

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

Facultad de Ciencias Químicas e Ingeniería

Maestría y Doctorado en Ciencias e Ingeniería



“CONSTRUCCIÓN DE REDES BAYESIANAS DE FORMA
AUTOMÁTICA PARA SISTEMAS TUTORES INTELIGENTES”

TESIS

PARA OBTENER EL GRADO DE

DOCTOR EN CIENCIAS

Presenta:

ALAN DAVID RAMÍREZ NORIEGA

Bajo la dirección de:

DR. J. REYES JUÁREZ RAMÍREZ

Co-dirigido por:

DR. JUAN JOSÉ TAPIA ARMENTA

TIJUANA, BAJA CALIFORNIA, MÉXICO

NOVIEMBRE DEL 2017

Dedicatoria

Para:

mis padres Silvia y David,

mi esposa Patricia,

mis hijos Leonardo y Jonathan.

Agradecimientos

A mi esposa Janeth Patricia y a mis hijos Leonardo y Jonathan por ser la motivación para que este proyecto de investigación se haya desarrollado de forma exitosa.

A mis Padres por su apoyo durante esta travesía y estar siempre presente apoyándome y motivándome en mis estudios con su ejemplo.

Al Dr. J. Reyes Juárez por su dirección de tesis y por haber confiado en mí para iniciar y culminar este proyecto.

A mi comité de tesis por el apoyo, la enseñanza y la atención brindada principalmente en la fase final de la investigación y del documento de tesis.

A mis compañeros de laboratorio con los que compartí muchos buenos momentos y aprendí muchas cosas de ellos.

A la Universidad Autónoma de Sinaloa, la Universidad Autónoma de Baja California, el Consejo Nacional de Ciencia y Tecnología, el Sindicato Nacional de Trabajadores de la Educación, la Secretaría de Educación Pública y el Colegio de Bachilleres del Estado de Sinaloa, que ha a través de sus recursos y prestaciones me permitieron estar lejos de mi hogar y mi trabajo para terminar esta tesis.

Por último a todos mis maestros e instituciones educativas que me han permitido llegar hasta aquí, siempre enseñando lo mejor de ellos y brindándome sus mejores consejos.

Gracias.

Resumen

Construcción de redes Bayesianas de forma automática para Sistemas Tutores Inteligentes

La Red Bayesiana (RB) es una técnica importante para inferir y representar el conocimiento en un Sistema Tutor Inteligente (STI); sin embargo, estos sistemas son complejos de construir. Autores han construido RBs basadas en ontologías para acelerar el proceso; sin embargo, no automatizan completamente el proceso, no siguen el estándar de ontologías OWL (por sus siglas en inglés Lenguaje de Ontología Web) o simplifican el dominio final. Este trabajo propone un método para construir RBs basadas en ontologías e información de Wikipedia para ser empleado en STIs. El método propuesto construye la parte cualitativa de la RB a través de clases y relaciones de ontologías; la parte cuantitativa se crea a partir de frecuencias, saltos y una medida de similitud entre los conceptos de la ontología representada por los artículos de Wikipedia. Este estudio consideró dos experimentos para determinar la correlación de resultados de nuestro método con expertos en el dominio; el método obtuvo una correlación positiva de 0.854 de acuerdo con la prueba de Pearson. Nuestro método construye automáticamente una RB para representar el conocimiento en STIs de manera similar a como lo harían los expertos.

Palabras claves: Red Bayesiana, Ontología, Sistema Tutor Inteligente, Tabla de Probabilidad Condicional, Wikipedia

Abstract

Construction of Bayesian networks automatically for Intelligent Tutoring Systems

The Bayesian Network (BN) is an important technique to infer and represent knowledge in an Intelligent Tutoring System (ITS); however, these systems are complex to build. Authors have built BNs based on ontologies to accelerate the process; nonetheless, they do not fully automate the process, do not follow the ontologies standard OWL (Web Ontology Language), or simplify the final domain. This work proposes a method to build BNs based on ontologies and Wikipedia information to be employed on ITSs. The proposed method constructs the qualitative part of the BN through classes and relations of ontologies; the quantitative part is created based on frequencies, hops, and a measure of similarity between concepts of the ontology represented by Wikipedia articles. This study considered two experiments to determine the results correlation of proposed method against domain experts; the method obtained a positive correlation of 0.854 according to the Pearson test. The method constructs a BN to represent the knowledge in ITSs, in a similar way as experts would, supporting the automatically build of these systems.

Keywords: Bayesian Network, Ontology, Intelligent Tutoring System, Conditional Probability Table, Wikipedia

Índice general

| | |
|--|------------|
| Índice de figuras | XII |
| Índice de tablas | XV |
| Lista de acronimos | XVI |
| 1 Introducción | 1 |
| 1.1 Antecedentes | 1 |
| 1.2 Planteamiento del problema | 4 |
| 1.3 Objetivos | 7 |
| 1.4 Preguntas de investigación | 8 |
| 1.5 Contribuciones esperadas | 9 |
| 1.6 Organización del documento de tesis | 9 |
| 2 Representación del conocimiento en tutores inteligentes | 11 |
| 2.1 Introducción | 12 |
| 2.2 Representación del conocimiento | 13 |
| 2.3 Representación del conocimiento en la enseñanza | 14 |
| 2.4 Sistemas tutores inteligentes | 16 |
| 2.5 Técnicas de representación del conocimiento | 18 |
| 2.5.1 Mapa de conceptos | 19 |
| 2.5.2 Red Bayesiana | 19 |

| | | |
|---------|---|----|
| 2.5.3 | Mapas cognitivos | 20 |
| 2.5.3.1 | Mapa cognitivo como red Bayesiana | 20 |
| 2.5.3.2 | Mapa cognitivo difuso | 20 |
| 2.5.4 | Grafos conceptuales | 21 |
| 2.5.5 | Mapas de conocimiento | 21 |
| 2.5.6 | Redes semánticas | 21 |
| 2.5.7 | Mapa de memoria | 22 |
| 2.6 | Características de la representación del conocimiento | 22 |
| 2.6.1 | Conceptos | 23 |
| 2.6.2 | Habilidades | 24 |
| 2.6.3 | Asociaciones | 25 |
| 2.6.4 | Inferencia | 27 |
| 2.6.5 | Tipo de conocimiento | 28 |
| 2.6.6 | Estructura jerárquica | 28 |
| 2.6.6.1 | Sucesores y predecesores | 29 |
| 2.6.6.2 | Categorías | 29 |
| 2.6.6.3 | Herencia | 30 |
| 2.6.6.4 | Representación gráfica | 30 |
| 2.7 | Comparativa entre las técnicas de representación del conocimiento | 30 |
| 2.7.1 | Conceptos y atributos | 31 |
| 2.7.2 | Asociaciones | 32 |
| 2.7.3 | Inferencia y tipo de conocimiento | 33 |
| 2.7.4 | Estructura jerárquica | 33 |
| 2.8 | Análisis de las técnicas | 34 |
| 2.9 | Redes Bayesianas frente a otras técnicas | 36 |
| 2.10 | Resumen del capítulo | 39 |

| | | |
|----------|---|-----------|
| 3 | Redes Bayesianas basadas en ontologías | 40 |
| 3.1 | Redes Bayesianas | 40 |
| 3.1.1 | Evaluación del conocimiento | 42 |
| 3.1.2 | Construcción manual de una red Bayesiana | 45 |
| 3.1.2.1 | Metodología | 46 |
| 3.1.2.2 | Implementación de la metodología | 46 |
| 3.1.2.3 | Red Bayesiana para el Proceso Software Personal | 51 |
| 3.1.3 | Formalización | 54 |
| 3.1.4 | Descomposición de una red Bayesiana | 55 |
| 3.2 | Ontologías | 56 |
| 3.2.1 | Componentes principales | 58 |
| 3.2.2 | Lenguaje de Ontología Web | 59 |
| 3.2.3 | Metodologías de desarrollo | 60 |
| 3.2.4 | Formalización | 60 |
| 3.2.5 | Descomposición de una ontología | 61 |
| 3.3 | Trabajos relacionados | 63 |
| 3.4 | Resumen del capítulo | 67 |
| 4 | Relación entre conceptos | 68 |
| 4.1 | Minería web | 69 |
| 4.1.1 | Minería de texto | 69 |
| 4.1.2 | Indexación de documentos de texto | 70 |
| 4.1.3 | El proceso de indexación | 70 |
| 4.2 | Relación semántica | 71 |
| 4.2.1 | Métodos basados en corpus | 73 |
| 4.2.2 | Métodos basados en conocimiento | 74 |
| 4.2.3 | Métodos basados en términos | 75 |

| | | |
|----------|---|-----------|
| 4.2.4 | Métodos basados en traslape de palabras | 76 |
| 4.3 | Wikipedia | 77 |
| 4.3.1 | Estructura | 78 |
| 4.3.2 | Extracción de información | 81 |
| 4.3.3 | Recursos y herramientas | 82 |
| 4.3.3.1 | Java Wikipedia Library | 83 |
| 4.4 | Resumen del capítulo | 84 |
| 5 | Generación de la red Bayesiana | 85 |
| 5.1 | Descripción general del método | 85 |
| 5.2 | Características del método | 88 |
| 5.2.1 | Parte cualitativa | 88 |
| 5.2.2 | Parte cuantitativa | 89 |
| 5.3 | Construcción de la red Bayesiana | 90 |
| 5.3.1 | Parte cualitativa | 90 |
| 5.3.2 | Parte cuantitativa | 90 |
| 5.3.2.1 | Relación cuantitativa entre nodos | 90 |
| 5.3.2.2 | Algoritmo para construir las tablas de probabilidad condicional | 93 |
| 5.4 | Método propuesto: Ontologías y Wikipedia | 94 |
| 5.4.1 | Parte cualitativa de la red Bayesiana utilizando ontologías | 94 |
| 5.4.1.1 | Obtención de variables e identificación de estados | 94 |
| 5.4.1.2 | Relaciones entre variables | 95 |
| 5.4.2 | Parte cuantitativa de la red Bayesiana con Wikipedia | 95 |
| 5.4.2.1 | Identificación de las variables del modelo usando Wikipedia | 96 |
| 5.5 | Resumen del capítulo | 96 |
| 6 | Experimentación y resultados | 97 |

| | | |
|---------|--|-----|
| 6.1 | Caso de estudio A: Programación orientada a objetos | 98 |
| 6.1.1 | Desarrollo de la ontología de programación orientada a objetos | 98 |
| 6.1.2 | Generación de pesos por expertos | 101 |
| 6.1.2.1 | Muestra | 101 |
| 6.1.2.2 | Encuesta | 102 |
| 6.1.2.3 | Procedimiento | 103 |
| 6.1.3 | Generación de pesos del método | 103 |
| 6.1.3.1 | Aspectos de implementación | 103 |
| 6.1.4 | Prueba estadística | 104 |
| 6.1.5 | Resultados | 105 |
| 6.1.5.1 | Expertos vs método | 105 |
| 6.1.5.2 | Construcción de una tabla de probabilidad condicional | 105 |
| 6.1.5.3 | Prueba estadística | 108 |
| 6.2 | Caso de estudio B: Tipos de lenguajes de programación | 108 |
| 6.2.1 | Parte cualitativa de la red Bayesiana por medio de ontologías | 109 |
| 6.2.1.1 | Obtención de variables e identificación de estados | 110 |
| 6.2.1.2 | Relaciones entre variables | 110 |
| 6.2.2 | Parte cuantitativa de la red Bayesiana por medio de Wikipedia | 110 |
| 6.2.2.1 | Identificación de las variables del modelo | 111 |
| 6.2.3 | Generación de pesos por expertos | 111 |
| 6.2.3.1 | Muestra | 111 |
| 6.2.3.2 | Encuesta | 112 |
| 6.2.3.3 | Procedimiento | 113 |
| 6.2.4 | Generación de pesos por el método | 114 |
| 6.2.4.1 | Procedimiento | 114 |
| 6.2.4.2 | Aspectos de implementación | 114 |

| | | |
|----------|---|------------|
| 6.2.5 | Prueba de hipótesis | 114 |
| 6.2.6 | Resultados | 115 |
| 6.2.6.1 | Expertos vs método | 115 |
| 6.2.6.2 | Red Bayesiana generada | 115 |
| 6.2.6.3 | Prueba estadística | 116 |
| 6.3 | Resumen del capítulo | 117 |
| 7 | Discusión de los resultados obtenidos | 118 |
| 7.1 | Caso de estudio A: Programación orientada a objetos | 118 |
| 7.2 | Caso de estudio B: Tipos de lenguajes de programación | 119 |
| 7.3 | Implicaciones en general | 119 |
| 7.3.1 | Implicaciones prácticas | 120 |
| 7.3.2 | Implicaciones teóricas | 120 |
| 8 | Conclusiones, aportaciones y trabajo futuro | 122 |
| 8.1 | Conclusiones | 122 |
| 8.2 | Aportaciones | 124 |
| 8.3 | Trabajo futuro | 125 |
| | Bibliografía | 125 |
| | Apéndices | 145 |
| A | Encuestas | 146 |
| A.1 | Encuesta sobre Orientación a Objetos | 146 |
| A.2 | Encuesta sobre Tipos de Lenguajes de Programación | 153 |
| B | Calculo de los pesos | 159 |
| B.1 | Pesos obtenidos en caso de estudio B | 159 |

| | |
|-----------------------|----|
| <i>ÍNDICE GENERAL</i> | XI |
|-----------------------|----|

| | |
|------------------------|------------|
| C Publicaciones | 161 |
|------------------------|------------|

| | |
|--|-----|
| C.1 Primer autor | 161 |
| C.1.1 Revistas con factor de impacto (JCR) | 161 |
| C.1.2 Conferencias y libros | 162 |
| C.2 Colaboraciones | 163 |
| C.2.1 Revistas con factor de impacto (JCR) | 163 |
| C.2.2 Conferencias y libros | 163 |

Índice de figuras

| | | |
|------|---|----|
| 2.1 | Bloque 1 de algoritmos computacionales. | 15 |
| 2.2 | Estructura jerárquica por medio de grafos. | 16 |
| 2.3 | Estructura jerárquica por medio de grafos con conceptos. | 17 |
| 2.4 | Elementos considerados de representación del conocimiento por STIs. | 23 |
| 3.1 | Modelo de Millán y Pérez-De-La-Cruz (2002). | 42 |
| 3.2 | Pregunta y su relación con los conceptos. | 43 |
| 3.3 | Conceptos, preguntas y sus probabilidades a priori. | 44 |
| 3.4 | Probabilidad a posteriori, pregunta 2 correcta. | 44 |
| 3.5 | Probabilidad a posteriori, pregunta 2 incorrecta. | 45 |
| 3.6 | Probabilidad a posteriori, pregunta 1 y 2 correctas. | 45 |
| 3.7 | Conceptos desglosados de PSP (Humphrey, 1997). | 47 |
| 3.8 | Conceptos de PSP sin agrupar. | 48 |
| 3.9 | Conceptos de PSP agrupados. | 49 |
| 3.10 | Tablas de probabilidad condicional para las variables. | 50 |
| 3.11 | Red Bayesiana con preguntas. | 51 |
| 3.12 | Tabla de probabilidad condicional para la pregunta P2. | 51 |
| 3.13 | Simulación de la red Bayesiana sin evidencia. | 52 |
| 3.14 | Simulación de la red Bayesiana con evidencia. | 53 |
| 3.15 | Ejemplo de red Bayesiana. | 55 |
| 3.16 | Ontología en Protégé. | 58 |

| | | |
|------|--|-----|
| 3.17 | Ejemplo de ontología sencilla. | 62 |
| 4.1 | Interfaz de Wikipedia. | 77 |
| 5.1 | Proceso de construcción de la red Bayesiana. | 86 |
| 5.2 | Algoritmos para construir la red Bayesiana. | 87 |
| 5.3 | Relaciones entre variables en una estructura de representación del conocimiento. | 91 |
| 6.1 | Ontología de la orientación a objetos | 100 |
| 6.2 | Sección empleada de la ontología de la orientación a objetos. | 100 |
| 6.3 | ¿Cuánto tiempo llevas programando en OO? | 101 |
| 6.4 | ¿Qué nivel de conocimiento consideras tener en OO? | 102 |
| 6.5 | ¿Cuánto tiempo tienes impartiendo POO? | 102 |
| 6.6 | Comparación entre los expertos y el método propuesto. | 107 |
| 6.7 | Ontología de los Tipos de Lenguajes de Programación. | 109 |
| 6.8 | ¿A qué rango de edad perteneces? | 112 |
| 6.9 | ¿Cuántos años tienes trabajando como docente en nivel universitario? | 112 |
| 6.10 | Comparación entre el método propuesto y los expertos. | 116 |
| 6.11 | Red Bayesiana generada con los resultados del método propuesto. | 116 |
| A.1 | Ontología representada jerárquicamente. | 147 |
| A.2 | Representación de árbol de la ontología. | 148 |
| A.3 | Conceptos de orientación a objetos. | 149 |
| A.4 | Conceptos de abstracción. | 149 |
| A.5 | Conceptos de clase. | 150 |
| A.6 | Conceptos de modificadores de clase. | 150 |
| A.7 | Conceptos de método. | 151 |
| A.8 | Conceptos de objeto. | 151 |

A.9 Conceptos de encapsulación. 151

A.10 Conceptos de jerarquía. 152

A.11 Conceptos de herencia. 152

A.12 Conceptos de herencia. 152

Índice de tablas

| | | |
|-----|--|-----|
| 2.1 | Comparativa concepto y atributos. | 31 |
| 2.2 | Comparativa asociaciones. | 32 |
| 2.3 | Comparativa inferencia y tipo de conocimiento. | 33 |
| 2.4 | Comparativa jerarquía. | 34 |
| 2.5 | Principal ventaja y desventaja de las técnicas. | 36 |
| 2.6 | Comparativa entre técnicas. | 37 |
| 2.7 | Comparativa entre RBs y MCD. | 38 |
| 3.1 | Artículos relacionados. | 63 |
| 3.2 | Comparativa entre estudios relacionados a ontologías y redes Bayesianas. | 64 |
| 4.1 | Estadísticas de Wikipedia. | 78 |
| 6.1 | Resultados de los expertos y el método propuesto (OO). | 106 |
| 6.2 | Tabla de probabilidad condicional para la variable <i>encapsulación</i> | 107 |
| 6.3 | Resultados de los expertos y el método propuesto (Tipos de lenguajes de programación). | 115 |
| B.1 | Valores obtenidos por factor | 160 |

Lista de acrónimos

| No. | Acrónimo | Significado |
|-----|----------|---|
| 1 | API | Interfaz de Programación de Aplicaciones |
| 2 | BCN | Mapa Cognitivo Bayesiana |
| 3 | HAL | Hiperespacio Análogo al Lenguaje |
| 4 | IA | Inteligencia Artificial |
| 5 | IPFP | Procedimiento de Ajuste Proporcional Iterativo |
| 6 | JWPL | Librería de Wikipedia de Java |
| 7 | LSA | Análisis Semántico Latente |
| 8 | MCD | Mapas Cognitivos Difusos |
| 9 | OWL | Lenguaje de Ontología Web |
| 10 | OO | Orientación a Objetos |
| 11 | PLN | Procesamiento de Lenguaje Natural |
| 12 | POO | Programación Orientada a Objetos |
| 13 | PSP | Proceso Software Personal |
| 14 | RB | Red Bayesiana |
| 15 | RDF | Marco de trabajo de Descripción de Recursos |
| 16 | RDFS | Esquema del Marco de trabajo de Descripción de Recursos |
| 17 | TF-IDF | Frecuencia de Termino-Frecuencia Inversa de Documento |
| 18 | TPC | Tablas de Probabilidad Condicional |
| 19 | W3C | Consortio Mundial de la Web |

Capítulo 1

Introducción

En este primer capítulo se aborda de forma breve los antecedentes del tema de tesis, así como la problemática que resuelve este estudio. Además, se trazan los objetivos, hipótesis y preguntas de investigación que guían el desarrollo de esta investigación. Después, se describe de forma general las contribuciones que esperamos al finalizar este proyecto. Finalmente, la organización del documento es descrita.

1.1. Antecedentes

El conocimiento es una combinación de experiencias, saberes, valores, información contextual, percepción e ideas que crean una estructura mental en el sujeto para evaluar e incorporar nuevas ideas, saberes y experiencias (Jennex *et al.*, 2014). La educación es el proceso hacia una persona para estimularla a desarrollar este conocimiento a través de sus capacidades cognitivas y físicas; también es llamado proceso enseñanza-aprendizaje (Rivas Navarro, 2008). Los profesores guían al estudiante durante este proceso, los educadores deben percibir las necesidades del estudiante para mejorar su forma de aprender. Sin embargo, en un ambiente de clases común, surgen problemas generados por la superioridad numérica de alumnos; algunos de estos problemas se refieren a la poca atención personalizada para el estudiante o un mismo tipo de enseñanza para todos los alumnos (Shishehchi y

Banihashem, 2011). Para mitigar estos problemas, algunos autores (Carbonell, 1970; Cataldi y Lage, 2010b; Santhi *et al.*, 2013) proponen el uso de sistemas de aprendizaje basados en computadoras para cubrir las necesidades del estudiante.

Un Sistema Tutor Inteligente (STI) es un sistema de enseñanza-aprendizaje por computadora que utiliza técnicas de Inteligencia Artificial (IA) para interactuar con los estudiantes y enseñarles (Millán y Pérez-De-La-Cruz, 2002), tal como lo haría un profesor con sus estudiantes. Existe una arquitectura generalizada para los STIs que considera cuatro módulos básicos (Carbonell, 1970):

- El módulo del tutor genera las interacciones de aprendizaje basadas en las dificultades de aprendizaje del estudiante.
- El módulo del dominio define el área de conocimiento que el STI enseñará.
- El módulo del estudiante define el conocimiento del estudiante en la sesión de trabajo.
- El módulo de la interfaz permite la interacción del estudiante con el STI.

Un problema importante en el desarrollo de STIs es la evaluación del aprendizaje del estudiante (Conejo *et al.*, 2001). Los STIs deben ser capaces de determinar con precisión y rapidez el nivel de aprendizaje del estudiante para decidir qué temas enseñarle. Para esto, se requiere una representación eficiente del conocimiento, así como un mecanismo de inferencia para determinar el grado de esa posesión de conocimiento.

Existen implementaciones matemáticas que actualmente se utilizan para trabajar la medición del aprendizaje, tales como: relación de contingencia, modelo de Rescorla-Wagner, redes Bayesianas y redes neuronales (Rodríguez Mañanes, 2009). Las más utilizadas son las dos últimas, sin embargo, las redes neuronales tienen las desventajas de no tener un razonamiento simbólico ni representación semántica. Las redes neuronales generalmente toman la forma de un modelo de caja negra debido a que sus relaciones no lineales de causa y efecto no son fácilmente interpretables, lo que hace difícil explicar sus resultados (Correa *et al.*, 2008).

La Red Bayesiana (RB) usa la teoría de la probabilidad como un marco de trabajo para administrar la incertidumbre, esta teoría se apoya de una descripción gráfica para representar la distribución probabilística entre variables permitiendo eficientemente combinar la propagación de probabilidades con un formalismo riguroso (Santhi *et al.*, 2013).

Esta teoría probabilística es considerada adecuada para representar el conocimiento por permitir un razonamiento sobre un ambiente de incertidumbre como el que muestra un tutor al querer determinar el conocimiento del alumno (Lima y Stump, 2010). Es decir, el profesor asigna una calificación al estudiante que representan su conocimiento, sin embargo, el profesor no sabe con precisión si esa calificación es mas adecuada de acuerdo a los temas o conceptos que aprendió. Cuando se asigna una calificación aprobatoria, tenemos la idea que el alumno entendió casi todos los temas, pero no sabemos con precisión cuáles si aprendió, cuáles no y en que grado los aprendió; el profesor solo estima con base en las evidencias de aprendizaje del alumno porque existe incertidumbre en este aspecto.

Una RB representa el conocimiento a través de nodos y relaciones de la misma forma que autores afirman que se organiza el conocimiento en la mente de las personas. Según Rivas Navarro (2008) los conocimientos están almacenados en redes organizadas por conceptos y palabras (nodos), conectados a otros por enlaces o trayectorias asociativas/semánticas, este esquema es el mismo empleado por las RBs. Además, Tabora (2010) afirma que las personas tienen la capacidad de predecir la siguiente acción, a partir de información probabilística; es decir, pueden realizar inferencias psicológicas en situaciones de incertidumbre sobre experiencias pasadas; de la misma forma en que funciona el razonamiento probabilísticos de las RBs. Estos argumentos dan fuerza para representar el conocimiento de los estudiantes con el objetivo de modelar el conocimiento y hacer inferencias sobre el mismo.

En años recientes, el desarrollo de STIs se ha enfocado en tres principales teorías para modelar al estudiante (Chrysafiadi y Virvou, 2013): RBs, lógica difusa y ontologías. Algunos autores de (Satar, 2012; Ting *et al.*, 2013; Victorio-Meza *et al.*, 2014) han concluido que las RBs son el enfoque más comúnmente usado en los STIs por las ventajas ofrecidas de su teoría. Además, Danaparamita y Gaol (2014) evaluaron dos modelos para STIs, uno basado en lógica difusa y otro en RBs, concluyendo que

las predicciones de las necesidades de aprendizaje de los estudiantes son mejores usando RBs.

Los STIs han sido apoyados por el uso de RBs con aportes tales como:

- Modelar el dominio de conocimiento a través de la estructura formada por sus nodos y relaciones, sirviendo como base para determinar que debe aprender el alumno (módulo del dominio).
- Representar el conocimiento del estudiante de los temas representados en el módulo del dominio a través de inferencias probabilísticas (módulo del estudiante).
- Tomar decisiones sobre que tema hay que enseñarle con base en la información del módulo del estudiante (módulo del tutor).

1.2. Planteamiento del problema

Una RB está formada por tres elementos principales: los nodos que son variables que representan el problema, las relaciones que son enlaces de influencia entre variables y las Tablas de Probabilidad Condicional (TPCs) que definen la probabilidad que un evento pueda ocurrir considerando la combinación de nodos y el valor de sus estados. Estos tres aspectos pueden ser divididos en una parte cualitativa, nodos y relaciones; y una parte cuantitativa, las TPCs.

Existen dos procedimientos genéricos para crear redes Bayesianas (López Puga, 2012): (1) Centrado en el juicio de expertos, donde se utiliza el conocimiento de expertos sobre un dominio particular para generar el modelo estadístico. (2) Basado en rutinas automáticas, donde se ponen en funcionamiento algoritmos que son capaces de construir una RB de un conjunto de datos.

La construcción de la red centrada en el juicio del experto es una tarea compleja (Cruz Ramírez y Martínez Morales, 2007; Yan y Wei, 2013; Ting *et al.*, 2013) por la interacción con el experto y la necesidad de plasmar su conocimiento en una estructura con nodos, relaciones y valores numéricos. Mientras que, la construcción de una RB de forma automática es generada en poco tiempo, la construcción manual implica más tiempo (Cruz Ramírez y Martínez Morales, 2007; Fenz, 2009; Yan y Wei,

2013) y mayor labor (Andrade, 2003; Thirumuruganathan y Huber, 2011).

La construcción de RBs automatizada por medio de minería de datos se basa en el análisis de gran cantidad de información para identificar variables, relaciones y TPCs, sin embargo, tiene algunas desventajas (Andrade, 2003; Devitt *et al.*, 2006; Xiao-xuan *et al.*, 2007):

- Incapacidad de hacer frente a los datos que faltan.
- Hacen supuestos sobre los datos de entrada o salida de la estructura.
- No integran fuentes de conocimiento estructurada del mundo actual.
- Si hay una gran cantidad de nodos en las redes, el espacio de búsqueda del algoritmo de aprendizaje es tan grande que a menudo conduce a una baja en la eficiencia del aprendizaje.

La construcción de RBs basados en minería de datos es adecuada para áreas donde no es necesario precisión en la parte cualitativa de la red (Gómes Flechoso, 1998), es decir los nodos y sus relaciones no son determinantes para el problema. Sin embargo, para el enfoque de este estudio, la minería de datos podría construir una estructura muy generalizada que no refleje las necesidades de representación del conocimiento en un ambiente educativo. Es decir, una estructura de temas adecuadamente jerarquizados para ser enseñados.

Otra forma de automatizar la construcción de RBs es a través de las ontologías; ambas representaciones comparten similitudes (Rettinger *et al.*, 2012; Rodrigues *et al.*, 2013; Yan y Wei, 2013; Chang *et al.*, 2013). En los últimos años, ha existido un aumento en el uso de las ontologías en sistemas inteligentes debido a que contienen una representación de conocimientos sobre un dominio y modelan la información semánticamente a través de nodos y enlaces (Yang y Calmet, 2005).

La construcción de RBs con base en ontologías representa uno de los enfoques más precisos para construir la parte cualitativa de la RB porque ambas estructuras están compuestas de nodos y relaciones. Sin embargo, el trabajo más complicado de la RB es construir la parte cuantitativa debido a que requiere información estadística que represente las relaciones entre las variables (Xiao-xuan *et al.*, 2007; Rodrigues *et al.*, 2013; Mkrtychyan *et al.*, 2016).

Existen trabajos para generar RBs a través de ontologías, estos esfuerzos han sido aplicados en el campo médico así como en otras áreas. Estos intentos pueden ser clasificados en tres tipos:

- Métodos manuales: Se enfoca en la ejecución humana de los procesos de la creación de las RBs por un análisis del conocimiento formalizado en ontologías (Pilato *et al.*, 2012; Pshenichny, 2014).
- Métodos semiautomáticos: Generan automáticamente alguna parte de la RB de una ontología dada, no obstante, este método requiere la interacción del usuario para ayudar en las partes restantes (Devitt *et al.*, 2006; Fenz *et al.*, 2009).
- Métodos automáticos: Son métodos que generan una RB integral sin la intervención del usuario, aunque con alcance limitado, trabajan una forma especial de entrada de la ontología generando una estructura de RBs predefinida simplificada (Yang *et al.*, 2005; Bucci *et al.*, 2011).

Analizando la literatura se observa que se proponen diferentes formas de construir una RB; en (Yang y Calmet, 2005; Yan y Wei, 2013) se propone la construcción de RBs extendiendo las ontologías, sin embargo, el trabajo manual está presente cuando la ontología es construida. Andrea y Franco (2009) plantearon construir RBs con instancias de la ontología. El enfoque de Andrea podría ser útil cuando la ontología tiene suficientes instancias, pero, la mayoría de las ontologías disponibles no cuentan con tales instancias, estas son distribuidas solamente con su diseño ontológico. Aun cuando la ontología tuviera instancias, el número de estas, no es representativo de las TPCs considerando el dominio educativo. Bucci *et al.* (2011) automatizaron el proceso de generación de RBs con una plantilla usando una base de datos de enfermedades y síntomas, sin embargo, este método reduce el dominio original de la ontología perdiendo información importante. Estos enfoques mencionados no están diseñados para ambientes educativos, por esta razón, una nueva propuesta es necesaria para solventar este hecho.

Hasta ahora, las investigaciones (Yang y Calmet, 2005; Andrea y Franco, 2009; Bucci *et al.*, 2011; Yan y Wei, 2013) no se han enfocado en el dominio educativo, los trabajos anteriores establecen bases para construir RBs tales como generar la estructura y propuestas para construir las TPCs. Sin embargo,

extrapolar estos métodos a otra área es inadecuado en la mayoría de los casos. Una RB, en el campo educativo, necesita precisión en los esquemas de conocimiento a enseñar (parte cualitativa) así como la definición de prioridades o dependencias del conocimiento (parte cuantitativa). Los modelos propuestos anteriormente no cubren estos aspectos. Por esta razón, es necesario un nuevo método de construcción de RBs que resuelvan los problemas en el área educativa.

Este estudio propone manejar el dominio de conocimiento implementado para la toma de decisiones en STIs. Si representamos el dominio de conocimiento del tutor de forma automática, el diseño de cursos para ser enseñados a través de la computadora se volverá menos complejo. Así, la inteligencia del tutor queda a cargo de la RB construida, esta red podrá hacer inferencia del estado de conocimiento y proponer temas acorde a las necesidades de aprendizaje de cada alumno.

1.3. Objetivos

- General
 - Desarrollar un método para construir las partes cualitativa y cuantitativa de una red Bayesiana de forma automática para ser empleado en Sistemas Tutores Inteligentes.
- Específicos
 1. Establecer un modelo para representar una ontología.
 2. Diseñar un algoritmo para obtener una relación cuantitativa entre dos conceptos.
 3. Diseñar un algoritmo para generar las tablas de probabilidad condicional.
 4. Validar la parte cuantitativa de la red Bayesiana con expertos del dominio.
 5. Desarrollar un modelo para crear la parte cualitativa de la red Bayesiana.
 6. Integrar los algoritmos que conforman los elementos individuales del método.

1.4. Preguntas de investigación

El cumplimiento de los objetivos están guiados por las preguntas de investigación definidas a continuación.

Nuestra propuesta contempla las ontologías como una estructura de conocimiento base, es necesario definir un modelo que represente los elementos que componen una ontología. Esta descomposición brinda flexibilidad para definir el método propuesto de este proyecto. Por tanto se formula la siguiente pregunta: **¿Cómo puede representarse un modelo que permita descomponer una ontología en sus elementos?**.

La literatura define distintos aspectos para obtener relaciones cuantitativas entre conceptos. Basado en los argumentos de Li *et al.* (2003), quienes afirman que la relación cuantitativa no sólo depende de múltiples factores, sino de la correcta combinación entre ellos; surge la siguiente pregunta: **¿Cuáles factores deben considerarse para obtener una relación cuantitativa entre conceptos?**.

Las relaciones cuantitativas entre conceptos permiten establecer pesos entre variables para construir TPCs. Estas tablas representan la probabilidad de que ocurra algún evento, en el presente estudio, determinar posesión de conocimiento de los conceptos relacionados. Por tanto, surge la siguiente pregunta: **¿Cómo debe diseñarse el algoritmo para construir las TPCs?**

La parte cuantitativa, representada por las TPCs, necesita ser sometida a pruebas para definir la similitud que existe con los expertos del dominio elegido. Para responder esto, es necesario definir experimentos que cumplan con las necesidades del proyecto. Estos aspectos son respondidos con la pregunta **¿Cómo deben validarse los resultados de la parte cuantitativa propuesta?**

La identificación de los elementos de la RB y de la ontología permiten desarrollar un modelo para construir la parte cualitativa de la red Bayesiana, las preguntas anteriores sirven para guiar la construcción de la parte cuantitativa, sin embargo, aún falta la parte cualitativa, aunque más sencilla requiere una formalización válida. Por tanto, surge la pregunta: **¿Cómo debe diseñarse el modelo para construir la parte cualitativa de la red Bayesiana?**

La propuesta de esta investigación se basa en módulos representados por algoritmos, sin embargo en las primeras fases del proyecto estos algoritmos están desarrollados de forma independiente, por tanto, se requiere integrar los algoritmos de este proyecto, por lo cual surge la pregunta: **¿Cómo pueden integrarse los algoritmos para construir la RB?**

1.5. Contribuciones esperadas

- **Un modelo matemático para representar una ontología.** La presente tesis presenta la necesidad de formalizar un modelo matemático que permita representar una ontología y sus elementos, con el objetivo de tener mayor flexibilidad al diseñar nuestro método.
- **Un algoritmo para obtener una relación cuantitativa entre dos conceptos.** Existen varios métodos para obtener tal relación, sin embargo, dejan fuera el dominio educativo. Un nuevo algoritmo es necesario para cubrir las necesidades presentadas en la educación.
- **Un algoritmo para generar TPCs.** El algoritmo genera TPCs considerando relaciones entre conceptos del punto anterior, este algoritmo maneja aspectos relacionados con la educación.
- **Un método para construir una RB de forma automática para ser empleada en STI.** Las metodologías actuales para desarrollo de RBs mediante ontologías no cubren las necesidades del área educativa, esta propuesta establece un método que tome en cuenta el dominio y los estándares de la estructura base. Además, la RB debe ser construida de forma similar como lo harían los expertos.

1.6. Organización del documento de tesis

El presente documento de tesis está estructurado de acuerdo a los siguientes capítulos:

- **Capítulo 2. Representación del conocimiento:** Esta sección explica el marco teórico que rodea a esta investigación, profundiza sobre las técnicas de representación de conocimientos existentes con un énfasis en las RBs y las ontologías. Además, se incluye los trabajos relacionados a la actual investigación.
- **Capítulo 3. Redes Bayesianas basadas en ontologías:** Este capítulo describe los fundamentos de las RBs así como los conceptos relacionados con las ontologías. En la parte final se describen los trabajos relacionados a esta investigación.
- **Capítulo 4. Relación entre conceptos:** Este capítulo es parte del marco teórico, define aspectos relacionados a la relación cuantitativa entre conceptos. También, profundiza sobre los aspectos de la fuente de conocimiento empleada para construir la parte cuantitativa de la RB.
- **Capítulo 5. Generación de la RB:** Esta sección desarrolla el modelo propuesto, además, desarrolla los modelos para describir las RBs y las ontologías. A su vez, este capítulo establece la formalización del método propuesto.
- **Capítulo 6. Experimentación y resultados:** Esta parte del documento describe dos casos de estudio, los cuales fueron abordados para validar los resultados del método propuesto, un primer caso enfocado a la programación orientada a objetos y el segundo enfocado a los tipos de lenguajes de programación.
- **Capítulo 7. Discusión:** Este capítulo se enfoca en dar interpretación a los resultados de la experimentación, teniendo como enfoque la parte cualitativa y cuantitativa de la RB.
- **Capítulo 8. Conclusión, aportaciones y trabajo futuro:** Por último, esta sección explica algunos puntos finales del proyecto de tesis así como los caminos o líneas que surgen de esta investigación. Las últimas secciones del documento despliegan las referencias empleadas que brindan el sustento teórico a este proyecto y los apéndices que describen las herramientas e instrumentos usados para alcanzar los objetivos de la tesis.

Capítulo 2

Representación del conocimiento en tutores inteligentes

El aprendizaje es un proceso individual, sin embargo, es inadecuado aplicar el mismo método a cada estudiante; para lidiar con esto, se han usado STIs para brindar una enseñanza personalizada. Para que la computadora pueda enseñar un tema es necesario mantener una base de información que pueda ser manejable y entendible por ella misma, esta es la representación del conocimiento.

Este capítulo describe una recopilación de técnicas para representar el conocimiento en STIs, con el fin de analizar las virtudes y defectos de cada una de ellas.

Las técnicas elegidas son las referenciadas en la literatura como técnica de representación del conocimiento de los STIs, destacando (Ramírez-Noriega *et al.*, 2017b): Mapas cognitivos difusos (MCD), RBs, redes semánticas, grafos, entre otras. En la parte final se describe un comparativa entre cada técnica con los factores determinados de cada una.

2.1. Introducción

La educación es el proceso para estimular a una persona a desarrollar sus capacidades cognitivas y físicas. El aprendizaje consiste en el proceso de asimilación de las experiencias, para que pasen a ser parte de nuestra vida y modifiquen de manera importante nuestra conducta y personalidad. Aprendemos de muchas fuentes: el hogar, los amigos, la iglesia entre otras, además de la educación formal. Ésta última es aquella efectuada por docentes profesionales valiéndose de las herramientas que postula la pedagogía para alcanzar sus objetivos (Rivas Navarro, 2008), para el caso de esta investigación enfocada en la enseñanza por computadora.

La tecnología tiene muchos años participando en el proceso educativo, recibe el nombre de tecnología educativa, ésta se define como el estudio y práctica ética para facilitar el aprendizaje y mejorar el desempeño a través de la creación, uso y administración apropiada de los recursos y procesos tecnológicos (Campos y Mata, 2011). El presente trabajo se enfoca en los STIs y otros sistemas que apoyan la enseñanza por medio de la computadora, sin embargo, todos tienen como objetivo la enseñanza de una forma distinta a la tradicional.

Un sistema de enseñanza por computadora necesita conocer lo que enseñará, es decir el dominio de conocimiento, este tiene que ser representado de alguna manera para que la computadora lo entienda, esto es llamado representación del conocimiento. La representación del conocimiento se ha tomado en cuenta desde hace años en las ciencias de la computación. Es importante considerarla para que las máquinas puedan automatizar y procesar el conocimiento además de realizar inferencias que les permitan tomar decisiones como lo hiciera un humano.

Diversos autores establecen que en la memoria semántica (Colace y De Santo, 2006; Rivas Navarro, 2008; Rodríguez, 2010) se define una jerarquización de conceptos con sus relaciones para organizar el conocimiento; de ahí surge la idea de representar el conocimiento por medio de una especie de grafos como lo hacen distintas técnicas: mapas de conceptos, RBs, mapas cognitivos, grafos conceptuales, mapas de conocimiento, redes semánticas y mapas de memoria. Algunos de estos enfoques son

relativamente nuevos y otros ya con años en la literatura.

2.2. Representación del conocimiento

Existen muchas definiciones sobre el conocimiento, sin embargo, tomaremos la siguiente: El conocimiento es una combinación de experiencias, saberes, valores, información contextual, percepción e ideas que crean una estructura mental en el sujeto para evaluar e incorporar nuevas ideas, saberes y experiencias (Jennex *et al.*, 2014). Esa definición se aplica más a nuestra necesidad por la estructura mental para representar la información.

Ramirez y Valdes (2012) concluyeron que distintas definiciones de conocimiento convergen en los siguientes puntos:

- El conocimiento está compuesto de unidades básicas, que son referidas como conceptos.
- Los conceptos tienen asociaciones o relaciones con otros conceptos.
- Los conceptos y relaciones construyen estructuras dinámicas las cuales tienden a convertirse en más estables a través del tiempo.

La representación del conocimiento es el medio por el cual expresamos cosas del mundo, es el medio de expresión y comunicación en el cual le indicamos a una computadora sobre el mundo (Davis *et al.*, 1993). Las cualidades que un buen lenguaje de representación del conocimiento deben tener son (Kesarwani y Misra, 2013):

1. Ser expresivo.
2. Ser conciso.
3. No tener ambigüedades.
4. Ser independiente del contexto: Lo que digas hoy debe ser interpretable mañana.

5. Ser eficiente: El conocimiento debe ser representado en un formato que es adaptable para las computadoras con procedimientos de inferencia prácticos.
6. Ser eficaz: Un procedimiento de inferencia que puede actuar sobre el mismo conocimiento para producir nuevo conocimiento.

2.3. Representación del conocimiento en la enseñanza

La educación en las instituciones educativas es organizada por materias o asignaturas, donde se agrupan los temas comunes de acuerdo a la carrera elegida. Estas materias están organizadas por medio de planes de estudios; que describen, entre otros elementos, los contenidos ha enseñar; aquí se establece un orden y una planeación para trabajar con los alumnos. Generalmente se elabora una dosificación para enseñar el conocimiento por clases.

El índice de contenido representa los temas que se tienen que aprender y su orden. En la Figura 2.1 se muestra el contenido para aprender algoritmos computacionales, como caso de ejemplo. La información está organizada por capítulos o unidades, temas y subtemas, vemos claramente una estructura jerárquica donde se establece categorías y se agrupan temas y subtemas. Este orden jerárquico está organizado por grados de complejidad, donde se ven los contenidos más sencillos de asimilar y con menor grado de razonamiento en un principio, para avanzar a los temas más complejos en fases posteriores.

Este orden está relacionado con la taxonomía de Bloom (De Bruyn *et al.*, 2011; Rodrigues y dos Santos, 2013), donde los primeros conocimientos son simples conceptos y los finales son complejos procesos de análisis, síntesis y evaluación. Podemos analizar los primeros temas donde se contestan preguntas como: ¿qué es un algoritmo?, ¿qué es una variable?, ¿Cuales son los tipos de algoritmos?, entre otras; son básicamente conceptos simples sin razonamiento. Al finalizar el curso el alumno muy probablemente tenga que elaborar pequeños programas en algún editor de algoritmos, poniendo a prueba su razonamiento y aplicando todos los conceptos anteriormente aprendidos.

Este tipo de organización podría ser representado también por grafos, los cuales son una representa-

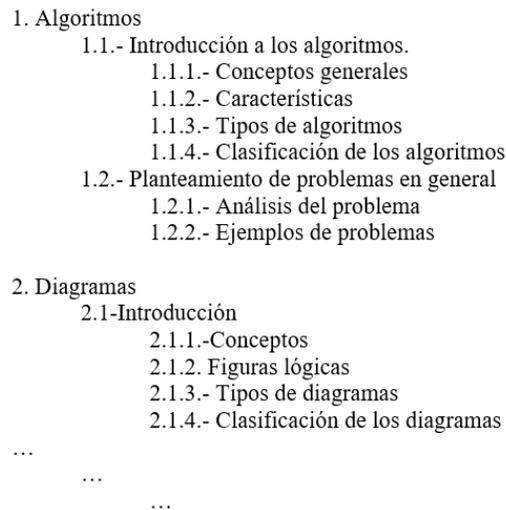


Figura 2.1: Bloque 1 de algoritmos computacionales.

ción de nodos y enlaces, así obtenemos una estructura fácil de analizar por una máquina por medio de alguna técnica de representación del conocimiento. La Figura 2.2 representa una estructura de árbol elaborada con base en los contenidos desplegados en la Figura 2.1. Sólo es considerado el capítulo 1 para simplificar la idea. El capítulo se convierten en el nodo raíz, de ahí surgen los temas, para dejar como nodos hoja a los subtemas.

De esta forma, se puede tener un mayor nivel de granularidad si agregamos los conceptos de cada subtema como se muestra en la Figura 2.3. Esta estructura, en su forma completa, debería estar representada en la memoria del estudiante al finalizar este bloque. Diversas teorías argumentan que la memoria semántica organiza la información en una estructura similar conformada por nodos y enlaces (Rivas Navarro, 2008), así estaríamos simulando de cierta forma lo que el alumno debería de tener en su mente.

Con un enfoque adecuado, esta representación basada en nodos y enlaces también podría servir de evaluación para determinar el conocimiento del alumno, el cual es usado por los STIs y otros sistemas asistidos por computadora. La evaluación es probablemente el elemento más importante para ayudar a aprender al alumno (Brown, 2004). Esta se refiere a una medida comparativa entre lo que el alumno aprendió y lo que debería de aprender (Milne y Witten, 2013), por medio de cualquier procedimiento o

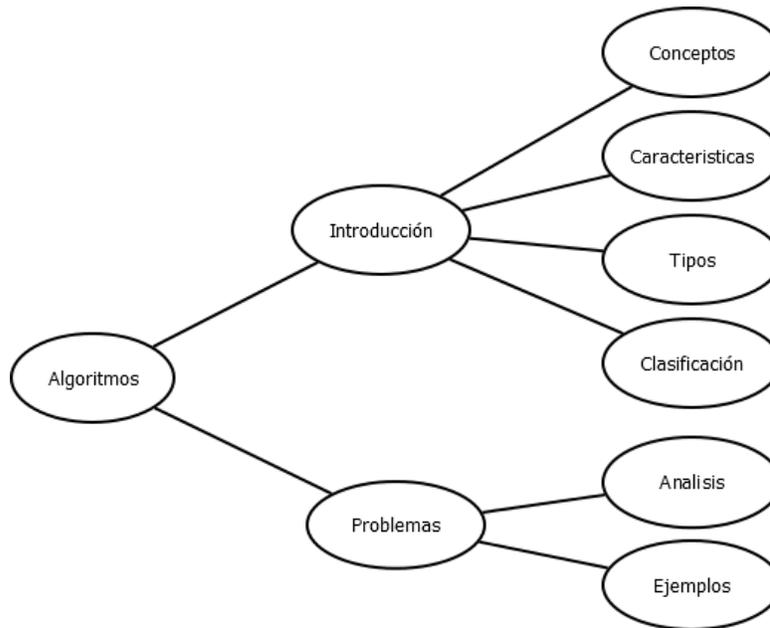


Figura 2.2: Estructura jerárquica por medio de grafos.

actividad diseñada para reunir información sobre el conocimiento, actitudes o habilidades del aprendiz (Kellaghan y Greaney, 2001). Sin embargo, los mecanismos de evaluación son sólo aproximaciones a la realidad, ya que no podemos saber a ciencia cierta qué ha aprendido el alumno. A este desconocimiento de la realidad se le llama incertidumbre (Kovalerchuk, 2013). Es decir, el profesor siempre tendrá incertidumbre sobre lo que el alumno ha aprendido.

2.4. Sistemas tutores inteligentes

El aprendizaje es un proceso individual, como tal, un modelo orientado a la estandarización se enfoca en: Enseñar el mismo contenido, aplicar el mismo método y permanecer en el mismo lugar, esto es aplicado a cada estudiante. Este proceso es inadecuado dadas las necesidades individuales de cada estudiante (Ramirez y Valdes, 2009). Buscando resolver estos y otros problemas, se han usado STIs para aproximarse a la enseñanza personalizada. Un STI, básicamente, es un sistema de software que trata de comportarse como un ser humano; Cataldi (Cataldi y Lage, 2010a) brinda tres definiciones de un tutor inteligente:

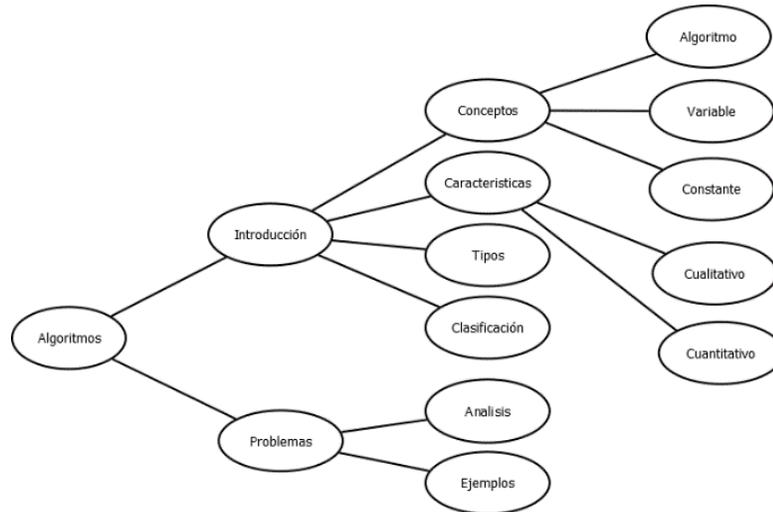


Figura 2.3: Estructura jerárquica por medio de grafos con conceptos.

1. “Es un sistema de software que utiliza técnicas de Inteligencia Artificial (IA) para representar el conocimiento e interactúa con los estudiantes para enseñárselo”.
2. “Sistemas que modelan la enseñanza, el aprendizaje, la comunicación y el dominio conocimiento del especialista y el entendimiento del estudiante sobre ese dominio”.
3. “Un sistema que incorpora técnicas de IA a fin de crear un ambiente que considere los diversos estilos cognitivos de los alumnos que utilizan el programa”.

Las definiciones convergen en un sistema que considera técnicas de IA para dotarlos de razonamiento que le permita tomar decisiones como lo hiciera un tutor de verdad.

El soporte de los STIs está compuesto principalmente por la retroalimentación que provee el sistema como respuesta a las acciones del estudiante. Algunos de los tipos más comunes son resaltar elementos que ya eran mostrados, ofrecer pistas sobre la actividad o solución (por ejemplo, sobre los errores cometidos), enseñar contenido del dominio o ejecutar de forma explícita pasos dentro de la solución (Castillo *et al.*, 2015). Sin embargo, el costo de desarrollar un STI es altamente alto (Hossein *et al.*, 2012)

18 *CAPÍTULO 2. REPRESENTACIÓN DEL CONOCIMIENTO EN TUTORES INTELIGENTES*

Los STIs permiten la emulación de un tutor humano para determinar qué enseñar, cómo enseñar y a quién enseñar. Esto lo hacen a través de (Victorio-Meza *et al.*, 2014)(Cataldi y Lage, 2010a):

- Un módulo del dominio: que define el dominio del conocimiento.
- Un módulo del estudiante: que es capaz de definir el conocimiento del estudiante en cada punto durante la sesión de trabajo.
- Un módulo del tutor: que genera las interacciones de aprendizaje basadas en las discrepancias entre el especialista y el estudiante.
- Una interfaz con el usuario: que permite la interacción del estudiante con un STI de una manera eficiente (conocimiento sobre cómo presentar los contenidos).

Un STI organiza un propósito educativo de acuerdo a un criterio pedagógico. Estos criterios determinan la representación del conocimiento en el modelo del dominio, el modelo del estudiante y su proceso de actualización. Una forma común de representación de conocimiento adoptada por los STIs es una red jerárquica, donde los nodos son conceptos y la relación entre nodos está conectada por arcos o flechas. Generalmente, esta estructura puede tomar la forma de red semántica, mapa conceptual o RB (Badaracco y Martínez, 2011).

A continuación se describen las técnicas más comunes de representación de conocimiento en STIs.

2.5. Técnicas de representación del conocimiento

Esta sección hace un análisis de las técnicas que se utilizan en los últimos años para representar el dominio de conocimiento de los STIs. Existen algunas otras técnicas, sin embargo, las listadas en esta sección son las referenciadas en la literatura con uso en la educación asistida por computadora, especialmente en los STIs. Las técnicas encontradas son las siguientes:

- Mapa de conceptos (Rollande y Grundspenkis, 2012; Hossein *et al.*, 2012; Liu y Lee, 2013)

- Red Bayesiana (Krudysz *et al.*, 2006; Yamna *et al.*, 2010; Sedki y Beaufort, 2012; Khodeir *et al.*, 2012a; Chrysafiadi y Virvou, 2013; Millán *et al.*, 2013)
- Mapas cognitivos (Ughev, 2014)
 - Como red Bayesiana (Sedki y Beaufort, 2012)
 - Difusos (Hossein *et al.*, 2012)
- Grafos conceptuales (Kavcic, 2004; Rollande y Grundspenkis, 2012)
- Mapas de conocimiento (Mishra *et al.*, 2012)
- Redes Semánticas (Badaracco y Martínez, 2011; Satar, 2012)
- Mapa de memoria (Ramirez y Valdes, 2012)

A continuación se define cada una de ellas.

2.5.1. Mapa de conceptos

Un mapa de conceptos es una técnica popular gráfica para modelar el conocimiento en el salón de clases (Hossein *et al.*, 2012; Liu y Lee, 2013). Esta técnica consiste en conceptos y preposiciones, las cuales esta usualmente encerradas por un círculo o rectángulo; flechas unidireccionales enlazan los conceptos, de este modo las flechas demuestran la relación entre ellos. El mapa de conceptos es una alternativa para resumir información en forma de diagrama (Liu y Lee, 2013).

2.5.2. Red Bayesiana

Una RB es un modelo gráfico probabilístico cuya estructura consiste en nodos enlazados con arcos dirigidos. Esto significa que las relaciones entre variables tienen un sentido (unidireccional) (Krudysz *et al.*, 2006). Los nodos representan variables aleatorias y los arcos entre los nodos indican relaciones causales entre variables. Cada nodo tiene asociada una TPC, estas tablas contienen una distribución de

probabilidades condicionales de cada resultado del nodo hijo a cada combinación de valores de los nodos padres (Baraldi *et al.*, 2015). El capítulo 3 brinda más detalles sobre este enfoque.

2.5.3. Mapas cognitivos

Los mapas cognitivos son una técnica utilizada para estructurar y administrar datos desordenados o complejos para su comprensión y fácil acceso. Surgen de la idea que en el curso del aprendizaje, un mapa de campo es desarrollado en el cerebro, por ello, el aprendizaje consiste en la enlace de conjuntos de neuronas en el sistema nervioso que funcionan como un mapa cognitivo, que es utilizado para hacer diversas representaciones de su entorno (Peña Ayala, 2007). Esta teoría sugiere que cada individuo tiene su propia conceptualización del conocimiento y que la estructura puede variar entre individuos (Yengin y Saerbeck, 2012). A pesar de que la conceptualización difiere de persona a persona, los mapas cognitivos son adecuados para comunicar sus propios conceptos y su relaciones (Yengin y Saerbeck, 2012; Rodrigues *et al.*, 2013).

2.5.3.1. Mapa cognitivo como red Bayesiana

Sedki et al. (Sedki y Beaufort, 2012) propusieron este enfoque combinando mapas cognitivos y RBs. Ellos primero crean un mapa cognitivo y luego lo transforman a un Mapa Cognitivo Bayesiano (BCM por sus siglas en inglés Bayesian Cognitive Map), el cual combina técnicas de modelado causal y la teoría de probabilidad Bayesiana. El BCM que ellos generan corresponde a un mapa cognitivo con modificaciones como relaciones directas, indirectas y circulares, y dependencia entre variables. Este es un nuevo modelo de reciente surgimiento en la literatura para representar conocimiento.

2.5.3.2. Mapa cognitivo difuso

Los mapas cognitivos representan el conocimiento causal sin indicar el grado de fortaleza de la relación causa-efecto (Leyva-Vazquez, 2013), cuando son combinados con lógica difusa surgen los MCD como una técnica de la IA. Los MCD son grafos difusos para representar razonamiento causal

capaz de codificar conocimiento usando conjuntos difusos. Los MCD mejoran los mapas cognitivos describiendo la fortaleza de la relación mediante el empleo de valores borrosos en el intervalo $[-1,1]$ (Hosseini *et al.*, 2012). Los nodos son conceptos causales y pueden modelar eventos, acciones, valores, metas o procesos. Constituyen una estructura de grafo difuso con retroalimentación para representar causalidad. Ofrecen un marco de trabajo más potente y flexible para representar el conocimiento humano y para el razonamiento de los sistemas expertos tradicionales que implementan explícitamente reglas de tipo "si-entonces" (Leyva-Vazquez, 2013; Baron *et al.*, 2014).

2.5.4. Grafos conceptuales

Los grafos conceptuales son un formalismo de representación del conocimiento desarrollado por Sowa (1984) y empleado por primera vez en terminología por Hook y Ahmad (1992). Los grafos permiten representar la estructura interconectada del conocimiento de una manera flexible, con diferentes configuraciones y puntos de acceso para llegar a los mismos datos. Por otro lado, los grafos permiten hacer inferencias y, por tanto, (re)generar conocimiento nuevo (Maroto García, 2007).

2.5.5. Mapas de conocimiento

Los mapas de conocimiento son grafos dirigidos guiando el proceso de evaluación por medio del aseguramiento del dominio de los temas pre-requisitos antes de avanzar a los siguientes temas. En el nivel más alto, un mapa de conocimiento representa la arquitectura del currículo entero. En un segundo nivel, el mapa de conocimiento representa cada curso como una interconexión de módulos. El tercer nivel muestra la arquitectura para los conceptos del módulo y sus relaciones (Mishra *et al.*, 2012).

2.5.6. Redes semánticas

Una red semántica es un grafo con nodos representando conceptos u objetos individuales y arcos representando relaciones o asociaciones entre los conceptos (Chandrasegaran *et al.*, 2013). Un objeto

puede ser cualquier ítem físico como un libro, un carro, un escritorio, o incluso una persona. Los nodos también pueden ser conceptos, eventos o acciones. Algunos de los arcos más comunes son del tipo es-un o tiene-un. La relación es-un es usada para mostrar una relación de clase en casos donde un objeto pertenece a una clase más grande o una categoría de objetos. Por otra parte, la relación tiene-un es usada para identificar características o atributos de nodo de objeto (Kesarwani y Misra, 2013).

2.5.7. Mapa de memoria

El mapa de memoria es un modelo de representación del conocimiento para conceptos y habilidades, su principal meta es representar la interacción de esos elementos en diferentes conceptos. Los mapas de memoria pueden ser visualizados como una red semántica, la principal diferencia entre los mapas de memoria y otros modelos es que los mapas de memoria están fuertemente enfocados en la flexibilidad del contexto, debido a este enfoque, los conceptos de mapa de memoria y habilidades deben tener una granularidad abierta sujeta a los criterios del modelador; un nivel arbitrario de atomicidad que se puede especificar para cada concepto y jerarquías dinámicas que pueden cambiar para diferentes dominios del conocimiento. La implementación del mapa de memoria es un grafo dirigido, muy similar a los tipos más flexibles de redes semánticas y a las ontologías (Ramirez y Valdes, 2012).

2.6. Características de la representación del conocimiento

Por medio de un análisis de las técnicas anteriormente citadas, fueron identificados aquellos elementos que son considerados para representar el conocimiento en el ambiente educativo. La Figura 2.4 despliega un diagrama con los elementos del conocimiento acorde a la literatura revisada.

A continuación se describen los elementos del diagrama.

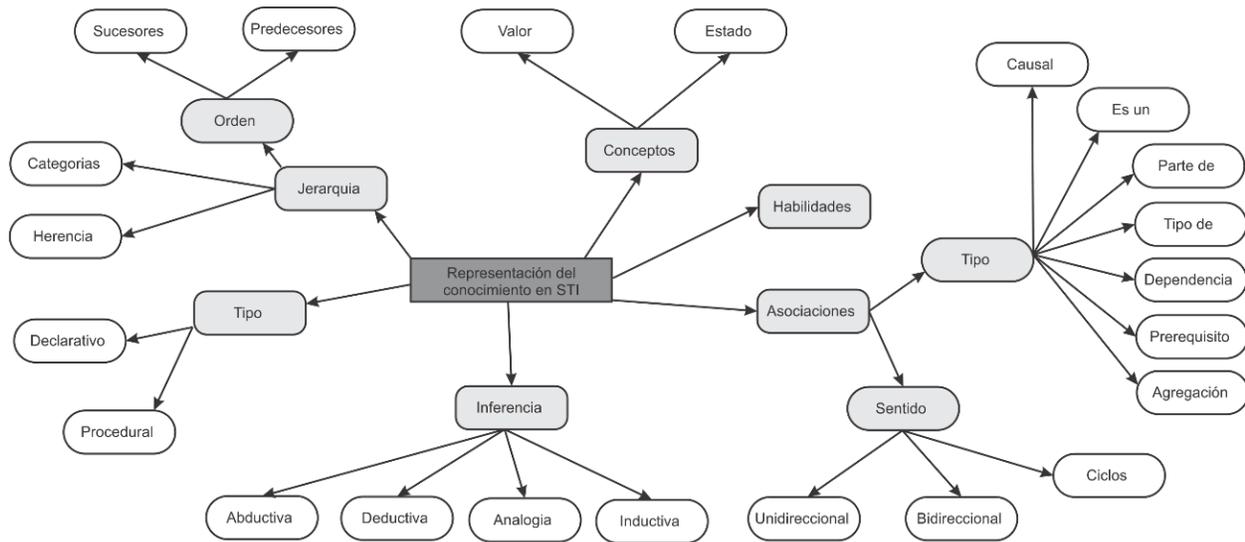


Figura 2.4: Elementos considerados de representación del conocimiento por STIs.

2.6.1. Conceptos

El concepto se forma mediante la sucesión de cuatro actividades cognitivas: análisis, abstracción, síntesis y generalización. El análisis es la facultad cognitiva que separa un objeto o fenómeno por sus elementos. La abstracción es la actividad cognitiva que enfoca la atención en algunos elementos del objeto o fenómeno. La síntesis es la composición mental de los elementos del objeto en un todo. La generalización es la actividad mental orientada a integrar las propiedades de varios objetos o fenómenos de la misma clase (Peña Ayala, 2007).

Un concepto es definido como la representación de un objeto mental con sus atributos, expresado a través de un lenguaje específico de la mente, el cual permite ser representado en símbolos y caracteres, por lo cual puede ser computable (Ramírez y Valdes, 2009). El concepto es una pieza elemental del conocimiento, en el sentido que no puede ser descompuesto en partes más pequeñas, a juicio del experto del dominio. El concepto es considerado la unidad básica del conocimiento (Millán y Pérez-De-La-Cruz, 2002).

Según Maroto (Maroto García, 2007) las características de los conceptos son:

- a) Los conceptos pertenecen al ámbito del pensamiento, y no al de los objetos de la realidad, si bien

surgen a partir de ella mediante la abstracción de sus características.

- b) Los conceptos permiten estructurar el conocimiento sobre los objetos.
- c) Los conceptos abarcan no sólo a objetos o entidades materiales, sino también a cualidades, acciones y relaciones. La terminología se interesa únicamente por aquellos conceptos relevantes para los ámbitos de especialidad, y no por los de la lengua general.
- d) Los conceptos desempeñan un papel en la comunicación.

El concepto es el elemento principal para todas las técnicas encontradas (Millán y Pérez-De-La-Cruz, 2002; Krudysz *et al.*, 2006; Ramirez y Valdes, 2009; Satar, 2012; Millán *et al.*, 2013; Liu y Lee, 2013) y es utilizado por cada una de ellas. Estos puede tener distintos tipos de atributos que complementan el entendimiento del concepto describiendo sus características (Ramirez y Valdes, 2009). Sin embargo, los trabajos relacionados a la educación asistida por computadora analizados se basan en dos aspectos principales:

1. Un aspecto cuantitativo definido como valor o peso de posesión de conocimiento (Ngo *et al.*, 2011; Yamna *et al.*, 2010; Khodeir *et al.*, 2012b; Mishra *et al.*, 2012; Millán *et al.*, 2013), el cual se refiere a un valor numérico asignado al nodo que representa el concepto.
2. Un aspecto cualitativo definido como estado (Mishra *et al.*, 2012) que se refiere a un valor discreto en que puede estar la posesión del conocimiento.

2.6.2. Habilidades

Una habilidad está definida como un proceso cognitivo que interactúa con uno o más conceptos, usualmente a través de la aplicación, la cual tiene un propósito específico y produce un resultado, siendo interno o externo. Las habilidades tienen diferentes grados de complejidad y pueden ser integradas con otras habilidades, debido a que usualmente son relacionadas con la aplicación o la acción tiende a ser

representadas por verbos (Millán y Pérez-De-La-Cruz, 2002; Ramirez y Valdes, 2009). Una habilidad también puede ser catalogada como un concepto depende de cómo el autor la desea tomar.

2.6.3. Asociaciones

También conocidas como relaciones o enlaces. El objetivo de la asociación es conocer como los conceptos están relacionados (Millán y Pérez-De-La-Cruz, 2002; Friss de Kereki Guerrero, 2003; Ramirez y Valdes, 2009; Mishra *et al.*, 2012; Satar, 2012; Millán *et al.*, 2013; Badaracco y Martínez, 2013). En terminología se suele hablar de relaciones conceptuales para referirse a los vínculos que se establecen entre conceptos dentro de un ámbito de especialidad. Las relaciones conceptuales son fundamentales para establecer sistemas que reflejan la estructuración del conocimiento especializado (Maroto García, 2007), es decir determina un significado semántica para el tipo de relación (Satar, 2012).

Las relaciones tienen como propiedades (Friss de Kereki Guerrero, 2003):

- La valencia, que describe la forma cuantitativa en que los elementos intervienen en la relación.
- Su funcionalidad, que señala un tipo de implementación de la relación; puede ser inyectiva y sobreyectiva.
- La cardinalidad, que determina el número de elementos que participan en la relación.

Tradicionalmente, han sido cuatro las clases de relaciones conceptuales que se han considerado fundamentales para la estructuración de un campo conceptual (con fines terminológicos). Estas relaciones pueden resumirse de la siguiente manera (Pérez Hernández, 2002):

- Relaciones genérico/específico: Relación jerárquica en la que se identifica a los conceptos por su pertenencia a una categoría, en la que un concepto genérico se considera superordinado de otros conceptos más específicos. Los conceptos subordinados comparten las características del concepto genérico, pero, además, poseen algunas peculiaridades propias que los diferencian y

26 CAPÍTULO 2. REPRESENTACIÓN DEL CONOCIMIENTO EN TUTORES INTELIGENTES

hacen más específicos. Se establece, por tanto, una relación que va en dos sentidos diferentes: vertical, es decir, la que se establece entre un concepto específico (subordinado) y su genérico (superordinado); y horizontal, la que sostienen varios conceptos específicos que poseen el mismo genérico y que se diferencian entre sí por poseer alguna característica distintiva (conceptos coordinados).

- Relaciones parte/todo: se refiere a la que existe entre conceptos que están formados por más de una parte y dichas partes constituyentes.
- Relaciones polivalentes: En las que se da cuenta de la posibilidad de que un concepto pueda colocarse en lugares diferentes en un mismo sistema conceptual.
- Relaciones complejas: Categoría en la que se engloban una serie de interrelaciones que se establecen entre conceptos en una jerarquía, pero que no pueden considerarse ni genéricas ni partitivas. Ejemplos de este tipo de relaciones son las de causa-efecto, actividad-lugar de realización, proceso-producto, etc.

La descripción anterior es una clasificación general; los tipos de relaciones encontradas en la revisión de literatura son:

- Causal (Millán y Pérez-De-La-Cruz, 2002; Krudysz *et al.*, 2006; Uglev, 2014; Sedki y Beaufort, 2012).
- Es un (Is-a) (Kesarwani y Misra, 2013) (Ben Messaoud *et al.*, 2013).
- Parte de (Part-of) (Kavcic, 2004; Satar, 2012; Ben Messaoud *et al.*, 2013; Badaracco y Martínez, 2013; Baazaoui-zghal y Ghezala, 2014).
- Tipo de (Type-of) (Satar, 2012)
- Dependencia (Depends-on, Dependency) (Kavcic, 2004; Liu y Wang, 2007).

- Prerrequisito de (Prerequisite-of) (Millán y Pérez-De-La-Cruz, 2002)
- Agregación (Aggregation) (Liu y Wang, 2007; Badaracco y Martínez, 2011)

Existen otras relaciones como (Maroto García, 2007; Rodríguez, 2010): Contextualidad espacio-temporal (“ocurrió en”), Atributo (“posee”), ejemplo (“tiene como instancia”), equivalencia/identidad (“es igual a”) entre otras. Existe una variedad de relaciones que pueden ser definidas de acuerdo a las necesidades de cada autor si la estructura lo permite, cada estructura tiene sus propias características y algunas como las redes semánticas permiten definir la relación entre conceptos.

Las relaciones son del tipo binaria, es decir de concepto a concepto, sin embargo pueden ser de tres tipos según el sentido de la asociación:

- Unidireccional: Relación de un concepto *A* hacia un concepto *B* (Krudysz *et al.*, 2006; Millán *et al.*, 2013; Ramirez y Valdes, 2012; Rodrigues *et al.*, 2013).
- Bidireccional: Relación de un concepto *A* hacia un concepto *B* y viceversa (Ramirez y Valdes, 2012; Rodrigues *et al.*, 2013).
- Ciclos: Relación de un concepto *A* con sí mismo, ya sea por medio de otros conceptos o a sí mismo. (Ramirez y Valdes, 2012; Sedki y Beaufort, 2012).

2.6.4. Inferencia

También se le llama razonamiento, se refiere a la obtención de deducciones o conclusiones en base al conocimiento ya establecido (Davis *et al.*, 1993; Arevalillo-Herráez *et al.*, 2013). Los principales tipos de inferencia son (Peña Ayala, 2007; Ramirez y Valdes, 2012):

- Abductiva: Consiste en descubrir el estado de la situación más probable dada una evidencia (Larrañaga y Moral, 2011).

- **Deductiva:** Se refiere a predecir un efecto dado una causa, así, podemos obtener respuestas sobre efectos dados la causa pero no podemos contestar porque los efectos son producidos (Sedki y Beaufort, 2012). Parte de lo general a lo particular (Friss de Kereki Guerrero, 2003).
- **Analogía:** Es un proceso de descomponer una idea en partes y buscar una idea o situación del dominio objetivo que comparta un subconjuntos de esos aspectos (Liu y Singh, 2004), así se podrá entender la idea por medio de similitudes entre casos.
- **Inductiva:** Varias ideas específicas previas se incluyen en una nueva, más general, de categoría u orden superior. Este tipo de progresión es de carácter ascendente (de abajo-arriba) (Rivas Navarro, 2008). Parte de lo particular y va a lo general (Friss de Kereki Guerrero, 2003).

2.6.5. Tipo de conocimiento

Existen dos tipos de conocimiento, el conocimiento declarativo y procedural. El conocimiento procedural es más complicado de representar que el declarativo. El conocimiento declarativo es el que se aplica en la educación, fácil de representar y estructurar; por tanto, es el tipo de conocimiento que enseñan los sistemas asistidos por computadoras (Millán y Pérez-De-La-Cruz, 2002; Krudysz *et al.*, 2006; Sharma y Kelkar, 2012; Victorio-Meza *et al.*, 2014). El conocimiento declarativo generalmente considera si una afirmación es verdadera o falsa, así, todo conocimiento declarativo es explícito.

El conocimiento procedural se refiere al orden que debe ser seguido para hacer una actividades o resolver un problema, aunque, aparentemente, este tipo de conocimiento se refiere a una serie de pasos a seguir como un algoritmo, representar ciertas actividades se vuelve muy complejo, por ejemplo, la habilidad de pasear en bicicleta (Ramirez y Valdes, 2012).

2.6.6. Estructura jerárquica

La información en las técnicas de representación de conocimiento son organizadas a través de un estructura jerárquica para compartir propiedades y formar superclases y subclases (Nguyen *et al.*, 2013),

así podremos organizar el conocimiento dividiendo el dominio en secciones (Larrañaga *et al.*, 2014) y elaborando categorías de conocimiento. En algunos casos esta información es dividida de acuerdo a la estructura educativa basada en unidades, temas y subtemas (Millán *et al.*, 2013; Liu y Lee, 2013), esto permite una organización similar a la de los libros y los planes educativos de las instituciones.

Tomando como base la estructura jerárquica, surgen elementos que son considerados por distintos autores como importantes debido a que establecen una estructura para mantener en orden los contenidos (Ramirez y Valdes, 2009; Tian, 2009; Mishra *et al.*, 2012; Badaracco y Martínez, 2013). Estos elementos son: sucesores, predecesores, categorías, herencias, entre otros, los cuales son definidos a continuación.

2.6.6.1. Sucesores y predecesores

Son dos percepciones relacionadas al orden de los conceptos en la estructura jerárquica; se refieren al conocimiento posterior o previo que el alumno debe tener dado un concepto, también pueden ser llamadas consecuentes y antecedentes. Los conceptos sucesores son elementos del conocimiento que se consideran subsecuentes al actual (Mishra *et al.*, 2012), es decir el concepto actual sirve como base para aprender los conceptos siguientes.

Los conceptos predecesores son considerados como elementos previos necesarios para comprender el actual. Es importante el conocimiento previo adquirido, porque sienta las bases para el posterior (Liu y Wang, 2007; Ramirez y Valdes, 2009; Mishra *et al.*, 2012). Estos dos aspectos están muy relacionados debido a los sucesores tienen predecesores y los predecesores tienen sucesores.

2.6.6.2. Categorías

La categorización es un proceso mediante el cual el objeto se encuadra dentro de una clase o categoría, implicando atribución de significado. Su clasificación como miembro de una categoría, supone asignarle características, atributos o propiedades de tal categoría, vinculando al objeto la información pertinente disponible en la memoria semántica sobre la misma (Rivas Navarro, 2008).

Las personas necesitan dividir el mundo en categorías o clases de objetos para dar sentido al propio

conocimiento. La organización propia de la memoria semántica está basada en un sistema similar al de categorías de conceptos, de tal modo que, los más concretos y específicos son miembros de conjuntos más amplios y generales. Esta estructuración jerárquica en categorías supraordenadas y subordinadas permite realizar inferencias y predicciones, esenciales en la cognición y la acción humana, partiendo de que las propiedades o atributos de las categorías más generales son también verdaderas para las más específicas o particulares semánticamente conectadas a aquellas (Rivas Navarro, 2008). Distintos autores hacen uso de este enfoque para organizar el conocimiento (Rosić *et al.*, 2006; Liu y Lee, 2013).

2.6.6.3. Herencia

Es un sistema de razonamiento que lleva a deducir propiedades de un concepto basándose en las propiedades de conceptos más altos en la jerarquía (Socorro *et al.*, 2008). Es decir, permite que los conceptos relacionados cuenten con los atributos de los conceptos con la cual se relacionan. Algunos de los autores que hacen uso del concepto son (Sharma y Kelkar, 2012; Pilato *et al.*, 2012; Satar, 2012; Ramirez y Valdes, 2012).

2.6.6.4. Representación gráfica

La estructura jerárquica fácilmente puede ser representada gráficamente, todos los enfoques aquí manejados comparten la ventaja de ser representados gráficamente. La representación gráfica permite la fácil interpretación de la organización del conocimiento a través de un estructura esquemática (Richardson y Domingos, 2006; Millán *et al.*, 2010; Liu y Lee, 2013; Rau *et al.*, 2015).

2.7. Comparativa entre las técnicas de representación del conocimiento

A continuación se establece una serie de tablas comparativas analizando los aspectos definidos en la sección anterior. Se Consideran agrupaciones como se establece en la Figura 2.4 para evitar tablas

2.7. COMPARATIVA ENTRE LAS TÉCNICAS DE REPRESENTACIÓN DEL CONOCIMIENTO 31

Tabla 2.1: Comparativa concepto y atributos.

| Enfoque | Conceptos | Atributos | |
|-------------------------------------|-----------|-----------|--------|
| | | Valor | Estado |
| Mapa de conceptos | x | | |
| Red Bayesiana | x | x | c |
| Mapas cognitivos como red Bayesiana | x | c | c |
| Mapas cognitivos difusos | x | x | x |
| Grafos | x | | |
| Mapas de conocimiento | x | x | x |
| Red semántica | x | | |
| Mapa de memoria | x | | |

demasiado grandes. Se consideró la organización en las áreas: concepto y sus atributos, asociaciones, inferencia y tipo de conocimiento, y por último los aspectos relacionados con la estructura jerárquica. Las tablas más abajo referenciadas consideran la misma simbología. La primer columna muestra los enfoques analizados, el resto los elementos del conocimiento que se están evaluado. La letra *x* corresponde aquellos elementos que son considerados en los artículos estudiados y la letra *c* corresponde a elementos no considerados en el estudio pero la técnica podría soportar si fuera requerido.

2.7.1. Conceptos y atributos

Cada uno de los enfoques está basado en la definición de conceptos y sus enlaces, por tanto todos cumplen con ese aspecto como marca la Tabla 2.1.

A diferencia de otras técnicas como las ontologías, los enfoques analizados sólo consideran los atributos del concepto referentes a su valor cuantificable o su estado cualitativo. En cuanto a su valor, técnicas como las RBs, los MCD o los mapas de conocimiento consideran el aspecto cuantitativo para definir un grado o probabilidad de posesión de conocimiento. Además, el atributo estado considera un aspecto discreto en cuanto a la posesión de conocimiento, es decir, podemos catalogar el conocimiento en estados como excelente, muy bien, bien, regular, no deseable o simplemente como aprobado o reprobado.

Tabla 2.2: Comparativa asociaciones.

| Enfoque | Asociaciones | Tipos de asociaciones | | | | | | | Sentido de la asociación | | | |
|-------------------------------------|--------------|-----------------------|-------|----------|---------|-------------|---------------|------------|--------------------------|----------------|---------------|--------|
| | | Causal | Es un | Parte de | Tipo de | Dependencia | Prerrequisito | Agregación | Definir la relación | Unidireccional | Bidireccional | Ciclos |
| Mapa de conceptos | x | | | | | x | | x | c | c | | |
| Red Bayesiana | x | x | | | | | | x | | x | | |
| Mapas cognitivos como red Bayesiana | x | x | | | | x | | x | | c | | x |
| Mapas cognitivos difusos | x | x | | | | | | x | | x | | |
| Grafos | x | | | x | | x | | | | c | c | |
| Mapas de conocimiento | x | | | x | | | | | | x | | |
| Red semántica | x | | | x | x | | | x | c | x | c | c |
| Mapa de memoria | x | | | | | | | | x | x | | |

Estos atributos agregan un valor extra a la representación del conocimiento al considerar la incertidumbre del dominio. La RB considera el aspecto cuantitativo para modelar el dominio de conocimiento, sin embargo, en los estudios analizados no contemplaron el aspecto cualitativo que fácilmente podrían representarlo. Los mapas cognitivos como redes Bayesianas no contemplan ni el valor cuantitativo ni el estado cualitativo en los estudios analizados. Por último los MCD y los mapas de conocimiento consideran tanto el aspecto cuantitativo como el cualitativo.

2.7.2. Asociaciones

Las asociaciones contempladas en cada uno de los enfoques varían bastante de acuerdo a la necesidad del autor, en la Tabla 2.2 encontramos siete distintos tipos (causal, es un, parte de, tipo de, dependencia, prerrequisito, agregación), una octava columna en el tipo de asociaciones muestra aquellas estructuras que pueden definir su propia relación. Las técnicas que muestran gran flexibilidad para representar el dominio dado que pueden definir sus relaciones son: los mapas de conceptos, los grafos conceptuales, las redes semánticas y los mapas de memoria.

A su vez las relaciones pueden ser de tres tipos, unidireccional, direccional y ciclos. Cada enfoque puede manejar al menos la relación unidireccional que son las más generalmente usadas, sólo las redes semánticas pueden manejar relaciones en ambos sentidos y ciclos. Los mapas cognitivos como redes

2.7. COMPARATIVA ENTRE LAS TÉCNICAS DE REPRESENTACIÓN DEL CONOCIMIENTO 33

Tabla 2.3: Comparativa inferencia y tipo de conocimiento.

| Enfoque | Inferencia | | | | Conocimiento | |
|-------------------------------------|------------|-----------|----------|-----------|--------------|------------|
| | Abductiva | Deductiva | Analogía | Inductiva | Declarativo | Procedural |
| Mapa de conceptos | | c | | | c | |
| Red Bayesiana | x | x | | | c | |
| Mapas cognitivos como red Bayesiana | x | x | | | c | |
| Mapas cognitivos difusos | x | x | | | c | |
| Grafos | | c | | | c | |
| Mapas de conocimiento | | c | | | c | |
| Red semántica | | c | | | c | |
| Mapa de memoria | | c | | | x | |

Bayesianas tiene la capacidad de manejar ciclos, aspecto que una RB simple no podría hacerlo por definición.

2.7.3. Inferencia y tipo de conocimiento

La Tabla 2.3 muestra la información referente a la inferencia de cada técnica y al tipo de conocimiento que puede representar. El tipo de inferencia está muy relacionada a la información cuantitativa que se pueda almacenar en la representación del dominio. Así las RBs, mapas cognitivos con redes Bayesianas y los MCD son los que hacen uso de esta ventaja por la naturaleza de su teoría, teniendo la posibilidad de hacer inferencias abductivas y deductivas. Los mapas de memoria tiene la capacidad de hacer inferencias deductivas aunque no fue requerido en los estudios analizados. En cuanto al tipo de conocimiento, todas las técnicas se enfocan en representar el conocimiento declarativo

2.7.4. Estructura jerárquica

Cada una de las técnicas se basan en el manejo de conceptos y enlaces, por tanto, todos tienen un representación gráfica un tanto similar basada en una estructura jerárquica. De los cuales se derivan algunos conceptos de su representación como: sucesores y predecesores, para llevar un orden de:

Tabla 2.4: Comparativa jerarquía.

| Enfoques | Sucesores | Predecesores | Jerarquía | Categorías | Herencia |
|-------------------------------------|-----------|--------------|-----------|------------|----------|
| Mapa de conceptos | c | x | c | x | c |
| Red Bayesiana | c | c | c | x | c |
| Mapas cognitivos como red Bayesiana | c | x | c | c | c |
| Mapas cognitivos difusos | c | c | x | c | c |
| Grafos | c | c | c | c | c |
| Mapas de conocimiento | x | x | x | c | c |
| Red semántica | c | c | x | c | x |
| Mapa de memoria | x | x | x | x | x |

aprendizaje de conceptos, jerarquía propia de la organización del conocimiento en forma de árbol, categorización para agrupación de conceptos con similitudes y por último herencia para tomar propiedades de los padres. Todos los elementos pueden ser representados en las técnicas, algunos lo hacen explícitamente y otros no, como dijimos anteriormente depende de lo que los autores quieren representar. Los mapas de memoria en los estudios considerados representan explícitamente cada uno de los factores considerados (ver Tabla 2.4).

2.8. Análisis de las técnicas

Como lo indica Davis *et al.* (1993), las técnicas de representación del conocimiento son una sustitución de la realidad, la única representación precisa del objeto es el objeto mismo, todas las demás son aproximaciones cercanas e imperfectas. Cada técnica de representación del conocimiento provee su propia visión de cómo debería verse el mundo, además definen qué aspectos son importantes plasmar y cuáles no para ser ignorados. Diferentes enfoques guían a diferentes metas y diferentes definiciones de éxito. Un lenguaje diseñado para expresar hechos declarativos no es necesariamente bueno para expresar información imperativa características de una estrategia de razonamiento.

Tomando en cuenta lo anterior, la decisión de cuál es la mejor técnica de representación del

conocimiento está dada por el problema a solucionar y los objetivos perseguidos. Así, cada técnica tiene sus ventajas y desventajas para modelar dominios de conocimiento. Sin embargo, todas comparten la capacidad de modelar el conocimiento por medio de conceptos y relaciones, formando una estructura jerárquica con las ventajas que esto conlleva.

Los mapas de conceptos y grafos conceptuales tienen entre sus ventajas poder representar el tipo de relación entre conceptos de la forma que desee el autor, brindando mayor representación semántica. Las mismas ventajas tienen los mapas de memoria, aunque al ser un enfoque relativamente nuevo no ha sido tan estudiado como los dos primeros, sin embargo, su fundamentación promete grandes alcances.

Las redes semánticas, entre las que están las ontologías, destacan su gran flexibilidad para representar la información dada su gran expresividad para modelar el mundo; tienen un gran respaldo de la W3C¹ al generar un estándar para construir estas estructuras (OWL). Entre todas las técnicas aquí mencionadas la consideramos la mejor cuando no hay que modelar la incertidumbre, lo cual es su principal desventaja no tomar los aspectos cuantitativos del mundo.

Los mapas de conocimiento argumentan buenos principios e incluso consideran pesos en los conceptos para modelar el conocimiento, sin embargo, no es una de las técnicas más usadas, los autores de este enfoque no argumenta el uso práctico de los pesos o el tipo de inferencia que podemos realizar con estos.

La RB es una técnica de razonamiento aproximado con alcance para modelar parte del mundo sin demasiada expresividad semántica, lo cual es su principal desventaja. Fueron desarrolladas para resolver dominios que manejan incertidumbre, área donde mejor se desenvuelven, además otra de sus ventajas es la inferencia abductiva y deductiva que se puede lograr con su representación. Para darle mayor flexibilidad semántica a este enfoque han surgido los mapas cognitivos como redes Bayesianas (Sedki y Beaufort, 2012), una de las grandes aportaciones de este fusión es el manejo de ciclos, no considerados por las red Bayesianas por definición (Pearl, 1988).

Por último los MCD, este enfoque toma las ventajas de los mapas cognitivos y la lógica difusa;

¹<https://www.w3.org/>

Tabla 2.5: Principal ventaja y desventaja de las técnicas.

| Enfoques | Ventaja | Desventaja |
|-------------------------------------|---|--|
| Mapa de conceptos | Definición de relaciones | Necesidad de inferencia más eficiente |
| Red Bayesiana | Inferencia abductiva y deductiva, | Baja representación semántica |
| Mapas cognitivos como red Bayesiana | Inferencia abductiva y deductiva | Enfoque nuevo y poco trabajado |
| Mapas cognitivos difusos | Inferencia abductiva y deductiva | Poca disponibilidad de herramientas de soporte |
| Grafos | Definición de relaciones | Necesidad de inferencia más eficiente |
| Mapas de conocimiento | Aspectos cuantitativos y cualitativos para representar el dominio | Necesidad de inferencia más eficiente |
| Red semántica | Alta expresividad semántica | No manejo de incertidumbre |
| Mapa de memoria | Expresividad semántica adecuada | Enfoque nuevo y poco trabajado |

al igual que las RBs, permite modelar dominios con incertidumbre por medio de sus relaciones causales, cuentan también con capacidad de inferencia abductiva y deductiva. A diferencia de los mapas cognitivos con redes Bayesianas mencionados anteriormente tiene bastante uso en la literatura para distintas áreas (Hossein *et al.*, 2012; Leyva-Vazquez, 2013; Baron *et al.*, 2014).

La Tabla 2.5 despliega la ventaja más importante de cada técnica así como su principal desventaja. Algunas técnicas prometen cosas importantes en un futuro pero son nuevas y poco trabajadas, otras ya tienen tiempo de uso, pero carece de elementos para representar ciertos dominios. Cada técnica aporta elementos para modelar algunos dominios, es consideración del autor decidir cuál es la mejor.

2.9. Redes Bayesianas frente a otras técnicas

Esta sección de la tesis busca determinar el alcance de la RB frente a las otras técnicas. Se seleccionaron 3 criterios para trabajar con la representación del conocimiento en STIs:

- Información cualitativa: Se refiere a la estructura jerárquica útil para representar variables y sus relaciones.

- Información cuantitativa: Incorporada dentro de la estructura jerárquica para lograr inferencia y grados de posesión del conocimiento.
- Teoría solida: Técnicas con bases bien fundamentas y con años de trabajo en diversas áreas y estudios.

La Tabla 2.6 muestra los resultados comparativos de cada técnica. La columna referente a la información cualitativa indica que cualquier técnica aquí presentada es candidata para este tipo de organización de la información. En la siguiente columna (información cuantitativa), los enfoques: redes Bayesianas simples y combinadas con mapas cognitivos, los MCD y los mapas de conocimiento cumplen con este punto. Finalmente las RBs y MCD son una técnica con bases bien fundamentas y con años de trabajo en diversas áreas y estudios. En resumen las RBs y los MCD son las que cumplen con los tres criterios especificados.

Tabla 2.6: Comparativa entre técnicas.

| Enfoque | Inf. Cualitativa | Inf. Cuantitativa | Teoría solida |
|-------------------------------------|------------------|-------------------|---------------|
| Red Bayesiana | X | X | X |
| Mapas cognitivos difusos | X | X | X |
| Mapas cognitivos como red bayesiana | X | X | |
| Mapas de conocimiento | X | X | |
| Grafos | X | | X |
| Mapa de conceptos | X | | |
| Red semántica | X | | X |
| Mapa de memoria | X | | |

Las RBs y MCD comparten bastantes elementos y están diseñados para objetivos muy similares, sin embargo, manejan algunas diferencias entre ellos. Un estudio comparativo elaborado por Cheah (Cheah *et al.*, 2008) posiciona en forma general mejor colocado a las RBs frente a los MCD (ver Tabla 2.7).

Tabla 2.7: Comparativa entre RBs y MCD.

| Enfoque | RB | MCD |
|---------------------------------------|----|-----|
| Usabilidad | | x |
| Manejo de ciclos | | x |
| Encadenamiento hacia adelante y atrás | x | x |
| Eficiente mecanismo de propagación | x | |
| Herramientas de soporte | x | |
| Aplicación en distintas áreas | x | |
| Aplicación de STIs | x | |

Los MCD al ser la combinación de dos enfoques requiere una curva de aprendizaje mayor al tener que conocer la teoría de la lógica difusa y de los mapas cognitivos, además estos no cuentan con variedad de herramientas de soporte, a diferencia de las RBs que tienen muy buen soporte en este aspecto, a pesar de esto y acorde al trabajo de Cheah, este enfoque tiene mayor usabilidad que las RBs, debido a que es más amigable y su representación es más directa, intuitiva y fácil de manejar. Una de las grandes ventajas es el manejo de ciclos carentes en la redes Bayesianas.

La comparativa elaborada por Cheah obtiene puntos a favor de las redes Bayesianas sobre los MCD donde destaca:

- El encadenamiento hacia adelante y hacia atrás.
- Eficientes mecanismos de propagación.
- Suficientes herramientas de soporte e implementación.
- El uso de esta técnica en distintas áreas.
- Emplea una teoría matemática bien formada, definida sobre axiomas básicos.
- La correctitud de sus mecanismos de inferencia es probable.

Chrysafiadi y Virvou (2013) realizaron una revisión de literatura de diez años enfocándose en el modelo del estudiante de los Sistemas Tutores Inteligentes donde concluyen que las redes Bayesianas son una de las técnicas más utilizadas; además Victorio-Meza *et al.* (2014) y Satar (2012) confirman el estudio anterior al concluir el uso de redes Bayesianas como los enfoque más comúnmente usados en sistemas tutores inteligentes.

Por estas razones concluimos que para modelar los aspectos del conocimiento para la toma de decisiones en STIs las RBs se adaptan mejor en forma general que otros enfoques similares.

2.10. Resumen del capítulo

En este capítulo recopila algunos enfoques para representar el conocimiento en los sistemas de enseñanza asistidos por computadoras. Todos los enfoques se basan en conceptos con sus relaciones para formar una estructura jerárquica. Cada uno de los enfoques ataca un problema en particular de acuerdo al dominio mostrando sus ventajas y desventajas para afrontar el problema.

Las técnicas como mapa de conceptos, grafos conceptuales, mapas de memoria y redes semánticas son útiles cuando querernos expresividad semántica para modelar el dominio y no ocupamos algún tipo de razonamiento avanzado.

Las redes Bayesianas simples y con mapas cognitivos, los mapas de conocimiento y los MCD representan la mejor opción cuando no deseamos expresar con mucho detalle el dominio y preferimos alto grado de razonamiento para realizar inferencias que permitan controlar la incertidumbre y otros aspectos cuantitativos.

Finalmente concluimos que los mejores enfoques para trabajar en STIs tomando en cuenta la información cualitativa, cuantitativa y el respaldo de una teoría solida con historia en la comunidad científica son las redes Bayesianas y los Mapas Cognitivos Difusos, aunque la primera muestra una ligera ventaja sobre la segunda.

Capítulo 3

Redes Bayesianas basadas en ontologías

A pesar que las RBs y las ontologías son técnicas relativamente nuevas en el campo de la computación, la construcción de RBs con base en ontologías tiene algunos antecedentes. Estos antecedentes no ha sido en el área educativa, sin embargo, si sientan bases que se pueden retomar. Esta sección analiza los trabajos relacionados a los objetivos de esta investigación. Antes de iniciar con la descripción de la revisión de literatura se explican los fundamentos de las RBs y de las ontologías.

La primer parte de este capitulo se enfoca en las RBs enfoca al área educativa dado que es el dominio de la investigación. La segunda parte describe a las ontologías así como una propuesta para formalizar su modelado. Finalmente se describen los trabajos relacionados a la construcción de RBs a través de ontologías.

3.1. Redes Bayesianas

Las RBs son una poderosa aproximación para la administración de la incertidumbre en la inteligencia artificial basado en la teoría de la probabilidad (Rivas Navarro, 2008)(Santhi *et al.*, 2013). Estas tienen gran relación con la estructura del conocimiento en la mente de las personas, según Rivas (Rivas Navarro, 2008) el conocimiento es una estructura que representa una red de conceptos organizados jerárquicamente, tal cómo se representa una red Bayesiana. Por tanto, se determina cómo una técnica

adecuada para representar el conocimiento de estudiantes por su similitud con la mente humana, sus mecanismos de inferencias probabilística, la documentación y herramientas de soporte, así como otras ventajas definidas en el capítulo anterior.

Una RB está formada por cuatro elementos principales: nodos, escalas de los nodos, arcos y TPCs (Pearl, 1988). Los nodos representan las variables aleatorias del mundo; la escala de nodos representa los posibles estados donde pueden caer los valores; los arcos representan las relaciones causa-efecto entre nodos; finalmente, las TPCs representan el grado de dependencia de cada padre respecto a su hijo. La Figura 3.3 muestra gráficamente estos elementos.

La RB considera una parte cualitativa y otra cuantitativa para ser construida. El aspecto cualitativo identifica las dependencias causales entre variables y crea la estructura final del modelo, es decir, las relaciones entre variables. El aspecto cuantitativo nos lleva a estimar los valores de los parámetros a priori y las TPCs (Xiao-xuan *et al.*, 2007; Misirli y Bener, 2014).

La RB es empleada en problemas que envuelven incertidumbre, es decir, problemas donde no hay un conocimiento completo de los estados del sistema. Sin embargo, podemos darle evidencias del estado del mundo a la RB para actualizar las probabilidades del resto del sistema. Estas redes pueden representar dos tipos de evidencias; una evidencia fuerte o concisa, la cual permite asignar el estado de una variable, es decir, tenemos la certeza absoluta de ese estado; y la evidencia parcial o virtual, la cual se genera en base a la evidencia fuerte para actualizar las probabilidades de los estados de las variables (Rodríguez García y Dolado, 2007). De esta forma, basados en la evidencia fuerte podemos hacer inferencia del estado de otras variables. Las probabilidades previas a la introducción de la evidencia se conocen como probabilidad *apriori*; una vez presentada la evidencia, la nueva evidencia propagada se denomina probabilidad *aposteriori* (Rodríguez García y Dolado, 2007).

En la bio-computación, la RB se ha utilizado para descubrir interacciones funcionales entre los genes, construcción de modelos de enfermedad y la interpretación del análisis de la expresión génica. En el cuidado de la salud, son populares para manipular la incertidumbre en el diagnóstico de enfermedades y para apoyar el análisis de diferentes terapias alternativas (Misirli y Bener, 2014). Otras investigaciones

han aplicado las RBs a áreas tales como: comercio electrónico, minería de datos, reconocimiento de patrones, procesamiento de imágenes y fusión de datos (Liu y Wang, 2007; Kammerdiner, 2009; Santhi *et al.*, 2013). Las redes de este tipo tienen un amplio uso hoy en día para muchos campos en la investigación por su flexibilidad para adaptarse a la resolución de diferentes problemas de diagnóstico y la determinación de patrones.

3.1.1. Evaluación del conocimiento

Millán (Millán y Pérez-De-La-Cruz, 2002) describe un enfoque para modelar al estudiante basado en RBs y tests adaptativos para evaluar su conocimiento. El modelo propuesto por Millán define una serie de nodos organizados jerárquicamente que representan conceptos que el alumno debe comprender; además, existen otros nodos representados por preguntas, estos sirven como evidencia del conocimiento del estudiante y alimentan el resto de los nodos. En el modelo desplegado en la Figura 3.1 se manejan 4 niveles de granularidad: unidades, temas, subtemas y conceptos. Además de otro conjunto de variables para reunir la evidencia representada cómo preguntas.

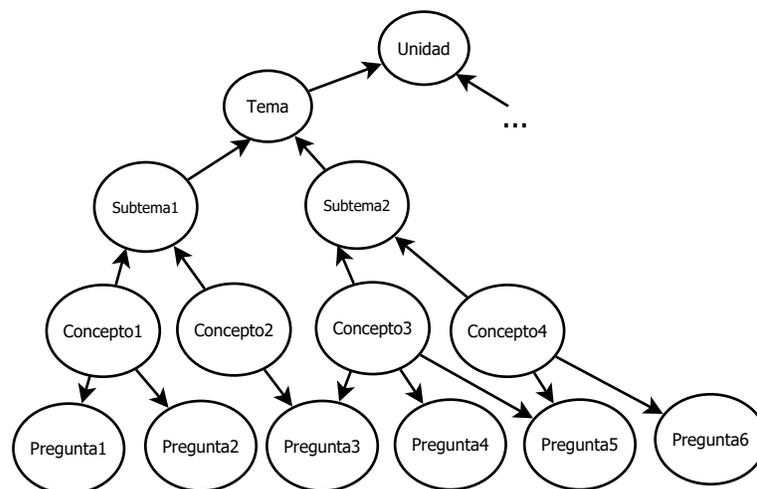


Figura 3.1: Modelo de Millán y Pérez-De-La-Cruz (2002).

El esquema básico de funcionamiento de las RBs para evaluar el conocimiento del alumno está representado en la Figura 3.2. Esta figura muestra dos variables, *Concepto1* y *Concepto2*; y dos

preguntas, *Pregunta1* y *Pregunta2*, con sus respectivas TPCs. La variable *Pregunta1* está relacionada con *Concepto1* y *Concepto2*, de la misma forma como lo está *Pregunta2*, sin embargo, cada pregunta tiene distintos pesos, lo que implica relaciones cuantitativas diferentes.

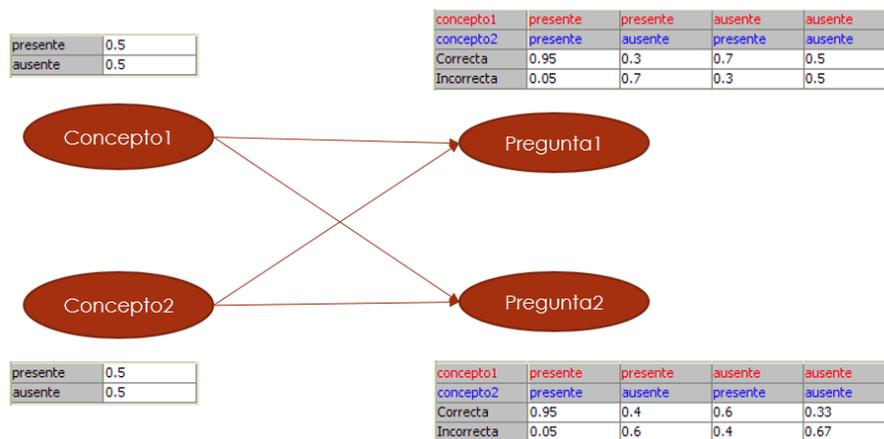


Figura 3.2: Pregunta y su relación con los conceptos.

Analizando las TPCs de las variables de la Figura 3.2, *Pregunta1* tiene 30 y 70% de relación con *Concepto1* y *Concepto2* respectivamente. Si ambos conceptos se dominan, se tendría un 95% de probabilidades de acertar correctamente la respuesta; no se tiene un 100% debido a que se considera que el alumno puede errar su respuesta aun conociéndola (error de dedo, distracción, confusión etc.). Si los conceptos no se dominan, se tiene un 50% de probabilidades de acertar correctamente, esto ocurre porque el alumno puede recurrir al azar para intentar responder; en esta pregunta las opciones de respuesta son dos, por tanto, tiene un 50% de acertar o errar.

En el caso de *Pregunta2* la relación con sus conceptos es 40% para *Concepto1* y 60% para *Concepto2*, si se conocen ambos 95%, si no se conoce ninguno 33%. A diferencia de *Pregunta1*, las opciones de respuesta son tres, por tanto tiene 33% de probabilidad de acertar y un 67% de fallar.

Los valores de los conceptos se encuentran en 50% de probabilidades de conocerlos; esto se le conoce como probabilidad a priori, es decir, antes de que cualquier evento ocurra. Cuando un evento ocurre, como contestar una pregunta, las probabilidades se actualizarán a favor (respuesta correcta) o en contra (respuesta incorrecta), de inicio sólo sabemos que puede o no conocerse el concepto.

Aplicando el teorema de Bayes, se determinan las probabilidades a priori de las preguntas (Pearl, 1988). Los resultados son representados en la Figura 3.3. Hay un 61.25 % de probabilidad de conocer la respuesta de la *Pregunta1* y 57% de probabilidades de conocer la respuesta de la *Pregunta2*.

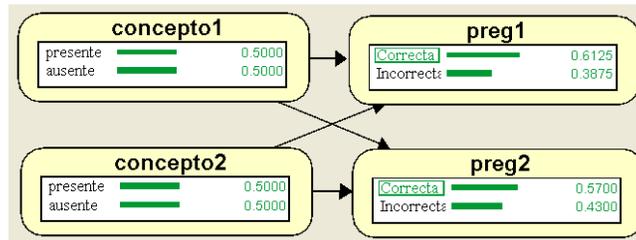


Figura 3.3: Conceptos, preguntas y sus probabilidades a priori.

Si ocurre un evento, es decir, contestar una pregunta; las probabilidades de las variables son actualizadas, a esto se les conoce como probabilidades a posteriori.

La Figura 3.4 muestra la *Pregunta2* contestada correctamente, por tanto, se da por hecho que se conoce la pregunta. La probabilidad de conocer la respuesta de la pregunta se actualiza a 100%, esto activa el algoritmo de propagación para actualizar a los nuevos valores de la RB. *Concepto1* cambio su valor de probabilidad a 59.21 % y *Concepto2* a 67.38 % de probabilidad de conocerse. Los conceptos tienen diferentes porcentajes, aunque dependen de la misma pregunta; esto se debe a los pesos, la *Pregunta2* está relacionada con el concepto1 en un 40 % y con el *Concepto2* en un 60 %. Por tanto, si se contesta correctamente la pregunta (cómo sucedió en el caso anterior), la probabilidad de conocer el concepto aumentará en mayor proporción con los que tengan mayor peso, en los otros conceptos también aumenta pero en una proporción menor.

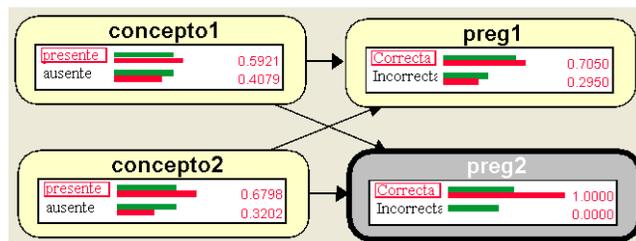


Figura 3.4: Probabilidad a posteriori, pregunta 2 correcta.

Por otra parte, la Figura 3.5 muestra la *Pregunta2* contestada de forma incorrecta, esto propaga

los valores de la siguiente forma: *Concepto1* 37.79% y *Concepto2* 26.16%. En este caso, ocurre lo inverso que en el párrafo anterior, cómo la *Pregunta2* está más influenciada por el *Concepto2*, este disminuye sus probabilidades en mayor proporción, para el *Concepto1* también las disminuye, pero en menor proporción.

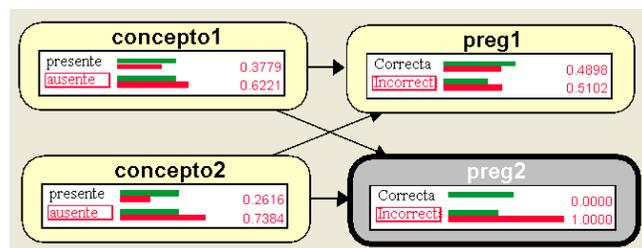


Figura 3.5: Probabilidad a posteriori, pregunta 2 incorrecta.

Si se contestan correctamente dos preguntas relacionadas a los mismos conceptos, las probabilidades de conocerlos aumentarán en proporción a su peso, cómo lo muestra la Figura 3.6.

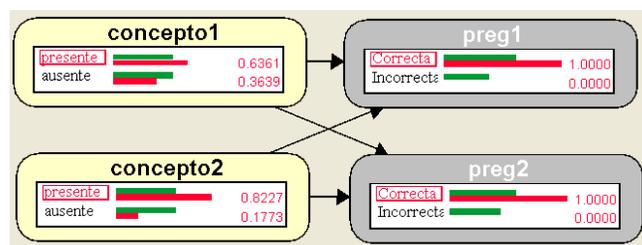


Figura 3.6: Probabilidad a posteriori, pregunta 1 y 2 correctas.

Lo descrito en los párrafos anteriores es la idea principal dentro del modelo de razonamiento basado en RBs. De esta forma, la RB ayuda a inferir el grado de conocimiento del alumno, determinando si el alumno conoce o no algún concepto o pregunta dadas las evidencia en las preguntas relacionadas a esos conceptos.

3.1.2. Construcción manual de una red Bayesiana

Esta sección describe cómo es construida una RB de forma manual, siguiendo la metodología descrita en (Ramírez-Noriega *et al.*, 2015).

3.1.2.1. Metodología

La metodología empleada divide el proceso de construcción de una RB para STIs en seis fases:

1. Definir el dominio de conocimiento: El área de trabajo es definida con base en las necesidades del problema.
2. Desarrollar una escala jerárquica del conocimiento: Esta fase clasifica al conocimiento en diferentes niveles de granularidad.
3. Construcción de la parte cualitativa de la RB: Los nodos de la RB son creados y se establecen dependencias entre estos nodos.
4. Diseño de la parte cuantitativa de la estructura base: Las TPCs son construidas de acuerdo al peso y las relaciones de los nodos.
5. Diseño de actividades para reunir evidencia: Se crea un banco de actividades representados por nodos, se asignan relaciones con los nodos ya definidos.
6. Creación de las TPCs para los nodos de actividad: Esta fase asigna probabilidades a los nodos de evidencia.

Las fases 1, 2, 3 y 5 crean la estructura de red (sección estructural o cualitativa) y las fases 4 y 6 calculan los valores de probabilidad estimados para cada nodo (parte paramétrica o cuantitativa). Las fases 4 y 6 podrían combinarse en una sola, sin embargo, se dividieron en dos fases para mejorar la organización y la claridad. Esta investigación en particular se enfoca en las primeras cuatro fases del modelo de Millán y Pérez-De-La-Cruz (2002).

3.1.2.2. Implementación de la metodología

Con el fin de desarrollar una RB con la metodología descrita, se ha utiliza una sección del libro Proceso Personal de Software (PSP por sus siglas en ingles Personal Software Process) (Humphrey,

1997), este tópico representa un curso impartido a estudiantes con carreras relacionadas con el área de computación. A continuación se resume el desarrollo de la RB:

- Fase 1: Como se mencionó anteriormente se eligió una sección del libro de PSP, esta metodología ha sido utilizada entre los ingenieros de software para mejorar la calidad en los procesos de desarrollo.
- Fase 2: La Figura 3.7 muestra los temas desglosados del capítulo 1 del libro de PSP (Humphrey, 1997). Los nombres de los temas originales están numerados; además, se definieron nombres más cortos para los temas, indicados por viñetas negras. El nivel más profundo, viñetas blancas, corresponde a los conceptos encontrados en este tema; si no hay conceptos en el tema, consideramos las viñetas negras como conceptos.

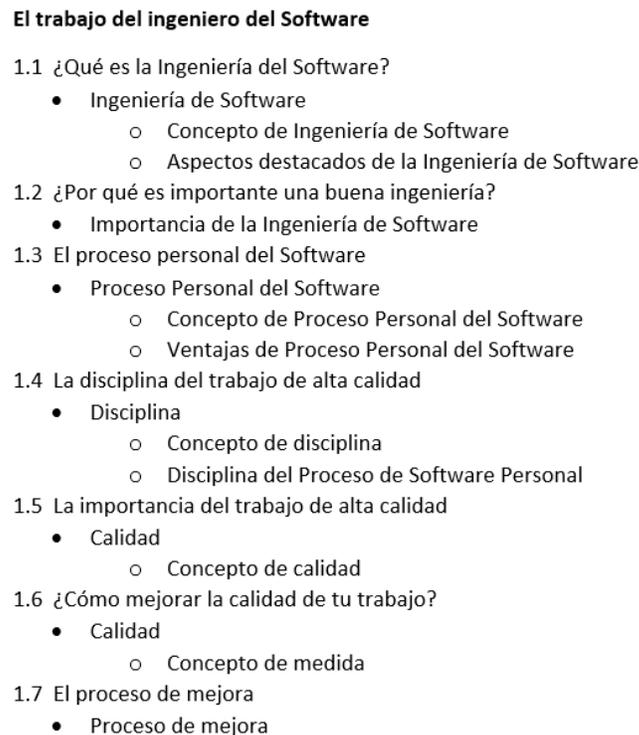


Figura 3.7: Conceptos desglosados de PSP (Humphrey, 1997).

- Fase 3: Con base en la escala jerárquica de la fase previa, se construyó la RB, los temas son resumidos con palabras claves para representar el nombre de las variables. En la Figura 3.8 los

los nodos grises representan conceptos, los temas 1.5 y 1.6 de la Figura 3.7 se fusionaron, de acuerdo a la visión del experto, para contener el mismo nodo *Calidad*. La Figura 3.9 representa a los nodos agrupados por temas o conceptos, un nodo puede ser subtema de otro tema si el experto encuentra la relación de dependencia requerida. El nodo *importancia_SI* se convirtió en parte del nodo *Ingeniería_Software*, *Disciplina* se convirtió en padre de *PSP* y *Proceso_mejora* se convirtió en parte de *Calidad*.

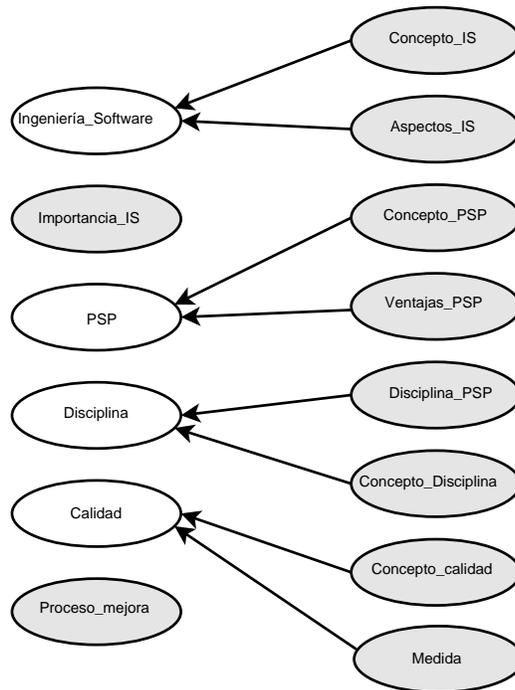


Figura 3.8: Conceptos de PSP sin agrupar.

- Fase 4: El conocimiento y la percepción del experto es empleada para definir los valores de los nodos y construir las TPCs. El ejemplo de la Figura 3.10 muestra las TPCs para los nodos *Concepto_IS*, *Ingeniería_Software* y *Capitulo1*.

Es fácil calcular las TPCs de los conceptos que son sólo padres, porque no tienen influencia de ningún otro nodo, el concepto *Concepto_IS* es un ejemplo de este caso. Suponemos que no se tienen referencias previas de dicho nodo que ayuden a determinar su valor, por lo tanto, se define un 50% de probabilidad para conocer el nodo, como se muestra en la tabla A de la Figura 3.10.

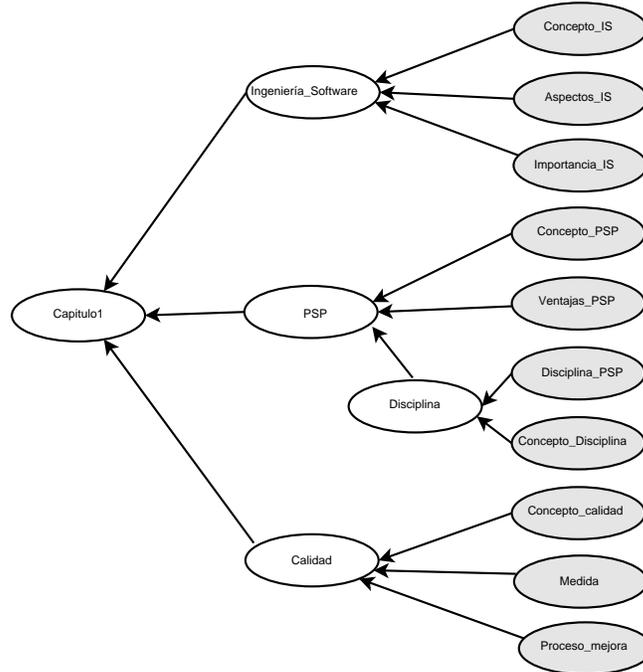


Figura 3.9: Conceptos de PSP agrupados.

El nodo del tema *Ingeniería_Software* tiene probabilidades representadas por la tabla B de la Figura 3.10, la tabla está basada en los valores asignados por el experto a cada uno de sus padres. En este caso, el nodo *Ingeniería_Software* tiene tres padres (*Concepto_IS*, *Aspectos_IS*, e *Importancia_IS*), cada uno de los cuales tiene la misma importancia (0.33). Los valores del nodo hijo en la columna *T* de la tabla B se definen añadiendo el valor de los padres cuando se considera poseedor del conocimiento, es decir, si es cierto. El nodo *Capitulo1* representa la unidad de estudio, para este nodo la TPC se construye de la misma manera que se describió en el párrafo anterior, teniendo en cuenta a sus padres (*Ingeniería_Software*, *PSP*, y *Calidad*). Aquí el valor de uno de los padres es diferente, por tanto, la tabla de probabilidad será diferente a la anterior.

- Fase 5: La Figura 3.11 muestra 11 preguntas que se desarrollaron con sus respectivos conceptos relacionados. Esta fase puede ser la parte que implica más tiempo del procedimiento, por la formulación de las preguntas, sus respuestas y los conceptos relacionados; aunque, con

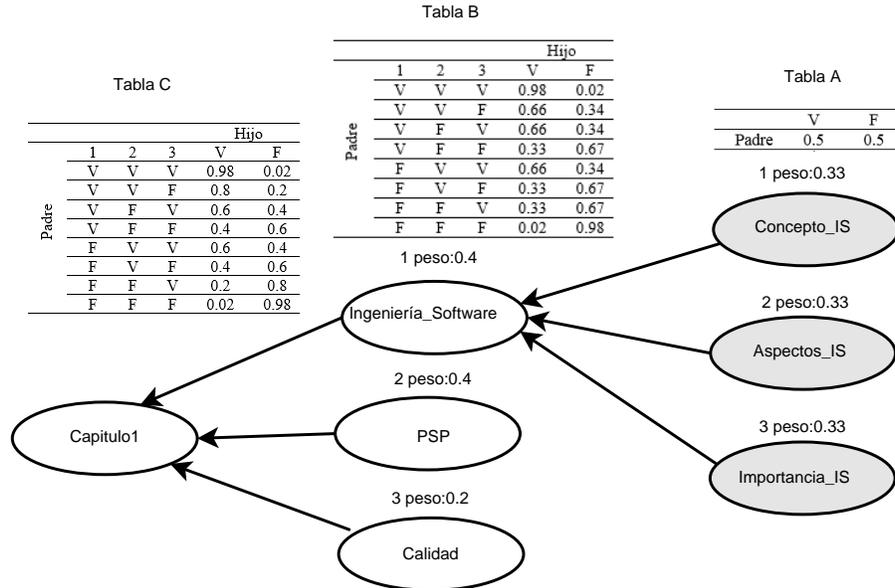


Figura 3.10: Tablas de probabilidad condicional para las variables.

experiencia en la enseñanza del tema, probablemente ya se tengan preguntas o actividades listas para ser empleadas.

Para ejemplificar, seleccionamos la pregunta dos ($P2$): ¿Cuáles son los aspectos que mejoran la eficacia del ingeniero de software? Esta pregunta se relaciona con el concepto de ingeniería de software y con sus aspectos, como lo demuestran las relaciones de la Figura 3.11, según la perspectiva del experto.

- Fase 6: La Figura 3.12 muestra la tabla de probabilidad de la pregunta $P2$, esta pregunta está relacionada con los conceptos *Concepto_IS* y *Aspectos_IS*, con un valor de 0.6 y 0.4, respectivamente. Cuando se conocen los dos conceptos, las probabilidad de poseer el conocimiento no es el 100% porque el alumno puede errar su respuesta incluso sabiéndola (tipografía, distracción, confusión, entre otros). Cuando los conceptos son desconocidos, calculamos los valores según el número de respuestas de la pregunta. En este caso, consideramos cuatro respuestas, de las cuales una es correcta, dando un 25% de probabilidad de adivinar sin poseer conocimiento.

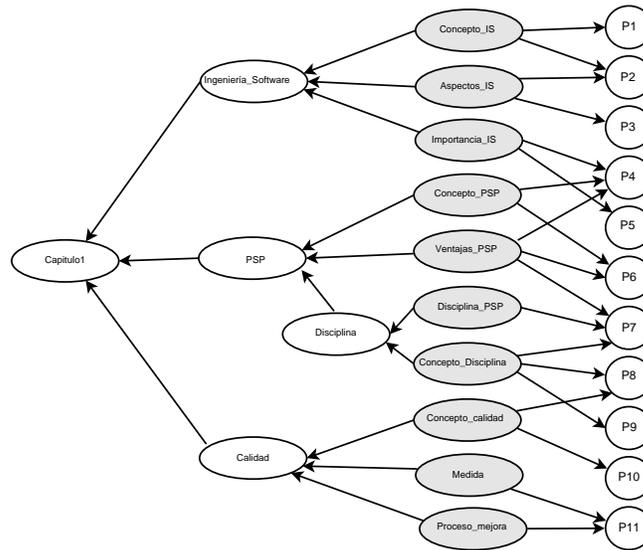


Figura 3.11: Red Bayesiana con preguntas.

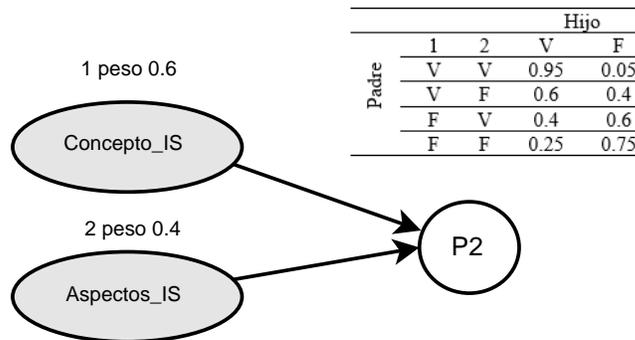


Figura 3.12: Tabla de probabilidad condicional para la pregunta P2.

3.1.2.3. Red Bayesiana para el Proceso Software Personal

En la sección anterior se generó una red que puede ser utilizada como una plantilla para expresar el conocimiento de PSP, la cual contiene una organización estructurada del conocimiento para el primer capítulo de ese tema. La RB de PSP está representada en la Figura 3.13, esta red fue desarrollada con la herramienta de software Genie ¹. La RB, en primera instancia, no contiene evidencia; es decir, contiene los valores a priori.

Para determinar si un concepto es conocido usando esta RB, es necesario definir los valores de

¹<https://download.bayesfusion.com/files.html?category=Academia>

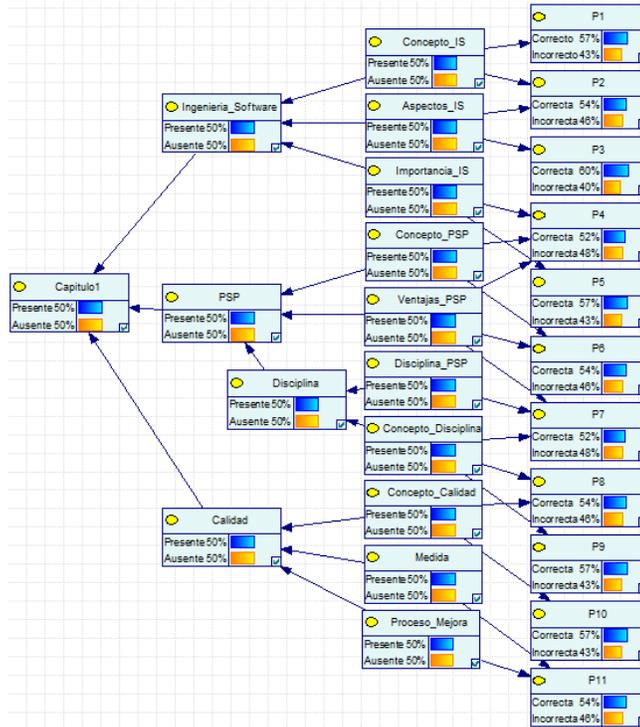


Figura 3.13: Simulación de la red Bayesiana sin evidencia.

probabilidad. Para este ejemplo, el valor de probabilidad por encima de 70% significa que un nodo es conocido, menor de 30% significa que el nodo es desconocido, dejando un rango de 30% a 70% para nodos que todavía son indeterminados. La última clasificación significa que necesitamos más pruebas para determinar los conceptos como conocidos o desconocidos.

Considerando que el estudiante responde a cuatro preguntas: P_2 , P_4 , P_7 y P_{11} ; respondió P_2 , P_4 y P_7 correctamente y P_{11} incorrectamente, esta información se muestra en la Figura 3.14. Con esta evidencia, los nodos padre (conceptos) aumentarán sus posibilidades de conocerse cuando las preguntas son contestadas correctamente. Los nodos *Concepto_IS* y *Ventajas_PSP* inicialmente tenían un 50% de posibilidades de ser conocidos (ver Figura 3.13), ahora, sus probabilidades se actualizan a 72% y 76% (ver Figura 3.14), respectivamente, debido a la evidencia dada cuando la pregunta P_2 , P_4 y P_7 se respondió correctamente. Lo anterior permite determinar un mayor grado de posesión del conocimiento para los nodos con probabilidades más altas.

Con base en la evidencia de las tres preguntas contestadas, se puede analizar el estado de las

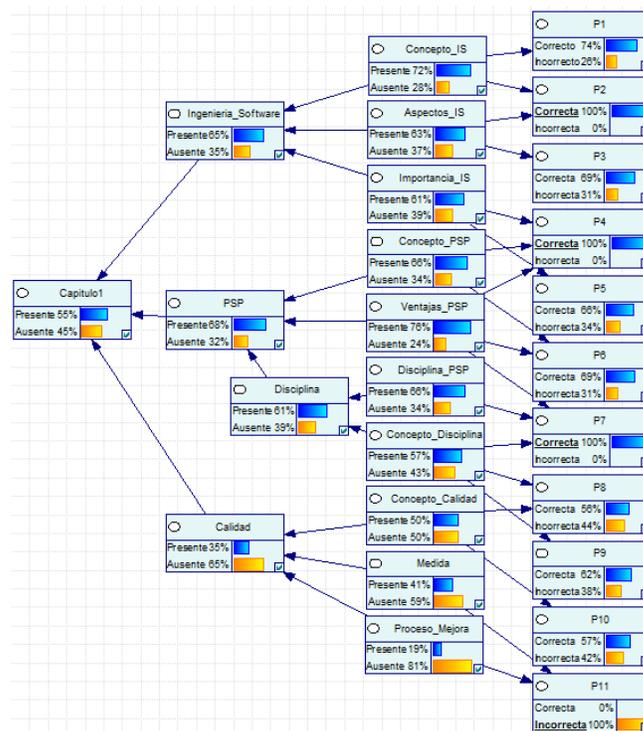


Figura 3.14: Simulación de la red Bayesiana con evidencia.

preguntas restantes (no contestadas aun). Con el objetivo de centrarse en los conceptos que la RB aún no han clasificado como conocidos (probabilidad $< 70\%$), se pueden descartar las preguntas con valores superiores al 70% . De modo que, la pregunta $P1$ puede ser descartada por su valor (74%). Así, se puede elegir mostrar al alumno sólo preguntas con una probabilidad menor al 30% de ser conocidas. También podemos hacer inferencias con los conceptos de la red, no sólo con las preguntas.

La misma situación ocurre cuando las preguntas se responden incorrectamente, pero en orden inverso. Por ejemplo, $P11$ fue contestada incorrectamente; por lo tanto, los valores con baja probabilidad se propagan, lo que significa que se tienen probabilidades bajas de conocer algunas preguntas o conceptos.

Este modelo no considera el promedio general para resumir el conocimiento del estudiante, el modelo prefiere tener un enfoque sobre conceptos clasificándolos en tres niveles: Conocido, desconocido e indeterminado. Por lo tanto, es necesario centrarse en conceptos que el estudiante no ha aprendido (desconocido) para trabajar con ellos. Si el nivel desconocido no tiene conceptos elegibles, se trabaja

con el siguiente nivel (indeterminado), teniendo en cuenta conceptos con valores de probabilidad más bajos. Así, la RB de esta sección permite inferir el conocimiento del estudiante para identificar debilidades y para proponer temas de instrucción.

3.1.3. Formalización

A continuación son descritas algunas definiciones y conceptos importantes de las RBs (Neapolitan, 1990):

Definición 3.1. Sea $N = \{n_1, \dots, n_{|N|}\}$ el conjunto de variables aleatorias que representan un problema. Una variable o nodo n puede tener múltiples estados reuniendo dos propiedades importantes: (1) Ser mutuamente excluyentes, es decir, un nodo sólo puede encontrarse en uno de sus estados en un momento dado. (2) Ser un conjunto exhaustivo, es decir, un nodo no puede tener ningún valor fuera de ese conjunto.

Definición 3.2. Un arco es una conexión entre dos nodos representando una dependencia entre ellos, el arco también puede ser llamado relación o enlace. Un arco es definido por el par ordenado de nodos (n_i, n_j) , donde $n_i, n_j \in N$.

Definición 3.3. El nodo n_i es padre del nodo n_j , si hay un arco (n_i, n_j) entre los nodos. El nodo n_j es hijo de n_i , si hay un arco (n_i, n_j) entre los dos nodos.

Definición 3.4. Sea $BN_{tupla} = \{N, DB, P\}$ una tupla de 3 elementos para representar una RB. El elemento N representa a las variables, este conjunto es descrito en la Definición 3.1.

Definición 3.5. Sea $DB = \{db_1, \dots, db_{|DB|}\}$ el conjunto de pares ordenados que representan la parte cualitativa de una RB. Un elemento en DB es representado como $db = (m, n)$, donde $m, n \in N$ y $m \neq n$. El primer elemento del par ordenado representa el nodo padre (inicio de la relación) y el segundo elemento representa el nodo hijo (fin de la relación).

Definición 3.6. Sea P la variable que representa la parte cuantitativa de una RB, esta variable representa las TPCs. Estas tablas especifican la distribución de la probabilidad condicional del descendiente, el nodo hijo; para todas las combinaciones de los estados de sus predecesores, los nodos padre (Laitila y Virtanen, 2016). La incertidumbre de la relación causal es representada por la TPC $P(n_q|\pi_q)$ asociada a cada nodo n_q , donde π_q es el conjunto de padres de N . Bajo una suposición de independencia condicional, la estructura gráfica de la RB permite una representación no ambigua de interdependencia entre variables. Esto guía a una de las más importantes características de las RBs: La distribución de probabilidad conjunta de $N = (n_1, \dots, n_{|N|})$ puede ser factorizada como un producto de las TPCs en la red $P(n_1, \dots, n_{|N|}) = \prod_{q=1}^{|N|} P(n_q|\pi_q)$ (Ding y Peng, 2004).

3.1.4. Descomposición de una red Bayesiana

Este caso descompone una RB en sus partes, con el objetivo de construir una RB desde cero basado en sus elementos mínimos. La Figura 3.15 representa uno de los casos comunes cuando se explica el funcionamiento de una RB. Este ejemplo modela las probabilidades de que la hierba este húmeda tomando en cuenta dos factores: lluvia y rociador. La variable lluvia está relacionada con la hierba húmeda por lógica común, pero también con el rociador por su función de regar la hierba. Así, dadas las estadísticas o datos obtenidos por un experto podemos construir las TPCs.

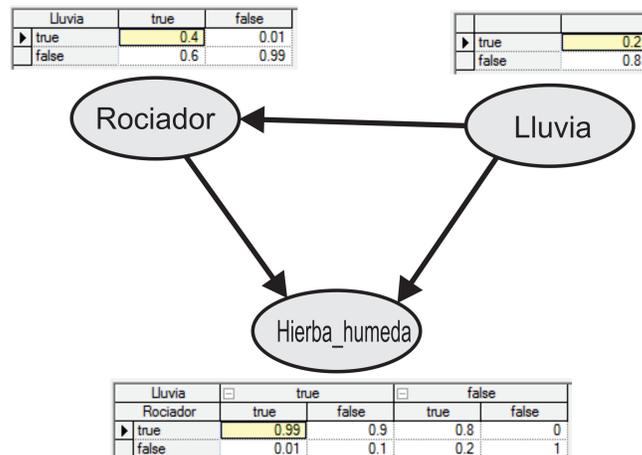


Figura 3.15: Ejemplo de red Bayesiana.

Los elementos principales de la RB: variables, relaciones y TPCs, son mostrados en la Figura 3.15. Basados en el modelo de 3 tuplas descrito de la Definición 3.4, podemos obtener: el conjunto de variables $N = \{Rociador, Lluvia, Hierba_humeda\}$, las relaciones $D = \{(Lluvia, Rociador), (Rociador, Hierba_humeda), (Lluvia, Hierba_humeda)\}$ y sus TPCs para cada variable $P = (0.4, 0.6, 0.01, 0.99)(0.2, 0.8)(0.99, 0.01, 0.9, 0.1, 0.8, 0.2, 0, 1)$. Con esta descomposición podemos enfocarnos en construir una RB considerando sus elementos particulares para construir el sistema general.

3.2. Ontologías

En la ciencia de la filosofía, una ontología es empleada para estudiar las cosas que existen en el universo y cómo estas entidades están compuestas (Abuhassam y AlMashaykhi, 2012). En ciencias de la computación, Gruber (1995) estipuló que una ontología es una especificación explícita de una conceptualización. La palabra conceptualización se refiere a un modelo abstracto de algún fenómeno que tiene identificado sus aspectos relevantes. La palabra explícita significa que todos los conceptos usados y las restricciones en su uso son explícitamente definidas (Ben Messaoud *et al.*, 2013).

Las ontologías han sido fuente de estudio y desarrollo en años recientes por las siguientes razones (Noy y McGuinness, 2001):

- Comparten el entendimiento común de la estructura de la información entre las personas o agentes de software.
- Permiten reutilizar un dominio de conocimiento.
- Realizan suposiciones explícitas de un dominio.
- Separan el dominio de conocimiento del dominio operacional.
- Analizan un dominio de conocimiento.

En el caso más simple, una ontología describe una jerarquía de conceptos (clases) enlazadas por relaciones taxonómicas (es-un, parte-de). En los casos más sofisticados, una ontología define un dominio de clases, propiedades o atributos para cada clase, instancias de clase o individuos, además de relaciones que mantienen unidas las clases; también es posible añadir algunos axiomas lógicos para restringir la interpretación de conceptos y expresar relaciones complejas entre conceptos (Ben Messaoud *et al.*, 2013).

Los criterios de diseño de ontologías son descritos a continuación Gruber (1995):

1. Claridad: Las definiciones deben ser objetivas e independientes del contexto social.
2. Coherencia: Los axiomas que define deben ser lógicamente consistentes.
3. Extensibilidad: La ontología debe prestarse para definir nuevos términos basada en el vocabulario existente.
4. Sesgo de codificación mínima: Se produce cuando una serie de opciones de representación se realizó únicamente a conveniencia de la notación o ejecución.
5. Compromiso ontológico mínimo: La ontología debe hacer la menor cantidad posible de afirmaciones sobre el mundo que está modelando.

Existen distintas herramientas para desarrollar ontologías, pero una muy utilizada y referenciada en la literatura es Protégé². Este es un editor de ontologías desarrollado por la Universidad de Stanford que puede ser ejecutado en una variedad de plataformas (Khan y Kumar, 2014). La Figura 3.16 representa una ontología en Protégé donde se organiza la información de maestros, estudiantes, grupos y materias. Esta estructura representa clases (Persona, Profesor, Estudiante, Módulo), relaciones entre objetos (enseña, estudia, subclass) y atributos (nombre, apellido, IDProfesor, IDEstudiante).

²<http://protege.stanford.edu/>

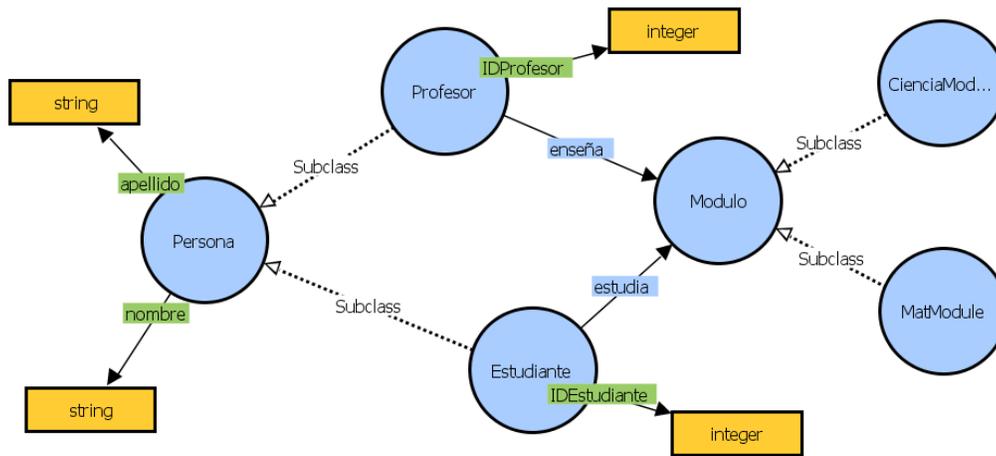


Figura 3.16: Ontología en Protégé.

3.2.1. Componentes principales

Los componentes de una ontología varían de acuerdo al dominio de interés y a las necesidades de los desarrolladores. Por lo general, entre los componentes de una ontología se encuentran los siguientes (Gruber, 1995):

- **Clases:** Son la base de la descripción del conocimiento en las ontologías, ya que expresan los conceptos (ideas básicas que se intentan formalizar) del dominio. Las clases usualmente se organizan en taxonomías pudiéndose aplicar mecanismos de herencia para su manejo.
- **Relaciones:** Representan las interacciones entre los conceptos del dominio. Las ontologías por lo general contienen relaciones binarias; el primer argumento de la relación se conoce como el dominio y el segundo como el rango.
- **Funciones:** Son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de una ontología.
- **Instancias:** Representan objetos determinados de un concepto.
- **Axiomas:** Se usan para modelar sentencias que son siempre ciertas. Los axiomas permiten, junto con la herencia de conceptos, inferir conocimiento que no esté indicado explícitamente en la

taxonomía de conceptos. En especial, los axiomas son usados para verificar la consistencia de la ontología; también son utilizados para inferir nuevo conocimiento.

3.2.2. Lenguaje de Ontología Web

El Lenguaje de Ontología Web (OWL por sus siglas en inglés *Ontology Web Language*) está destinado para ser usado cuando la información contenida en un documento necesita ser procesada por aplicaciones, opuesto a las situaciones donde el contenido necesita ser procesado sólo por humanos. EL Consorcio Mundial de la Web (W3C por sus siglas en inglés *World Wide Web Consortium*) da soporte a OWL, este lenguaje es una extensión del RDF (*Resource Description Framework*) y RDFS (*RDF Schema*). OWL convierte sus representaciones a lógica descriptiva, función que no puede ser realizada por RDF y RDFS; además, este lenguaje define una capa de conceptos que provee mejores características de propósitos de recuperación (Khan y Kumar, 2014). OWL maneja tres sublenguajes que pueden ser empleados según las necesidades de la problemática:

- OWL Lite: Está diseñado para aquellos usuarios que necesitan principalmente una clasificación jerárquica y restricciones simples.
- 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).
- OWL Full: está dirigido a usuarios que quieren máxima expresividad y libertad sintáctica de RDF sin garantías computacionales.

Existen distintos tipos de sintaxis para OWL, RDF/XML y OWL/XML son los estándares de la W3C; sin embargo, existen otras sintaxis que tratan de facilitar el entendimiento sobre el lenguaje destacando la sintaxis Manchester (Safia y Aicha, 2014).

3.2.3. Metodologías de desarrollo

Una metodología de desarrollo de ontologías cumple con un conjunto de principios, procesos, prácticas, métodos y actividades usadas para diseñar, construir, evaluar y desplegar ontologías (Gasevic *et al.*, 2009). Hay distintas metodologías a seguir tales como: Methontology (Fernández-López *et al.*, 1997), On-To-Knowledge (Sure *et al.*, 2004), DILIGENT (Pinto *et al.*, 2004), Ontology Development 101 (Noy y McGuinness, 2001), entre otras. No existe un mejor método para construir una ontología, sin embargo, en términos prácticos y concisos, desarrollar una ontología incluye (Noy y McGuinness, 2001):

- Definir clases en la ontología.
- Organizar las clases en una estructura taxonómica.
- Definir propiedades y describir valores permitidos para esas propiedades.
- Llenar los valores de las propiedades para construir instancias.

3.2.4. Formalización

A continuación se muestran algunas definiciones tomadas de (Ramírez-Noriega *et al.*, 2017c) para representar una ontología:

Definición 3.7. Sea $ont = \{C, B, A, D, T, R\}$ el conjunto de elementos que forman una ontología, cada elemento tiene una función especial para representar un componente de una ontología.

Definición 3.8. Definición 8.- Sea $C = \{c_1, \dots, c_{|C|}\}$ el conjunto de clases que componen el dominio de una ontología.

Definición 3.9. Sea $B = \{b_1, \dots, b_{|B|}\}$ el conjuntos de propiedades o atributos que pueden ser agregadas a las clases de una ontología.

Definición 3.10. Sea $AD = \{ad_1, \dots, ad_{|AD|}\}$ el conjunto que relaciona una propiedad con un tipo de dato. Un elemento de AD es representado como $ad = (a, d)$, donde $a \in A$ y $d \in D$. El conjunto A representa las propiedades de datos y el conjunto D representa los tipos de datos en la ontología.

Definición 3.11. Sea $T = \{t_1, \dots, t_{|T|}\}$ el conjunto que representa la estructura de una ontología, este conjunto representa las clases y sus relaciones. Un elemento de T es representado como una tupla de 3 elementos $t = (c_1, b, c_2)$, donde $c_1, c_2 \in C$, y $b \in B$. El primer elemento de la tupla representa el dominio, esto es, el inicio de la relación; el segundo es el tipo de relación entre las clases; finalmente, el tercero representa el rango, esto es, el fin de la relación.

Definición 3.12. Sea $R = \{l_1, \dots, l_{|R|}\}$ el conjunto que establece las propiedades de dato, un elemento de R representa una tupa de dos elementos $l = (c, V)$, donde $c \in C$ y $V \subseteq AD$, el primer elemento representa las clases y el segundo los atributos.

Estas definiciones pueden representar clases, relaciones y atributos. Las ontologías podrían tener instancias, sin embargo, esta información no es necesaria para los objetivos de esta investigación.

3.2.5. Descomposición de una ontología

La Figura 3.17 representa una ontología que modela brevemente dos tipos de árboles, el dominio de conocimiento está representado por el nodo principal que en este caso es *rbol*. Esta clase principal tiene dos subclases *frutal* y *forestal*, especificando que un árbol puede ser de varios tipos. Cada clase tiene atributos que la caracterizan, representados con una relación y un rectángulo en un extremo que representa el tipo de dato.

La representación gráfica de la ontología en la Figura 3.17 puede ser representada con el modelo expuesto en la sección previa (Definición 3.7):

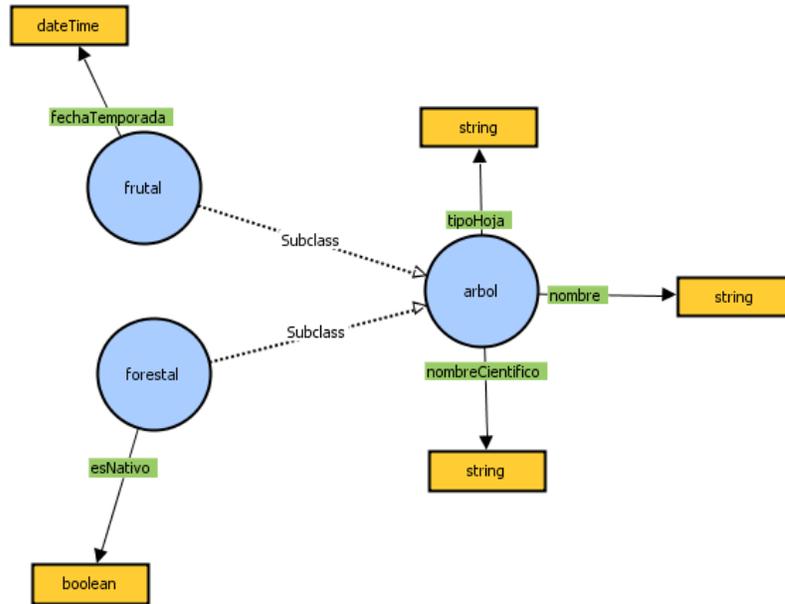


Figura 3.17: Ejemplo de ontología sencilla.

$$C = \{arbol, forestal, frutal\}$$

$$V = \{subclass\}$$

$$A = \{nombre, tipoHoja, nombreCientifico, esNativo, fechaTemporada\}$$

$$D = \{string, boolean, dateTime\}$$

$$CA = \{(nombre, string), (nombreCientifico, string), (tipoHoja, string), \\ (esNativo, boolean), (fechaTemporada, dateTime)\}$$

$$T = \{(forestal, Subclass, arbol), (frutal, Subclass, arbol)\}$$

$$R = \{(arbol, \{ca_1, ca_2, ca_3\}), \\ (frutal, \{ca_5\}), (forestal, \{ca_4\})\}$$

La descomposición anterior permite comprobar que el modelo expuesto puede contener los ele-

mentos de una ontología. Con las unidades mínimas de la estructura origen (ontología) y destino (RB), podemos establecer una conversión de forma menos compleja.

3.3. Trabajos relacionados

Existen esfuerzos para generar RBs basadas en ontologías, estos han sido aplicados tanto al campo médico como a otras áreas, sin embargo, el dominio educativo del proyecto de esta investigación no ha sido cubierto. Los artículos analizados son descritos en la Tabla 3.1, abarcan desde el años 2005 hasta el 2016.

Tabla 3.1: Artículos relacionados.

| Autor (año) | Nombre de artículo |
|-------------------------------|--|
| Yang y Calmet (2005) | OntoBayes: An Ontology-Driven Uncertainty Model |
| Devitt <i>et al.</i> (2006) | Constructing Bayesian Networks Automatically Using Ontologies |
| Ding <i>et al.</i> (2006) | BayesOWL: Uncertainty modeling in semantic web ontologies |
| Fenz <i>et al.</i> (2009) | Ontology-based generation of Bayesian network. |
| Leontidis y Halatsis (2009) | Supporting Learner's needs with ontology-based Bayesian network |
| Andrea y Franco (2009) | Extending ontology queries with Bayesian network reasoning |
| Bucci <i>et al.</i> (2011) | Ontologies and Bayesian Networks in Medical Diagnosis. |
| Ishak <i>et al.</i> (2011) | Ontology-based generation of object oriented bayesian networks |
| Pilato <i>et al.</i> (2012) | Integration of ontologies and Bayesian networks for maritime situation awareness |
| Yan y Wei (2013) | Development of a novel asset management system for power transformers based on Ontology |
| Pshenichny (2014) | Knowledge engineering in volcanology: practical claims and general approach |
| Ferreira <i>et al.</i> (2016) | An Automatic and Dynamic Student Modeling Approach for Adaptive and Intelligent Educational Systems using Ontologies and Bayesian Networks |

La Tabla 3.2 muestra una comparativa de los artículos encontrados. La tabla muestra los datos

Tabla 3.2: Comparativa entre estudios relacionados a ontologías y redes Bayesianas.

| Autores (Año) | V | E | R | TPC | S | EOWL | MD | MA | SE | AU | ATPC |
|-------------------------------|---|---|---|-----|---|------|----|----|----|----|--|
| Yang y Calmet (2005) | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | - | - | ✓ | Información se agrega en ontología |
| Devitt <i>et al.</i> (2006) | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | ✓ | - | ✓ | - | Manualmente |
| Ding <i>et al.</i> (2006) | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | - | - | ✓ | D-IPFP |
| Fenz <i>et al.</i> (2009) | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | ✓ | - | ✓ | - | Manualmente |
| Leontidis y Halatsis (2009) | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | - | - | ✓ | Información se agrega en ontología |
| Andrea y Franco (2009) | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ✓ | - | - | ✓ | Relaciones abox, tbox. A través de frecuencia de datos |
| Bucci <i>et al.</i> (2011) | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | - | - | ✓ | Estructura predefinida |
| Ishak <i>et al.</i> (2011) | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | ✓ | - | ✓ | - | Manualmente |
| Pilato <i>et al.</i> (2012) | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ | ✓ | - | - | Manualmente |
| Yan y Wei (2013) | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | - | - | ✓ | E-IPFP |
| Pshenichny (2014) | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ | ✓ | - | - | Manualmente |
| Ferreira <i>et al.</i> (2016) | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ | ✓ | - | - | Manualmente |

autores y su año de publicación. Las siguientes columnas muestran si el método de conversión identifica: variables (*V*), escala de las variables (*E*), relaciones de las variables (*R*) y TPCs (*TPC*). Después la columna *S* representa si se toma en cuenta la semántica de las palabras al construir las TPCs, seguido se despliega la columna *EOWL* para identificar si el estándar OWL es extendido. Algunos métodos pueden devolver una versión reducida del dominio, para eso se utilizó la columna *MD* para aquellos estudios que mantienen el dominio intacto. Las siguientes tres columnas representan si el método es manual (*MA*), semiautomático (*SE*) o automático (*AU*). Por último, una breve descripción del algoritmo o método para construir las tablas de probabilidad condicional.

La literatura clasifica los esfuerzos de conversión de RB a ontologías en tres tipos:

- Métodos manuales: Estos métodos se enfocan en la construcción por medio del humano de los procesos de creación de las RBs mediante un análisis del conocimiento formalizado en ontologías

(Pilato *et al.*, 2012; Pshenichny, 2014).

- Métodos semiautomáticos: Estos enfoques generan automáticamente una parte de la RB a partir de una ontología, pero requieren la interacción del usuario para ayudar en las partes restantes (Fenz *et al.*, 2009; Devitt *et al.*, 2006).
- Métodos automáticos: Estos métodos buscan obtener una RB sin la intervención humana valiéndose de distintos enfoques (Yang *et al.*, 2005; Bucci *et al.*, 2011)

Las investigaciones más destacadas organizadas dentro de estas categorías son las siguientes.

Pilato *et al.* (2012) y Pshenichny (2014) utilizaron métodos manuales para resolver los problemas en sus dominios. El primer estudio se enfocó en el área del cuidado de situaciones marítimas y el segundo en el campo de la vulcanología. Estos estudios inicialmente generaron una ontología como representación del conocimiento; luego, convirtieron manualmente la ontología en una RB; finalmente, hicieron inferencias probabilísticas y ontológicas en el dominio para resolver sus problemas.

Los métodos semiautomáticos generan automáticamente una parte de la RB, pero, necesitan información adicional para completar su representación. Las investigaciones de Devitt *et al.* (2006) y Fenz *et al.* (2009) no generan completamente las TPCs de la RB y necesitan intervención manual para modelar esta parte. Ishak *et al.* (2011) establecieron un método para convertir ontologías a RB orientadas a objetos; sin embargo, no construyeron las TPCs, el estudio consideró sólo la parte cualitativa. Estos tres estudios tienen en común que construyen las variables, relaciones y estados de las variables, sin embargo, no construyen las TPCs.

Los modelos automáticos no asumen ninguna interacción con los expertos para modelar las RBs. Algunos modelos pueden convertir automáticamente una ontología a una RB a costa de extender el estándar OWL, sin embargo, un experto debe crear manualmente la ontología y los valores probabilísticos; además, las ontologías que siguen el estándar OWL tendrían que ser descartadas debido a la incompatibilidad. Ejemplos de estos tipos de modelos son: Yang y Calmet (2005) propusieron una extensión a las ontologías para apoyar la incertidumbre, Leontidis y Halatsis (2009) utilizaron

ontologías y RBs para almacenar e inferir el conocimiento afectivo del alumno, Yan y Wei (2013) trabajaron en el diagnóstico de fallas en transformadores de potencia y Ding *et al.* (2006) desarrollaron BayesOWL, este último estudio propuso una modificación al algoritmo IPFP (iterativo proportional fitting procedure) algoritmo para construir las TPCs. Todas estas obras ampliaron el estándar OWL para resolver sus problemas de conversión.

Andrea y Franco (2009) propusieron construir la RB automáticamente sin alterar el estándar OWL. Este trabajo argumentó que la información para construir la RB se encuentra en las mismas ontologías. La información semántica de conceptos puede proporcionar la dependencia de probabilidad condicional entre nodos, y la frecuencia de datos de instancias puede determinar la distribución de probabilidad. Esto funciona bien con ontologías con muchas instancias, sin embargo, pocas ontologías se distribuyen con instancias. Además, este método sería inexacto en el ámbito educativo.

Bucci *et al.* (2011) utilizaron un enfoque para convertir ontologías a RBs en el diagnóstico médico siguiendo el objetivo de mantener el estándar OWL. Este estudio manejó una estructura predefinida de dos niveles para disminuir la complejidad de la conversión, sin embargo, este modelo modifica el dominio original dando como resultado una versión reducida de la estructura, por lo tanto, la información valiosa podría perderse en la conversión.

Esta tesis propone elaborar un método automático en cada fase de desarrollo para automatizar la construcción de una RB considerando una ontología. Este proyecto busca mantener el estándar OWL en la reutilización de ontologías; mantener el dominio original y desarrollar TPCs similares a las desarrolladas por expertos en el área educativa. Estos trabajos relacionados no cumplen con todos estos aspectos en conjunto. Nuestro trabajo impacta esfuerzos tal como (Ferreira *et al.*, 2016). Estos investigadores utilizaron una ontología y una RB para evaluar el conocimiento en un sistema educativo adaptativo, sin embargo, la construcción de ambas estructuras fue manual.

3.4. Resumen del capítulo

Esta sección explicó a detalle la forma de trabajar de las RBs con enfoque a la educación, además, se definieron algunos aspectos formales de la teoría. Estos aspectos son importantes porque posteriormente serán referenciados para definir el modelo propuesto. Por otra parte, esta sección también describió a las ontologías, que caen dentro de las redes semánticas explicadas en el capítulo previo, a la vez se definió una propuesta propia de la investigación para modelar una ontología. Por último, se analizaron los trabajos relacionados de construcción de RBs a través de ontologías.

El siguiente capítulo explica algunos aspectos teóricos referente al método propuesto por este proyecto.

Capítulo 4

Relación entre conceptos

Este capítulo tiene como objetivo describir aquella teoría, métodos y procesos que permitirán establecer la relación numérica entre conceptos para construir las TPCs de una RB. Esta investigación considera la información textual como el tipo de información más rico de Internet, por otra parte, la teoría en la educación es empleada para darle soporte a los aspectos prácticos, jugando un rol muy importante al principio de la enseñanza de un tópico. Dados estos fundamentos se considero tomar la información de Wikipedia para cumplir con los objetivos de este proyecto. De tal forma es necesario describir los aspectos que rodean el minado de texto, las propuestas actuales para medir las relaciones entre conceptos y profundizar en la organización de la información de Wikipedia.

Evaluar la similitud entre conceptos ha sido ampliamente estudiado y es una cuestión central y común en muchas áreas de investigación como: psicología, lingüística, ciencia cognitiva, biomedicina e inteligencia artificial. En el área de lingüística computacional se utiliza para: procesamiento del lenguaje natural, adquisición de conocimiento, recuperación de información y desambiguación del sentido de palabras (Jiang *et al.*, 2015).

4.1. Minería web

La minería web ayuda a encontrar nuevos patrones y conocimiento de la información de la web. Este tipo de minería puede ser categorizada en tres tipos (Liu, 2011):

- Minería de la estructura de la web: Descubre conocimiento útil de los enlaces.
- Minería del contenido de la web: Extrae información útil del contenido de las páginas.
- Minería del uso de la web: Se refiere al descubrimiento de patrones de acceso de los usuarios sobre sus registros de uso.

Esta investigación se enfoca principalmente en el minado del contenido web y secundariamente sobre la estructura de la web. El uso del contenido, basado en corpus (conjunto de información textual), para obtener relaciones entre palabras se basa en la distribución de co-ocurrencias, sin embargo, pierde la relación con otros documentos, lo cual es abordado por el análisis de la estructura de la web. Por tanto, una combinación entre estos dos enfoques es adecuada para subsanar las carencias individuales. Se dejó de lado la minería del uso de la web, debido al enfoque de análisis un poco más tardado y con fines distintos a los objetivos del proyecto.

4.1.1. Minería de texto

El minado de texto, una variante de minado de datos pero basado en documentos de texto, es una disciplina importante del procesamiento del lenguaje natural. El análisis de grandes cantidades de datos textuales pre-procesados lingüísticamente no es sólo un requisito previo para la construcción de recursos léxicos tales como ontologías y diccionarios relacionados con el lenguaje, también tiene aplicaciones directas en el procesamiento automatizado de texto.

A continuación se explica el proceso de preparación de los datos para explotarlos.

4.1.2. Indexación de documentos de texto

Para poder implementar mecanismos de recuperación sobre una colección de documentos de texto es necesario obtener una representación de los mismos. Dicha representación responde a la implementación de un conjunto de criterios mediante los cuales se obtienen los términos y las relaciones entre éstos. Toda implementación de un sistema de recuperación de información comienza con esta tarea de procesamiento del corpus. Esto se debe a que no todos los términos que componen un documento son igualmente representativos de su contenido; cuestiones como su posición, la cantidad de ocurrencias, su función lingüística, entre otras, definen el grado de importancia de cada uno de los términos. El resultado es una representación de la información, que es computacionalmente adecuada para los procesos siguientes y que generalmente se describe como *indexación de la colección* (Tolosa, 2008).

El proceso de indexación puede realizarse desde dos enfoques (Tolosa, 2008):

- Basado en métodos no lingüísticos: Se utilizan técnicas estadísticas para análisis de frecuencias y cálculo de pesos de los términos, análisis de probabilidades para determinación de multipalabras y técnicas de agrupamiento (clustering) destinadas a la detección y extracción de relaciones.
- Basado en métodos lingüísticos: Se utilizan técnicas derivadas del procesamiento del lenguaje natural (PLN), las que pretenden imitar el comportamiento de los indizadores humanos.

En el enfoque de esta investigación se brinda especial atención a los métodos no lingüísticos.

4.1.3. El proceso de indexación

La indexación nos permite identificar los conceptos que representan el contenido de un texto de una forma computacionalmente manejable. Para esto el texto debe someterse a una limpieza que elimine información innecesaria de acuerdo a los objetivos. A continuación describimos el proceso:

1. Sólo minúsculas: Convertimos todos el texto a sólo minúsculas para evitar distintas palabras con diferencia en mayúsculas o minúsculas (Bhowmik, 2008).

2. Eliminación de símbolos: Se eliminan todos los símbolos (;?!., etc.) encontrados en el texto.
3. Manejo de Stop-Words: Consisten en palabras comunes tales como: Determinadores (a, the, this), preposiciones (in, from, into, of, on), conjunciones (and, but, after, since, as) (Fox, 1989; Bhowmik, 2008). Este tipo de palabras aparecen siempre en las primeras posiciones en las frecuencias de palabras ordenadas de un texto (Zipf, 1950). En el enfoque no lingüístico que se maneja, se puede omitir este tipo de palabras que se consideran innecesarias.
4. Aplicar Stemming o lematización: Esta técnica sirve para reducir variantes morfológicas de las formas de una palabra a raíces comunes o lexemas (Bhowmik, 2008). Básicamente, esto consiste en remover el plural, el tiempo, o los atributos finales de la palabra. Por ejemplo cars=car, musical=music.

Una vez aplicado este proceso de limpieza se procede a obtener las frecuencias de los conceptos, determinando cuáles son los más importantes.

4.2. Relación semántica

Por lo general asociamos la palabra *banco* con términos como *dinero* o *moneda*. Somos capaces de identificar fácilmente la relación entre *banco* y *dinero* o *moneda* por nuestro conocimiento y experiencia individuales. Sin embargo, el uso de algoritmos para relacionar de manera precisa y efectiva dos términos diferentes es un problema fundamental e importante (Yang y Kao, 2011).

Identificar la relación semántica de dos palabras es una tarea importante para la recuperación de la información, el procesamiento del lenguaje natural y la minería de texto. Sin embargo, debido a la diversidad de significado de una palabra, la relación semántica de dos palabras es todavía difícil de evaluar con precisión en virtud de los corpus limitados (Yang y Kao, 2011). Los primeros trabajos que abordan la problemática de cómo determinar la similitud existente entre dos objetos o entidades de información se basan en el uso de las características similares y diferentes que tienen los objetos a

comparar (Montoy Iribe, 2010).

Las entidades de información a comparar pueden ser objetos más abstractos que una figura geométrica, como son los conceptos y a su vez dichos conceptos pueden pertenecer a un concepto más general, formando así, una estructura de conceptos o estructura conceptual. Existen distintas estructuras que permiten almacenar y organizar conceptos, como: grafos, taxonomías y árboles (Montoy Iribe, 2010).

La noción de similitud se refiere a identificar conceptos que tienen características en común. La similitud semántica se entiende como el grado de proximidad taxonómica entre conceptos, términos o palabras. En otras palabras, la similitud semántica nos dice cuan cerca numéricamente están dos conceptos, porque ellos comparten algunos aspectos de su significado. Técnicamente las medidas de similitud evalúan un puntaje numérico que cuantifica su proximidad como una función de la evidencia semántica observada en uno o algunas fuentes de conocimiento (Jiang *et al.*, 2015).

Utilizar varias fuentes para obtener la medida de similitud semántica es un enfoque que se ha trabajado anteriormente. Li *et al.* (2003) usaron una combinación de métricas asumiendo que la similitud semántica depende no sólo de múltiples factores, sino de la correcta combinación entre ellos. Estos autores propusieron que la similitud entre dos términos está determinada por una función que considera tres atributos:

- El camino más corto entre dos términos (longitud).
- La profundidad o altura de un término (profundidad).
- El contenido de información.

Las dos primeras se consiguen con la taxonomía léxica, y la tercera con un corpus.

Básicamente, los métodos que utilizan similitud semántica se pueden dividir en: basados en corpus, basados en conocimiento, basados en términos y basados en traslape de palabras (Bollegala *et al.*, 2011; Hossein Zadeh y Reformat, 2013; Álvarez Carmona, 2014).

4.2.1. Métodos basados en corpus

Las medidas basadas en corpus, como su nombre lo indica, utilizan grandes colecciones de datos para obtener estadísticas de los textos y a partir de ellas se calcula el grado de similitud de los textos de entrada. Un ejemplo de las medidas basadas en corpus son las que utilizan la medida TF-IDF (Zhiqiang *et al.*, 2009; Biuk-aghay y Ng, 2014). Pueden subdividirse en (Álvarez Carmona, 2014):

- Hiperespacio Análogo al Lenguaje (HAL). HAL crea un espacio semántico a través de las co-ocurrencias de las palabras en un corpus. Palabra por palabra se construye una matriz M donde m_{ij} representa que tanto co-ocurió la palabra i con la palabra j en la colección de datos. El valor en cada posición de la matriz M se calcula de forma acumulativa. Así, mientras un par de palabras obtenga un valor alto en la matriz, es evidencia de que se refieren a lo mismo dentro del contexto
- Análisis Semántico Latente (LSA). LSA es la técnica basada en corpus más popular. Esta técnica asume que las palabras que semánticamente son similares co-ocurrirán en pequeños pedazos del texto. Para capturar eso, se construye una matriz donde los renglones representan los párrafos de textos y las columnas representan sólo palabras. Posteriormente, como esta matriz regularmente es muy grande, se utiliza una técnica matemática llamada descomposición en valores singulares y es usada para reducir las dimensiones de la matriz tratando de mantener la similitud original. Al final las palabras se comparan de forma vectorial utilizando el coseno del ángulo que forman los vectores a comparar.
- Información Mutua-Recuperación de la información. Es un método para calcular similitud entre palabras. Esta técnica utiliza métodos avanzados de búsqueda de consultas para calcular probabilidades, mientras más co-ocurren palabras cerca de otras en una colección de páginas web más alto es su grado de relación.

4.2.2. Métodos basados en conocimiento

Las medidas basadas en conocimiento utilizan recursos léxicos y semánticos que contienen conocimiento humano de las palabras como diccionarios o tesauros. Estos enfoques se basan en proximidad de palabras de acuerdo a una escala taxonómica definida, una de las escalas más populares disponibles es la de WordNet (Miller, 1995). A continuación se describen las medidas más comunes (Bollegala *et al.*, 2011; Álvarez Carmona, 2014):

- Resnik (1995) devuelve un valor continuo que denota qué tan similares son dos palabras, indicando que entre mayor sea el valor devuelto mayor es la similitud entre las palabras. Este cálculo está basado en el contenido de información en la red, es decir, a partir del nodo en común más cercano a las palabras que se están comparando. Esta función es en su totalidad dependiente de cómo se construyó la red de conocimiento por lo que cualquier cambio en la colección hace que el resultado de similitud entre dos palabras pueda cambiar de forma abrupta. Resnik la define como: $sim = IC(lcs)$, donde $IC(lcs)$ es el nodo donde se encuentra el ancestro más cercano en común.
- Lin (1998) combinó la información semántica estructural de una taxonomía léxica y el contenido de información de un corpus en un modelo no lineal. Este autor propone una medida de similitud que usa la longitud de la ruta más corta, profundidad y densidad local en una taxonomía. Sus experimentos reportan un alto coeficiente de correlación de Pearson de 0.8914 en el conjunto de datos de prueba Miller y Charles (1991). Lin definió la similitud entre dos conceptos como la información que es común a ambos conceptos y la información contenida en cada concepto individual.
- Jiang y Conrath (1997) retornan un valor continuo denotando la similitud entre un par de palabras. Este método está basado en las palabras que se van a comparar y el ancestro más cercano a las dos palabras. El cálculo de esta función está definido como: $sim = 1/(IC(s1) + IC(s2) - 2 * IC(lcs))$. Donde, $IC(s1)$ es el nodo en la red donde se encuentra la palabra uno. $IC(s2)$ es el nodo en la red

donde se encuentra la palabra dos. Finalmente $IC(lcs)$ es el nodo donde se encuentra el ancestro más cercano en común.

- Leacock *et al.* (1998) propusieron un enfoque para medir la similitud semántica como la ruta más corta usando una jerarquía para sustantivos en Wordnet (Miller, 1995). Esta medida determina la similitud semántica entre dos synsets (conceptos) definiendo la ruta más corta y escalando usando la profundidad de la taxonomía: $sim = (c_1, c_2) = -\log(np(c_1, c_2)/2d)$. Donde $np(c_1, c_2)$ muestra la ruta más corta entre los synsets y D es la profundidad máxima de la taxonomía.
- Wu y Palmer (1994) Muy similar a la medida anterior, también se calcula la ruta más corta entre las palabras comparadas. La diferencia recae en el uso de uno de los ancestros de las palabras en la red. En la medida anterior se utiliza el ancestro en común más cercano; en esta medida se utilizará el ancestro más cercano a alguno de las dos palabras. Esta medida se define como: $sim = (2 * profundidad(LCS)) / (profundidad(IC(s_1)) + profundidad(IC(s_2)))$.
- La longitud de ruta (Path length) (Budanitsky y Hirst, 2006) devuelve un valor continuo denotando que tan similares son dos palabras, con base en la trayectoria más corta que conecta las palabras en el árbol obtenido únicamente a través de relaciones del tipo es-un. El resultado está acotado en el rango de cero a uno, donde cero significa nula similitud y uno significa máxima similitud. Esta medida se define como: $sim = SP(IC(s_1), IC(s_2))$. Donde, $SP()$ es la función que devuelve el número de aristas del camino más corto de dos palabras en el árbol de WordNet.

4.2.3. Métodos basados en términos

- N-gramas. Los n-gramas son subsecuencias de n palabras tomados de las cadenas de textos a comparar. El cálculo de similitud mediante n-gramas consiste en dividir el número de n-gramas que comparten ambas cadenas entre el número total de n-gramas. Es decir: $sim_{ngramas}(t_1, t_2) = \frac{|ngramas(t_1) \cap ngramas(t_2)|}{|ngramas(t_1) \cup ngramas(t_2)|}$

- Distancia euclidiana. Se define como la distancia más corta entre dos puntos. Para utilizar esta distancia, los textos se deben representar en un espacio euclidiano mediante vectores, donde cada dimensión de los vectores representa una palabra de los textos. La distancia euclidiana está definida como: $dis_e(t_1, t_2) = \sqrt{\sum_{i=1}^n (V_{t_{1i}} - V_{t_{2i}})^2}$.

Donde: n es la cardinalidad del vocabulario existente para comparar ambas cadenas. V_{t_1} y V_{t_2} son los vectores que representan a las cadenas t_1 y t_2 respectivamente. $V_{t_{1i}}$ y $V_{t_{2i}}$ son los i -ésimos elementos en los vectores que representan a las cadenas t_1 y t_2 respectivamente.

- Distancia Manhattan. La métrica usual de la geometría euclidiana es reemplazada por una nueva métrica en la cual la distancia entre dos puntos es la suma de las diferencias (absolutas) de sus coordenadas utilizando los mismos vectores que en la medida anterior. Esta distancia está definida como: $dis_m(t_1, t_2) = \sum_{i=1}^n |V_{t_{1i}} - V_{t_{2i}}|$.
- Similitud coseno. Al igual que las medidas anteriores, en esta métrica se utiliza una representación vectorial. El resultado es el coseno del ángulo entre los dos vectores a comparar; mientras menor sea el ángulo, mayor será la similitud. El valor del ángulo resultante de entre dos vectores se define como: $\theta = \cos^{-1}(\frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|})$

4.2.4. Métodos basados en traslape de palabras

- Coeficiente de Dice. Este coeficiente se basa en la teoría de conjuntos. Toma el número de las palabras que comparten ambas cadenas y los divide entre el número total de la suma de las palabras del texto uno y dos. Es decir: $sim_D(t_1, t_2) = 2 \frac{|t_1 \cap t_2|}{|t_1| + |t_2|}$.

Esta métrica está normalizada, cero es nula similitud mientras que uno se refiere a la máxima similitud.

- Coeficiente de Jaccard. Parecido al coeficiente anterior, este coeficiente se obtiene al dividir la intersección de términos entre la unión de los mismos, es decir: $sim_J(t_1, t_2) = \frac{|t_1 \cap t_2|}{|t_1 \cup t_2|}$

- Coeficiente de coseno. Este coeficiente se obtiene dividiendo la cardinalidad de la unión de los dos conjuntos entre la raíz cuadrada del producto de las cardinalidades de los conjuntos considerados. Este coeficiente se define como: $sim_C(t_1, t_2) = \frac{|t_1 \cap t_2|}{\sqrt{|t_1| |t_2|}}$

4.3. Wikipedia

Wikipedia ¹ es la enciclopedia colaborativa más grande del mundo (Figura 4.1). Esta enciclopedia ofrece una cantidad de recursos de alta calidad en términos de contenido. La disponibilidad y frescura de los datos hacen que Wikipedia sea interés de una variedad de actividades de investigación (Zhiqiang *et al.*, 2009; De Medio *et al.*, 2016). Wikipedia fue lanzada en 2001 con la meta de construir una enciclopedia gratuita en varios idiomas.



Figura 4.1: Interfaz de Wikipedia.

La Tabla 4.1 muestra el crecimiento de los trabajos de investigación relacionados con Wikipedia desde su surgimiento hasta el año 2014 (Biuk-aghai y Ng, 2014). Esta enciclopedia es ampliamente utilizada por la comunidad científica en diversas áreas como clusterizado de texto, desambiguación de palabras y resolución de correferencias (Nakamura *et al.*, 2014). Esta información es hasta mediados del 2014, por tanto, a estas fechas se tendrán muchas más publicaciones relacionadas con Wikipedia.

¹https://en.wikipedia.org/wiki/Main_Page

Tabla 4.1: Estadísticas de Wikipedia.

| Año | ACM DL | IEEE xplore | Web of Science | Google Scholar |
|-------|--------|-------------|----------------|----------------|
| 2001 | 0 | 0 | 0 | 7 |
| 2002 | 0 | 0 | 1 | 2 |
| 2003 | 0 | 0 | 1 | 4 |
| 2004 | 0 | 0 | 0 | 14 |
| 2005 | 4 | 0 | 10 | 38 |
| 2006 | 16 | 5 | 21 | 142 |
| 2007 | 27 | 12 | 50 | 228 |
| 2008 | 47 | 32 | 95 | 377 |
| 2009 | 128 | 25 | 128 | 432 |
| 2010 | 137 | 45 | 102 | 548 |
| 2011 | 149 | 39 | 107 | 548 |
| 2012 | 137 | 31 | 93 | 563 |
| 2013 | 105 | 26 | 115 | 508 |
| 2014 | 48 | 13 | 43 | 277 |
| Total | 798 | 228 | 766 | 3661 |

Wikipedia está basado en el software MediaWiki como un proyecto open-source. Su contenido entero es fácilmente obtenible de archivos SQL y XML. Estos son liberados esporádicamente cada semanas con actualizaciones (Medelyan *et al.*, 2009). Actualmente Wikipedia incluye más de 5 millones de artículos en su versión en inglés, además de 750 artículos añadidos diariamente.

4.3.1. Estructura

La estructura de Wikipedia sigue las secciones (Medelyan *et al.*, 2009; Jiang *et al.*, 2015):

- **Artículos:** Es la unidad básica de información en Wikipedia. Estos son escritos en formato de texto libre, siguiendo un conjunto de directrices editoriales para promover la consistencia y cohesión. Cada artículo describe un concepto simple.
- **Páginas de redirección:** Estas páginas son usadas para redirigir la consulta hacia un artículo. Generalmente se usan cuando se consulta un nombre alternativo al título de la página.
- **Páginas de desambiguación:** Este tipo de página contiene una lista de términos posibles para la

consulta dada, debido a que la consulta es ambigua. Aquí se manejan los homónimos, es decir términos con el mismo nombre pero distinto significado

- Enlaces: Cada artículo se agregan enlaces hacia otros artículos con relación similar. Estos están embebidos en el contenido del artículo, brindan información más completa sobre conceptos descritos en el texto.
- Estructura de categorías: A través de las categorías se establece una red semántica. Los artículos son asignados a categorías como forma de organización. La categorización es una especie jerarquía de árbol, aunque en la práctica permite múltiples categorizaciones.
- Plantillas: es un fragmento de código de Wikipedia que puede ser reutilizado por muchas páginas distintas, sin necesidad de copiar todo el código nuevamente. Corresponden a un tipo de páginas especiales, destinadas a utilizarse dentro de otras.
- Infoboxes: Son plantillas usadas para coleccionar y presentar un subconjunto de información sobre un tema. Es un pequeño documento estructurado con un conjunto de pares atributo-valor. Representa un resumen de información.

Estructura de un artículo en Wikipedia(Medelyan *et al.*, 2009; Jiang *et al.*, 2015):

- Sección introductoria: Todo artículo comienza con una sección introductoria, y toda sección introductoria comienza con una definición o caracterización lo más concisa posible del concepto o tema del artículo. En general, en la primera oración también se acostumbra hacer una mención en negritas del título mismo del artículo. Tras la definición, la sección introductoria del artículo debe: Poner en su contexto el tema del artículo, Ofrecer la información más relevante de este. Mostrar por qué el tema del artículo es notable e interesante. La entradilla generalmente no debe tener más de cinco párrafos de largo. Una vez finalizada, se debe dejar una única línea vacía para ayudar a los demás editores a distinguirla del resto del artículo. La sección introductoria puede contener también otros elementos adicionales, como enlaces de desambiguación, etiquetas

de mantenimiento, fichas, imágenes y una tabla de contenidos. Algunos de estos elementos son explicados a continuación.

- **Desambiguación:** A veces, varios artículos pueden tener títulos muy similares, esto puede causar que muchos lectores lleguen a un artículo por error, mientras buscan algún otro. Para ayudar a estos lectores a encontrar su camino, a veces se inserta antes de la introducción una advertencia que informa al lector sobre la existencia de una página de desambiguación que contiene un listado de otros artículos con títulos similares.
 - **Índice:** El índice o tabla de contenidos aparece de forma automática cuando hay más de tres secciones.
- **Secciones y subsecciones:** Este está dividido en secciones y subsecciones para organizar el contenido y para ayudar al lector a encontrar lo que busca. Cada sección o subsección se introduce mediante un título, y existen seis niveles distintos de secciones y subsecciones.
- **Enlaces a artículos principales:** Cuando el texto de una sección o subsección es un resumen de otro artículo que provee un tratamiento más detallado del tema, entonces debe aparecer debajo del título de la sección un enlace al artículo en cuestión.
 - **Véase también:** Una vez terminado el desarrollo del artículo, la sección Véase también proporciona al lector enlaces a otros artículos de la Wikipedia que puedan resultarle de interés. Esta sección consiste sólo en enlaces a artículos, ordenados por medio de una lista normal (no numerada). En general deben evitarse los enlaces a artículos que todavía no existen.
 - **Notas y referencias:** La función de esta sección es desplegar las notas al pie que hayan sido insertadas en el cuerpo del artículo mediante las etiquetas `<ref>` y `</ref>`. Dichas notas se utilizan para citar los trabajos que justifican las afirmaciones particulares del artículo, así como para insertar las notas aclaratorias que hubieran sido incómodas de incluir dentro del texto principal.

- **Bibliografía utilizada:** En esta sección deben aparecer los libros, artículos, publicaciones y demás que hayan sido utilizados de manera general para redactar el artículo, y que por lo tanto permitan al lector verificar el contenido de aquel.
- **Bibliografía adicional:** En esta sección pueden incluirse textos recomendados que no se hayan mencionado en el cuerpo del artículo y que tampoco se hayan citado en la sección de Notas y referencias o de Bibliografía utilizada.
- **Enlaces externos:** Se incluyen enlaces a sitios web ajenos a Wikipedia que sean relevantes al contenido del artículo. El título de esta sección se escribe en plural, aun cuando haya listado un solo sitio web.
- **Categorías:** Describe los enlaces a las categorías en las que se desea incluir al artículo
 - **Ejemplo:** Se agregan ejemplos para comprender el artículo.
 - **Véase también:** Son enlaces a otros artículos en Wikipedia relacionados con el actual.

4.3.2. Extracción de información

Wikipedia ofrece un conocimiento mucho más amplio que taxonomías existentes tal como Wordnet Miller (1995); Sánchez-bermejo (2013), por otra parte, está estructurada de una manera más definida que la información que puede existir en la Web. El conocimiento que alberga Wikipedia está consensuado por un gran número de personas y se actualiza de manera ágil, además esta traducida en numerosas lenguas. Es por esto que una medida de cálculo de similitud semántica basada en esta fuente parece resultar, en principio, la más idónea y eficaz, cubriendo no sólo multitud de dominios y escenarios sino que también puede utilizarse para procesar términos de diferentes idiomas (Sánchez-bermejo, 2013).

Wikipedia tiene algunos atributos que la hacen ideal para la extracción de información (Buzzi y Leporini, 2009):

- Al manejar contenido del artículo se pueden extraer conceptos relacionados al tema.

- Manejo de páginas de desambiguación que nos permite encontrar fácilmente sinónimos o homónimos.
- Provee una jerarquía taxonómica de categorías para agrupar los artículos.
- Ofrece infoboxes, son cajas de información resumida organizada en una sección aparte

El acceso a la información de Wikipedia puede realizar a través de acceso directo de las páginas por Internet, como rastreadores de datos, o la descarga de base de datos.

4.3.3. Recursos y herramientas

Existen varias herramientas para explotar la información en Wikipedia. Estas han sido ampliamente utilizadas por la comunidad científica, A continuación listamos API's utilizadas para el acceso a Wikipedia:

- Wikipedia API: Esta la provee la fundación WikiMedia para acceder al arreglo de datos a través de REST API. Acceder a esta API es más lento que las versiones almacenadas localmente. Sin embargo, su gran ventaja es que siempre está actualizada con la información del momento (Sen *et al.*, 2014).
- JWPL: Es una biblioteca de funciones para java que que explota la información en Wikipedia (Java Wikipedia Library por sus siglas en ingles). Es gratuita y tienen como principales ventajas el rápido y eficiente acceso a la información y un analizador para la sintaxis MediaWiki (Zesch *et al.*, 2008).
- Wikipedia Miner: Es una herramienta de accesos al rico nivel semántico de Wikipedia. Esta provee un acceso simplificado y orientado a objetos a la estructura y contenido de Wikipedia, además mide cómo los términos y conceptos están relacionados con otros (Milne y Witten, 2013)

- Lucene ²: Es una API de código abierto para recuperación de información. Es útil para cualquier aplicación que requiera indexado y búsqueda a texto completo. Lucene ha sido ampliamente usado por su utilidad en la implementación de motores de búsquedas.

Existen otros recursos para obtener información de Wikipedia como (Medelyan *et al.*, 2009; Sen *et al.*, 2014): DBpedia, YAGO, QuaALiM, Koru, Wikify, Bliki y otros más. Sin embargo, dado el enfoque de la investigación nos hemos enfocado en una API para acceder con mayor flexibilidad a la información de Wikipedia.

Se destaca el uso de JWPL en artículos científicos enfocados a la relación semántica (Milne y Witten, 2013; Ferschke, 2014; Jiang *et al.*, 2015) para resolver problemas propios de cada investigación. Por tanto, nos inclinamos por esta herramienta por las características que mencionamos a continuación.

4.3.3.1. Java Wikipedia Library

La biblioteca de funciones de Wikipedia de java (JWPL) ofrece una interfaz de programación basada en java, accediendo a toda la información de Wikipedia en diferentes lenguajes de una forma estructurada. Esto incluye un analizador de etiquetas de MediaWiki para un profundo análisis del contenido de páginas. JWPL ofrece métodos de acceso a propiedades como: enlaces, plantillas, categorías, texto y otras propiedades (Ferschke, 2014).

La estructura original de Wikipedia está optimizada para la búsqueda de artículos por palabras claves, la cual es utilizada por millones de usuarios en línea. Sin embargo, el diseño de un API para procesamiento de lenguaje natural tiene que dar soporte a un amplio acceso a rutas, incluyendo iteraciones sobre todos los artículos, una sintaxis de consulta, así como un eficiente acceso a la información como enlaces, categorías y redireccionamientos. Así que JWPL opera sobre una base de datos optimizada que es creada de los archivos disponibles por Wikimedia Foundation para mejorar el procesamiento.

Las ventajas de este tipo de arquitecturas son (Zesch *et al.*, 2008):

²<https://lucene.apache.org/core/>

- Se logra una mayor eficiencia computacional para tareas a gran escala de Procesamiento de Lenguaje Natural (por sus siglas en inglés NLP): Es una consecuencia del acceso a las bases de datos usando mecanismos de indexación propios para una rápida recuperación
- Reproducir los resultados de investigaciones: Al poder descargarse versiones previas de Wikipedia nos da la posibilidad de recrear procesamientos pasados. Una desventaja ante Wikipedia en línea que se actualiza constantemente.
- Una forma fácil de usar una interfaz de programación orientada a objetos: Principalmente se cuenta con los objetos: Wikipedia, Page y Category. El primero es para establecer una conexión con la base de datos. El segundo nos permite el acceso a artículos individuales y sus características (títulos, texto plano, texto no plano infoboxes, enlaces de salida, de entrada entre otras más). El último objeto representa las categorías que maneja Wikipedia donde se agrupan los artículos u otras categorías.

4.4. Resumen del capítulo

El proceso considera el análisis de documentos de texto, por lo cual, se describieron conceptos de minería de datos y los procesos de tratamiento del texto para que la computadora procese la información. Por otra parte, se describieron los principales métodos de relación semántica entre conceptos de la literatura con el objetivo de reutilizar elementos que den soporte a nuestra propuesta. Por último, se explicaron algunos aspectos relacionados a Wikipedia, la fuente de conocimiento para los casos de estudio; así como algunas herramientas para explotar la información.

El siguiente capítulo desarrolla el método propuesto empleando términos descritos en los capítulos 2, 3 y 4.

Capítulo 5

Generación de la red Bayesiana

Esta sección hace énfasis los detalles de la propuesta, se emplean dos términos de las RBs: las partes cualitativa y cuantitativa. La primera se refiere a la estructura de la RB, es decir, como esta conformada por las variables y su relaciones; la segunda parte se refiere a las TPCs, una estructura que representa la probabilidad que ocurra cada combinación entre las variables de la parte cualitativa.

El proceso de desarrollo de la propuesta fue dividido en algoritmos con el objetivo de modularizar la problemática, aunque al final el método se representa como uno solo. En la primer sección se explica el método en forma general, después se definen las características que dan forma a la propuesta. Finalmente, se puntualiza como se incorpora la propuesta con las ontologías y una enciclopedia en línea.

5.1. Descripción general del método

La propuesta de esta investigación es representada en la Figura 5.1, se despliega el proceso de conversión de ontología a RB a través de 3 algoritmos principales:

- Algoritmo 1, relación cuantitativa entre nodos: Este algoritmo obtiene una relación numérica entre los conceptos la cual represente el grado de relación entre ellos.

- Algoritmo 2, Creación de las TPCs: Con base en las relaciones determinadas en el algoritmo anterior se construyen las TPCs, las cuales representan la probabilidad de que ocurra algún evento considerando cada combinación posible entre los nodos relacionados.
- Algoritmo 3, Integración de algoritmos: Los tres algoritmos desarrollados se unen para formar el método propuesto, donde las salidas de los primeros algoritmos son las entradas de los segundos.
 - Algoritmo 3.1, Creación de la estructura de la red: Este algoritmo construye la parte estructural de la RB que constituye los nodos y sus relaciones

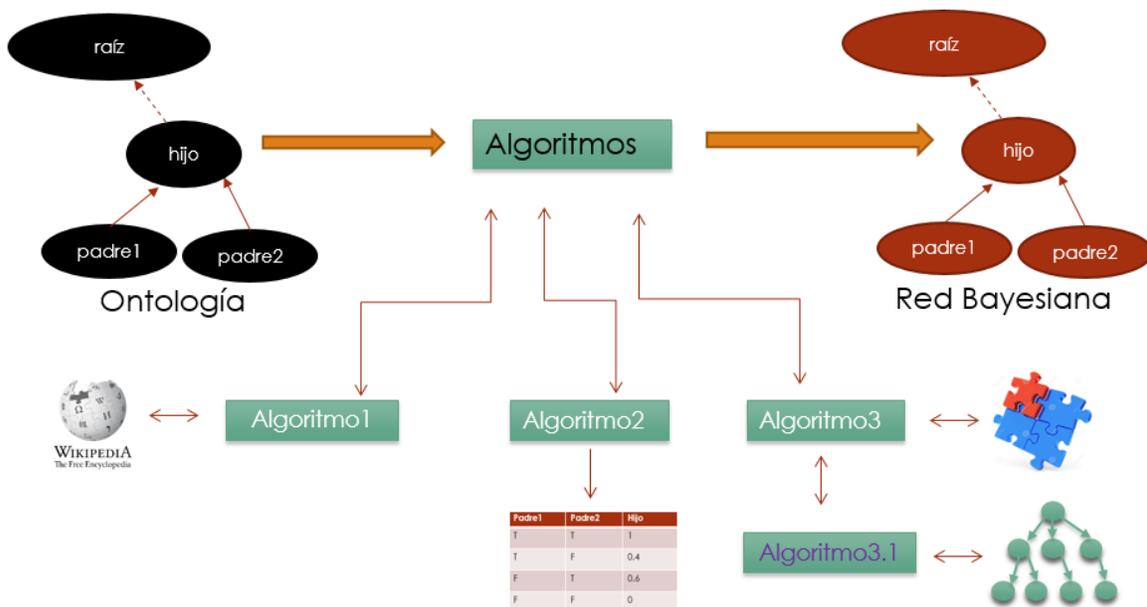


Figura 5.1: Proceso de construcción de la red Bayesiana.

El proceso, en forma general, empieza con una ontología que represente conocimiento de algún tema de enseñanza. El primer paso es calcular la relación entre nodos (Algoritmo 1), después construir las TPCs con los valores de la relación (Algoritmo 2), finalmente, construir los nodos y relaciones que forman la estructura de la red integrando los algoritmos anteriores (Algoritmo 1 y 2). La Figura 5.2 representa el resultado de cada algoritmo.

De forma básica, el método busca en documentos de texto información estadística que permita definir la relación entre los nodos. A través de la literatura se seleccionaron una serie de atributos que

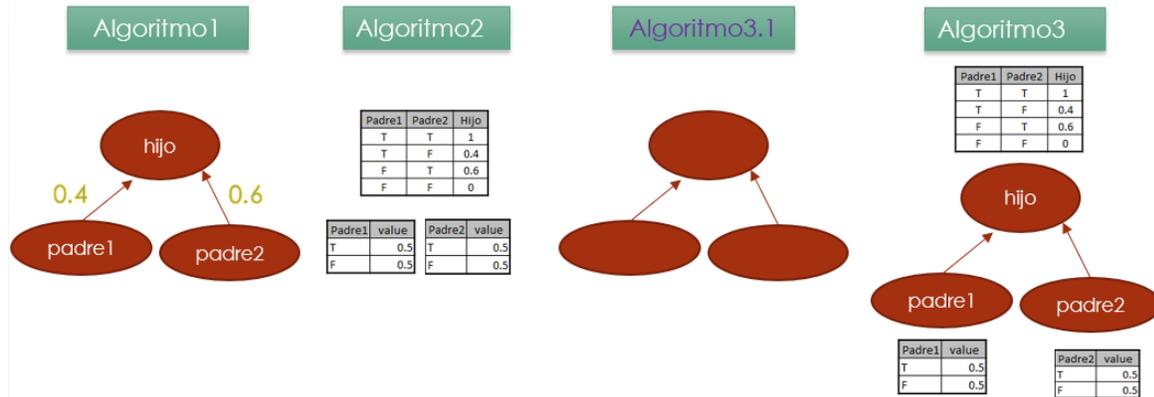


Figura 5.2: Algoritmos para construir la red Bayesiana.

diversas investigaciones fundamentan que son importante como medidas de similitud entre conceptos.

Frecuencia del concepto (positivo): La idea básica de este factor es que las palabras claves de un texto recaen en aquellos términos que más se repiten, previa aplicación a un procesos de limpieza como: la eliminación de palabras de parada (Fox, 1989) o el algoritmo de Porter (1980). Por tanto, un concepto es importante si se repite varias veces en un documento (Li *et al.*, 2003; Bhowmik, 2008; Baoyao *et al.*, 2009). Este factor es de bastante uso en diversas técnicas basadas en el algoritmo ID-TFD (Zhiqiang *et al.*, 2009). Para este enfoque se considera frecuencias del concepto padres en el concepto hijo.

Saltos del concepto padre hasta el concepto hijo (negativo): Basados en la estructura de Wikipedia y no en el contenido calculamos esta medida. En un estructura jerárquica de árbol los conceptos que impliquen menos saltos entre los nodos son los mayormente relacionados (Leacock y Chodorow, 1998; Li *et al.*, 2003; Zhang *et al.*, 2012; Jindal *et al.*, 2014). La relación entre dos conceptos (A y B) esta ampliamente relacionada si dentro del contenido del concepto A se encuentra un enlace directo hacia el concepto. Esta relación decrementará de acuerdo al número de saltos que existan entre los dos conceptos.

Incidencia de conceptos en el contenido de ambos artículos (positivo): Esto se refiere a que si dos términos están relacionados compartirán palabras dentro de su significado (Lesk, 1986; Zhang

et al., 2012). Se basa en la idea de obtener la frecuencia de palabras de dos textos, estarán mayormente relacionados tanto compartan mayores palabras comunes entre ellos.

5.2. Características del método

Nuestra propuesta desarrolló una RB a través de ontologías e información de Wikipedia, sin embargo, el método propuesto en este documento podría aplicarse utilizando diferentes fuentes de conocimiento. En esta sección se explican las características mínimas para aplicar nuestro método utilizando diferentes fuentes de conocimiento. Esta descripción es dividida en las dos partes básica de una RB, la parte cualitativa y la parte cuantitativa.

5.2.1. Parte cualitativa

La estructura cualitativa debe representar un dominio del conocimiento exacto, esta representación es un dominio descompuesto a la unidad mínima, el concepto. Los conceptos están relacionados entre sí para formar conocimiento, por lo tanto, dos elementos son al menos necesarios para representar el conocimiento:

- **Nodos:** Representan conceptos de un dominio de conocimiento.
- **Relaciones:** Representan enlaces entre nodos, las cuales deben ser unidireccionales.

Las características mínimas de la parte cualitativa se representan como:

Definición 5.1. *Sea $KD = \{K, I\}$ el conjunto de elementos que forman la parte cualitativa del conocimiento de un dominio. Donde el conjunto $K = \{k_1, \dots, k_{|K|}\}$ representan los nodos o conceptos del dominio de conocimiento, y el conjunto $I = \{i_1, \dots, i_{|I|}\}$ representa los pares ordenados para formar la estructura de conocimiento. Un elemento en I es representado como $i = (k, z)$; donde $k, z \in K$ y $k \neq z$, el elemento k representa el inicio de la relación y el elemento z el fin de la relación.*

La estructura del conocimiento no debe tener ciclos para cumplir con las definiciones de la RB, además, los nodos deben converger en un único nodo que representa el tema principal de la estructura, este nodo se llama *raiz*.

5.2.2. Parte cuantitativa

La estructura cualitativa representa la información estadística para describir el peso de las relaciones entre los nodos. El supuesto principal es que el texto contiene información que puede ser expresada a través de estadística. Los algoritmos de análisis de texto pueden generar información estadística más exacta cuanto más información textual se tenga.

Una enciclopedia es una acumulación de conocimiento sistematizado y estructurado que facilita el aprendizaje. Después de Internet, una enciclopedia en línea es uno de los mayores repositorios de información que podemos consultar; por lo tanto, representa la mejor fuente de información basada en texto que se puede explotar para generar datos estadísticos en para cumplir los objetivos de esta tesis.

Una enciclopedia debe cumplir algunos aspectos mínimos para explotar el conocimiento con el método propuesto:

- Artículos: Estos deben tener un título y describir mínimamente el contenido en modo texto. Cuanto más desarrollado sea el contenido, más precisa será su estadística.
- Enlaces entre artículos: Los artículos deben tener referencias a otros artículos a través de enlaces.

Las características mínimas de la parte cuantitativa son representadas como:

Definición 5.2. Sea $G = \{g_1, \dots, g_{|G|}\}$ el conjunto de artículos que representan una enciclopedia. Un elemento en G es representado como $g = \{e, o, U\}$, cada artículo g mínimamente posee tres atributos: título (e), contenido (o) y enlaces (U).

5.3. Construcción de la red Bayesiana

Una RB es construida de acuerdo a la definición 3.4 $BN_{tupla} = \{N, DB, P\}$. Se pueden asignar elementos de nuestro método (Definiciones 5.1 y 5.2) a los componentes de BN_{tupla} para generar una RB.

5.3.1. Parte cualitativa

Las variables de la RB se pueden obtener como $N = K$, donde K representa los conceptos de un dominio de conocimiento. Las relaciones de la RB se puede obtener como $DB = I$, donde I representa las relaciones entre variables.

5.3.2. Parte cuantitativa

La parte cuantitativa es difícil de construir porque la estructura no define explícitamente la información cuantitativa; por lo tanto, necesitamos una fuente de conocimiento para generarla. Esta investigación considera que la información de una enciclopedia en línea ofrece la mejor fuente para producir la parte cuantitativa de la RB. La siguientes partes del documento describen la propuesta de nuestro método.

5.3.2.1. Relación cuantitativa entre nodos

La Figura 5.3 muestra las variables para representar el problema y cómo se relacionan. El nodo raíz de la estructura está representado por la variable r ; la variable h representa un nodo compuesto por otros nodos; finalmente, las variables p son nodos que forman otro nodo.

Sea $X = \{r, h, \{P\}\}$ el conjunto de elementos que forman la estructura básica para obtener los pesos de las relaciones, la estructura del conocimiento se descompone en muchos conjuntos X para calcular pesos. El conjunto $P = \{p_1, p_2, \dots, p_{|P|}\}$ representa los nodos padres que influyen en un nodo hijo.

Los elementos del conjunto X se relacionan de la siguiente manera (ver Figura 5.3):

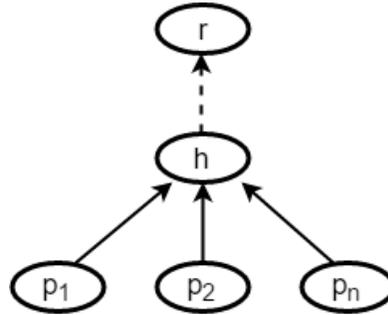


Figura 5.3: Relaciones entre variables en una estructura de representación del conocimiento.

- El elemento *raíz* está influenciado por el elemento *hijo*, (h, r) , en este caso r es un hijo y h actúa como padre.
- El elemento *hijo* está influenciado por un elemento de P , (p, h) , en este caso h es un hijo y p actúa como padre.

El objetivo es encontrar la relación cuantitativa entre los padres y el hijo. Cada elemento de X corresponde a un artículo que describe un concepto. El elemento $x \in X$, puede ser tratado por el título (concepto que representa el artículo) o el contenido (el texto del artículo incluyendo enlaces, referencias y atributos). Cuando nos referimos al concepto, será denotado como $ct(x)$ y para el contenido como $cd(x)$. Lo mismo se aplica a los elementos de P , para el concepto como $ct(p)$ y para el contenido como $cd(p)$.

Este estudio define la Ecuación (5.1) para obtener la relación cuantitativa entre variables. Esta ecuación considera tres factores: el factor fr_{norma} obtiene la medida de relación normalizada considerando las frecuencias, el factor j_{norma} obtiene la medida normalizada de relación teniendo en cuenta el coeficiente de Jaccard (Jaccard, 1912), y el factor s_{norma} obtiene la medida normalizada de la relación considerando la escala jerárquica de Wikipedia. Las siguientes secciones explican estos factores.

$$rel(p, h) = \frac{1}{3}(fr_{norma}(p) + j_{norma}(p) + s_{norma}(p)) \quad (5.1)$$

Relación cuantitativa considerando las frecuencias (Factor fr)

Este factor, llamado factor fr , calcula un valor que representa un elemento de P buscando frecuencias de términos en todos los nodos directamente relacionados; nodos hermanos, el nodo hijo y el nodo raíz. Este factor define la Ecuación (5.2) para encontrar esta relación. Donde $fr(ct(x), cp(p))$ es la búsqueda de frecuencia del término p en el dominio X . El proceso se repite para cada elemento de X obteniendo la suma de resultados. El elemento $ct(x)$ debe ser diferente de $ct(p)$ porque P es un subconjunto de X . Los resultados se normalizan mediante la Ecuación (5.3).

$$fr(p) = \sum_{x \in X} (fr(ct(x), cp(p))) \quad (5.2)$$

$$fr_{norma}(p) = \frac{fr(p)}{\sum_{q \in P} fr(q)} \quad (5.3)$$

Relación cuantitativa considerando la coincidencia de conceptos (factor j)

Este factor establece una relación cuantitativa entre dos conjuntos, W_1 y W_2 . Las palabras encontradas en el contenido del artículo p representan W_1 , y las palabras encontradas en el contenido del artículo h representan W_2 . La Ecuación (5.4) representa la relación obtenida a través del coeficiente de Jaccard (Jaccard, 1912). La ecuación representa la intersección de ambos conjuntos divididos por la unión de ambos conjuntos. Los resultados se normalizan mediante la Ecuación (5.5).

$$j(W_1, W_2) = \frac{|W_1 \cap W_2|}{|W_1 \cup W_2|}. \quad (5.4)$$

$$j_{norma}(p) = \frac{j(p)}{\sum_{q \in P} fr(q)} \quad (5.5)$$

Relación cuantitativa considerando saltos entre conceptos (Factor s)

Este factor se refiere al número de saltos entre dos conceptos, $conc_1$ y $conc_2$. Los saltos son el número de enlaces que deben seguirse para pasar de una página de Wikipedia a otra. Los conceptos

$conc_1$ y $conc_2$ representan a $cp(p)$ y $cp(h)$, respectivamente. La Ecuación (5.6) estima la relación.

$$s(conc_1, conc_2) = (w - d(conc_1, conc_2))(1/x_r) \quad (5.6)$$

La ecuación es soportada por $x_r = \text{round}(x_{mean})$ y $w = x_r + 1$, donde x_r representa el promedio redondeado de saltos para ir de una página a otra página en Wikipedia, Milne y Witten (2008) y Arola y Wysocki (2012) definieron este número en cuatro. La variable w representa un límite de saltos, cualquier número de saltos de una página a otra mayor que w se considerará una relación nula. Los resultados se normalizan mediante la Ecuación (5.7).

$$s_{norma}(p) = \frac{s(p)}{\sum_{q \in P} fr(q)} \quad (5.7)$$

5.3.2.2. Algoritmo para construir las tablas de probabilidad condicional

El Algoritmo 1 recibe como entrada el peso de las relaciones entre los nodos (*weights*) y la salida es una TPC (*CPT*) para el nodo hijo en cuestión.

En la primera sección del algoritmo, las variables son declaradas e inicializadas (de la línea 2 a la 6). En la segunda sección, dos ciclos regulan la generación de la matriz *CPT*, el primer bucle (línea 7) controla las filas, y el segundo bucle (línea 9) controla las columnas. El objetivo es agregar pesos asignados por columna; La variable booleana *flag* controla la asignación de pesos en la matriz. Estas asignaciones se mezclan con la Ecuación $jumps \leftarrow (2^{rows-x})/2$. La matriz final es de $2 \times (2^{length(weights)})$, donde la primera posición de las filas representa el estado *present* (*presente* o verdadero), y la segunda posición representa el estado *absent* (*ausente* o falso).

El Algoritmo 1 genera una TPC para un único nodo, este algoritmo es invocado para cada nodo hijo de la red.

Algoritmo 1 Construcción de una TPC para un nodo hijo.

```

1: function BUILD_CPT(weights[])
2:   rows ← length(weights)
3:   columns ← 2rows
4:   initialize CPT[present][columns] with zeros
5:   initialize CPT[absent][columns] with ones
6:   cont ← 0; flag ← true
7:   for all x ∈ {0, 1, ..., (rows - 1)} do
8:     jumps ← (2rows-x)/2
9:     for all y ∈ {0, 1, ..., (columns - 1)} do
10:      if flag then
11:        CPT[present][y] ← CPT[present][y] + weights[x]
12:        CPT[absent][y] ← 1 - CPT[present][y]
13:      end if
14:      cont ← cont + 1
15:      if cont ≥ jumps then
16:        if flag then
17:          flag ← false
18:        else
19:          flag ← true
20:        end if
21:        cont ← 0
22:      end if
23:    end for
24:  end for
25:  return CPT
26: end function

```

5.4. Método propuesto: Ontologías y Wikipedia

Esta sección relaciona el método propuesto con una estructura de conocimiento, como las ontologías, que permita formar la parte estructural de la RB y con una fuente de información, como Wikipedia, para obtener las TPCs.

5.4.1. Parte cualitativa de la red Bayesiana utilizando ontologías

La ontología es una estructura representada por nodos y relaciones, esta estructura cumple las características mínimas para ser la parte cualitativa de la RB. A continuación es descrito cómo se obtienen las variables, los estados y las relaciones de la RB basado en la ontología

5.4.1.1. Obtención de variables e identificación de estados

Las variables representan el conocimiento y se miden usando una escala de 0 a 1, que representan probabilidades de poseer conocimiento o la falta de él. Una variable necesita dos estados; el estado

actual *presente* determina el grado de conocimiento de los estudiantes, y el estado *ausente* representa lo opuesto al primer estado (Ramírez-Noriega *et al.*, 2017c).

Teniendo en cuenta la representación de la ontología (definición 3.7) y de la RB (definición 3.4), existe una relación directa entre las clases de ontología, set C de la definición 3.8, y las variables de la RB, conjunto N de la definición 3.1. Cada conjunto representa la unidad fundamental en cada modelo, son similares y se puede establecer una relación como $N = C$, donde $C = \{c_1, c_2, \dots, c_{|C|}\}$, por lo tanto $N = \{n_1 = c_1, n_2 = c_2, \dots, n_{|C|} = c_{|C|}\}$.

Hay dos estados predefinidos para medir el conocimiento del estudiante; *presente* y *ausente*, cada variable contiene ambos estados.

5.4.1.2. Relaciones entre variables

Esta sección clasifica las variables en niveles de granularidad formando una escala jerárquica. Las relaciones entre variables en una RB (definición 3.5) se pueden obtener con base en el conjunto T de la ontología (definición 3.11). El elemento $db_i = (m \leftarrow c_1, n \leftarrow c_2)$ donde $c_1, c_2 \in t_i, t_i \in T, db_i \in DB$.

El tipo de relación de la RB no es necesaria porque las relaciones en una RB son siempre causales, por lo tanto, el método no considera la información semántica.

5.4.2. Parte cuantitativa de la red Bayesiana con Wikipedia

Wikipedia se compone de elementos tales como: artículos, páginas de redirección y desambiguación, enlaces, categorías, infoboxes entre otros (Medelyan *et al.*, 2009). Esta enciclopedia cumple los criterios mínimos de nuestro modelo (definición 5.1); artículos que explican conceptos a través de su contenido con enlaces a otros conceptos relacionados, además, los artículos tienen un título. La probabilidad de encontrar un concepto dentro del índice de Wikipedia es alta debido a un gran número de artículos, esto hace que Wikipedia sea una fuente de conocimiento apropiado para nuestras necesidades.

5.4.2.1. Identificación de las variables del modelo usando Wikipedia

La parte cuantitativa requiere identificar elementos básicos del modelo en la ontología; r , h , y P , estas variables representan el conjunto X . El elemento central de la ontología se denomina raíz r , una de las características del modelo es que la ontología elegida debe converger en un nodo. Los nodos que actúan como rango en una posición dada se consideran nodo hijo h , el grupo de nodos que sirven como dominio del mismo rango se consideran como el conjunto P . El conjunto $X = \{r, h, P\}$ representa un subconjunto de la ontología, por lo que, la ontología se divide en subconjuntos X para determinar el peso de las relaciones.

5.5. Resumen del capítulo

Este capítulo describió los tres algoritmos principales en que se dividió la propuesta, el proceso se enfocó en construir primeramente la parte cualitativa de la RB, después la parte cuantitativa, para al final, un algoritmo de integración construya completamente la RB. Además, se contemplaron 3 factores para determinar la relación cuantitativa entre conceptos.

Con las definiciones de este capítulo y las definidas en secciones anteriores se formalizó y explicó el método propuesto. La parte final del capítulo relaciona la formalización con las ontologías y Wikipedia.

Capítulo 6

Experimentación y resultados

La sección de experimentación fue dividida en dos casos:

- Caso de estudio A: El objetivo de este caso de estudio es construir las TPCs, parte cuantitativa, por medio de la relación entre los conceptos de una ontología y la información de Wikipedia. Los resultados del método se centran en entornos educativos, por lo tanto, los resultados se comparan con los expertos del ámbito educativo a través de la prueba de Pearson.
- Caso de estudio B: El objetivo de este caso de estudio es probar el método basado en ontologías para construir las partes cualitativas y cuantitativas de una RB para emplearla en STIs. Este caso emplea el modelo base con los requerimientos mínimos de la parte cualitativa y la fuente de conocimiento para construir la parte cuantitativa. La RB resultante debe mostrar un alto grado de correlación con la opinión experta del dominio.

Como se aprecia en la descripción anterior, el primer caso se enfoca solamente en probar la parte cuantitativa. El segundo caso se enfoca en probar ambas partes de la RB, la parte cualitativa y la cuantitativa, probando de nuevo la misma parte que en el primer caso y desarrollando de forma completa una RB. En ambos casos comparamos las opiniones de los expertos con los resultados del método propuesto, pero en distintos dominios.

6.1. Caso de estudio A: Programación orientada a objetos

Este caso de estudio está diseñado para construir las TPCs de una RB, se prueba que los valores asignados a las TPCs por el método propuesto cuentan con un significativo grado de correlación con los valores asignados por los expertos.

En primer lugar, se seleccionó una ontología del dominio Orientación a Objetos (OO) del artículo (Ramírez-Noriega *et al.*, 2017c). Esto sirve de base para obtener las variables y relaciones para la RB. Se realizó un cuestionario a los maestros que enseñan Programación Orientada a Objetos (POO). Los resultados de la encuesta ayudaron a determinar cuantitativamente la asociación entre un concepto para aprender otro. Luego, el método determinó cuantitativamente las mismas relaciones entre los conceptos que los expertos. Finalmente, se realizó una prueba de correlación entre los datos obtenidos para medir la exactitud del método propuesto.

El experimento fue diseñado y ejecutado para verificar que el valor de las relaciones utilizadas para construir una TPC sea similar a las opiniones de los expertos para este dominio. El método para la construcción de RBs para entornos educativos es nuevo y no hay un conjunto de datos para comparar los resultados.

La opinión de los expertos es la mejor manera de validar estos tipos de estructuras que tratan de representar un dominio a menudo subjetivo. Fenz (2009) validó sus resultados de esta manera utilizando expertos para su evaluación. Este estudio confirmó la importancia de los expertos y estableció una prueba estadística para verificar la correlación.

6.1.1. Desarrollo de la ontología de programación orientada a objetos

La ontología se desarrolló desde cero, sin embargo, se puede tomar cualquier ontología y usar la metodología propuesta. La ontología fue desarrollada a través de Methontology (Fernández-López *et al.*, 1997), siguiendo las fases de especificación, conceptualización, formalización, implementación y evaluación. Methontology es una de las metodologías más comprensibles para construir ontologías

(Gasevic *et al.*, 2009). Ésta ha sido aplicada en diferentes artículos enfocados en la construcción de ontologías para distintos dominios (Pribyl *et al.*, 2012; Tabares García y Jiménez Builes, 2014; Muñoz García *et al.*, 2014; John, 2014).

El dominio a modelar en la ontología, como se mencionó anteriormente, es la Orientación a Objetos (OO), la cual es un campo en la ciencia de la computación. La OO es un paradigma de programación organizado alrededor de objetos y datos. Es un tema importante en las carreras relacionadas con programación. Una ontología puede representar dominios del conocimiento pudiendo ser analizada por la computadora para la toma de decisiones. Esta representación es la base del conocimiento que un estudiante debe tener, de esta manera, los algoritmos pueden hacer inferencia para definir qué enseñar al estudiante.

Las clases y la jerarquía de la ontología fueron obtenidas a través de un análisis de un curso sobre POO, con el soporte de un grupo de profesores con experiencia en el área. El aspecto central para desarrollar la ontología fue la organización del conocimiento con respecto a la estructura del curso, considerando cuáles conceptos son necesarios para aprender otros conceptos. Se inició desglosando los conceptos más generales de la OO (abstracción, encapsulación, jerarquía y modularidad) (West, 2004; Kendal, 2010). Después, se agregaron los conceptos secundarios de la OO (conurrencia, persistencia y tipificado); así, hasta completar la construcción de la ontología.

La ontología parcial es presentada en la Figura 6.1. En la estructura se muestran las clases y sus relaciones semánticas, el concepto central es *orientacion_objetos*, todas las clases convergen en este concepto. Hay diferentes tipos de flechas que representan el tipo de relación tales como parte de (62.7%), es un (27.4%), tipo de (4.0%) y derivado de (5.9%).

La ontología completa fue desarrollada en el software Protégé¹, el cual es un editor de ontología gratuito además de un framework para construir sistemas inteligentes. Protégé es apoyada por una comunidad activa de profesores, gobierno y usuarios incorporados, quienes usan Protégé para crear soluciones basadas en conocimiento en áreas diversas como biomedicina, comercio en línea y modelado

¹<http://protege.stanford.edu/>

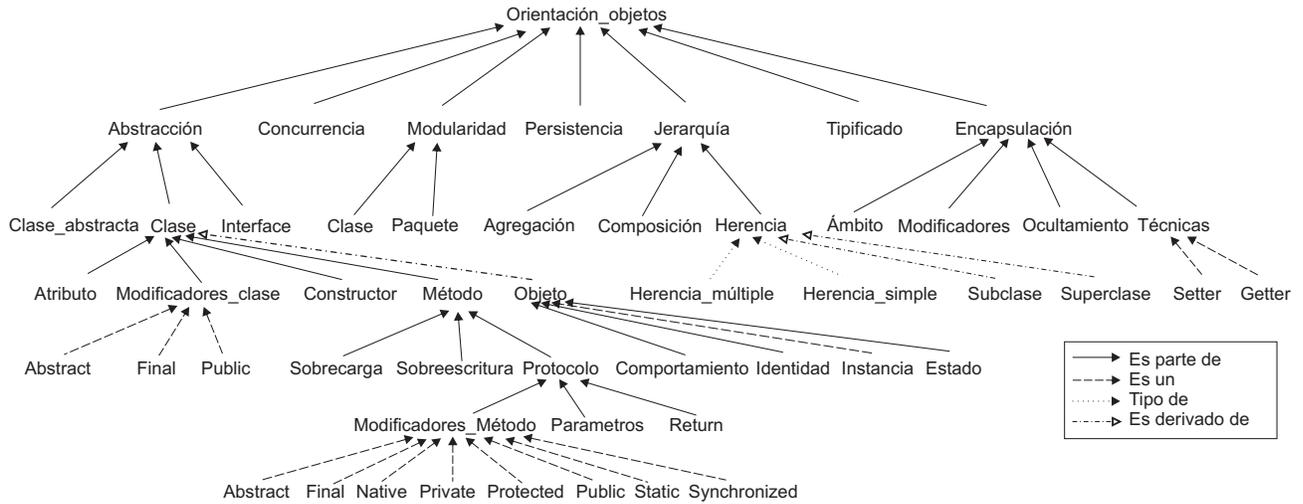


Figura 6.1: Ontología de la orientación a objetos

organizacional.

Para los fines de este estudio, la ontología fue reducida a la versión en la Figura 6.2, considerando solamente aquellas clases que están representadas por un artículo en Wikipedia. Esto, debido a que es necesario un contenido desarrollado del concepto para su minado. Esta sección puede apreciarse con más detalles en Ramírez-Noriega *et al.* (2017a).

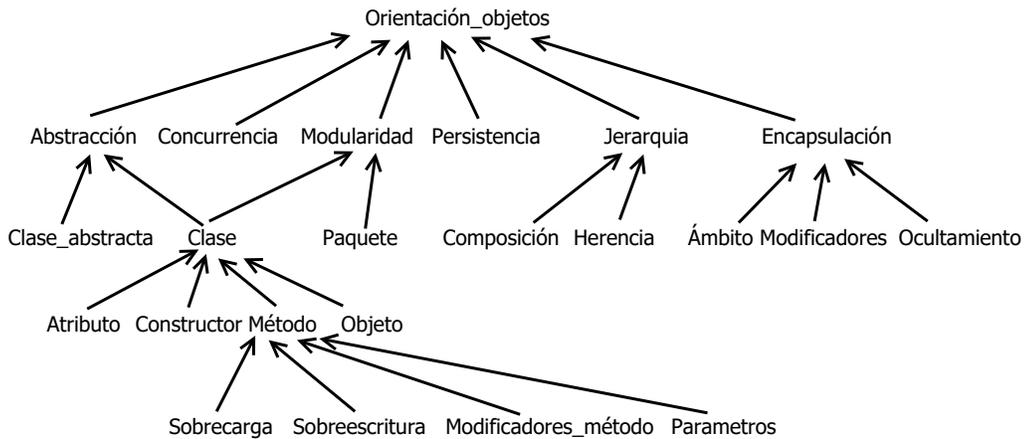


Figura 6.2: Sección empleada de la ontología de la orientación a objetos.

6.1.2. Generación de pesos por expertos

El objetivo de este segmento del experimento es determinar por expertos la importancia de un concepto particular para aprender otro concepto general.

6.1.2.1. Muestra

El instrumento está destinado a los profesores que enseñan el curso POO. La ontología representa una sección del conocimiento de este curso. El tipo de muestreo utilizado es no probabilístico porque la representación de la población no se requiere en este experimento. La muestra consideró a 20 profesores que imparten el curso en las universidades de México.

El tiempo programando de la muestra corresponde en su mayoría a profesores con experiencias de 5-6 años, sin embargo, el 20% de la muestra corresponde a profesores con más de 15 años de experiencias (ver Figura 6.3).

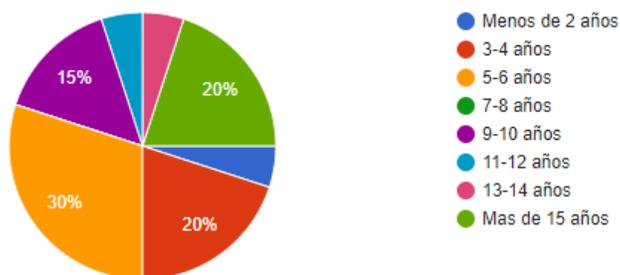


Figura 6.3: ¿Cuánto tiempo llevas programando en OO?

La Figura 6.4 representa el nivel de conocimiento que la muestra considera tener. La muestra está muy distribuida por lo que podemos considerar que los profesores se consideran a sí mismos con un conocimiento arriba del promedio del tema.

Los profesores tienen poco tiempo impartiendo la materia de POO, un gran porcentaje no pasa de los 6-7 años de experiencias, aun cuando entre la muestra el 15% tienen mucha experiencia (más de 14 años).

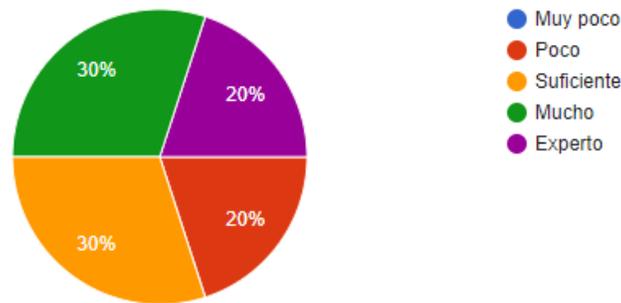


Figura 6.4: ¿Qué nivel de conocimiento consideras tener en OO?

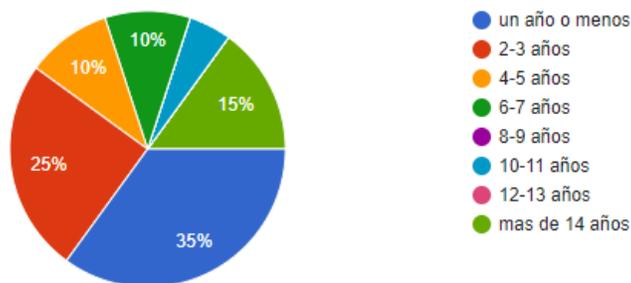


Figura 6.5: ¿Cuánto tiempo tienes impartiendo POO?

6.1.2.2. Encuesta

La encuesta determina en qué medida un concepto es importante para comprender otro. Se redactaron preguntas del tipo:

- ¿Hasta qué punto crees que el concepto *abstracción* es importante para entender el concepto *orientación – objetos*? (relación *abstracción_orientación – objetos*)
- ¿Hasta qué punto crees que el concepto *constructor* es importante para entender el concepto *clase*? (relación *constructor-clase*)

En el apéndice A.1 se despliega la encuesta completa.

6.1.2.3. Procedimiento

Cada relación de la ontología estableció una pregunta. La escala de evaluación consideró valores de uno a siete, donde uno fue considerado como el menos importante y siete como el más importante. El proceso del experimento consistió en los siguientes pasos:

1. Una lista de universidades que enseñan OO fue hecha para identificar a los profesores candidatos para contestar el cuestionario.
2. El instrumento fue creado con base en la ontología.
3. Se creó una versión en línea de la encuesta utilizando Google Forms².
4. Los profesores respondieron al cuestionario a través de Internet. No se consideró ningún límite de tiempo y no se requirió información personal de los participantes.
5. Finalmente, los datos obtenidos fueron analizados e interpretados.

6.1.3. Generación de pesos del método

Las relaciones obtenidas de la ontología fueron la entrada del método. Se evaluaron las mismas relaciones que los expertos para determinar el peso de cada clase. El archivo OWL de la ontología de la orientación a objetos fue la entrada del método.

6.1.3.1. Aspectos de implementación

El método propuesto fue implementado en el lenguaje de programación Java con el apoyo de la interfaz de programación JWPL (por sus siglas en inglés Java Wikipedia Library). JWPL proporciona acceso a toda la información de Wikipedia en diferentes idiomas de forma estructurada. Esto incluye un analizador de etiquetas de MediaWiki para el análisis de contenido de página en profundidad.

²<https://www.google.com/forms/about/>

JWPL ofrece métodos de acceso a propiedades tales como enlaces, plantillas, categorías, texto y otras propiedades (Zesch *et al.*, 2008).

6.1.4. Prueba estadística

Las variables *metodo* y *expertos* fueron consideradas. La variable *metodo* es el grado de relación numérico entre las clases generadas a través de nuestro algoritmo. La variable *expertos* es el grado de relación numérica entre las clases obtenido a través del grupo de maestros universitarios.

Las pruebas de Pearson y Spearman fueron consideradas para medir la correlación de los datos. Estas dos pruebas muestran si las variables están relacionadas y en qué grado lo están. Las pruebas Levene y Kolmogorov-Smirnov son necesarias para decidir si se emplea el test Pearson o Spearman. La prueba de Levene describe una situación en la cual el término de error es el mismo a través de todos los valores de las variables independientes, esto es llamado homocedasticidad. La prueba de Kolmogorov-Smirnov verifica que los datos tengan una distribución normal. Si las pruebas de Levene y Kolmogorov-Smirnov son aprobadas se usa la prueba de Pearson si no se usa la prueba de Spearman.

La prueba se basa en la hipótesis de que la combinación de los tres factores propuestos mostrará resultados cercanos a los valores estimados por el grupo de profesores. Esto se verifica mediante una prueba de correlación en la que se espera obtener una correlación superior a 0.8 con un nivel de confianza del 95 %.

Las hipótesis son:

- H_0 : Las variables *metodo* y *expertos* no tienen correlación.

- H_1 : Las variables *metodo* y *expertos* tienen correlación.

6.1.5. Resultados

6.1.5.1. Expertos vs método

Los resultados de los expertos y el método se muestran en la Tabla 6.1. Las relaciones se describen en la columna con el mismo nombre, se agregó un identificador (columna *Id*) para referencias posteriores. La opinión de los expertos con valores no normalizados se muestra en la columna *Valor*. La columna *Peso (expertos)* muestra el peso de la relación teniendo en cuenta la ecuación $weight(P, c_i) = c_i / \sum_{i=1}^m c_i$. Donde c_i es el valor del elemento i del conjunto C ($C = \{c_1, \dots, c_m\}$). c se refiere a los conceptos que componen otro concepto, donde todos los elementos de C componen el mismo concepto principal. En otras palabras, las variables *composicion* y *herencia* comparten el mismo hijo (*jerarquia*). Por lo tanto, el valor máximo se divide entre estas dos variables de acuerdo con la proporción del valor del experto.

Los valores devueltos por el método aparecen en la columna *Peso (Algoritmo)*. Estos valores no requieren ningún procesamiento; el método ya los muestra normalizados.

En la Figura 6.6 se representa una interpretación visual, este diagrama muestra los resultados de la Tabla 6.1. El eje x representa las relaciones considerando su identificador. El eje y representa el peso obtenido por cada relación. La línea continua representa los resultados normalizados de los expertos. La línea punteada representa los resultados del método.

6.1.5.2. Construcción de una tabla de probabilidad condicional

Esta parte muestra cómo se construye una TPC basada en el conocimiento de los expertos. La variable *encapsulación* fue considerada para este ejemplo. El algoritmo 1 y sus pesos mostrados en la Tabla 6.1 se consideran para este proceso. La variable *encapsulación* está formada por tres variables; *ámbito*, *ocultamiento* y *modificadores*. Por lo tanto, las relaciones creadas son *encapsulación – ámbito*, *encapsulación – ocultamiento*, y *encapsulación – modifiers*.

La variable *encapsulación* tiene tres padres, por lo tanto hay 3 relaciones con sus pesos. Esto representa la variable de entrada *peso* en el algoritmo 1. El valor para el estado *presente* es 8 (2^3), más

Tabla 6.1: Resultados de los expertos y el método propuesto (OO).

| Id | Relación | Expertos | | Algoritmo |
|----|-----------------------------|----------|------|-----------|
| | | Valor | Peso | Peso |
| 1 | OO-abstracción | 6.80 | 0.19 | 0.22 |
| 2 | OO-concurrencia | 4.70 | 0.13 | 0.12 |
| 3 | OO-encapsulación | 6.45 | 0.18 | 0.13 |
| 4 | OO-jerarquía | 6.50 | 0.19 | 0.07 |
| 5 | OO-modularidad | 5.90 | 0.17 | 0.08 |
| 6 | OO-persistencia | 4.60 | 0.13 | 0.04 |
| 7 | abstracción-clase_abstracta | 6.40 | 0.48 | 0.36 |
| 8 | abstracción-clase | 6.95 | 0.52 | 0.64 |
| 9 | modularidad-clase | 6.55 | 0.51 | 0.61 |
| 10 | modularidad-paquete | 6.35 | 0.49 | 0.39 |
| 11 | jerarquía-composición | 6.05 | 0.47 | 0.44 |
| 12 | jerarquía-herencia | 6.85 | 0.53 | 0.56 |
| 13 | encapsulación-ámbito | 6.25 | 0.33 | 0.24 |
| 14 | encapsulación-ocultamiento | 6.40 | 0.34 | 0.43 |
| 15 | encapsulación-modificadores | 6.25 | 0.33 | 0.33 |
| 16 | clase-atributo | 6.90 | 0.25 | 0.14 |
| 17 | clase-constructor | 6.75 | 0.25 | 0.20 |
| 18 | clase-método | 6.90 | 0.25 | 0.29 |
| 19 | clase-objeto | 6.70 | 0.25 | 0.38 |
| 20 | método-sobrecarga | 6.30 | 0.25 | 0.23 |
| 21 | método-sobreescritura | 6.05 | 0.24 | 0.26 |
| 22 | método-modificadores_método | 6.25 | 0.25 | 0.22 |
| 23 | método-parámetros | 6.80 | 0.27 | 0.30 |

8 para el estado *ausente*. La matriz de soporte (*matrix*) representada en el algoritmo 1 se muestra en la Tabla 6.2.

Los valores 0.24, 0.43, y 0.33 de la Tabla 6.2 se toman de la Tabla 6.1. Estos valores pueden apreciarse cuando una variable es *presente* pero los otros dos son *ausente*. Por ejemplo, cuando las variables *ámbito* y *ocultamiento* son *ausente* pero la variable *modificadores* es *presente*, entonces el valor de *encapsulación* es 0.33 (0 + 0 + 0.33, números en negrita), igual a la relación 15 en la Tabla 6.1.

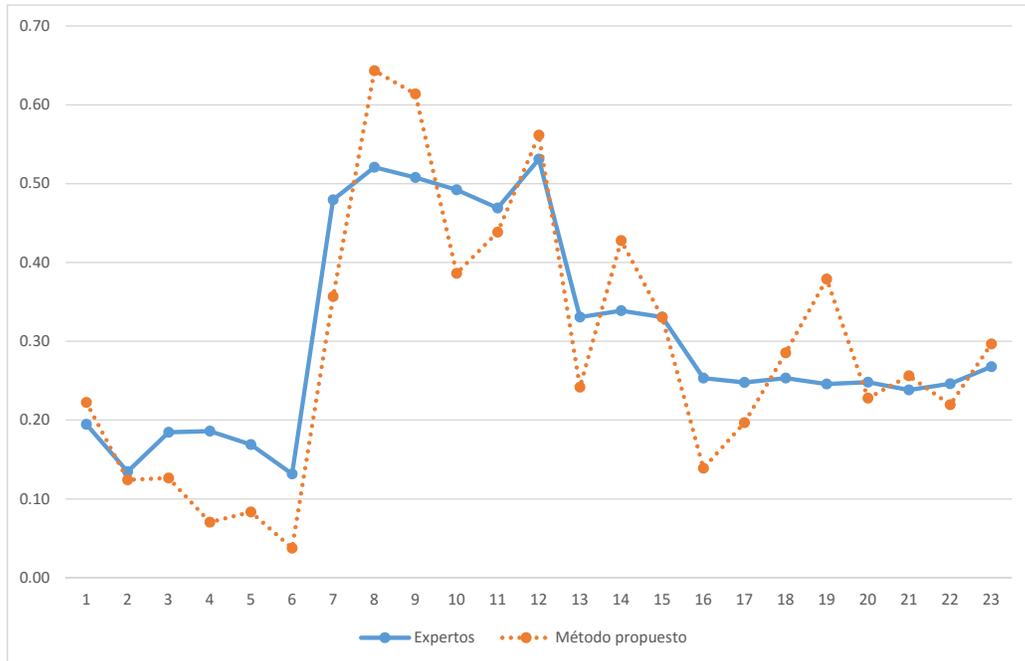


Figura 6.6: Comparación entre los expertos y el método propuesto.

Tabla 6.2: Tabla de probabilidad condicional para la variable *encapsulación*.

| ámbito | presente | | | | ausente | | | |
|--------------------------|----------|------|---------|------|----------|------|-------------|-----|
| | 0.24 | 0.24 | 0.24 | 0.24 | 0.0 | 0.0 | 0.0 | 0.0 |
| ocultamiento | presente | | ausente | | presente | | ausente | |
| | 0.43 | 0.43 | 0.0 | 0.0 | 0.43 | 0.43 | 0.0 | 0.0 |
| modificadores | p | a | p | a | p | a | p | a |
| | 0.33 | 0.0 | 0.33 | 0.0 | 0.33 | 0.0 | 0.33 | 0.0 |
| presente (Σ) | 1.0 | 0.67 | 0.57 | 0.24 | 0.76 | 0.43 | 0.33 | 0.0 |
| ausente ($1 - \Sigma$) | 0.0 | 0.33 | 0.43 | 0.76 | 0.24 | 0.57 | 0.67 | 1.0 |

Los valores para obtener el estado *presente* de la TPC se obtienen a través de una suma de columnas como se muestra en la Tabla 6.2. El valor *ausente* se obtiene a través del complemento del valor *presente*. Se construye una TPC para cada variable de red siguiendo la estructura de la ontología.

6.1.5.3. Prueba estadística

Las variables *metodo* y *expertos* aprobaron la prueba de Levene y Kolmogorov-Smirnov con un nivel de confianza superior al 95%. Por lo tanto, utilizamos la prueba de Pearson.

La prueba de Pearson mostró una correlación entre las variables de 0.895 (N=23). La prueba arrojó un *p* – *valor* por debajo del nivel de significancia ($0.001 < 0.05$). Por lo tanto, aceptamos H_1 ; las variables *metodo* y *expertos* tienen correlación.

6.2. Caso de estudio B: Tipos de lenguajes de programación

Esta sección desarrolla un estudio de caso para emplear la propuesta de investigación. El estudio de caso considera tres aspectos:

1. ¿Cuál es el dominio del conocimiento?
2. ¿Cuál es la estructura base para construir la parte cualitativa de la RB?
3. ¿Cuál es la fuente del conocimiento para crear la parte cuantitativa de la RB?

Las preguntas uno y dos están relacionadas porque el dominio del conocimiento debe estar representado en una estructura con nodos y relaciones. Para este estudio de caso, el dominio es tipos de lenguaje de programación, el cual es un tema importante para los estudiantes que comienzan clases de programación. Abuhassan (Abuhassam y AlMashaykhi, 2012) desarrolló una ontología que considera en parte los tipos de lenguajes de programación. La ontología descrita en la Figura 6.7 se basa en la obra de Abuhassan, se eligió esta ontología porque se adapta a las necesidades reales de aprendizaje de los estudiantes de programación, es una ontología desarrollada por otro autor que demuestra que el método puede reutilizar ontologías hechas para otros fines.

La ontología original tiene un mayor número de clases y relaciones, para este artículo la ontología se redujo a una sola parte. Para complementar la ontología, se añadió un lenguaje a la programación orien-

tada a objetos (C#), dos lenguajes a la programación estructural (Ada y Algol) y dos a la programación lógica (Datalog y Mercury).

La fuente de conocimiento, correspondiente a la pregunta 3, es Wikipedia³. La versión en inglés tiene más de cinco millones de artículos escritos organizados por categorías. Aunque la validez de la información de Wikipedia ha sido cuestionada, estudios como Limongelli *et al.* (2015) verificaron un nivel de precisión cercano a la Enciclopedia Británica⁴.

Wikipedia tiene una extensa comunidad científica que analiza su corpus de información y estructura desde su aparición en 2001 (Biuk-aghai y Ng, 2014). Varios autores han mostrado la importancia de Wikipedia en el campo educativo (Forte y Bruckman, 2006) empleándolo para: aprender en parejas, aprendizaje colaborativo, intercambio de información, entre otras actividades (Hämäläinen *et al.*, 2009; Mareca y Bosch, 2011).

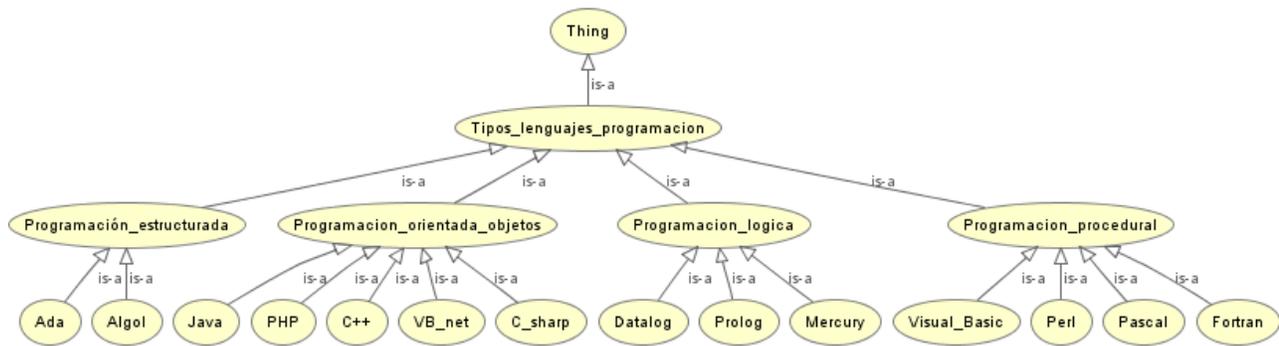


Figura 6.7: Ontología de los Tipos de Lenguajes de Programación.

Antes de proceder con el estudio de caso, explicaremos algunos aspectos del tratamiento de la ontología y Wikipedia para el experimento.

6.2.1. Parte cualitativa de la red Bayesiana por medio de ontologías

La ontología es una estructura representada por nodos y relaciones, esta estructura cumple las características mínimas para ser la parte cualitativa de la RB. A continuación se describe cómo es el

³<https://en.wikipedia.org>

⁴<https://www.britannica.com/>

proceso de adaptación al modelo.

6.2.1.1. Obtención de variables e identificación de estados

Las variables representan el conocimiento y se miden con una escala de 0 a 1; estas representan probabilidades de poseer conocimiento o la falta de él. Una variable necesita dos estados; el estado *presente* determina el grado de conocimiento del estudiante, y el estado *ausente* representa lo opuesto al primer estado (Ramírez-Noriega *et al.*, 2017c).

Teniendo en cuenta la representación de la ontología (definición 3.7) y de la RB (definición 3.4); existe una relación directa entre las clases de la ontología, conjunto C (definición 3.8), y las variables de la RB, conjunto N (definición 3.1). Cada conjunto representa la unidad fundamental en cada modelo, son similares y se puede establecer una relación como $N = C$. Donde $C = \{c_1, c_2, \dots, c_{|C|}\}$, por lo tanto, $N = \{n_1 = c_1, n_2 = c_2, \dots, n_{|C|} = c_{|C|}\}$.

Hay dos estados predefinidos para medir el conocimiento del estudiante; *presente* y *ausente*, cada variable contiene ambos estados.

6.2.1.2. Relaciones entre variables

Esta sección clasifica las variables en niveles de granularidad formando una escala jerárquica. Las relaciones entre variables en una RB (definición 3.5) se pueden obtener a través del conjunto T de la ontología (definición 3.11). El elemento $db_i = (m \leftarrow c_1, n \leftarrow c_2)$ donde $c_1, c_2 \in t_i; t_i \in T; db_i \in DB$.

El tipo de relación de la RB no es necesaria porque las relaciones en esta estructura son siempre causales, por lo que el método no considera la información semántica.

6.2.2. Parte cuantitativa de la red Bayesiana por medio de Wikipedia

Wikipedia cumple con los criterios mínimos de nuestro modelo (definición 5.1); artículos que explican los conceptos a través de su contenido con enlaces a otros conceptos relacionados, y estos artículos tienen un título. La probabilidad de encontrar un concepto dentro del índice de Wikipedia

es alta debido al gran número de artículos que posee; esto hace que Wikipedia sea una fuente de conocimiento apropiado para nuestras necesidades.

6.2.2.1. Identificación de las variables del modelo

La parte cuantitativa requiere identificar los elementos básicos del modelo en la ontología; r , h , y P , estas variables representan el conjunto X . El elemento central de la ontología se denomina raíz r . Los nodos que actúan como un rango en una posición dada se consideran nodo hijo h , el grupo de nodos que sirven como dominio del mismo rango se consideran como el conjunto P . El conjunto $X = \{r, h, P\}$ representa un subconjunto de la ontología, por lo que la ontología se divide en subconjuntos X para determinar el peso de las relaciones.

Considerando la Figura 6.7, se muestra el siguiente ejemplo. La variable raíz r sería el nodo *Tipos_lenguajes_programacion*, la variable secundaria h sería el nodo *Programacion_procedural*, y los padres serían $P = \{Fortran, Perl, Visual_Basic, Pascal\}$. De esta manera, podemos aplicar el método para generar la TPC para h mediante el cálculo de relaciones.

6.2.3. Generación de pesos por expertos

Esta sección define como se obtuvo la opinión del experto en el dominio a través de una encuesta.

6.2.3.1. Muestra

El muestreo aplicado es no probabilístico porque la representatividad de la población no se requiere en este experimento. La muestra consideró 26 a profesores que imparten cursos relacionados con la programación en universidades de México. Este tipo de muestra es el más adecuada para evaluar los resultados porque el dominio es educativo. Fenz (2009) aplicó un método de evaluación similar para validar su trabajo.

Los profesores encuestados en su mayoría (76.9%) están entre 36 y 50 años, lo que indica que no son profesores jóvenes recién egresados de la universidad (ver Figura 6.8). El tiempo de experiencia de

los profesores impartiendo clases en universidad está muy distribuido teniendo profesores que apenas empiezan (3.8 %) mientras otros cuentan con más de 25 años de actividad (11.5 %) (ver Figura 6.9).

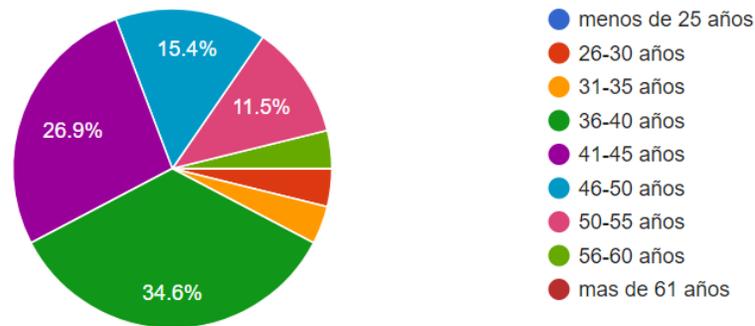


Figura 6.8: ¿A qué rango de edad perteneces?

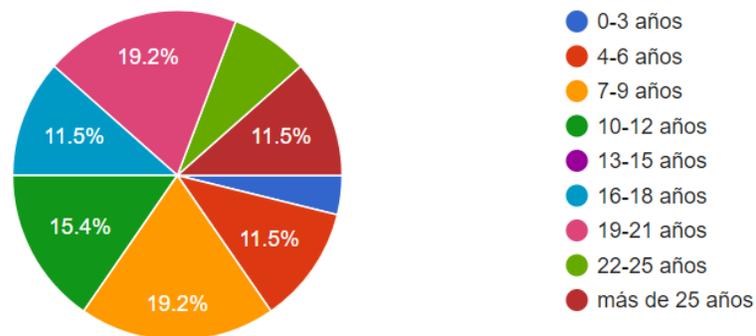


Figura 6.9: ¿Cuántos años tienes trabajando como docente en nivel universitario?

Por otra parte, la mayoría de los profesores encuestados dijeron tener experiencia en lenguajes como C++, Java, Pascal y C#, en ese orden. Mientras, sus lenguajes de programación favoritos son Java, C# y PHP, en ese orden.

6.2.3.2. Encuesta

Se diseñó una encuesta considerando las clases y las relaciones en la ontología. Una relación está formada por dos clases y una relación semántica; de esta forma, cada relación representa una pregunta: ¿Hasta qué punto distribuye el tiempo de lenguaje p para enseñar el tema h ?, donde p y h son conceptos de la ontología.

La encuesta consideró las preguntas dentro de un caso hipotético, en el que se pidió a los maestros que enseñaran un curso a un grupo de estudiantes sobre diferentes tipos de lenguajes de programación. Este curso describía los paradigmas de programación, así como los principales lenguajes, considerando su importancia, características y funcionamiento en aspectos teóricos.

El objetivo fue determinar una proporción generalizada entre temas y subtemas del contenido basado en la estructura ontológica. La encuesta no consideró el tiempo en horas para evitar inconsistencias o contradicciones, el instrumento manejó el tiempo en porcentaje (respuestas de 0 a 100); por lo tanto, la encuesta puede adaptarse fácilmente a cualquier tiempo. Para más detalles de la encuesta se puede consultar el apéndice A.2.

6.2.3.3. Procedimiento

El proceso de la encuesta consistió en los siguientes pasos:

1. Una lista de universidades fue hecha para enviar la encuesta a profesores del área de computación.
2. El instrumento fue creado con base en la ontología considerando una pregunta por cada relación de la ontología.
3. Se creó una versión en línea de la encuesta utilizando Google Forms⁵.
4. Los profesores respondieron al cuestionario a través de Internet. No se consideró ningún límite de tiempo y no se requirió información personal de los participantes.
5. Finalmente, los datos obtenidos fueron analizados e interpretados con el software estadístico SPSS⁶.

⁵<https://www.google.com/forms/about/>

⁶<http://www.ibm.com/analytics/us/en/technology/spss/>

6.2.4. Generación de pesos por el método

6.2.4.1. Procedimiento

Las relaciones obtenidas de la ontología fueron la entrada del método. Se evaluaron las mismas relaciones que los expertos para determinar el peso de cada clase. El archivo OWL de la ontología de la orientación a objetos fue la entrada del método.

6.2.4.2. Aspectos de implementación

El método propuesto fue implementado en el lenguaje de programación Java con el apoyo de la interfaz de programación JWPL (por sus siglas en inglés Java Wikipedia Library).

6.2.5. Prueba de hipótesis

La prueba se basa en la hipótesis de que la combinación de los tres factores propuestos mostrará resultados cercanos a los valores estimados por los profesores. La prueba de Pearson verifica tal similitud, este estudio tiene como objetivo obtener una correlación mayor de 0.8 con un nivel de confianza de 95 %.

La hipótesis considera dos variables, la primera variable *metodo* se define como el grado de relación cuantitativa entre las clases generadas por el método propuesto, la segunda variable *expertos* se define como el grado de relación cuantitativa entre las clases obtenidas por el grupo de la universidad profesores

Las hipótesis son las siguientes:

- H_0 : Las variables *metodo* y *expertos* no tienen correlación.
- H_1 : Las variables *metodo* y *expertos* tienen correlación.

6.2.6. Resultados

6.2.6.1. Expertos vs método

Los datos de la Tabla 6.3 muestran los resultados del experimento. La tabla muestra las relaciones de la ontología en la figura 6.7, hay 18 pares de relaciones que organizan la información de los expertos y el método propuesto. El apéndice B.1 muestra con detalle como se obtuvieron los valores de la tabla.

Los resultados de la Tabla 6.3 se trazan en la Figura 6.10. El eje x tiene las relaciones evaluadas con sus identificadores y el eje y tiene los valores alcanzados en la evaluación. La línea continua representa a los expertos y la línea punteada representa nuestro método. El caso ideal es que los puntos se traslapen, sin embargo, algunos casos son similares y otros no.

Tabla 6.3: Resultados de los expertos y el método propuesto (Tipos de lenguajes de programación).

| Id | Relaciones | Método propuesto | Expertos |
|----|---|------------------|----------|
| 1 | Tipos_lenguajes_programación-Programación_estructurada | 0.21 | 0.20 |
| 2 | Programación_estructurada-Algol | 0.47 | 0.53 |
| 3 | Programación_estructurada-Ada | 0.53 | 0.47 |
| 4 | Tipos_lenguajes_programación-Programación_procedural | 0.18 | 0.20 |
| 5 | Programación_procedural-Perl | 0.21 | 0.13 |
| 6 | Programación_procedural-Fortran | 0.27 | 0.13 |
| 7 | Programación_procedural-VB | 0.27 | 0.27 |
| 8 | Programación_procedural-Pascal | 0.25 | 0.47 |
| 9 | Tipos_lenguajes_programación-Programación_orientada_objetos | 0.34 | 0.41 |
| 10 | Programación_orientada_objetos-VB.NET | 0.12 | 0.06 |
| 11 | Programación_orientada_objetos-C# | 0.20 | 0.16 |
| 12 | Programación_orientada_objetos-C++ | 0.27 | 0.2 |
| 13 | Programación_orientada_objetos-Java | 0.25 | 0.46 |
| 14 | Programación_orientada_objetos-PHP | 0.16 | 0.12 |
| 15 | Tipos_lenguajes_programación-Programación_Lógica | 0.27 | 0.19 |
| 16 | Programación_Lógica-Prolog | 0.52 | 0.74 |
| 17 | Programación_Lógica-Datalog | 0.27 | 0.11 |
| 18 | Programación_Lógica-Mercury | 0.21 | 0.15 |

6.2.6.2. Red Bayesiana generada

La Figura 6.11 representa la RB construida sobre la base de los resultados de nuestro método. El peso de las relaciones en la Tabla 6.3 se toman para construir las TPCs, los pesos se muestran en la

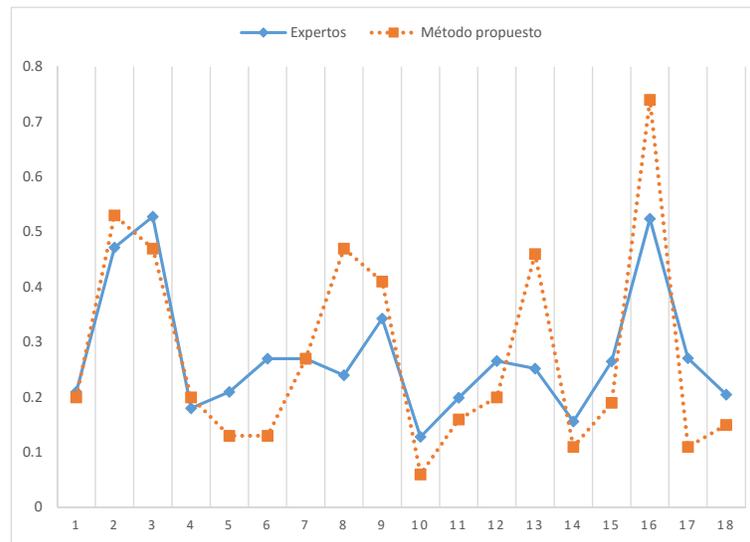


Figura 6.10: Comparación entre el método propuesto y los expertos.

figura sobre las relaciones entre los nodos. Esta red contiene las mismas clases de la ontología y la misma estructura, pero se han agregado las TPCs para convertirla en una RB.

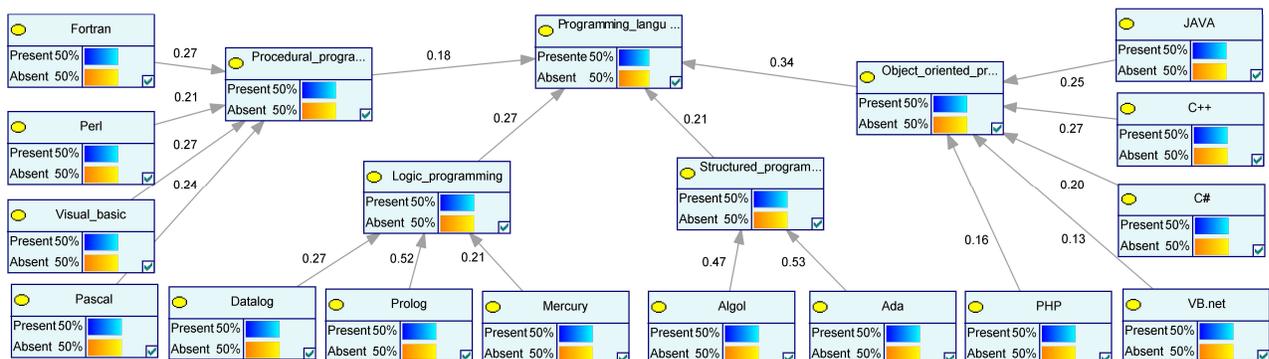


Figura 6.11: Red Bayesiana generada con los resultados del método propuesto.

6.2.6.3. Prueba estadística

La prueba de Pearson mostró una correlación de 0.812 entre las variables ($N=18$). La prueba arrojó un p – valor por debajo del nivel de significancia ($0.000 < 0.05$); por lo tanto, aceptamos H_1 : Las variables *metodo* y *expertos* tienen correlación.

6.3. Resumen del capítulo

Este capítulo probó el método propuesta en un ambiente real bajo dos casos de estudio. El primer caso se enfocó en validar la construcción de las TPCs a través de una ontología construida ad hoc. El segundo caso validó la construcción de la parte cualitativa y cuantitativa de la RB con una ontología tomada de la literatura. Para ambos casos se empleo la prueba de Pearson para determinar correlación en la propuesta y expertos del dominio; para el primer caso profesores de programación orientada a objetos y para el segundo profesores con relación al tema lenguajes de programación.

Capítulo 7

Discusión de los resultados obtenidos

7.1. Caso de estudio A: Programación orientada a objetos

El método se enfrenta a la difícil tarea de determinar con precisión los valores de la misma manera que los expertos. La decisión de los expertos se basa en la subjetividad y, a menudo, no muestra acuerdo en sus opiniones. Por esta razón, sus opiniones se promediaron para obtener un consenso en general.

Una correlación de 0.895 entre los resultados de los expertos y el método fue obtenida aplicando la prueba de Pearson; este valor representa una alta correlación positiva en el rango $[-1,1]$. También se obtuvo una confiabilidad superior al 99%. Una correlación mayor a 0.8 es aceptable considerando la subjetividad que aparece en las estimaciones de las relaciones (Hernández Sampieri *et al.*, 2014).

El método utilizado se justifica por el conocimiento implícito dentro de Wikipedia. Los artículos tienen conocimiento representado por su información textual, además de una estructura jerárquica que brinda información en cuanto a la dependencia y cercanía de los conceptos de los artículos. La combinación adecuada de estos factores proporciona una medida de la relación entre las clases. El conocimiento en Wikipedia es generado y evaluado por personas con experiencia en el dominio que llegaron a un consenso sobre el contenido de los artículos; por lo tanto, se puede establecer un grado de relación entre lo que un experto piensa (expertos encuestados) y lo que otros escriben (información de

Wikipedia).

Los resultados obtenidos indican que la parte cuantitativa de una RB puede construirse sobre la base de ontologías e información de Wikipedia. Esta estructura es útil para fines educativos con resultados similares a los expertos.

7.2. Caso de estudio B: Tipos de lenguajes de programación

Según los resultados de la Figura 6.10, el método estimó resultados similares al experto, sin embargo, se aplicó la prueba de Pearson para evitar la subjetividad de los resultados. La prueba mostró una correlación positiva de 0.812 entre los tratamientos; esto indica una correlación aceptable y mayor al límite establecido por la hipótesis (0.8).

La variación de los resultados de los expertos y el método representan la diferencia de opinión de los editores de los artículos de Wikipedia y la muestra de este estudio. La primera opinión es representativa de varios países, y la segunda opinión representa a profesores de México. Es difícil predecir con precisión las opiniones de la gente porque las mismas características pueden proporcionar diferentes interpretaciones por las personas. Sin embargo, cuanto más se acerca la estimación a la idea experta, mejor será el sistema de predicción.

La RB en la Figura 6.11 representa el resultado del proceso. El método puede generar RBs para ser utilizado como dominio de conocimiento en STIs. Las investigaciones (Ting *et al.*, 2013; Khodeir *et al.*, 2012b; Ramirez-Noriega *et al.*, 2016) se han basado en RBs para representar el conocimiento del dominio de los STIs; esto es útil para reducir el tiempo de construcción de tales sistemas inteligentes.

7.3. Implicaciones en general

A continuación se muestran las implicaciones de ambos casos de estudio, los cuales componen el método propuesto.

7.3.1. Implicaciones prácticas

El método contribuye al área de tutores inteligentes automatizando la construcción de RBs. La construcción de los STIs es laboriosa, sin embargo, con métodos automatizados como la propuesta, los tiempos de producción se acortan. Esta propuesta impacta de diferentes maneras los módulos básicos de un STI:

- El módulo de dominio recibe el mayor impacto generando la estructura de conocimiento (RB) que representa el tema a ser enseñado.
- El módulo de estudiante se beneficia representando el conocimiento del alumno en la estructura del conocimiento.
- Finalmente, el módulo tutor se basa en el conocimiento del estudiante representado en la RB para tomar decisiones sobre la próxima acción en el proceso de enseñanza del STI.

7.3.2. Implicaciones teóricas

Este trabajo formalizó la construcción de una RB basado en una ontología. El método construyó la parte cualitativa (variables y relaciones) de la RB conservando la misma estructura de la ontología, lo cual se aprecia en las Figuras 6.7 y 6.11 donde los mismos nodos y los enlaces se conservan en ambas estructuras.

La ontología y la información de Wikipedia generaron la parte cuantitativa de la RB, lo que permitió construir TPCs de manera similar a los expertos. El método propuesto mantiene el estándar OWL creando la parte cuantitativa de una fuente externa; esto permite reutilizar otras ontologías. El estándar OWL se respeta cuando se consideran sólo las clases de la ontología con su dominio y rango, estos elementos pertenecen al estándar OWL por defecto; por lo tanto, la propuesta no requiere otro aspecto que no se tenga en cuenta en el estándar OWL.

En la revisión de la literatura no hubo precedentes para construir RBs automáticamente con un enfoque en la educación. Los trabajos relacionados en otras áreas mostraron limitaciones en no generar

las TPCs (Devitt *et al.*, 2006; Fenz *et al.*, 2009; Ishak *et al.*, 2011) completamente automatizadas; los autores que automatizaron el proceso rompieron el estándar OWL (Yang y Calmet, 2005; Yan y Wei, 2013), disminuyeron el dominio Bucci *et al.* (2011), o sus propuestas son poco prácticas para representar un dominio educativo (Andrea y Franco, 2009). El método propuesto trató estos problemas construyendo las TPCs automáticamente, respetando el estándar OWL, y enfocado en un campo educativo.

Capítulo 8

Conclusiones, aportaciones y trabajo futuro

8.1. Conclusiones

Este estudio desarrolló un método para construir la parte cualitativa y cuantitativa de una RB para STIs. En los estudios de caso, el método consideró la minería de texto en artículos de Wikipedia para construir la parte cuantitativa (caso A y B) y las clases y relaciones de ontologías para construir la parte cualitativa (caso B). El experimento comparó la evaluación de expertos del área educativa elegida con los resultados del método, la prueba de Pearson obtuvo una correlación entre ambos tratamientos de 0.895 para el primer caso y 0.812 para el segundo.

Esta investigación desarrolló un modelo de una ontología a través de la teoría de conjuntos descomponiendo la ontología en sus elementos básicos: clases, propiedades de objetos, propiedades de datos y tipos de propiedades de datos; para después formar la ontología a través de la relaciones entre estos elementos.

La relación cuantitativa entre conceptos fue definida con base en tres factores: Frecuencia del concepto (positivo), saltos del concepto padre hasta el concepto hijo (negativo), incidencia de conceptos en el contenido de ambos artículos (positivo), estos conceptos mostraron las correlaciones más cercanas con las opiniones de los expertos ante otros factores considerados. Con estos factores se propuso un

algoritmo para construir las TPCs.

La validación de la parte cuantitativa fue hecha con expertos del dominio, la carencia de conjuntos de datos para este problema tan particular llevó a dicha decisión. Los resultados se validaron considerando correlación de los valores entre conceptos determinados por los expertos y el método.

La parte cualitativa fue diseñada con el modelo de la ontología propuesto y el modelo de la RB de la literatura. Se consideraron las clases de la ontología para obtener las variables de la RB y el rango y dominio de las clases de la ontología para obtener las relaciones entre variables de la RB. De esta forma se obtuvo la parte cualitativa de la RB.

Distintos algoritmos desarrollados, sin embargo, fueron embebidos en un solo modelo formalizado matemáticamente, de esta forma puede ser implementado en cualquier lenguaje de programación.

Este trabajo estableció una relación entre los expertos del dominio y el método propuesto. Los expertos fueron profesores que dieron su opinión sobre la importancia de los conceptos en el aprendizaje de otros. El método consistió en extraer la estructura de Wikipedia y sus artículos basados en las clases y relaciones de una ontología.

De acuerdo con los resultados, este estudio afirma que las frecuencias, saltos y coincidencias entre conceptos permiten obtener medidas entre conceptos similares a expertos del mismo dominio. La prueba de Pearson mostró una precisión en promedio de 0.854 entre los resultados del método propuesto y los resultados de los expertos, esta correlación permite construir TPCs con alto grado de fiabilidad para ser utilizado en STIs.

La educación tiene fases muy teóricas como la mayoría de los artículos de Wikipedia; esta información se puede utilizar para obtener el peso de las relaciones entre conceptos de una RB; la presente investigación se enfocó en analizar aspectos teóricos de los conceptos principalmente. Por otra parte, existen otras áreas donde no se podría aplicar la propuesta; por ejemplo: diagnosticar una enfermedad o asignar un crédito bancario, la información de Wikipedia es inadecuada porque no hay conocimiento específico o claro para construir RBs para este tipo de problemas; es necesario encontrar otras fuentes de información distinta para resolverlos.

8.2. Aportaciones

Las aportaciones más destacadas de esta investigación son:

- Un modelo matemático para representar una ontología (sección 3.2.4): Un modelo matemático fue propuesto para representar una ontología, fue desarrollado con el objetivo de descomponer una ontología en sus partes para los objetivos de este proyecto; sin embargo, puede ser reutilizado para otras necesidades.
- Un algoritmo para obtener una relación cuantitativa entre dos conceptos (sección 5.3.2.1): Una de las aportaciones más importantes fue desarrollar este algoritmo, nos permitió construir las TPCs de la RB para automatiza el proceso a través de tres factores.
- Un algoritmo para generar las TPCs (sección 5.3.2.2): Con los pesos de las relaciones calculadas, se desarrolló un algoritmo para construir las TPCs y construir la parte cuantitativa de la RB.
- La construcción automática de un módulo de un STI (Módulo de dominio): Los módulos del tutor, del estudiante, del dominio y de interfaces constituyen de forma genérica un STI. La aportación más impactante de nuestra investigación repercute en el modelo del dominio, sin embargo de forma indirecta también lo hace en el modelo del estudiante y del tutor.
- La construcción automática de cuatro fases de la metodología propuesta en Ramírez-Noriega *et al.* (2015). En la referencia citada, se establece una metodología que consta de seis fases para construir una RB de forma manual para el área educativa, nuestra propuesta automatiza las cuatro primeras fases.
- La construcción de RBs de forma automática para dominios educativos: No se encontraron otras investigaciones que automatizaran la construcción automática de la RB para la educación, sin embargo, si existen trabajos para otras áreas, mejorando esos métodos en nuestro dominio se construyó la RB sin modificar el estándar OWL y sin modificar el dominio original de la ontología.

Estas contribuciones fortalecen el desarrollo de los STIs basados en RBs mediante la automatización de parte del proceso de construcción; sin embargo, la propuesta puede ser empleada en otras áreas que basan el conocimiento en documentos de texto.

8.3. Trabajo futuro

A continuación se describen las líneas de trabajo que esta investigación ha generado:

- El modelo funciona con conceptos que tienen un artículo con contenido en Wikipedia: La enciclopedia seleccionada tiene un gran número de artículos de diferentes áreas; sin embargo, algunos temas muy especializados aún no están definidos. Pocas enciclopedias tienen más artículos que Wikipedia, sólo una búsqueda en Internet podría superar este tipo de desventaja ante la pérdida de la construcción colaborativa del artículo que ofrece Wikipedia. Esto ofrece un área de oportunidad para buscar sobre la Web los conceptos no desarrollados en Wikipedia.
- Construcción de las dos fases restantes del modelo: El método propuesto construye una parte importante para el modelo del dominio basado en RBs; sin embargo, para ser aplicado directamente en un STI, todavía es necesario desarrollar nodos de preguntas o actividades y relacionarlos con los conceptos de la red. Surge otra área de oportunidad para determinar un algoritmo para clasificar preguntas automáticamente a los nodos correspondientes.
- Mejora de la correlación con expertos: Aunque los resultados de esta investigación son alentadores, para el trabajo futuro se van a mejorar algunos aspectos del algoritmo para aumentar la correlación con los expertos.
- Construcción de un STI con un dominio basado en la propuesta de esta investigación: Esto brindaría la posibilidad de probar ante estudiante y de manera práctica los alcances del modelo, más allá de las opiniones de los expertos.

Bibliografía

- Abuhassam, I. y AlMashaykhi, A. (2012). Domain Ontology for Programming Languages. *Journal of Computations & Modelling*, 2(4):75–91.
- Álvarez Carmona, M. Á. (2014). *Detección de similitud semántica en textos cortos*. Tesis doctoral, Instituto Nacional de Astrofísica, Óptica y Electrónica.
- Andrade, J. L. J. (2003). BayesN: Un Algoritmo para Aprender Redes Bayesianas Clasificadoras a partir de datos. *Universidad Veracruzana*.
- Andrea, B. y Franco, T. (2009). Extending ontology queries with Bayesian network reasoning. En *International Conference on Intelligent Engineering Systems, INES 2009*, pp. 165–170.
- Arevalillo-Herráez, M., Arnau, D., y Marco-Giménez, L. (2013). Domain-specific knowledge representation and inference engine for an intelligent tutoring system. *Knowledge-Based Systems*, 49:97–105.
- Arola, K. y Wysocki, A. (2012). *Composing Media Composing Embodiment*. Utah State University Press.
- Baazaoui-zghal, H. y Ghezala, H. B. (2014). A Fuzzy-Ontology-Driven Method for A Personalized Query Reformulation. En *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1640–1647, Beijing, China.

- Badaracco, M. y Martínez, L. (2011). An intelligent tutoring system architecture for competency-based learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6882 LNAI(PART 2):124–133.
- Badaracco, M. y Martínez, L. (2013). A fuzzy linguistic algorithm for adaptive test in Intelligent Tutoring System based on competences. *Expert Systems with Applications*, 40(8):3073–3086.
- Baoyao, Z., Ping, L., Yuhong, X., y Wei, L. (2009). Wikipedia-graph based key concept extraction towards news analysis. *IEEE Conference on Commerce and Enterprise Computing, CEC 2009*, pp. 121–128.
- Baraldi, P., Podofilini, L., Mkrtychyan, L., Zio, E., y Dang, V. N. (2015). Comparing the treatment of uncertainty in Bayesian Networks and fuzzy expert Systems used for a Human reliability analysis application. *Reliability Engineering & System Safety*, 138:176–193.
- Baron, H. B., Rojas, M. M., Diaz, J. T., y Contreras, A. V. (2014). Graph Isomorphism in Fuzzy Cognitive Maps for Monitoring of Game-Based Learning. *2014 International Conference on Intelligent Networking and Collaborative Systems*, pp. 304–310.
- Ben Messaoud, M., Leray, P., y Ben Amor, N. (2013). Active learning of causal Bayesian networks using ontologies: A case study. *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8.
- Bhowmik, R. (2008). Keyword extraction from abstracts and titles. En *Conference Proceedings - IEEE SOUTHEASTCON*, pp. 610–617.
- Biuk-aghai, R. P. y Ng, K. K. (2014). A method for Automated Document Classification Using Wikipedia-Derived Weighted Keywords. En *Data and Software Engineering (ICODSE), 2014 International Conference on*, pp. 1–6, Bandung. IEEE.

- Bollegala, D., Matsuo, Y., e Ishizuka, M. (2011). A Web Search Engine-Based Approach to Measure Semantic Similarity between Words. *IEEE Transactions on knowledge and data engineering*, 23(7):977–990.
- Brown, S. (2004). Assessment for learning. *Learning and teaching in higher education*, (1):81–89.
- Bucci, G., Sandrucci, V., y Vicario, E. (2011). Ontologies and Bayesian networks in medical diagnosis. *Proceedings of the Annual Hawaii International Conference on System Sciences*, pp. 1–8.
- Budanitsky, A. y Hirst, G. (2006). Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Comput. Linguist.*, 32(1):13–47.
- Buzzi, M. y Leporini, B. (2009). Editing Wikipedia content by screen reader: easier interaction with the Accessible Rich Internet Applications suite. *Disability and rehabilitation. Assistive technology*, 4:264–275.
- Campos, I. S. y Mata, A. C. U. (2011). Analysis of the Debate on the Impact of Technological Media in Learning Processes. *Actualidades Investigativas en Educación*, 11(1):1–22.
- Carbonell, J. R. (1970). AI in CAI: an artificial intelligence approach to computer assisted instruction. *IEEE transaction on Man. Machine System*, 11:190–202.
- Castillo, E. J. A., Estrada, L. G., y Senti, V. E. (2015). Domain Knowledge Representation For Programming Teaching. *IEEE Latin America Transactions*, 13(5):1528–1533.
- Cataldi, Z. y Lage, F. J. (2010a). Dimensiones para Evaluacion de Sistemas Tutores Inteligentes. En *XII Workshop de Investigadores en Ciencias de la Computación*, pp. 715–719.
- Cataldi, Z. y Lage, F. J. (2010b). Modelado del Estudiante en Sistemas Tutores Inteligentes. *Revista Iberoamericana de Tecnología en Educación y Educación enTecnología*, 5:29–38.

- Chandrasegaran, S. K., Ramani, K., Sriram, R. D., Horváth, I., Bernard, A., Harik, R. F., y Gao, W. (2013). The evolution, challenges, and future of knowledge representation in product design systems. *Computer-Aided Design*, 45(2):204–228.
- Chang, Y. S., Hung, W. C., y Juang, T. Y. (2013). Depression diagnosis based on ontologies and bayesian networks. *Proceedings - 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013*, pp. 3452–3457.
- Cheah, W.-P., Kim, K.-Y., Yang, H.-J., Kim, S.-H., y Kim, J.-S. (2008). Fuzzy Cognitive Map and Bayesian Belief Network for Causal Knowledge Engineering: A Comparative Study. *The KIPS Transactions:PartB*, 15(2):147–158.
- Chrysafiadi, K. y Virvou, M. (2013). Student modeling approaches: A literature review for the last decade. *Expert Systems with Applications*, 40(11):4715–4729.
- Colace, F. y De Santo, M. (2006). A Tutoring Tool Based on Bayesian Approach. *Advanced Learning Technologies, 2006. Sixth International Conference on*, pp. 109–113.
- Conejo, R., Millán, E., Pérez, J., y Trella, M. (2001). Modelado del alumno : un enfoque bayesiano. *Revista Iberoamericana de Inteligencia Artificial*, 12:50–58.
- Correa, M., Bielza, C., Pamies-Teixeira, J., y Alique López, J. (2008). Redes Bayesianas vs redes neuronales en modelos para la predicción del acabado superficial.
- Cruz Ramírez, N. y Martínez Morales, M. (2007). Un algoritmo para generar redes Bayesianas a partir de Datos estadísticos. *Primer Encuentro Nacional de Computación*, 1.
- Danaparamita, M. y Gaol, F. L. (2014). Comparing Student Model Accuracy with Bayesian Network and Fuzzy Logic in Predicting Student Knowledge Level. *International Journal of Multimedia and Ubiquitous Engineering*, 9(4):109–120.

- Davis, R., Shrobe, H., y Szolovits, P. (1993). What is a knowledge representation? *AI magazine*, 14(1):17–33.
- De Bruyn, E., Mostert, E., y Van Schoor, a. (2011). Computer-based testing - The ideal tool to assess on the different levels of Bloom's taxonomy. *2011 14th International Conference on Interactive Collaborative Learning, ICL 2011 - 11th International Conference Virtual University, VU'11*, (September):444–449.
- De Medio, C., Gasparetti, F., Limongelli, C., Sciarrone, F., y Temperini, M. (2016). Mining prerequisite relationships among learning objects. *Communications in Computer and Information Science*, 618:221–225.
- Devitt, A., Danev, B., y Matusikova, K. (2006). Constructing Bayesian Networks Automatically using Ontologies. En *Second Workshop on Formal Ontologies Meets Industry (FOMI 2006)*, Trento, Italy. Applied Ontology 0.
- Ding, Z. y Peng, Y. (2004). A probabilistic extension to ontology language OWL. *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*, 00(C):1–10.
- Ding, Z., Peng, Y., y Pan, R. (2006). BayesOWL: Uncertainty modeling in semantic web ontologies. *Studies in Fuzziness and Soft Computing*, 204:3–29.
- Fenz, S. (2009). An ontology-based approach for constructing Bayesian networks. *Proceedings of the International Conference on Complex, Intelligent and Software Intensive Systems, CISIS 2009*, 73(i):712–717.
- Fenz, S., Tjoa, a. M., y Hudec, M. (2009). Ontology-based generation of bayesian networks. En *Proceedings of the International Conference on Complex, Intelligent and Software Intensive Systems, CISIS 2009*, pp. 712–717.

- Fernández-López, M., Gómez-Pérez, A., y Juristo, N. (1997). Methontology: from ontological art towards ontological engineering. *Assessment*, SS-97-06:33–40.
- Ferreira, H. N. M., Brant-ribeiro, T., Ara, R. D., Dorc, F. A., y Cattelan, R. G. (2016). An Automatic and Dynamic Student Modeling Approach for Adaptive and Intelligent Educational Systems using Ontologies and Bayesian Networks. En *2016 IEEE 28th International Conference on Tools with Artificial Intelligence*.
- Ferschke, O. (2014). *The Quality of Content in Open Online Collaboration Platforms Approaches to NLP-supported Information Quality Management in Wikipedia*. Tesis doctoral.
- Forte, A. y Bruckman, A. (2006). From Wikipedia to the classroom: exploring online publication and learning. En *Proceedings of the 7th International Conference on Learning Sciences*, pp. 182–188. International Society of the Learning Sciences.
- Fox, C. (1989). A stop list for general text. *ACM SIGIR Forum*, 24(1-2):19–21.
- Friss de Kereki Guerrero, I. (2003). *Modelo para la Creación de Entornos de Aprendizaje basados en técnicas de Gestión del Conocimiento*. Tesis doctoral, Universidad Politecnica de Madrid.
- Gasevic, D., Djuric, D., y Devedzic, V. (2009). *Model Driven Engineering and Ontology Development*. Springer-Verlag.
- Gómes Flechoso, A. J. (1998). *Inducción de conocimiento con incertidumbre en bases de datos relacionales borrosas*. Tesis doctoral, Escuela Técnica Superior de Ingenieros de Telecomunicación.
- Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5-6):907–928.
- Hämäläinen, H., Ikonen, J., y Porras, J. (2009). Using Wiki for collaborative studying and maintaining personal learning diary in a computer science course. *Proceedings - 2009 9th IEEE International Conference on Advanced Learning Technologies, ICALT 2009*, pp. 679–680.

- Hernández Sampieri, R., Fernández Collado, C., y Baptista Lucio, M. (2014). Análisis de datos cuantitativos. En Hill, M., editor, *Metodología de la Investigación*, capítulo 10, p. 632. 6ta edición.
- Hossein, M., Zarandi, F., Khademian, M., y Minaei-bidgoli, B. (2012). A Fuzzy Expert System Architecture for Intelligent Tutoring Systems : A Cognitive Mapping Approach. *Journal of Intelligent Learning Systems and Applications*, 4:29–40.
- Hossein Zadeh, P. y Reformat, M. Z. (2013). Fuzzy Semantic Similarity in Linked Data Using Wikipedia Infobox. En *IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), 2013 Joint*, pp. 395–400, Edmonton, AB. IEEE.
- Humphrey, W. S. (1997). *Introduction to the Personal Software Process*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
- Ishak, M. B., Leray, P., y Amor, N. B. (2011). Ontology-based generation of object oriented bayesian networks. *CEUR Workshop Proceedings*, 818:9–17.
- Jaccard, P. (1912). The Distribution of the Flora in the Alpine Zone. *New Phytologist*, 11(2):37–50.
- Jennex, M. E., Smolnik, S., y Croasdell, D. (2014). Knowledge Management Success in Practice. *2014 47th Hawaii International Conference on System Sciences*, pp. 3615–3624.
- Jiang, J. J. y Conrath, D. W. (1997). Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. En *In Proceedings of International Conference Research on Computational Linguistics (ROCLING X)*, p. 15, Taiwan.
- Jiang, Y., Zhang, X., Tang, Y., y Nie, R. (2015). Feature-based approaches to semantic similarity assessment of concepts using Wikipedia. *Information Processing & Management*, 51(3):215–234.
- Jindal, V., Bawa, S., y Batra, S. (2014). A review of ranking approaches for semantic search on Web. *Information Processing and Management*, 50(2):416–425.

- John, S. (2014). Development of an Educational Ontology for Java Programming (JLEO) with a Hybrid Methodology Derived from Conventional Software Engineering Process Models. *International Journal of Information and Education Technology*, 4(4):308–312.
- Kammerdiner, A. (2009). Bayesian networks Bayesian Networks. En Floudas, C. A. y Pardalos, P. M., editores, *Encyclopedia of Optimization SE - 32*, pp. 187–196. Springer US.
- Kavcic, A. (2004). Fuzzy student model in InterMediActor platform. En *26th International Conference on Information Technology Interfaces, 2004.*, pp. 297–302.
- Kellaghan, T. y Greaney, V. (2001). *Using assessment to improve the quality of education*. United Nations Educational, Scientific and Cultural Organization, Paris, France.
- Kendal, S. (2010). *Object Oriented Programming using Java*. Bookboon.
- Kesarwani, P. y Misra, A. (2013). Selecting Integrated Approach for Knowledge Representation by Comparative Study of Knowledge Representation Schemes. *International Journal of Scientific and Research Publications*, 3(2):1–5.
- Khan, J. A. y Kumar, S. (2014). Deep Analysis for Development of RDF , RDFS and OWL Ontologies with Protege. En *Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization*, pp. 1–6.
- Khodeir, N., Wanas, N., Hegazy, N., y Darwish, N. (2012a). Bayesian based student knowledge modeling in intelligent tutoring systems. *Proceedings - 2012 6th IEEE International Conference on E-Learning in Industrial Electronics, ICELIE 2012*, pp. 12–17.
- Khodeir, N., Wanas, N., Hegazy, N., y Darwish, N. (2012b). Bayesian based student knowledge modeling in intelligent tutoring systems. En *Proceedings - 2012 6th IEEE International Conference on E-Learning in Industrial Electronics, ICELIE 2012*, pp. 12–17.

- Kovalerchuk, B. (2013). Quest for Rigorous Intelligent Tutoring Systems under Uncertainty : Computing with Words and Images. pp. 685–690.
- Krudysz, G. A. ., Sam Li, J., y McClellan, J. H. . (2006). Web-Based Bayesian Tutoring System. En *Digital Signal Processing Workshop, 12th - Signal Processing Education Workshop, 4th*, pp. 129–134.
- Laitila, P. y Virtanen, K. (2016). Improving Construction of Conditional Probability Tables for Ranked Nodes in Bayesian Networks. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1691–1705.
- Larrañaga, M., Conde, A., Calvo, I., Elorriaga, J. a., y Arruarte, A. (2014). Automatic generation of the domain module from electronic textbooks: Method and validation. *IEEE Transactions on Knowledge and Data Engineering*, 26(1):69–82.
- Larrañaga, P. y Moral, S. (2011). Probabilistic graphical models in artificial intelligence. *Applied Soft Computing*, 11(2):1511–1528.
- Leacock, C. y Chodorow, M. (1998). *Combining Local Context and WordNet Similarity for Word Sense Identification*.
- Leacock, C., Miller, G. A., y Chodorow, M. (1998). Using corpus statistics and WordNet relations for sense identification. *Computational Linguistics*, 24(1):147–165.
- Leontidis, M. y Halatsis, C. (2009). Supporting learner’s needs with an ontology-based bayesian network. *Proceedings - 2009 9th IEEE International Conference on Advanced Learning Technologies, ICALT 2009*, pp. 579–583.
- Lesk, M. (1986). Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone. *Proceedings of the 5th annual international conference on Systems documentation*, pp. 24–26.

- Leyva-Vazquez, M. (2013). *Modelo de ayuda a la toma de decisiones basado en mapas cognitivos difusos*. Tesis doctoral, Universidad de las Ciencias Informáticas.
- Li, Y., Bandar, Z. a., y McLean, D. (2003). An approach for measuring semantic similarity between words using multiple information sources. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):871–882.
- Lima, S. y Stump, S. (2010). How to obtain knowledge evidences in a student model based on Bayesian Network. En *Information Systems and Technologies (CISTI), 2010 5th Iberian Conference on*, pp. 1–4.
- Limongelli, C., Gasparetti, F., y Sciarrone, F. (2015). Wiki course builder: A system for retrieving and sequencing didactic materials from Wikipedia. En *2015 International Conference on Information Technology Based Higher Education and Training, ITHET 2015*, p. 5.
- Lin, D. (1998). An Information-Theoretic Definition of Similarity. En *Proceedings of the Fifteenth International Conference on Machine Learnin*, pp. 296–304. Morgan Kaufmann Publishers Inc.
- Liu, B. (2011). *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Data-Centric Systems and Applications. Springer.
- Liu, H. y Singh, P. (2004). ConceptNet - a practical commonsense reasoning tool-kit. *BT Technology Journal*, 22(4):211–226.
- Liu, S.-H. y Lee, G.-G. (2013). Using a concept map knowledge management system to enhance the learning of biology. *Computers & Education*, 68:105–116.
- Liu, Z. y Wang, H. (2007). A Modeling Method Based on Bayesian Networks in Intelligent Tutoring System. En *Proceedings of the 11th International Conference on Computer Supported Cooperative Work in Design*, pp. 967–972.

- López Puga, J. (2012). Cómo Construir y Validar Redes Bayesianas con Netica. *Revista Electrónica de Metodología Aplicada*, 17:1–17.
- Mareca, P. y Bosch, V. A. (2011). Editing the Wikipedia: Its role in science education. En *6th Iberian Conference on Information Systems and Technologies (CISTI 2011)*, pp. 366–370.
- Maroto García, M. d. I. N. (2007). *Las Relaciones Conceptuales en la Terminología de los Productos Cerámicos y su Formalización Mediante un Editor de Ontologías*. Tesis doctoral, Universitat Jaume I Departament.
- Medelyan, O., Milne, D., Legg, C., y Witten, I. H. (2009). Mining meaning from Wikipedia. *International Journal of Human Computer Studies*, 67(9):716–754.
- Millán, E., Descalço, L., Castillo, G., Oliveira, P., y Diogo, S. (2013). Using Bayesian networks to improve knowledge assessment. *Computers & Education*, 60(1):436–447.
- Millán, E., Loboda, T., y Pérez-De-La-Cruz, J. L. (2010). Bayesian networks for student model engineering. *Computers and Education*, 55(4):1663–1683.
- Millán, E. y Pérez-De-La-Cruz, J. L. (2002). A Bayesian diagnostic algorithm for student modeling and its evaluation. *User Modeling and User-Adapted Interaction*, 12:281–330.
- Miller, G. A. (1995). WordNet: a lexical database for English. *ACM Communications of the ACM*, 38(11):39–41.
- Miller, G. A. y Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6(1):1–28.
- Milne, D. y Witten, I. H. (2008). Learning to link with Wikipedia. En *Proceeding of the 17th ACM Conference on Information and Knowledge Management (CIKM '08)*, pp. 509–518.
- Milne, D. y Witten, I. H. (2013). An open-source toolkit for mining Wikipedia. *Artificial Intelligence*, 194:222–239.

- Mishra, M., Mishra, V. K., y Sharma, H. R. (2012). Intellectual ability planning for intelligent tutoring system in computer science engineering education abstract. *Proceedings - 2012 3rd National Conference on Emerging Trends and Applications in Computer Science, NCETACS-2012*, pp. 26–30.
- Misirli, A. T. y Bener, A. B. (2014). Bayesian networks for evidence-based decision-making in software engineering. *IEEE Transactions on Software Engineering*, 40(6):533–554.
- Mkrtychyan, L., Podofilini, L., y Dang, V. N. (2016). Methods for building Conditional Probability Tables of Bayesian Belief Networks from limited judgment: An evaluation for Human Reliability Application. *Reliability Engineering and System Safety*, 151:93–112.
- Montoy Iribe, N. E. (2010). *Uso de Medidas de Similitud Semántica para Procesamiento de Información no Estructurada*. Tesis doctoral, Instituto Politécnico Nacional.
- Muñoz Garcia, A. C., Sandia Saldivia, B., y Monzòn Pàez, G. (2014). An ontological model of collaborative learning in interactive distance education. *Red de Revistas Científicas de América Latina, el Caribe, España y Portugal*, 18(61):449–460.
- Nakamura, T., Shirakawa, M., Hara, T., y Nishio, S. (2014). Semantic Similarity Measurements for Multi-lingual Short Texts Using Wikipedia. *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, pp. 22–29.
- Neapolitan, R. E. (1990). *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. John Wiley & Sons, Inc., New York, NY, USA.
- Ngo, T. L., Lee, H., y Mizukawa, M. (2011). Building basic level knowledge network. En *IEEE SSCI 2011: Symposium Series on Computational Intelligence - RISS 2011: 2011 IEEE Workshop on Robotic Intelligence in Informationally Structured Space*, pp. 113–118.
- Nguyen, T. A., Raspitzu, A., y Aiello, M. (2013). Ontology-based office activity recognition with

- applications for energy savings. *Journal of Ambient Intelligence and Humanized Computing*, 4(5):1–15.
- Noy, N. F. y McGuinness, D. L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. *Stanford Knowledge Systems Laboratory*, p. 25.
- Peña Ayala, A. (2007). *Un modelo del estudiante basados en mapas cognitivos*. Tesis doctoral, Instituto Politécnico Nacional.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc, San Mateo, CA, USA.
- Pérez Hernández, C. (2002). *Explotación de los corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento*. Málaga, España.
- Pilato, G., Augello, A., Missikoff, M., y Taglino, F. (2012). Integration of ontologies and Bayesian networks for maritime situation awareness. *Proceedings - IEEE 6th International Conference on Semantic Computing, ICSC 2012*, pp. 170–177.
- Pinto, H. S., Staab, S., y Tempich, C. (2004). DILIGENT : Towards a fine-grained methodology for DIstributed , Loosely-controlled and evolvinG Engineering of oNTologies. *16Th European Conference on Artificial Intelligence - Ecai*, pp. 393–397.
- Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, 40(3):211–218.
- Pribyl, P., Fábera, V., y Faltus, V. (2012). Domain-Oriented Ontology for ITS System. En *2012 ELEKTRO*, pp. 364–368.
- Pshenichny, C. a. (2014). Knowledge engineering in volcanology: practical claims and general approach. *Journal of Volcanology and Geothermal Research*, 286:78–92.
- Ramirez, C. y Valdes, B. (2009). A general knowledge representation model for the acquisition of skills and concepts. En *2009 8th IEEE International Conference on Cognitive Informatics*, pp. 412–417.

- Ramirez, C. y Valdes, B. (2012). A General Knowledge Representation Model of Concepts. En *Advances in Knowledge Representation*, pp. 43–76.
- Ramírez-Noriega, A., Juárez-Ramírez, R., Huertas, C., y Martínez-Ramírez, Y. (2015). A Methodology for building Bayesian Networks for Knowledge Representation in Intelligent Tutoring Systems. En Juárez Ramírez, R., Guerra García, C., y Oktaba, H., editores, *Congreso Internacional de Investigación e Innovación en Ingeniería de Software 2015*, pp. 124–133, San Luís Potosí.
- Ramírez-Noriega, A., Juárez-Ramírez, R., Jiménez, S., Inzunza, S., Navarro, R., y López-Martínez, J. (2017a). An Ontology of the Object Orientation for Intelligent Tutoring Systems. En *2017 5th International Conference in Software Engineering Research and Innovation*, pp. 1–10, Merida, Yucatan, México. IEEE.
- Ramírez-Noriega, A., Juárez-Ramírez, R., Jiménez, S., y Martínez-Ramírez, Y. (2017b). *Knowledge Representation in Intelligent Tutoring System*, pp. 12–21. Springer International Publishing, Egypt.
- Ramírez-Noriega, A., Juárez-Ramírez, R., Jiménez, S., Martínez-Ramírez, Y., y Armenta, J. (2017c). *Building a Bayesian Network for Object Oriented Programming with Experts' Knowledge*, pp. 267–276. Springer International Publishing, Cham.
- Ramirez-Noriega, A., Juarez-Ramirez, R., Martinez-Ramirez, Y., Samantha, J., y Sergio, I. (2016). *Using Bayesian Networks for Knowledge Representation and Evaluation in Intelligent Tutoring Systems*, volumen 1. Springer International Publishing Switzerland.
- Rau, M. a., Michaelis, J. E., y Fay, N. (2015). Connection making between multiple graphical representations: A multi-methods approach for domain-specific grounding of an intelligent tutoring system for chemistry. *Computers & Education*, 82:460–485.
- Resnik, P. (1995). Using Information Content to Evaluate Semantic Similarity in a Taxonomy. En *IJCAI'95 Proceedings of the 14th international joint conference on Artificial intelligence*, volumen 1, pp. 448–453, Montreal, Quebec, Canada. Morgan Kaufmann Publishers Inc.

- Rettinger, A., Lösch, U., Tresp, V., D'Amato, C., y Fanizzi, N. (2012). Mining the semantic web: Statistical learning for next generation knowledge bases. *Data Mining and Knowledge Discovery*, 24(3):613–662.
- Richardson, M. y Domingos, P. (2006). Markov logic networks. *Machine learning*, 62(1-2):107–136.
- Rivas Navarro, M. (2008). *Procesos cognitivos y aprendizaje significativo*. BOCM, Madrid.
- Rodrigues, A. N. y dos Santos, S. C. (2013). A Systems Approach to Managing Learning based on the Revised Blooms Taxonomy. *43rd Frontiers in Education Conference*.
- Rodrigues, F. H., Bez, M. R., y Flores, C. D. (2013). Generating Bayesian networks from medical ontologies. *2013 8th Computing Colombian Conference, 8CCC 2013*.
- Rodríguez, R. J. (2010). Herramientas informáticas para la representación del conocimiento Software. *Subjetividad y pocesos cognitivos*, 14(1712):217–232.
- Rodríguez García, D. y Dolado, J. (2007). Redes Bayesianas en la Ingeniería del Software. En *Técnicas cuantitativas para la gestión en la ingeniería del software*. Netbiblo.
- Rodríguez Mañanes, F. (2009). *Redes Bayesianas y redes neuronales como modelos del aprendizaje causal*. Tesis doctoral, Instituto de Neurociencias de Castilla Y León.
- Rollande, R. y Grundspenkis, J. (2012). Representation of study program as a part of graph based framework for tutoring module of intelligent tutoring system. *2012 2nd International Conference on Digital Information Processing and Communications, ICDIPC 2012*, pp. 108–113.
- Rosić, M., Glavinić, V., y Stankov, S. (2006). Intelligent tutoring interoperability for the new web. *Telecommunication Systems*, 32(2-3):193–207.
- Safia, B.-B. y Aicha, M. (2014). Poss-OWL 2: Possibilistic Extension of OWL 2 for an Uncertain Geographic Ontology. *Procedia Computer Science*, 35:407–416.

- Sánchez-bermejo, A. A. M. (2013). *Similitud semántica entre conceptos de Wikipedia*. Tesis doctoral, Universidad Carlos III de Madrid.
- Santhi, R., Priya, B., y Nandhini, J. (2013). Review of intelligent tutoring systems using bayesian approach. *CoRR*, abs/1302.7.
- Satar, A. (2012). Using of Intelligence Tutoring Systems For Knowledge Representation in Learning & Teaching Process. *Kufa for Mathematics and Computer*, 1(5):1–13.
- Sedki, K. y Beaufort, L. B. D. (2012). Cognitive Maps and Bayesian Networks for Knowledge Representation and Reasoning. En *24th International Conference on Tools with Artificial Intelligence*, pp. 1035–1040.
- Sen, S., Lesicko, M., Weiland, A., Gold, R., Li, Y., Hillmann, B., Li, T. J.-J., y Hecht, B. (2014). Wiki-Brain: Democratizing computation on Wikipedia. *Proceedings of the 10th International Symposium on Open Collaboration*, p. 10.
- Sharma, T. y Kelkar, D. (2012). A Tour Towards Knowledge Representation Techniques. *International Journal of Computer Technology and Electronics Engineering (IJCTEE)*, 2(2):131–135.
- Shishehchi, S. y Banihashem, S. Y. (2011). Learning Content Recommendation for Visual Basic . Net Programming Language based on Ontology. *Journal of Computer Science*, 7(2):188–196.
- Socorro, R., Simón, A., Valdés, R., Fernández, F. O., Rosete, A., Moreno, M., Leyva, E., y Pina, J. (2008). Las ontologías en la representación del conocimiento.
- Sure, Y., Staab, S., y Struder, R. (2004). On-To-Knowledge Methodology. *Handbook on Ontologies*, pp. 117–132.
- Tabares García, J. J. y Jiménez Builes, J. A. (2014). Ontology for the evaluation process in higher education. *Revista Virtual Universidad Católica del Norte*, (42):68–79.

- Taborda, H. (2010). Modelos bayesianos de inferencia psicológica: ¿Cómo predecir acciones en situaciones de incertidumbre? *Universitas Psychologica*, 9(2):495–507.
- Thirumuruganathan, S. y Huber, M. (2011). Building Bayesian Network based expert systems from rules. *2011 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 3002–3008.
- Tian, L. (2009). The design of domain knowledge model in teaching based on multi-agent. En *2009 2nd IEEE International Conference on Computer Science and Information Technology*, pp. 168–170.
- Ting, C.-Y., Cheah, W.-N., y Ho, C. C. (2013). Student Engagement Modeling Using Bayesian Networks. En *2013 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2939–2944.
- Tolosa, G. (2008). Introducción a la Recuperación de Información. pp. 1–149.
- Uglev, V. (2014). Implementation of Decision-making Methods in Intelligent Automated Educational System Focused on Complete Individualization in Learning. *AASRI Procedia*, 6:66–72.
- Victorio-Meza, H., Mejia-Lavalle, M., y Ortiz, G. R. (2014). Advances on Knowledge Representation of Intelligent Tutoring Systems. En *2014 International Conference on Mechatronics, Electronics and Automotive Engineering*, pp. 212–216.
- West, D. (2004). *Object Thinking*. Microsoft Press.
- Wu, Z. y Palmer, M. (1994). Verbs Semantics and Lexical Selection. En *Proceedings of the 32Nd Annual Meeting on Association for Computational Linguistics, ACL '94*, pp. 133–138, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Xiao-xuan, H. X.-x. H., Hui, W. H. W., y Shuo, W. S. W. (2007). Using Expert's Knowledge to Build Bayesian Networks. *International Conference on Computational Intelligence and Security Workshops (CISW 2007)*, pp. 220–223.

- Yamna, E., Mellouli, K., y Willemin, P. H. (2010). A multicriteria Bayesian intelligent tutoring system MBITS. En *Proceedings of the 2010 10th International Conference on Intelligent Systems Design and Applications, ISDA'10*, pp. 714–719.
- Yan, L. y Wei, C. H. (2013). Development of a novel asset management system for power transformers based on Ontology. En *2013 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pp. 1–6.
- Yang, K.-a., Yang, H.-j., Yang, J.-d., y Kim, K.-h. (2005). Bio-ontology Construction Using Object-oriented Paradigm 1. En *Proceedings of the 12th Asia-Pacific Software Engineering Conference (APSEC'05)*.
- Yang, W. T. y Kao, H. Y. (2011). Measuring semantic relatedness using wikipedia revision information in a signed network. *Proceedings - 2011 Conference on Technologies and Applications of Artificial Intelligence, TAAI 2011*, pp. 69–74.
- Yang, Y. Y. y Calmet, J. (2005). OntoBayes: An Ontology-Driven Uncertainty Model. *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, 1.
- Yengin, I. y Saerbeck, M. (2012). Structuring automated learning discussions using dialog games and cognitive maps. En *2012 International Conference on Information Technology Based Higher Education and Training, ITHET 2012*.
- Zesch, T., Müller, C., y Gurevych, I. (2008). Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary. En *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pp. 1646–1652, Marrakech, Morocco. European Language Resources Association (ELRA).

- Zhang, X., Asano, Y., y Yoshikawa, M. (2012). Mining and explaining relationships in Wikipedia. *IEICE Transactions on Information and Systems*, E95-D(7):1918–1931.
- Zhiqiang, L., Werimin, S., y Zhenhua, Y. (2009). Measuring Semantic Similarity between Words Using Wikipedia. *2009 International Conference on Web Information Systems and Mining*, pp. 251–255.
- Zipf, G. K. (1950). Human behavior and the principle of least effort. *Journal of Clinical Psychology*, 6(3):306.

Apéndices

Apéndice A

Encuestas

A.1. Encuesta sobre Orientación a Objetos

El objetivo de este instrumento es determinar la importancia que tiene un concepto (específico) para aprender otro concepto (general). Se trabaja con el dominio Orientación a Objetos (OO). Está destinada a profesores que imparten la materia de Programación Orientada a Objetos (POO). La Figura A.1 representa parte de la jerarquía de conocimiento establecida para el curso POO; en la figura, la orientación a objetos representa el concepto más general, conforme se profundiza en la jerarquía los conceptos se vuelven más específicos.

La encuesta trata de determinar la importancia que existe entre dos conceptos (relaciones binaria). Por ejemplo, ¿Qué tan importante es el concepto abstracción para comprender el concepto orientación a objetos (abstracción-orientación a objetos)? o ¿Qué tan importante es el concepto constructor para comprender qué es una clase (constructor-clase)?. Se establece una pregunta por cada relación binaria de la jerarquía propuesta. Además, se define una escala para su evaluación.

Jerarquía de conceptos sobre orientación a objetos

Las Figura A.1 y A.2 representan la ontología desarrollada en Protégé.

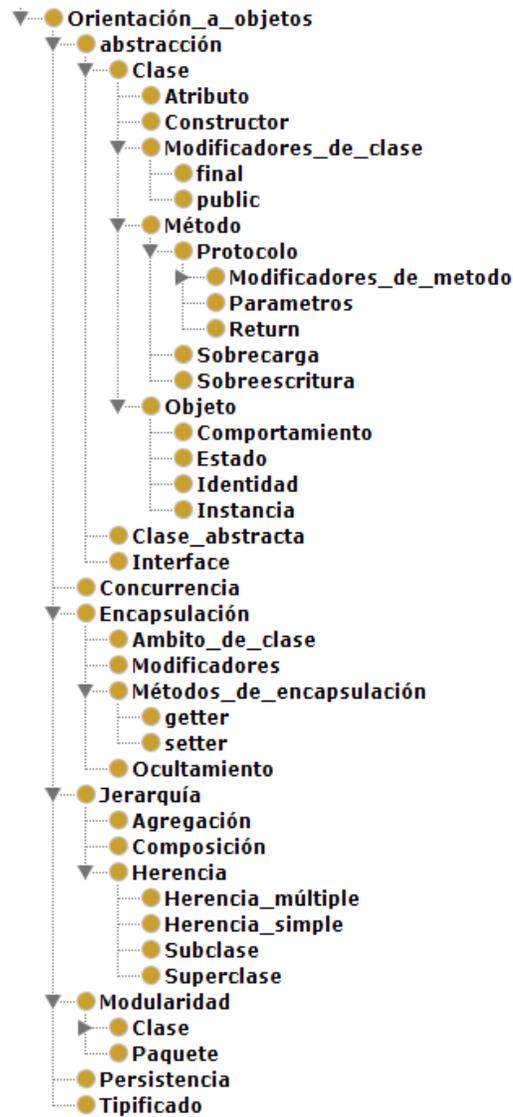


Figura A.1: Ontología representada jerárquicamente.

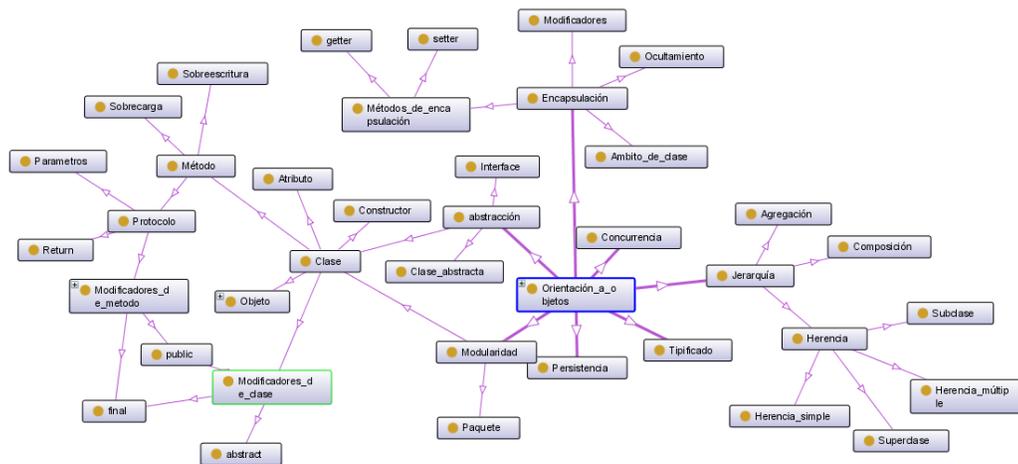


Figura A.2: Representación de árbol de la ontología.

Preguntas

1. ¿En cuál escuela laboras?

- Opción única: Instituto Tecnológico de Tijuana, Universidad Autónoma de Sinaloa, Universidad de Colima, Instituto Tecnológico de los Mochis, Universidad Autónoma de Baja California, Instituto Tecnológico de Hermosillo, Universidad Autónoma de Ciudad Juárez, Universidad Autónoma de Yucatán, Universidad Autónoma de Zacatecas, Universidad Autónoma Metropolitana, Universidad del Papaloapan, Universidad Nacional Autónoma de México, Universidad Popular Autónoma del Estado de Puebla, Universidad Tecnológica de la Mixteca, Universidad Veracruzana, Universidad Politécnica de San Luis Potosí, Universidad Iberoamericana, Otra.

2. ¿Cuánto tiempo llevas programando en OO?

- opción única: Menos de 2 años, 3-4 años, 5-6 años, 7-8 años, 9-10 años, 11-12 años, 13-14 años, mas de 15 años.

3. ¿Qué nivel de conocimiento consideras tener en OO?

- Opción única: Muy poco, poco, suficiente, mucho, experto.

4. ¿Cuánto tiempo tienes impartiendo POO?

- Opción única: Un año o menos, 2-3 años, 4-5 años, 6-7 años, 8-9 años, 10-11 años, 12-13 años, mas de 14 años.

5. En la escala presentada ¿Qué tan importante consideras los siguientes conceptos para que el estudiante pueda aprender el concepto ORIENTACIÓN A OBJETOS? (Figura A.3)

| | 1 (Poco importante) | 2 | 3 | 4 | 5 | 6 | 7 (Demasiado importante) |
|---------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------------|
| Abstracción | <input type="radio"/> |
| Concurrencia | <input type="radio"/> |
| Encapsulación | <input type="radio"/> |
| Jerarquía | <input type="radio"/> |
| Modularidad | <input type="radio"/> |
| Persistencia | <input type="radio"/> |
| Tipificado | <input type="radio"/> |

Figura A.3: Conceptos de orientación a objetos.

6. En la escala presentada ¿Qué tan importante consideras los siguientes conceptos para que el estudiante pueda aprender el concepto ABSTRACCIÓN? (Figura A.4)

| | 1 (Poco importante) | 2 | 3 | 4 | 5 | 6 | 7 (Demasiado importante) |
|-----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------------|
| Clase | <input type="radio"/> |
| Clase abstracta | <input type="radio"/> |
| Interface | <input type="radio"/> |

Figura A.4: Conceptos de abstracción.

7. En la escala presentada ¿Qué tan importante consideras los siguientes conceptos para que el estudiante pueda aprender el concepto CLASE? (Figura A.5)

| | 1 (Poco importante) | 2 | 3 | 4 | 5 | 6 | 7 (Demasiado importante) |
|------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------------|
| Atributo | <input type="radio"/> |
| Modificadores de clase | <input type="radio"/> |
| Método | <input type="radio"/> |
| Constructor | <input type="radio"/> |
| Objeto | <input type="radio"/> |

Figura A.5: Conceptos de clase.

8. En la escala presentada ¿Qué tan importante consideras los siguientes conceptos para que el estudiante pueda aprender el concepto MODIFICADORES DE CLASE? (Figura A.6)

| | 1 (Poco importante) | 2 | 3 | 4 | 5 | 6 | 7 (Demasiado importante) |
|----------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------------|
| Abstract | <input type="radio"/> |
| Final | <input type="radio"/> |
| Public | <input type="radio"/> |

Figura A.6: Conceptos de modificadores de clase.

9. En la escala presentada ¿Qué tan importante consideras los siguientes conceptos para que el estudiante pueda aprender el concepto MÉTODO? (Figura A.7)

| | 1 (Poco importante) | 2 | 3 | 4 | 5 | 6 | 7 (Demasiado importante) |
|--------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------------|
| Return | <input type="radio"/> |
| Parametros | <input type="radio"/> |
| Modificadores de métodos | <input type="radio"/> |
| Sobrecarga | <input type="radio"/> |
| Sobreescritura | <input type="radio"/> |

Figura A.7: Conceptos de método.

10. En la escala presentada ¿Qué tan importante consideras los siguientes conceptos para que el estudiante pueda aprender el concepto OBJETO? (Figura A.8)

| | 1 (Poco importante) | 2 | 3 | 4 | 5 | 6 | 7 (Demasiado importante) |
|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------------|
| Instancia | <input type="radio"/> |
| Identidad | <input type="radio"/> |
| Estado | <input type="radio"/> |
| Comportamiento | <input type="radio"/> |

Figura A.8: Conceptos de objeto.

11. En la escala presentada ¿Qué tan importante consideras los siguientes conceptos para que el estudiante pueda aprender el concepto ENCAPSULACIÓN? (Figura A.9)

| | 1 (Poco importante) | 2 | 3 | 4 | 5 | 6 | 7 (Demasiado importante) |
|-------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------------|
| Agregación | <input type="radio"/> |
| Composición | <input type="radio"/> |
| Herencia | <input type="radio"/> |

Figura A.9: Conceptos de encapsulación.

12. En la escala presentada ¿Qué tan importante consideras los siguientes conceptos para que el estudiante pueda aprender el concepto JERARQUÍA? (Figura A.10)

| | 1 (Poco importante) | 2 | 3 | 4 | 5 | 6 | 7 (Demasiado importante) |
|-------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------------|
| Agregación | <input type="radio"/> |
| Composición | <input type="radio"/> |
| Herencia | <input type="radio"/> |

Figura A.10: Conceptos de jerarquía.

13. En la escala presentada ¿Qué tan importante consideras los siguientes conceptos para que el estudiante pueda aprender el concepto HERENCIA? (Figura A.11)

| | 1 (Poco importante) | 2 | 3 | 4 | 5 | 6 | 7 (Demasiado importante) |
|-------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------------|
| Herencia múltiple | <input type="radio"/> |
| Herencia simple | <input type="radio"/> |
| Subclase | <input type="radio"/> |
| Superclase | <input type="radio"/> |

Figura A.11: Conceptos de herencia.

14. En la escala presentada ¿Qué tan importante consideras los siguientes conceptos para que el estudiante pueda aprender el concepto MODULARIDAD? (Figura A.12)

| | 1 (Poco importante) | 2 | 3 | 4 | 5 | 6 | 7 (Demasiado importante) |
|---------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------------|
| Clase | <input type="radio"/> |
| Paquete | <input type="radio"/> |

Figura A.12: Conceptos de herencia.

A.2. Encuesta sobre Tipos de Lenguajes de Programación

Buenos días (tardes):

Se está trabajando en un estudio que servirá para elaborar un artículo de investigación relacionada con Sistemas Tutores Inteligentes.

Se pide tu ayuda para que contestes algunas preguntas que no te tomarán mucho tiempo. Tus respuestas serán confidenciales y anónimas. Las opiniones de todos los encuestados serán incluidas en el artículo en forma general, no se comunicarán datos individuales.

Considere el siguiente caso hipotético: Se te solicita dar un curso a un grupo de estudiantes sobre los tipos de lenguajes de programación. Este curso se basa en describir los paradigmas de programación así como sus principales lenguajes considerando su importancia, características y su funcionamiento en aspectos teóricos.

El objetivo es determinar una proporción generalizada entre los temas y subtemas del contenido. Este instrumento no maneja tiempos en horas o clases para evitar incoherencias o contradicciones, se manejan porcentajes sobre el curso, de esta forma es fácil adaptarse a cualquier tiempo considerado.

A continuación se muestran los temas considerados:

Tipos de lenguajes de programación

- Programación estructurada

- Algol
- Ada

- Programación procedural

- Perl
- Fortran
- VB

- Pascal
- Programación orientada a objetos
 - VB. NET
 - C#
 - C++
 - Java
 - PHP
- Programación lógica
 - Prolog
 - Datalog
 - Mercury

Se esta conscientes que ningún profesor conocerá todos los lenguajes de programación para estimar los tiempos con la precisión que ustedes quisieran. Sin embargo, se solicita contestar lo mas razonablemente posible.

Datos generales

1. ¿A qué rango de edad perteneces?

- Opción única: menos de 25 años, 26-30 años, 31-35 años, 36-40 años, 41-45 años, 46-50 años, 50-55 años, 56-60 años, mas de 61 años.

2. ¿Cuántos años tienes trabajando como docente en nivel universitario?

- Opción única: 0-3 años, 4-6 años, 7-9 años, 10-12 años, 13-15 años, 16-18 años, 19-21 años, 22-25 años, más de 25 años.

3. ¿En cuál escuela laboras?

- Opción de selección múltiple: Instituto Tecnológico de Tijuana, Universidad Autónoma de Sinaloa, Universidad de Colima, Instituto Tecnológico de los Mochis, Universidad Autónoma de Baja California, Instituto Tecnológico de Hermosillo, Universidad Autónoma de Ciudad Juárez, Universidad Autónoma de Yucatán, Universidad Autónoma de Zacatecas, Universidad Autónoma Metropolitana, Universidad del Papaloapan, Universidad Nacional Autónoma de México, Universidad Popular Autónoma del Estado de Puebla, Universidad Tecnológica de la Mixteca, Universidad Veracruzana, Universidad Politécnica de San Luis Potosí, Universidad Iberoamericana, otra.

4. ¿Con cuáles lenguajes de programación tienes experiencia?

- Opción de selección múltiple: Algol, Ada, Pascal, Visual basic, Fortran, Perl, Visual basic .net, C#, C++, Java, PHP, Prolog, Datalog, Mercury, Otro.

5. ¿Cuáles es (son) tu(s) lenguaje(s) de programación favorito(s) ?

- Opción de selección múltiple: Algol, Ada, Pascal, Visual basic, Fortran, Perl, Visual basic .net, C#, C++, Java, PHP, Prolog, Datalog, Mercury, Otro.

Tipos de lenguajes de programación

1. ¿Para enseñar el tema tipos de lenguajes de programación como distribuyes el tiempo (en porcentaje, la suma de la distribución debe ser 100) del subtema ...

- Programación estructurada?
 - Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 1000 %.
- Programación procedural?

- Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 1000 %.
- Programación orientada a objetos?
 - Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 1000 %.
- Programación lógica?
 - Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 1000 %.

Programación procedural

1. ¿Para enseñar el tema programación procedural, como distribuyes el tiempo (en porcentaje, la suma de la distribución debe ser 100) del lenguaje ...
 - Perl?
 - Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 1000 %.
 - Fortran?
 - Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 1000 %.
 - Visual Basic?
 - Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 1000 %.
 - Pascal?
 - Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 1000 %.

Programación orientada a objetos

1. ¿Para enseñar el tema programación orientada a objetos, como distribuyes el tiempo (en porcentaje, la suma de la distribución debe ser 100) del lenguaje ...

■ Visual Basic .NET?

- Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 100 %.

■ C#?

- Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 100 %.

■ C++?

- Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 100 %.

■ Java?

- Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 100 %.

■ PHP?

- Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 100 %.

Programación estructurada

1. ¿Para enseñar el tema programación estructurada como distribuyes el tiempo (en porcentaje, la suma de la distribución debe ser 100) del lenguaje ...

■ Algol?

- Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 1000 %.

- Ada?

- Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 1000 %.

Programación lógica

1. ¿Para enseñar el tema programación lógica como distribuyes el tiempo (en porcentaje, la suma de la distribución debe ser 100) del lenguaje ...

- Prolog?

- Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 1000 %.

- Datalog?

- Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 1000 %.

- Mercury?

- Opción única: 0 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 1000 %.

Sección final

1. ¿Como enfrentaste la situación de asignar porcentajes a lenguajes de programación desconocidos o poco conocidos para ti?

- Respuesta abierta.

Apéndice B

Calculo de los pesos

B.1. Pesos obtenidos en caso de estudio B

Este anexo describe el cálculo de los pesos considerando los factores:

- Relación cuantitativa considerando las frecuencias (Factor f_r , Ecuación 5.2)
- Relación cuantitativa considerando la coincidencia de conceptos (factor j , Ecuación 5.4)
- Relación cuantitativa considerando saltos entre conceptos (Factor s , Ecuación 5.6)

Con fines demostrativos solo se tomó el primer nivel de la Figura 6.7, considerando el nodo raíz (*Tipos_lenguajes_programación*) y los elementos que componen a este nodo (*Programación_Lógica*, *Programación_orientada_objetos*, *Programación_procedural* y *Programación_estructurada*).

La Tabla B.1 muestra los resultados obtenidos a partir de los factores. Esta tabla muestra en la primer columna las relaciones consideradas para el caso de estudio B; cada factor considera dos valores, el valor sin normalizar y el valor normalizado. La última columna muestra el valor final de la relación. Cada factor llega al resultado desplegado en la columna *Sin norm.* a través de un proceso propio de cada factor.

Tabla B.1: Valores obtenidos por factor

| Relación | Factor j | | Factor s | | Factor fr | | Valor |
|---|-----------|-------|-----------|-------|-----------|-------|-------|
| | Sin norm. | Norm. | Sin norm. | Norm. | Sin norm. | Norm. | |
| Tipos_lenguajes_programación-Programación_Lógica | 0.348 | 0.245 | 1 | 0.250 | 12 | 0.300 | 0.27 |
| Tipos_lenguajes_programación-Programación_orientada_objetos | 0.433 | 0.305 | 1 | 0.250 | 19 | 0.475 | 0.34 |
| Tipos_lenguajes_programación-Programación_procedural | 0.269 | 0.189 | 1 | 0.250 | 4 | 0.100 | 0.18 |
| Tipos_lenguajes_programación-Programación_estructurada | 0.370 | 0.260 | 1 | 0.250 | 5 | 0.125 | 0.21 |

El valor de la columna *Norm.* es obtenido a través de la Ecuación (5.5). Donde P contiene los valores sin normalizar de cada relación $P = \{0.348, 0.433, 0.269, 0.370\}$, por tanto, el valor normalizado de $p = \text{Tipos_lenguajes_programación} - \text{Programación_Lógica}$ está dado por (B.1), el valor calculado para dicha relación es 0.245. El proceso es repetido para cada elemento de P.

$$j_{norma}(p) = 0.245 = \frac{0.348}{0.348 + 0.433 + 0.269 + 0.370} \quad (\text{B.1})$$

Para cada columna *Norm.* se emplea la misma fórmula, solo que cambian los valores de P de acuerdo al factor.

El valor final es calculado con la Ecuación (5.1), sustituyendo los valores para $p = \text{Programación_Lógica}$ y $h = \text{Tipos_lenguajes_programación}$ tenemos un valor redondeado en la relación de 0.27 (Ecuación (B.2)).

$$rel(p, h) = 0.27 = \frac{1}{3}(0.245 + 0.250 + 0.300) \quad (\text{B.2})$$

La misma ecuación es aplicada para obtener el valor final del resto de las relaciones. Los resultados de cada relación son mostrados en la Tabla 6.3.

Apéndice C

Publicaciones

C.1. Primer autor

C.1.1. Revistas con factor de impacto (JCR)

- **Ramírez-Noriega A.**, Juárez-Ramírez R. , Jiménez S., Martínez-Ramírez Y. , Figueroa F. (2017). Determination of the course sequencing to intelligent tutoring systems using an ontology and Wikipedia. *Journal of Intelligent & Fuzzy Systems* (aceptado, por publicar)

- **Ramírez-Noriega A.**, Juárez-Ramírez R. , Jiménez S., Martínez-Ramírez Y. Inzunza S. (2017). ASHuR: Evaluation of the Summary-Content Relation without Human Reference Using ROUGE, *Computing and Informatics* (aceptado, por publicar).

- **Ramírez-Noriega, A.**, Juárez-Ramírez, R., & Martínez-Ramírez, Y. (2017). Evaluation module based on Bayesian networks to Intelligent Tutoring Systems. *International Journal of Information Management*, 37(1), 1488–1498. <http://doi.org/10.1016/j.ijinfomgt.2016.05.007>

C.1.2. Conferencias y libros

- **Ramírez-Noriega, A.**, Juárez-Ramírez, R., Jiménez, S., Inzunza, S., Navarro, R., & López-Martínez, J. (2017). An Ontology of the Object Orientation for Intelligent Tutoring Systems. In 2017 5th International Conference in Software Engineering Research and Innovation (pp. 1–10). Merida, Yucatan, México: IEEE (aceptado, por publicar).
- **Ramírez-Noriega, A.**, Juárez-Ramírez, R., Jiménez, S., Martínez-Ramírez, Y., & Armenta, J. (2017). Building a Bayesian Network for Object Oriented Programming with Experts' Knowledge. In Á. Rocha, A. M. Correia, H. Adeli, L. P. Reis, & S. Costanzo (Eds.), *Recent Advances in Information Systems and Technologies: Volume 1* (pp. 267–276). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-56535-4_27
- **Ramírez-Noriega, A.**, Juárez-Ramírez, R., Navarro, R., & López-Martínez, J. (2016). Using Bayesian networks to obtain the task's parameters for schedule planning in scrum. In *Proceedings - 2016 4th International Conference in Software Engineering Research and Innovation, CONISOFT 2016*. <http://doi.org/10.1109/CONISOFT.2016.33>
- **Ramírez-Noriega, A.**, Juárez-Ramírez, R., Jiménez, S., & Martínez-Ramírez, Y. (2017). Knowledge Representation in Intelligent Tutoring System. In A. E. Hassanien, K. Shaalan, T. Gaber, A. T. Azar, & M. F. Tolba (Eds.), *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016* (pp. 12–21). Egypt: Springer International Publishing. http://doi.org/10.1007/978-3-319-48308-5_2
- **Ramírez-Noriega, A.**, Juárez-Ramírez, R., Martínez-Ramírez, Y., Samantha, J., & Sergio, I. (2016). Using Bayesian Networks for Knowledge Representation and Evaluation in Intelligent Tutoring Systems. *New Advances in Information Systems and Technologies, (Vol. 1)*. Springer International Publishing Switzerland. <http://doi.org/10.1007/978-3-319-31232-3>
- **Ramírez-Noriega, A.**, Juárez-Ramírez, R., Huertas, C., & Martínez-Ramírez, Y. (2015). A

Methodology for building Bayesian Networks for Knowledge Representation in Intelligent Tutoring Systems. In R. Juárez Ramírez, C. Guerra García, & H. Oktaba (Eds.), Congreso Internacional de Investigación e Innovación en Ingeniería de Software 2015 (pp. 124–133). San Luís Potosí.

C.2. Colaboraciones

C.2.1. Revistas con factor de impacto (JCR)

- López-Martínez, J., **Ramírez-Noriega, A.**, Juárez-Ramírez, R., Licea, G., & Jiménez, S. (2017). User stories complexity estimation using Bayesian networks for inexperienced developers. *Cluster Computing*, 1–14. <http://doi.org/10.1007/s10586-017-0996-z>
- Jiménez, S., Juárez-Ramírez, R., Castillo, V. H., & **Ramírez-Noriega, A.** (2017). Integrating affective learning into intelligent tutoring systems. *Universal Access in the Information Society*, 1–14. <http://doi.org/10.1007/s10209-017-0524-1>
- Quezada, A., Juárez-Ramírez, R., Jiménez, S., **Ramírez-Noriega, A.**, Inzunza, S., & Garza, A. A. (2017). Usability Operations on Touch Mobile Devices for Users with Autism. *Journal of Medical Systems*, 41(11), 184. <http://doi.org/10.1007/s10916-017-0827-z>

C.2.2. Conferencias y libros

- López-Martínez, J., **Ramírez-Noriega, A.**, Juárez-Ramírez, R., Licea, G., & Martínez-Ramírez, Y. (2017). Analysis of Planning Poker Factors between University and Enterprise. In 2017 5th International Conference in Software Engineering Research and Innovation. Merida, Yucatan, México: IEEE.
- Jiménez, S., Juárez-Ramírez, R., Castillo, V. H., **Ramírez-Noriega, A.**, Inzunza, S., & Quezada,

- A. (2017). A Conceptual Framework for Learning System Evaluation. In 2017 5th International Conference in Software Engineering Research and Innovation. Merida, Yucatan, México: IEEE.
- López-Martínez, J., Juárez-Ramírez, R., **Ramírez-Noriega, A.**, Licea, G., & Navarro-Almanza, R. (2017). Estimating User Stories' Complexity and Importance in Scrum with Bayesian Networks. In Á. Rocha, A. M. Correia, H. Adeli, L. P. Reis, & S. Costanzo (Eds.), *Recent Advances in Information Systems and Technologies: Volume 1* (pp. 205–214). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-56535-4_21
 - Quezada, A., Juárez-Ramírez, R., Jiménez, S., **Ramírez-Noriega, A.**, & Inzunza, S. (2017). An Empirical Study on Usability Operations for Autistic Children. In Á. Rocha, A. M. Correia, H. Adeli, L. P. Reis, & S. Costanzo (Eds.), *Recent Advances in Information Systems and Technologies: Volume 2* (pp. 628–638). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-56538-5_64
 - Jiménez, S., Juárez-Ramírez, R., Castillo, V. H., **Ramírez-Noriega, A.**, & Inzunza, S. (2017). Affectivity Level for Intelligent Tutoring System Based on Student Stereotype. In Á. Rocha, A. M. Correia, H. Adeli, L. P. Reis, & S. Costanzo (Eds.), *Recent Advances in Information Systems and Technologies: Volume 2* (pp. 701–710). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-56538-5_71
 - Jiménez, S., Juárez-Ramírez, R., Castillo Topete, V., & **Ramírez-Noriega, A.** (2017). Affective Dialogue Ontology for Intelligent Tutoring Systems: Human Assessment Approach. In A. E. Hassanien, K. Shaalan, T. Gaber, A. T. Azar, & M. F. Tolba (Eds.), *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016* (pp. 608–617). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-48308-5_58
 - Inzunza, S., Juárez-Ramírez, R., & **Ramírez-Noriega, A.** (2016). User and context information in context-aware recommender systems: A systematic literature review. *Advances in Intelligent*

Systems and Computing (Vol. 444). http://doi.org/10.1007/978-3-319-31232-3_61

- Jimenez, S., Juarez-Ramirez, R., Castillo, V. H., & **Ramírez-Noriega, A.** (2016). An affective learning ontology for educational systems. *Advances in Intelligent Systems and Computing* (Vol. 444). http://doi.org/10.1007/978-3-319-31232-3_106

- Juárez-ramírez, R., Jiménez, S., **Ramírez-noriega, A.**, Inzunza, S., Navarro, C. X., Ciencias, F. De, . . . Tijuana, B. C. (2015). Mejora de la Usabilidad de Sistemas de Software: Un Modelo orientado hacia la Motivación del Usuario. *Congreso Internacional de Investigación E Innovación En Ingeniería de Software 2015*, 89–99.