideran a la hora de realizar el diagnóstico [Paterson, 2009].

También, existen factores que no se reflejan como síntoma pero que son condiciones que predisponen la existencia de la enfermedad. Entre estos factores se encuentran factores genéticos y de raza, así como también el historial clínico de la mascota [Paterson, 2009].

4.2.2. Catálogo de Enfermedades

En esta sección se presentará el análisis de la información de las enfermedades de la piel en los perros necesario para la estructura de la ontología planteada.

Para este análisis, se hecho uso del libro “Manual of Skin Diseases of the Dog and Cat” de Sue Paterson, el cual da una clasificación de las enfermedades de la piel de perros y gatos, especificando los subtipos y enfermedades dentro de estos [Paterson, 2009].

Se ha elaborado un cuadro con lo expuesto en el libro para poder así identificar los tipos de enfermedades a abordar en el modelo de ontología que plantearemos puesto que abarcar todos los tipos no sería factible en el tiempo del ciclo regular.

El criterio para seleccionar estos tipos depende tanto de que estos sean los más comunes en los perros y del criterio del experto veterinario que nos indique las más relevantes. En un inicio se había planteado abarcar solo las enfermedades de la piel de la raza Shar Pei;

sin embargo, se ha visto la posibilidad de quitar esta limitación puesto que las enfermedades específicas para esta raza son pocas y la estructura de enfermedades planteada provee un mayor campo de enfermedades comunes, no solo en dicha raza, sino a varias en general. En los anexos se muestra la tabla estructurada de las enfermedades de la piel. Adicionalmente se ha construido una ontología basada en enfermedades gastro-intestinales, la tabla correspondiente se encuentra también en los anexos.

Como siguiente paso en el análisis, se identificará los síntomas característicos de cada enfermedad para estructurarlos de manera que puedan expresarse como parte de una ontología.

4.2.3. Conclusión

Se ha analizado la información existente de las enfermedades de la piel en base al conocimiento del experto veterinario y a un libro guía, el cual nos ha permitido seleccionar las enfermedades que se cubrirán en la ontología y que orientarán el diagnóstico de las enfermedades. También, con este análisis se tiene una idea más clara de lo que se va a diseñar de acuerdo a las especificaciones que tiene cada enfermedad y los criterios que se necesitan para poder identificarla y evaluarla.

4.3. Resultado Esperado 4: Diseño de la Base de Conocimiento

En este capítulo se presentará el diseño de la ontología basado en el análisis realizado, en las ontologías revisadas y en la metodología CommonKADS. Como se especificó en el informe de ontologías aplicables se utilizará la plantilla de diagnóstico propuesta por CommonKADS. Además, se seguirán ejemplos de ontologías existentes para definir un modelo de clasificación de enfermedades y otras entidades importantes para realizar las inferencias. Se presentarán dos modelos, el esquema del dominio y la estructura de inferencias.

El primero muestra las entidades y las relaciones entre estas como parte de la representación del dominio. El segundo muestra roles, inferencias y funciones de transferencia a utilizarse para realizar inferencias en base a la ontología a construir. Se hizo una adaptación de la plantilla propuesta agregando elementos particulares del caso a tratar. En el

documento de diseño de la inferencia presentado en el capítulo 5 se tratará a profundidad la plantilla propuesta.

4.3.1. Esquema del Dominio

El modelo de conocimiento planteado provee un conjunto de construcciones modeladas para especificar el esquema del dominio de una aplicación. En particular, siguiendo la notación provista por UML para la diagramación de clases [Schreiber, 2000].

A continuación se muestra la esquema a desarrollar en el dominio de la veterinaria, la cual posee las clases necesarias que nos permitirán realizar el diagnóstico, las relaciones y la cardinalidad entre estas clases. Se ha considerado estas clases debido a que son las más relevantes para el experto veterinario a la hora de diagnosticar.

Las clases determinadas son:

- **Paciente:** Representa al paciente a diagnosticar la enfermedad. Como atributos se ha identificado inicialmente: nombre, raza, edad, género y tamaño del perro.
- **Enfermedad:** Representa al conjunto de enfermedades de la piel. En el esquema planteado, sólo se ha considerado tres tipos de enfermedades de la piel, puesto que el espectro de enfermedades es muy amplia. La selección de estos tres tipos de enfermedades se hará a criterio de las más comunes según lo indique el experto veterinario. Inicialmente se ha propuesto tomar las enfermedades del tipo: causadas por hongos (*fungus skin disease*), por bacterias (*bacterial skin disease*) y por virus, rickettsias y protozoarios (*viral, rickettsial and protozoal diseases*).
- Dentro de los atributos identificados se han identificado: nombre, periodo y descripción de la enfermedad.
- **Observable:** Representa los rasgos presentes en la mascota enferma que ayudaran a identificar a la enfermedad que la aqueja. Estos observables pueden ser:

Síntoma: Representa los síntomas característicos de las enfermedades. Esta clase tiene como atributo : nombre y descripción.

Condición: Son las condiciones en la que se encuentra la mascota enferma y que el veterinario toma nota a la hora de examinar al paciente. Estas pueden

agrupar subclases que definan tipos especiales de condiciones como la clase Temperatura Corporal.

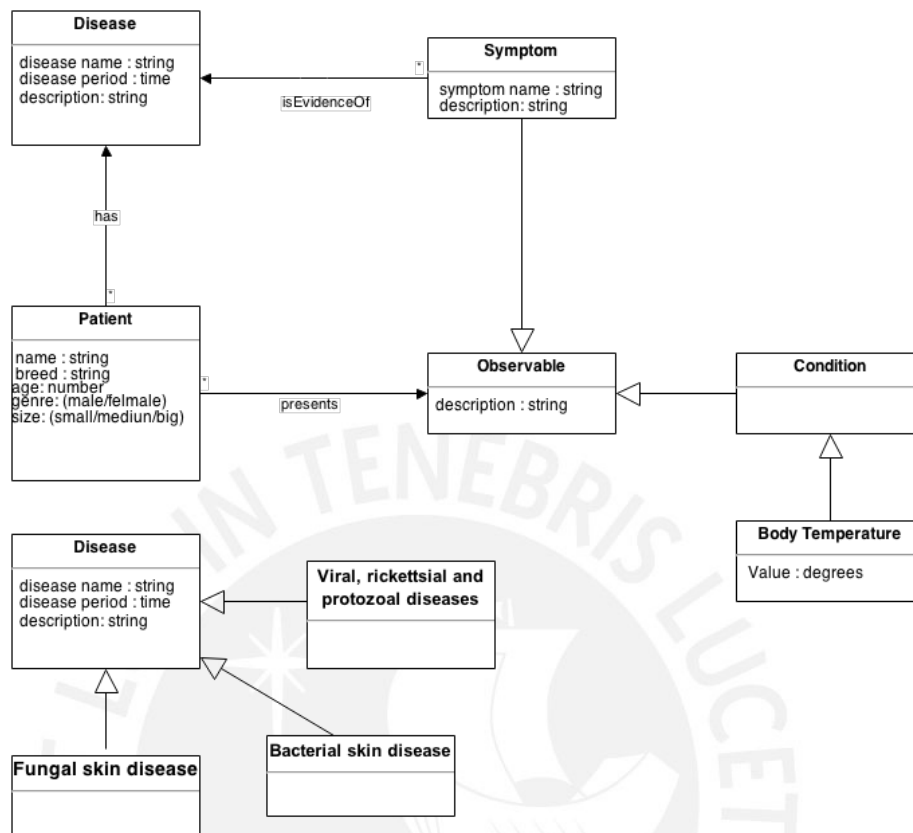


Imagen 7: Esquema del dominio de la veterinaria para el diagnóstico

Las clases principales y básicas para realizar un diagnóstico son enfermedad y síntomas, puesto que lo que se quiere saber es cuáles enfermedades posiblemente estén afectando al perro y que la información que se posee o ha identificado el dueño son los síntomas. Las demás clases permiten realizar un diagnóstico a más detalle y poder discernir cuál de las posibles enfermedades es la que realmente tiene la mascota.

Las relaciones entre estas clases son: Enfermedad-Síntoma, debida que los síntomas son la evidencia de la existencia de una o muchas enfermedades. Enfermedad-Paciente, debido a que un paciente presenta una enfermedad la cual va a ser diagnosticada. Paciente-Observable, debido a que el paciente presenta manifestaciones observables de la enfermedad, donde existen dos tipos de observable: los síntomas y las condiciones.

4.3.2. Estructura de Inferencias

A continuación se muestra el diagrama de la estructura de inferencias que se utilizará para realizar el diagnóstico. Se utilizarán cuatro tipos de inferencias que permitan llegar a un diagnóstico final. El diagrama muestra las inferencias en elipses y los roles de entrada y roles de salida, en recuadros. Se tienen una función de transferencia, representada por un recuadro de bordes redondeados y dos roles especiales, Hypothesis y Observable. La función de transferencia permite obtener información externa en respuesta a una petición, esto para obtener nueva evidencia.

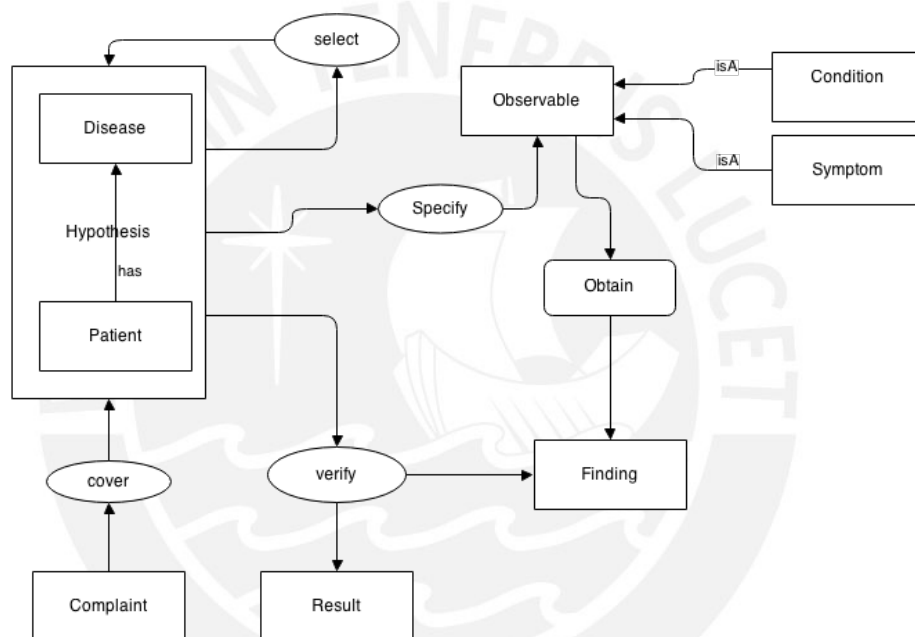


Imagen 8: Imagen propia, adaptada del modelo de diagnóstico de CommonKADS [Schreiber, 2000]

El rol Hypothesis representa un posible resultado, este se expresa en función de las clases enfermedad y paciente, donde un paciente puede tener una o más enfermedades. Varios posibles resultados se agruparán en un conjunto denominado diferencial de hipótesis. Cada hipótesis del diferencial es un resultado potencial que indica que un paciente determinado tiene una enfermedad específica. Se evaluará cada hipótesis mediante las inferencias presentadas en el diagrama y se llegará a un resultado, que es el conjunto de hipótesis después de haber descartado todas las posibilidades nulas. Este resultado se

ordenará asignando prioridades a las hipótesis de acuerdo a su valor de verdad.

El rol Observable puede ser de dos tipos, Condition y Symptom, el primero se refiere a una condición del paciente independiente de la enfermedad, por ejemplo, la temperatura corporal o el tamaño. El segundo es un síntoma asociado a una enfermedad existente en la base de conocimiento.

Cada una de las inferencias cumple un rol específico que permitirá llevar a cabo el diagnóstico. La plantilla propone construir un conjunto de hipótesis basadas en una queja que da inicio al ciclo de inferencia [Schreiber, 2000]. Este conjunto de hipótesis se validará iterativamente contrastando cada hipótesis con nuevos hallazgos en base a observables, evidencias, obtenidos en cada iteración [Schreiber, 2000]. El ciclo termina cuando no se puede descartar más hipótesis del conjunto o se tiene una sola hipótesis [Schreiber, 2000]. Se podría cortar la ejecución cuando se tienen N hipótesis para ofrecer un número específico de enfermedades.

La inferencia *cubrir* recibe como entrada el rol queja y genera el diferencial de hipótesis [Schreiber, 2000]. La inferencia *seleccionar* permite escoger una hipótesis candidata para obtener un hallazgo en base a un factor de preferencia a determinarse [Schreiber, 2000]. Esta hipótesis permite especificar nuevos observables a consultar al usuario y aseverar nuevos hechos o hallazgos [Schreiber, 2000]. Con esta nueva evidencia se recorre el diferencial validando las hipótesis y descartando las que sean anuladas por la nueva evidencia [Schreiber, 2000].

4.3.3. Conclusión

En esta sección se ha definido dos estructuras que permitirán construir y diseñar la ontología para soportar las inferencias que permitirán simular el proceso de diagnóstico. Además, se realizó una revisión de ontologías existentes que representen un dominio médico y estén orientadas al diagnóstico de enfermedades. En base a las ontologías revisadas se ha diseñado una estructura básica que permitirá construir la ontología. Se ha propuesto utilizar la plantilla de diagnóstico definida en la metodología CommonKADS y adaptarla para solucionar el problema en cuestión. Se presentaron las inferencias que de-

berán implementarse, la ontología deberá soportar las inferencias planteadas. Con estas herramientas es posible comenzar la construcción de la base de conocimiento. En resultados posteriores se estudiarán más a fondo las inferencias y cómo se implementarán para el prototipo propuesto.

4.4. Resultado Esperado 5: Construcción de la Base de Conocimiento

En esta sección se explicará el funcionamiento uso de la herramienta Protégé para la construcción de la base de datos del conocimiento. Se mencionan las principales clases de la definidas dentro de la ontología. Además se explicará cómo se definió las relaciones que tendrán estas clases para poder tener acceso entre ellas. También se explicará cómo insertar a los individuos en las respectivas clases y a relacionar los individuos. Los puntos mencionados se desarrollan con el objetivo de establecer una guía que permita comprender el uso de Protégé y la implementación realizada del conocimiento veterinario.

4.4.1. Sobre la herramienta Protégé

Protégé es una plataforma libre y de código abierto que proporciona un conjunto de herramientas para la construcción de modelos de dominio y las aplicaciones basadas en el conocimiento de las ontologías [Knublauch et al., 2004]. Provee dos formas de trabajo en Web y en Desktop (actualmente usada en el proyecto) [Knublauch et al., 2004].

Protégé Desktop brinda un soporte completo para el OWL 2 (Lenguaje de Ontología Web), y conexiones directas en memoria a la descripción de los razonadores lógicos como HermiT y Pellet [Knublauch et al., 2004].

Además, es compatible con la creación y edición de una o varias ontologías en un único espacio de trabajo a través de una interfaz de usuario completamente personalizable [Knublauch et al., 2004]. Las herramientas de visualización permiten una navegación interactiva de las relaciones de la ontología [Knublauch et al., 2004]. Las operaciones de refactorización incluyen la combinación de ontología, mover axiomas entre ontologías, renombrar múltiples entidades y más [Knublauch et al., 2004].

4.4.2. Definición de las Clases

El primer paso es crear las clases básicas y necesarias de donde se pueda obtener la información para realizar el diagnóstico. Estas clases son *Disease* y *Symptom*, y heredarán de *Thing*. Casi todas las clases que se crean son subordinadas a *Thing*, la cual esta ubicada en la sección “*Classes*”.

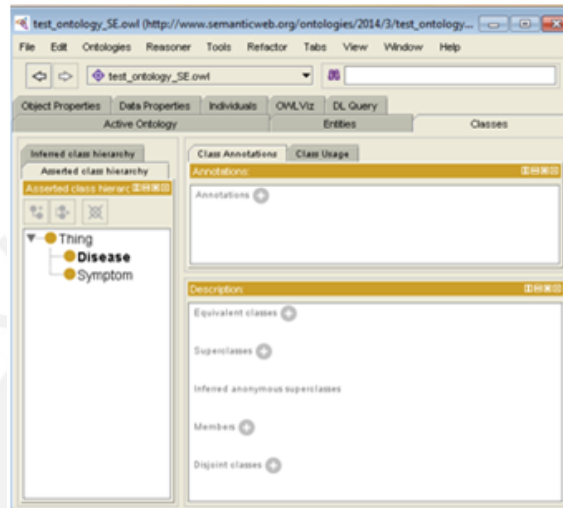


Imagen 9: Definición de las clases *Disease* y *Symptom* en Protégé

4.4.3. Relaciones entre Clases

El siguiente paso es definir la relación que existe entre las clases. La relación que vincula a *Symptom* y *Disease* es *evidenceOf* puesto que los síntomas son evidencia de que el paciente presenta una enfermedad.

Para definir esta relación, esta es creada desde la sección “*Object Properties*”. También se define la dirección en que afecta la relación, donde “*Domain*” es la clase que da inicio a la relación y “*Range*” es en quien recae la relación. Para este proyecto, el *Domain* es *Symptom* y el *Range* es *Disease*.

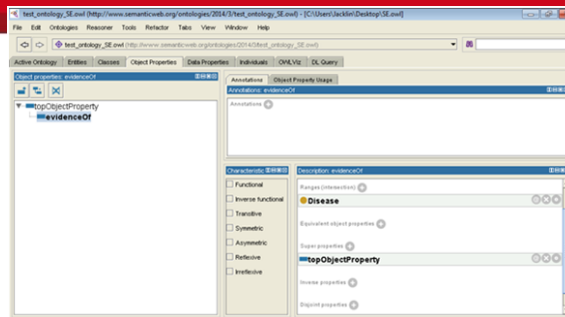


Imagen 10: Definición de la relación *evidenceOf* en Protégé

4.4.4. Inserción de Individuos

Una vez que se tiene las clases y la relación entre estas, se procede a insertar individuos en cada clase. Esto se realiza desde la sección “*Individuals*”, donde hay que especificar a qué clase pertenece el individuo insertado.

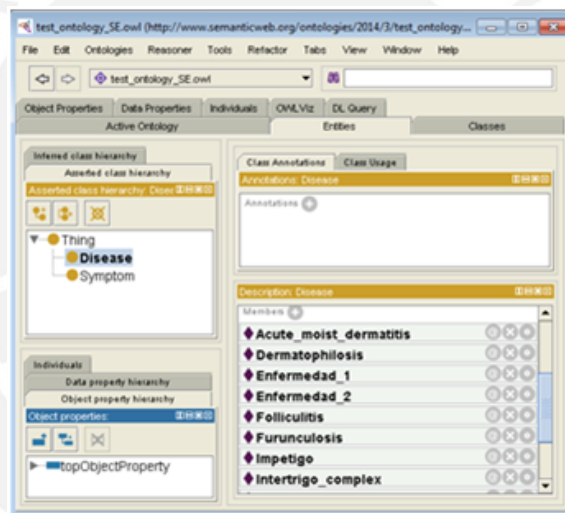


Imagen 11: Creación de los individuos de la clase *Disease* en Protégé

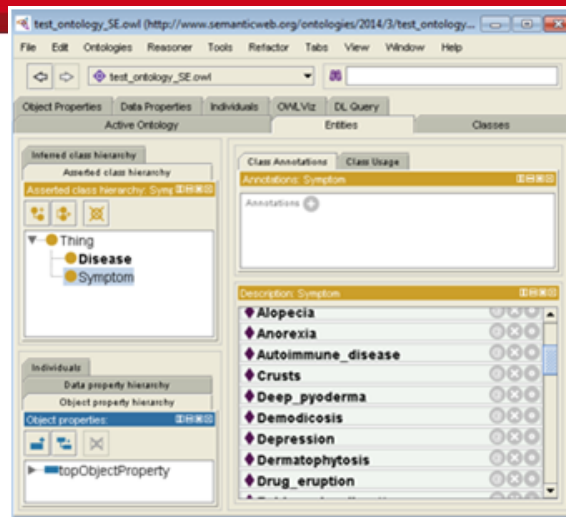


Imagen 12: Creación de los individuos de la clase *Symptom* en Protégé

4.4.5. Relaciones entre Individuos

Para relacionar individuos, esta se realiza desde el individuo de la clase - *Domain*. En el individuo de dicha clase que se elija a relacionar, existe la opción “*Object property assertions*”, en la cual se puede seleccionar la relación - “*Object property*” que se desea usar y el individuo de la clase - *Range* al que también afectará la relación.

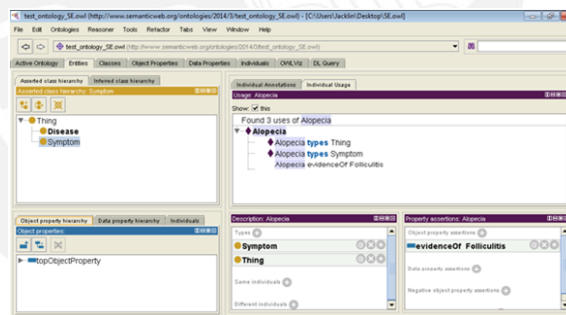


Imagen 13: Asignación de la relación *evidenceOf* entre el síntoma *Alopecia* y la enfermedad *Folliculitis* en Protégé

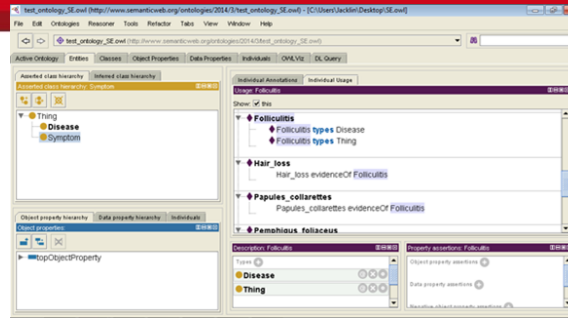


Imagen 14: Relación de los síntomas asociados a la enfermedad *Folliculitis* en Protégé

4.4.6. Conclusión

Se realizó la construcción de la base de datos del conocimiento exitosamente. Se tienen opciones adicionales que se podrían agregar a la base de datos del conocimiento con la intención de explotar su uso y beneficios ofrecidos; entre estas tenemos: causa, tratamiento y signo de la enfermedad. Protégé desktop demostró ser de gran utilidad en cuestiones de creación y manipulación de la ontología, es una herramienta bastante intuitiva de fácil uso para quienes desarrollan ontologías. Esta base de datos del conocimiento está basada en los datos necesarios y básicos en el diagnóstico de enfermedades en el campo de la veterinaria. Con este avance se procederá a realizar la integración en una herramienta piloto de la base de conocimiento y el motor de inferencia.

5. Capítulo 5

5.1. Introducción

En este capítulo se presentará el análisis y diseño del motor de inferencia junto a los resultados del proceso de construcción de este. Se hará una revisión de los procesos mentales que sigue el razonamiento humano, del proceso de diagnóstico en general y de la plantilla correspondiente a la tarea de diagnóstico de la metodología CommonKADS. Se presentará además el proceso de diseño y construcción del motor de inferencia que incluye la construcción de cada una de las funciones propuestas en esta plantilla.

5.2. Resultado Esperado 6: Análisis del Razonamiento para el Diagnóstico

En esta sección se presentará brevemente cómo es posible representar computacionalmente el proceso de razonamiento específicamente el de diagnóstico. Además se presentará la definición de la tarea de diagnóstico según la metodología CommonKADS.

5.2.1. Procesos Mentales

El artículo Human Problem Solving: The state of the theory in 1970 presenta una estrategia para alcanzar la comprensión del razonamiento humano y los procesos que permiten llevar a cabo la resolución de problemas complejos. En este artículo se sientan las bases para la comprensión del sistema humano de resolución de problemas con el fin de poder replicarlo de manera computacional [Simon and Newell, 1971].

Para construir una teoría de resolución de problemas es necesario predecir el comportamiento de un solucionador de problemas, entender que procesos se dan y que mecanismos permiten realizar estos procesos [Simon and Newell, 1971]. Además se debe conocer los fenómenos internos y externos al problema que lo afectan [Simon and Newell, 1971]. Finalmente, se debe tener en cuenta las habilidades de resolución de problemas que son adquiridas y las que posee el solucionador de problemas [Simon and Newell, 1971].

Los autores señalan, como características del sistema de resolución de problemas humano, la capacidad de leer, almacenar, borrar y comparar patrones para resolver complejas tareas no numéricas [Simon and Newell, 1971]. El solucionador de problemas construye un espacio de problema para enfrentar una tarea, este espacio es una representación de esta tarea [Simon and Newell, 1971]. Son características del sistema humano de resolución de problemas realizar los procesos racionales uno por uno y no en paralelo, el acceso rápido a una memoria a corto plazo bastante limitada y el acceso lento a una memoria a largo plazo prácticamente ilimitada [Simon and Newell, 1971].

Para llegar a una solución es necesario recorrer los estados de conocimiento del espacio del problema hasta que un estado contenga la respuesta [Simon and Newell, 1971]. Incluso problemas simples pueden tener espacios de problemas inmensos, por ejemplo, el

problema de las torres de Hanoi tiene 81 estados para solo 4 discos y el ajedrez llega a altas potencias de 10^{20} [Simon and Newell, 1971]. El poder de las heurísticas utilizadas en el razonamiento humano reside en su capacidad para examinar pequeñas regiones prometedoras del espacio del problema en busca de la solución [Simon and Newell, 1971]. Esto se hace estructurando la tarea en un espacio de problema que facilite realizar una búsqueda selectiva entre las posibles soluciones [Simon and Newell, 1971].

5.2.2. Diagnóstico Veterinario

Según el artículo “Frames and Boxes: Is this how veterinarians also think?” el diagnóstico veterinario sigue una tendencia al uso de algoritmos predeterminados desplazando al aprendizaje tradicional [Milani, 2008]. Estos algoritmos consisten en agregar los síntomas a un árbol de decisiones y pruebas para luego realizar pruebas que los lleven a un diagnóstico [Milani, 2008].

La autora señala que una práctica para un diagnóstico recomendable es: observar al paciente, hacer preguntas y escuchar a las respuestas [Milani, 2008]. El uso de algoritmos predefinidos es un método eficiente de diagnóstico pero existen problemas que no encajan en estos marcos de referencia [Milani, 2008].

Considera que un problema del uso de estas herramientas es que desmotiva a los estudiantes de refinar su creatividad y pensamiento crítico [Milani, 2008]. El problema es que el médico veterinario, por falta de estas habilidades, trate de hacer encajar el problema del paciente a la fuerza [Milani, 2008]. Estos algoritmos permiten reducir la cantidad de variables que debe afrontar el veterinario; de ser muchas, sobrepasan su capacidad; siendo pocas, llevan a una respuesta muy encasillada que podría no ser la más adecuada [Milani, 2008].

La herramienta construida deberá proveer el soporte que permita al veterinario lidiar con esta gran cantidad de variables para que pueda centrarse en la observación y análisis del paciente. No deberá pretender encajar el proceso de razonamiento limitando las respuestas. Deberá permitir al veterinario explotar sus potenciales de razonamiento crítico permitiéndole acceder al conocimiento rápidamente, realizando sugerencias sujetas a su

evaluación, permitiendo hacer un seguimiento al razonamiento de la herramienta que retroalimenta al del médico y ofreciendo múltiples posibilidades en vez de una respuesta única y definitiva.

5.2.3. Tarea de Diagnóstico Segun CommonKADS

El método de diagnóstico a utilizarse estará basado en la plantilla propuesta por la metodología CommonKADS para la tarea de diagnóstico. Segun CommonKADS la tarea de diagnóstico consiste en encontrar un fallo que causa que un sistema no funcione correctamente, plantean como ejemplo clásico el diagnóstico de un dispositivo electrónico [Schreiber, 2000].

Segun la plantilla la tarea de diagnóstico debería contar con un modelo del comportamiento del sistema pero en algunos casos se puede reducir a una tarea de clasificación reemplazando el modelo de comportamiento por una asociación directa entre síntomas y fallas [Schreiber, 2000]. Esta estrategia de utilizar un modelo de causas puede potenciarse agregando elementos de probabilidad, esto implica realizar una extracción de conocimiento sobre las probabilidades asociadas a cada relación causal.

5.2.4. Conclusión

Tras revisar el razonamiento humano para la resolución de problemas se determinó que una alternativa es representar el problema de manera estructurada y recorrer espacios prometedores de esta estructura en búsqueda de la solución. Las ontologías se utilizaron para construir el espacio de problemas. La plantilla propuesta por CommonKADS se utilizará para recorrer este espacio en busca de la solución. Se debe tener en cuenta que la herramienta debe reducir las variables que maneja el veterinario dejándole libertad para explorar diversas posibilidades, evitando encasillarlo, en la medida de lo posible. Con la seguridad de poder reproducir mediante simulación los procesos mentales de solución de problemas abordamos la tarea de diseño del motor de inferencia.

5.3. Resultado Esperado 7: Diseño del Motor de Inferencia

Esta sección está enfocada en el diseño del motor de inferencia que permitirá realizar el diagnóstico basado en la ontología construida. Como se ha explicado en capítulos

anteriores el algoritmo a utilizar para realizar las inferencias es el propuesto por la metodología CommonKADS en la plantilla de la tarea de diagnóstico. A continuación se detallarán las inferencias presentadas en el modelo a nivel de implementación.

5.3.1. Plantilla de Diagnóstico CommonKADS

La siguiente imagen muestra la definición de la tarea diagnóstico propuesta por CommonKads, en la sección Control-Structure se muestra como interactúan los elementos para llegar a una solución [Schreiber, 2000].

```

TASK diagnóstico;
ROLES:
INPUT:
  queja: 'hallazgo que inicia el proceso de diagnóstico';
OUTPUT:
  fallas: 'las fallas que pueden haber causado la queja';
  evidencia: 'la evidencia recolectada durante el diagnóstico';
END TASK diagnóstico;

TASK-METHOD cobertura-por-causas:
REALIZES: diagnóstico;
DECOMPOSITION:
  INFERENCES: cubrir, seleccionar, especificar, verificar;
  TRANSFER-FUNCTIONS: obtener;
ROLES:
INTERMEDIATE:
  diferencial: 'conjunto de soluciones candidatas';
  hipótesis: 'solución candidata';
  resultado: 'booleano que indica el resultado de una prueba';
  hallazgo-esperado: 'información que se esperaría encontrar';
  hallazgo-actual: 'información observada en la práctica'
CONTROL-STRUCTURE:
  WHILE NUEVA-SOLUCION cubrir(queja -> hipótesis) DO
    diferencial := hipótesis ADD diferencial;
  END WHILE
  REPEAT
    seleccionar( diferencial -> hipótesis );
    especificar( hipótesis -> observable );
    obtener( observable -> hallazgo );
    evidencia := hallazgo ADD evidencia;
  FOR-EACH hipótesis IN diferencia DO
    verificar( hipótesis + evidencia -> resultado);
    IF resultado == falso
      THEN diferencial := diferencial SUBTRACT hipótesis;
    END IF
  END FOR-EACH
  UNTIL
    SIZE diferencial <= 1 OR 'no quedan más observables';
  END REPEAT
  fallos := diferencial;
END TASK-METHOD
  
```

Imagen 15: Traducido de la especificación de la tarea de diagnóstico de la metodología CommonKADS [Schreiber, 2000]

Este muestra la definición de la tarea de diagnóstico. La tarea recibe de entrada una queja, que es un observable del sistema que produce malestar a un agente [Schreiber, 2000]. Para el problema planteado esta queja viene a ser un síntoma de alguna enfermedad. Como salida se tienen dos elementos, las fallas que resultan del diagnóstico del sistema y la evidencia que soporta la elección de estas fallas [Schreiber, 2000]. Para el

problema planteado, sería la enfermedad o posibles enfermedades y el conjunto de síntomas que se utilizó para descartar y aceptar las hipótesis.

En la imagen se puede ver también el proceso que permitirá llevar a cabo la tarea de diagnóstico. CommonKADS propone una cobertura por causas para llevar a cabo tareas de diagnóstico [Schreiber, 2000]. Esta se utilizará en conjunto con la ontología que provee un catálogo de relaciones entre causas y manifestaciones.

Se definen las siguientes funciones de inferencia: cubrir, seleccionar, especificar y verificar. Además se cuenta con una función de transferencia, obtener, cuyo papel es proveer información externa sobre algún observable y su presencia en el sistema.

Se definen como roles intermediarios: diferencial, hipótesis, resultado, hallazgo esperado, hallazgo actual. El diferencial de hipótesis es el conjunto de hipótesis candidatas para convertirse en una falla [Schreiber, 2000]. Una hipótesis es una posible falla que deberá ser contrastada con la evidencia [Schreiber, 2000]. Un resultado es el resultado de realizar una prueba para descartar o reafirmar una hipótesis [Schreiber, 2000]. Un hallazgo es nueva información obtenida, forma parte de la evidencia, puede ser esperado o actual dependiendo de su origen, inferido o proporcionado por un agente externo [Schreiber, 2000].

Las funciones de inferencia propuestas son las siguientes:

La primera, cubrir (cover), recibe una queja y obtiene todos los posibles fallos asociados a esa queja y las carga al diferencial [Schreiber, 2000]. Para el problema planteado obtiene todas las enfermedades asociadas a un síntoma inicial.

La segunda, seleccionar (select), selecciona una hipótesis del diferencial basada en algún factor de preferencia asociado al conocimiento representado [Schreiber, 2000]. Para el presente problema se escogerá la enfermedad con menos síntomas en común, de ser posible con un síntoma distintivo.

La tercera, especificar (specify), consiste en obtener una entidad observable que permitirá limitar el número de posibles respuestas [Schreiber, 2000]. Este observable puede

dar información sobre la hipótesis seleccionada o sobre el resto de hipótesis [Schreiber, 2000]. Para el presente problema se buscará un síntoma que permita descartar o reforzar la enfermedad seleccionada.

La cuarta, verificar (verify), recibe, mediante la función de transferencia obtener, información externa sobre el observable determinado por la función anterior [Schreiber, 2000]. Con esta nueva evidencia se recorre el diferencial evaluando cada hipótesis y descartando las que no concuerden con la evidencia [Schreiber, 2000]. Para el presente problema se consultará sobre la existencia o no de un síntoma y se adicionará el resultado al conjunto de evidencia sobre síntomas presentes.

La estructura de control presenta un flujo básico y muestra cómo estas funciones de inferencia interactúan entre sí para llevar a cabo el diagnóstico.

5.3.2. Diagrama de Diseño

A continuación se presenta el diagrama de diseño que permitirá construir el motor de inferencia e integrarlo al prototipo. En este se muestran las tres clases principales, la rutina principal, el motor de inferencia y el acceso a la ontología.

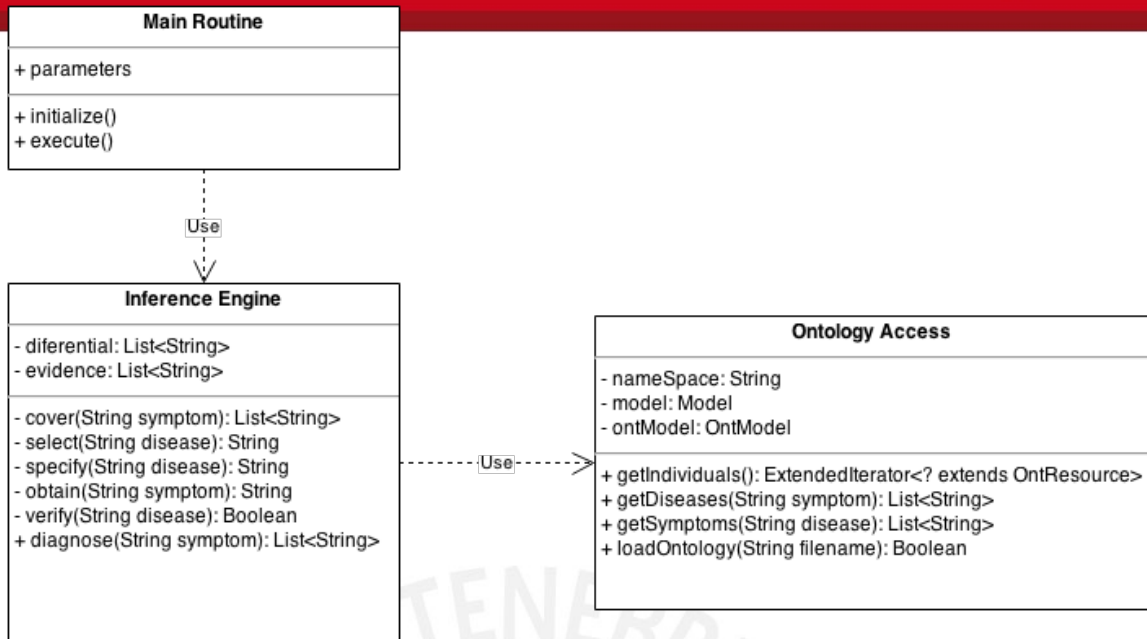


Imagen 16: Diagrama de diseño construido por los autores.

Como se puede observar en el diagrama la rutina principal utilizará el motor de inferencia que realizará la mayor parte del trabajo. Este implementará la estructura de control presentada previamente y utilizará la clase de acceso a la ontología para acceder al conocimiento representado. Se implementarán funciones de acceso que devolverán cadenas que sirven como identificadores para síntomas y enfermedades. A continuación se explicará a detalle cómo se implementarán las funciones de inferencia.

5.3.3. Implementación de cada Función de Inferencia

5.3.3.1. Cubrir

```

COVER ( Queja Y)
  ListaEnfermedades LX := { X / X ∈ Enfermedad ∧ es_causa_de(X,Y) }
  WHILE (LX != vacio)
    hipótesis := LX.cabeza
    diferencial := hipótesis ADD diferencial
    LX:= LX cola
  END WHILE
END COVER

```

Imagen 17: Pseudocódigo para la función cubrir.

5.3.3.2. Seleccionar

```

SELECT ( ListaEnfermedades LX)
  ListaSintomas LY := { Y / Y ∈ Sintoma ∧ es_causa_de(X,Y) ∧ X ∈ LX }
  ContadorSintomas CLY := vacio
  MenorValor M = ∞
  MenorElemento MY = vacio
  WHILE (LY != vacio)
    IF ( existe(CLY[LY.cabeza] ) THEN
      CLY[LY.cabeza] = CLY[LY.cabeza] + 1
    ELSE
      CLY[LY.cabeza] = 1
    END IF
    IF( CLY[LY.cabeza] < M) THEN
      M = CLY[LY.cabeza];
      MY = LY.cabeza
    END IF
  END WHILE
  EnfermedadSeleccionadas SX = {X / X ∈ Enfermedad ∧ es_causa_de(X,MY)}
  RETURN SX.cabeza
END SELECT

```

Imagen 18: Pseudocódigo para la función seleccionar.

5.3.3.3. Especificar

```

SPECIFY ( Enfermedad SX )
  ListaSintomas LY := { Y / Y ∈ Sintoma ∧ es_causa_de(SX,Y) }
  MenorValor M = ∞
  MenorElemento MY = vacio
  WHILE (LY != vacio)
    IF( CLY[LY.cabeza] < M) THEN
      M = CLY[LY.cabeza];
      MY = LY.cabeza
    END IF
    LY = LY cola
  END WHILE
  RETURN MY
END SPECIFY

```

Imagen 19: Pseudocódigo para la función especificar.

5.3.3.4. Verificar

```

VERIFY (Enfermedad X )
  Boolean resultado = verdadero;
  FOR EACH (evidencia AS value)
    IF es_causa_de ( value.sintoma, X) THEN
      IF ( ! value.presente ) THEN
        resultado = falso;
        RETURN resultado
      END IF
    END IIF
  END FOR EACH
  RETURN resultado;
END VERIFY

```

Imagen 20: Pseudocódigo para la función verificar.

5.3.4. Conclusión

Se ha seguido la plantilla propuesta por la metodología commonKADS para la tarea de diagnóstico y se ha detallado cómo es que esta se implementará para el proyecto. Se tiene las clases principales de diseño, motor de inferencia y acceso a la ontología, que soportarán la estructura de diagnóstico. Se cuenta con una descripción de las funciones de inferencia a nivel de lógica y pseudocódigo. Con estas herramientas bien definidas se puede proceder a la parte de construcción del motor de inferencia.

5.4. Resultado Esperado 8: Construcción del Motor de Inferencia

En esta sección se explicará el funcionamiento general de la librería Jena. Se mencionan las principales clases de la librería y su funcionalidad. Además se explicará cómo acceder a las clases definidas en la ontología, a los individuos correspondientes a cada clase y a las propiedades definidas en la ontología. También se explicará como obtener los individuos relacionados a otro individuo en base a las propiedades. Se detallará la implementación de las inferencias y se presentarán los resultados del proceso.

5.4.1. Librería Jena

Se utilizó la versión 2.11.1 de la librería Jena para el presente proyecto. La librería se incluye en el proyecto como una librería externa cuyos principales archivos Jar son jena-arq, jena-core, jena-iri, jena-sdb y jena-tdb. Jena implementa varias clases que permiten trabajar el manejo de archivos RDF y permite trabajar con OWL a través de una extensión y el uso de razonadores que permitan procesar la ontología [Foundation, 2014].

Una clase importante de la librería es Model representa el modelo de la ontología y permite realizar la lectura de un archivo mediante el método read. Este permite cargar la ontología a una clase de Java para poder trabajar con esta y la lee en varios formatos («RDF/XML», «N-TRIPLE», «TURTLE» (o «TTL») y «N3») siendo «RDF/XML» el formato por defecto [Foundation, 2014].

Jena provee una clase especial para el trabajo con ontologías, la clase OntModel, esta se implementa sobre la clase InfModel que a su vez hereda de la clase Model [Foundation, 2014]. Se puede utilizar la clase ReasonerRegistry y la función getOWLReaso-

ner() para obtener un razonador sobre Owl por defecto [Foundation, 2014]. También se puede utilizar la clase ModelFactory y la función createOntologyModel(PelletReasonerFactory.THE_SPEC) para instanciar un modelo que use el razonador Pellet [Foundation, 2014]. De esta manera se tiene un modelo que puede cargar una ontología a partir de un archivo, utilizando la función read(), y acceder al conocimiento almacenado.

5.4.2. Acceso a las Clases, Individuos y Propiedades de la Ontología

El acceso a las clases de la ontología se hace a través del modelo declarado, usando la función getOntClass(uri) que recibe el uri (unified resource identifier) de la clase [Foundation, 2014]. Este se construye concatenando el namespace de la ontología y el nombre de la clase. Este devuelve un objeto Clase [Foundation, 2014].

La recuperación de individuos se realiza a través del objeto Clase y la función listInstances() que devuelve un iterador en una lista de individuos correspondientes a la clase [Foundation, 2014]. También se puede recuperar un individuo específico por su uri, desde el modelo de la ontología, utilizando la función getOntResource(uri) [Foundation, 2014].

Una propiedad se puede obtener por su uri desde el modelo de la ontología utilizando la función getProperty(uri) [Foundation, 2014]. Los individuos asociados por esta propiedad se puede obtener a través del modelo de la ontología utilizando la función listStatements(s, p, v) que recibe tres parámetros [Foundation, 2014]. El primero es el sujeto, el segundo la propiedad y el tercero el valor de esta propiedad [Foundation, 2014].

Los tres elementos pueden ser llamados con valor null, en este caso la función devuelve todas las afirmaciones existentes [Foundation, 2014]. Si se declara la propiedad p, la función devuelve todas las afirmaciones que existen respecto a esta propiedad [Foundation, 2014]. Si se declara el valor de la propiedad, se busca todas las afirmaciones en las que se obtenga este objeto como valor final [Foundation, 2014]. Si se declara el sujeto, se busca todas las afirmaciones en las que se aplique una propiedad correspondiente al sujeto [Foundation, 2014].

5.4.3. Implementación de las Inferencias y Resultados

Se implementó las inferencias basadas en el pseudocódigo presentado en el documento de diseño. Cada función de inferencia fue desarrollada independientemente y realiza la función que se especificó en los documentos previos. Frente a la posibilidad de realizar menor procesamiento agrupando tareas en una sola se decidió dejar cada función de inferencia dentro del rol definido para dejar abierta la posibilidad de realizar optimizaciones a nivel de cada inferencia.

El proceso principal sigue la estructura de control planteada en la metodología CommonKADS. Este recibe un síntoma inicial, la queja, y utiliza la inferencia *cover* para llenar la estructura diferencial que contiene una lista de enfermedades candidatas a solución. También recibe el número de enfermedades que se desea como respuesta.

Luego se genera el ciclo de filtro de resultados, mientras el diferencial contenga más respuestas de las deseadas y se obtenga nueva evidencia se seguirá buscando descartar enfermedades. La inferencia *select* recupera la enfermedad más prometedora, en este caso, con mayor probabilidad de ser descartada pues se desea reducir el diferencial al tamaño indicado. Este criterio se incluirá en el conocimiento encapsulado en la ontología.

La inferencia *specify* selecciona un síntoma correspondiente a la enfermedad elegida buscando que este permita descartar o reforzar la presencia de la enfermedad. Consideramos que el hecho de que un síntoma característico de una enfermedad no esté presente, es decir, que el usuario indique que no está presente, es suficiente evidencia para descartar la enfermedad, sin embargo su presencia no la afirma definitivamente; se convierte en evidencia fuerte de su presencia.

Finalmente la función *obtain* obtiene información sobre la presencia del síntoma seleccionado y esta se agrega a la evidencia. La función *verify* se encarga de contrastar todas las enfermedades contra esta nueva evidencia y descarta las que estén en contradicción con esta.

A continuación se presenta la salida en consola de una prueba para el motor llamando a

la función de diagnóstico con la queja *Lesions* para 1, 3 y 5 enfermedades resultantes. En el tercer caso se devolverá el diferencial completo pues para el síntoma *Lesions* existen 5 enfermedades.

```

Patient has the symptom Painful_to_touch: Y/N
Y
Added to evidence: Patient shows Painful_to_touch
The symptom Painful_to_touch was checked.
Patient has the symptom Mucocutaneous_erythema: Y/N
Y
Added to evidence: Patient shows Mucocutaneous_erythema
The symptom Mucocutaneous_erythema was checked.
Patient has the symptom Intertrigo_complex: Y/N
N
Added to evidence: Patient does not show Intertrigo_complex
The hypothesis Mucocutaneous_pyoderma was discarded.
Patient has the symptom Ulceration: Y/N
N
Added to evidence: Patient does not show Ulceration
The hypothesis Methicillin_resistant_Staphylococcus_aureus was discarded.
Patient has the symptom Erythema_multiforme: Y/N
N
Added to evidence: Patient does not show Erythema_multiforme
The hypothesis Furunculosis was discarded.
Patient has the symptom Zinc_responsive_dermatitis: Y/N
Y
Added to evidence: Patient shows Zinc_responsive_dermatitis
The symptom Zinc_responsive_dermatitis was checked.
Patient has the symptom Seborrheic_disease: Y/N
N
Added to evidence: Patient does not show Seborrheic_disease
The hypothesis Dermatophilosis was discarded.
*****RESPONSE*****
Acute_moist_dermatitis
    
```

Imagen 21: Resultado de la función diagnóstico para síntoma *Lesions* y 1 enfermedad.

```

Patient has the symptom Painful_to_touch: Y/N
Y
Added to evidence: Patient shows Painful_to_touch
The symptom Painful_to_touch was checked.
Patient has the symptom Mucocutaneous_erythema: Y/N
N
Added to evidence: Patient does not show Mucocutaneous_erythema
The hypothesis Mucocutaneous_pyoderma was discarded.
Patient has the symptom Ulceration: Y/N
N
Added to evidence: Patient does not show Ulceration
The hypothesis Methicillin_resistant_Staphylococcus_aureus was discarded.
*****RESPONSE*****
Furunculosis
Dermatophilosis
Acute_moist_dermatitis
    
```

Imagen 22: Resultado de la función diagnóstico para síntoma *Lesions* y 3 enfermedades.

```

*****RESPONSE*****
Mucocutaneous_pyoderma
Methicillin_resistant_Staphylococcus_aureus
Furunculosis
Dermatophilosis
Acute_moist_dermatitis
|
    
```

Imagen 23: Resultado de la función diagnóstico para síntoma *Lesions* y 5 enfermedades.

5.4.4. Conclusión

Se finaliza la construcción del motor de inferencia de acuerdo al diseño especificado, probando la utilidad de la librería Jena para procesar el conocimiento almacenado en la ontología. Se mostró las principales funciones de las clases Model y OntModel para cargar la ontología, asignar un razonador y acceder a los elementos y relaciones. Se demuestra también que la plantilla de diagnóstico de CommonKADS puede adaptarse para trabajar sobre ontologías. Con este resultado se puede proceder a construir el prototipo que integre los resultados obtenidos.

6. Capítulo 6

6.1. Introducción

En este capítulo se presentará el prototipo construido para integrar las herramientas construidas en los resultados previos. Estos dos últimos resultados apuntan a desarrollar una herramienta simple que haga uso de la base de conocimiento y el motor de inferencia. Se presentan los principales aspectos de la herramienta y la interfaz que permitirá interactuar con la herramienta.

6.2. Resultado Esperado 9: Construcción del Prototipo

En esta sección se presentará el prototipo que integrará los resultados previos en una herramienta que permita evidenciar cómo es posible dar soporte al diagnóstico veterinario haciendo uso de la base de conocimiento y el motor de inferencia. Se planteará un conjunto de funcionalidades básicas que cubrirá la herramienta. Se presentará brevemente el diseño de la herramienta donde se podrá ver cómo se realizó la integración de los resultados previos.

6.2.1. Funcionalidades

El objetivo principal del prototipo es permitir probar el funcionamiento del motor de inferencia y la base de conocimiento integrados en una herramienta que dé soporte al diagnóstico permitiendo la interacción con el usuario. No se trata de un sistema basado en conocimiento, es solo una herramienta provisional que permita probar los resultados

principales.

Las principales funcionalidades que debe cubrir el prototipo son las siguientes:

- Debe permitir al usuario ingresar un síntoma que de inicio a la tarea de diagnóstico del motor de inferencia.
- Debe permitir al usuario ingresar un número que delimite la cantidad de enfermedades que se devolverá como respuesta.
- Debe permitir al motor de inferencia pedir información externa sobre la presencia o ausencia de síntomas.
- Debe permitir al usuario entregar información sobre la presencia o ausencia de síntomas.
- Debe permitir al usuario hacer seguimiento al razonamiento del motor de inferencia.
- Debe permitir al motor de inferencia mostrar en pantalla el resultado obtenido.

6.2.2. Diseño

A continuación se presenta el diagrama de diseño del prototipo.

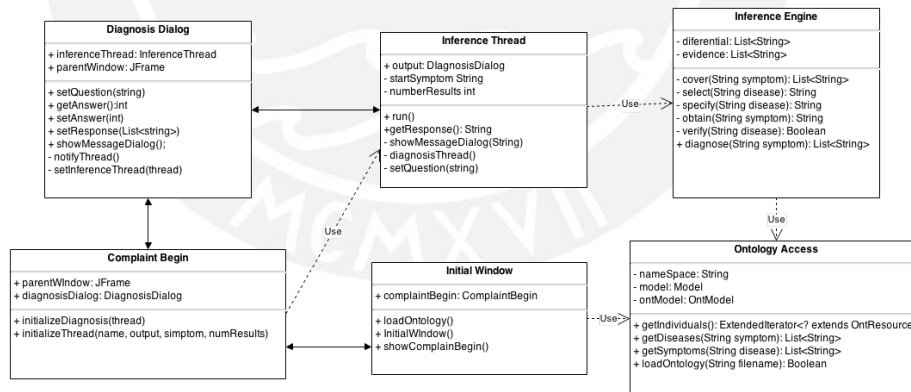


Imagen 24: Diagrama de diseño del prototipo elaborado por los autores.

En el diagrama se puede observar las vistas y las funciones que permiten que estas se comuniquen con el núcleo a través del hilo de inferencia. La ventana inicial da paso al ingreso de parámetros y una vez iniciado el diagnóstico la ventana de diálogo se encarga de la comunicación con el motor de inferencia.

El núcleo, conformado por la base de conocimiento y el motor de inferencia, se presentaron previamente en la sección correspondiente al resultado 8. La diferencia está en la comunicación que se realiza a través del hilo.

6.2.3. Conclusión

Se tiene una definición concisa del prototipo, se ha tenido en cuenta lo suficiente para que cumpla el objetivo planteado. Se presentaron las funcionalidades básicas y un diagrama que permite examinar los principales componentes del prototipo. Se pensaron funcionalidades extra como búsqueda de síntomas, navegación libre entre las preguntas del motor de inferencia cambiando las entradas, información complementaria sobre tratamiento y descripción. Estas funcionalidades se plantean como mejoras y se deja abierta la posibilidad de ampliar el prototipo construido a un sistema completo.

6.3. Resultado Esperado 10: Interfaz Gráfica

En esta sección se presentará la interfaz gráfica creada para mostrar el proceso de diagnóstico mediante el uso de inferencias basadas en ontologías. Se presentará los prototipos de las principales pantallas en la interfaz, así como el flujo de las pantallas con las que interactuará el experto que use la herramienta.

6.3.1. Diseño

6.3.1.1. Ventana de Inicio

La ventana inicial de prototipo de herramienta de diagnóstico muestra un mensaje de bienvenida, una breve descripción de la herramienta y un botón “Start” con el que se dará inicio al proceso de diagnóstico. Esta permite elegir entre dos ontologías construidas para demostrar que la base de conocimiento puede cambiarse con facilidad.



Imagen 25: Ventana inicial de prototipo de la herramienta de diagnóstico

6.3.1.2. Ventana de Inicio

Es la ventana que dará inicio al proceso diagnóstico. En está se puede apreciar las indicaciones a seguir: primero, se ingresará un síntoma representativo de la mascota; luego se indicará el número de posibles enfermedades que se espera tener como resultado.

El síntoma que se ingresará debe pertenecer al conjunto de enfermedades registradas en la base de datos del conocimientos y el número de resultados esperados debe ser un entero positivo.

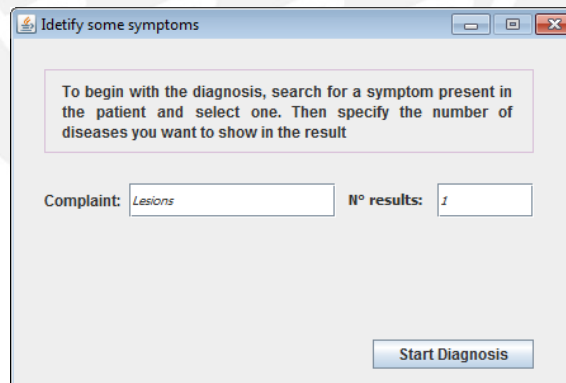


Imagen 26: Ventana de configuración de la herramienta de diagnóstico

6.3.1.3. Ventana de Evaluación de Síntomas

Esta ventana permite al experto aceptar o rechazar los posible síntomas que tendría la mascota. Los síntomas que aparecen en la ventana provienen de un análisis del primer síntoma registrado en la ventana de configuración. Permite descartar las enfermedades que no posee o no se asemejan al que realmente afecta a la mascota.

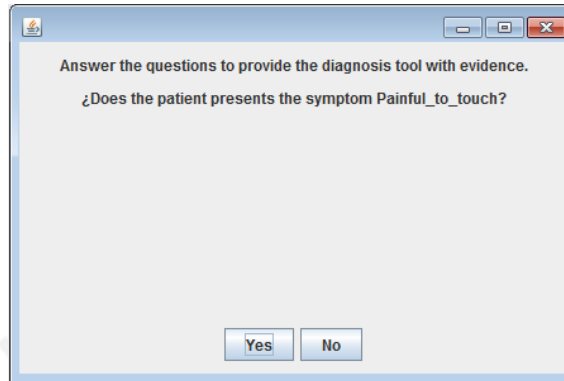


Imagen 27: Ventana de aceptación y rechazo de síntomas

6.3.1.4. Ventana de Resultados

Es la ventana final del proceso de diagnóstico, donde se muestra la enfermedad o conjunto de enfermedades posibles que podrían estar afectando a la mascota. También se muestra un botón "Back", que permitirá realizar todo el proceso de diagnóstico desde la ventana de configuración.



Imagen 28: Ventana de resultados

6.3.1.5. Mensajes

Durante el uso de la Ventana de aceptación y rechazo de síntomas, se muestran diferentes mensajes que permiten hacer seguimiento al razonamiento del motor de inferencia.

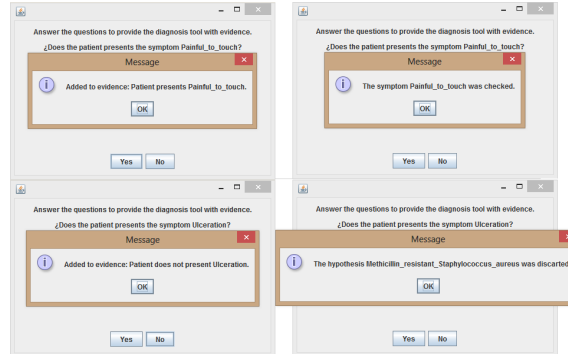


Imagen 29: Diversos mensajes de la herramienta

6.3.2. Flujo

La imagen mostrada a continuación es un diagrama de estado que explica el funcionamiento de la herramienta. El proceso se inicia mediante la identificación de un síntoma representante y proporcionando un número deseado de enfermedades sugeridas. Una vez validado, se selecciona un síntoma representativo y su presencia se verifica a través de entradas externas indicadas por el usuario. Si el síntoma no está presente, las enfermedades relacionadas se descartan; de lo contrario, la evidencia de las enfermedades relacionadas se fortalece. En ambos casos, se muestran mensajes de retroalimentación a los usuarios. Este proceso se repite hasta que no hayan más pruebas o se alcance el número de resultados.

Esto permite al usuario descartar de forma interactiva o validar los síntomas presentes que están relacionados con la actual hipótesis considerada en el proceso de diagnóstico. Después de cada validación, el usuario puede revisar las enfermedades relacionadas que se vieron afectados por las nuevas evidencias.

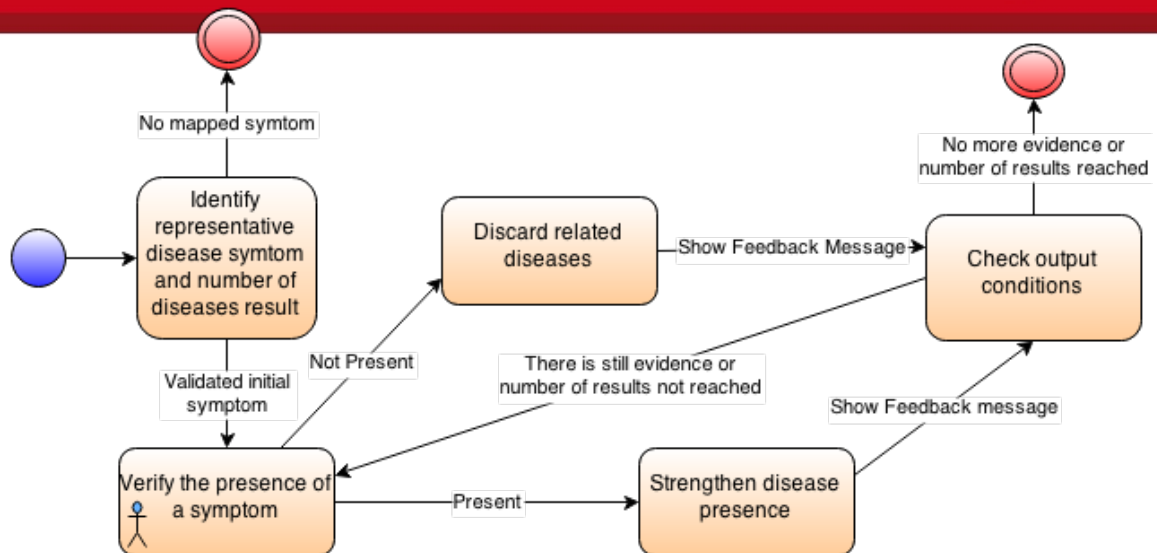


Imagen 30: Diagrama de estados funcionales de la herramienta de diagnóstico.

6.3.3. Resultados

El sistema se validó contra el criterio del veterinario usando una ontología de enfermedades gastrointestinales. Después de elegir dos enfermedades y responder en base a su experiencia, tres respuestas se obtuvieron, las dos enfermedades que escogió y una tercera que reconoció como válida de acuerdo a sus respuestas.

La retroalimentación recibida apunta a desarrollar una interfaz más flexible para interactuar con el conocimiento. El prototipo actual es muy estricto pues guía el diagnóstico y le da la facilidad al veterinario de probar varias combinaciones durante el diagnóstico. La opinión general del veterinario es optimista, dada la falta de automatización de los procesos de las clínicas veterinarias en el país, resulta interesante la posibilidad de desarrollar una herramienta que amplíe las posibilidades durante el diagnóstico. También se recibieron indicaciones de usabilidad, el veterinario debe poder interactuar con el animal durante la evaluación, tipear puede resultar poco práctico. Un enfoque basado en sugerencias y opciones para seleccionar puede solucionar este problema.

Finalmente, el sistema se probó con 30 casos históricos y se logró obtener un diagnóstico en el 70 % de los casos, esto en base a la información anotada por el veterinario para

cada caso. La ontología debe ser especializada con ayuda de un médico veterinario y el prototipo debe adaptarse para ofrecer una interfaz más flexible.

6.3.4. Conclusión

Se realizó exitosamente la construcción de la interfaz gráfica con la intención de mostrar visualmente como se realiza el diagnóstico haciendo uso de inferencias basadas en ontologías. Se tienen opciones adicionales a este prototipo de herramienta que se podrían agregar con intención de mejorar la interacción con el usuario; entre estas opciones tenemos: tener un botón de retroceso en la “Ventana de aceptación y rechazo de síntomas” en caso uno se equivoque en decidir la presencia o no presencia de algún síntoma que proponga la herramienta; también, en la “Ventana de configuración” un campo de texto con autocompletado que brinde sugerencias de síntomas registrados en la base de datos de conocimiento. Las opiniones de veterinarios son en general optimistas con críticas a la flexibilidad de la herramienta. Dado que es un prototipo, quedan pendientes cambios a la interfaz y el acceso al conocimiento que mejorarán este aspecto. Se realizaron pruebas con casos reales obteniendo un 70 % de casos diagnosticados en base a la información anotada por el veterinario. Con esta herramienta se completa la integración del motor de inferencia y la base de datos del conocimiento desarrollados.

7. Capítulo 7

7.1. Discusión de Resultados

7.1.1. Introducción

En este capítulo se presenta la discusión de resultados alcanzados a lo largo del proyecto. Para cada objetivo esperado se evaluará hasta qué punto se logró solucionar el problema planteado, en que contribuyen los resultados dentro del estado del arte y que posibilidades quedan abiertas para el futuro en base a los resultados actuales. Finalmente se realizarán sugerencias para abordar el mismo problema con variaciones a nivel de métodos de inferencia, herramientas y dominio del conocimiento.

7.1.2. Objetivo Específico 1:

- Problema: El problema detectado en el planteamiento de la problemática fue la falta de herramientas orientadas a brindar soporte al diagnóstico veterinario.
- Solución: Se propuso como solución buscar y proponer herramientas relacionadas a las ontologías y los posibles razonadores sobre ontologías que permitan desarrollar una herramienta que dé soporte al diagnóstico veterinario.
- Logros: Entre las ontologías revisadas se encuentran las ontologías de la base de datos de OBO. También se exploró las posibilidades que ofrece el lenguaje OWL. Se evaluaron los formatos OBO y OWL con sus respectivos editores OBO-Edit y Protégé. Por el lado de las inferencias se revisó la herramienta Jena y la herramienta BaseVisor, ambas construidas en Java. Se utilizó Jena y se hizo una revisión de los razonadores compatibles con Jena: Pellet, Fact y Racer, de los cuales se utilizó Pellet.
- Propuesta: Se deja abierta la posibilidad de utilizar herramientas propuestas como BaseVisor, otro razonador u otra propuesta de ontología. Además sería interesante evaluar más a profundidad las diferencias entre los formatos OBO y OWL e incluso intentar elaborar una interfaz que relacione ambos formatos.

7.1.3. Objetivo Específico 2

- Problema: En la problemática se mencionó que el conocimiento para diagnóstico es complejo.
- Solución: Utilizar ontologías para representar el conocimiento complejo y hacerlo computable.
- Logros: Existen distintos métodos para representar el conocimiento, se utilizaron ontologías con el objetivo de aprovechar su potencial, la reutilización de elementos para reducir el tiempo de desarrollo. Se documentó el proceso de construcción de una base de conocimiento utilizando la herramienta Protégé. Se siguió la metodología CommonKADS para el diseño del núcleo de la ontología. Como resultado se obtuvo una ontología conteniendo el conocimiento necesario para realizar el diagnóstico en un formato computable.

- Propuesta: Es interesante explorar las posibilidades de diagnóstico en distintos dominios, se puede realizar diagnóstico de otro tipo de enfermedades, otra especie o cualquier tipo de sistema que presente fallas por mal funcionamiento de sus componentes. Además se podría replicar el procedimiento de construcción de una ontología más compleja a detalle siguiendo la metodología CommonKads (capítulos 4, 5 y 7).

7.1.4. Objetivo Específico 3

- Problema: Otro problema mencionado en la problemática es que el razonamiento que realiza el veterinario para realizar un diagnóstico es complejo.
- Solución: El objetivo planteado es utilizar métodos de inferencia basados en ontologías para realizar un diagnóstico. Se investigó el proceso mental racional del humano y las necesidades de la veterinaria a nivel de soporte al diagnóstico. Se encontró una propuesta para la tarea de diagnóstico en la metodología CommonKADS, metodología para la construcción de sistemas basados en conocimiento.
- Logros: Se documentó el proceso de construcción de un motor de inferencia que trabaje sobre ontologías. Se documentó el uso de la API Jena para acceder y procesar el conocimiento. Se utilizó la plantilla correspondiente a la tarea diagnóstico propuesta por CommonKADS.
- Propuesta: Se deja como posibilidad futura el cambiar los métodos de inferencia utilizados, agregar elementos probabilísticos. También se puede modificar la estructura de inferencia o buscar maneras de optimizar las funciones de inferencia existentes.

7.1.5. Objetivo Específico 4

- Problema: Este objetivo apunta a integrar los resultados previos para contribuir a solucionar el problema planteado.
- Solución: La propuesta de solución es construcción de un prototipo de una herramienta que brinde soporte al diagnóstico veterinario utilizando inferencia basada en ontologías.

- Logros: El prototipo de la herramienta demuestra que es factible apuntar a desarrollar un sistema de diagnóstico que utilice inferencia basada en ontologías. Se documentó el prototipo y la interfáz gráfica del mismo. Se construyó un ejemplo concreto utilizando la plantilla de diagnóstico propuesta en la metodología CommonKADS obteniendo un 70% de casos diagnosticados sobre 30 casos reales y opiniones optimistas de médicos veterinarios con críticas a la flexibilidad de la herramienta.
- Propuesta: Se deja abierta la posibilidad de construcción de un sistema completo, una ampliación del prototipo que abarque aspectos más generales como administración de pacientes e historial clínico y permita una navegación libre por el conocimiento de la ontología, este se realizará como proyecto personal. También es muy recomendable la construcción de una herramienta web que brinde orientación sobre posibles malestares comunes que aquejan a las mascotas. Además de la tarea de diagnóstico, CommonKADS tiene definiciones para las tareas de clasificación, asesoramiento, monitoreo, síntesis, diseño de configuraciones, asignación, planeamiento y programación de tareas. Estas pueden explorarse como propuestas de proyectos basados en esta metodología.

8. Capítulo 8

8.1. Conclusión

Habiendo recorrido un largo camino de investigación sobre un área de la gestión del conocimiento en un inicio desconocida pero de gran interés se llega al fin del presente proyecto. La motivación de este proyecto fue el interés personal por la representación de conocimiento, la simulación de procesos mentales racionales y un interés personal en la medicina veterinaria. Se buscó contribuir al problema de la falta de soporte tecnológico para la tarea de diagnóstico veterinario. Se tiene como objetivo general del proyecto aplicar métodos y técnicas inferencia sobre ontologías para contribuir a solucionar este problema. A continuación se presentan las conclusiones del proyecto agrupadas por objetivos específicos.

Respecto al objetivo específico 1, la revisión de herramientas aplicables, se encontraron herramientas útiles para la representación y procesamiento de conocimiento. Se revisó

el lenguaje de ontologías de la web semántica OWL y las ontologías OBO (*Open Biological and Biomedical Ontologies*) con su formato OBO-Format. El lenguaje usado para la construcción de la base de datos del conocimiento es OWL DL modelado en la herramienta Protégé. Se eligió sobre OBO por ser más estable, mejor documentado y con mayor compatibilidad de herramientas, además, se cuenta con gran parte del contenido de OBO en formato OWL.

Para la realización del motor de inferencia se usara el lenguaje Java con la librería Jena, para manejar la estructura de la ontología, y el razonador lógico Pellet. Se revisó también la herramienta BaseVIsor que también permite trabajar como ontologías pero, a diferencia de Jena, no cuenta con clases especializadas para el procesamiento de ontologías.

El objetivo específico 2 apuntaba a encapsular el conocimiento necesario para realizar el diagnóstico veterinario. Se logró diseñar y construir la base de datos del conocimiento con las clases necesarias para realizar el diagnóstico, siendo las principales *enfermedad* y *síntoma* en los roles de *observable del sistema* y *estado del sistema* propuestos por CommonKADS. Se evaluaron de 3 tipos de enfermedades de la piel: causadas por hongos (*fungus skin disease*), por bacterias (*bacterial skin disease*) y por virus, rickettsias y protozoarios (*viral, rickettsial and protozoal diseases*). Todo esto en base a la plantilla del CommonKADS y a la información de libros especializados en este tipo de enfermedades, y usando la herramienta Protégé para modelarla.

El objetivo específico 3, simular el razonamiento de un médico veterinario utilizando métodos de inferencia basada en ontologías, se alcanzó mediante la construcción de un motor de inferencia que trabaja sobre la base de conocimiento construida. Se presentó un resumen de las características del razonamiento humano para la resolución de problemas: procesamiento de patrones y reducción de espacio de problemas. También se evaluó la perspectiva de la medicina veterinaria para el mejoramiento de la tarea de diagnóstico, se plantea dejar de lado tareas que pueden automatizarse para dar paso al pensamiento crítico.

Se presentó el diagrama de inferencia adaptado de la plantilla CommonKADS con las definiciones de las principales funciones de inferencia: *cubrir*, *seleccionar*, *especificar*,

verificar. También se explicó el proceso de diagnóstico según la plantilla, se realiza un levantamiento de enfermedades o hipótesis en base a un síntoma o queja inicial, se seleccionan enfermedades en base a síntomas y, sobre estas, síntomas a evaluar. Esto teniendo en cuenta las enfermedades y síntomas que permitan descartar una enfermedad. Los síntomas seleccionados deben ser característicos de la enfermedad para que permitan realizar el descarte, esto es un diagnóstico diferencial. Además se presentaron salidas en consola de los resultados obtenidos con el motor para un síntoma determinado y un número de respuestas deseadas ingresado por el usuario.

Finalmente, el objetivo específico 4 consiste en una herramienta que permita integrar los resultados obtenidos y ver su funcionamiento. Se logró construir un prototipo con las funcionalidades básicas: Ingresar parámetros iniciales, ingresar información sobre la presencia o ausencia de síntomas, hacer seguimiento a las decisiones tomadas en el proceso de inferencia y mostrar las conclusiones obtenidas por el motor. Este cuenta, como principales componentes, con una ventana principal, una ventana de ingreso de parámetros iniciales, ventana de diálogo, un hilo que comunica la interfaz con el motor de inferencia y el núcleo conformado por el motor de inferencia y una interfaz de acceso a la ontología. El prototipo fue evaluado por un médico veterinario quien se mostró optimista frente al potencial de la herramienta. Hubo observaciones sobre el acceso al conocimiento y el proceso de diagnóstico, este debería ser más flexible permitiendo al veterinario probar varias combinaciones de síntomas. Se probó la herramienta con 30 casos reales logrando diagnosticar un 70 % en base a la información recogida por un médico veterinario.

Se llega al final del proyecto con la satisfacción de haber cumplido los objetivos planteados y haber realizado una contribución a la solución de la problemática planteada. El prototipo construido demuestra que es factible aplicar la inferencia basada en ontologías a la construcción de soluciones que ataquen el problema de brindar soporte al diagnóstico en medicina veterinaria.

Referencias

- [Bechhofer, 2003] Bechhofer, S. (2003). Owl reasoning examples. *University of Manchester*.
- [Bouamrane et al., 2010] Bouamrane, M.-M., Rector, A., and Hurrell, M. (2010). Experience of using owl ontologies for automated inference of routine pre-operative screening tests. In *The Semantic Web—ISWC 2010*, pages 50–65. Springer.
- [Casado-Lumbreras et al., 2012] Casado-Lumbreras, C., Rodríguez-González, A., Álvarez Rodríguez, J. M., and Colomo-Palacios, R. (2012). Psydis: Towards a diagnosis support system for psychological disorders. *Expert Systems with Applications*, 39(13):11391 – 11403.
- [Chandrasekaran et al., 1999] Chandrasekaran, B., Josephson, J. R., and Benjamins, V. R. (1999). What are ontologies, and why do we need them? *Intelligent Systems and Their Applications, IEEE*, 14(1):20–26.
- [Corcho and Gómez-Pérez, 2000] Corcho, O. and Gómez-Pérez, A. (2000). A roadmap to ontology specification languages. In *Knowledge Engineering and Knowledge Management Methods, Models, and Tools*, pages 80–96. Springer.
- [Decker et al., 1999] Decker, S., Erdmann, M., Fensel, D., and Studer, R. (1999). *Ontobroker: Ontology based access to distributed and semi-structured information*. Springer.
- [Foundation, 2014] Foundation, T. A. S. (Copyright © 2011–2014). Jena documentation overview.
- [García-Crespo et al., 2010] García-Crespo, Á., Rodríguez, A., Mencke, M., Gómez-Berbís, J. M., and Colomo-Palacios, R. (2010). Oddin: Ontology-driven differential diagnosis based on logical inference and probabilistic refinements. *Expert Systems with Applications*, 37(3):2621–2628.
- [Gómez, 2007] Gómez, Leonardo y Atehortua, C. y. O. (2007). La influencia de las mascotas en la vida humana. *Revista Colombiana de Ciencias Pecuarias*, 20:377–386.
- [Gruber, 1995] Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies*, 43(5):907–928.

- [Guarino, 1998] Guarino, N. (1998). *Formal Ontology in Information Systems: Proceedings of the First International Conference (FIOS'98), June 6-8, Trento, Italy*, volume 46. IOS press.
- [Guha et al., 1998] Guha, R. V., Lassila, O., Miller, E., and Brickley, D. (1998). Enabling inferencing. Published as <http://www.w3.org/TandS/QL/QL98/pp/enabling.html>.
- [Guo and Kraines, 2009] Guo, W. and Kraines, S. (2009). Discovering relationship associations in life sciences using ontology and inference. pages 10–17. cited By (since 1996)4.
- [Hendler, 2004] Hendler, J. A. (2004). Frequently asked questions on w3c's web ontology language (owl). "<http://www.w3.org/2003/08/owlfaq>" Accessed: 2013-03-02.
- [Houe et al., 2011] Houe, H., Gardner, I. A., Nielsen, L. R., et al. (2011). Use of information on disease diagnoses from databases for animal health economic, welfare and food safety purposes: strengths and limitations of recordings. *Acta Veterinaria Scandinavica*, 53(Suppl 1):S7.
- [Knublauch et al., 2004] Knublauch, H., Ferguson, R. W., Noy, N. F., and Musen, M. A. (2004). The protégé owl plugin: An open development environment for semantic web applications. In *The Semantic Web—ISWC 2004*, pages 229–243. Springer.
- [Lian et al., 2012] Lian, H. H., Bao, W. X., and Wang, Y. H. (2012). Animal diseases diagnosis expert system based on hsmc-svm. *Applied Mechanics and Materials*, 198:1036–1041.
- [Maetschke et al., 2012] Maetschke, S. R., Simonsen, M., Davis, M. J., and Ragan, M. A. (2012). Gene ontology-driven inference of protein–protein interactions using inducers. *Bioinformatics*, 28(1):69–75.
- [Matheus et al., 2006] Matheus, C., Dionne, B., Parent, D., Baclawski, K., and Kokar, M. (2006). Basevisor: A forward-chaining inference engine optimized for rdf/owl triples. In *Digital Proceedings of the 5th International Semantic Web Conference, ISWC*.
- [McGuinness et al., 2004] McGuinness, D. L., Van Harmelen, F., et al. (2004). Owl web ontology language overview. *W3C recommendation*, 10(2004-03):10.

- [Milani, 2008] Milani, M. (2008). Frames and boxes: Is this how veterinarians also think? *The Canadian Veterinary Journal*, 49(2):195.
- [Noy and McGuinness, 2001] Noy, N. and McGuinness, D. L. (2001). Ontology development 101. *Knowledge Systems Laboratory, Stanford University*.
- [Parsia and Sirin, 2004] Parsia, B. and Sirin, E. (2004). Pellet: An owl dl reasoner. In *Third International Semantic Web Conference-Poster*, volume 18.
- [Paterson, 2009] Paterson, S. (2009). *Manual of skin diseases of the dog and cat*. John Wiley & Sons.
- [Pearl, 1988] Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.
- [Rodríguez et al., 2009] Rodríguez, A., Mencke, M., Alor-Hernandez, G., Posada-Gomez, R., Gomez, J. M., and Aguilar-Lasserre, A. A. (2009). Medboli: Medical diagnosis based on ontologies and logical inference. In *eHealth, Telemedicine, and Social Medicine, 2009. eTELEMED'09. International Conference on*, pages 233–238. IEEE.
- [Schreiber, 2000] Schreiber, G. (2000). *Knowledge engineering and management: the CommonKADS methodology*. the MIT Press.
- [Serrano, 2008] Serrano, César y Arcila, V. (2008). La importancia social del profesional en medicina veterinaria. *REDVET. Revista electrónica de Veterinaria*, IX(6).
- [Simon and Newell, 1971] Simon, H. A. and Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist*, 26(2):145.
- [Smith et al., 2007] Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungall, C. J., et al. (2007). The obo foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, 25(11):1251–1255.
- [Studer et al., 1998] Studer, R., Benjamins, V. R., and Fensel, D. (1998). Knowledge engineering: principles and methods. *Data & knowledge engineering*, 25(1):161–197.
- [Studer et al., 2000] Studer, R., Decker, S., Fensel, D., and Staab, S. (2000). Situation and perspective of knowledge engineering. *Knowledge Engineering and Agent Technology*. IOS Press, Amsterdam.

- [Uschold et al., 1996] Uschold, M., Gruninger, M., et al. (1996). Ontologies: Principles, methods and applications. *Knowledge engineering review*, 11(2):93–136.
- [Van Harmelen and McGuinness, 2004] Van Harmelen, F. and McGuinness, D. L. (2004). Owl web ontology language overview. *World Wide Web Consortium (W3C) Recommendation*.
- [Vera Ramirez, 2009] Vera Ramirez, N. (2009). El negocio de las mascotas en el Perú. "http://www.foyel.com/paginas/2009/10/886/el_negocio_de_las_mascotas_en_peru/" Accessed: 2012-09-23.

