

Técnicas de Representación del Conocimiento

Inteligencia Artificial 2

1^{do} Angely Mendez

Escuela de Informática

Universidad Nacional de Trujillo

Trujillo, Perú

t052701020@unitru.edu.pe

2^{ero} Ciara Mendez

Escuela de Informática

Universidad Nacional de Trujillo

Trujillo, Perú

t022700920@unitru.edu.pe

Resumen—Este documento es una investigación que describe tres Técnicas de Representación del Conocimiento, las cuales son: Redes Semánticas, Marcos (Frames) y Ontologías. Este artículo es la recopilación de varias investigaciones a nivel nacional e internacional, en los que abordan un problema, describen de manera formal y computacionalmente el dominio sobre el cual se quiere operar, considerando tanto a los objetos como a las relaciones entre ellos combinando estructuras de datos y procedimientos interpretativos que, empleados correctamente hicieron que un programa exhiba un comportamiento inteligente; respecto a las mencionadas técnicas, se explican sus definiciones y dos investigaciones de cada una, con la finalidad de entender la importancia de las ciencias de la computación en la vida diaria.

Palabras claves:—Representación del conocimiento, Redes Semánticas, Marcos, Ontologías, ciencias de la computación.

I. INTRODUCCIÓN

Los seres humanos son buenos para comprender, razonar e interpretar el conocimiento. Y utilizando este conocimiento, pueden realizar diversas acciones en el mundo real. Pero, ¿cómo podemos representar este conocimiento?

En inteligencia artificial, el conocimiento se puede representar de varias formas empleando varias técnicas, según la estructura del conocimiento o la perspectiva del diseñador o incluso el tipo de estructura interna utilizada. Una representación efectiva del conocimiento debe ser lo suficientemente

rica como para incluir el conocimiento requerido para resolver el problema. Debe ser natural, compacto y mantenible.

Entonces este informe presenta información al respecto, el cual está organizado de la siguiente manera: en primer lugar, se explican los conceptos teóricos: definiciones de las tres técnicas, luego se da énfasis en dos investigaciones para cada una de ellas y para finalizar las conclusiones más relevantes.

II. TÉCNICAS DE REPRESENTACIÓN DEL CONOCIMIENTO

II-A. Redes Semánticas

Están basadas en la idea de que los objetos o los conceptos pueden ser unidos por alguna relación. Estas relaciones se representan usando una liga que conecte dos conceptos. Los nodos y las ligas pueden ser cualquier cosa, dependiendo de la situación a modelar. [2]. En un grafo o red semántica los elementos semánticos se representan por nodos como se muestra en la Figura 1.

Elementos de la representación:

- Las **instancias** se representan por constantes.
- Las **clases** se representan por constantes.
- Las **relaciones** clase-superclase se representan por hechos de la forma es un ($> \text{clase} >$, $< \text{super-clase} >$).

- Las relaciones instancia–clases se representan por hechos de la forma inst ($\langle instancia \rangle, \langle clase \rangle$).
- Cada propiedad se representa por un predicado binario de la forma prop ($\langle instanciaoclase \rangle, \langle propiedad \rangle, \langle valor \rangle$).
- La constante inicio representa la clase inicial de la jerarquía.
- Las propiedades de una instancia es una lista de pares **atributo–valor**.

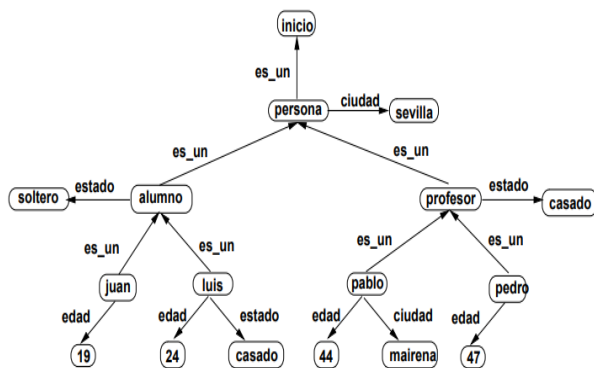


Figura 1. Ejemplo Redes Semánticas- Tomado Universidad de Sevilla, España.

A continuación, se presentan dos investigaciones donde se evidencia la aplicación de la técnica Redes Semánticas:

1. En la investigación “*Análisis de causas de retraso en proyectos de construcción complejos: un enfoque de análisis de redes semánticas*”, [11].

Problema: Los retrasos se encuentran entre los adversarios más cruciales para el éxito y el rendimiento de los proyectos de construcción, lo que hace que el análisis y la gestión de retrasos sean una tarea crítica para los gerentes de proyectos. Esta tarea será muy complicada en proyectos de gran envergadura como la construcción, que suelen consistir en una red compleja de entidades heterogéneas en continua interacción. Los enfoques y métodos tradicionales para el análisis de las demoras y sus causas han sido criticados por su capacidad para manejar proyectos complejos y por considerar las interrelaciones entre las causas de las demoras.

Resultados: Al abordar esta brecha, esta investigación presenta un enfoque alternativo para el análisis de las causas del retraso mediante la adopción del método de análisis de redes semánticas (SNA). El documento informa los resultados de una investigación de retrasos en proyectos de construcción en el sector Petróleo-Gas-Petroquímico usando SNA. Tal como se muestra en la Figura 2, las causas pertenecen a los conceptos Group A: Retrasos relacionados con los procesos de contratación, Group B: Retrasos relacionados con las negociaciones iniciales, C: Retrasos relacionados con el proceso de planificación, D: Retrasos relacionados con el proceso de control, donde se examinó y discutió empíricamente la capacidad del método para identificar y clasificar las causas de las demoras, el análisis ayudó a una identificación más precisa del principal factor que causa retrasos en los proyectos de construcción petroquímicos iraníes, las ‘*deficiencias de las negociaciones iniciales*’, una etapa en la que los puntos de vista y expectativas contradictorias de los diferentes actores ejercen presión sobre el proceso.

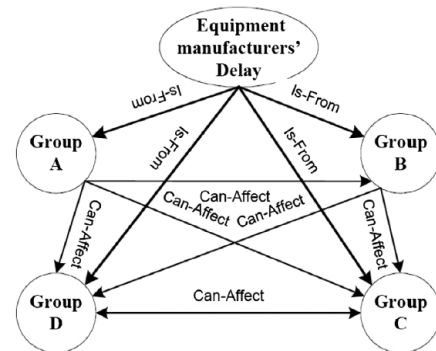


Figura 2. Red semántica de causas de retraso de IPI.

Importancia: El documento argumenta que el SNA conduce a una **comprensión más integral** de las principales causas de retraso en proyectos grandes y complejos, lo que permite una mejor identificación y mapeo de las interrelaciones entre estos factores discretos lo que puede ayudar a los gerentes a seleccionar las medidas

apropiadas para eliminarlas.

2. Otra investigación es “*Aplicación de un método para el análisis de las redes semánticas en pacientes que sufrieron un accidente cerebro vascular*”, [10].

Problema: Numerosas investigaciones han estudiado las alteraciones en las relaciones semánticas entre conceptos producidas por diversas patologías, centrándose en su mayoría en pacientes con demencia (Alzheimer en particular) y esquizofrenia y ninguna en accidente Cerebro Vascular (ACV); por ello se presenta esta aplicación de un método de estimación de distancias semánticas, denominado DISTSEM, en pacientes que sufrieron un Accidente Cerebro Vascular (ACV). Este método se basa en la Teoría de Propagación de la Activación, la cual supone que la memoria semántica se organiza en forma de una red de acuerdo con líneas de similitud semántica. Así, cuantas más propiedades en común haya entre dos conceptos, más cerca se encontrarán en la red.

Resultados: Los resultados ponen de manifiesto que algunos pacientes luego de sufrir un ACV realizan estimaciones de proximidad entre conceptos, distintas a las esperables de acuerdo al desempeño del grupo control. Las respuestas no fueron homogéneas. Se hallaron sujetos que establecieron escasas asociaciones entre los pares de palabras. También se hallaron sujetos que encontraron más vinculaciones que el grupo control, incluso encontrando similitudes en aquellos pares que no tenían vinculación semántica y que utilizaron justificaciones para realizar sus estimaciones, que no se corresponden con categorías semánticas sino que se basan principalmente en la funcionalidad. La lectura de estos resultados a la luz de la Teoría de Propagación de la Activación indica que dichos pacientes tienden a fallar en el proceso de búsqueda dentro de la red, particularmente en la regulación de la propagación de la activación que es necesaria para realizar la tarea.

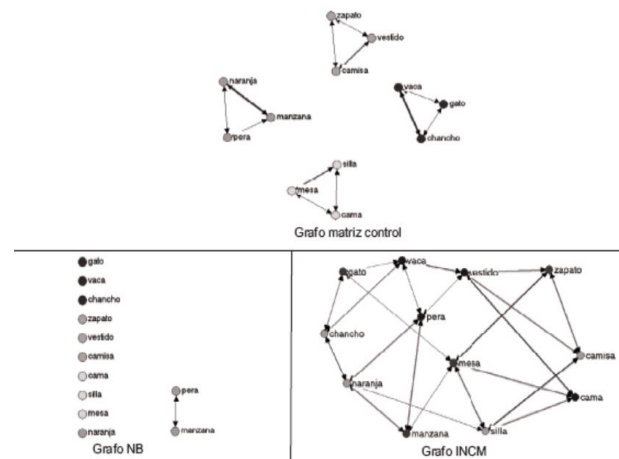


Figura 3. Grafo Ilustrativo de la Red semántica.

Importancia: El método DISTSEM permite capturar una red semántica conformada a partir de las estimaciones de distancias semánticas realizadas por un sujeto en función de un universo de pares de conceptos presentados por el examinador. Mediante la aplicación del método es posible constituir una matriz semántica, describirla y analizarla. El Método DISTSEM cuenta con una sistematización informática (Infosem) para hacer más amigable y versátil su utilización en diferentes ámbitos e intereses de investigación. Su principal desarrollo hasta el momento ha sido en torno a la evaluación educativa.

II-B. Marcos (Frames)

Los frames son una estructura en la cual se pueden representar **valores, restricciones, procesos, atributos, relaciones etc.** Tienen relaciones de pertinencia y herencia, es decir especialización y generalización (por lo que se parecen a la programación orientada a objetos) [3].

Elementos de la representación:

- Las **instancias** se representan por constantes.
- Las **clases** se representan por constantes.
- Cada **propiedad** se representa por una igualdad de la forma $\langle \text{atributo} \rangle : \langle \text{valor} \rangle$

- Las relaciones clase–superclase se representan por hechos de la forma clase ($\langle \text{clase} \rangle$, $\langle \text{sup-clase} \rangle$, [$\langle \text{prop} - 1 \rangle$, ..., $\langle \text{prop} - n \rangle$])
- Las relaciones instancia–clase se representan por hechos de la forma instancia ($\langle \text{clase} \rangle$, $\langle \text{sup-clase} \rangle$, [$\langle \text{prop} - 1 \rangle$, ..., $\langle \text{prop} - n \rangle$])
- La constante inicio representa la clase inicial de la jerarquía.
- Las propiedades de una instancia es una lista de pares **atributo–valor**.

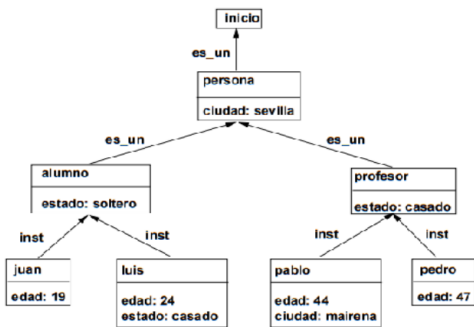


Figura 4. Ejemplo Representación del Conocimiento (Marcos o Frames) - Universidad Nacional de Tucumán.

Una de las ideas intuitivas detrás de los Frames, es que la memoria se basa mucho en estereotipos (propiedades típicas de los objetos), tal como se muestra en la Figura 4.

A continuación, se presentan dos investigaciones donde se evidencia la aplicación de esta técnica:

1. La investigación “*Marcos predicativos asociados al concepto signo y síntoma en textos sobre medicina en español*”, [7].

Problema: En la redacción de textos sobre medicina, una precisa selección de verbos facilita la comprensión de terminología y las relaciones conceptuales de este dominio. En este artículo se pretendió extraer verbos que normalmente coocurren con sintagmas nominales que representan el concepto **signo y síntoma** en textos médicos en español y describir sus patrones. Con una metodología de corpus y la plataforma *SketchEngine* se elaboro un inventario de dichos verbos y se organizo

semánticamente en un **esquema de marcos y submarcos**, a los que se han asociado plantillas con información semántica y sintáctica para las predicaciones verbales que incluyen signo y síntoma como argumentos.

Resultados: Tras el análisis de patrones y líneas de concordancia y la consulta a un experto médico, se propuso un esquema con siete marcos diferentes para la escena. una persona tiene signos y síntomas, probablemente asociados a una enfermedad, como se muestra en la Figura 5. Refleja las predicaciones y categorías conceptuales ‘activadas’ en este contexto y sus posiciones en la estructura argumental. Se han seleccionado verbos frecuentes que estaban presentes en todos los subcorpus, recuperados con las fórmulas de búsqueda descritas en el apartado 2.2. Y se han aplicado al proyecto de investigación CombiMed (Léxico Combinatorio en Medicina: cognición, texto y contexto) con vistas a la creación de un inventario de base semántica en español para la redacción y traducción de textos médicos.

A. SIGNOS/SÍNTOMAS EN EL PACIENTE		
1. Aparición y existencia de signos/síntomas		
SIGNO/SÍNTOMA	<aparece>, <aumenta>, <disminuye>, <empeora>, <dura> <desaparece>	-- TIEMPO --
2. Tener, percibir y sentir signos/síntomas		
PACIENTE	<tiene>, <percibe>, <experimenta>, <tolera>, <permanece> <asintomático>	SIGNO/SÍNTOMA
SIGNO/SÍNTOMA	<molesta>	PACIENTE
3. Signos/síntomas relacionados con la enfermedad		
SIGNO/SÍNTOMA ENFERMEDAD	<se asocia>, <curso>	a ENFERMEDAD con SIGNO/SÍNTOMA
4. La enfermedad causa signos/síntomas		
ENFERMEDAD	<causa>	SIGNO/SÍNTOMA
B. SIGNOS Y SÍNTOMAS EN UN ENTORNO MÉDICO		
5. Signos y síntomas en el diagnóstico de la enfermedad		
PACIENTE PROFESIONAL SANITARIO SIGNO Y SÍNTOMA	<describe>, <experimenta>, <halla>, <indica>, <descarta>	SÍNTOMA SIGNO ENFERMEDAD
6. Prevención y control de signos y síntomas		
PROFESIONAL SANITARIO TRATAMIENTO	<vigila>, <previene>, <controla>	SIGNO Y SÍNTOMA
7. Alivio de síntomas (tratamiento sintomático)		
TRATAMIENTO, HÁBITO SALUDABLE, ALIMENTACIÓN	<alivia>, <detiene>, <suprime>	SÍNTOMA

Figura 5. Marcos y submarcos de las predicaciones verbales que incluyen SIGNO Y SÍNTOMA como argumentos.

Importancia: En estudios previos se ha señalado la importancia de patrones léxico-sintácticos de verbos en el lenguaje especializado. Con lo que se espera que la representación de estos marcos predicativos en español beneficie a científicos, redactores técnicos y traductores

que comunican avances científicos en lengua española, y por ende, al público lego para una mayor comprensión respecto a signo y síntoma.

2. Otra investigación es “*Tierra común, Marcos y Ranuras para la comprensión en Sistemas de Diálogo*”, [1]

Problema: Este artículo presenta un sistema de diálogo para capacitar a los médicos para dar malas noticias. La originalidad de este trabajo radica en su representación del conocimiento. Toda la información conocida antes del diálogo (el universo del discurso, el contexto, el escenario del diálogo) así como el conocimiento transferido del médico al paciente durante la conversación se representa en una estructura de conocimiento compartido denominada terreno común, que constituye el núcleo del sistema.

Resultados: Centrándose en dos aspectos: la representación del conocimiento para comprender y generar comportamientos del paciente. Formalmente, el terreno común está formado por un conjunto de tramas, que reúnen información, denominada ranuras, como se ilustra en la Figura 6. Un valor de ranura puede ser atómico (por ejemplo, Nombre, Edad, etc.) o (por ejemplo, Persona, Patología). Cada espacio es ponderado, en escala: obligatorio, importante, opcional. Y una información puede depender de otro marco. Por ejemplo, el médico no puede describir ninguna remediación antes de haber presentado la patología.

Importancia: El uso de técnicas de diálogo para capacitar a los médicos representa un caso de uso fundamental tanto en la perspectiva de desarrollar nuevas técnicas de diálogo como en términos de aplicación. La representación basada en marcos ofrece primero la posibilidad de utilizar técnicas de clasificación que identifiquen directamente el marco que se va a instanciar. Gracias a ello se propuso un método original de llenado de ranuras, basado en la información de terreno común y semántica distribucional. La generación de las reacciones del

agente y su adaptación al discurso del médico aprovecha directamente la representación.

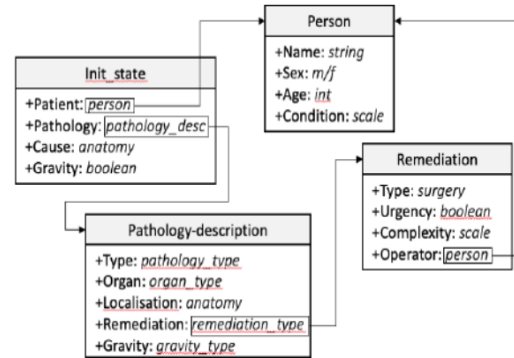


Figura 6. Ejemplo de marcos y ranuras utilizados en la investigación.

II-C. Ontologías

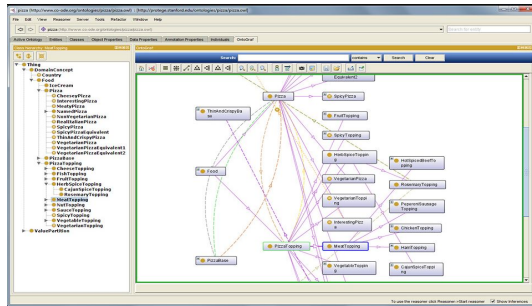
Las ontologías son herramientas útiles para la comunicación entre expertos en un área determinada. Se puede formalizar conceptos en un dominio para que luego sean validadas por computadores y obtener respuestas más favorables a la hora de consultarlas. [5]

Formalmente las ontologías son usadas para representar el conocimiento, la razón por la cual se usan estos esquemas, es debido a que nos permite realizar búsquedas de información con carácter semántico, con la finalidad de obtener resultados mucho más óptimos. [8]

Una búsqueda semántica se puede definir como un proceso de optimización de los resultados obtenidos por un motor de búsqueda, utilizando las propiedades de la semántica. En este contexto, las ontologías nos ofrecen mecanismos de relación conceptual que nos permiten transformar la consulta en una más enriquecida y representada de manera estructurada, para luego compararla con la ontología que maneja el motor utilizado y así devolver mejores resultados [9]

Para la representación de ontologías se utilizan diferentes lenguajes que han ido evolucionando en el tiempo, la finalidad de estos lenguajes es de representar a las mismas de manera que sean entendibles por el computador y por el usuario, por ejemplo: RDF/XML y OWL y entornos como Protege. (Ver

Figura 7.)

Figura 7. Ejemplo de ontología en Protegé <http://protege.stanford.edu/>

A continuación, dos investigaciones respecto a la técnica de representación del conocimiento de ontologías :

1. En la tesis “*Diseño de un modelo para la Recuperación de Documentos basado en Ontologías en el Dominio de la Ingeniería Informática*” [4], se propuso el uso de ontologías como estructura básica para la representación del conocimiento por su eficacia en realizar la recuperación de documentos. Además, se planteó utilizar un proceso de etiquetación semántica de documentos para relacionar cada documento digital con - al menos - una entidad de la ontología con la finalidad de poder realizar búsquedas mediante el uso de etiquetas y lenguaje natural. (Ver Figura 8)

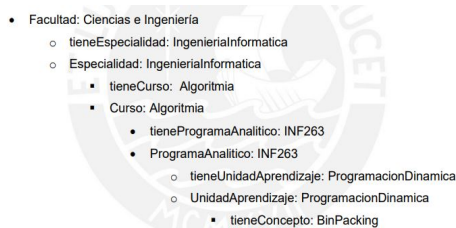


Figura 8. Un ejemplo de hechos sacado de la ontología.

Problema: En la especialidad de Ingeniería Informática de la PUCP existe la necesidad en los estudiantes de buscar documentos que contengan información sobre los cursos de la carrera.(ámbito o dominio del que se habla es bastante amplio). Actualmente, se cuenta con repositorios digitales de documentos, que almacenan imágenes

escaneadas o documentos en formato PDF, a la que los alumnos y profesores tienen acceso. Sin embargo, actualmente estos documentos no proporcionan alguna manera de reconocer su contenido y, así, facilitar su recuperación.

Importancia: Ante ello se presentó el análisis del dominio de Ingeniería Informática, para proponer una estructura ontológica capaz de soportar y almacenar todas las relaciones e inferencias producidas en el dominio. (Ver Figura 9)

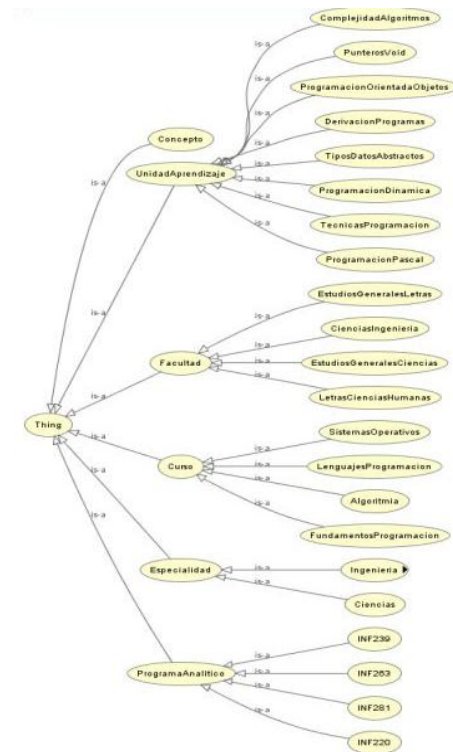


Figura 9. Ontología planteada en el dominio de la Ingeniería Informática.

Resultados: En el análisis del dominio, se establecieron una serie de restricciones a medida que se fue desarrollando la ontología, esto con la finalidad de reducir incongruencias. Luego de probar el aplicativo, se observó que el modelo planteado cubre en gran medida las necesidades planteadas. Aun no se puede cumplir con realizar consultas en lenguaje natural. Debido a que el procedimiento de análisis se propuso, a manera que, limitemos al usuario a utilizar consultas que directamente

sean entidades de la ontología en la estructura.

2. En el artículo de investigación titulado “*Seguridad informática de un sistema basado en ontología de control vía Web de dispositivos domóticos*” [6].



Figura 10. Estructura de sistema domótico.

Los sistemas domóticos (Ver Figura 10) son un conjunto de sistemas de automatización de una vivienda con fines de gestión energética, seguridad, bienestar y comunicación; integrados por controladores, actuadores y sensores, pudiendo ser gestionados remotamente desde un servidor web, tal y como lo afirman García Sánchez y Moreno Martín (2013, p. 8), y mediante una ontología.

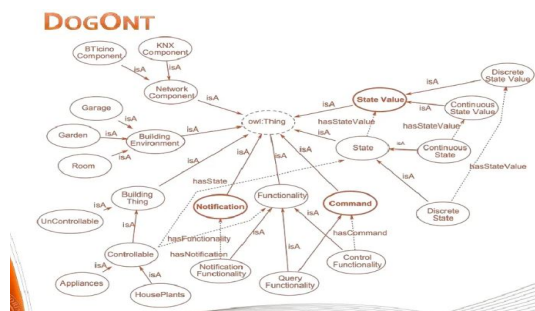


Figura 11. Ejemplo de sistema domótico.

Importancia: Las ontologías son componentes de la Web semántica que, implementan conceptos relacionados, razonamiento y reglas de inferencia en la Web. Un ejemplo es DogOnt (Ver Figura 11) desarrollada como una ontología que representa conceptos de ambientes y dispositivos referidos a una vivienda domótica, de

manera remota desde un servidor web y, en una configuración adecuada, reducir el consumo de energía de múltiples viviendas, teniendo en cuenta la reutilización de componentes ante usuarios concurrentes en un mismo ambiente con actividades parecidas o prioritarias.

Problema: Sin embargo, la complejidad inherente al desarrollo de aplicaciones web de hoy, como las basadas en ontologías que controlarían remotamente a los sistemas domóticos en viviendas automatizadas, trae consigo una falla en la seguridad informática, vulnerabilidad que es propia de la Web semántica ya que los ataques informáticos, en un futuro próximo, se centrarán en modificar el significado de conceptos incluidos en una ontología de la Web semántica para permitir la intrusión en los sistemas, por ejemplo, de automatización de vivienda.

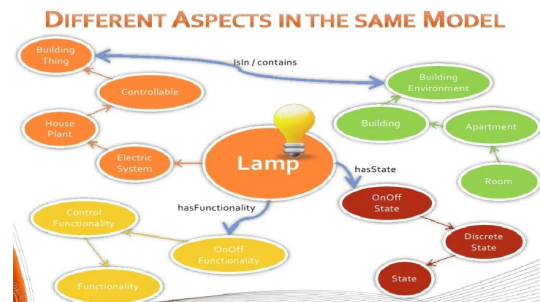


Figura 12. Ontología en DogOnt.

Resultados: Por eso se concluyó que, para mitigar el riesgo informático de una ontología controladora de sistemas domóticos de manera remota expuesta desde un servidor web, es necesario un criptosistema simétrico basado en el lenguaje XML Encryption con el Advanced Encryption Standard con un marco de desarrollo y ECC Mixed Coordinates Cryptography para encriptar el mensaje apoyándose en OWL / OWL-S, debido a su capacidad de asegurar, mediante clases privadas, las reglas de inferencia y consultas en lenguaje plano que intercambian el servidor web con el controlador central del sistema domótico.

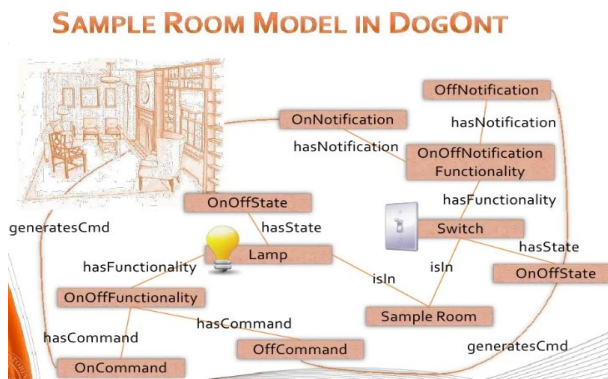


Figura 13. Ejemplo de sistema domótico en una habitación usando herramienta de desarrollo de Ontología DogOnt.

III. CONCLUSIONES

Este informe presentó información relevante acerca de tres técnicas de Representación del Conocimiento empleados en la Inteligencia Artificial, explicándose también los conceptos teóricos relacionados a ellas. Además de permitir conocer sobre algunas investigaciones que se vienen realizando año a año, entendiendo las diversas aplicaciones en la vida diaria. Concluimos que la mayoría analiza la problemática y dependiendo de sus características se establecieron los procesos así como la técnica a emplear, para representar información o conocimiento del mundo real en una forma que se pueda entender y usar para resolver problemas de la vida real. Finalmente, las investigaciones expuestas en este trabajo de recopilación, sirven como antecedentes y base a futuros investigadores a nivel nacional e internacional, que pueden ser usados tanto en el área comercial o académico.

REFERENCIAS

[1] P. Blache, M. Houlès, et al. Common ground, frames and slots for comprehension in dialogue systems. In *CS & IT Conference Proceedings*, volume 11. CS & IT Conference Proceedings, 2021.

[2] D. de Ciencias de la Computación e Inteligencia Artificial Universidad de Sevilla. Redes semánticas y marcos. URL <https://www.cs.us.es/~jalonso/cursos/ia2>

-02/temas/tema-4.pdf. (Accessed on 10/29/2022).

[3] G. J. F. de Ciencias Exactas y Tecnología Universidad Nacional de Tucumán. Representación del conocimiento, 2018. URL <https://acortar.link/e283MK>. (Accessed on 29/10/2022).

[4] H. Gómez. *Diseño de un modelo para la recuperación de documentos basado en ontologías en el dominio de la ingeniería informática*. PhD thesis, Tesis pregrado. Pontificia Universidad Católica del Perú, Perú., 2014.

[5] E. Greengrass. Information retrieval: A survey. URL <https://redirect.cs.umbc.edu/csee/research/cadip/readings/IR.report.120600.book.pdf>. 30 November 2000.

[6] L. V. U. Huiman. Seguridad informática de un sistema basado en ontología de control vía web de dispositivos domóticos. *PUEBLO CONTINENTE*, 26(2):501–503, 2016.

[7] C. I. López Rodríguez. Marcos predicativos asociados al concepto signo y síntoma en textos sobre medicina en español. *Revista signos*, 53(103):392–418, 2020.

[8] H. Melgar Sasieta. Un modelo para la visualización de conocimiento basado en imágenes semánticas, 2011.

[9] B. Rujiang and L. Junhua. Improving documents classification with semantic features. In *2009 Second International Symposium on Electronic Commerce and Security*, volume 1, pages 640–643. IEEE, 2009.

[10] L. Y. Vivas. Aplicación de un método para el análisis de las redes semánticas en pacientes que sufrieron un accidente cerebro vascular. *Interdisciplinaria*, 27(1):147–162, 2010.

[11] B. Zarei, H. Sharifi, and Y. Chaghoei. Delay causes analysis in complex construction projects: a semantic network analysis approach. *Production Planning & Control*, 29(1):29–40, 2018.