

UniversidadeVigo

Roberto Agustín García Vélez

TESIS DOCTORAL

*Contribuciones a la gestión del conocimiento en el ámbito de la Educación Superior mediante técnicas de Inteligencia Artificial*

2019

UniversidadeVigo

## TESIS DOCTORAL

*Contribuciones a la gestión del conocimiento en el ámbito de la Educación Superior mediante técnicas de Inteligencia Artificial*

Roberto Agustín García Vélez  
2019

UniversidadeVigo

EIDO  
Escola Internacional  
de Doutoramento

# Universidade de Vigo

International Doctoral School

Autor:  
Roberto Agustín García Vélez

TESIS DOCTORAL:  
Contribuciones a la gestión del conocimiento en  
el ámbito de la Educación Superior mediante  
técnicas de Inteligencia Artificial

Dirigida por:  
Dr. Martín López Nores  
Dr. José Juan Pazos Arias

2019



# Agradecimientos

Esta tesis doctoral no hubiera sido posible sin la colaboración de muchas personas. Me gustaría agradecer su apoyo y agradecer a todos los cómplices por su confianza, tiempo y asesoramiento.

Mi más profunda gratitud y admiración a mis directores Dr. Martín López Nores y Dr. José Juan Pazos Arias, quienes han sido las personas más influyentes en mi formación académica. Gracias por su tiempo, su apoyo incondicional, su motivación y su carisma, en aquellos momentos difíciles han sido la parte fundamental para alcanzar este objetivo. Sin duda alguna, su calidad humana y académica son un gran ejemplo que seguir, gracias por su amistad.

De igual forma, un agradecimiento muy especial a la profesora del grupo de Servicios para la Sociedad de la Información, del Departamento de Ingeniería Telemática de la Universidad de Vigo, la Dra. Yolanda Blanco Fernández, por compartir su saber y experiencia vinculada al trabajo desarrollado.

Mi gratitud entera, a las distinguidas autoridades de la Universidad Politécnica Salesiana, en especial al Ph.D. Padre Javier Herrán, al Ph.D. Luis Tobar Pesántez y al Magister César Vásquez Vásquez —rector, vicerrector y vicerrector de sede de la Universidad Politécnica Salesiana del Ecuador (UPS)— por confiar en mi trabajo y porque a través de su apoyo financiero fue posible desarrollar mis estudios doctorales en la Universidad de Vigo.

Un inmenso agradecimiento al Ph.D. Vladimir Robles, Ph.D. Fernando Pesántez, Ph.D. Yaroslava Robles, amigos y colaboradores del grupo GIIATA, por apoyar mi investigación con sus competencias en la investigación científica. Adicionalmente y de manera muy especial al Ing. Jorge Galán, quién trabajó arduamente en el desarrollo del ecosistema semántico empleado en esta tesis. Al igual expreso mi gratitud al Ing. Daniel Calle, colaborador en la parte experimental del proceso.

Esta disertación es el resultado de una reflexión colectiva de un grupo transdisciplinario, conducido por mi persona al proyecto de “*Contribuciones a la gestión del conocimiento en el ámbito de la Educación Superior*”, entre la Carrera de Ingeniería en Sistemas y la Carrera de Psicología. Por ello, expreso mi gratitud hacia la directora de Carrera de Ingeniería en Sistemas, Mgs. Bertha Tacuri Capelo, por su colaboración y a la psicóloga Mgs. María Eugenia Barros, cuyo saber brindó un gran aporte a este trabajo.

Estoy agradecido a las institución, docentes, y estudiantes de las diferentes Carreras partícipes: Carrera de Ingeniería en Sistemas, Carrera de Electrónica y la Carrera de Telecomunicaciones en los períodos 2017-2018 y 2018-2018, quienes se unieron a este trabajo y colaboraron para hacerlo posible.

Finalmente, muchas personas dentro y fuera de España me brindaron el apoyo emocional para enfrentar esta tarea desafiante. Deseo agradecer a mis colegas y amigos de doctorado, Gustavo Bravo Quezada, Esteban Ordóñez Morales, Luis Serpa y Fernando Guerrero. Un agradecimiento especial a mis amigos de estancias Hilario, Fátima, Miran, Carmelo, Carmen, Sheila y Juan.

En lo personal, además de la gratitud por todo el apoyo académico, quiero expresar mi inmensa gratitud para con Martín y su maravillosa familia Susi, Andrea y su precioso hijo Diego, sin olvidar a sus amables padres. Gracias por todo.

Sin duda alguna, la familia es un pilar fundamental y por ello, mi enorme gratitud a Marianita Vélez, Roberto García, mis padres, a María Eugenia, Jhonny, Adriana y Gabriela, mis hermanos, quienes fueron mi soporte emocional durante los momentos más difíciles en el progreso de este trabajo, con sus palabras y el ejemplo diario de perseverancia, son una parte maravillosa de mi vida.

Por último, pero no menos importante a mi maravillosa y amada esposa Verónica y mi dulce y querido hijo Agustín. Gracias por brindarme su tiempo y su apoyo para alcanzar esta meta juntos.

Muchas gracias.

# Dedicatoria

*“Con especial dedicación a mi esposa Verónica y mi niño Agustín, el centro de mi vida,  
por su amor incondicional y por hacer de cada día un motivo de alegría.”*



# Resumen

Las instituciones de Educación Superior cuentan por lo general con plataformas digitales para el manejo de la información correspondiente a los perfiles de los estudiantes, docentes, personal administrativo, autoridades y demás, organizadas en forma de bases de datos relacional. Igualmente, emplean diferentes aplicaciones que generan continuamente una gran cantidad de datos. Tal información usualmente está distribuida entre diferentes departamentos, cada uno de los cuales cuenta a menudo con sus propias tecnologías de información y de almacenamiento, de modo que no siempre manejan las mismas estructuras de repositorios digitales, bases de datos o formatos de datos.

La información forma parte vital del funcionamiento de las instituciones, ya que a través de los diferentes sistemas se hace posible que los diferentes actores del ecosistema académico puedan realizar consultas, análisis, programación de actividades, gestión de recursos, toma de decisiones y demás acciones. La falta de integración, sin embargo, es fuente habitual de errores e ineficiencias, y es por ello que a lo largo de las últimas décadas se han registrado numerosos aportes en la línea de implementar sistemas de Inteligencia Artificial que promuevan la interconexión de los múltiples repositorios, la integración y consolidación de la información dispersa, y el análisis automatizado de grandes volúmenes de datos. El objetivo último es convertir la información en *conocimiento*, al que se le pueda sacar el máximo partido en el sentido de optimizar procesos y maximizar resultados.

El punto de partida de esta tesis doctoral fue que las propuestas basadas en el uso de las tecnologías existentes para brindar apoyo a la gestión del conocimiento, tanto de la información administrativa de las instituciones de Educación Superior como de la información académica, presentan notables limitaciones, tales como: (i) no tomar en cuenta las complejas relaciones que existen entre las informaciones que poseen los diferentes departamentos, (ii) no dividir y explicitar el desarrollo que los diferentes actores del sistema académico toman a lo largo del tiempo, (iii) mantener el conocimiento disperso por múltiples formatos no estructurados y repositorios digitales no conectados, (iv) no generar relaciones semánticas en base al histórico de la información de los estudiantes, para desarrollar estrategias que mejoren el rendimiento académico, y (v) no contemplar una definición formal del conocimiento asociado al ámbito de la Educación Superior que faculte mejorar los procesos que exploran desarrollar herramientas de soporte.

La tesis se planteó con el objetivo general de forjar las bases de un completo ecosistema de apoyo a la administración y gestión de la información que se genera entorno al ámbito de la Educación Superior, desde una perspectiva fundamentada en el modelado semántico del conocimiento y el empleo de innovadoras técnicas de

Inteligencia Artificial, brindando un diseño altamente flexible e interoperable. Las contribuciones de este trabajo de investigación se enmarcan en tres grandes áreas:

- **Modelado formal del conocimiento del ecosistema académico de Educación Superior.** En primer lugar, se ha creado e implantado una red ontológica para integrar la información generada en las distintas áreas del entorno académico, dando cuenta de las cuestiones de representación, interoperabilidad y consolidación.
- **Formación eficiente de grupos de trabajo colaborativo.** Aprovechando el modelado formal de los estudiantes y de los planos de estudios a nivel micro y macroscópico, hemos implementado un sistema basado en algoritmos genéticos con el que apoyar la formación de grupos de trabajo colaborativo que potencien los resultados de aprendizaje.
- **Predicción del rendimiento académico.** Igualmente, con base en el modelado formal del conocimiento del ecosistema académico, hemos llevado a cabo un estudio sobre el potencial de aplicar técnicas de aprendizaje profundo para la predicción del rendimiento académico de los estudiantes, a partir de características socioeconómicas, con el objetivo de apoyar estrategias de planificación y dotación de recursos.

Las contribuciones se han validado en el marco concreto de la Universidad Politécnica Salesiana (UPS) de Ecuador, con acceso a los datos acumulados en sus sistemas de información desde el inicio de su actividad, más otros recabados al efecto de nuevos experimentos, en los que se ha contado con la participación activa de estudiantes y docentes de dicha universidad. Las técnicas y los resultados se consideran de aplicación general a otras universidades del ámbito de Latinoamérica, debido a las similitudes entre los modelos estructurales y organizativos; mirando más allá, los fundamentos de las distintas propuestas también serían fácilmente trasladables a la realidad de las instituciones de Educación Superior en otras partes del mundo.

# Abstract

Higher Education institutions usually have digital platforms for the management of information corresponding to the profiles of students, teachers, administrative staff, authorities and others, organized in the form of relational databases. Likewise, they use different applications that continuously generate large amounts of data. Such information is usually distributed among different departments, each of which often has its own information and storage technologies, so that they do not always handle the same structures of digital repositories, databases or data formats.

The information is a vital part of the functioning of the institutions, since through the different systems it is possible for the different actors of the academic ecosystem to make queries, analyze, program activities, manage resources, make decisions and other actions. The lack of integration, however, is a common source of errors and inefficiencies, and that is why throughout the last decades there have been numerous contributions in the line of implementing Artificial Intelligence systems that promote the interconnection of the multiple repositories, the integration and consolidation of dispersed information, and the automated analysis of sizable volumes of data. The ultimate goal is to turn information into *knowledge*, which can be used to optimize processes and maximize results.

The starting point of this doctoral thesis was that the proposals based on the use of existing technologies to provide support for knowledge management, both of the administrative information of Higher Education institutions and of academic information, present notable limitations, such as: (i) not taking into account the complex relationships that exist between the information held by the different departments, (ii) not dividing and making explicit the development that the different actors of the academic system take over time, (iii) maintaining knowledge dispersed over multiple unstructured formats and unconnected digital repositories, (iv) not generating semantic relationships based on the historical information of the students, to develop strategies that improve academic performance, and (v) not contemplating a formal definition of the knowledge associated with the field of Higher Education that empowers the development of support tools.

The thesis was posed with the general objective of forging the foundations of a complete ecosystem to support the management of the information that is generated around the field of Higher Education, from a perspective based on semantic modeling of knowledge and the use of innovative Artificial Intelligence techniques, providing a highly flexible and interoperable design. The contributions of this research work fall into three major areas:

- **Formal modeling of knowledge of the academic ecosystem of Higher Education.** In the first place, an ontological network has been designed and

implemented to integrate the information generated in the different areas of the academic environment, dealing with the issues of representation, interoperability and consolidation.

- **Efficient formation of collaborative work groups.** Taking advantage of the formal modeling of students and the levels of studies at the micro- and macroscopic level, we have implemented a system based on genetic algorithms with which to support the formation of collaborative work groups that enhance learning outcomes.
- **Prediction of academic performance.** Also based on the formal modeling of knowledge of the academic ecosystem, we have carried out a study on the potential of applying *deep learning* techniques for the prediction of the academic performance of students, based on socioeconomic characteristics, with the objective of supporting planning strategies and resource management.

The contributions have been validated in the context of Universidad Politécnica Salesiana (UPS) in Ecuador, with access to the data accumulated in their information systems since the beginning of its activity, plus others collected for the purpose of new experiments, which have enjoyed active participation of students and teachers of the University. The techniques and results are considered of general application to other universities in the Latin American area, due to the similarities between the structural and organizational models; beyond that, the foundations of the different proposals would be easily transferable to the reality of Higher Education institutions in other parts of the world, too.





# Índice general

<b>Índice general</b>	<b>XI</b>
<b>Índice de figuras</b>	<b>XV</b>
<b>Índice de tablas</b>	<b>XVII</b>
<b>I Introducción</b>	<b>1</b>
<b>1 Ámbito, objetivos y planteamiento de la tesis</b>	<b>3</b>
1.1 Gestión del conocimiento . . . . .	4
1.2 Gestión del conocimiento en las instituciones de Educación Superior	6
1.3 Problemática principal de la tesis . . . . .	7
1.4 Objetivo y contribuciones . . . . .	9
1.5 Estructura y organización de la memoria . . . . .	10
<b>II Estado del arte</b>	<b>13</b>
<b>2 Estado del arte en áreas seleccionadas de Inteligencia Artificial</b>	<b>15</b>
2.1 Introducción a la Inteligencia Artificial . . . . .	15
2.1.1 Breve historia de la IA . . . . .	16
2.2 Tecnologías de Web Semántica . . . . .	18
2.2.1 Arquitectura de la Web Semántica . . . . .	20
2.2.2 Redes ontológicas y la metodología NeOn . . . . .	26
2.3 Redes Neuronales Artificiales . . . . .	29
2.3.1 Estructura de una Red Neuronal Artificial . . . . .	29
2.3.2 Perceptrón simple . . . . .	32
2.3.3 Redes de Hopfield . . . . .	34
2.3.4 Redes de Kohonen . . . . .	36
2.3.5 Perceptrón multicapa . . . . .	39
2.4 Algoritmos genéticos . . . . .	42
2.4.1 Estructura básica de un algoritmo genético . . . . .	43
2.4.2 Representación de cromosomas . . . . .	43
2.4.3 Operadores genéticos . . . . .	44
2.4.4 Tipos de algoritmos genéticos . . . . .	46
2.4.5 Propagación genética: Teorema de los esquemas . . . . .	47
2.4.6 Fortalezas y debilidades de los algoritmos genéticos . . . . .	49

<b>3</b>	<b>Técnicas de IA en las instituciones de Educación Superior</b>	<b>51</b>
3.1	Introducción . . . . .	51
3.2	Juegos serios como elementos de refuerzo en el aprendizaje . . . . .	52
3.3	Uso de simuladores en las instituciones de Educación Superior . . . . .	54
3.4	Sistemas de tutoría inteligentes . . . . .	60
3.5	Modelado del conocimiento en el ámbito de las IES . . . . .	61
3.6	Sistemas de predicción y toma de decisiones en el ámbito académico . . . . .	64
3.7	Apoyos para la formación de grupos de aprendizaje colaborativos . . . . .	68
<b>III</b>	<b>Contribuciones</b>	<b>75</b>
<b>4</b>	<b>Aportes a la gestión de conocimiento en instituciones de Educación Superior</b>	<b>77</b>
4.1	Introducción . . . . .	77
4.2	Elicitación de requerimientos . . . . .	78
4.3	Diseño de la red ontológica OEA . . . . .	83
4.4	Implementación del gestor de conocimiento para la red ontológica OEA . . . . .	90
4.5	Población de la red . . . . .	90
4.6	Optimización de motor de reglas . . . . .	97
4.7	Resultados de migración y tiempos de carga . . . . .	102
4.8	Trazabilidad de los perfiles . . . . .	102
4.9	Comprobación de inferencias <i>sameAs</i> en registros de estudiante y profesor . . . . .	108
4.10	Sumario . . . . .	111
<b>5</b>	<b>Aportes a la formación de grupos y predicción de rendimiento académico</b>	<b>113</b>
5.1	Introducción . . . . .	113
5.2	Arquitectura del sistema . . . . .	114
5.2.1	Interfaz de usuario . . . . .	116
5.2.2	Gestión del perfil del estudiante . . . . .	116
5.2.3	Gestión del conocimiento . . . . .	117
5.2.4	Sistema experto . . . . .	120
5.3	Componente de OEA para la inferencia de rasgos de personalidad . . . . .	120
5.3.1	Formalización del componente ontológico . . . . .	121
5.4	Algoritmo genético para formación de grupos de aprendizaje colaborativo . . . . .	124
5.4.1	Criterios para la formación de los grupos de aprendizaje colaborativo . . . . .	125
5.4.2	Descripción general del modelo <i>Big 5</i> . . . . .	128
5.4.3	Detalles del algoritmo propuesto . . . . .	130
5.5	Aprendizaje profundo para predicción de rendimiento académico . . . . .	132
5.5.1	Análisis de parámetros socioeconómicos de los estudiantes . . . . .	133
5.5.2	Selección de parámetros . . . . .	135
5.5.3	Valores de salida . . . . .	139
5.5.4	Balance de datos y regularización . . . . .	139
5.5.5	Modelado e implementación de la red . . . . .	140
5.6	Sumario . . . . .	142

---

<b>6</b>	<b>Evaluación experimental de los aportes</b>	<b>143</b>
6.1	Logística de experimentación . . . . .	143
6.2	Evaluación del algoritmo genético . . . . .	145
6.3	Evaluación de la predicción de rendimiento académico . . . . .	152
6.4	Sumario . . . . .	156
<b>IV</b>	<b>Conclusiones</b>	<b>159</b>
<b>7</b>	<b>Conclusiones y trabajos futuros</b>	<b>161</b>
7.1	Conclusiones . . . . .	161
7.1.1	Sobre el modelado formal del conocimiento . . . . .	161
7.1.2	Sobre la formación de grupos de aprendizaje colaborativo . . . . .	162
7.1.3	Sobre la predicción del rendimiento académico . . . . .	162
7.1.4	Perspectiva general . . . . .	163
7.2	Trabajo futuro . . . . .	163
<b>V</b>	<b>Bibliografía</b>	<b>167</b>
	<b>Bibliografía</b>	<b>169</b>



# Índice de figuras

2.1	Las capas de la arquitectura de la Web Semántica. . . . .	21
2.2	Sistema general de una red neuronal artificial . . . . .	29
2.3	Funciones de activación dentro de las RNA. . . . .	30
2.4	Diagrama del perceptrón simple . . . . .	32
2.5	Ejemplo de clases linealmente divisibles. . . . .	33
2.6	Arquitectura de una red Hopfield con $N$ nodos. . . . .	34
2.7	Estructura básica de una red neuronal competitiva de Kohonen. . . . .	37
2.8	Estructura de un perceptrón multicapa con dos capas ocultas. . . . .	40
4.1	Esquema modular de la red ontológica OEA. . . . .	87
4.2	Captura de los módulos de Protégé que contienen las clases de la red OEA. . . . .	89
4.3	Capas, herramientas y técnicas del nuevo repositorio semántico de la UPS. . . . .	91
4.4	Esquema ontológico del mapeo de estudiantes. . . . .	94
4.5	Esquema ontológico del mapeo de docentes. . . . .	95
4.6	Instancias resultantes del mapeo de estudiantes y docentes. . . . .	96
4.7	Expansión de un conjunto de triplas inicial por medio del conjunto de reglas <i>OWL Horst</i> . . . . .	100
4.8	Análisis de inferencias por regla de <i>OWL Horst</i> . . . . .	101
4.9	Expansión de un conjunto de triplas inicial por medio del conjunto de reglas <i>OWL Horst</i> modificado. . . . .	102
4.10	Análisis de inferencias por regla del conjunto <i>OWL Horst</i> modificado. . . . .	103
4.11	Diagrama ontológico del dominio del perfil personal. . . . .	105
4.12	Diagrama ontológico del dominio del perfil académico. . . . .	106
4.13	Diagrama ontológico del dominio del perfil socioeconómico. . . . .	107
4.14	Diagrama ontológico del dominio del perfil psicológico. . . . .	107
4.15	Captura de instancias de los perfiles de un estudiante en GraphDB. . . . .	109
4.16	Código SPARQL de inferencia de fusión de perfiles. . . . .	110
4.17	Visualización de la fusión de perfiles disjuntos de una persona que cumplió el rol de estudiante y profesor. . . . .	110
5.1	Arquitectura del sistema. . . . .	115
5.2	Captura de uno de los cuestionarios ofrecidos sobre Moodle. . . . .	117
5.3	Visualización de los resultados psicológicos de un estudiante. . . . .	118
5.4	Visualización de ponderación de los aspectos psicológicos de un curso. . . . .	118
5.5	Organización de los grupos de aprendizaje colaborativo. . . . .	119

---

5.6	Modulos correspondientes al componente psicológico de la red ontológica. . . . .	124
5.7	Ilustración de las reglas que rigen la composición de grupos. . . . .	131
6.1	Distribución de rasgos de personalidad globales. . . . .	149
6.3	Distribución de rendimiento académico por carrera. . . . .	150
6.2	Distribución de rasgos de personalidad por carrera. . . . .	151
6.4	Convergencia de algunas configuraciones de ejemplo del perceptrón multicapa a la hora de predecir aprobados/reprobados en asignaturas individuales. . . . .	154

# Índice de tablas

2.1	Áreas de aplicación de la Inteligencia Artificial. . . . .	18
2.2	Posibles métricas para calcular distancias en redes de Kohonen. . . . .	38
4.1	Sección 1 perteneciente al documento de elicitación de requerimientos. . . . .	79
4.2	Sección 2 perteneciente al documento de elicitación de requerimientos. . . . .	80
4.3	Muestra de los términos de grupos más importantes de la Sección 3 perteneciente al documento de elicitación de requerimientos. . . . .	81
4.4	Extracto de los requerimientos funcionales. . . . .	82
4.5	Análisis de variables descriptivas de los recursos ontológicos candidatos. . . . .	84
4.6	Axiomas reutilizados de la ontología VIVO. . . . .	86
4.7	Axiomas de los módulos de la red OEA. . . . .	88
4.8	Reportes de origen procedentes del sistema SNA. . . . .	92
4.9	Estructura del modelo de datos del archivo 6_Estudiantes.csv. . . . .	97
4.10	Estructura del modelo de datos del archivo 1_y_9_Docentes.csv. . . . .	97
4.11	Conjunto de reglas del dialecto OWL Horst. . . . .	98
4.12	Modificaciones realizadas sobre el conjunto de reglas <i>OWL Horst</i> . . . . .	100
4.13	Resultados del proceso de migración de elementos no ontológicos de diversas fuentes. . . . .	104
4.14	Resultados de la población de clases según el dominio de los perfiles. . . . .	105
5.1	Resultado de la consulta SPARQL asociada al cuestionario IPIP-NEO-120. . . . .	123
5.2	Datos de los registros socioeconómicos de los estudiantes. . . . .	133
5.3	Resultado de la selección y normalización de los datos. . . . .	137
6.1	Planificación de las actividades colaborativas . . . . .	144
6.2	Resumen de factores, facetas y rasgos de personalidad. . . . .	145
6.3	Resumen de las 6 facetas de cada rasgo. . . . .	146
6.4	Resumen de una faceta de cada uno de los 5 rasgos. . . . .	146
6.6	Medias y desviaciones típicas de las notas grupales (NG), según distintos métodos de agrupamiento. . . . .	147
6.5	Registros extraídos de la base de datos experimental. . . . .	148
6.7	Aplicación del modelo GLM a la nota del trabajo grupal. . . . .	152
6.8	Hiperparámetros y rangos de valores. . . . .	153
6.9	Detalles de las configuraciones del perceptrón multicapa representadas en la Figura 6.4. . . . .	155



## **Parte I**

# **Introducción**



# Capítulo 1

## Ámbito, objetivos y planteamiento de la tesis

La evolución acelerada de las Tecnologías de la Información y la Comunicación (TICs) ha marcado el desarrollo de la historia de la humanidad desde mediados del siglo XX, proporcionando facilidad para el ejercicio de las tareas cotidianas del individuo [194]. Los cambios cualitativos producto de la innovación tecnológica han permeado todos los ámbitos de la vida política, social, económica y cultural con la denominada “Era de la Información”, definida por Castell (1999) como “*el periodo vinculado a los avances de las tecnologías de la información y la comunicación, en el cual el movimiento de la información se volvió más rápido que el movimiento físico*” [40].

Las TICs se han desarrollado a partir de los avances científicos, ejerciendo un rol globalizador promotor de una visión sobre la sociedad, que hace posible abordar los problemas comunes desde una nueva perspectiva de vivir, pensar y actuar. En su conjunto, ofrecen herramientas de información y comunicación que son el resultado del proceso integrador que surgió con el desarrollo de Internet y los dispositivos electrónicos. No obstante, de acuerdo con Tubella y Requena (2005) [251], la importancia de las TICs no radica precisamente en la tecnología sino en el alcance que dan al conocimiento, la comunicación y la información. En este sentido, han facilitado herramientas al diseño estructurado de nuevas formas de relación entre los individuos en diversos contextos culturales, que confluyen en la nueva sociedad global. Salinas (2003) [223] decía que “*el objeto de la sociedad de la información no está centrado en la capacidad del manejo de la información sino en las estrategias institucionales para transformar esa información en conocimiento*”.

El término “*Sociedad del Conocimiento*” fue utilizado por primera vez por Peter Drucker en su análisis sobre la sociedad postcapitalista [76], haciendo referencia a la necesidad de colocar al conocimiento en el núcleo central de las teorías económicas para la generación de riquezas. A partir de sus consideraciones surgieron importantes investigaciones [265] que ponen de manifiesto la fuerte correlación que existe entre el desarrollo de las sociedades del conocimiento y la sostenibilidad de las mismas. Decían Dattakumar *et al.* [67] que “*una sociedad del conocimiento es una sociedad que se alimenta de su diversidad y sus capacidades, siendo por tanto necesario esforzarse por consolidar nuevas formas de desarrollo, adquisición y difusión*”.

De acuerdo con [35], la sociedad del conocimiento es el resultado de la etapa evolutiva que sigue a la sociedad de la información, a la cual se accede mediante el aprovechamiento de las oportunidades generadas por las TICs. No obstante, según expone la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), “*la sociedad del conocimiento tiene la capacidad para identificar, producir, tratar, transformar, difundir y utilizar la información con vistas a crear y aplicar los conocimientos necesarios para el desarrollo humano y social*” [86], de ahí que jueguen un papel central “*los sistemas de gestión dedicados a producir conocimiento*”.

Durante décadas, el uso de las TICs ha permitido registrar, almacenar y recuperar información en múltiples ámbitos, pero no fue hasta mediados de los 90 que se empezó a vislumbrar un uso más eficiente de la misma, mediante procedimientos que permitirían establecer relaciones con los registros de información previa [4]. Así se iniciaron importantes investigaciones sobre la creación de conocimiento asistida por TICs, entendiendo “conocimiento” como “*un proceso dialéctico de carácter complejo que se desarrolla en etapas consecutivas directamente vinculadas entre sí; proceso que parte de la contemplación viva al pensamiento abstracto, siguiendo a la práctica; siendo esta la vía dialéctica del conocimiento de la verdad, es decir del conocimiento de la realidad objetiva*” [28]. Desde esta perspectiva, la información se puede examinar desde una orientación sintáctica (el volumen) o desde una orientación semántica (el significado), existiendo un amplio consenso en que esta última es la más importante [4, 114]. Las claves que ha de habilitar la tecnología pasan, pues, por conseguir que la información pueda ser comparada con otros elementos de igual ámbito, por permitir establecer conexiones o asociaciones con más información y por realizar predicciones de consecuencias o hechos futuros.

Desarrollar plenamente estas capacidades será clave para brindar ventajas competitivas en cualquier organización, y así son muchos los autores que han catalogado al conocimiento como el recurso intangible más valioso o como el insumo de vital importancia, mediante el cual se puede incrementar el rendimiento organizacional [1, 4, 67, 114]. En la presente tesis se abordan cuestiones relativas a la gestión del conocimiento de manos de las instituciones de Educación Superior, que son sin lugar a dudas uno de los pilares de la *Sociedad del Conocimiento*.

## 1.1. Gestión del conocimiento

En general, una de las claves para la innovación en la Sociedad del Conocimiento reside en cómo las organizaciones pueden crear conocimiento nuevo y diferente al ya existente a partir de la investigación de nuevas estrategias, o bien mediante la adecuada implementación de tecnologías innovadoras, y poder adquirir así la capacidad de síntesis necesaria. Bajo este paradigma, se trata de conocimiento vivo, que permanentemente se encuentra en un ciclo actualizable acorde con las necesidades de la sociedad [4].

Un aspecto muy importante en torno a la generación de conocimiento es que no todo el conocimiento nuevo de una organización presenta una ventaja competitiva, sino sólo aquellos elementos que aporten a la creación de valor organizacional. Aquel conocimiento nuevo que no es relevante para la organización no debe ser almacenado, ya que genera lentitud operacional. La *gestión del conocimiento* (del inglés *Knowledge Management* o GC) al día de hoy es un concepto ampliamente

utilizado en cualquier tipo de organización, que hace referencia precisamente a la capacidad de generar nuevos métodos para resolver problemas [4, 67].

Los orígenes de la GC se remontan a finales de la última década del siglo pasado, principalmente en el ámbito empresarial. En el transcurso del tiempo se han dado una amplia variedad de definiciones asociadas a la gestión del conocimiento, entre las que se pueden destacar las proporcionadas en [66, 69, 140]. En particular, Davenport [68] se refiere a un proceso que se lleva a cabo de manera sistemática y que comprende la búsqueda, organización, filtrado y socialización de la información con el objeto de mejorar su comprensión en un área específica. Desde una perspectiva estrictamente empresarial, Tejedor [243] considera la GC como un elemento clave para la competitividad en la Sociedad del Conocimiento, abarcando el conjunto de procesos que facilitan la utilización del conocimiento como un factor determinante para añadir y generar valor.

En relación con las etapas que comprenden el proceso de gestión del conocimiento, Medina [169] inicia con la identificación del conocimiento que se considera accesible o activo, diferenciando los conocimientos explícitos y los conocimientos tácitos. Los primeros representan la *memoria corporativa*, que permanecerán archivados y dispuestos para quien puedan sacarles partido, mientras que los segundos son un complemento destinado a identificar a las personas que poseen los primeros, mediante un índice o mapa de localización. Desde este punto de vista, la función de la gestión del conocimiento es hacer accesible el conocimiento a todos los miembros de la organización. Por lo tanto, las últimas etapas del proceso de GC consisten en la combinación del conocimiento para la creación de un nuevo conocimiento que será finalmente transferido, asimilado y utilizado por todos los miembros de la organización garantizando el desarrollo de su capacidad de innovación empresarial.

Por otra parte, Nonaka [191] habla del conocimiento como única fuente de ventaja competitiva segura y duradera. Teniendo como fundamentos las teorías del aprendizaje por experiencias, sostiene que el secreto del éxito de las empresas japonesas consiste en saber lo que es el conocimiento, su papel en la organización de negocios y qué deben hacer las empresas para explotarlo. Subraya los pasos de almacenamiento, transferencia y la necesaria transformación del conocimiento. En tal sentido, organiza estas etapas en lo que denomina la *“conversión del conocimiento resultado de la interacción sistémica entre el conocimiento tácito y el conocimiento explícito mediado por la socialización, externalización, combinación e internalización del mismo”*. Vinculado a esto, Duttay De Meyer [78] incorpora un elemento innovador a la GC, al definirla como la habilidad de las personas por entender y manejar información utilizando la tecnología y la compartición del conocimiento. Entre tanto, reconoce el gran aporte que desempeñan las TICs en cualquier organización al momento de facilitar el oportuno y eficiente proceso de la GC.

También considerando las TICs como pivote, Dalkir [66] argumenta que se trata de la coordinación intencionada y sistemática de las personas, la tecnología, los procesos y la estructura de una organización cuya finalidad no es otra que añadir valor por medio de la reutilización y la innovación. La coordinación a la cual se refiere es posible creando, compartiendo y aplicando conocimiento, así como nutriendo las valiosas lecciones aprendidas y las mejores prácticas en la memoria corporativa para fomentar el ciclo de aprendizaje continuo de la organización. Independientemente del tipo de organización, la GC permite generar ventajas competitivas, generar estrategias, identificar y valorar los recursos y capacidades disponibles de una organización. En la actualidad, la GC ha llegado a ser una actividad primordial para la

adaptabilidad de las organizaciones, para su supervivencia y competitividad. Esto en virtud de reconocer que se vive en un contexto cada vez más exigente, cambiante y discontinuo, por lo cual prima la necesidad de seguir desarrollando la capacidad de adaptación a través de la gestión de conocimiento.

En concordancia con esto, se ha acentuado cada vez más que las TICs pueden facilitar las actividades de gestión del conocimiento, no sólo por medio de la capacidad para la codificación del conocimiento, sino también facilitando diversas formas de comunicación fáciles e interactivas que permitan el intercambio y la creación de conocimiento. Esto es más visible en la investigación sobre el papel que las tecnologías de medios sociales pueden desempeñar para facilitar las actividades de gestión del conocimiento. Recientemente, las tecnologías de *Big Data* (manejo de grandes volúmenes de datos) y *data mining* (minería de datos) se han incorporado en el espectro de las tecnologías de gestión del conocimiento, favoreciendo el descubrimiento de conocimiento en grandes cantidades de información y, por lo tanto, respaldando los procesos de generación de conocimiento en las organizaciones. En efecto, los postulados expuestos nos llevan de vuelta a los orígenes de la gestión del conocimiento a través de las TICs, ya que el análisis de estos datos es complementado a través de la incorporación o uso de diferentes técnicas de Inteligencia Artificial que serán objeto de atención en el siguiente capítulo de esta tesis: aprendizaje automático, redes neuronales, sistemas recomendadores, algoritmos genéticos, etc.

## 1.2. Gestión del conocimiento en las instituciones de Educación Superior

En los últimos años, dentro del ámbito de la Educación Superior, la gestión de conocimiento se ha ido tornando como una actividad ardua y compleja. El principal motivo de esta situación es la gran cantidad de datos, de diversa índole, que se generan continuamente, producto de las diversas actividades desarrolladas por los diferentes actores y procesos del sistema académico: estudiantes, docentes, personal administrativo, autoridades, etc. [95,180]. En este contexto, la holística de la gestión del conocimiento de la Educación Superior tiene una gran magnitud, y su misión principal es poder generar o mantener un vínculo entre la información que se genera en los diferentes subsistemas y poder aprovechar el potencial de este valioso recurso, pero permitiendo que cada instancia conserve su autonomía de gestión.

Como mecanismo de soporte a esta actividad, las instituciones de Educación Superior (en adelante, IES) buscan continuamente implementar adecuados avances tecnológicos como parte integral de su estrategia para dar mejor respuesta a situaciones cotidianas. El análisis y la implementación de los diferentes avances tecnológicos dentro de las estructuras de las instituciones educativas ha permitido incrementar la componente de innovación en el entorno educativo, proporcionando soporte para recuperar, acceder y transformar el conocimiento asociado al ámbito académico y administrativo.

Los referentes teóricos, de manera general, reconocen una adecuada gestión de conocimiento a nivel institucional. Esta gestión la conforman un conjunto mucho más amplio de aplicaciones independientes, que de manera específica gestionan información cubriendo aspectos muy particulares relacionados con el entorno aca-

démico. Este tema presume un gran interés, dado que al tener una diversidad de formatos no estructurados de diferentes fuentes de información se genera una restricción de interoperabilidad entre ellos. Tal particularidad reduce la capacidad de administrar y explotar de manera eficiente todo el potencial de esta información que pudiese transformarse en conocimiento, para ser utilizado de manera oportuna y relevante por los diferentes actores del ecosistema académico. Si bien en la actualidad existen varios sistemas de gestión del conocimiento en el campo de la Educación Superior centrados en resolver estas dificultades (sistemas transaccionales que permiten asociar información entre diferentes subsistemas, sistemas de minería de datos y *data warehouse*, entre otros), la información recuperada carece de relaciones y anotaciones de consenso que permitan a los subsistemas identificar incluso si dos piezas de información se relacionan con el mismo tema o entidad.

En virtud de estos antecedentes, con el propósito de ayudar a las universidades a garantizar una gestión eficiente, compartir, buscar y reutilizar el conocimiento, en esta tesis se ha considerado la oportunidad de emplear tecnología semántica mediante la creación de una ontología académica e incorporar información semántica a los datos que son generados dentro del ecosistema educativo. Esta concepción permite romper los esquemas tradicionales de cómo se veía a la información como simples datos a pasar a ser entidades con su propia conceptualización, estableciendo relaciones independientemente de con qué aplicación hayan sido generados. Sobre esta base, la tesis presenta procedimientos para evaluar la capacidad de resolver algunas situaciones que son de suma importancia en los entornos educativos.

Sobre la base de estas evidencias teóricas y empíricas para el desarrollo de la presente investigación se analizan las fortalezas de la gestión del conocimiento en la Educación Superior respaldada por tecnologías semánticas, que serán debidamente analizadas en el siguiente capítulo. De igual manera, se sugieren algunas vías para apoyar esta actividad dentro de las instituciones de Educación Superior. La idea de utilizar la tecnología semántica es permitir a las universidades mejorar los mecanismos de búsqueda de información y producción de conocimiento.

### 1.3. Problemática principal de la tesis

Desde el punto de vista organizativo, las instituciones de Educación Superior cuentan por lo general con plataformas digitales para el manejo de la información correspondiente a los perfiles de los estudiantes, docentes, personal administrativo, autoridades y demás, organizadas en forma de bases de datos relacionales. Igualmente, emplean diferentes aplicaciones que generan continuamente una gran cantidad de datos [182]. La información que se genera usualmente está distribuida entre diferentes departamentos, de acuerdo con el tipo de información que gestionan [208]. Cada departamento cuenta a menudo con sus propias tecnologías de información y de almacenamiento, por lo tanto no siempre manejan las mismas estructuras de repositorios digitales, bases de datos o formatos [180].

La mayor cantidad de datos generados en torno a las IES son derivados de la interacción entre docentes y alumnos a través del proceso enseñanza-aprendizaje [182, 185]. Estos datos son el resultado del desarrollo de diversas tareas (evaluaciones, proyectos, etc.) previamente planificadas dentro de un programa curricular o plan de estudios, y ejecutadas durante el transcurso de cada período académico. El resto

de datos corresponden al área administrativa, relacionada a diversos departamentos tales como los siguientes:

- **Departamento de Recursos Humanos.** Encargado de gestionar toda la información asociada a los perfiles profesionales de quienes laboran dentro de la institución, así como de la gestión de permisos, vacaciones, capacitaciones, tipos de contrato, duración de contratos, y otras actividades asociadas a esta área.
- **Departamento de bienestar universitario.** Responsable de administrar la información de los estudiantes referente a su estado social, situación familiar, número de integrantes de la familia, situaciones relacionadas con tipos de discapacidad, problemas afectivos, situación laboral, ... entre otros aspectos que pudiesen condicionar el contexto y el rendimiento de un estudiante.
- **Departamento financiero.** Responsable de manejar toda la información financiera institucional asociada a los ingresos y egresos económicos. Dependiendo del tipo de institución educativa (pública o privada), maneja las aportaciones por parte de los estudiantes y sus cuentas para con la institución.
- **Departamento académico.** Encargado de registrar y gestionar toda la información proveniente de las actividades académicas en el proceso enseñanza-aprendizaje, donde queda constancia explícita de los registros académicos de los estudiantes, con cada una de sus evaluaciones, a nivel micro-curricular (por actividad, tarea, prueba, proyecto, etc.) y macro-curricular (promedio por asignatura, año, o general).

Dependiendo de la estructura de cada IES pudiesen existir departamentos con otras denominaciones u otros adicionales, pero lo común en cualquier institución es que cada departamento se encarga de administrar y gestionar información acorde a su función. Esta información forma parte vital de las IES, ya que a través de los diferentes sistemas que utilizan en cada departamento permiten que los diferentes actores del ecosistema (personal administrativo, docente, estudiantes, etc.) puedan realizar consultas, análisis, programación de actividades, gestión de recursos, toma de decisiones y más acciones asociadas al entorno académico.

Dentro de este marco, las instituciones de Educación Superior pueden beneficiarse significativamente de la aplicación de sistemas de Inteligencia Artificial que promuevan la interconexión de múltiples repositorios de información que repercutan directamente en aspectos tales como el dimensionamiento de recursos, la planificación de las asignaturas por parte del claustro profesoral, en la selección de asignaturas por parte de los estudiantes, en la conformación de grupos de trabajo, en la organización de la carga horaria de unos y otros, etc.

Las propuestas basadas en el uso de las tecnologías existentes para brindar apoyo al adecuado modelado y gestión del conocimiento, tanto de la información administrativa de las instituciones de Educación Superior como de la información académica, presentan algunas limitaciones relacionadas en el campo de aplicación. Estas limitaciones se generan ya sea por el tratamiento para abordar el problema o por las metodologías que utilizan. A continuación, de manera previa al análisis más detallado que se llevará a cabo en el siguiente capítulo de esta tesis, se exponen algunos de los aspectos más importantes que no se han considerado o limitaciones que presentan en los desarrollos actuales:

- No toman en cuenta las complejas relaciones que existen entre la información que posee cada uno de los diferentes departamentos de la institución.
- No permiten dividir y explicitar el desarrollo que los diferentes actores del sistema académico de Educación Superior (alumnos-docentes) toman a lo largo del tiempo.
- Gran parte del conocimiento está disperso por la diversidad de formatos no estructurados que se utilizan, al igual que por la existencia de diferentes bases de datos o repositorios digitales no conectados.
- No posibilitan generar una relación semántica en base al histórico de la información de los estudiantes, para desarrollar estrategias que mejoren el rendimiento académico disminuyendo la tasa de repitencia o deserción estudiantil.
- No existe una definición formal del conocimiento asociado al ámbito de Educación Superior que faculte mejorar los procesos que exploran desarrollar herramientas de soporte.

Es posible observar que estos criterios son primordiales para el diseño y construcción de cualquier sistema de soporte al contexto de la Educación Superior, puesto que si no son considerados, van en detrimento de la calidad de las actividades que se llevan a cabo dentro del ámbito académico y administrativo de la institución.

## 1.4. Objetivo y contribuciones

La presente tesis tiene como objetivo general forjar las bases de un completo ecosistema de apoyo a la administración y gestión de la información que se genera en torno al ámbito de la Educación Superior, desde una perspectiva fundamentada en el modelado semántico del conocimiento y el empleo de innovadoras técnicas de Inteligencia Artificial, brindando un diseño altamente flexible e interoperable.

Con esta propuesta se espera contribuir de manera eficiente a la denodada labor que llevan adelante las instituciones de Educación Superior y, particularmente, a los aspectos relacionados con el modelado formal del conocimiento, la predicción del rendimiento académico, y la eficiente formación de grupos de trabajo colaborativo. Con ello, se busca que los docentes, estudiantes y personal administrativo puedan llevar a cabo de una forma más efectiva sus múltiples actividades y tareas, aspecto que permitirá que se promueva una mejor y más eficiente gestión del conocimiento generado en el contexto de la Educación Superior, en especial con aspectos específicos asociados al rendimiento académico.

Las contribuciones de este trabajo de investigación se enmarcan en tres grandes áreas:

- **Modelado formal del conocimiento del ecosistema académico de Educación Superior.** En primer lugar, se ha creado e implantado una red ontológica para integrar la información generada en las distintas áreas del entorno académico, dando cuenta de las cuestiones de representación, interoperabilidad y consolidación.
- **Formación eficiente de grupos de trabajo colaborativo.** Aprovechando el modelado formal de los estudiantes y de los planos de estudios a nivel micro-

y macroscópico, hemos implementado un sistema basado en algoritmos genéticos con el que apoyar la formación de grupos de trabajo colaborativo que potencien los resultados de aprendizaje.

- **Predicción del rendimiento académico.** Igualmente, con base en el modelado formal del conocimiento del ecosistema académico, hemos llevado a cabo un estudio sobre el potencial de aplicar técnicas de aprendizaje profundo (“*deep learning*” en inglés) para la predicción del rendimiento académico de los estudiantes, a partir de características socioeconómicas, con el objetivo de apoyar estrategias de planificación y dotación de recursos.

Las contribuciones se han validado en el marco concreto de la Universidad Politécnica Salesiana (UPS) de Ecuador, con acceso a los datos acumulados en sus sistemas de información desde el inicio de su actividad, más otros recabados al efecto de nuevos experimentos, en los que se ha contado con la participación activa de estudiantes y docentes de dicha universidad. Las técnicas y los resultados se consideran de aplicación general, en primer lugar, a otras universidades del ámbito de Latinoamérica debido a las similitudes entre los modelos estructurales y organizativos; no obstante, los fundamentos de las distintas propuestas serían fácilmente trasladables también a la realidad de las instituciones de Educación Superior en otras partes del mundo.

## 1.5. Estructura y organización de la memoria

La memoria de la tesis se encuentra organizada en función de la siguiente estructura:

- En la Parte I se aborda una introducción y descripción general del trabajo propuesto, incluyendo la problemática tratada (Capítulo 1).
- En la Parte II se da una introducción a la Inteligencia Artificial y a las principales técnicas empleadas en esta tesis (Capítulo 2), junto con una recopilación, análisis y categorización de las principales contribuciones de dichas técnicas en instituciones de Educación Superior (Capítulo 3).
- La Parte III contiene los capítulos que describen cada una de las contribuciones realizadas al ámbito académico.
  - En el Capítulo 4 se exponen todos los elementos que tienen que ver con el diseño y el modelado del conocimiento del ecosistema académico, a través de una red ontológica.
  - En el Capítulo 5 se presenta el empleo de la red ontológica como base para desarrollar dos propuestas de aplicación de Inteligencia Artificial al área que nos ocupa: algoritmos genéticos para dirigir la formación de grupos de aprendizaje colaborativo, y redes neuronales de aprendizaje profundo para predicción de rendimiento académico.
  - El Capítulo 6 presenta los resultados de una serie de experimentos, llevados a cabo con la participación de docentes y estudiantes universitarios, para validar las dos propuestas del punto anterior en diversos escenarios.

- La Parte IV contiene el Capítulo 7, donde se describen las principales conclusiones extraídas de esta tesis y se especifican algunas líneas de acción futura.



**Parte II**

**Estado del arte**



## Capítulo 2

# Estado del arte en áreas seleccionadas de Inteligencia Artificial

*Las instituciones de Educación Superior, al igual que muchas otras organizaciones, para mantenerse a la vanguardia han optado por incorporar diferentes tecnologías de la información y comunicaciones que les permitan ser más eficientes dentro de sus procesos. Bajo esta realidad, en el contexto académico las que mayor demanda han presentado, específicamente, son las tecnologías derivadas de la Inteligencia Artificial, las cuales han permitido mejorar y complementar el ecosistema de la Educación Superior. En el presente capítulo se describen algunas de las tecnologías derivadas de la inteligencia artificial que se emplean como soporte en este ámbito, exponiendo un conjunto de usos y aplicaciones y describiendo con mayor énfasis aquellas técnicas que se emplearán como insumo para el desarrollo de las contribuciones de esta tesis.*

### 2.1. Introducción a la Inteligencia Artificial

La Inteligencia Artificial (IA) es un área de la informática que se ocupa del diseño de sistemas de computación que simulan características asociadas a la inteligencia y el comportamiento humano, tales como comprensión del lenguaje, razonamiento, aprendizaje, resolución de problemas, etc. Desde el ámbito académico, la IA es el área de la ciencia que estudia el desarrollo de técnicas y algoritmos para crear máquinas con capacidades similares a las que posee el ser humano, tales como el razonamiento lógico, conocimiento, planificación, percepción (audición, visión, etc.) o capacidad de procesamiento del lenguaje natural, entre otras [162]. Expuesto de otra manera, la IA busca construir elementos que desarrollen comportamientos que, si fuesen desarrollados por un ser humano se considerarían *inteligentes* [23]. Desde el ámbito de la ingeniería se promueve la creación de sistemas, dispositivos o máquinas que integren un comportamiento inteligente.

Generalmente, los principios de IA se aplican en procesos de reconocimiento de patrones, tratamiento de datos e identificación de sistemas. Se trata de un área altamente interdisciplinar, a la que pueden vincularse disciplinas tan variadas co-

mo la educación, la neurociencia, la biomedicina, la robótica, la computación y la psicología, entre otras.

### 2.1.1. Breve historia de la IA

Suele citarse como punto de partida para la IA una conferencia sobre informática celebrada en el *Dartmouth College* (Nuevo Hampshire, Estados Unidos) en 1956 [56,176], aunque el marco científico-tecnológico estaba propiciado por trabajos teóricos de investigadores anteriores como Charles Babbage (inventor de la máquina analítica, 1843) [211], Kurt Gödel (teorema de la incompletitud, 1930) [232], Alan Turing (máquina universal, 1936) [115], Norbert Wiener (cibernética, 1943) [34] y John von Neumann (arquitectura del computador, 1950) [42]. En la conferencia de 1956, Allen Newell y Herbert Simon dieron a conocer su programa *Logic Theorist*, que se consagró como el primer software de inteligencia artificial por su capacidad de demostrar una parte de los teoremas sobre la lógica matemática usando métodos heurísticos [56,187].

El teorema de la incompletitud de Gödel fue de gran importancia, ya que estableció un conjunto de teoremas de lógica matemática que indican las limitaciones relacionadas a un sistema basado en procedimientos lógicos y reglas. Mediante estos avances, se realizaron las primeras aproximaciones en una máquina por determinar un comportamiento inteligente, que sirvieron de base para el artículo publicado por Alan Turing en 1950 en la revista *Mind*, bajo el título de “*Computing machinery and intelligence*” [253]. En este artículo se plantea un test de inteligencia para máquinas, según el cual una máquina mostraría un comportamiento inteligente en la medida en que fuese capaz de mantener un diálogo con una persona sin que esta pudiese distinguir si se trata de un humano o un ordenador. El *test de Turing* puso de manifiesto los incipientes intentos de lograr una definición objetiva de la inteligencia en aquellos años.

Para el año de 1954, fue introducida la máquina IBM 704 como la primera de producción en masa, y con ella se desarrollaron numerosos lenguajes de programación específicamente orientados a implementar sistemas de inteligencia artificial. Posteriormente, en 1958, en el Massachusetts Institute of Technology (MIT), John McCarthy y Marvin Minsky fundaron el Laboratorio de Inteligencia Artificial [148], donde comenzaron a trabajar en lo que denominaron *List Processing Language* (LISP), un lenguaje de computación que pronto se convirtió en el estándar para el desarrollo de la inteligencia artificial y para el ámbito de la investigación.

Tras las incipientes propuestas en el entorno de la IA de los años 50, en la década de los 60 se realizó un gran esfuerzo de formalización matemática de los métodos empleados por dichos sistemas. Más tarde, como respuesta al test de Turing, en los años 70 se produjo el nacimiento del campo del *procesamiento del lenguaje natural* (*Natural Language Processing*, *NLP*, por sus siglas en inglés) [50], con el objetivo de crear sistemas capaces de mantener una conversación y generar respuestas inteligentes al interactuar con una persona. El NLP abrió diversas áreas de investigación entre la inteligencia artificial y la lingüística, como por ejemplo la *desambiguación semántica*, esto es, el discernimiento de las múltiples acepciones que puede tener una misma palabra en función del contexto en que se utiliza [113,129]. Uno de los hitos principales fue el desarrollo de la *lógica difusa* (*Fuzzy Logic*), que hace posible que las máquinas puedan manejar conceptos vagos o imprecisos del mundo real en una escala no binaria entre lo verdadero y lo falso [273]. Esta particularidad permite a las

máquinas trabajar con información imprecisa o que no se encuentra completamente definida, pudiendo así juzgar y evaluar la información que es “*imprecisa, flexible y analógica*” [237].

A lo largo de la historia los intereses en el área de la inteligencia artificial se han ido redefiniendo, ya no solamente enfocados a abordar el test planteado por Turing, sino a complementar otros aspectos asociados al empleo de la IA en diversos ámbitos productivos. Así, para los años 80 el uso de la IA se centró en desarrollar los primeros sistemas comerciales, principalmente orientados a soluciones de producción, contabilidad, calidad y control de procesos. Mediante el desarrollo de estos sistemas aparecieron los denominados *sistemas expertos* (*Expert Systems, ES*, por sus siglas en inglés) [82], que brindaban soporte en tareas de toma de decisiones y diagnóstico en función de la información aportada por profesionales humanos, en diversas áreas de aplicación. Un sistema experto simula los procesos de aprendizaje, de memorización, de razonamiento, de comunicación y de acción que realizaría un profesional con conocimientos especializados [3, 128]. De manos de los sistemas expertos se generalizó el uso de *redes bayesianas*, grafos acíclicos dirigidos cuyos nodos corresponden a variables aleatorias y cuyos arcos describen dependencias causales entre ellos, más o menos acusadas dependiendo de los valores recogidos en una tabla de probabilidades condicionales [74, 130].

En la década de 1990 surgió una nueva categoría, denominada como *agentes inteligentes* (*Intelligent Agents, IA*, por sus siglas en inglés) [266], referida a programas capaces de percibir su entorno, procesar tales estímulos y comportarse en el contexto de forma racional, posiblemente en cooperación con otros. Mediante el uso de sensores, los agentes inteligentes son capaces de percibir su medio ambiente y de comportarse o responder de una forma adecuada, asociando a cada estímulo una acción que se llevará a cabo por medio de dispositivos actuadores. Por otra parte, se profundizó significativamente en la problemática del *aprendizaje automático* (*Machine Learning*), que emplea técnicas estadísticas y algoritmos diversos para brindar a una máquina la capacidad de aprender emulando el comportamiento biológico de reconocimiento, aprendizaje y generalización. Los sistemas de aprendizaje automático se entrenan manipulando datos basados en la experiencia (a menudo de manera progresiva), sin que todas las respuestas estén explícitamente programadas; por ello se les conoce también como *sistemas de inducción al conocimiento* [173, 224].

Desde entonces y hasta la actualidad, el interés se ha centrado en desarrollar e implementar sistemas inteligentes que posean la capacidad de analizar y gestionar grandes volúmenes de datos de manera oportuna y eficiente, a fin de ayudar a las personas y a todo tipo de organizaciones a hacer frente a la exposición a una cantidad creciente de información. Así, la evolución de las técnicas de la IA aplicada ha permitido concebir novedades tales como el *Big Data* [210], la minería de datos (*Data Mining, DM*, por sus siglas en inglés) [214], la gestión del conocimiento (*Knowledge Management, KM*) [247] y la inteligencia empresarial (*Business Intelligence, BI*) [47]. Estos mecanismos buscan poder extraer información relevante de volúmenes de información antaño inmanejables con la finalidad de mejorar y optimizar tomas de decisiones y procesos en todos los niveles de la vida de una persona y la actividad de una organización. Todas estas tecnologías persiguen explorar grandes bases de datos de manera automática o semiautomática, con la finalidad de descubrir nuevo conocimiento que no se puede detectar mediante la exploración tradicional [27] y explotar en la práctica lo aprendido. La clave estriba en identificar patrones repetitivos o tendencias que permitan adquirir conocimiento nuevo, útil y relevante, con la

finalidad de comprender el comportamiento de los datos en un determinado entorno y, a partir de ahí, mejorar los procesos organizacionales y dar soporte a la toma de decisiones [157,164].

Actualmente, son muchos los campos del conocimiento en los que se emplean sistemas de IA. La Tabla 2.1 resume apenas un puñado de campos y aplicaciones, destacando el área que ha sido objeto de la investigación de esta tesis: la educación. A continuación se revisan los principios operativos de algunas de las tecnologías más relevantes (tecnologías de Web Semántica, redes neuronales artificiales y algoritmos genéticos), de manera previa al análisis de trabajos anteriores relacionados con su aplicación a problemas concretos de interés en el ámbito específico de la Educación Superior.

Tabla 2.1: Áreas de aplicación de la Inteligencia Artificial.

<i>Área</i>	<i>Aplicaciones</i>
Robótica	Sistemas de visión artificial Sistemas de rehabilitación Robots humanoides
Ingeniería	Cálculo de estructuras Automatización de procesos de producción Diagnóstico de fallos Sistemas de control
Medicina	Soporte para el diagnóstico Procesado de imágenes biomédicas Medicina personalizada Análisis de señales fisiológicas
Biología	Redes neuronales biológicas Genética médica y molecular Estudio de estructuras biológicas
Economía	Estudio de riesgos Análisis financiero y bursátil Mercadotecnia y fidelización de clientes
Educación	<i>Modelado del conocimiento académico</i> <i>Sistemas de tutoría inteligentes</i> <i>Predicción de rendimiento académico</i> <i>Conformación de grupos de trabajo colaborativo</i> <i>Juegos serios como soporte de aprendizaje</i>

## 2.2. Tecnologías de Web Semántica

La adopción de Internet por parte de más de la mitad de la población mundial (según *Global Digital Reports* [260]) ha cambiado la forma en que las personas se relacionan y comunican a todos los niveles, tanto personal como profesionalmente. La cantidad de información accesible a través de la web no deja de crecer, y surge la necesidad de estructurarla de manera que sea posible acceder de manera eficaz al conocimiento que interese en cada momento. En esta línea, la *Web Semántica* surgió como un conjunto de actividades, lideradas por el W3C (*World Wide Web Consortium*), para crear tecnologías de publicación de datos que hagan el contenido de la

web más fácilmente procesable por máquinas, habilitando así mecanismos de Inteligencia Artificial que interpreten el significado de la información a través de procesos análogos a los que emplean los seres humanos para dar sentido a las expresiones lingüísticas, visuales y sonoras. El objetivo último es potenciar los mecanismos de búsqueda de información, estableciendo referencias comunes que permitan compartir y reutilizar conocimiento entre personas y organizaciones de todo tipo [244].

A la hora de plantear que un programa informático interprete una determinada información, es necesario partir de una conceptualización del dominio de aplicación concreto en que esa información cobra significado. Según Genesereth y Nilsson [96], la conceptualización de un cuerpo de conocimiento que se representa formalmente ha de definir objetos y conceptos, además de las relaciones existentes entre ellos. El significado de “*representación*” está estrechamente relacionado con la sintaxis y la semántica: la sintaxis juega el papel de normalizar el conjunto de símbolos que se pueden utilizar y las reglas de uso de esos símbolos, mientras que la semántica ayuda a poder interpretar el significado de las expresiones formadas por dichos símbolos [22].

Mediante una estructura relacional extensible, Genesereth y Nilsson explican la naturaleza de la conceptualización en el campo de la inteligencia artificial, definiendo una tupla (D,R) en donde:

- D representa el universo de discurso.
- R es el conjunto de relaciones sobre el universo de discurso.

A modo de ejemplo de esa estructura, presentamos a continuación un escenario similar al propuesto por Guarino [104] para la administración de recursos humanos en una gran compañía de software, sólo que replanteado para una red de investigación en un instituto de Educación Superior. Se toma como universo de discurso a cada una de las personas que conforman la red mediante un número único y un prefijo con la letra I, mientras que las relaciones unarias y binarias se representan con la letra R:

- $D = \{I00001, \dots, I00004, \dots\}$
- $R = \{Persona, Investigador, Estudiante-Investigador, coopera-con\}$

Extendiendo el conjunto de relaciones R sobre el universo de discurso D se modela la conceptualización de la realidad que se está plasmando. Partiendo de las relaciones unarias definimos que la relación *Persona* se extiende sobre todo el conjunto D, mientras que *Investigador* y *Estudiante-Investigador* se extiende en un subconjunto de D. La relación binaria *coopera-con* se articula como un conjunto de tuplas que especifica la colaboración que tiene un elemento del universo de discurso con otro elemento.

- $Person = D$
- $Investigador = \{\dots, I00004, \dots\}$
- $Estudiante-Investigador = \{\dots, I00012, \dots\}$
- $coopera-con = \{\dots, (I00004, I00012), \dots\}$

Partiendo del mecanismo que ofrece una conceptualización formal, se han propuesto diferentes sistemas de representación de conocimiento, que Santos [22, 199] clasificaba conforme se indica a continuación:

- *Lógica*. La representación lógica se basa en determinar si un argumento es aceptable o no, sabiendo que depende de un conjunto de premisas. Definir un lenguaje lógico implica tanto la descripción de la sintaxis como de la semántica. La sintaxis se refiere a las posibles combinaciones de uso de los símbolos en sentencias válidas, mientras que la semántica implica la relación que tiene cada representación de las sentencias con la conceptualización de base. Un lenguaje de primer orden puede representar tanto objetos como entidades, propiedades de las entidades y relaciones entre entidades, lo cual basta para modelar un dominio de la realidad que son representaciones de fácil entendimiento para los humanos pero a su vez responden a una lógica proposicional que puede generar inferencias.
- *Redes semánticas*. Las redes semánticas son estructuras basadas en grafos que unen conceptos por medio de relaciones. Los conceptos se representan por medio de palabras que se interrelacionan. Las relaciones introducen el concepto de *herencia*, puesto que una clase B que se relaciona por medio de “*subclase*” con una clase A hereda todos los atributos de esta. Las redes semánticas nacieron con el propósito de entender de una manera más fácil el lenguaje natural, representando el conocimiento que está implícito en las oraciones.
- *Marcos*. Los marcos (en inglés, *frames*) tomaron como punto de partida a las redes semánticas para su desarrollo, pero son mucho más estructurados y, por tanto, más versátiles y menos propensos a problemas de interoperabilidad o de expresividad limitada. Son un método de representación de conocimiento en el que una clase es definida por un conjunto de atributos que tienen ciertos valores, que incluso pueden tomar de otro marco. También son estructuras jerárquicas, de modo tal que los valores por defecto o los valores específicos que define una clase para cada atributo se pueden especializar en cualquiera de sus subclases.
- *Guiones*. Los guiones (en inglés, *scripts*) son estructuras que definen eventos que componen situaciones representando conocimiento. Un guión está compuesto por objetos, agentes, condiciones de entrada, secuencias de escenas y resultados. Un ejemplo habitual (tomado de [22]) es el de un cliente en un restaurante, donde podemos decir que el cliente fue a un restaurante X, pidió un plato del menú, pago la cuenta y salió; con una conceptualización subyacente adecuada, se puede inferir que el cliente tenía dinero y hambre.

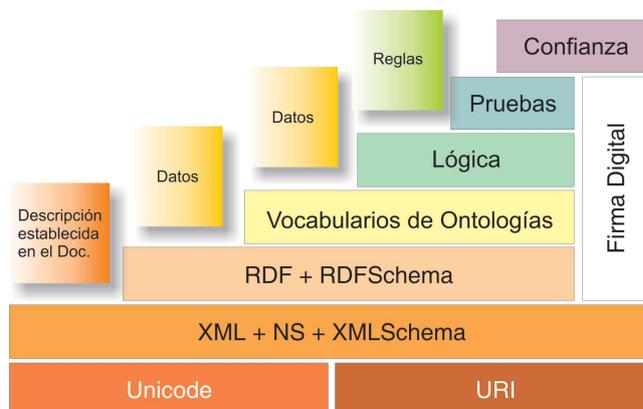
### 2.2.1. Arquitectura de la Web Semántica

Sobre la base que definen las conceptualizaciones y los sistemas de representación de conocimiento, la arquitectura de la Web Semántica —debida a Tim Berners-Lee [25]— se divide en múltiples capas, entendidas con una extensión de la Web tradicional. Dichas capas se muestran en la Figura 2.1 y se describen en las siguientes subsecciones.

#### 2.2.1.1. Capa XML

La capa de XML (*eXtensible Markup Language*) es la base común utilizada en el dominio de la Web. Se trata de un lenguaje de marcado derivado de SGML (*Standard*

Figura 2.1: Las capas de la arquitectura de la Web Semántica. [25]



*Generalized Markup Language*), que utiliza etiquetas (*tags*) que pueden estar anidadas. Cada etiqueta que se abre tiene una correspondiente etiqueta de cierre, de modo que el contenido encerrado entre ambas forma lo que se denomina *elemento*. En este sentido, XML es similar al lenguaje HTML que define el contenido y la estructura de una página web. HTML, no obstante, no está diseñado para ser entendible por una máquina, mientras que ese es precisamente el objetivo de XML: definir un lenguaje que especifica cada una de las piezas de información que puede contener un elemento, junto con restricciones de los tipos de datos que puede representar un objeto. Un documento XML está formado de un prólogo, un número de elementos y opcionalmente un epílogo.

### 2.2.1.2. Capa RDF y RDF Schema

XML no se ocupa del cómo ha de interpretarse el significado de cada objeto u elemento de un documento. Ese papel corresponde a las normas RDF<sup>1</sup> y RDF Schema (RDFS)<sup>2</sup> como el medio formal para interpretar la sintaxis que se define en XML [14].

RDF tiene una sintaxis abstracta que refleja un modelo de datos basado en gráficos simples y una semántica formal con una noción de vinculación rigurosamente definida, que proporciona una base para deducciones bien fundamentadas [14]. RDF está diseñado para representar información de manera mínimamente restrictiva y flexible. Se puede usar en aplicaciones aisladas, donde los formatos diseñados individualmente pueden ser más directos y fáciles de entender, pero su generalidad permite hacer que la información sea accesible para el uso por terceros. Como características más reseñables de su diseño, se pueden mencionar las siguientes [9]:

- Tiene un modelo de datos simple, que las aplicaciones pueden procesar y manipular fácilmente, y que es independiente de cualquier sintaxis de serialización específica.

<sup>1</sup><https://www.w3.org/RDF/>

<sup>2</sup><https://www.w3.org/TR/rdf-schema/>

- Tiene una semántica formal que proporciona una base confiable para razonar sobre el significado de una expresión RDF. En particular, apoya nociones de vinculación rigurosamente definidas que proporcionan una base para definir reglas de inferencia.
- El vocabulario es totalmente ampliable, ya que se basa en URIs. Las referencias de URI se utilizan para nombrar todo tipo de cosas en RDF. El otro tipo de valor que aparece en los datos RDF es el *literal*.
- RDF tiene un formato de serialización recomendado en XML, que se puede usar para codificar el intercambio de información entre aplicaciones.
- RDF puede usar valores representados según tipos de datos definidos con XML Schema, lo que ayuda al intercambio de información entre RDF y otras aplicaciones XML.
- Para facilitar el funcionamiento a escala de Internet, RDF es un marco de mundo abierto que permite a cualquier persona hacer declaraciones sobre cualquier recurso.

**N-Triple:** La forma más simple de representación de información expresada en RDF es la llamada *N-Triple*, la cual utiliza el texto completo de cada elemento encerrado entre los símbolos de *mayor que* y *menor que*. La sintaxis de los componentes debe ser una *tripleta* (también *tripla*) representada por el sujeto, predicado y objeto seguido de un punto para denotar que es el final de la declaración. Un ejemplo de representación de una tripleta se presenta a continuación:

```
<http://www.research-ups.edu.ec/person0001>
<http://www.w3.org/2000/01/rdf-schema#type>
<http://www.research-ups.edu.ec/ontology/Person>.
```

**Notation 3 RDF (N3):** La Notation 3 utiliza la visualización de tripletas como N-Triple pero con la introducción de *qnames*, que son prefijos de abreviación. Como primer paso, se definen los enlaces a las URI que son locales y globales de la siguiente manera:

```
@prefix rdf: <http://www.w3.org/2000/01/rdf-schema#type>
@prefix ro: <http://www.research-ups.edu.ec/ontology/>
@prefix con: <http://www.research-ups.edu.ec/>
```

Partiendo de estas abreviaturas se puede reescribir el ejemplo en N-Triple de la siguiente forma, que redundante en una mayor legibilidad:

```
con:person0001 rdf:type ro:Person
```

Una práctica muy común en N3 para representar múltiples declaraciones con el mismo sujeto es la de escribir la primera declaración como sujeto, predicado y objeto, pero usando punto y coma en lugar de punto. Después de la primera declaración, generalmente, se da un espacio de tabulación y se escribe el siguiente predicado y

objeto, poniendo al final punto y coma. Solo para la última declaración se utiliza el punto, como se muestra a continuación:

```
con:person0001 rdf:type ro:Person;
ro:name "Cristina Juna";
ro:age "2018-11-11T12:00:00Z"xsd:TimeStamp.
```

**RDF/XML:** La notación de N3 es una de las más compactas y convenientes, pero muchas infraestructuras web ya vienen trabajando con la representación de información por medio de XML. La serialización de XML para RDF (RDF/XML) del ejemplo anterior se presenta a continuación:

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/2000/01/rdf-schema#type"
xmlns:ro="http://www.research-ups.edu.ec/ontology/"
xmlns:con="http://www.research-ups.edu.ec/">
  <ro:Person
rdf:about=con:person0001 >
    <ro:name>"Cristina Juna"</ro:name>
    <ro:age>22</ro:age>
  </ro:Person>
</rdf:RDF>
```

Como RDF es independiente del dominio de aplicación, los usuarios deben definir su contexto por medio de RDF Schema, con la posibilidad de definir los tipos de objetos, relaciones y dominios que corresponda [14].

Los axiomas propuestos tanto en RDF como en RDF Schema son utilizados para el razonamiento automático, el cual requiere un sistema lógico de primer orden para su funcionamiento. Por ello, se ha dotado de una semántica directa a manera de tripletas para solventar inferencias como  $Type(?x, ?y)$  [14].

En un sistema basado en inferencias la semántica directa es tan eficaz como las interacciones de sus patrones de axiomas básicos. Es así como en RDF Schema se puede decir que en una subclase C de B, en donde B es subclase de A, todas las instancias en C son elementos de las clases B y A —no puede haber discusión sobre este postulado de inferencia puesto que el axioma básico ya es un consenso descrito en el esquema para el dominio conceptual dado. Este sistema de inferencia encuentra su utilidad cuando en la Web Semántica se combinan datos provenientes de diferentes fuentes [9].

### 2.2.1.3. Ontologías

La palabra "*ontología*" puede ser descrita desde dos puntos de vista: desde la filosofía se puede decir que las ontologías tratan con la estructura de la realidad, explicando la causalidad y propiedades del ser, mientras que para la rama de la Inteligencia Artificial se definen como "*modelos formales que instrumentalizan la información en artefactos computacionales tomando en cuenta que lo que existe es lo que puede ser representado*" [104]. Gruber decía en 1993 que "*una ontología es una especi-*

*cación explícita de una conceptualización*” [101], mientras que Boster matizó que una ontología es “*la especificación de una conceptualización compartida*” [32]. Studer mezcló ambos puntos de vista definiendo a las ontologías como “*la especificación formal y explícita de una conceptualización compartida*” [236].

En la Web Semántica, las ontologías (también llamadas “*vocabularios*” por muchos autores) definen los conceptos y las relaciones que se utilizan para describir y representar un área de interés. Se utilizan para organizar los conceptos relevantes para una aplicación en particular, capturar las relaciones entre ellos y definir cualesquiera restricciones afecten a su uso. En la práctica, las ontologías pueden ser muy complejas (con varios miles o millones de términos) o muy simples (describiendo solo uno o dos conceptos). La tendencia es usar la palabra “*ontología*” para una colección de términos más compleja y posiblemente más formal, mientras que se emplea “*vocabulario*” cuando no se usa necesariamente un formalismo tan estricto o solo en un sentido muy amplio.

La función de las ontologías en la Web Semántica es ayudar a la integración de datos cuando, por ejemplo, pueden existir ambigüedades en los términos empleados en conjuntos de datos distintos, o cuando un poco de conocimiento adicional puede llevar a la inferencia (por deducción, abducción, inducción, etc.) de nuevo conocimiento, de forma supervisada por humanos o completamente autónoma. Como ejemplo, podemos mencionar el uso de ontologías en la atención médico-sanitaria. Los profesionales las utilizan para representar el conocimiento sobre los síntomas, las enfermedades y los tratamientos, mientras que las empresas del medicamento las utilizan para representar información sobre dosis, alergias, efectos secundarios, etc. La combinación de las múltiples fuentes de conocimiento así representadas permite vislumbrar para el futuro próximo un amplio abanico de aplicaciones de sistemas expertos como herramientas de apoyo a la toma de decisiones: diagnóstico, tratamiento, experimentación, investigación epidemiológica, etc.

Otro tipo de ejemplo es usar ontologías para organizar el conocimiento. Bibliotecas, museos, gobiernos, medios de comunicación, compañías privadas, aplicaciones de redes sociales y otras comunidades que administran colecciones de libros, ítems del patrimonio histórico o cultural, informes, glosarios, entradas de blog y otros artículos pueden usar ontologías como formalismos estándar, para aprovechar el poder de los datos vinculados.

Depende de la aplicación el cómo se utilicen las ontologías complejas. Algunas aplicaciones pueden decidir no usar vocabularios pequeños, y dependen de la lógica del programa de aplicación. Otras aplicaciones pueden optar por utilizar vocabularios muy simples y permitir que un entorno general de la Web Semántica utilice esa información adicional para identificar los términos. Algunas aplicaciones dependen de la existencia de acuerdos sobre terminología, sin imposición por un sistema lógico. Finalmente, otras aplicaciones pueden necesitar ontologías que habiliten sofisticados procedimientos de razonamiento sistemático, para consolidar el conocimiento existente y generar conocimiento nuevo.

Para satisfacer estas diferentes necesidades, el W3C ofrece una amplia paleta de técnicas para definir distintos tipos de vocabularios en un formato estándar, tales como el *Sistema de Organización de Conocimiento Simple* (SKOS, por sus siglas en inglés), el *Lenguaje de Ontología Web* (OWL) y el *Formato de Intercambio de Reglas* (RIF). La elección entre estas diferentes tecnologías depende de la complejidad y el rigor que requiere una aplicación específica.

A modo de ejemplo, un librero puede querer integrar datos provenientes de diferentes editores. Los datos se pueden importar a un modelo RDF común, por ejemplo, mediante el uso de convertidores a las bases de datos de los editores. Sin embargo, una base de datos puede usar el término “*autor*”, mientras que la otra puede usar el término “*creador*”. Para completar la integración, se debe agregar una definición adicional a los datos RDF, describiendo el hecho de que la relación descrita como “*autor*” es la misma que “*creador*”. Esta información adicional es, de hecho, un vocabulario (o una ontología), aunque extremadamente simple.

En un caso más complejo, la aplicación puede necesitar una ontología más detallada como parte de la información adicional. Esto puede incluir una descripción formal de cómo los autores deben ser identificados de manera única (por ejemplo, en un entorno de los Estados Unidos de Norteamérica, refiriéndose a un número único de seguridad social), cómo los términos utilizados en esta aplicación en particular se relacionan con otros conjuntos de datos en la web (por ejemplo, Wikipedia o información geográfica), cómo el término “*autor*” (o “*creador*”) se puede relacionar con términos como “*editores*”, etc.

Diversos autores han apuntado que es importante diferenciar entre *bases de conocimiento* y *ontologías*: aunque las dos son repositorios de conocimiento, las ontologías están definidas formalmente en un lenguaje codificable que puede ser portable, leído e interpretado por máquinas, mientras que una base de conocimiento no necesariamente está explicitada o codificada en un lenguaje. Estas propiedades ayudan a representar la información en forma de conceptos y axiomas, los cuales pueden ser manipulados por sistemas autónomos e incluso brindar ciertas inferencias a partir del conocimiento almacenado [33].

#### 2.2.1.4. Capa lógica

La capa lógica de la Web Semántica cumple la función de generar nuevo conocimiento a partir de los axiomas propuestos en una ontología, cubriendo así las limitaciones de lenguajes como OWL en torno a la gestión de datos. Los sistemas basados en reglas son empleados comúnmente en la lógica de negocios [9], utilizando alguno de los siguientes formalismos:

- RuleML (*Rule Markup Language*) es un lenguaje que se desarrolló originalmente en XML y posteriormente transferido a otros formatos como JSON. Está especificado a través de esquemas para documentos web, con el propósito de permitir la reutilización y compartir conocimiento [31].
- SWRL (*Semantic Web Rule Language*) es un lenguaje de reglas que combina subconjuntos de las versiones DL y Lite de OWL con formalismos de RuleML, de manera que vincula reglas tipo Horn con la expresividad de la base de conocimiento de OWL. Las reglas SWRL tienen una estructura compuesta por un cuerpo y una cabecera, en donde si las condiciones del antecedente se cumplen entonces también las del consecuente se deben cumplir [124].
- SPIN (*SPARQL Inference Notation*) es una combinación de conceptos de lenguajes orientados a objetos, reglas y lenguajes de consulta, para vincular clases que tienen consultas SPARQL capturando restricciones y reglas que formalizan el comportamiento de una clase. Se pueden representar por medio de tripletas que utilizan sintaxis SPARQL SPIN [142].

### 2.2.1.5. Capa de comprobación y confianza

De acuerdo con Antoniou [14], la capa superior de la arquitectura de la Web Semántica cumple con la función de realizar las pruebas desde las capas inferiores y validarlas, recurriendo a mecanismos de encriptación para saber si la información de una fuente es legítima. Al estar en el nivel más alto, la valoración de la calidad de la información que se proporciona está en función de la confianza de la fuente. También se debe considerar que los razonadores hacen inferencias tomando información de diferentes fuentes, lo cual implica un juicio de confianza de las fuentes consultadas. Estos juicios sobre las fuentes deben ser asumidos por los agentes que automáticamente deben decidir cuál fuente es la más adecuada [15].

### 2.2.2. Redes ontológicas y la metodología NeOn

Debido a que los recursos ontológicos se encuentran distribuidos en la red y parte del dominio de estos esquemas se superponen, se puede decir que este comportamiento forma una red de ontologías relacionadas por diferentes meta-relaciones. Las ontologías materiales se desarrollan bajo un contexto estrecho que cubre su dominio, por lo cual su desarrollo generalmente no considera un marco en el que las aplicaciones semánticas están orientadas a esta red. Con este antecedente podemos decir que una ontología no es un recurso independiente, más bien es parte de una red de recursos semánticos interconectados.

Una red ontológica entonces se define como un conjunto de ontologías interrelacionadas entre sí por metarrelaciones. Cada una de esas ontologías es definida no solo por su contenido sino que también especifica su procedencia, propósito y relación con otras ontologías y recursos semánticos [240]. La forma más común en la que una ontología forma parte de una red ontológica se da cuando tiene una definición de otra ontología para poder definir su propio modelo, lo cual crea una dependencia, esto se puede realizar mediante primitivas de OWL como `owl:imports` que permite fusionar declaraciones de dos modelos diferentes.

En vista de la existencia de gran número de ontologías, se hace cada vez más imprescindible un proceso de reutilización de acuerdo con varios factores como: ontologías que ya están creadas, ontologías modulares, instancias de ontologías, recursos no ontológicos, diseños de patrones ontológicos, etc. Con esta premisa tenemos que tratar la construcción de una ontología como una red en la que sean posibles procesos de desarrollo y administración encargados por diferentes personas e incluso de diferentes organizaciones, las cuales pueden basarse en marcos metodológicos que permita y fomente el desarrollo colaborativo de la red ontológica.

La Metodología NeOn es utilizada para el desarrollo de ontologías y redes ontológicas, presentando una serie de escenarios base para favorecer la reutilización y evolución dinámica de redes ontológicas en ambientes distribuidos [240]. El desarrollo de ontologías hoy en día se centra en procesos de reutilización, donde se puede tomar en cuenta el uso de:

- Otras ontologías, como son DOLCE [26], SUMO [205] y Kowien [6].
- Módulos de ontología.
- Patrones de diseño de ontologías y sentencias ontológicas.
- Recursos no ontológicos como son tesauros, bases de datos y esquemas de clasificación.

Una metodología madura para el desarrollo de ontologías debería cubrir procesos de desarrollo y modelos de ciclo de vida, así como métodos, técnicas y herramientas que pueden ser usadas como soporte en el proceso de desarrollo. Los marcos metodológicos son ampliamente aceptados en áreas maduras como la ingeniería del conocimiento y la ingeniería de software.

La metodología NeOn presenta nueve escenarios que soportan procesos comunes que surgen a la hora de construir una ontología, pero también se utilizan a la hora de construir redes ontológicas en ambientes distribuidos. En estos procesos participan diferentes actores como expertos en el dominio y desarrolladores de ontologías.

**Escenario 1: De la especificación a la implementación.** Este escenario describe la construcción de una ontología desde cero, en donde se debe realizar las especificaciones de requerimientos de la ontología generando un documento de especificación de requisitos. Posteriormente, se lleva a cabo una búsqueda de documentos, tanto ontológicos como no ontológicos. Tomando el documento de especificación de requerimientos ontológicos y los resultados de búsqueda de recursos, se procede a realizar la planificación del ciclo de vida de la ontología y los recursos humanos necesarios, con el fin de llevar a cabo las actividades de conceptualización, formalización e implementación de la ontología.

**Escenario 2: Reutilización y reingeniería de recursos no ontológicos.** El escenario plantea la idea de que la reutilización de recursos no ontológicos es algo necesario para optimizar recursos, pero esto implica una reingeniería para transformarlos a una ontología. Para llevar a cabo este escenario se tiene primero que realizar el proceso de reutilización y selección de recursos no ontológicos; luego se evalúa el conjunto de recursos candidatos y se realiza la selección de los recursos más apropiados. En un segundo proceso se busca transformar los recursos no ontológicos más apropiados a ontologías por medio de una ingeniería inversa, generando el modelo conceptual de los recursos no ontológicos y realizando finalmente una actividad de ingeniería hacia adelante para generar la ontología.

**Escenario 3: Reutilización de recursos ontológicos.** Los desarrolladores de ontologías son conscientes del desarrollo de múltiples ontologías en Internet, por lo cual es ideal la aplicación del principio de reutilizar ontologías enteras o conceptos que se acoplen a un dominio nuevo. El escenario parte de la condición en la cual se tienen a disposición recursos ontológicos que se han mostrado de utilidad para solventar requisitos similares. El proceso de reutilización de recursos ontológicos parte de la búsqueda de ontologías que serán evaluadas y comparadas para seleccionar las ontologías más convenientes para el dominio en cuestión, y por último se tienen que integrar las ontologías o partes de ellas a una nueva red ontológica.

**Escenario 4: Reutilización y reingeniería de recursos ontológicos.** El escenario se da cuando se cuenta con recursos ontológicos que son útiles para la construcción de nuestra red semántica pero con la particularidad de que estos recursos no nos son útiles en la forma en la que están diseñados y modificarse para que calcen en nuestro propósito. El escenario consta de los procesos de reutilización de recursos ontológicos más adecuados para nuestros propósitos y posteriormente se lleva a

cabo el proceso de reingeniería de recursos ontológicos y finalmente se utiliza los recursos ontológicos resultantes como entradas del Escenario 1.

**Escenario 5: Reutilización y fusión de recursos ontológicos.** El escenario es apropiado cuando el desarrollador tiene a su disposición dos o más recursos ontológicos que cubren el mismo dominio, por lo cual se busca formar un nuevo recurso ontológico que combine las dos en una superposición, o en otro caso se busca realizar la alineación de los elementos ontológicos para la nueva red. Para cumplir con este propósito, primero se realiza el proceso de selección de recursos ontológicos llevando a cabo todos los procesos del Escenario 3, y tomando en cuenta que todos los recursos ontológicos seleccionados son válidos y se realizan las actividades de alineación ontológica y fusión de ontologías. Finalmente, se utilizan los recursos ontológicos resultantes como entradas del Escenario 1.

**Escenario 6: Reutilización, fusión y reingeniería de recursos ontológicos.** En este escenario el desarrollador de la ontología cuenta con varios recursos ontológicos que están bajo un mismo dominio, por lo cual se busca fusionarlos para formar un nuevo recurso ontológico, pero el recurso tiene que ser modificado para que cumpla con nuevos propósitos. Primero se tiene que realizar el proceso de selección de recursos ontológicos como en el Escenario 3; posteriormente se realizan las actividades de alineación y fusión de ontologías como en el Escenario 5, para realizar con los recursos ontológicos resultantes una reingeniería como se explicó para el Escenario 4. Por último, se utilizan los recursos ontológicos resultantes del Escenario 4 como entradas del Escenario 1.

**Escenario 7: Reutilización de patrones de diseño ontológico.** Los *patrones de diseño de ontologías* (PDO) han surgido para ayudar a los desarrolladores de ontologías como una guía y un modelo de codificación de buenas prácticas, proporcionando tres tipos de beneficios: reutilización, guía y comunicación. Los desarrolladores de ontologías pueden aplicar PDOs en las actividades de conceptualización e implementación de la ontología para buscar soluciones modernas en repositorios.

**Escenario 8: Reestructuración de los recursos ontológicos.** Este escenario se da cuando la conceptualización de la red ontológica tiene que corregirse y reorganizarse para cumplir con nuevos requisitos, por lo cual se tiene que llevar a cabo actividades de reestructuración, incluyendo labores de modularización, poda, enriquecimiento, extensión y especialización. Tales actividades se pueden llevar a cabo de manera combinada y en cualquier orden.

**Escenario 9: Localización de los recursos ontológicos.** Este escenario se presenta cuando la ontología final tiene que estar en diferentes idiomas, por lo cual después de formar la red ontológica en un determinado idioma se debe proceder a realizar las siguientes actividades de localización: selección de los recursos lingüísticos más adecuados, selección y obtención de etiquetas localizadoras, evaluación de la traducción de etiquetas y actualización de la ontología.

## 2.3. Redes Neuronales Artificiales

Las redes neuronales artificiales (RNA o ANN por las siglas en inglés) —también denominadas por algunos autores *sistemas conexionistas*— son un paradigma de procesamiento y aprendizaje automático inspirado en un conglomerado de unidades neuronales simples que intentan modelar el funcionamiento de los cerebros de los animales [256]. Se trata de sistemas que permiten emular el proceso de sinapsis que se produce en un cerebro biológico, donde un estímulo (información de entrada) atraviesa una maraña de neuronas para producir un valor de salida. Estos estímulos tienen establecidos un peso numérico, en el que se codifica el conocimiento que la RNA posee sobre un determinado problema. Las RNA no son programadas de manera explícita y tienen la cualidad de aprender por sí mismas. Sobresalen en áreas donde identificar soluciones o características no es fácil de expresar mediante la programación convencional [270].

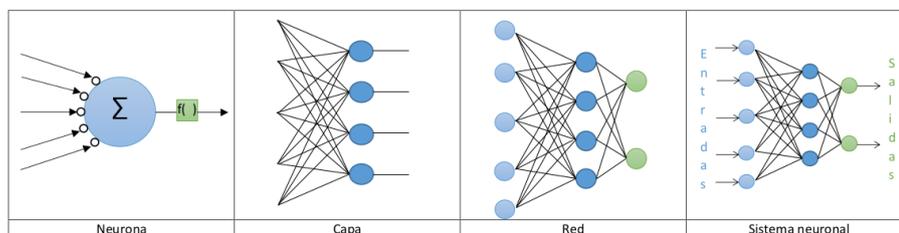
### 2.3.1. Estructura de una Red Neuronal Artificial

Las RNA tienen la capacidad de emular tres cualidades del sistema nervioso biológico:

- *Procesamiento en paralelo*: esencial para brindar al sistema la capacidad de poder realizar una gran cantidad de operaciones de cálculo en un tiempo reducido [54].
- *Distribución de memoria*: permite que en el sistema la información ocupe diferentes posiciones de memoria bien definidas, de modo que si una sinapsis resulta dañada, se pierde sólo una porción de la información [171].
- *Adaptabilidad*: el sistema se ajusta fácilmente a su entorno modificando sus sinapsis, aprendiendo de la experiencia y generalizando a partir de casos particulares [38].

Como se puede apreciar en la Figura 2.2, el elemento principal de una RNA es la neurona artificial. Las neuronas se organizan en capas, que a su vez constituyen una red neuronal. Por último, una red neuronal (o un conjunto de ellas), junto con las interfaces de entrada y salida, constituyen un sistema completo.

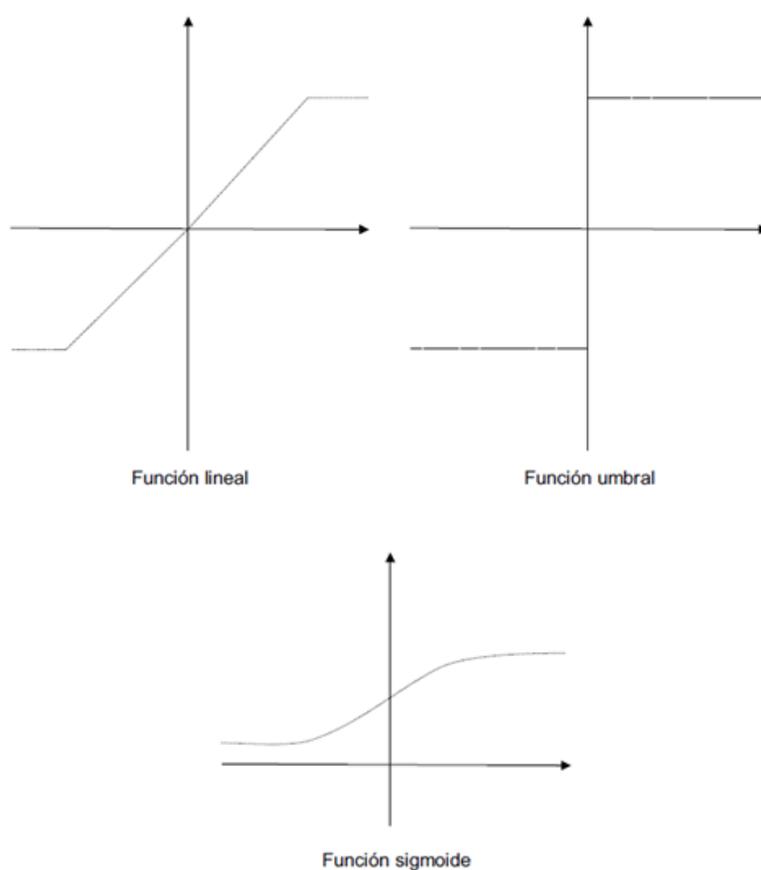
Figura 2.2: Sistema general de una red neuronal artificial



Algo común en cualquier sistema de RNA es que la salida generada por una neurona es producto de dos situaciones:

- La *regla de propagación*, que normalmente corresponde a la suma ponderada de las entradas multiplicadas por los pesos, para proporcionar el valor del potencial postsináptico integrando la información proveniente de las distintas neuronas de la capa precedente.
- La *función de activación*, que decide, en función de la estimulación total a la entrada, qué valor se presentará a la salida. Existen múltiples opciones, como funciones de tipo escalón para tener salidas discretas, funciones lineales para tener salidas proporcionales a las entradas en un determinado rango, y funciones sigmoideas o gaussianas para introducir efectos de no linealidad [141,179] (véase Figura 2.3).

Figura 2.3: Funciones de activación dentro de las RNA.



Comúnmente, dentro del estado del arte las RNA han sido clasificadas en base a su topología, haciendo alusión al número de capas disponibles en la red, la cantidad de neuronas presentes por capa, y el tipo de conectividad entre ellas. Así, en función de su topología, se habla de los siguientes tipos:

- *Red neuronal monocapa* es aquella que sólo tiene una capa de neuronas, que funcionan al mismo tiempo como entrada y salida del sistema. El perceptrón

simple [217] y la red Hopfield [121] son unas de las redes más populares en esta categoría [179].

- La *red neuronal multicapa* es una generalización de la anterior, en tanto que dispone de conjuntos de neuronas jerarquizadas en distintos niveles, con al menos uno de entrada y otro de salida. Este tipo de redes pueden estar total o parcialmente conectadas, y se pueden a su vez clasificar atendiendo a la manera en que se conectan sus capas:
  - Las *redes con conexión hacia adelante (feedforward)* contienen solo conexiones entre capas hacia adelante. Como casos de este tipo de redes podemos mencionar el perceptrón multicapa (MLP) [107], la red Madaline [222], las máquinas de Boltzmann [112] y Cauchy [241], la memoria lineal adaptativa [123], los modelos de Kohonen [165] y las redes backpropagation [97].
  - Las *redes con conexión hacia atrás (feedback)* se diferencian de las anteriores por la existencia de retroalimentación, sea entre neuronas de diferentes capas, de la misma capa o sobre sí mismas. Esta estructura las hace especialmente adecuadas para estudiar sistemas no lineales. Las más habituales son las redes bicapa, tales como las redes neuronales ART [18], las memorias asociativas bidireccionales (BAM) [145], el Neocognitron [261] y el Cognitron [93].

En tanto que las redes *feedforward* presentan siempre soluciones estables, los modelos con *feedback* entregan soluciones variables o dinámicas, con riesgo de no convergencia en el aprendizaje [200].

El aprendizaje consiste en presentar patrones a la red neuronal, para cambiar los pesos de las conexiones sinápticas usando una determinada regla. En base a este criterio se estipula la siguiente clasificación:

- El *aprendizaje supervisado* presenta un esquema de entrenamiento controlado por un agente *supervisor*. En el mundo real representa al problema a modelar que se visibiliza mediante un conjunto de datos de entrada y salida. El supervisor guía el aprendizaje en la red y conoce las salidas deseadas correspondientes a cada entrada. Si la salida no es la esperada, se ajustan los pesos de cada neurona para, de forma iterativa, ir aproximando las respuestas adecuadas. Podemos nombrar como redes que se basan en este tipo de entrenamiento *backpropagation* [97], el perceptrón multicapa [107], la red Madaline [222], las redes de Boltzmann [112], el perceptrón simple [217], la memoria asociativa bidireccional [145], la red de Cauchy [241] y las redes de Adaline [263].
- En *aprendizaje no supervisado* no requiere que se le presenten los patrones objetivo de salida, debido a que la lógica de computación y la regla de modificación de las conexiones dan lugar a salida consistentes. Cuando la red procesa patrones muy similares, produce la misma salida para todos ellos, sirviendo así propósitos de clasificación. En este tipo de aprendizaje se cimentan las redes de memorias asociativas bidireccionales (BAM) [145], el neocognitrón [261], el cognitrón [93], los modelos de Kohonen [165] y las redes Hopfield [200], entre otros modelos. Este tipo de aprendizaje conoce dos formas básicas:

- *Aprendizaje por componentes principales*: Consiste en identificar las características principales que son comunes a muchos patrones de entrada. Con ese fin, un cierto número de neuronas participan en la representación del patrón de entrada.
  - *Aprendizaje competitivo*: Hace referencia a que las neuronas compiten por representar determinados patrones de entrada.
  - *Aprendizaje reforzado*: Se basa en la noción de condicionamiento por refuerzo, potenciando las conductas deseadas y penalizando las no deseadas, lo cual se consigue a base de aumentar los pesos sinápticos si se acierta a la salida y reducirlos en caso contrario.
- Finalmente, se habla de *redes híbridas* como aquellas que representan una combinación de las dos anteriores, aplicando una función de mejora para facilitar el proceso de convergencia. Para este caso cabe mencionar las redes de base radial [179].

A continuación se presentan detalles sobre el funcionamiento de algunas de las RNA antes mencionadas, que han sido particularmente relevantes para el desarrollo de la presente tesis.

### 2.3.2. Perceptrón simple

El perceptrón simple es un tipo de red neuronal monocapa híbrida de aprendizaje supervisado. Es una red unidireccional donde la información fluye desde las  $x_n$  entradas hacia la única  $f(x)$  salida, como se puede observar en la Figura 2.4.

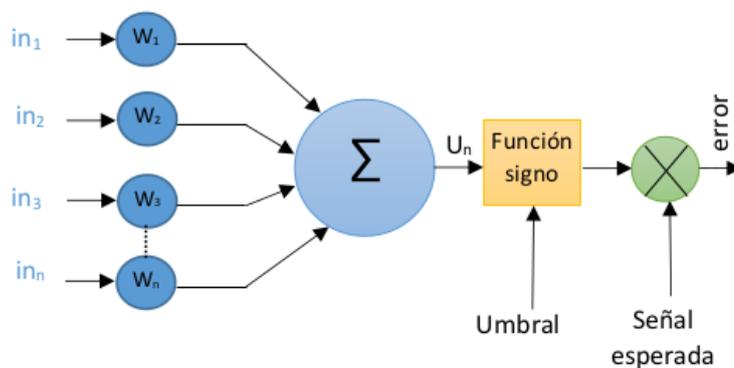


Figura 2.4: Diagrama del perceptrón simple

Este tipo de red obtiene un valor de salida que se obtiene al sumar las señales de entrada (un vector  $X_i$ ) y multiplicar por los pesos correspondientes (vector de pesos  $W_i$ , que pueden iniciar con un valor de 0 o generados al azar), menos el valor de umbral, como se aprecia en la Ecuación 2.3.1:

$$x = \sum_{i=1}^n w_i x_i - \Theta \quad (2.3.1)$$

En base al valor de salida la función de activación asumirá un valor de 1 en el caso que  $f(x)$  sea mayor a cero; caso contrario, toma el valor 0 (Ecuación 2.3.2):

$$f(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{caso contrario} \end{cases} \quad (2.3.2)$$

El perceptrón simple, al ser un tipo de red supervisado, necesita conocer los valores esperados para cada una de las entradas presentadas a la red, por lo que se manejan pares de entrada-salida:

$$\{in_1, out_1\}, \{in_2, out_2\}, \dots, \{in_n, out_n\} \quad (2.3.3)$$

Cuando las entradas ( $in$ ) son expuestas a la red, las salidas ( $out$ ) son comparadas con el valor esperado ( $Vv$ ), para lo cual el valor de salida o valor calculado ( $Vc$ ) está dada por la Ecuación 2.3.4:

$$Vc = \sum_{i=1}^n w_i x_i \quad (2.3.4)$$

Dentro del proceso de entrenamiento, el factor clave es poder ajustar los valores de los pesos  $W_i$  que permiten relacionar eficientemente las entradas con la salida. Para ello, la red se encarga de calcular iterativamente el nuevo valor peso de cada  $w_i$  en la etapa de aprendizaje. Este procedimiento de convergencia, también denominado como *regla del perceptrón*, es la primera regla para actualizar los pesos según la Ecuación 2.3.5:

$$w_{i+1} = w_i + (\eta * \delta * x_i) \quad (2.3.5)$$

- $w_i$ : valor del peso anterior.
- $\eta$ : coeficiente de aprendizaje.
- $\delta$ : error cometido (diferencia entre el valor calculado y el valor deseado).
- $x_i$ : valor de la señal de entrada.

El perceptrón simple es muy utilizado en diferentes problemas de clasificación, donde se tiende a tener resultados perfectos si los patrones de análisis son linealmente separables [29,175,192], como el caso que representa la Figura 2.5.

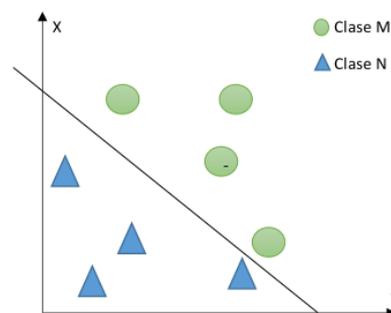


Figura 2.5: Ejemplo de clases linealmente divisibles.

Como limitación, cabe destacar que el algoritmo del perceptrón no puede converger hacia un error nulo si los patrones que se analizan no son linealmente separables.

### 2.3.3. Redes de Hopfield

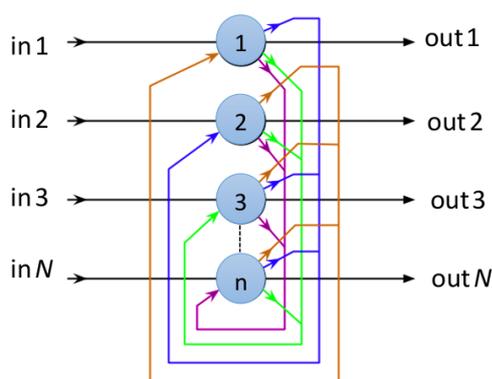
Las redes de Hopfield, también denominadas *redes recurrentes o recursivas*, fueron propuestas en el año 1982 por el físico John Hopfield [121]. Se fundamentan en los siguientes aspectos:

- Comportamiento dinámico (cambiante en función del tiempo).
- Capacidad de memorizar relaciones.
- Posibilidad de almacenar información.
- Fácil implementación en hardware analógico.

La red de Hopfield emplea un sistema de operación enfocado en el almacenamiento de la información imitando el comportamiento de una memoria asociativa —esto es, una memoria que permite recuperar información a partir de conocimiento parcial de su contenido, sin conocer su ubicación de almacenamiento. Como se describe en [64, 121], la memoria asociativa permite que la red inicie a partir de un estado inicial, denominado *información de partida*, y posteriormente evolucione hasta alcanzar un estado estable, identificado por el patrón más semejante a la información inicial. El propósito es recuperar un patrón que se almacenó previamente en su estructura, a partir de una muestra parcial (incompleta) o ruidosa (distorsionada) del patrón original [84, 121]. En base a los patrones inicialmente almacenados, la red identifica aquel que sea más semejante al patrón de entrada que se le presente.

La arquitectura inicial de Hopfield consta de  $n$  neuronas binarias, donde cada una se conecta con todas las demás a través de conexiones laterales, pero no consigo misma como se aprecia en la Figura 2.6.

Figura 2.6: Arquitectura de una red Hopfield con  $N$  nodos.



La matriz de pesos  $W = (W_{ij})$  de tamaño  $n \cdot n$ , en la cual  $W_{ij}$  simbolizan el peso de la conexión entre las neuronas  $i$  y  $j$ . La matriz tiene algunas particularidades tales como las siguientes:

- La matriz debe ser simétrica en cuanto al peso de conexión entre dos neuronas; eso quiere decir que debe tener el mismo valor en ambas direcciones  $W_{ij} = W_{ji} \quad \forall i,j$ .
- Los componentes de la diagonal de la matriz deben ser iguales a 0, como corresponde a que no existe conectividad de ninguna neurona consigo misma:  $W_{ii} = 0 \quad \forall i$ .

Los estados posibles generalmente son 1 o -1, o bien 1 o 0. La activación de una neurona en la red de Hopfield se realiza mediante la siguiente regla:

$$s_i \leftarrow \begin{cases} +1 & \text{si } \sum_j w_{ij}s_j \geq \theta_i, \\ 0 \text{ o } -1 & \text{caso contrario} \end{cases}$$

donde:

- $w_{ij}$  es la intensidad del peso de conexión entre la neurona  $j$  a la neurona  $i$  (el peso de la conexión).
- $s_j$  representa el estado de la neurona  $j$ .
- $\theta$  es el umbral de la neurona  $i$ .

Por otra parte, la actualización de la red de Hopfield puede realizarse de dos maneras diferentes [121,122]:

- *Asíncrona*: solo se actualiza una neurona a la vez. Esta neurona se puede elegir al azar o se puede imponer un orden predefinido desde el principio.
- *Síncrona*: todas las neuronas se actualizan al mismo tiempo. Esto requiere un reloj central para el sistema para mantener la sincronización.

Las redes de Hopfield poseen un valor escalar relacionado con cada estado de la red, llamado “energía”, representado en la Ecuación 2.3.6.

$$E = -\frac{1}{2} \sum_{i,j} w_{ij}s_i s_j + \sum_i \theta_i s_i \quad (2.3.6)$$

Bajo este concepto de *energía* se garantiza que las unidades sean elegidas aleatoriamente para actualizar sus valores de activación, de tal forma que la red convergerá a estados que sean mínimos locales de la función de energía [65]: si un estado es un mínimo local en la función, entonces se corresponde con un estado estable de la red.

El entrenamiento de una red de Hopfield consiste en bajar la energía de los estados que la red debería “recordar”. Esto proporciona a la red la capacidad de utilizarse como un sistema de memoria de contenido direccionable; es decir, la red convergerá a un “estado recordado” si se le entrega solo parte del estado. Este tipo de red se puede usar para recuperar de una entrada distorsionada o incompleta al estado entrenado que es más similar a esa entrada. Por ello se denomina memoria asociativa: porque recupera recuerdos sobre la base de la similitud. Por ejemplo, si se entrena una red de Hopfield con cuatro unidades que sea un mínimo de energía representado por el estado (1, -1, 1, -1) y se le asigna a la red el estado (1, -1, -1, -1),

está convergerá al estado (1, -1, 1, -1). Por consiguiente, la red está oportunamente entrenada cuando la energía de los estados que debe recordar son mínimos locales.

La inicialización de las redes de Hopfield se ejecuta estableciendo los valores de las neuronas al patrón de inicio deseado. El procedimiento de actualización se realiza hasta que la red converja a un patrón determinado. En este tipo de redes la convergencia está generalmente garantizada, pues Hopfield demostró que los atractores de este sistema dinámico no lineal son estables, no periódicos o caóticos como en otros sistemas. Por consiguiente, dentro del contexto de las redes de Hopfield, un *patrón atractivo* será un estado estable final, un patrón que no puede modificar ningún valor dentro de sí mismo durante el proceso de actualización.

El procedimiento consta de los siguientes pasos:

- (i) Se inicializan las conexiones de los patrones de entrada.

$$w_{i,j} = \sum_{p=0}^n x_i^p x_j^p \quad i \neq j \quad w_{i,j} = 0 \quad i = j;$$

- $i, j$  simbolizan diferentes patrones de entrada.
- $p$  simboliza una categoría de patrones donde  $X$  es el valor representante.
- $m$  simboliza el número de patrones.

- (ii) Se procede a actualizar las neuronas de la capa de procesamiento.
- (iii) Si se ha modificado el estado de la red hasta alcanzar su convergencia, o en su caso se ha alcanzado el número máximo de iteraciones, se finaliza.
- (iv) Si no se ha alcanzado la convergencia de la red o finalizado el número de iteraciones, se retorna al paso 2.

Como ventajas de las redes de Hopfield cabe comentar las siguientes:

- No existe un tiempo de entrenamiento, debido a que no disponen de un proceso adaptativo; simplemente se calcula la matriz de pesos asociados a la red.
- Al funcionar como memorias asociativas, tienen la cualidad de ser bastantes tolerantes al ruido [111].
- Son particularmente estables cuando sus patrones no se asemejan demasiado [111].

Como desventajas, destaca el hecho de que el número de patrones a almacenar o aprender es limitado, en comparación con el número de nodos de la red. Según indicaba el propio Hopfield [111, 122], esto se debe a que el número de clases debe ser menor a 0.15 veces al número de nodos en la red. Asimismo, la red se torna muy inestable cuando los patrones son muy semejantes entre sí [111].

### 2.3.4. Redes de Kohonen

Las redes de Kohonen —también denominadas como *mapas autoorganizados* (*Self-Organizing Map*, SOM) fueron inventadas por Teuvo Kohonen a inicios de la década de 1980. Este tipo de RNA crea un modelo de los datos de entrada para

agruparlos empleando criterios de similitud, mediante un procedimiento iterativo con un conjunto de datos y modificaciones para asemejarse a los mismos [143].

Las redes de Kohonen poseen un arquitectura que consta de dos capas:

- *Capa de entrada o sensorial:* consta de  $n$  neuronas, una en función de cada variable de entrada (comúnmente un conjunto o vector de características).
- *Capa de salida o de competición:* representada por  $m$  neuronas, es la responsable del procesamiento y de formar el mapa de rasgos de dos dimensiones. Estas neuronas, aunque no están conectadas entre sí, son capaces de generar con cierto grado de influencia conexiones laterales de excitación e inhibición sobre las neuronas más próximas o vecinas.

Las conexiones entre las neuronas de la red son hacia adelante, y cada patrón de entrada está interconectado a través de pesos sinápticos con todas las neuronas de la capa de competencia, como se puede apreciar en la Figura 2.7.

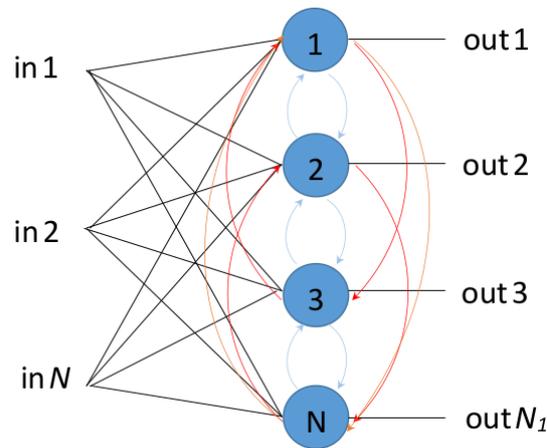


Figura 2.7: Estructura básica de una red neuronal competitiva de Kohonen.

A cada neurona de la capa de salida se le otorga un vector de pesos de dimensión del vector de entrada, como se puede apreciar en la Ecuación 2.3.7.

$$W_{ij} = [W_{ij}^1, W_{ij}^2, W_{ij}^3, \dots, W_{ij}^n] \quad (2.3.7)$$

La red identifica qué patrones de entrada semejantes activan las neuronas siguientes en la capa de competencia. Esto se ejecuta en el instante en que los pesos en la capa de competencia relacionados a esa entrada son similares a ese patrón de entrada. La similitud se obtiene a través de aplicar medidas de semejanza, tales como las que se presentan en la Tabla 2.2.

La neurona que posea mayor nivel de similitud en relación al patrón de entrada será la ganadora en la capa de competencia. Posteriormente, se lleva a cabo la actualización de los pesos de las neuronas de la capa de competencia, permitiendo que la neurona vencedora se asemeje más a la entrada en las siguientes iteración, hasta el número de iteraciones establecido [228]. La ecuación para la actualización

Distancia	Ecuación
Euclídea	$d(w, x) = \sqrt{\sum_{k=1}^l (w_{ij}^k - x^k)^2} \quad (2.3.8)$
Minkowski	$d(w, x) = \left[ \sum_{k=1}^l (w_{ij}^k - x^k)^\lambda \right]^{\frac{1}{\lambda}} \quad (2.3.9)$
Manhattan	$d(w, x) = \sum_{k=1}^l  (w_{ij}^k - x^k)  \quad (2.3.10)$

Tabla 2.2: Posibles métricas para calcular distancias en redes de Kohonen.

de los pesos sinápticos es la siguiente:

$$W_{lk}^n = W_{lk}^n + \alpha * (x^n - W_{lk}^n) \quad (2.3.11)$$

donde:

- $lk$  representa la neurona vencedora,
- $n = 1, 2, 3, \dots, l$  representa el índice de cada parámetro de un patrón de entrada.

Se obtienen como resultados células ganadoras semejantes cuando se tienen patrones de entrada similares, pese a que no se garantiza que estén relativamente próximas en la capa de competencia.

Para determinar el conjunto de neuronas del vecindario que necesitan modificar sus pesos sinápticos y con qué intensidad, es necesario aplicar la interacción lateral mediante la función de vecindad. Con este procedimiento, la excitación de neuronas se produce también en el entorno de la neurona ganadora.

Las redes de Kohonen tienen dos modos de operación.

- En la etapa de *entrenamiento* por cada neurona se recibe un vector de entrada, para construir el mapa interno de pesos.
- En la etapa de *ejecución* o *mapeo* se trata de clasificar cada nueva entrada.

El procedimiento consta de los siguientes pasos:

- (i) Se inicializan los valores  $W_{ij}$  con pesos pequeños al azar entre 0 y 1, para la entrada  $j$  y la salida  $i$ .
- (ii) Se presenta un vector de entrada  $E_l : (e_1, e_2, e_3, \dots, e_n)$ , cada cierto período de tiempo hasta que converja la red.

- (iii) Se determina la neurona ganadora empleando una de las funciones de distancia, identificando aquella con mayor similitud del vector de pesos y la entrada.
- (iv) Se procede a actualizar las conexiones entre la capa de entrada y la neurona ganadora, actualizando su vecindario a la par.
- (v) Si los datos que se expusieron en la entrada son mayores al umbral, se retorna al paso 2; en caso contrario, termina el procedimiento.

Como ventajas de las redes de Kohonen, cabe citar las siguientes:

- Al presentar una estructura de configuración simple, así como su particular dinámica de entrenamiento, son una herramienta sofisticada para aplicaciones en problemas relacionados con la clasificación de patrones y la agrupación de datos [64].
- Disminuyen el tiempo de entrenamiento y los recursos de cómputo, en comparación con procedimientos de *clustering* tradicionales.
- Presentan baja sensibilidad al ruido en los datos.

Entre los puntos débiles, por el contrario, cabe mencionar:

- Requieren ser entrenadas con un gran número de ejemplos.
- Es complejo poder determinar qué pesos de entrada utilizar.
- Requieren que los puntos cercanos se comporten de manera similar.
- El mapeo pueda dar como resultado grupos divididos.

### 2.3.5. Perceptrón multicapa

Un perceptrón multicapa (*Multilayer Perceptron*, *MLP* por sus siglas en inglés) es un tipo de red neuronal artificial alimentada hacia adelante. Esta red emplea un mecanismo de aprendizaje supervisado denominado “*backpropagation*” o “*regla delta generalizada*” para el entrenamiento [220].

Una MLP consta de, al menos, tres capas de nodos:

- *Capa de entrada*: Conformada por aquellas neuronas que ingresan los patrones de entrada en la red. Estas neuronas no realizan ningún procesamiento.
- *Capas ocultas*: Constituidas por aquellas neuronas que tienen entradas provenientes de capas anteriores, y cuyas salidas pasan a neuronas de capas posteriores.
- *Capa de salida*: Representada por aquellas neuronas cuyos valores de salida conciernen a las salidas de la red.

En la Figura 2.8 se muestra la estructura de una red neuronal MLP con dos capas ocultas. El conjunto de neuronas extiende la señal hasta la salida. Para llevar a cabo tal propagación, cada neurona debe realizar una combinación de las señales que ingresan de las neuronas de la capa anterior empleando los respectivos pesos

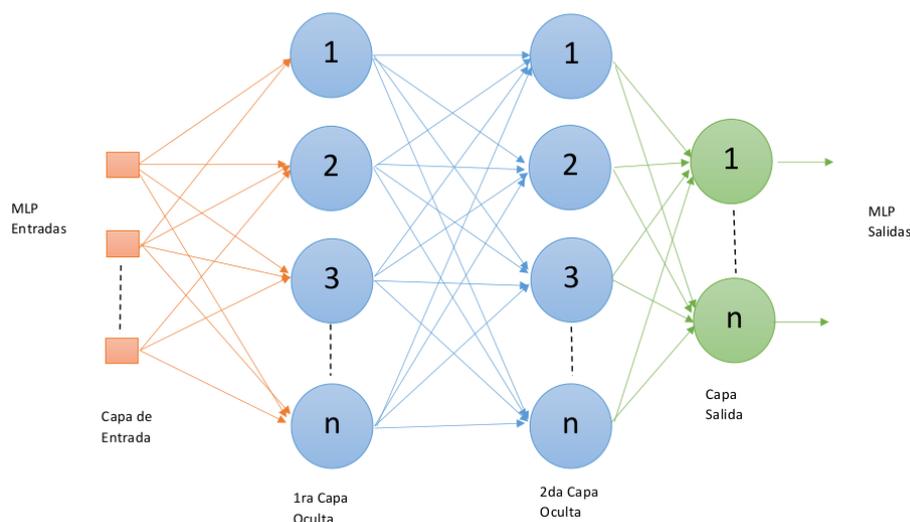


Figura 2.8: Estructura de un perceptrón multicapa con dos capas ocultas.

sinápticos. El algoritmo de aprendizaje es responsable de ajustar las conexiones (i.e. los pesos) durante el proceso de entrenamiento.

Excepto por los nodos de entrada, cada nodo es una neurona que utiliza una función de activación no lineal. Sus múltiples capas y su activación no lineal distinguen al MLP de un perceptrón lineal, y son la característica clave para que la RNA pueda distinguir datos que no son linealmente separables [62].

Es necesario estudiar el número de neuronas que conforman las capas de la red, tanto para la capa de entrada, el número de capas ocultas con sus respectivos números de neuronas, y la capa de salida, en relación al problema que se pretende resolver.

Las dos funciones de activación más comunes son ambas del tipo sigmoide, y están descritas por la Ecuación 2.3.12 y la Ecuación 2.3.13:

$$y(w_i) = \tanh(w_i) \quad (2.3.12)$$

$$y(w_i) = (1 + e^{-w_i})^{-1} \quad (2.3.13)$$

La primera ecuación corresponde a una tangente hiperbólica, que varía entre 1 y -1, en tanto que la segunda es una función logística con comportamiento similar, pero cuyo rango está entre 0 a 1. En ambas ecuaciones  $y(i)$  es el resultado computado por la  $i$ -ésima neurona y  $w_i$  se corresponde con la suma ponderada de las conexiones de entrada. Entre otras funciones de activación alternativas se pueden mencionar las funciones *rectificador* y *softplus*.

El aprendizaje acontece en el perceptrón al modificar los pesos de conexión después de que se procesa cada parte de los datos, en comparación de la cantidad de error en la salida en relación con el resultado esperado. Este entrenamiento se lleva a cabo como una generalización del algoritmo de mínimos cuadrados en el perceptrón lineal.

La identificación del error en el nodo de salida  $j$  en el punto  $n$ -ésimo de datos (proceso de entrenamiento) se puede realizar mediante  $e_j(n) = t_j(n) - c_j(n)$ , donde  $t$  es el valor verdadero y  $c$  es el valor calculado por el perceptrón. Los pesos de los nodos se ajustan en función de las correcciones que minimizan el error en toda la salida, dado por la Ecuación 2.3.14:

$$\varepsilon(n) = \frac{1}{2} \sum_j e_j^2(n) \quad (2.3.14)$$

El algoritmo permite retrotraer el resultado desde la salida hacia la entrada mediante un proceso enfocado en minimizar la función de coste, para así optimizar los valores de los pesos. Aplicando el método de descenso por gradiente, el cambio en cada peso estará dado por la Ecuación 2.3.15:

$$\Delta v_{ji}(n) = -\eta \frac{\partial \varepsilon(n)}{\partial w_j(n)} c_i(n) \quad (2.3.15)$$

donde:

- $c_i$  es la salida de la neurona anterior.
- $\eta$  es la tasa de velocidad de aprendizaje, que se escoge para garantizar que los pesos converjan rápidamente en una respuesta, sin oscilaciones.

La derivada que se calcula depende del campo local inducido  $w_j$  y se puede expresar como sigue:

$$-\frac{\partial \varepsilon(n)}{\partial w_j(n)} = e_j(n) \phi w_j(n) \quad (2.3.16)$$

donde  $\phi$  es el derivado de la función de activación descrita anteriormente. El análisis es más difícil para el cambio en las ponderaciones de un nodo oculto, pero se puede mostrar que el derivado relevante es el indicado en la Ecuación 2.3.17:

$$-\frac{\partial \varepsilon(n)}{\partial w_j(n)} = \phi w_j(n) \sum_k -\frac{\partial \varepsilon(n)}{\partial w_k(n)} v_{kj}(n) \quad (2.3.17)$$

Esto depende del cambio en los pesos del nodo  $k$ -ésimo, que representa la capa de salida. Entonces, para cambiar los pesos de capa ocultos, los pesos de capa de salida cambian de acuerdo con la derivada de la función de activación, por lo que este algoritmo representa una propagación hacia atrás de la función de activación [108].

El algoritmo de aprendizaje de una red perceptrón multicapa se articula en los pasos siguientes:

- (i) Se inicializan los pesos con valores aleatorios próximos a cero.
- (ii) Se procede a presentar un patrón  $n$  de entrenamiento, que se propaga hasta la salida y se adquiere el resultado calculado por la red  $c_n$ .
- (iii) Se analiza el error que se comete en la red a nivel de cada neurona, en cada capa oculta.
- (iv) Se procede a aplicar la regla de propagación hacia atrás para actualizar los pesos, iniciándose desde la capa de salida hasta alcanzar la capa de entrada.

(v) El proceso itera hasta alcanzar un mínimo de error establecido.

El perceptrón multicapa es una de las RNA más utilizadas en la actualidad, debido entre otras a las siguientes ventajas [64, 250]:

- Tienen una capacidad de clasificación mayor que otras RNAs, al identificar hasta  $2^m$  clases siendo  $m$  el número de neuronas en la capa de salida.
- Presentan un alto grado de tolerancia a fallos.
- Si se entrenan adecuadamente, pueden alcanzar un elevado nivel de asertividad en la relación entre dos conjuntos de datos.

Como desventajas, cabe mencionar las siguientes:

- El MLP no puede extrapolar tan bien como otras RNA: si durante el proceso de aprendizaje no fue entrenada adecuadamente o con suficientes datos, los resultados serán imprecisos.
- El proceso de entrenamiento puede tener un alto nivel de complejidad, sobre todo cuando las clases representadas por puntos están próximos entre sí [107].
- Una red MLP suele requerir un gran esfuerzo computacional durante el proceso de entrenamiento, derivado de multiplicar el número de pesos sinápticos de la red por el número de capas de entrenamiento y por el número de patrones de entrenamiento utilizados en cada capa [64], amén del cálculo de cada función de activación.

## 2.4. Algoritmos genéticos

Los algoritmos genéticos (GA, por sus siglas en inglés) son un procedimiento evolutivo centrado en encontrar una solución en espacios complejos, inspirados en los procesos de selección natural y evolución genética [116, 119].

El mecanismo de funcionamiento parte siempre de una población inicial de individuos a los cuales se denomina *cromosomas*, que representan soluciones posibles a un problema en concreto. Cada cromosoma de la población es evaluado mediante una función de aptitud, y solo aquellos cromosomas que superan la evaluación son considerados para reproducirse en la siguiente generación. A estos cromosomas seleccionados se les aplican diferentes operadores genéticos (selección, cruce y mutación) [212] que, a través del tiempo, permiten generar nuevas poblaciones de individuos mejorados (nuevos cromosomas). Estos nuevos cromosomas se aproximan iterativamente a una solución cada vez mejor del problema en cuestión.

Los GA se han utilizado con éxito para generar soluciones de alta calidad en problemas de búsqueda y optimización [153, 207, 246]. Su principal ventaja se debe a su capacidad de exploración y explotación de información acumulada sobre un espacio de búsqueda (población de individuos o cromosomas mejor dotados), y mediante esta metodología dirigir las próximas búsquedas hacia los mejores subespacios. Este método resulta más idóneo donde se trabaja con un gran conjunto de posibles soluciones donde los métodos clásicos de búsqueda no son apropiados.

Este tipo de algoritmos emplean un vocabulario particular haciendo uso de términos propios de la genética:

- Un *cromosoma* es una cadena, estructura, con una secuencia de elementos que representa (codifica) una solución o un individuo.
- El *genotipo* permite denotar el contenido genético de un individuo (codificación específica de su cromosoma).
- El *fenotipo* es un término empleado para identificar las soluciones representadas por los cromosomas.
- Un *gen* es cada uno de los elementos que conforman la cadena, estructura o codificación que emplea un cromosoma.
- Un *alelo* representa cada uno de los valores o estados que puede tomar un gen dentro de un cromosoma.

### 2.4.1. Estructura básica de un algoritmo genético

Como se señaló anteriormente, un algoritmo genético comienza generando aleatoriamente un determinado número de cromosomas (grupos de individuos pertenecientes a una población), que se codifican como un conjunto de soluciones posibles al problema de estudio. A este conjunto de cromosomas se les denomina *población inicial* y su tamaño representa un parámetro esencial en el proceso.

El algoritmo, a continuación, empieza a valorar la calidad de cada uno de los cromosomas de la población en base a una función de aptitud. Una vez identificado el nivel de aptitud de cada cromosoma, se les somete a todos a un proceso natural de selección, mediante el cual se determina cuáles son los cromosomas más idóneos para participar en el proceso de reproducción de la nueva *generación*.

Existen diversos métodos para la selección de los cromosomas, pero usualmente se realiza otorgando mayor probabilidad a escoger aquellos cromosomas que fueron mejor evaluados. Los cromosomas escogidos son sometidos a operadores genéticos (cruce y mutación), responsables de generar una nueva descendencia o nuevos cromosomas que contengan las mejores características de sus progenitores.

Del conjunto de cromosomas escogidos y de la descendencia generada, se escogerá un número suficiente de cromosomas que permitan obtener una nueva población (un conjunto de nuevas soluciones mejoradas). Este proceso de sustitución de los cromosomas de la generación anterior por nuevos cromosomas se denomina *reemplazo*.

Todo el ciclo de evaluación de aptitud, selección, cruce, mutación y reemplazo se repite iterativamente sobre la población hasta que se cumple un criterio de parada preestablecido, ya sea por alcanzar la solución óptima (o una aproximación suficiente buena) o por alcanzar un número de iteraciones máximas establecidas.

### 2.4.2. Representación de cromosomas

Un cromosoma se refiere a una cadena de cierta longitud donde se almacena toda la información genética de un individuo, o, en su caso la codificación de la solución a un problema. La elección de una representación conveniente asociada al problema que se considere se convierte en una pieza clave del funcionamiento del algoritmo, particularmente por dos razones [98]:

- (i) Los AG operan manipulando la representación codificada del cromosoma o solución del problema.
- (ii) La manera de representación del cromosoma puede limitar en gran medida la bondad que tiene un sistema en función de su entorno [146].

Con frecuencia se ha utilizado la codificación binaria, donde los cromosomas están conformados por cadenas de bits (ceros y unos). A este tipo de representación se denominan algoritmos genéticos con codificación binaria. Sin embargo, existen otros tipos de codificación, entre las que cabe citar los vectores de números reales [203, 249, 267], vectores de números enteros [36, 89, 126], listas ordenadas o expresiones representadas como árboles [5, 90, 125], entre otras.

### 2.4.3. Operadores genéticos

Todos los algoritmos genéticos poseen tres operadores básicos como son el operador de selección, el operador de cruce y el operador de mutación. Cada uno de dichos operadores generan una fuerte influencia en la mejora de la aptitud de la evolución de los cromosomas o siguientes poblaciones.

El operador de selección tiene su analogía con el proceso natural de competición por conseguir la supervivencia y el apareamiento; el cruce es análogo a la reproducción sexual (cruce de dos cromosomas) donde se combinan las cargas genéticas de los cromosomas o individuos que se aparean; y el operador de mutación simula aleatoriamente diferentes alteraciones que por diversas causas se producen en el código genético de algunos cromosomas. A continuación se explicará su funcionamiento de manera más detallada.

#### 2.4.3.1. Operador de selección

El operador de selección es el primer operador empleado dentro de los algoritmos genéticos. Su objetivo principal se centra en seleccionar los mejores cromosomas o individuos de una población. Este operador no crea ninguna solución nueva, en su lugar selecciona las soluciones relativamente buenas de una población y descarta las soluciones restantes.

El proceso para identificar qué soluciones de una población son buenas o malas se realiza de acuerdo con la aptitud de la solución. La idea principal es que una solución con una mejor valoración de aptitud tenga una mayor probabilidad de selección. No obstante, los operadores de selección difieren en la forma en que las copias se asignan a mejores soluciones. Algunos operadores clasifican la población según la aptitud y eligen de manera determinista las mejores soluciones, mientras que otros asignan una probabilidad de selección a cada solución según la aptitud y hacen una copia utilizando esa distribución de probabilidad.

Existen varios operadores, entre los cuales cabe destacar los siguientes:

- *Selección por torneo*: se eligen subgrupos de cromosomas de la población, y los cromosomas de cada subgrupo pugnan entre ellos. Sólo se escoge a un cromosoma de cada subgrupo para la reproducción.

- *Selección por probabilidad o ruleta*: dependiendo de qué tan buena solución represente proporcionalmente un cromosoma tendrá una mayor probabilidad de ser elegido [98,116].
- *Selección elitista*: emplea un procedimiento que garantiza que la mejor solución de una población estará presente en la siguiente generación [72].

Los operadores de selección están caracterizados por su “*presión selectiva*” o “*tiempo de toma*” [16], que hace referencia al tiempo que requiere para producir una población uniforme. Un operador con una presión selectiva alta disminuye la diversidad en la población más rápidamente que los operadores con una presión selectiva baja, lo que puede llevar a una convergencia prematura hacia soluciones subóptimas. Una presión selectiva alta limita las capacidades de exploración y explotación de la población.

#### 2.4.3.2. Operador de cruce

Al aplicar el operador de cruce se persigue el intercambio de información genética entre individuos [71,116], mecanismo que juntamente con el operador de selección provocan la transmisión y combinación a la “*descendencia*” de aquellos genotipos que se han expresado en cromosomas exitosos. Por este motivo, el operador cruce también suele ser denominado como *operador de recombinación*. Una propiedad importante del operador de cruce es que permite explotar el espacio de búsqueda asociado a los cromosomas padre.

En un inicio es difícil conocer si los nuevos cromosomas (hijos) o descendientes generados poseen una mayor aptitud que sus cromosomas progenitores (padres), pero debido a la lógica aplicada de selección y cruce de los cromosomas con mejor aptitud, inevitablemente, con el tiempo propiciarán a identificar grupos de genes que presuman mejores características deseables, que se transmitirán a su descendencia [159]. Por lo tanto, el operador de cruce ayuda a promover la convergencia hacia las mejores combinaciones de la carga genética ya presentes en la población hacia las próximas generaciones. Existen diferentes operadores de cruce, por ejemplo:

- *Cruce por un punto*: se eligen dos cromosomas que actuarán como progenitores; luego se escoge al azar un punto que divida a los cromosomas en dos partes, y se intercambian la segundas secciones entre ellos.
- *Cruce multipunto*: se seleccionan  $n$  puntos de cruce para obtener la descendencia por intercambio alternado de genes de los progenitores.
- *Cruce segmentado*: es una derivada del cruce multipunto en la que el número de puntos de cruce es variable. Se define una probabilidad —en vez de un valor específico— de que se realice o no el cruce al llegar a cierto gen de la cadena durante la reproducción.

A mayores, se pueden mencionar el cruce por alfabetos de cardinalidad, el cruce por orden, el cruce segmentado, etc. [16]

#### 2.4.3.3. Operador de mutación

El operador de mutación [156] modifica aleatoriamente uno o más elementos (genes) de un cromosoma para incrementar la diversidad estructural de la población.

Su rol en los algoritmos genéticos es reponer código genético perdido o no explorado en la población. De esta forma, se garantiza que la probabilidad de alcanzar cualquier punto del espacio de búsqueda nunca sea nula. Otra bondad de este operador es que al permitir explorar un espacio de solución más amplio minimiza la posibilidad de convergencia a un mínimo local no adecuado.

Cada gen en cada cromosoma puede estar expuesto a una mutación de acuerdo con un criterio de control, denominado *probabilidad de mutación*. La condición de cómo operará el proceso de mutación está directamente relacionada al tipo de representación que se esté empleando para los cromosomas. Por ejemplo, si la representación del cromosoma fuese una cadena con codificación binaria, cada uno de los genes podrían mutar su valor o estado a un solo valor (intercambiando de 0 a 1 y viceversa).

De igual forma que para los operadores de selección y cruce, existe un amplio abanico de posibles operadores de mutación, tales como los siguientes:

- *Operador con mutación uniforme*: se centra en modificar un valor del cromosoma con un valor obtenido al azar. Este puede ser utilizado para genes binarios, enteros y de coma flotante.
- *Operador con mutación no uniforme*: se basa en asignar probabilidades de mutación diferentes a cada gen que conforma el cromosoma [172]. Con ello se puede disponer de probabilidades ajustadas al significado de cada gen, proporcionando al algoritmo un conocimiento concreto del problema a resolver.
- *Operador con mutación gaussiana*: se enfoca en cambiar un valor del cromosoma mediante un método de obtención de números aleatorios con distribución normal [116].

Entre otros operadores de mutación, se pueden mencionar la mutación polinomial, la mutación sobre alfabetos de alta cardinalidad o la mutación por intercambio de posiciones.

#### 2.4.4. Tipos de algoritmos genéticos

Referente a los algoritmos genéticos se pueden analizar dos tipos de modelos, como se expone en [98, 117, 262]:

- *Modelo clásico*. Para cada generación, a través de la selección de los padres de la población anterior y mediante la ejecución de los operadores genéticos sobre ellos, se genera una población completa con nuevos individuos. La población antigua es reemplazada directamente por la nueva población.
- *Modelo estacionario*. Para cada nueva generación se seleccionan dos padres de la población (mediante cualquier tipo de muestreo) y se procede a aplicar los operadores genéticos. Dependiendo del número de nuevos cromosomas obtenidos (uno o dos hijos), reemplazan en igual número de cromosomas de la población, que usualmente son aquellos que tienen peor adaptación. Este modelo genera una presión selectiva alta, al reemplazar a los peores, lo que ocasiona que converja rápidamente.

### 2.4.5. Propagación genética: Teorema de los esquemas

Se denomina como *esquema* a un cromosoma modelo,  $H$ , capaz de observarse replicado en sus semejantes y de ser genéticamente replicable en sus descendientes [117,118]. Analizando una población  $P$  de cadenas construidas sobre un alfabeto binario  $B = \{0, 1\}$ , se dispone de un conjunto de cromosomas binarios.

Sea  $P(t)$  la población en el momento (iteración o generación)  $t$  constituida por los cromosomas  $C_1, C_2, C_3, \dots, C_n$  y sea  $B^+ = \{0, 1, *\}$  el alfabeto para señalar los esquemas incluidos en los cromosomas (individuos) y poblaciones, donde el carácter  $*$  puede simbolizar tanto el valor 0 como el valor 1. De esta forma, por ejemplo, considerando el esquema  $H = (*01**1*)$ , y analizando los cromosomas  $C_1 = (0010010)$ ,  $C_2 = (1011110)$  y  $C_3 = (0110010)$  se observa que los cromosomas  $C_1$  y  $C_2$  lo validan, cosa que no sucede con el cromosoma  $C_3$ .

Los esquemas exhiben ciertas propiedades, como las que se exponen a continuación [166,271]:

- El *orden* hace referencia al número de posiciones con valores fijos presentes en un esquema  $H$  (esto es, los valores distintos de  $*$ ). Se representa mediante  $o(H)$ . Como ejemplo,  $H = (*01**1*)$  tiene orden 3.
- La *longitud* hace referencia a la longitud definida entre la primera y la última posición en un esquema  $H$ . Se representa mediante  $\delta(H)$ . Ejemplos:  $\delta(*01**1*) = 5 - 2 = 3$  y  $\delta(101**1*) = 6 - 2 = 4$ .

La propagación genética se fundamenta en el principio denominado *bloques de construcción*, comprendiendo por bloque de construcción una cadena dentro de un cromosoma que se emplea como modelo para generar nuevos cromosomas.

Una solución eficiente a un problema será aquella que esté constituida por buenos bloques de construcción. El operador de cruce es el responsable de combinar bloques buenos que se identifiquen en los distintos progenitores, los que traspasarán a sus descendientes para mejorar los valores de aptitud. Por otra parte, la *presión selectiva* es la encargada de filtrar que solo los bloques más idóneos se conserven, y poco a poco se vaya convergiendo a una óptima solución.

El denominado *teorema de los esquemas* permite comprender cómo el mecanismo del GA, a medida que transcurre el tiempo de ejecución, va generando bloques óptimos. Es oportuno analizar cómo cada fase de un GA produce un resultado sobre los esquemas:

- *Resultado de la reproducción.* Sea  $s = s(H, t)$  el conjunto de las observaciones del esquema  $H$  en la población  $C(t)$ . Un cromosoma (individuo o cadena) según su aptitud es elegido para su reproducción —más específicamente, se puede exponer que  $C_i$  será escogido con una probabilidad  $p_i = \frac{f_i}{\sum_{j=0}^n (f_j)}$ . Posterior al reemplazo se espera obtener:

$$s(H, t + 1) = s(H, t) * \left( \frac{f(H)}{\sum_{j=0}^n (f_j)} \right) * N$$

donde  $N$  representa la dimensión de la población, y  $f(H)$  corresponde a la aptitud promedio de los cromosomas que representan al esquema  $H$ . Por lo

tanto, un esquema se incrementa manteniendo una relación entre su aptitud promedio y la aptitud de la población. De esta manera, los esquemas con valor promedio mayor al aptitud promedio de la población dispondrán de un mayor número de muestras en la siguiente generación. Por contra, aquellos esquemas con menor aptitud promedio que la aptitud promedio de la población dispondrán de un menor número de muestras en la siguiente generación.

- **Resultado del cruce.** Siendo  $H_1 = **11**0*$  y  $H_2 = **1***00$  dos esquemas descritos por el cromosoma  $C_i = 00111100$ , se realiza la operación de cruce en un punto, considerando que el producto del cruce se realiza en una posición aleatoria entre dos cromosomas padres (cromosomas binarios), donde se intercambia el contenido genético entre ellos. Por ejemplo el cromosoma  $C$  es seleccionado para la reproducción con punto de corte 4:

$$- H_1 = **11|**0*$$

$$- H_2 = **1*|**00$$

$$- C_i = 0011|1100$$

Se puede observar que  $H_2$  se encontrará presente por lo menos en uno de los cromosomas hijos, en tanto que  $H_1$  no, debido de que dicho esquema se eliminará desde un inicio. Al analizar  $H_1 = **11|**0*$  se tiene  $\delta(H) = 4$ . Considerando que el punto de corte es elegido uniformemente, entre las  $L - 1 = 7$  posiciones viables, siendo  $L$  la distancia, la probabilidad de que  $H_1$  se elimine es  $P_e = \delta(H_1)/(L - 1) = 5/7$ , mientras que la probabilidad de vida sería  $P_v = 1 - P_e = 2/7$ . Por ello, se puede expresar de manera general que la probabilidad de que un esquema sobreviva al realizarse el cruce es:

$$P_v = 1 - \delta(H)/(L - 1)$$

Por lo tanto, se puede enunciar que si se designa a  $P_c$  como la probabilidad de cruce, entonces:

$$s(H, t + 1) = s(H, t) * \left( \frac{f(H)}{\sum_{j=0}^n (f_j)} \right) * \left( 1 - \frac{P_c * \delta(H)}{L - 1} \right)$$

- **Resultado de la mutación.** Sea  $P_m$  la probabilidad de mutación de un gen, y sea  $L$  el tamaño de un cromosoma binario. La probabilidad de que el esquema  $H$  subsista al proceso de modificar al azar un gen parte de dicho cromosoma viene dada por  $(1 - P_m)^{o(H)}$ , de lo que se deduce lo siguiente:

$$(H, t + 1) = s(H, t) * \left( \frac{f(H)}{\sum_{j=0}^n (f_j)} \right) * \left( 1 - \frac{P_c * \delta(H)}{L - 1} \right) * (1 - P_m)^{o(H)}$$

Sobre este resultado se fundamenta el *teorema de los esquemas* [116], que se enuncia como sigue: “*esquemas cortos, de bajo orden y con adaptación por encima de la media, ... reciben un incremento exponencial de cromosomas en las siguientes generaciones*”. Como resultado derivado, se tiene la denominada *hipótesis de los bloques constructores* [98]: “*La búsqueda realizada por un algoritmo genético con codificación binaria es cuasi-óptima a causa de la yuxtaposición de los bloques constructores.*”

### 2.4.6. Fortalezas y debilidades de los algoritmos genéticos

Los GA se han utilizado ampliamente en un rango de problemas asociados a ámbitos tan diversos como la medicina, la ingeniería, la economía y la educación, destacando entre sus ventajas las siguientes:

- Son algoritmos que se pueden hibridar fácilmente.
- Tienen la cualidad de aprender a identificar o descartar soluciones en función de su calidad.
- Permiten encontrar soluciones para problemas de difícil solución de una manera viable, rápida y eficiente.
- Generalmente permiten generar varias soluciones lo suficientemente próximas a la óptima para un problema en particular.

Entre las desventajas, por su parte, cabe citar:

- *Problema de convergencia prematura hacia mínimos locales subóptimos* [150]. Este inconveniente reside en el inadecuado mantenimiento del equilibrio entre explotar lo que es mejor mediante el cruce, en el momento actual, y explorar posibilidades mediante la mutación que pueden conllevar a un mejor resultado. Para solventar este inconveniente se han desarrollado varias soluciones [85, 227].
- En problemas de alto grado de complejidad, la función de evaluación puede ser excesivamente costosa en relación al tiempo y los recursos.
- Pueden existir casos en los que, dependiendo de los atributos que se empleen para la evaluación, el algoritmo podría no terminar convergiendo.
- El diseño y el desarrollo de la función de aptitud y la elección de los criterios de selección y mutación, entre otros, requieren de cierta experiencia y conocimiento del problema para conseguir buenos resultados.



## Capítulo 3

# Aplicaciones de técnicas de Inteligencia Artificial en las instituciones de Educación Superior

*Durante los últimos años, las instituciones de Educación Superior (en adelante, IES) han experimentado grandes cambios con el objetivo de mejorar los procesos y metodologías de enseñanza-aprendizaje, con la finalidad de formar profesionales cada vez más capacitados para afrontar los nuevos desafíos que supone una sociedad en constante cambio. Bajo este contexto, las TICs han realizado aportes significativos desde diferentes frentes, pero sigue vigente la posibilidad de crear y proponer nuevas herramientas que sirvan de soporte para mejorar el ecosistema educativo. Un elemento de innovación a gran escala en la educación en la actualidad consiste en el empleo de la Inteligencia Artificial para ayudar tanto a docentes como estudiantes a mejorar su desempeño durante el proceso educativo. Este capítulo pretende dar una visión global del acercamiento de la IA (interdisciplinaria de por sí) y las ciencias cognitivas (psicología, neurociencias, lingüística, sociología, matemáticas, física, etc.) a la Educación Superior, orientado a promover el desarrollo de entornos de aprendizaje adaptativos y otras componentes flexibles, personalizados, inclusivos, amigables y efectivos.*

### 3.1. Introducción

Como se explicó en la Sección 1.4, la presente tesis tiene como objetivo forjar las bases de un completo ecosistema semántico de apoyo a la administración y gestión de la información que se genera en torno al ámbito de la Educación Superior, desde una perspectiva fundamentada en el modelado del conocimiento y el empleo de innovadoras técnicas de inteligencia artificial. Las contribuciones hacen uso de las técnicas descritas con más profusión en el Capítulo 2, a saber: redes ontológicas, redes neuronales y algoritmos genéticos. A fin de situar debidamente estas contribuciones en el estado del arte, a lo largo de las siguientes páginas se realizará una revisión bibliográfica de trabajos previos en aplicación de IA a las IES, desde un

punto de vista general pero, nuevamente, prestando especial atención a las áreas que guardan mayor relación con los capítulos 4, 5 y 6:

- (i) Modelado del conocimiento en las instituciones de Educación Superior.
- (ii) Sistemas de predicción de rendimiento académico.
- (iii) Formación de grupos de aprendizaje colaborativo.

Previamente, se revisarán trabajos en materia de juegos serios y uso de simuladores en Educación Superior que, si bien comportan elementos de IA simples, son representativos de los avances registrados en la aplicación de TICs a esta área. A estos puntos les siguen los sistemas de tutoría inteligentes y los tres arriba mencionados.

### 3.2. Juegos serios como elementos de refuerzo en el aprendizaje

Los juegos serios (JS o SG, acrónimo de *Serious Games* en inglés) son juegos digitales desarrollados con un propósito eminentemente educativo. Los JS permiten promover la educación al plantear retos que pueden ser cuantificados para conocer el nivel académico de aprendizaje del estudiante, y pueden impactar positivamente sobre el aprendizaje mejorando los aspectos cognitivos, conductuales, afectivos y motivacionales, mejorando el rendimiento dentro de las aulas. Mediante el uso de los juegos serios se aplican las teorías pedagógicas que explicitan que el estudiante construye mejor su conocimiento a través del paradigma de “*aprender haciendo*”. El factor de mantener la motivación a través del uso de los JS permite potenciar el aprendizaje del estudiante. Los JS han cobrado gran relevancia para el desarrollo de competencias, habilidades y destrezas en diversas disciplinas o carreras de formación profesional, permitiendo ante todo adaptar las metodologías de los docentes a las características de los estudiantes, a sus necesidades, maneras de aprender y conocimientos.

Calador [37] describió en 2018 la experiencia de implementación de juegos serios en el desarrollo de una materia de contabilidad de gestión, permitiendo con ello la evaluación de la adquisición de competencias genéricas y específicas contempladas en los contenidos de la materia. El juego empleado fue “*Platform Wars Simulation*”, desarrollado por el Instituto Tecnológico de Massachusetts. Se trata de una investigación empírica que demuestra la efectividad de los JS en el proceso de aprendizaje en ambientes universitarios y, particularmente, en el área de gestión. La aplicación de cuestionarios pre- y post-actividad permite constatar que no existen barreras tecnológicas ni demográficas, además de ser una herramienta pedagógica que contribuye significativamente para el desarrollo de competencias instrumentales, interpersonales y sistémicas relacionadas con el área estudiada.

En el año 2013, Muñoz-Arteaga *et al.* [106] propusieron un modelo de integración de juegos serios educativos en un objeto de aprendizaje que abordaba una variedad de aspectos culturales, con el fin de enseñar y motivar en la educación pluricultural de los estudiantes universitarios. El objetivo de utilizar tales objetos de aprendizaje era facilitar el consumo de recursos digitales en entornos virtuales para que los diferentes usuarios de distintas regiones y costumbres accediesen desde diferentes

dispositivos a los servicios y repositorios multiculturales de varias universidades mexicanas. El modelo se compone de los siguientes elementos:

- *Teoría* con la información de los temas, conceptos y formulas a estudiar.
- *Autoevaluación* mediante preguntas sobre la temática.
- *Objetivos y temas relacionados* para describir la finalidad del contenido.
- *Práctica* para integrar el videojuego para realizar actividades con retroalimentación.
- *Metadatos* que describen en detalle cada elemento de multiculturalidad, tradiciones, costumbres y ubicación geográfica.

En respuesta a las exigencias en la innovación de métodos para la enseñanza en el Espacio Europeo de Educación Superior (EEES), en 2015 Poy-Castro *et al.* [209] plantearon el diseño y evaluación de un juego serio para la promoción del trabajo en equipo en estudiantes universitarios. La investigación giraba en torno al análisis de la adquisición y desarrollo de diversas competencias en una actividad de control de tráfico aéreo, donde los jugadores debían combinar una serie de acciones, estrategias, decisiones y negociaciones para evitar colisiones entre ellos. El estudio contribuyó a una mayor comprensión del potencial de los juegos simuladores en la motivación de experiencias para los estudiantes. Se evidenció que el sistema proporcionaba capacidad de desarrollar las competencias de trabajo en equipo, resultando más eficiente que a través a los métodos tradicionales de enseñanza.

En 2015, Díaz y Queiruga [80] incursionaron a través de su investigación en videojuegos inmersivos y juegos de plataforma para contextos educativos. El primero, *eQuino*, es un juego en 3D que se enfoca en la inmersión sensorial a través del desarrollo del arte y la imaginativa por medio del ambiente ecuestre. La línea gráfica se reprodujo lo más fielmente posible para promover el reconocimiento del ambiente y afianzar el aprendizaje de las actividades, destacando tres componentes principales: el ambiente que recreaba, los personajes en primera persona, y las actividades o desafíos que son representaciones de las actividades que los estudiantes realizaban en el centro formativo. El segundo videojuego, *RAICES*, era un videojuego social de plataformas cuya temática giraba entorno a pueblos originarios. Este género se reveló ideal para ser utilizado en las redes sociales, porque implica poner en contacto a todos los miembros de un determinado grupo de estudiantes de una clase, motivando a la cooperación y al trabajo en equipo. Ambos videojuegos fueron pensados para mejorar y enriquecer las experiencias educacionales en el aula y fuera de ella, y la experimentación demostró que contribuyen a disminuir el estrés y la ansiedad que el alumnado pueda sentir ante la necesidad de interiorizar nuevos conocimientos, destrezas o habilidades. Como resultado, se comprobó que el juego les resultaba “*divertido*” y “*motivador*”, al tiempo que se pudo constatar que los estudiantes continuaban las partidas fuera del aula (en el caso de RAICES) y del centro formativo (en el caso de eQuino), jugando durante su tiempo de ocio para afianzar los conocimientos y destrezas.

En la misma línea, en el año 2016, Valencia *et al.* [255] presentaron un juego donde el estudiante desempeña el rol de un jefe de proyecto de desarrollo de software, con trabajadores localizados en distintas partes del mundo, y tiene que hacer frente a problemas de coordinación. Un ejemplo podría ser la propia deslocalización de los

equipos, que implica problemas de comunicación, coordinación y control, así como aquellos derivados de las diferencias culturales de los participantes del equipo. Estos eventos son producidos de manera aleatoria. El juego cuenta con distintos escenarios, en distintos niveles de dificultad para que el estudiante adquiera conocimientos de forma gradual. Además, la aplicación permite a un profesor crear escenarios personalizados, por si quiere que el alumno practique un escenario en concreto y a su vez éste podrá consultar la ayuda cuando no sepa cómo superar cierto escenario. El estudiante debe conseguir desarrollar todas las fases que componen cada módulo, cumpliendo con las condiciones en el tiempo y el presupuesto.

Desde otro punto de vista, en el marco del contexto educativo japonés para la enseñanza del español como lengua extranjera, García Joao [94] desarrolló una importante propuesta de integración de las TIC y ludificación, planteando nuevas vías para la mejora del proceso de enseñanza-aprendizaje mediado por *entornos virtuales del aprendizaje* (EVA). La estrategia de la propuesta se sustenta en investigaciones previas desarrolladas por el Grupo de Investigación para la Didáctica del Español de la Universidad de Tokio. El escenario de aplicación lo constituye un grupo de seis centros de Educación Superior. El éxito de la propuesta gira en torno a la consideración de las características del entorno educativo, elementos como la infraestructura digital de cada centro, la formación tanto de docentes como de estudiantes de nuevas tecnologías, motivación por parte de los estudiantes y disponibilidad de apoyo por parte de otros departamentos o unidades administrativas para el uso de recursos digitales.

El desarrollo de competencias desde el punto de vista técnico es fundamental para la formación de profesionales en carreras de Ingeniería de la Computación, Tecnologías de la Información, Ciencias de la Información e Ingeniería de Software. No obstante, como se ha expuesto en trabajos anteriores, el Espacio Europeo de la Enseñanza de Educación Superior recomienda la formación de habilidades y competencias transversales o *“soft skills”* como el trabajo en equipo, capacidad de comunicación, negociación, liderazgo, etc. en pos de un mejor desempeño en el ámbito laboral. En este sentido, Ortiz [197] planteó en 2018 el desarrollo de competencias personales y habilidades sociales en la Ingeniería Informática a través de la puesta en práctica de dos juegos serios, denominados *“Ball Point Game”* y *“Dealing with Difficult People”*. La experiencia llevada a cabo en la Escuela Superior de Ingeniería de Cádiz fue evaluada por los estudiantes de tercer y cuarto año a través del modelo *MEEGA+* [206], teniendo como indicadores la experiencia del usuario y la percepción del aprendizaje. Los datos recogidos fueron analizados mediante técnicas propias de la estadística descriptiva, resultando una propuesta positiva para el fomento de las competencias transversales y profesionales.

### 3.3. Uso de simuladores en las instituciones de Educación Superior

La simulación hace referencia al proceso de experimentar con un modelo. Para comprender el proceso de simulación es necesario definir, diferenciar y establecer las relaciones entre dos conceptos clave [242]: por una parte, por *sistema* se comprende el conjunto de objetos o ideas que se relacionan entre sí de cara a la consecución de un determinado objetivo; por otra, el *modelo* permite la representación de conjunto

de ideas, sistemas u objeto con la intención de explicar, mejorar o entender un sistema —en este sentido, un objeto A es un modelo del objeto B para el observador C, si C puede usar A para responder a preguntas que le interesan acerca de B [174]. La *simulación* supone una serie de pasos para diseñar un sistema real para llevar a cabo experiencias que permitan aprender y comprender su comportamiento así como evaluar las diferentes estrategias para el funcionamiento del sistema [229].

La revisión bibliográfica del uso de simuladores en el ámbito educativo para el desarrollo de competencias y mejora del proceso de enseñanza-aprendizaje permitió obtener grandes e interesantes hallazgos a través de investigaciones centradas en diferentes disciplinas como la medicina, la administración, la enfermería, la psicología y la ingeniería. Al desarrollar trabajos que involucran a seres humanos directamente, las previsiones a tener en cuenta para la formación profesional ameritan tomar cautela por los efectos que un servicio de baja calidad puede provocar en la vida de otras personas. Para controlar estos riesgos y hacer posible la realización de prácticas controladas sin límite, las escuelas de medicina y enfermería vienen desarrollando desde hace algunos años múltiples tipos de simulaciones, que van desde el empleo de maniqués (e.g. para ejercicios de respiración artificial) hasta software que simula la interacción con pacientes y la aplicación de procedimientos. De igual manera, en el área de formación de administradores, ingenieros, contadores, ... es posible encontrar simulaciones sobre creación de empresas, manejo de conflictos, liderazgo, toma de decisiones, manejo de las ventas, entre otros. Las simulaciones se fundamentan en el método de “*aprender haciendo*” y existe evidencia de que pueden hacer del aprendizaje un proceso más efectivo, permitiendo a los estudiantes practicar sus conocimientos y desarrollar en un entorno real de una manera segura, libre de las consecuencias en la vida real de los errores. No obstante, para aumentar las posibilidades de éxito en cuanto a las habilidades y competencias para el campo laboral, se requiere un énfasis en todos los componentes del sistema educativo, no solo en el aspecto de la simulación como tal [105].

A continuación se resumen una serie de investigaciones que hacen referencia a los avances de la educación en cuanto al uso de simuladores para el desarrollo de habilidades y destrezas y contribuir a las mejoras del proceso de enseñanza-aprendizaje, basados en TICs y con aportes de técnicas de IA.

El surgimiento del *constructivismo* como teoría del aprendizaje dio lugar a la idea de la construcción activa del conocimiento por parte de los estudiantes mediante la interacción con el entorno y a través de un proceso de reorganización de sus estructuras mentales [127, 252]. Desde este punto de vista, el aprendizaje se centra en el estudiante como constructor de conocimientos mediados por el facilitador (docente). Por su parte, autores del constructivismo social resaltan la importancia del contexto, pero incorporan la necesaria interacción entre el individuo y su entorno para el desarrollo de la cognición y el aprendizaje que es producto de la actividad, el contexto y la cultura en la cual se forma. En atención a estas consideraciones teóricas, Pantoja, Amaya, Sotomayor, Chávez y Vidal presentaron en 2018 el diseño de un “*simulador de ficha clínica electrónica*” (SFC) como herramienta para el fortalecimiento del aprendizaje activo [202]. El simulador se basa en una aplicación web que proporciona a los estudiantes y profesores herramientas para crear sus propias fichas clínicas o trabajar sobre otras ya elaboradas, centradas en las distintas actividades de la disciplina. La validación del simulador se hizo a través de una prueba piloto en el transcurso de la asignatura coordinada por parte del equipo promotor y la participación de un grupo de 16 estudiantes. Para la valoración se aplicó una

encuesta de opinión, con la cual se evaluó de forma positiva el SFC destacando la utilidad, la facilidad para el manejo y la accesibilidad. El simulador constituyó una herramienta complementaria para la adquisición del conocimiento y la promoción del aprendizaje activo.

Con el objetivo de exponer una serie de experiencias en cuanto a la incorporación de simuladores en el proceso de enseñanza-aprendizaje de la física, química y electrónica; Ariel y Forero [213] presentaron la descripción de un modelo metodológico educativo para el uso de simuladores en software libre, para identificación y documentación de recursos de software, y para el diseño de guías de implementación sustentado en teorías modernas sobre el aprendizaje. Del diseño de prototipo de guía para el uso educativo de simuladores de física a nivel superior resultó el proyecto denominado “*MODELLUS: Interactive modelling with Mathematics*”. Para la validación de la propuesta se aplicó una matriz *DOFA* [43] a dos grupos de estudiantes pertenecientes al programa de Tecnologías en Obras Civiles en modalidad a distancia, los cuales participaron en el proceso formativo. Los resultados permitieron conocer la existencia de software libre como una alternativa de fácil acceso para las instituciones educativas, bajo costos y de calidad en la mejora de la formación de profesionales.

En este mismo orden de ideas, la Universidad Católica San Antonio de Murcia (UCAM) integró dentro de los diferentes *practicum* clínicos la simulación como itinerario formativo en el programa de enfermería, según una estrategia para introducir metodologías centradas en los alumnos. En el trabajo titulado “*La simulación clínica como herramienta pedagógica*”, Juguera *et al.* [135] analizaron en 2014 la percepción de los alumnos de segundo y tercer curso del Grado en Enfermería acerca de la simulación clínica. Se trata de una investigación de tipo cualitativa que utilizó un cuestionario semiestructurado para el desarrollo de entrevistas. Los estudiantes indican que lograron la adquisición de competencias como el trabajo en equipo, comunicación y destacan la importancia de la simulación para la corrección de errores.

También en las escuelas de administración, contabilidad y mercadeo el uso de simuladores educativos ha generado una importante herramienta de análisis para el diseño y la ejecución de sistemas complejos. En el marco de una escuela de administración de negocios, León y Cañas [155] describieron las características y experiencias de estudiantes y profesores en cuanto al uso del denominado “*Simulador Gerencial de Operaciones*”, que para el año 2014 cumplía 6 años de aplicación. Los estudiantes no contaban con experiencias previas en ambientes de operaciones o logística, lo cual limitaba el desarrollo de una serie de capacidades y habilidades que les permitieran desempeñarse como ejecutivos exitosos. La aplicación simulaba la organización de empresas que participaran en una subasta pública de productos para lograr una adjudicación dependiendo de la empresa que proponga el precio más bajo. Una vez asignada la adjudicación, se da inicio a la producción poniendo en práctica los conocimientos de la ingeniería y la administración vistos en clase promoviendo el pensamiento, el desarrollo de habilidades y capacidad de análisis de situaciones y toma de decisiones. La empresa que obtenga más recursos resulta ganadora. Finalmente, los estudiantes deben realizar un informe sobre la experiencia, exponiendo los aspectos más relevantes, las decisiones tomadas, su aprendizaje y conclusiones.

Otra investigación desarrollada por Fiad y Galarza en 2015 [87] evaluó la implementación del “*Laboratorio Virtual de Química General*” en el aprendizaje sobre de

nociones básicas de química en la Universidad Nacional de Catamarca (UNCa). La problemática que se planteaba está relacionada con las dificultades generales que se presentan en la enseñanza-aprendizaje de las Ciencias Naturales, marcadas por bajos rendimientos, desmotivación y, en última instancia, deserción estudiantil. La problemática conlleva al análisis de aspectos relacionados con la enseñanza tradicional de las Ciencias Naturales, el poco uso de estrategias didácticas innovadoras, entre otras, y no sólo con las capacidades intelectuales de los estudiantes [19]. En atención a estas situaciones se diseñaron los “*Laboratorios Virtuales de Química (LVQs)*” como herramientas pedagógicas basadas en entornos virtuales que simulaban laboratorios para el ensayo de procesos químicos. Para el experimento desarrollado se tomó en consideración la evaluación del período 2010-2014, constatando los errores de concepto que tenían la mayoría de los estudiantes. Se trabajó con todos los alumnos ingresantes en 2014, dividiéndolos en un grupo control (GC) y otro grupo experimental (GE). En el primer parcial, un porcentaje alto, representado por el 90% de los estudiantes del GE, respondió de manera correcta el ítem que correspondía a las cantidades atómico-moleculares, mientras que en GC sólo lo hizo el 45% de los alumnos. La valoración del experimento constata que los alumnos del GE desarrollaron habilidades cognoscitivas gracias a la interacción con el simulador; de hecho, los estudiantes obtuvieron un desarrollo de aprendizaje alto, al tiempo que mostraron una actitud positiva hacia los conceptos trabajados y la estrategia para desarrollar la clase.

También desde las ciencias médicas se ha avanzado notablemente en los últimos años con la utilización de aplicaciones de la IA en pro de la mejora de la enseñanza. Dentro de esta perspectiva, Coro *et al.* [58] presentaron en 2015 los resultados obtenidos en el diagnóstico relacionados con los elementos didácticos que deben orientar el uso de tecnología háptica, basada en la interacción mediante el tacto, con realidad virtual en la carrera de Odontología. Se puso en práctica un simulador de última generación que supone el desarrollo de una situación educativa compleja: “*Moog Simodont Dental Trainer*”, un proyecto transversal desarrollado por fases en las asignaturas Odontopediatría I y II y Odontología Restauradora II. De acuerdo con el criterio de expertos, el simulador es un recurso didáctico que puede ser de gran provecho en el campo educativo. Sobre la base de la metodología de *Investigación Acción* (IA) [57], se llevó a cabo un estudio que tenía como objetivo crear una comunidad de prácticas, con una muestra de 22 docentes, de una población total de 45 profesores que aplicaban al proceso de simulación. En el proceso de experimentación, se emplearon técnicas de investigación propias de la investigación con metodología cualitativa [167] como son la observación, aplicación de entrevistas en profundidad, diarios reflexivos del investigador, registros audiovisuales, la comparación, el contraste, el análisis y la sistematización de los datos aportados por los profesores que son parte de la muestra. Dentro de los resultados pueden mencionarse los siguientes:

- El instructor de simulación juega un papel medular, debe estar al tanto y tener el control permanente de la situación simulada para no desperdiciar ninguna de las propiedades adicionales del simulador y poder generar nuevas oportunidades de aprendizaje significativo [81]. Se hace estrictamente necesario dominar y comprender los principios didácticos y cognitivos que rigen el uso del simulador.

- Simodont permite ir haciendo un registro sobre la evolución de las destrezas, de detectar el origen y la razón de los errores, así como magnificar las manio-bras intra-orales y generar documentos para archivos; sin costos en materiales.
- Los profesores consultados destacan la capacidad de combinación del uso de la tecnología de punta con estrategias de aprendizaje como métodos del caso, *feedback* y *role playing*, sólo destinadas a especialidades centradas en campos operatorios amplios, visibles y de fácil acceso.
- Se valora como un elemento significativo del proceso de simulación el tra-bajo de equipo y la interacción y contradicción existente entre los modelos mentales de alumnos e instructores.
- Dentro de las desventajas observadas se encuentra la insuficiente formación del cuerpo docente en el uso de las TICs, lo cual, de acuerdo con sugerencias de los encuestados puede ser resuelto con métodos híbridos.

Por su parte, Forrero y Giraldo [88] presentaron en 2016 los resultados obteni-dos con el uso de un modelo de simulación del proceso de fabricación de bicicletas en un curso de Ingeniería Industrial de la Universidad Nacional de Colombia. El proceso consta de la simulación de dos situaciones de operatividad del sistema y la comparación con el desempeño, teniendo como indicador la cantidad de bicicletas fabricadas así como los cánones con que se utilizan tres áreas vinculadas con el proceso. El estudio sienta sus fundamentos teórico-metodológicos en la enseñanza basada en la resolución de problemas, ubicando al estudiante en el centro del pro-ceso, asumiendo un papel activo de su aprendizaje [196]. En respuesta al estudio de las condiciones y requerimientos de formación, el proceso de construcción y si-mulación del modelo se estructuró por etapas consecutivas, a saber: definición del sistema/proceso a modelar, construcción del modelo de simulación, definición del modelo de decisión, representación del modelo de decisión en la simulación, inter-faz usuario-modelo y experimentación con el modelo. La fase de interacción con el modelo, dado a través de hojas electrónicas, permitió a los estudiantes la com-prensión de las principales relaciones entre causas y efecto en el proceso. El informe resalta la importancia del sistema modelado y su uso como soporte didáctico en las asignaturas vinculadas con la gestión de sistemas de producción.

Bajo esta misma perspectiva, la Universidad de la Sierra Sur de Oaxaca (México), se propuso la simulación clínica como una herramienta educativa para la formación en el área de la enfermería con alcances de la modernización y tecnificación del sistema de salud mexicano, tanto público como privado. En este sentido, Martínez, Hernández y Jiménez (2016) [168] explicaban que la dotación en infraestructura y equipamiento especializados permite al estudiante de diferentes niveles practicar procedimientos y técnicas de Enfermería en un ambiente controlado y seguro, me-diante casos clínicos preprogramados con variaciones en los niveles del estado de salud de los pacientes. Para la valoración del simulador hospitalario se realizó un estudio descriptivo de corte transversal durante tres meses y se aplicó la prueba “*chi cuadrado*” con el objeto de establecer las diferencias significativas entre las variables propuestas. De acuerdo con resultados obtenidos en un análisis descriptivo, el uso del simulador constituye una herramienta importante para generación de competen-cias y habilidades en los estudiantes de enfermería; no obstante, se hace necesaria la supervisión permanente del docente para la retroalimentación de los procesos simulados.

Partiendo de un análisis de los métodos tradicionales de enseñanza-aprendizaje y sus consecuencias en cuanto al rendimiento estudiantil, la permanencia, retención o deserción estudiantil, así como la calidad de los procesos formativos de las instituciones de Educación Superior del Ecuador, los investigadores Cárdenas, Cajamarca y Mantilla (2017) [63] establecieron mediante un trabajo con enfoque cualitativo el impacto que genera tanto para estudiantes como para profesores la implementación de tecnologías como el simulador *Monte Carlo* de @RISK. La propuesta se desarrolló con estudiantes de la asignatura Administración Financiera de la carrera de Administración de Empresas con modalidad de estudios presenciales. La metodología consistió en un estudio de caso y la participación un grupo de control y un grupo de simulación. A través del uso de las TIC's y la implementación del simulador se evidenciaron cambios significativos en la enseñanza desde el punto de vista cognitivo, a nivel práctico y en las actitudes de los estudiantes, repercutiendo de manera positiva en el proceso de enseñanza. Dentro de las recomendaciones destaca la necesaria formación, motivación y concienciación del claustro docente en relación con las TIC y la capacitación en el uso de simuladores educativos entre otros recursos tecnológicos.

Otra experiencia innovadora es el uso de simuladores de negocio como estrategia de aprendizaje adaptativo desarrollada en la Universidad de La Salle, Colombia. De acuerdo con investigadores de la *Education Growth Advisors* (EGA) [188], el aprendizaje adaptativo es un enfoque para la creación de experiencias de aprendizaje personalizado y mediado por la utilización de un esquema de datos computerizados. El principio diferenciador del método de aprendizaje adaptativo es la aproximación no lineal del proceso de instrucción, retroalimentación y corrección, lo cual hace posible ajustar el aprendizaje (contenidos y recursos) de acuerdo con el nivel de desempeño mostrado y las interacciones del estudiante. En este sentido, Vargas [170] presentó en 2018 evidencias del uso del simulador *Business Global: The Company* a través del análisis del modelo de evaluación aplicado en relación con las actividades desarrolladas y resultados alcanzados por los estudiantes. Business Global simula un escenario de acción de cinco compañías dedicadas a las tecnologías del hogar que compiten entre sí. Con el propósito de lograr la asimilación de los conocimientos vinculados al área administrativa, el desarrollo del ejercicio estuvo estructurado en fases que involucran el alcance de niveles cognitivos de manera consecutiva: conocer, comprender, compartir y crear. De acuerdo con la estructura presentada, se organizan equipos de tres o cuatro estudiantes y el proceso inicia con el registro de los estudiantes en la plataforma *Company Game* para la asignación de una industria. Sobre la base de los resultados de una matriz FODA, los participantes deben generar un plan de negocio y un mapa estratégico con la determinación de objetivos y decisiones proyectadas para cuatro años. El seguimiento a las acciones planteadas se realiza desde un *Objeto Virtual de Aprendizaje* (OVA), diseñado para la experiencia en el aula y desde lo que se denominó el *cuadro de mando integral* usado como tablero. A partir de los resultados allí plasmados se genera una realimentación en tiempo real que permite evaluar y ajustar el proceso de acuerdo con el desempeño de manera grupal e individual. El ejercicio concluye con una retroalimentación de los resultados del informe, el cual hace referencia a los avances en el desempeño de cada participante y las dificultades en el desarrollo de las actividades.

A través de las investigaciones desarrolladas se puede valorar la experiencia positiva, que permite destacar que mediante la adaptabilidad de los simuladores a las condiciones y necesidades de los estudiantes se tiene un componente innovador para

la transformación del proceso de enseñanza-aprendizaje que promueve un aumento en el desempeño académico, una mayor motivación hacia los contenidos y mayor seguimiento en tiempo real por parte de los docentes. Otro elemento a destacar es la optimización de los procesos y sus efectos sobre el aprovechamiento del factor tiempo mediante el uso de simuladores.

### 3.4. Sistemas de tutoría inteligentes

Los *sistemas de tutoría inteligentes* (del inglés *Intelligent Tutoring Systems* o ITS) son herramientas tecnológicas que se han creado con el objetivo de coadyuvar en las labores de los docentes que desempeñan tareas y/o funciones de tutor y así simplificar el aprendizaje de los estudiantes. Los ITS se fundamentan en ser aplicaciones que emplean técnicas de IA para modelar el conocimiento a impartir, así como las características de los estudiantes, para identificar las falencias de cada uno y reforzar el aprendizaje como si se dispusiese de un docente tutor.

Mitrovic, Olhsson y Barrow presentaron en el año 2013 [178] un ITS denominado SQL-TUTOR, con la capacidad de generar una retroalimentación positiva basado en un conjunto de restricciones. En el experimento se pudo constatar que, aunque el rendimiento entre el grupo de control y experimentación no fue muy significativo, los estudiantes que emplearon SQL-TUTOR se tomaron aproximadamente un 40% menos de tiempo en adquirir los mismos conocimientos que el grupo de control.

A través de su investigación, Latham, Crockett y McLean dieron a conocer en 2014 su ITS conversacional, denominado *Oscar CITS*, que empleaba un novedoso algoritmo de adaptación basado en los estilos de aprendizaje de Felder-Silverman. Oscar CITS analiza el comportamiento del estudiante universitario durante cada sesión de trabajo para predecir su estilo de aprendizaje y adaptar de manera única el material de tutoría en consonancia. Los resultados experimentales sugieren que los estudiantes a quienes se les presentó el material de aprendizaje acorde con sus estilos de aprendizaje individuales alcanzaban un mejor rendimiento (en promedio, un 12% mejor) [151].

En la investigación de Gorrostieta, Gonzáles y López (2014) se presentó otro ITS que utilizaba la web para asesorar a los estudiantes en el desarrollo de borradores de temas de investigación. El ITS emplea tecnologías de lenguaje natural para poder automatizar la evaluación de la escritura y poder realizar una retroalimentación. Adicionalmente, hace uso de un analizador léxico que permite tener un nivel adecuado de variedad y sofisticación en los escritos de los estudiantes. Mediante un estudio empírico se logró demostrar que los estudiantes que emplearon el ITS mejoraron los tres aspectos léxicos asociados: densidad, diversidad y sofisticación, en contraste con aquellos que no lo utilizaron. Tres años más tarde, los mismos autores presentaron una nueva versión del ITS bajo el nombre de *Oscar FLSP (Fuzzy Learning Style Prediction)*. En esta versión emplean árboles de decisión difusos, que les permitían construir una serie de modelos de predicción que combinan las variables de comportamiento con todas las dimensiones de los estilos de aprendizaje de Felder-Silverman. Esta combinación permite mitigar el inconveniente de presentar eventualmente material no acorde al estilo de aprendizaje del estudiante, pues la metodología realiza una predicción más exacta referente a los estilos de aprendizaje de hasta un 28% comparado con la primera versión.

Hooshyar *et al.* [120], por su parte, presentaron en 2017 un ITS bayesiano denominado *OGITIS*, dirigido a la adquisición de conocimientos individuales de programación informática. El sistema emplea juegos en línea que les permiten a los estudiantes desarrollar sus habilidades de programación y de búsqueda de información en línea. Para ello, emplean juegos de mesa en la web que permiten mejorar la capacidad de los estudiantes mediante la resolución de ciertos problemas. Se realizó un estudio con 79 estudiantes universitarios, y los resultados indicaron que un 75.4% de estudiantes manifestaron que *OGITIS* colaboró en su aprendizaje al facilitar contenidos adecuados, en tanto que el 74.1% y 75.3% expusieron que *OGITIS* mejoró su programación informática y sus destrezas de búsqueda de información. En la misma línea, el mismo grupo de investigadores desarrolló en 2018 el sistema *SITS*, también basado en redes bayesianas para la gestión de la incertidumbre en el proceso de toma de decisiones, pero incorporando un procedimiento automático de conversión de texto a diagramas de flujo, el cual permite involucrar a los programadores inexpertos con el fin de mejorar sus habilidades de resolución de problemas. En el proceso de experimentación se observó que los estudiantes del grupo de experimentación obtuvieron un promedio superior en un 28% al grupo de control.

### 3.5. Modelado del conocimiento en el ámbito de las IES

Como se explicó en el Capítulo 1, el conocimiento representa uno de los factores más importantes dentro de una institución, ya que proporciona una ventaja competitiva. En la última década se han desarrollado varias propuestas tecnológicas para poder administrarlo de una manera eficiente. En el ámbito universitario específicamente, se han presentado varias alternativas de modelado ontológico, teniendo en cuenta varias relaciones entre los diferentes conceptos que componen la estructura de las IES. Entre las ventajas de emplear las ontologías en el contexto universitario está la posibilidad de reutilización e integración de la información de una manera más eficiente, facilitando el acceso oportuno y la recuperación de la misma.

Una propuesta de una plataforma e-learning fue presentada en 2004 por Guangzuo *et al.* [103], integrando conceptos como las ontologías, la computación ubicua, la Web Semántica y la computación en red. Esta arquitectura de aprendizaje electrónico, que lleva por nombre *OntoEdu*, tiene entre sus principales cualidades la adaptación del entorno acorde al perfil del estudiante, la composición automática del contenido, la facilidad de permitir adherir nuevas funciones educativas, y una ontología académica como núcleo del sistema. Se puede evidenciar que es una arquitectura que aprovecha las características de flexibilidad y escalabilidad de las ontologías para mejorar el proceso de enseñanza. Se trata en realidad de una red ontológica simple, constituida por una ontología de contenido y una ontología de actividades. Los resultados de los experimentos demostraron que *OntoEdu* era una arquitectura de plataforma educativa viable y flexible.

Una plataforma que permite gestionar la información referente a proyectos de docencia e investigación basada en ontologías es presentada por Taehee *et al.* [152] en 2006, resaltando la importancia que tiene la descripción de todos los atributos de los proyectos y cómo a partir de dicha información semántica se pueden realizar búsquedas y descubrimientos de nuevo conocimiento. Este enfoque permite abordar

el problema de clasificar los resultados de la búsqueda mediante palabras clave, al modelar la ontología como una red bayesiana de confianza.

Dada la relevancia del conocimiento en cualquier organización o institución, un trabajo relacionado con la explotación del conocimiento experiencial que puede ser empleado en varios dominios es propuesto por Sannin y Szczerbicki (2007) [225]. Estos autores plantean un sistema de conocimiento para la toma de decisiones en compañías e instituciones educativas, para lo cual emplean la ontología SOUPA [48] extendida y ADN decisional, permiten analizar de forma profunda la decisiones de los eventos de una organización.

Por otra parte, Laoufi et al. [149] en 2011, proponen el diseño de una memoria organizacional universitaria de enseñanza en línea mediante el uso de ontologías. Su propuesta ontológica no sólo se enfoca en la administración del conocimiento pedagógico (cursos, libros electrónicos, etc.), sino también en el conocimiento científico, técnico y administrativo del capital humano de la universidad. Un hecho muy interesante en la creación de la ontología es que todos los actores del ecosistema académico están capacitados para poder explicitar y adherir sus conocimientos, lo que promueve una motivación colaborativa para desarrollarlos. Esta metodología ha permitido en particular que los diversos actores experimenten la elicitación y una cultura colaborativa de creación e indexación de conocimientos. Adicionalmente, su arquitectura de memoria organizacional universitaria tiene la bondad de integrar diferentes ontologías y, además, permitir el razonamiento y la recuperación integral del conocimiento institucional de una manera inteligente.

Henning (2014), en su trabajo de investigación [110], propone un mecanismo para adecuar a las plataformas de aprendizaje masivo, con el fin de que sean más eficientes supliendo una mayor número de necesidades de aprendizaje. Su propuesta incluye un novedoso formato de metadatos, conjuntamente con un motor que interprete las anotaciones semánticas, al igual que las acciones del estudiante. En base a esta información, posteriormente puede generar recomendaciones didácticamente significativas que permiten establecer un conjunto de pasos, que guíen el camino de aprendizaje del estudiante de forma personalizada. Su objetivo final es preservar la libertad de elección de cada alumno para eliminar la rigidez observada en los entornos de aprendizaje masivos actuales.

El empleo de una ontología de enfermedades para describir los genes humanos en relación de las enfermedades fue propuesto por Yu et al. [272] en el año 2014. Esta investigación permite proporcionar los cálculos de similitud semántica entre los términos de la ontología de enfermedades y los genes que permiten a los biólogos explorar las similitudes de las enfermedades y las funciones de los genes en la perspectiva de la enfermedad. Su estudio incluye un enriquecimiento basado en el modelo hipergeométrico y de conjunto de genes. Esta investigación permitió a los biólogos verificar la relevancia de la enfermedad en un experimento biológico e identificar asociaciones de enfermedades relacionadas.

Abelló et al. describieron en su investigación [2] (2015) que el uso del proceso analítico en línea (OLAP) para el análisis de la información va mucho más lejos de las consultas a la información local, por lo cual, se debería consultar de fuentes externas produciendo un “*OLAP exploratorio*”. En este punto la Web Semántica toma un papel muy importante, ya que permite conocer la interpretación precisa de los datos al unificar conceptos para quitar la ambigüedad. Es así que en este trabajo los autores describen la relación entre *modelos multidimensionales* (MD) y

las tecnologías de la web semántica (SW), así como también el proceso de ETL (extracción, transformación y carga) con la finalidad de capturar la semántica pasiva, admitir la inferencia activa y promover un mejor razonamiento en los datos.

Otra experiencia innovadora en el empleo de tecnologías semánticas fue propuesta por Karan et al. [138] 2015, que demuestran como la web semántica juega el rol de una herramienta que permite la interoperabilidad entre los dominios de modelado de información de edificios (BIM) y los sistemas de información georeferenciados (GIS). Su enfoque se cimenta sobre tres componentes principales: la construcción de ontologías, integración semántica mediante formatos y estándares de datos interoperables, y el acceso a fuentes de información heterogéneas, proporcionando estándares para la interoperabilidad de información entre estos dos dominios, diferentes pero relacionados.

El trabajo desarrollado por Chung y Kim en 2016 [52] se apoya en el uso de las tecnologías semánticas para plantear el diseño de una ontología de aprendizaje que integra y conceptualiza las estructuras de conocimiento multinivel como son, el currículo, el programa de estudios, la materia de aprendizaje y los materiales. Este enfoque permite mejorar la facilidad de uso del currículo y el plan de estudios mediante el diseño y la implementación de ontologías de aprendizaje, generando de esta forma un método de integración y clasificación del programa de estudios basado en la definición del modelo semántico del programa institucional. Adicionalmente, esta propuesta facilita la secuenciación de conceptos adaptativos y el uso compartido de programas.

Boella en el 2016, mediante su proyecto de investigación propone el desarrollo del software *Eunomos*, como un avanzado documento legal y un sistema de gestión del conocimiento en leyes, mediante el empleo del denominado “*XML legislativo*” y ontologías legales [30]. Su propuesta toma como directriz los desafíos de la investigación legal en un entorno cada vez más complejo, multinivel y multilingüe. Desde esta perspectiva empleando la tecnología semántica *Eunomos* ayuda a los usuarios a disminuir la sobrecarga de información para conseguir la información legal que se requiere de una forma organizada y estructurada, permitiendo realizar un seguimiento del estado de la ley pertinente en cualquier tema consultado. Este sistema adicionalmente colabora con los investigadores y profesionales del dominio legal, gestionando información compleja e incorporando investigaciones de vanguardia en torno a material de informática jurídica.

Una aproximación a la gestión del conocimiento en procesos de diseño de plataformas colaborativas en gestión de la información fue presentada por Jurado y Bustamante [136] en 2017. Su propuesta se basa en el empleo de una ontología que permite especificar las relaciones y dependencias entre los patrones obtenidos a través de instrumentalizar un catálogo completo de patrones genéricos recopilados de diferentes plataformas colaborativas. En su trabajo resaltan los beneficios y oportunidades al contar con una ontología de diseño para plataformas colaborativas, pues permite mejorar la visibilidad de la plataforma y de sus servicios, la incorporación de elementos fundamentales, mejorar la accesibilidad y el manejo y la experiencia por parte del usuario entre otras.

Por su parte, Mora, Piedra y Tenesaca (2017) presentan un trabajo enfocado en la construcción de una red de ontologías relacionadas al ámbito de Educación Superior denominada “*Datos académicos vinculados*” (LAD) [180]. La finalidad de esta ontología es poder representar los planes de enseñanza de la institución, permitien-

do representar la estructura académica, y a la vez convirtiéndose en un modelo para la planificación de los cursos desde una perspectiva semántica. Emplean una metodología de trabajo para la creación de ontologías basada en los mejores criterios para la representación del conocimiento académico, logrando de esta manera crear una red ontológica de planificación docente. Su proyecto presenta la virtud de emplear recursos ontológicos y no ontológicos, promoviendo la creación de un nuevo vocabulario acorde al dominio académico.

En 2017, Vandenbussche et al. [257] presentaron un catálogo de esquemas ontológicos que ayudan a la difusión de vocabulario común a ser reutilizado en modelos formales. La plataforma adicionalmente ofrece métricas de la reutilización de axiomas entre esquemas para conocer el grado en que se están uniendo los diferentes esquemas. Otro servicio que dispone es la presentación de un almacenamiento versionado o histórico de los modelos, y la garantía de documentos bien formados dentro de sus repositorios. Las innovaciones más relevantes que se identificaron en este trabajo son:

- Una alta calidad de conjuntos de datos disponibles mediante múltiples métodos de acceso.
- El análisis por parte de expertos, exponiendo por primera vez las relaciones entre los vocabularios y su historial de versiones.
- La consideración de la propiedad semántica en términos de búsqueda de puntuación relevante.

Muñoz, López y Díaz también propusieron en 2018 un modelo de gestión del conocimiento para la enseñanza agrícola basado en ontologías. Este modelo ontológico de gestión de conocimiento describe los procesos y tecnologías inteligentes que soportan la enseñanza en el área agrícola desde la visión aprender haciendo y trabajo colaborativo en conjunto con las comunidades. Para ello, describen los elementos que definen el conocimiento de una unidad de producción agropecuaria, que les permita incorporar el *know-how* de la gestión de conocimiento y aprendizaje articulado con las TIC aplicado a la gestión educativa-productiva. Mediante este modelo se estructuran las bases tecnológicas y de conocimiento necesario para la enseñanza de la agricultura en un contexto universitario [184].

### 3.6. Sistemas de predicción y toma de decisiones en el ámbito académico

Las instituciones de Educación Superior son responsables de formar profesionales que generen un desarrollo económico y social de un país. Dentro de esta labor, usualmente han enfrentado problemas tales como la deserción y repitencia estudiantil, situaciones que impactan negativamente tanto a nivel personal, como de progreso del país. Con la finalidad de prevenir e intervenir en estas problemáticas, se han desarrollado un sin número de trabajos predictivos de rendimiento académico, con diferentes enfoques y considerando diversos factores. La mayoría de estas investigaciones se han desarrollado mediante la aplicación de diferentes técnicas de IA (minería de datos, Big Data y algoritmos genéricos, entre otros) sobre los datos generados durante el proceso académico de los estudiantes.

De principio, se puede destacar el estudio realizado por Osmanbegović y Suljić en el año 2012 [198], con una comparación entre tres algoritmos de aprendizaje supervisados (redes bayesianas, perceptrón multicapa y J48) aplicando minería de datos para identificar el algoritmo más eficiente en cuanto a la predicción del rendimiento de los estudiantes. Este estudio contó con la participación de 257 estudiantes, vinculados a la Facultad de Economía en Tuzla. A través de su estudio, se pudo constatar que las redes bayesianas alcanzaron una mejor ponderación de predicción con un 76.8% de exactitud frente a un 73.9% del algoritmo J48 y un 71.2% de las redes neuronales perceptrón multicapa. Además, lograron determinar que tanto la nota del examen de admisión, como el material de estudio y el número de horas a la semana de dedicación, son atributos trascendentales para la predicción de un eficiente rendimiento académico.

También en 2012, Romero et al. [215] realizaron un trabajo de investigación en la misma línea con la participación de 280 estudiantes universitarios. En su estudio emplearon el software WEKA como herramienta para el análisis de los datos. WEKA les brindó la posibilidad de aplicar y comparar los diferentes resultados obtenidos entre 10 diferentes algoritmos de predicción. Estos algoritmos fueron aplicados sobre un repositorio que contenía los resultados de estudiantes aprobados, reprobados y retirados de una promoción en particular. Como información relevante de su investigación se pudo apreciar que los mejores algoritmos para la predicción fueron Naive Bayes, PART y J48, con un porcentaje de asertividad del 65%. Adicionalmente, también se pudo constatar que independientemente del tipo de algoritmo que se empleó, los atributos que mayor significancia tenían para la predicción del resultado académico fueron: asistencia a clases teóricas y prácticas, nota de admisión, expectativas de aprobación, recursos y materiales de estudio, y nivel de integración con los compañeros. Como conclusión, los autores exponen que aunque lograron obtener información muy relevante, los porcentajes de predicción o clasificación continúan siendo muy bajos.

Con un esquema muy similar en el 2014, Ruby y David [219] realizaron una investigación sobre una población de 165 estudiantes, que incluía información personal del estudiante conjuntamente con su registro académico. La base de datos de los estudiantes estaba conformada por diversos atributos como: puntaje teórico, puntaje de laboratorios, promedio general, ingresos familiares, nivel de educación de los padres, actividades extracurriculares, etc. Para la selección de los atributos más relevantes emplearon técnicas de análisis estadístico como: chi cuadrado, ganancia de información, radio de ganancia, correlación y regresión. Mediante este análisis observaron que los atributos con alto impacto sobre el rendimiento eran: medios de estudio, cursos previos, promedios anteriores, actividades extracurriculares e ingresos familiares. Su propuesta se enfocó en usar estos atributos para realizar la clasificación y la predicción de rendimiento del estudiante usando la herramienta de minería de datos WEKA. Después de aplicar varios algoritmos de minería de datos, los resultados demostraron que el perceptrón multicapa mostró el mejor resultado de un 74.8% exactitud de predicción, seguida por ID3, que mostraba un 73%, NB-Tree and REPTree (algoritmos basados en árboles de clasificación), que mostraron una exactitud de 71%, y otros algoritmos de clasificación como J48, *SimpleCard* y *Decision Table*, que mostraron una exactitud de precisión de 68.8%, 69.5% y 68.1%, respectivamente.

Con la finalidad de brindar una información oportuna a la administración de las IES para identificar los estudiantes en riesgo de fracaso estudiantil, Mishra et al. en

2014 desarrollaron un modelo de predicción de rendimiento empleando diferentes algoritmos de clasificación sobre la base de la integración académica, social y emocional de los alumnos [177]. Este modelo de predicción empleó los algoritmos J48 y *Random Tree* sobre los registros de los estudiantes de una IES, para predecir el rendimiento del tercer semestre. El desempeño de los dos algoritmos fue satisfactorio; sin embargo, el algoritmo que presentó una mejor predicción de rendimiento fue el *Random Tree* con un 94.42%, frente a un 88.4% del algoritmo J48.

En el mismo año 2014, Natek y Zwilling realizaron un estudio para determinar los principales atributos relacionados al éxito académico, empleando dos diferentes tecnologías de minería de datos [186]. Para su trabajo de investigación emplearon los registros de tres períodos académicos 2011-2012 (42 estudiantes), 2012-2013 (32 estudiantes) y 2013-2014 (32 estudiantes), con una participación total de 106 estudiantes. Tras el análisis realizado empleando los algoritmos *RepTree* y J48, se pudo establecer que atributos tales como tipo de estudios, edad, empleo y tipos de estudios realizado por los estudiantes pueden considerarse como factor clave para la predicción del éxito de los estudiantes en su formación académica. Otra importante conclusión fue que el algoritmo *RepTree* fue menos sensible al tratar valores misteriosos o perdidos que J48, por lo que la precisión de la predicción en el conjunto de pruebas se realizó mejor. Sin embargo, en el conjunto de datos de entrenamiento se logró demostrar un rendimiento ligeramente menor de *RepTree* que la tasa de predicción de J48 (97% vs 98%).

Naser *et al.* en 2015 desarrollaron y entrenaron un modelo de red neuronal artificial [201] con la finalidad de predecir el rendimiento estudiantil. Este modelo predictivo fue implementado utilizando la topología de perceptrón multicapa, tomando como entrada diez características (créditos aprobados, promedio de rendimiento, género, entre otras) vinculadas con el rendimiento académico. Para el proceso de entrenamiento y validación de la red neuronal emplearon una base de datos con 150 registros de estudiantes, de los cuales 90 registros se emplearon en el proceso de entrenamiento, 45 registros para pruebas y los 10 registros y los restantes para el cruce de validación. Los resultados demostraron que el modelo era capaz de predecir correctamente el rendimiento de un estudiante con un 84.6% de seguridad.

En 2016, Amrieh, Hamrini y Aljariah realizaron un estudio en el que aplicaron una combinación de clasificadores base y métodos de conjunto de datos, para predecir la categoría final a la que pertenecería un estudiante según sus atributos (bajo, medio, alto) y como consecuencia mitigar el índice de fracaso [13]. Los clasificadores básicos empleados fueron árboles de decisión, redes neuronales artificiales y redes bayesianas, mientras que los métodos de conjunto abarcaron *Bagging*, *Boosting* y *Random Forest*. El sistema fue entrenado manejando 16 características de 480 registros de estudiantes. Posteriormente al entrenamiento del sistema, se realizó una experimentación con 25 estudiantes para validar su potencialidad de predicción. El sistema tuvo una exactitud del 80% al catalogar adecuadamente según el rendimiento real (bajo, medio o alto) de los estudiantes.

Por su parte, Alsheddy y Habib (2017) emplearon el clasificador J48 para generar un árbol de decisión que les permitiría predecir el rendimiento de los estudiantes, en función de las variables sociodemográficas y los resultados obtenidos en el primer año del programa de estudios [11]. El clasificador seleccionado en su investigación también fue empleado para reconocer qué cursos críticos se pueden emplear como indicadores para el rendimiento de los estudiantes. Por otra parte, hacen uso del algoritmo de *maximización de expectativa* (EM), como algoritmo de agrupamiento

para descubrir los atributos de los estudiantes retirados del programa de estudios. Su proceso de investigación contó con un conjunto de datos de 1980 estudiantes. Para el procesamiento de datos y aplicación de los diferentes algoritmos hicieron uso de la herramienta WEKA. Uno de los resultados revela que los estudiantes quienes han obtenido un rendimiento muy bajo en el primer año de programa de estudios tienen un 85.8% de probabilidades de retirarse.

Una propuesta también de 2017 fue realizada por Almarabeh a través de un estudio para analizar y evaluar el rendimiento de los estudiantes a través de la utilización de diversas técnicas de clasificación de minería de datos mediante el uso de WEKA [10]. Dentro de su investigación consigue determinar que la precisión de los algoritmos está en relación con el tamaño y naturaleza de los datos. Para su análisis empleó un conjunto de datos de 225 estudiantes, de los cuales consideró específicamente 10 atributos. Sobre este conjunto de datos aplicó cinco diferentes algoritmos clasificadores, empleando también diferentes medidas de rendimiento para poder contrastar los resultados. Como conclusión a su estudio expresó que el clasificador de red bayesiano es el que mejor precisión tuvo en la predicción del rendimiento con un 92% en comparación con los otros clasificadores: *NaiveBayes*, ID3, J48 y *Neural Network*, que alcanzaron un 91%, 88%, 91% y 90.2%, respectivamente.

En 2018, Son y Fugita plantearon un método denominado MANFIS-S empleando múltiples conjuntos de parámetros [233]. La concepción de conjuntos de parámetros múltiples les permitió aproximar el modelo con varios parámetros significativos para garantizar el adecuado rendimiento del sistema. Plantean la idea como una estrategia de aprendizaje empleando un entrenamiento global y uno local:

- El entrenamiento global es responsable de identificar, rectificar y lograr un subconjunto de parámetros significativos para el proceso de entrenamiento. Así, se enfoca en seleccionar un conjunto de parámetros al azar con los cuáles se entrenaron el sistema desde el primer hasta el último registro de su base datos.
- El entrenamiento local se plantea en función de dos parámetros (premisa y consecuencia), empleando la técnica del gradiente descendente y la técnica de optimización de enjambre de partículas.
- Finalmente, para poder categorizar un nuevo registro de un estudiante emplean *Fuzzy-K-nearest Neighbor* y así poder realizar la predicción del rendimiento académico.

El modelo MANFIS-S fue validado experimentalmente en comparación con ANFIS, MANFIS, OneR y Random Tree empleando un conjunto de datos del repositorio de la Universidad de Ciencias de VNU. El experimento demostró que MANFIS-S fue superior sobre los otros algoritmos en términos de precisión.

Con la finalidad de complementar el aporte tecnológico al ámbito académico, en 2019, Yang y Li [268] desarrollaron tres herramientas de análisis asociadas al ámbito académico:

- Una matriz de atributos de los estudiantes denominada SAM, con lo cual pueden realizar el análisis del rendimiento.
- Una herramienta de estimación del rendimiento mediante el uso de una red neuronal de propagación hacia atrás basada en la clasificación, que permite

estimar el rendimiento de los estudiantes en función del conocimiento previo, al igual que el rendimiento de otros estudiantes que poseen características similares.

- Unos indicadores de progreso de los estudiantes y predicados de relación causal basados también en una red neuronal de propagación hacia atrás, permitiendo identificar cuánto afectaría un indicador en el rendimiento del estudiante.

El proceso de experimentación lo realizaron Yang y Li con 60 estudiantes, donde pudieron valorar que los mecanismos empleados brindaban resultados correctos y precisos, y sobre todo permitían conocer el progreso del estudiante.

### 3.7. Apoyos para la formación de grupos de aprendizaje colaborativos

Dillenbourg y Schneider [75] definen el aprendizaje colaborativo como “*la resolución de problemas en la que dos o más estudiantes construyen de forma interactiva una solución coordinada a dicho problema, y son capaces de discernir en la tarea colaborativa las actividades en las que la tarea se divide y de manera poder resuelta de forma independiente por los individuos del grupo*”. Conforme a Kagan [137], existen cuatro principios básicos que determinan el aprendizaje cooperativo:

- La **interdependencia positiva** implica que un beneficio para un estudiante se asocia con el beneficio para el grupo completo. Como tal, para un aprendizaje colaborativo efectivo, es relevante que los estudiantes comprendan que el éxito de cada miembro del equipo depende del éxito de cada uno de los otros miembros, y que si uno falla, todos lo hacen [137].
- La **responsabilidad individual** hace alusión a que cada estudiante del grupo es responsable de su propio aprendizaje y de contribuir al aprendizaje de los compañeros.
- La **participación igualitaria** estipula que a ningún estudiante se le debe autorizar dominar a un grupo, desde ninguna perspectiva social o académica, y a su vez a ningún estudiante se le debe permitir estar inactivo, o interferir en el trabajo de los otros estudiantes del grupo. La participación igualitaria no se realiza automáticamente y se deben emplear las medidas adecuadas para garantizar que ocurra, tales como la asignación de turnos o la división del trabajo.
- Por último, la **interacción simultánea** es el resultado de seleccionar y organizar a los estudiantes en grupos pequeños, colocarlos cara a cara y establecer una tarea grupal de tal forma que todos los estudiantes del grupo precisen trabajar fusionados para conseguir una solución. Esto difiere a un contexto de clases tradicional en el que todos los estudiantes observan al frente, trabajan de manera independiente y pasan la mayor parte del tiempo inactivos sentados en silencio, atendiendo a la exposición del docente.

Las peculiaridades del aprendizaje colaborativo están avaladas por un conjunto de investigaciones científicas y se fundamentan en las teorías de los psicólogos educativos. Por lo tanto, el aprendizaje colaborativo está amparado por teorías tanto cognitivas [100] como no cognitivas [195]. Se destaca, por ejemplo, que permite a los estudiantes tomar decisiones en el proceso de aprendizaje, contribuye a minimizar la competencia o rivalidad recia, ayuda a que las tareas de aprendizaje sean placenteras, e inclusive no obliga a todos los estudiantes a progresar al mismo ritmo. También se indica que estimula la ejecución de las teorías del aprendizaje social, al brindar la oportunidad de que los estudiantes piensen en voz alta y exploren sus emociones en grupos pequeños de discusión. Así, según Grusec [102], los estudiantes tienden a aprender de la interacción entre unos y otros, incluyendo conceptos como el aprendizaje observacional, el modelado y la imitación.

Johnson y Johnson [132] indicaron que pese a que los métodos de enseñanza-aprendizaje con estructuras individualistas y competitivos poseen indiscutiblemente su lugar en el programa de formación, estos tienen que ser equilibrados a través del aprendizaje colaborativo. Numerosos académicos de diferentes orígenes y disciplinas, han validado sus planteamientos en estudios en diferentes contextos y entornos con todo tipo de estudiantes, con parámetros de edad, sexo, nacionalidad y antecedentes culturales diversos. Por ejemplo, Costaguta [59], después de revisar 20 estudios relacionados con el aprendizaje colaborativo, declaró que había producido efectos positivos significativos en el 65% de los casos, y que solo dos estudios informaron un mayor rendimiento para el grupo de comparación. Lie [158] indicó que las experiencias de aprendizaje colaborativo promovían un mayor rendimiento y productividad en comparación con el aprendizaje competitivo o individualizado, mientras que Slavin [231] demostró que con el uso del aprendizaje colaborativo se puede lograr un incremento sustancial en relación a parámetros tales como el rendimiento estudiantil, la aceptación de estudiantes con discapacidades académicas, las relaciones étnicas y la autoestima. En el área concreta del aprendizaje de idiomas, Kuo [147] y Froman [92] confirmaron que los grupos colaborativos elevan las oportunidades para que los estudiantes produzcan y comprendan el lenguaje, a través de los ejercicios y comentarios de sus compañeros. Otras muchas evidencias de la mayor satisfacción por parte de docentes y estudiantes se pueden encontrar en [8, 73, 77, 139, 153, 193, 274].

Con todo, aunque las bondades del aprendizaje colaborativo estén bien estipuladas, las mejoras están condicionadas por la forma en la que los estudiantes del grupo trabajan juntos [59, 144, 181, 234]. De hecho, Johnson [133] manifestó que muchos de los resultados improductivos del trabajo grupal se deben a la inadecuada asignación de los miembros en la fase de composición del grupo. En este sentido, aunque no existe un único método perfecto de asignar a los estudiantes a un grupo, sí existen varias propuestas de uso habitual, empleando diferentes criterios como el nivel de desempeño, el género, las características de la familia y el origen escolar del estudiante, el origen étnico, las motivaciones, las actitudes, los intereses, la personalidad, el estilo de aprendizaje o la capacidad de liderazgo [41, 60]. Es en este punto donde se considera que la Inteligencia Artificial puede proporcionar las mejores ayudas, exprimiendo todo el conocimiento disponible sobre toda esa cantidad de parámetros y analizando sus interdependencias [8, 153]. A continuación se recoge un análisis de las más relevantes investigaciones llevadas a cabo en esta línea hasta la fecha. En cada una las investigaciones se describe el criterio empleado para la generación de grupos, así como la forma en que se aplica:

- En el trabajo desarrollado por Wang [259] en el año 2007, el autor propone un sistema de agrupación denominado DIANA que emplea la lógica de los algoritmos genéticos para evitar la creación de grupos excesivamente débiles. El algoritmo emplea como criterio de agrupación el estilo de pensamiento de cada estudiante (legislativo, ejecutivo o judicial) propuesto por Sternberg (1998). Su algoritmo crea grupos heterogéneos de estudiantes cuyos estilos de pensamiento fuesen lo más diferentes posibles. El experimento contó con la participación de 66 estudiantes, quienes conformaron grupos de tres. 10 grupos fueron creados al azar y los restantes 12 grupos empleando el sistema. Los resultados demostraron que los grupos generados por el algoritmo fueron más eficientes en completar la actividad a un nivel estadísticamente más significativo (61.4% frente a un 41.8%). Adicionalmente, la diferencia en cuanto al rendimiento intergrupar fue más pequeño, demostrando un mejor balance y equilibrio académico a nivel inter- e intragrupal.
- Otro tipo de algoritmo fue desarrollado por Christodoulopoulos y Papanikolaou en 2007, quienes exponen una herramienta de formación de grupos basada en la web que facilita la generación de grupos homogéneos y heterogéneos apoyado en los estilos de aprendizaje de los estudiantes [51]. La particularidad de su enfoque consiste en facilitar al docente modificar manualmente los grupos; además, también les permite a los estudiantes negociar la agrupación. La generación de los grupos se basa en el algoritmo *Fuzzy C-Means* y la creación de grupos heterogéneos se basa en un algoritmo de selección aleatoria. La herramienta proporciona una opción para negociar los equipos propuestos con los estudiantes. Sin embargo, para la negociación es obligatoria la interacción directa con el docente.
- Una investigación centrada en los niveles de comprensión e intereses de los estudiantes fue formulado por Lin et al [160] en 2010, para automatizar la composición de los grupos de aprendizaje. Su aporte principal fue el desarrollo de un algoritmo denominado *optimización de enjambre de partículas mejorado* (EP-SO). Este algoritmo permite generar eficientes grupos basados en dos criterios: información acerca del nivel de entendimiento y el interés de los estudiantes. Su proceso de experimentación se basó en conjuntos de datos autogenerados para la simulación, con los cuales se simulaba la formación de grupos de 5 estudiantes, tanto los niveles de comprensión y de interés sobre el tema fueron generados al azar, previo a la aplicación del algoritmo para la creación de los grupos.
- Soportado por la investigación de Belbin [20], quien expuso que la formación adecuada de los equipos de trabajo se realiza en función de los roles de los miembros que conforman el grupo, Yannibelli y Amandi [269] en el 2012 propusieron emplear los roles de los estudiantes como criterio de agrupación, para crear grupos de aprendizaje bien equilibrados de acuerdo con los roles de sus miembros. Para ello, desarrollaron un algoritmo evolutivo de aglomeraciones determinista que cumplía con este objetivo. El algoritmo era capaz de diseñar diferentes alternativas para la conformación de los grupos, y de evaluar cada alternativa acorde al criterio mencionado. La evaluación se daba sobre el conocimiento de los roles de los estudiantes. Para validar su propuesta, realizaron experimentos computacionales en 10 conjuntos de datos, con diferentes niveles de complejidad, obteniendo resultados óptimos en los 4

primeros conjunto de datos, y resultados casi óptimos en los 6 conjuntos de datos siguientes.

- Un enfoque sinérgico para la formación de los grupos colaborativos fue propuesto por Cocea y Magoulis [53] en 2012, empleando la metodología del razonamiento basado en casos para modelar el comportamiento del estudiante, y la agrupación como técnica de minería de datos. Los autores experimentaron la propuesta empleando un entorno de aprendizaje denominado *eXpresser* empleado para el ámbito de generalización de las matemáticas. Con el razonamiento basado en casos pudieron identificar las estrategias de solución aplicadas por cada estudiante, en tanto que el mecanismo de agrupación les permitió detectar qué estudiantes aplicaron similares estrategias. Mediante el uso de métricas heterogéneas y coeficientes de semejanza se pudo medir la similitud entre los comportamientos de los estudiantes. Esta información fue útil para informar al docente cómo constituir los grupos según su comportamiento y estrategias de solución.
- Con el afán de conformar inter-homogéneos e intra-heterogéneos grupos eficientes de aprendizaje, Moreno desarrolló en 2012 una solución con un algoritmo genético [181]. La particularidad de su algoritmo fue que permitía considerar tantas características de los estudiantes como se desee, planteando el problema de agrupación desde un enfoque de optimización multiobjetivo. El experimento se realizó con 135 estudiantes, y optó por emplear tres características: nivel de conocimiento, estimación de las habilidades comunicativas y las habilidades de liderazgo. En base a estos atributos se crearon 11 grupos empleando el método propuesto, 7 grupos generados aleatoriamente y 11 grupos conformados por autoorganización. Su experimento no solo demostró que el método propuesto alcanzaba el objetivo de crear grupos inter-homogéneos e intra-heterogéneos con mejor rendimiento, sino que también las características consideradas promovieron de manera positiva el desarrollo de las actividades.
- Zheng, en 2013, propuso un método de composición dinámica iterativa para perfeccionar el proceso de formación de los grupos de aprendizaje [275]. Su metodología se centra en analizar históricos de los trabajos colaborativos empleando técnicas de minería de datos, para conocer la evaluación otorgada por el docente, y el registro de las interacciones de los miembros del grupo. Su método primero obtiene los perfiles de los estudiantes mediante la aplicación de cuestionario recabando la información asociada a los atributos de liderazgo, conocimiento previo, interés por la asignatura, actitud de trabajo grupal, autoconfianza, timidez y género. Durante el proceso de aprendizaje colaborativo se registran todas las interacciones intragrupales, así como las valoraciones obtenidas. Al finalizar la actividad colaborativa se analizan los registros generados mediante tecnologías de minería de datos, identificando qué atributos de los perfiles de estudiantes que conformaron el grupo obtuvieron un mejor resultado. Finalmente, a través de la utilización del mecanismo de árboles de decisión identifican las características de aquellos grupos que tuvieron un buen rendimiento, y las adhieren como reglas en el proceso de formación para los siguientes grupos.
- Por otra parte, Tien, Lin, Chang y Chu (2013) sugieren una estrategia de agrupación que considera múltiples características de los estudiantes empleando

nuevamente un algoritmo genético, el cual permite constituir grupos balanceados de aprendizaje colaborativo inter-homogéneos e intra-heterogéneos [246]. Para ello, cada característica o atributo considerado de un estudiante es representado por un valor numérico, y de forma independiente cada característica posee su propio rango de valores. Estos autores para validar su propuesta emplean conjuntos de datos auto-generados de entrenamiento. Los resultados experimentales de su simulación demuestran que el método de agrupamiento propuesto es eficaz, eficiente y robusto, contrastado con los métodos de conformación al azar o auto-agrupación.

- También en 2013, una propuesta apoyada en la integración de políticas asistida por computadora que facilita la generación automática de equipos cuasi-óptimos fue presentada por Alberola *et al.* [7], basados en inteligencia colectiva, generación de estructuras de coalición y aprendizaje bayesiano. En su propuesta, después de cada actividad colaborativa, los estudiantes clasificaban a sus compañeros según los roles de la taxonomía de Belbin [21]. Posteriormente, esta información era utilizada por el algoritmo de creación de la estructura de coalición para calcular la siguiente estructura de equipo. Las simulaciones permitieron observar que, mientras los estudiantes no tuviesen grandes dificultades para clasificar a sus pares, la política era capaz con pocas interacciones de mejorar la calidad de las estructuras del equipo, convergiendo gradualmente a la solución óptima.
- Desde otra perspectiva, Balmaceda *et al.* [17] en 2014 propusieron un sistema inteligente que modela la formación de grupos como un problema de satisfacción de restricciones ponderadas (“*Weighted Constraint Satisfaction Problem*”, WSCP, por sus siglas en inglés). Emplean un asistente que considera tres características de los estudiantes: estilos psicológicos, roles de equipo y redes sociales. La ventaja de modelar la formación de los grupos como un WSCP es que consiguen integrar y explorar de forma natural las restricciones y preferencias a nivel de estudiantes y grupos. La finalidad del asistente es generar grupos de aprendizaje basados en el equilibrio de roles del equipo y la distribución de los estilos psicológicos. Para su propuesta en cuanto a los roles de equipo considera el modelo de los 8 roles propuestos por Mumma [183], en tanto que para identificar los estilos psicológicos emplearon el cuestionario *Myers-Briggs Type Indicator* (MBTI)<sup>1</sup>.
- Bergey y King, en 2014, presentaron una herramienta de apoyo para generar grupos de estudiantes heterogéneos con la máxima diversidad posible [24]. Su herramienta emplea algoritmos genéticos con criterios de agrupación que incluyen atributos como experiencia laboral, rasgos de personalidad, demografía, género, título y rendimiento académico. Esta herramienta fue utilizada en un contexto universitario, y se pudo demostrar que el rendimiento de los grupos óptimamente equilibrados fue superior que los creados manualmente.
- Por su parte, Duque *et al.* [77] propusieron en 2015 un método que analiza la colaboración y la interacción entre los alumnos utilizando un conjunto de indicadores o variables sobre cómo resuelven las tareas académicas. Para ello, aplica el concepto de “*profundidad de datos*”, que utiliza como una medida de la proximidad de los valores de los indicadores de análisis para un estudiante

---

<sup>1</sup><http://www.personalitypathways.com>

con respecto a los valores que los mismos indicadores tienen para los demás estudiantes. Su propuesta ofrece la ventaja de combinar dos perspectivas diferentes en el proceso de formación de nuevos grupos. En primera instancia, es factible crear grupos heterogéneos con relación a algunas habilidades o actitudes de los estudiantes donde poseen diferentes valores en ciertos indicadores. En segunda instancia, es factible crear grupos homogéneos aplicando simultáneamente un criterio en el que los estudiantes consiguieron valores similares en otros indicadores.

- Otra investigación asociada a la adecuada conformación de grupos que hace uso de un algoritmo genético fue desarrollada por Pininghof *et al.* [207] en 2015. Para ello, establecen como parámetros o variables del algoritmo las características personales asociadas a las inteligencias múltiples, los estilos de aprendizaje, liderazgo y la asertividad de cada estudiante. Utilizan el operador de cruce y de mutación como operadores de la información genética, en tanto que para el mecanismo de supervivencia emplean la estrategia del elitismo. Como resultado, los autores destacan que el éxito de este enfoque se fundamenta en la importancia de identificar las características adecuadas personales junto con expertos de diferentes disciplinas. Esta propuesta se cimentó en la interacción multidisciplinaria de psicólogos, expertos en educación y profesionales en desarrollo de software.
- Adicionalmente, en el mismo año 2016, Kardan *et al.* presentaron un modelo matemático eficaz para resolver un problema de optimización especial [139]. En dicho modelo se desarrolla una función objetivo y se establece que el conjunto de restricciones sean lineales, para lo cual las variables consideradas sólo pueden tomar los valores de 0 y 1. Su propuesta de formación de grupos mediante el enfoque de programación de enteros binarios se resuelve en un tiempo aceptable. Este enfoque fue comparado con el algoritmo *squeaky wheel* y el algoritmo de agrupamiento de *K-means*, siendo más exitoso en términos de generar los mejores resultados, satisfacción y rendimiento de los participantes.
- Por su parte, Srba y Bielikova (2015) describen un método novedoso basado en el enfoque de tecnología grupal (*Group Technology* en inglés) [234], conformando grupos de estudiantes compatibles en base a sus características individuales, e incorporando un proceso de retroalimentación al finalizar la actividad colaborativa. Esta información permite mejorar iterativamente la formación de los siguientes grupos. La aplicación del método se llevó a cabo mediante de la integración con la plataforma PopCorn, que proporciona a los estudiantes de los grupos creados un conjunto de herramientas de colaboración en tiempo real. Los resultados de este trabajo de investigación permiten concluir que la incorporación de los comentarios de la colaboración de los estudiantes permite alcanzar una mejor calidad en el aprendizaje colaborativo e iterativamente ir ajustando los parámetros de entrada para mejorar el soporte de la creación de los siguientes grupos.
- En el trabajo presentado por Lescano *et al.* [153] en 2016, los autores proponen el empleo de un algoritmo genético para la formación automática de los grupos de aprendizaje fundamentado en los estilos de aprendizaje de sus miembros. Su algoritmo es capaz de diseñar diferentes alternativas con su respectiva predicción de rendimiento, para lo cual emplean reglas de asociación

obtenidas mediante un proceso de minería de datos del historial académico. En su proceso de experimentación lograron demostrar que el operador de mutación por intercambio produjo mejores resultados asociados al promedio académico y un menor valor para la desviación estándar.

- Sadegui *et al.* [221] en el mismo 2016 desarrollaron un modelo matemático que se basa en los intereses de los estudiantes asociados con un área de conocimiento. Desarrolla una matriz en la cual los estudiantes exponen una valoración sobre las diversas áreas de interés. Posterior a ello, mediante una programación binaria generan grupos de estudiantes en base a dos restricciones: restricciones de igualdad y restricciones de desigualdad, que ayudan a identificar en qué área tienen mayor o menor interés, para finalmente poder conformar los grupos de aprendizaje asociados por su área de interés.
- Una última herramienta destacada fue presentada por Alberola *et al.* [8] en 2016. Al igual que otras anteriores, esta herramienta crea los grupos en base a la teoría de los roles de Belbin [20]. Los roles son deducidos de la retroalimentación de los estudiantes con respecto a sus compañeros al finalizar cada actividad. Emplean la técnica bayesiana para manejar la incertidumbre referente a los roles de los estudiantes, y el enfoque de generación de estructuras de coalición, que es compatible con la programación lineal para calcular la configuración óptima de cada equipo. Tras el proceso de experimentación se pudo constatar que el empleo de la herramienta permitió mejorar la dinámica de los equipos y la satisfacción de los estudiantes.

En conjunto, a pesar del rol crítico que tiene la creación de los grupos en términos de apuntalar la efectividad del aprendizaje colaborativo, la consideración tanto del rendimiento académico como de los rasgos de personalidad por medio de ayudas tecnológicas no se ha avanzado completamente en términos prácticos. Es necesario resaltar que las relaciones y las dependencias entre los atributos de la personalidad y el desempeño son muy complejas e implican dimensiones no estructuradas y de razonamiento incierto. Todo esto elementos hacen que el poder establecer un procedimiento eficiente para la formación de grupos de aprendizaje colaborativo sea muy desafiante e incluso difícil de modelar utilizando métodos convencionales basados en la inferencia y el razonamiento exactos.

**Parte III**

**Contribuciones**



## Capítulo 4

# Aportes a la gestión de conocimiento semántico en instituciones de Educación Superior

*Aunque existen varias propuestas de administración del conocimiento de las instituciones de Educación Superior, ninguna divide y explicita el desarrollo que los actores del ecosistema toman a lo largo del tiempo. Nuestra propuesta en este capítulo se basa en construir una representación semántica del ecosistema académico, mediante la implementación de una red ontológica del mismo basada en la metodología NeOn y la reutilización de elementos de propuestas anteriores. Esta red ontológica se ha utilizado como un instrumento para apoyar al cuerpo académico y administrativo de la Universidad Politécnica Salesiana, sentando asimismo las bases para el desarrollo de los demás aportes de esta tesis gracias a las capacidades de interoperabilidad y centralización de la información.*

### 4.1. Introducción

Como se explicó en los capítulos anteriores, las instituciones de Educación Superior generan periódicamente grandes volúmenes de datos asociados al ámbito de académico, que a menudo se gestionan de manera subóptima. Así, es habitual que la información se distribuya entre varios subsistemas, con diferentes bases de datos y formatos no estructurados adecuadamente, que limitan las posibilidades de reutilizar información más allá del propósito con el que fue capturada. Adicionalmente, el manejo de múltiples modelos de entidad-relación implica que los datos en cada uno de esos repositorios deben ser interpretados para derivar su significado, generalmente dando lugar a ambigüedades. Esta particularidad limita la capacidad de administrar y explotar todo el potencial de la información, en forma de conocimiento que pudiese ser utilizado de manera oportuna y adecuada por los diferentes actores del ecosistema académico (alumnos, docentes, personal administrativo y autoridades).

Si bien es cierto que en años recientes se han propuesto varias soluciones de gestión del conocimiento en el ámbito de la Educación Superior (e.g. sistemas transaccionales que permiten asociar la información entre varios subsistemas), frecuentemente la información recuperada no contiene relaciones semánticas, que posibilite a dichos subsistemas reconocer incluso si dos fragmentos de información están relacionadas con el mismo tema o entidad. Asimismo, las propuestas de modelado semántico no han alcanzado a explicitar las acciones que toman los diferentes actores del ecosistema a lo largo del tiempo, ni han integrado 4 perfiles que consideramos de gran importancia en la caracterización de los mismos:

- El **perfil personal** ha de capturar información de identificación y contacto básica, incluyendo cédula de identidad, domicilio, teléfono, correo electrónico, fecha y lugar de nacimiento, etc.
- El **perfil académico** ha de modelar las competencias adquiridas y el rendimiento académico de un estudiante a lo largo de un proceso educativo, al igual que las especialidades y las métricas de calidad docente alcanzadas por un profesor.
- El **perfil psicológico** ha de caracterizar de manera precisa la personalidad de cada estudiante o docente.
- El **perfil socio-económico** ha de caracterizar la situación social y económica de un estudiante o docente, atendiendo a tantos factores como la institución estime relevantes.

En respuesta a esas carencias, en este capítulo presentamos una nueva propuesta en forma de red ontológica, que nos permite garantizar una gestión del conocimiento más eficiente y con mejores capacidades expresivas, así como de interoperabilidad, búsqueda y reutilización de la información. Dicha red ontológica, que denominaremos por su acrónimo OEA (de “*Ontología del Ecosistema Académico*”) permite enlazar la información producida y conducirla a un nivel superior, de modo que pueda llegar a las personas adecuadas, en el momento adecuado, para facilitar labores de análisis y toma de decisiones como las que se presentarán en el Capítulo 5.

Como quedó dicho en el Capítulo 2, el mayor desafío en la construcción de ontologías consiste en la organización de los conceptos para permitir a las personas mapearlos con símbolos en la computadora, así como para permitir a los sistemas informáticos inferir nuevo conocimiento útil de los hechos establecidos. Con el fin de representar las múltiples conceptualizaciones que surgen a partir de los procesos académicos de las IES y unificar la información dispersa en varios repositorios, recurrimos a la metodología NeOn para llevar a cabo la elicitación de requerimientos y dirigir la reutilización de recursos ontológicos, el alineamiento de componentes y la creación de nuevo recursos. El proceso seguido se describe en las dos secciones, para luego dar paso a detalles sobre la migración a la nueva red ontológica de las bases de datos manejadas previamente por la Universidad Politécnica Salesiana.

## 4.2. Elicitación de requerimientos

La primera etapa especificada en la metodología NeOn tiene como finalidad establecer los usos previstos, los usuarios finales y los requerimientos que se deben

cumplir en la construcción de la ontología. En este proceso intervinieron dos tipos de actores: expertos en el dominio académico (responsables de definir las pautas y estrategias para el diseño de los programas analíticos y los procesos académicos) e ingenieros ontológicos, con experticia en el desarrollo de modelos formales y enriquecimiento semántico de datos.

En esta etapa se exigen una serie de procesos: identificación de requerimientos, agrupamiento de requerimientos funcionales, validación de requerimientos, extracción de terminología, etc. A continuación presentamos la información relacionada en diferentes tablas informativas. Para empezar, la Tabla 4.1 resume la actividad de identificación del propósito, alcance y lenguaje de implementación, además de establecer como usuarios finales involucrados al personal administrativo, a los docentes y a los directivos de la parte académica de la institución; en la parte final se enumeran los usos previstos.

Tabla 4.1: Sección 1 perteneciente al documento de elicitación de requerimientos.

<b>1</b>	<b>Propósito</b>
Desarrollar una red ontológica del ecosistema universitario que permita la interoperabilidad de la información alrededor del ecosistema académico de la Universidad Politécnica Salesiana y sirva de modelo para formalizar e instrumentalizar un repositorio semántico como una base sólida de conocimiento.	
<b>2</b>	<b>Alcance</b>
La ontología debe centrarse en el dominio del ecosistema académico de la Universidad Politécnica Salesiana y en las nomenclaturas nacionales e internacionales de instituciones de Educación Superior.	
<b>3</b>	<b>Lenguaje de implementación</b>
Se ha seleccionado OWL por ser el lenguaje ontológico más empleado en la Web Semántica.	
<b>4</b>	<b>Usuarios finales previstos</b>
<b>Usuario 1.</b> Expertos del dominio que desempeñan cargos estratégicos en las IES.	
<b>Usuario 2.</b> Docentes que participan de los procesos académicos de las IES.	
<b>5</b>	<b>Usos previstos</b>
<b>Uso 1.</b> Generar perfiles de los actores inmersos en los procesos académicos. <b>Uso 2.</b> Buscar características producidas en la interacción académica entre estudiantes y docentes del ecosistema universitario que influye en el rendimiento del trabajo colaborativo. <b>Uso 3.</b> Medir el desempeño académico de los estudiantes de la Universidad Politécnica Salesiana en sus asignaturas. <b>Uso 4.</b> Permitir el análisis predictivo del rendimiento académico. <b>Uso 5.</b> Pre-clasificar características de los estudiantes que van a ser tomadas como variables en un algoritmo genético para la generación dinámica de grupos óptimos de trabajo. <b>Uso 6.</b> Facilitar la asignación de asignaturas a docentes en función de sus perfiles.	

Con la intención de captar los requerimientos se empleó el instrumento de las denominadas *preguntas de competencia* [239] como se presenta en la Tabla 4.2, en donde se identifican los dominios de preguntas según el tipo de usuario que se entrevistó. Dentro de los requerimientos no funcionales se definió el uso de estructuras comúnmente utilizadas por los profesores, concretamente documentos PDF para los sílabos y los programas analíticos de materias. Cada uno de los entrevistados fue grabado para recolectar la información y esta fue transcrita para el análisis posterior.

Tabla 4.2: Sección 2 perteneciente al documento de elicitación de requerimientos.

6	Requerimientos de la ontología
a)	<b>Requerimientos no funcionales</b>
	<b>RNF1.</b> La ontología debe estar basada en la terminología utilizada en la plantilla de los sílabos de los Docentes. <b>RNF2.</b> La ontología debe estar basada en la terminología utilizada en la plantilla de los programas analíticos de las asignaturas.
b)	<b>Requerimientos funcionales - Preguntas de competencias</b>
	Grupo 1 (Usuario 1 y Usuario 2)
	<b>PC1.</b> ¿Quiénes son los actores del ecosistema universitario que influye en el rendimiento del trabajo colaborativo de los estudiantes? <b>PC2.</b> ¿Cuáles son los factores académicos con los que un estudiante interactúa al estudiar una materia? <b>PC3.</b> ¿Cuáles son las modalidades de oferta de una carrera? <b>PC4.</b> ¿Cuáles son los factores académicos que varían en la modalidad en la que se matricula un estudiante en una materia? <b>PC6.</b> ¿Cuáles son los factores de una asignatura que comprueban el aprendizaje de un alumno? <b>PC7.</b> ¿Qué es una metodología de enseñanza? <b>PC8.</b> ¿Cómo identificar qué alumnos pueden trabajar mejor en grupo para disminuir la deserción?
	Grupo 2 (Usuario 1)
	<b>PC9.</b> ¿Cuál ha sido la evolución de los planes analíticos a lo largo de los periodos académicos? <b>PC10.</b> ¿Cuál es la estructura de los programas analíticos de las mallas anteriores al periodo 49? <b>PC11.</b> ¿Cuál es la estructura de los programas analíticos de las mallas siguientes al periodo 48? <b>PC12.</b> ¿La universidad maneja nomenclaturas específicas que siguen algún estándar internacional o nacional para crear los programas analíticos? <b>PC13.</b> ¿Qué relación tiene el sílabo con la malla curricular? <b>PC14.</b> ¿Cuál es la estructura en la que se oferta una carrera, sus menciones, la malla curricular y sus asignaturas? <b>PC15.</b> ¿Qué es el sílabo? <b>PC16.</b> ¿Para qué sirve el sílabo? <b>PC17.</b> ¿Qué es el programa analítico de la asignatura? <b>PC18.</b> ¿Para qué sirve el programa analítico de la asignatura? <b>PC19.</b> ¿El sílabo y las calificaciones de los estudiantes se relacionan? <b>PC20.</b> ¿Qué es una unidad de organización curricular?

6	Requerimientos de la ontología
Grupo 2 (Usuario 1)	
PC21.	¿Qué son los campos de formación?
PC22.	¿Los contenidos procedimentales son habilidades?
PC23.	¿Las habilidades están normadas?
PC24.	¿El contenido actitudinal son igual que los valores?
PC25.	¿Los valores están normados?

Para el análisis de las respuestas a las preguntas de competencias de los actores se realizó un conteo de frecuencia de palabras con un filtro que eliminaba todo lo que no fuesen sustantivos, verbos y adjetivos, por medio de la herramienta *linguakit*<sup>1</sup>. También se realizó un agrupamiento de términos con el fin de llegar a la trazabilidad del pre-glosario de términos que muestra la Tabla 4.3.

Tabla 4.3: Muestra de los términos de grupos más importantes de la Sección 3 perteneciente al documento de elicitación de requerimientos.

7	Pre-glosario de términos	
(a)	Términos principales respondidas en las preguntas de competencia	
TG1. (Álgebra, Álgebra lineal, Asignatura, Cátedra, Materia)	PC22, PC12, PC1, PC9, PC14, PC20, PC21, PC28, PC6, PC3, PC23	
TG2. (Alumno, Estudiante)	PC7, PC9, PC1, PC3, PC4, PC6, PC16, PC20, PC21, PC22	
TG3. (Aprendizaje, Resultado, Resultado de aprendizaje)	PC4, PC12, PC13, PC2, PC5, PC6, PC9, PC23, PC24, PC26	
TG4. (Aprovechamiento, Registro de calificación, Matrícula)	PC4, PC19	
TG5. (Área básica, Básico)	PC21, PC9	
TG6. (Área curricular, Cambio de matriz productiva)	PC1, PC16, PC2	
TG7. (Área de especialización, Campo, Campo de formación)	PC9, PC27, PC21	
TG8. (ESPE, San Francisco, Universidad, UPS)	PC11, PC3, PC8, PC9, PC12, PC15, PC22, PC25	
TG9. (Área de titulación, Carrera, Carrera de administración, Carrera de computación, Carrera de mecánica, Ciencias de la computación, Computación, Comunicación social, Contabilidad)	PC9, PC1, PC2, PC11, PC12, PC14, PC16, PC22, PC25, PC21	
TG10. (Malla, Malla anterior, Malla antigua, Malla curricular, Modelo curricular, Plan curricular, Plan de carrera, Proyecto curricular)	PC13, PC14, PC3, PC1, PC9, PC25, PC11	

<sup>1</sup><https://linguakit.com/>

7 Pre-glosario de términos	
(a)	Términos principales respondidas en las preguntas de competencia
TG11. (Evaluación, Evaluación de aprovechamiento, Proceso de evaluación)	PC2, PC9, PC16, PC19, PC21, PC25, PC4
TG12( Hora, Hora clase, Hora docencia, Horas de experimentación, Horas para estudiante)	PC17, PC23, PC25, PC22, PC9
TG13. (Docente, Pedagogo, Profesor)	PC1, PC3, PC6, PC8, PC9,PC4, PC7, PC12, PC15,PC16, PC19, PC21, PC22
TG14. (Programa académico, Programa analítico)	PC28, PC3, PC11, PC13,PC17, PC25, PC18

Como resultado de estas actividades se concluye con una matriz de requerimientos, de la cual se muestra un extracto en la Tabla 4.4. La primera columna representa el **término de grupo** asignado; en la segunda columna aparecen los **términos asociados** a ese significado y en la tercera se indica el código de pregunta de competencia al cual se asoció el término:

Tabla 4.4: Extracto de los requerimientos funcionales.

Término de Grupo	Términos asociados	Preguntas de competencia
TG4	Docente / Actividad	PC1, PC3, PC6
TG5	Estudiante / Alumno	PC7, PC9, PC1, PC3, PC4, PC6, PC16, PC20, PC21, PC22
TG12	Área de conocimiento / Área de estudio	PC12, PC16, PC24, PC14
TG41	Profesor / Pedagogo	PC1, PC3, PC6
TG52	Graduado	PC27
TG73	Plan curricular / Modelo curricular / Proyecto curricular	PC13, PC14, PC3, PC1, PC9, PC25, PC11
TG87	Micro-curriculum	PC18, PC24
TG100	Periodo de clases	PC11

Toda esta terminología fue avalada por los expertos del dominio, y los requerimientos expuestos sentaron la base para el desarrollo de las demás fases de construcción de la red ontológica.

### 4.3. Diseño de la red ontológica OEA

Partiendo de las preguntas de competencias y el pre-glosario, según lo estipulado por la metodología NeOn, pudimos seleccionar diferentes tácticas para crear nuestra red ontológica empleando los elementos más convenientes. Finalmente, optamos por la estrategia de la reingeniería de recursos no ontológicos y la reutilización de recursos ontológicos, tratando de incorporar aquellos que cubriesen los conceptos presentes en los requerimientos.

En cuanto al proceso de transformación de recursos no ontológicos, se realizó la solicitud y autorización de la información en digital correspondiente a un reporte de un programa analítico de una asignatura, y un reporte de una malla curricular o plan curricular, en donde se definieron las clases clave para la formalización del modelo.

Por otra parte, referente a la etapa de selección de recursos ontológicos que decidimos incorporar a nuestra red, recurrimos a buscadores de ontologías como LOV<sup>2</sup>, encontrando mediante un procedimiento manual 5 ontologías candidatas a la etapa de selección. Como criterios de selección, se exponen en la Tabla 4.5 algunas variables, entre las que se encuentran el propósito y alcance de las ontologías, así como la cobertura de términos de grupo y las preguntas de competencias que coinciden con el pre-glosario de términos. Luego tenemos características más descriptivas de la población de axiomas entre clases, propiedades de objetos, propiedades de datos e instancias que posee. La cobertura de términos se calculó mediante un proceso de extracción, transformación y carga del pre-glosario de términos a un repositorio en la herramienta *GraphDB*<sup>3</sup>, donde posteriormente se utilizó el módulo denominado *similarity* para comparar las ontologías candidatas en base al recuento de coincidencias entre los términos y axiomas.

---

<sup>2</sup><https://lov.linkeddata.es/>

<sup>3</sup><http://graphdb.ontotext.com/>

Tabla 4.5: Análisis de variables descriptivas de los recursos ontológicos candidatos.

Ontología	Propósito	Alcance	Cobertura de requerimientos funcionales			Características de la ontología		
			Preguntas de competencia	Términos de grupo	Clases	Propiedades de objeto	Propiedades de datos	Instancias
The Bibliographic Ontology (BIBO) <sup>4</sup>	Proporcionar una descripción para citas y referencias bibliográficas.	Describe la parte bibliográfica de documentos como libros y artículos científicos.	12	9	58	67	0	12
Semantic Web for Research Communities (SWRC) <sup>5</sup>	Representación de estructuras relacionadas con la investigación.	Modela personas, organizaciones, publicaciones y la relación que existe entre ellas.	11	2	70	79	0	0
DDI-RDF Discovery Vocabulary (DISCO) <sup>6</sup>	Descubrimiento de conjuntos de metadatos relacionadas con documentos científicos.	Representa en un modelo formal de metadatos de documentos de investigación y datos de encuestas.	3	1	16	40	0	0

<sup>4</sup><http://bibliontology.com/><sup>5</sup><http://ontoware.org/swrc><sup>6</sup><https://www.ddialliance.org/Specification/RDF>

Ontología	Propósito	Alcance	Cobertura de requerimientos funcionales			Características de la ontología		
			Preguntas de competencia	Términos de grupo	Clases	Propiedades de objeto	Propiedades de datos	Instancias
VIVO Core Ontology (VIVO) <sup>7</sup>	Representación de metadatos sobre procesos de investigación.	Modela conceptualizaciones de carácter investigativo y académicos como documentos científicos, estructuras organizacionales y proyectos de investigación.	25	28	146	107	0	4
Teaching Core Vocabulary Specification (TEACH) <sup>8</sup>	Representación de términos interrelacionados en un curso.	Modela seminarios, cursos y actores de los cursos, definiendo la relación que existen entre ellos.	14	3	11	28	0	0

De entre las opciones de la Tabla 4.5 seleccionamos finalmente VIVO, puesto que sobresalía en torno al número de preguntas de competencia y términos de pre-glosario cubiertos, conteniendo además una gran cantidad de axiomas propios del dominio de la docencia y la investigación. Analizando más a fondo su estructura, se encontró que VIVO se define sobre una superontología denominada *Basic Formal Ontology* (BFO)<sup>9</sup> que, a modo de marco de clasificación, ayuda a integrar otras ontologías como *Friend of a Friend* (FOAF)<sup>10</sup>, vCard<sup>11</sup>, *Simple Knowledge Organization*

<sup>7</sup><https://duraspace.org/vivo/>

<sup>8</sup><http://linkedscience.org/teach/ns/>

<sup>9</sup><http://basic-formal-ontology.org/>

<sup>10</sup><http://xmlns.com/foaf/spec/>

<sup>11</sup><https://www.w3.org/TR/vcard-rdf/>

*System (SKOS)*<sup>12</sup>, *Relational Ontology*<sup>13</sup> e *Information Artifact Ontology (AIO)*<sup>14</sup>. Esto permitió que tomásemos BFO como ontología medular para la adición de nuevos axiomas faltantes para el dominio de nuestra red.

En la Tabla 4.6 se presentan los axiomas utilizados para el módulo VIVO de nuestra red ontológica, que toma en cuenta un dominio menor del total de la ontología original, ajustado a nuestros propósitos. Es así que se toman 17 clases, 17 propiedades de objetos y 7 propiedades de datos de la totalidad que indica la Tabla 4.5.

Tabla 4.6: Axiomas reutilizados de la ontología VIVO.

Clases	Propiedades de los objetos	Propiedades de los datos
swo:Software	skos:related	vcard:birthdate
vcard:Individual	ro:has participant	vcard:email
foaf:Person	bfo:realizes	vcard:telephone
vcard:Name	bfo:realized in	vcard:middleName
vcard:Email	ro:bearer of	vcard:familyName
vcard:Phone	ro:inheres in	vcard:givenName
vcard:Address	ro:participates in	rdfs:label
vivo:Course	vivo:dateInterval	
vivo:Teacher Role	vivo:start	
vivo:Student Role	vivo:end	
vivo:Date Time Interval	vivo:date Time Precision	
vivo:Date Time Value	skos:broader	
vivo:University	skos:narrower	
vivo:Academic Year	skos:is top concept in scheme	
skos:Concept Scheme	skos:have top concept	
vivo:Awarded Degree	vivo:assigns	
vivo:Academic Degree	vivo:assignedBy	

En la Figura 4.1 se presenta el esquema de nuestra red, en donde hemos reutilizado las ontologías vCard, FOAF, BFO y VIVO por medio de diferentes nexos y selección de componentes acordes al dominio necesario, a través de un proceso de alineación y modularización de las ontologías.

Con la cobertura inicial de nuestros requerimientos por medio de la reutilización de los módulos de VIVO y BFO, procedimos a formalizar los requerimientos y términos del pre-glosario en un modelo formal integrado por 6 módulos, como se presenta en la Tabla 4.7. Cada uno de los módulos cubre un dominio específico en la red OEA:

- **Módulo Micro Curricular (MC):** describe las estructuras académicas que se forman a partir de las carreras que se componen de diferentes mallas curriculares, a las cuales pertenecen varias materias que responden a diferentes

<sup>12</sup><https://www.w3.org/TR/2008/WD-skos-reference-20080829/skos.html>

<sup>13</sup><http://www.obofoundry.org/ontology/ro.html>

<sup>14</sup><http://www.obofoundry.org/ontology/iao.html>

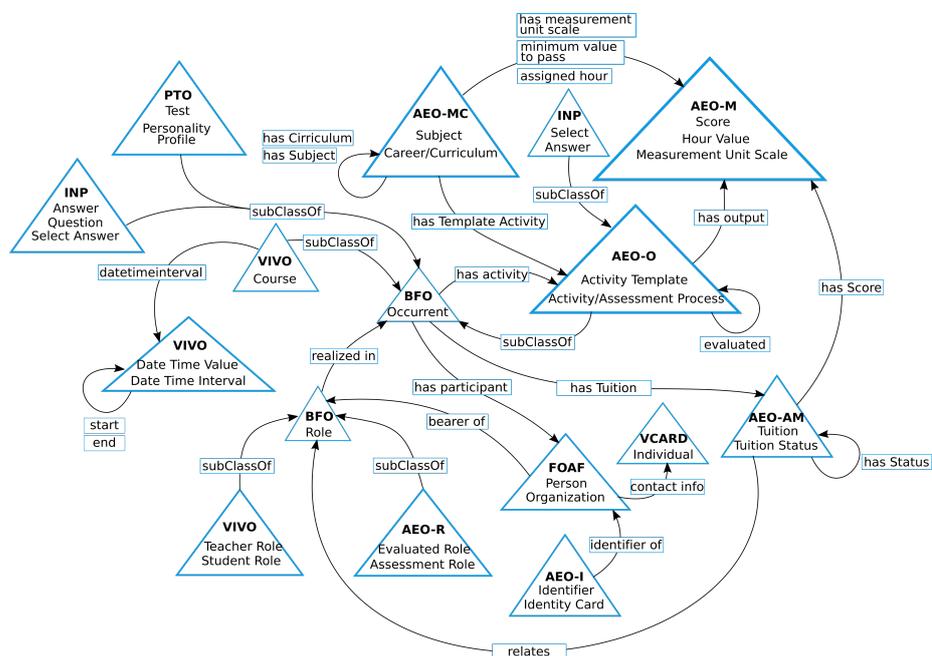


Figura 4.1: Esquema modular de la red ontológica OEA.

menciones. Cada materia ofertada tiene un sílabo, en donde se proponen diferentes estructuras correspondientes a los contenidos del programa analítico.

- **Módulo Identifier (I):** modela los identificadores tanto para trazabilidad de la información proveniente de diferentes fuentes, como axiomas para la identificación de las personas para conformar estrategias de unificación de perfiles.
- **Módulo Academic Metrics (AM):** modela los indicadores académicos que certifican el proceso de calificación. Dentro de estos se encuentran las matrículas, grados académicos otorgados y actividades realizadas en torno a estructuras del sílabo.
- **Módulo Metrics (M):** describe las métricas más comunes utilizadas por otros módulos, tales como métricas en determinadas unidades y métricas temporales (e.g. horas de clase).
- **Módulo Occurrent (O):** describe los procesos espacio-temporales propios de la enseñanza, como el proceso de evaluación o la realización de actividades.
- **Módulo Role (R):** modela los roles en torno a los procesos en los que participan los actores (e.g. rol de evaluador o rol de evaluado dentro de un proceso de calificación de una actividad).

Tabla 4.7: Axiomas de los módulos de la red OEA.

Clases		Propiedades de los objetos		Propiedades de los datos	
I	aeo:System Identity	I	aeo:has identifier	I	aeo:hour xsd:int
I	aeo:IdentifyCard	P	aeo:has living address	I	aeo:identifier xsd:string
MC	aeo:Career	MC	aeo:has curriculum	M	aeo:maxScore xsd:int
MC	aeo:Curriculum	MC	aeo:has subject	M	aeo:score xsd:int
MC	aeo:Subject Analytic Program	MC	aeo:has unit	M	aeo:minScore xsd:int
MC	aeo:Common Offer Subject	MC	aeo:minimum value to pass	M	aeo:nameScale xsd:string
MC	aeo:Subject	M	aeo:assigned total hours		
MC	aeo:Unit	M	aeo:assigned hours		
MC	aeo:Training Field	MC	aeo: has level		
MC	aeo:Mode	MC	aeo:has syllabus		
M	aeo:Score	MC	aeo:has quintile		
M	aeo:Hour Value	O	aeo:has template activity		
M	aeo:Teaching Component Hour	AM	aeo:has learning structure		
MC	aeo:Learning Experimentation aeo:Hour				
M	aeo:Self-Employment Hour				
MC	aeo:Level				
MC	aeo:Sede				
AM	aeo:Tuition				
AM	aeo:Tuition Status				
M	aeo:Score				
M	aeo:Measurement Unit Scale				
MC	aeo:Syllabus				
MC	aeo:Quintille				
O	aeo:Assessment Process				
R	aeo:Assessment Role				
R	aeo:Evaluated Role				
AM	aeo:Tertiary Level				
AM	aeo:Postgraduate Level				
AM	aeo:Template Activity				
AM	aeo:Learning Structure				
AM	aeo:Partial				

Cada uno de los módulos de la red ontológica pertenece a un contexto en específico, permitiéndonos gestionar de mejor manera los diferentes componentes importados a nuestra red. Para la construcción de los módulos se utilizó la herramienta *Protégé*<sup>15</sup>. Como se presenta en la Figura 4.2, cada una de las ontologías se desarrolló por separado, teniendo en cuenta la alineación de los componentes mediante nexos que son clases de otros módulos. Para verificar el funcionamiento de la red, como primer paso se importaron todos los módulos en un mismo proyecto y finalmente corrimos el proyecto sobre el razonador *Hermit*<sup>16</sup> para comprobar que no existiesen inconsistencias de los axiomas.

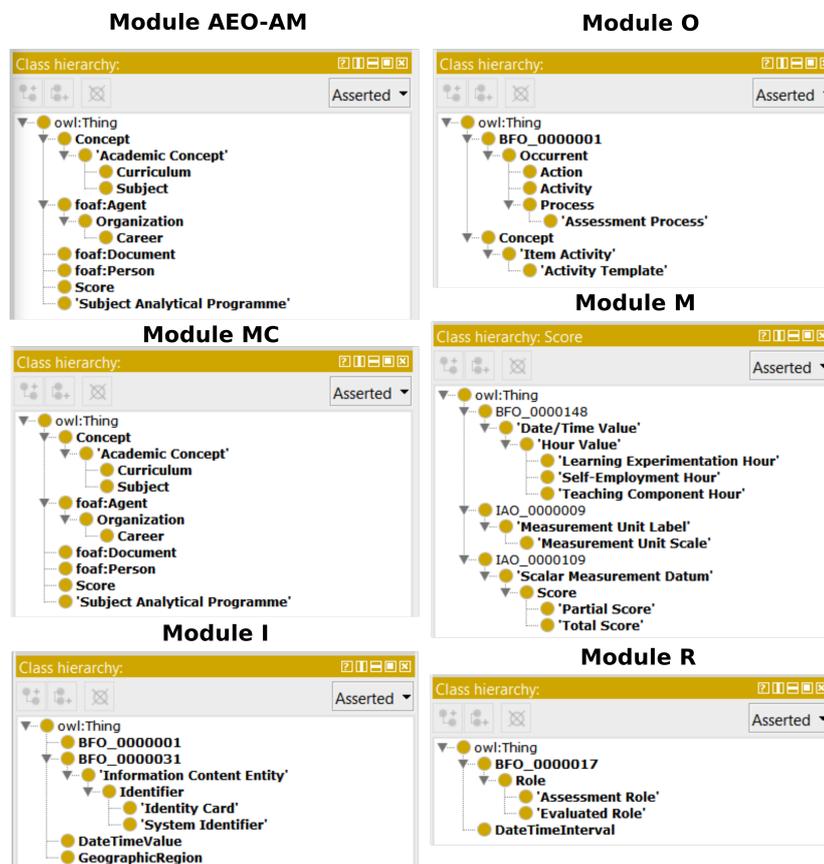


Figura 4.2: Captura de los módulos de Protégé que contienen las clases de la red OEA.

<sup>15</sup><https://protege.stanford.edu/>

<sup>16</sup><http://www.hermit-reasoner.com/>

#### 4.4. Implementación del gestor de conocimiento para la red ontológica OEA

El sistema de herramientas y técnicas que se planteó para la conformación del nuevo repositorio semántico de la Universidad Politécnica Salesiana se presenta en la Figura 4.3. En el proceso de conformación de una base de conocimiento intervienen tres capas:

- **Capa de origen:** define las fuentes de información que se tomarán en cuenta para poblar la base de conocimiento. Las fuentes de información aparecen caracterizadas por su estructura, desde sistemas relacionales de bases de datos (como el empleado por el Sistema Nacional Académico, SNA) hasta estructuras más simples tales como conjuntos de datos en hojas de cálculo, etc.
- **Capa de análisis:** cumple la función de extracción de información a partir de la capa de origen, con procedimientos diferenciados en función de las fuentes de información. Seguidamente de extraer la información, esta se transforma en una estructura de mapa y dependiendo de a qué estructura ontológica se haya de transformar, pasan por el *analizador de modelo* respectivo y se transforman en triplas en notación Turtle. Como paso final, se cargan las estructuras por medio del motor de carga, que divide las sentencias de triplas en varios hilos y por medio del API de **RDF4J**<sup>17</sup> se insertan a la base de triplas en **GraphDB**.
- **Capa de conocimiento:** cumple la función de almacenar y producir nuevo conocimiento, representado en estructuras RDF de tipo *sujeto-predicado-objeto*. En nuestra implementación se corresponde con la herramienta GraphDB, una base de datos de grafos, que es compatible con los estándares de la W3C y permite almacenar sentencias RDF y ejecutar axiomas de múltiples tipos.

Después de la conformación de una base de conocimiento, se cuenta con dos capas adicionales que conforman una herramienta para la administración del conocimiento:

- **Capa de servicio:** tiene el rol de interfaz de comunicación entre el repositorio de conocimiento y otras aplicaciones externas que consulten y agreguen nueva información de/a este. Las operaciones a la base de conocimiento se realizan a través de una interfaz de consultas en lenguaje *SPARQL* por medio de la API de RDF4J.
- **Capa de app:** da soporte a la creación de distintas aplicaciones que hagan uso de la base de conocimiento, tales como las que configuran los aportes de investigación en IA de esta tesis, expuestas en el Capítulo 5.

#### 4.5. Población de la red

En esta sección se describen los pasos de los procesos realizados para la carga de los datos atesorados por la Universidad Politécnica Salesiana en la red ontológica

<sup>17</sup><http://rdf4j.org/>

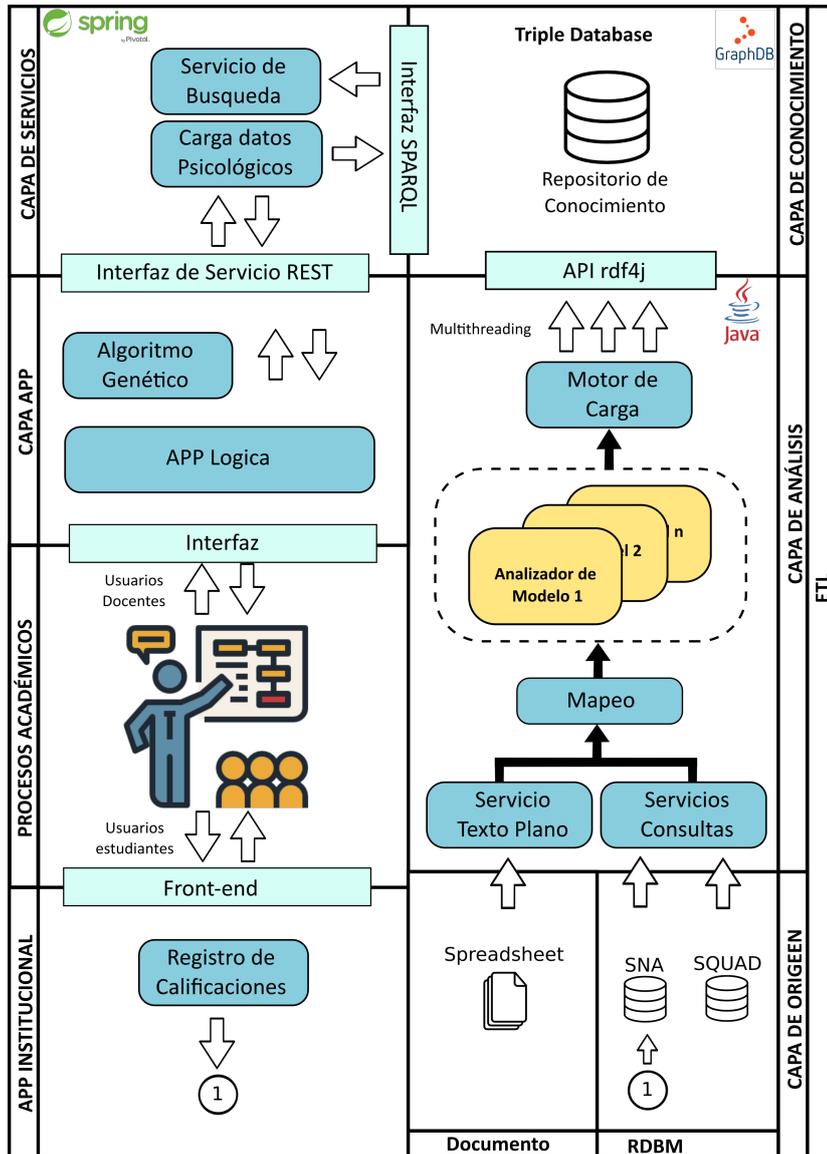


Figura 4.3: Capas, herramientas y técnicas del nuevo repositorio semántico de la UPS.

OEA, desde la obtención de los datos fuentes hasta el proceso de mapeo y carga al repositorio de triplas.

Los reportes de origen fueron obtenidos del Sistema Nacional Académico (SNA) de la Universidad Politécnica Salesiana. En la Tabla 4.8 figuran dos columnas: la primera hace referencia al nombre del reporte y la segunda a la descripción de la información que contiene.

Tabla 4.8: Reportes de origen procedentes del sistema SNA.

<b>Nombre del Reporte</b>	<b>Descripción</b>
1_y_9_Docentes.csv	Información de perfil de las personas que cumplen el rol de docente en la universidad.
2_Tercer Nivel.csv	Información de los títulos de tercer nivel de los docentes.
3_Posgrado.csv	Información de los títulos de postgrado de los docentes.
4_Docentes Tercer Nivel.csv	Registros intermedios entre Docente y Tercer Nivel.
5_Docentes Posgrado.csv	Registros intermedios entre Docente y Posgrado.
5.1_Historial Contratos.csv	Información del periodo en el que una persona (Docente) cumple un rol en la UPS.
6_Estudiante.csv	Información de perfil de las personas que cumplen el rol de estudiante en la universidad.
7_Cuenca.csv	Registro de notas finales de un estudiante en un curso de un programa analítico en la sede Cuenca.
7_Guayaquil.csv	Registro de notas finales de un estudiante en un curso de un programa analítico en la sede Guayaquil.
7_Quito.csv	Registro de notas finales de un estudiante en un curso de un programa analítico en la sede Quito.
8_Cuenca.csv	Registro de notas parciales por actividad de un estudiante en un curso de un programa analítico de ofertas curriculares nueva en la sede Cuenca.
8_Guayaquil.csv	Registro de notas parciales por actividad de un estudiante en un curso de un programa analítico de ofertas curriculares nueva en la sede Guayaquil.
8_Quito.csv	Registro de notas parciales por actividad de un estudiante en un curso de un programa analítico de ofertas curriculares nueva en la sede Quito.
10_Quintil.csv	Registros de la ubicación de un estudiante en cuanto a parámetros socioeconómicos.
11_Silabo Asignatura.csv	Registro del silabo de una asignatura.
12_Docente Silabo Asignatura Cuenca.csv	Registro intermedio entre el silabo y el docente de la asignatura en la sede Cuenca.
12_Docente Silabo Asignatura Guayaquil.csv	Registro intermedio entre el silabo y el docente de la asignatura en la sede Guayaquil.
12_Docente Silabo Asignatura Quito.csv	Registro intermedio entre el silabo y el docente de la asignatura en la sede Quito.
13_14_Bibliografia (textos_lecturas).csv	Registros de la bibliografía del programa analítico
15_Criterio de Evaluación.csv	Registros de los criterios de evaluación de un indicador de logro

Nombre del Reporte	Descripción
16_Organización de Aprendizaje Criterio de los componentes de Aprendizaje.csv	Criterio de organización de aprendizaje de un indicador de logro
17_Tipo Organización de Aprendizaje.csv	Tipo de organización de aprendizaje utilizado en los indicadores de logro.
18_Organización Aprendizaje Cuenca.csv	Registro de la organización de aprendizaje en una unidad de un programa analítico en la sede Cuenca
18_Organización Aprendizaje Guayaquil.csv	Registro de la organización de aprendizaje en una unidad de un programa analítico en la sede Guayaquil
18_Organización Aprendizaje Quito.csv	Registro de la organización de aprendizaje en una unidad de un programa analítico en la sede Quito
19_Malla Microcurricular.csv	Malla micro curricular de una carrera
20_Asignaturas.csv	Registro de asignaturas
21_Oferta Asignatura.csv	Registro de asignaturas equivalentes
22_Carrera.csv	Registro de carreras universitarias
23_Unidad de Organización Curricular.csv	Registro de unidades organizacionales por programa analítico de asignatura
24_Campo de Formación.csv	Campo de formación del programa analítico de la asignatura
25_Modalidad.csv	Modalidad de programa analítico de la asignatura
26_Programa Analítico Asignatura.csv	Registro de los programas analíticos de asignatura de una malla curricular
27_Unidad Resultado Aprendizaje.csv	Registros de la unidad de un programa analítico de asignatura con el resultado de aprendizaje
28_Contenido de la Unidad.csv	Registros de los contenidos de las unidades
29_Unidad.csv	Unidad de un programa analítico de asignatura
30_Resultado de Aprendizaje.csv	Resultado de aprendizaje de una unidad
31_Indicador de Logro.csv	Indicador de logro de una unidad
32_Titulo.csv	Título según la malla microcurricular
33_Itinerario.csv	Itinerario de la malla microcurricular
34_Itinerario Programa Analítico Asignatura.csv	Registros de la correspondencia de un itinerario a una malla microcurricular
35_Malla Curricular Título.csv	Registros de la correspondencia de un título a una malla microcurricular
36_Area Conocimiento .csv	Registros de la clasificación de áreas del conocimiento

Para explicar la migración del modelo de datos de estos archivos, mostramos en primer lugar la información que contienen dos de ellos, concretamente los denominados *6\_Estudiantes.csv* y *1\_y\_9\_Docentes.csv*. La Tabla 4.9 y la Tabla 4.10 muestran ejemplos de datos. La Figura 4.4 y la Figura 4.5 ilustran los esquemas ontológicos de mapeo de estos formatos de datos a entidades de nuestra red ontológica, y la Figura 4.6 muestra las instancias resultantes para los registros concretos de la Tabla 4.9 y la Tabla 4.10.

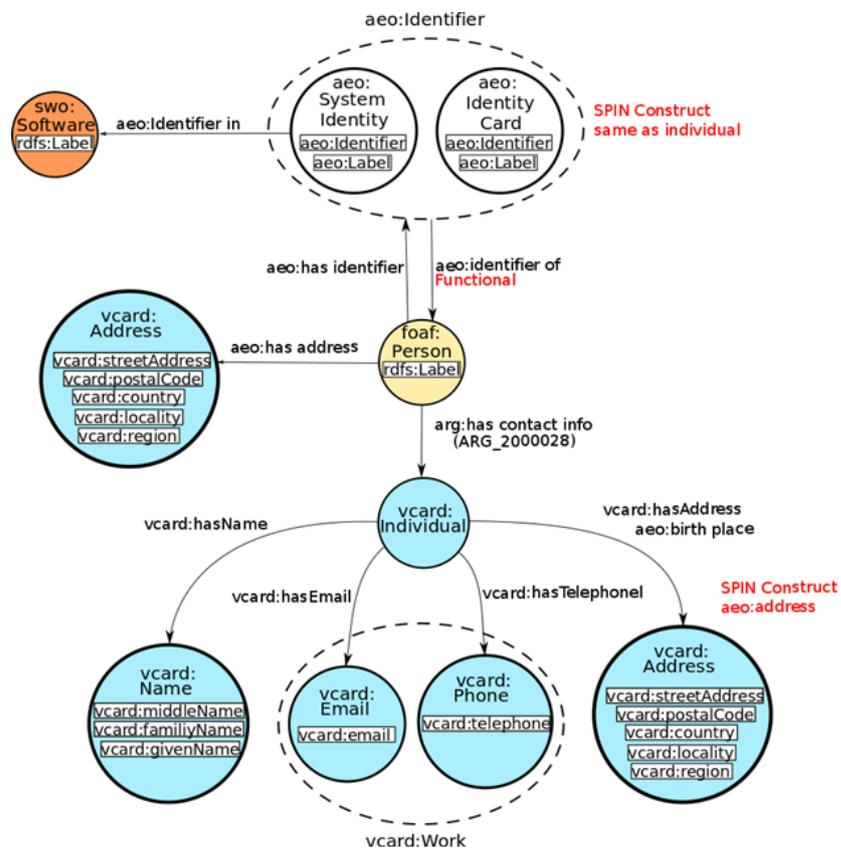


Figura 4.4: Esquema ontológico del mapeo de estudiantes.

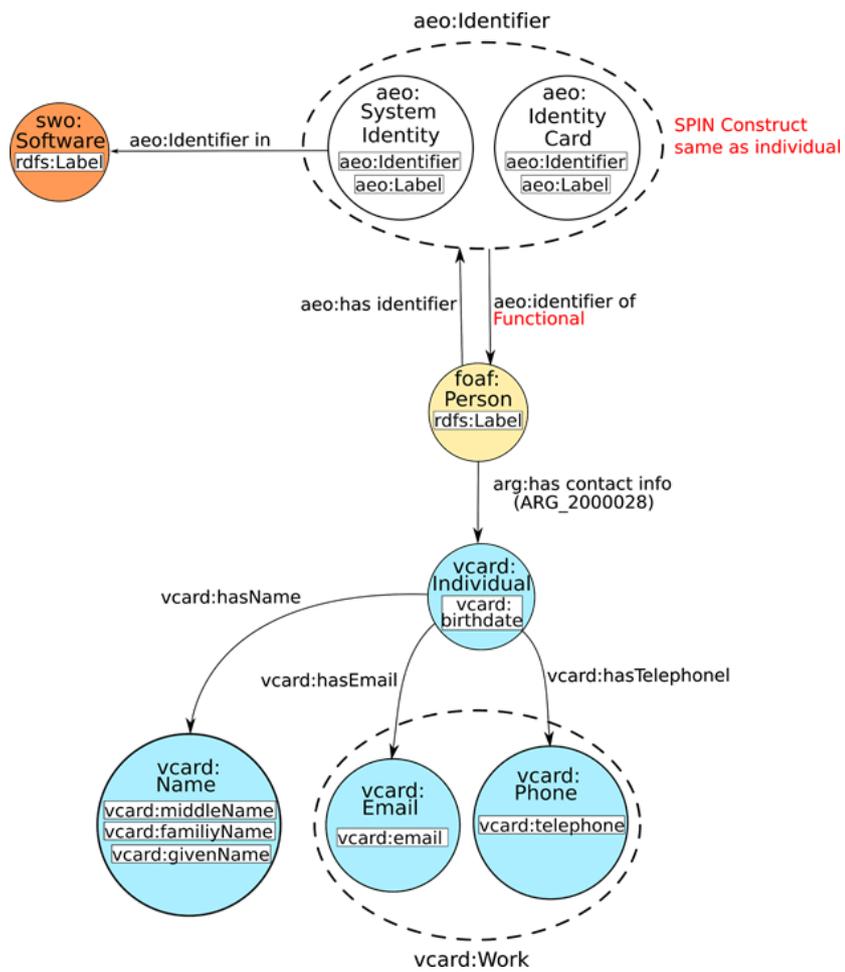


Figura 4.5: Esquema ontológico del mapeo de docentes.

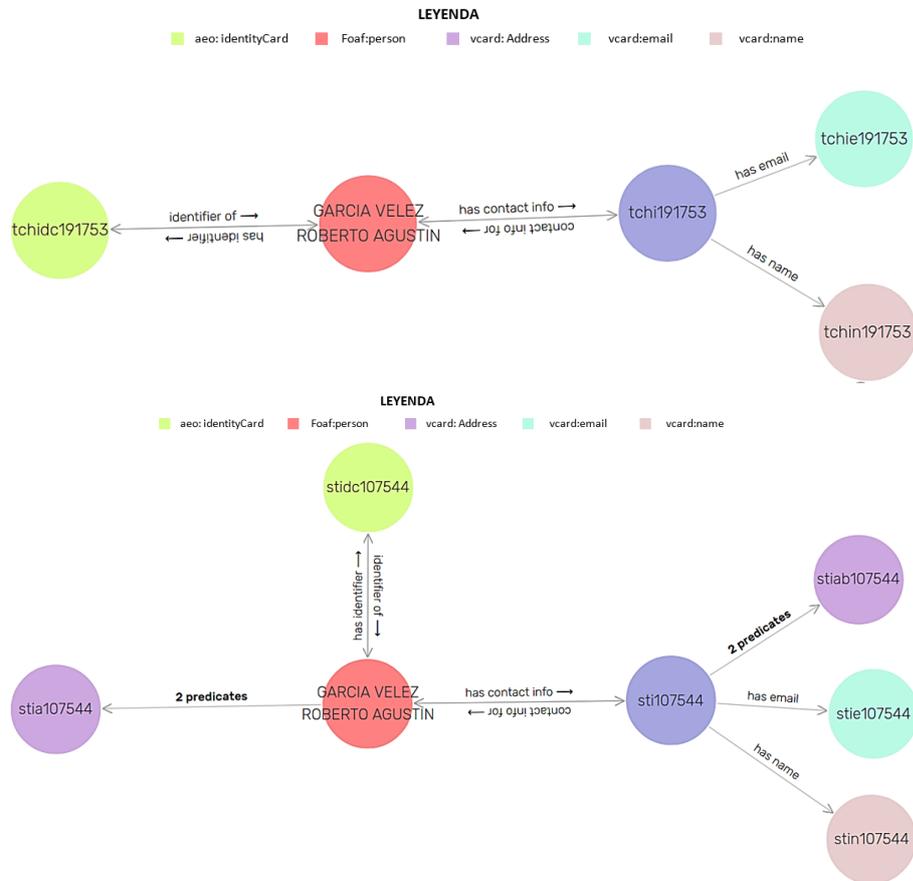


Figura 4.6: Instancias resultantes del mapeo de estudiantes y docentes.

Tabla 4.9: Estructura del modelo de datos del archivo 6\_Estudiantes.csv.

Atributo	Valor
<b>CODIGO_ESTUDIANTE</b>	225079
<b>NOMBRES</b>	ROBERTO AGUSTÍN
<b>APELLIDOS</b>	GARCÍA VÉLEZ
<b>FECHA_NACIMIENTO</b>	03/03/1981 00:00:00
<b>PAIS_NAC</b>	ECUADOR
<b>PROVINCIA_NAC</b>	AZUAY
<b>CIUDAD_NAC</b>	CUENCA
<b>IDENTIFICACION</b>	0103650891
<b>PROVINCIA_DOM</b>	AZUAY
<b>CIUDAD_DOM</b>	CUENCA

Tabla 4.10: Estructura del modelo de datos del archivo 1\_y\_9\_Docentes.csv.

Atributo	Valor
<b>CODIGO_DOCENTE</b>	1
<b>NOMBRES</b>	ROBERTO AGUSTÍN
<b>APELLIDOS</b>	GARCÍA VÉLEZ
<b>FECHA_NACIMIENTO</b>	03/03/1981 00:00:00
<b>IDENTIFICACION</b>	0103650891

## 4.6. Optimización de motor de reglas

En repositorios semánticos destinados a almacenar gran cantidad de triplas, a fin de garantizar la escalabilidad surge la obligación de acotar el proceso de inferencia desencadenado por cada nueva información que se añade, conforme a las necesidades de razonamiento que se pretendan satisfacer. Cabe explicar que cada nueva tripla semántica a un repositorio como GraphDB da lugar a que se comprueben las condiciones de aplicabilidad de cada regla OWL a tener en cuenta, y que la ejecución de aquellas para las cuales se cumplen dichas condiciones provoca la adición de nuevas triplas, que a su vez pueden dar lugar a nuevas inferencias, y así sucesivamente.

Para ilustrar el problema en el caso concreto de la importación de las fuentes de información de la Universidad Politécnica Salesiana mencionadas en las secciones anteriores, realizamos un experimento simple en el que se tomaron apenas los 10 primeros registros, los 10 últimos y 20 intermedios al azar del total de cada fuente. Esa información, traducida a los formatos de nuestra red ontológica, daba lugar a más de 14.000 triplas, que con el conjunto de reglas por defecto de GraphDB (las del dialecto de OWL denominado *OWL Lite*) se convertían en más de 102.000.

Revisando este asunto con los ingenieros ontológicos y manteniendo a mano la documentación de GraphDB (véase, en particular, la información sobre los dialectos de OWL de <http://www.graphdb.ontotext.com/documentation/8.0/free/introduction-to-semantic-web.html>.) concluimos que las necesidades de inferencia del entorno académico de la Educación Superior quedan satisfechas con las reglas del conjunto denominado OWL Horst, introducido en [245]. Se trata de un dialecto más limitado que *OWL Lite*, pero que se ha demostrado suficiente en otros

dominios similares al que nos ocupa [254]. Su conjunto de reglas es el que recoge la Tabla 4.11:

Tabla 4.11: Conjunto de reglas del dialecto OWL Horst.

Id	Condición	Consecuente
rdfs2	a b c [Constraint b != <rdf:type>] b <rdfs:domain> d	a <rdf:type> d
rdfs3	a b c b <rdfs:range> d	c <rdf:type> d
rdfs6	a <rdf:type> <rdfs:Property>	a <rdfs:subPropertyOf> a
rdfs7	a b c b <rdfs:subPropertyOf> d [Constraint b != d]	a d c
rdfs8_10	a <rdf:type> <rdfs:Class>	a <rdfs:subClassOf> a
rdfs12	a <rdf:type> <rdfs:ContainerMembershipProperty>	a <rdfs:subPropertyOf> <rdfs:member>
rdfs13	a <rdf:type> <rdfs:Datatype>	a <rdfs:subClassOf> <rdfs:Literal>
proton_ Transitive Over	a <psys:transitiveOver> b c a d d b e e <rdf:type> t	c a e
proton_ TransProp	a <rdf:type> <owl:TransitiveProperty>	a <psys:transitiveOver> a
proton_ TransProp Induct	a <psys:transitiveOver> a	a <rdf:type> <owl:TransitiveProperty>
Proton_ roleHolder	a <pext:roleHolder> b b <rdf:type> <pext:Agent> a <pext:roleIn> c	b <pext:involvedIn> c
owl_invOf	a b c b <owl:inverseOf> d	c d a
owl_invOfBy SymProp	a <rdf:type> <owl:SymmetricProperty>	a <owl:inverseOf> a
owl_Sym PropBy Inverse	a <owl:inverseOf> a	a <rdf:type> <owl:SymmetricProperty>
owl_Funct Prop	a <rdf:type> <owl:FunctionalProperty> [Constraint a != <rdf:type>] b a c b a d [Constraint c != d, d != c] [Cut]	c <owl:sameAs> d
owl_Inv Funct Prop	a <rdf:type> <owl:InverseFunctionalProperty> [Constraint a != <rdf:type>] b a c d a c [Constraint b != d, d != b] [Cut]	b <owl:sameAs> d

<b>Id</b>	<b>Condición</b>	<b>Consecuente</b>
owl_Equiv ClassBy SubClass	a <rdfs:subClassOf> b [Constraint b != a] b <rdfs:subClassOf> a [Cut]	a <owl:equivalentClass> b
owl_Equiv vPropBy SubProp	a <rdfs:subPropertyOf> b [Constraint b != a] b <rdfs:subPropertyOf> a [Cut]	a <owl:equivalentProperty> b
owl_Attr By HasVal	a <owl:hasValue> b a <owl:onProperty> c d <rdf:type> a	d c b
owl_sub classBy Intersect1	a <owl:intersectionOf> b	a <onto:_interOf> b
owl_subclass By Intersect2	a <onto:_interOf> b b <rdf:first> c b <rdf:rest> d	a <rdfs:subClassOf> c a <onto:_interOf> d
owl_type By Intersect_1	a <onto:_allTypes> b c <owl:intersectionOf> b c <rdf:type> d [Constraint d != <owl:Restriction>]	a <rdf:type> c
owl_type By Intersect_2	a <rdf:first> b a <rdf:rest> <rdf:nil> c <rdf:type> b a <rdf:type> h [Constraint h != <owl:Restriction>]	c <onto:_allTypes> a
owl_type By Intersect_3	a <rdf:first> b d <rdf:type> b a <rdf:rest> c d <onto:_allTypes> c d <rdf:type> h [Constraint h != <owl:Restriction>]	d <onto:_allTypes> a
owl_type By SomeVal	a <rdf:type> b c <owl:onProperty> d c <owl:someValuesFrom> b e d a	e <rdf:type> c
owl_type ByAllVal	a <owl:onProperty> b c <rdf:type> a a <owl:allValuesFrom> d c b e	e <rdf:type> d
owl_type ByHasVal	a <owl:onProperty> b a <owl:hasValue> c d b c	d <rdf:type> a

Como se puede observar en la Figura 4.7, con el conjunto de reglas *OWL Horst*, a partir de las más de 14.000 triplas que componen la muestra seleccionada se infirieron un total de 58.485 triplas nuevas (informa GraphDB de que la ratio de expansión es 5,05).

A partir de este primer análisis se procedió a capturar información a través de GraphDB sobre el uso de cada regla, a fin de identificar cuáles eran las que más inferencias generaban y determinar si eran realmente necesarias. Como se puede

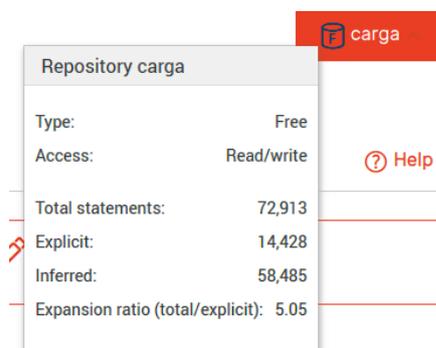


Figura 4.7: Expansión de un conjunto de triplas inicial por medio del conjunto de reglas *OWL Horst*.

observar en la gráfica de la Figura 4.8, la regla *proton\_TransitiveOver\_1* genera 25.000 triplas, mientras que *proton\_Transitive-Over\_2* genera más de 16.000. Ya en menor medida, contribuían también *owl\_subclassByIntersect\_2\_2*, *owl\_typeBy Intersect\_3\_0* y *owl\_typeBy Intersect\_3\_1*.

Partiendo de estas observaciones, los ingenieros ontológicos recomendaron aplicar al conjunto de reglas de *OWL Horst* los cambios que recoge la Tabla 4.12, eliminando reglas de inferencia por intersecciones y limitando determinados razonamientos de transitividad. Los nuevos cálculos de expansión de la información suministrada se muestran en la Figura 4.9, reflejando una disminución de la tasa de expansión a 2,1.

Tabla 4.12: Modificaciones realizadas sobre el conjunto de reglas *OWL Horst*.

ID	Condición	Consecuente
proton_TransitiveOver	a <psys:transitiveOver> b c a d d b e e <rdf:type> t [Constraint t != <owl:Restriction>, t != <rdfs:Class>]	c a e
owl_subclassByIntersect1		Omitida
owl_subclassByIntersect2		Omitida
owl_typeByIntersect_1		Omitida
owl_typeByIntersect_2		Omitida
owl_typeByIntersect_3		Omitida

Analizando la nueva gráfica de barras sobre el uso de cada regla (Figura 4.10) podemos ver cómo se han limitado significativamente los picos de inferencias a través de *proton\_Transitive\_Over\_1* y *proton\_Transitive\_Over\_2*, sin que se viesen afectadas las capacidades de gestión de conocimiento que pretendíamos habilitar —reflejadas en la implementación de los aportes de esta tesis en materia de aplicaciones de Inteligencia Artificial, detallados en el Capítulo 5.

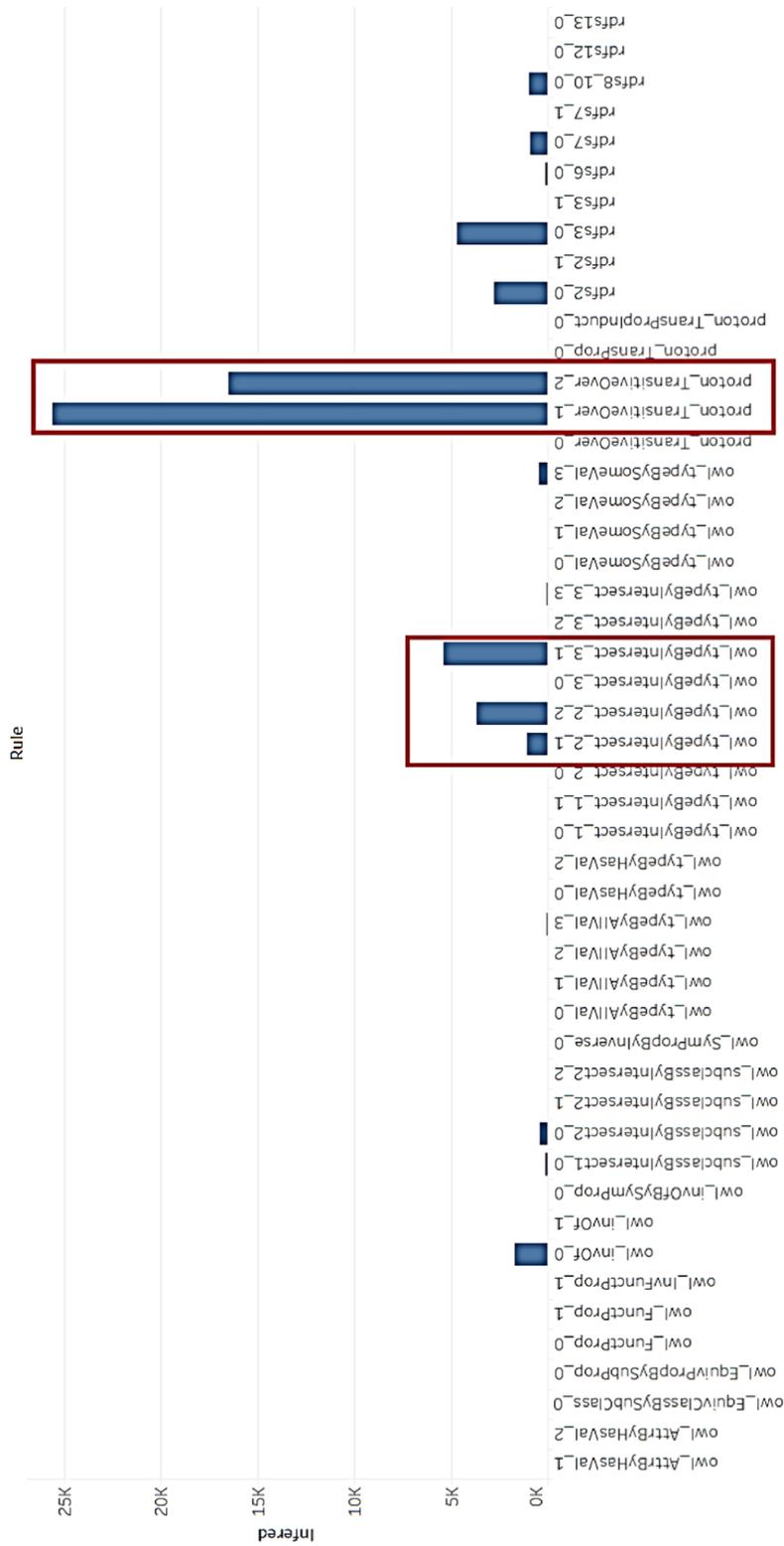
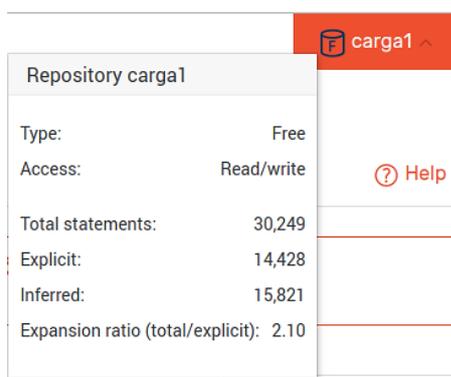


Figura 4.8: Análisis de inferencias por regla de OWL Horst.



The screenshot shows a web interface with a red header bar containing a logo and the text 'carga1'. Below the header is a white box with the title 'Repository carga1'. Inside this box is a table with the following data:

Type:	Free
Access:	Read/write
Total statements:	30,249
Explicit:	14,428
Inferred:	15,821
Expansion ratio (total/explicit):	2.10

To the right of the table is a red button with a question mark icon and the text 'Help'. A red rectangular box highlights the 'Explicit' and 'Inferred' rows of the table.

Figura 4.9: Expansión de un conjunto de triplas inicial por medio del conjunto de reglas *OWL Horst* modificado.

## 4.7. Resultados de migración y tiempos de carga

El resultado del proceso de migración de las bases de datos previas de la Universidad Politécnica Salesiana fue una base de conocimiento semántico —acorde a las definiciones de nuestra red ontológica OEA— de más de **52 millones de triplas**. La Tabla 4.13 recopila datos de número de triplas y tiempos de carga correspondientes a distintos ficheros de origen. Los conjuntos de datos más pesados en tiempo de ejecución se corresponden con las calificaciones de las tres sedes de la UPS en Cuenca, Quito y Guayaquil, con un histórico de la información de 24 años de procesos académicos (en conjunto, los procesos de transformación y carga se tomaron aproximadamente 2 horas y media con ellos, generando el 91,1% del total de la población de triplas). Era un resultado esperable, en tanto que el producto entre el número total de estudiantes por las notas obtenidas en sus procesos académicos conforma el eje principal de todo el proceso académico.

## 4.8. Trazabilidad de los perfiles

Para comprobar que la integración de la información de diferentes fuentes se encuentre alineada a los perfiles de las personas en función de las tres vertientes que pretendemos abarcar (académica, psicológica y socioeconómica), cuestionamos el tipo de información que debe tener cada dominio tomando como centro a la persona. Así, de 112.174 instancias de personas que se encuentran en nuestro repositorio, se considera para este plan experimental a 101.943 instancias de personas que tuvieron el rol de estudiante en algún proceso académico de la Universidad Politécnica Salesiana.

Como perfil básico de información personal tenemos las clases *Address*, *Individual*, *Email*, *Identity Card* y *Phone*, que del total de los datos del perfil ocupan un 4,3% con respecto al total de instancias. En la Figura 4.11 se presenta el esquema de la ontología correspondiente a dicha información.

Por su parte, el perfil académico es el más denso en cuanto a instancias, con un 93,7% de la población de acuerdo al dominio de los perfiles (véase Tabla 4.14).



Figura 4.10: Análisis de inferencias por regla del conjunto *OWL Horst* modificado.

Tabla 4.13: Resultados del proceso de migración de elementos no ontológicos de diversas fuentes.

Sistema de origen	Modelo de origen	Número de filas	Número de columnas	Número de triplas de salida	Tiempo de ejecución
Sistema académico UPS	6_Estudiante.csv	99879	12	2219214	12:22.291s
Sistema académico UPS	22_Carrera.csv	131	2	912	8.664s
Sistema académico UPS	19_Malla Microcurricular.csv	42243	3	18081	9.370s
Sistema académico UPS	20_Asignaturas.csv	6312	2	44185	10.052s
Sistema académico UPS	26_Programa_Analitico_Asignatura.csv	1068	9	1365	3.622s
Sistema académico UPS	7_Calificaciones Generales_Cuenca.csv	908336	5	11242658	50:38.740s
Sistema académico UPS	7_Calificaciones Generales_Quito.csv	2004905	5	24969279	1:12:31.578s
Sistema académico UPS	7_Calificaciones Generales_Guayaquil.csv	710621	5	8791448	26:26.051s
Sistema académico UPS	1_y_9_Docentes.csv	2941	6	33318	7.527s
Sistema académico UPS	11_Silabo Asignatura.csv	3269	3	16811	8.085s
Sistema académico UPS	12_Docente Sílabo Asignatura Cuenca.csv	237433	2	234305	52.175s
Sistema académico UPS	12_Docente Sílabo Asignatura Guayaquil.csv	29273	2	198979	42.141s
Sistema académico UPS	12_Docente Sílabo Asignatura Quito.csv	79212	2	502729	1:41.456s
Sistema de bienestar estudiantil de la UPS	10_Quintil.csv	74778	3	1118442	3:10.265s
Datos de personalidad	Reporte_NEO_PI-R.csv	183	5	5115	13.747s

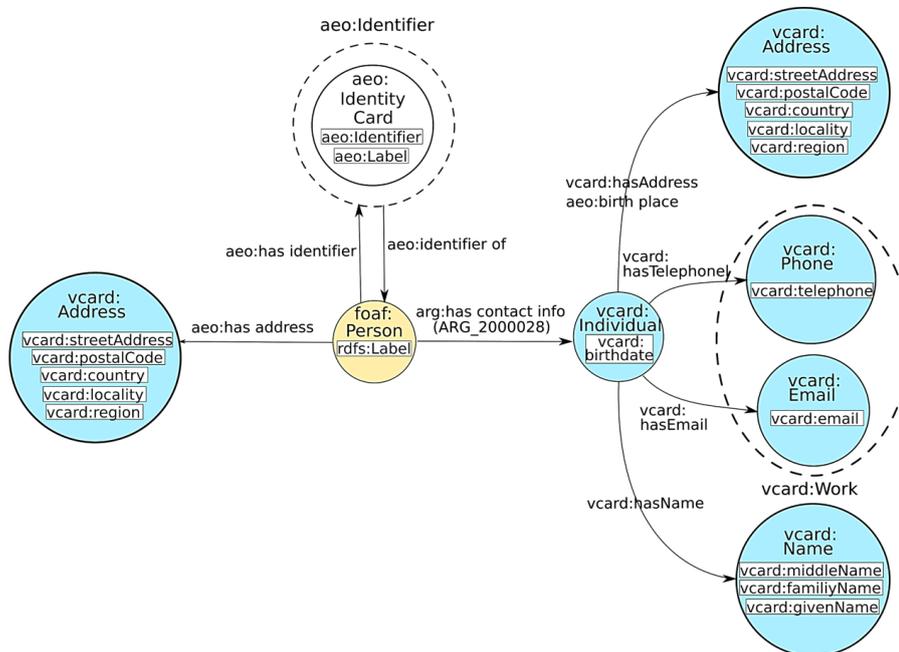


Figura 4.11: Diagrama ontológico del dominio del perfil personal.

Este perfil considera elementos en torno al proceso de aula que están conformados, primeramente, por su participación dentro de un proceso académico de aula, representado por instancias de la clase **Course**. En estas instancias, una persona puede tener el rol de alumno o docente, modelado como instancias de la clase **Role** que, a su vez, tienen un registro de notas por medio de la clase **Tuition**, que posee las instancias de la puntuación obtenida y el estado aprobado o reprobado. Este esquema se puede visualizar en la Figura 4.12, en donde se detalla también que el proceso académico pertenece a una materia ofertada en un plan de estudios.

Tabla 4.14: Resultados de la población de clases según el dominio de los perfiles.

Dominio	Clase	Instancias
Perfil personal	vcard:Individual	101943
Perfil personal	aeo:Identity Card	101943
Perfil académico	ero:Student Role	3637311
Perfil académico	aeo:Tuition	3637311
Perfil académico	aeo:Score	3637311
Perfil socioeconómico	aeo:Evaluated Role	74795
Perfil socioeconómico	aeo:Quintile 3	41974
Perfil socioeconómico	aeo:Quintile 4	9853
Perfil socioeconómico	aeo:Quintile 2	21121
Perfil socioeconómico	aeo:Quintile 1	1717
Perfil socioeconómico	aeo:Assessment Process	74795
Perfil académico	vivo:Course	19972
Perfil socioeconómico	aeo:Quintile 5	130

Dominio	Clase	Instancias
Perfil psicológico	prt:Personality Profile	183
Perfil psicológico	vivo:DateTimeValue	183
Perfil psicológico	aeo:Score	915
Perfil personal	vcard:Email	101943
Perfil personal	vcard:Name	101943
Perfil personal	vcard:Address	99653

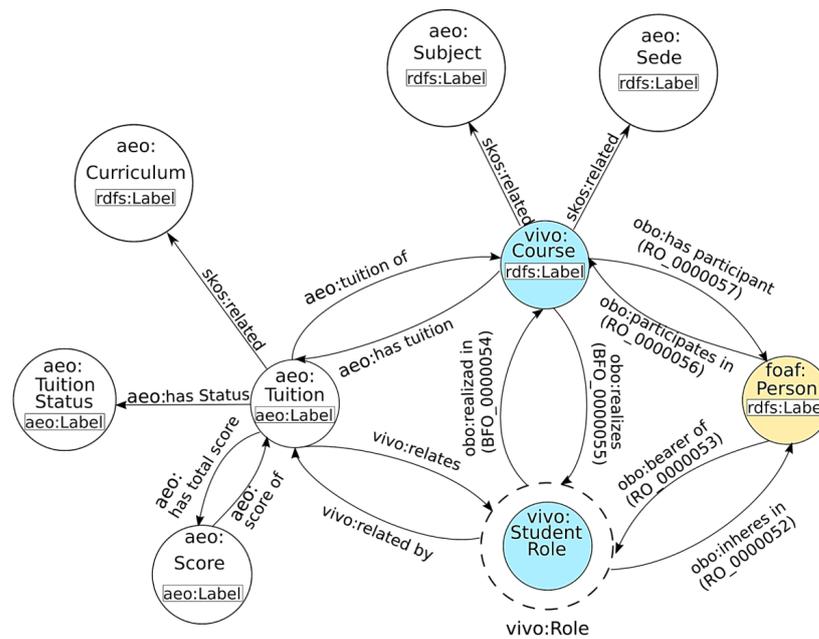


Figura 4.12: Diagrama ontológico del dominio del perfil académico.

Los conceptos más relevantes que se manejan en el perfil socioeconómico se muestran en la Figura 4.13. Los datos parten de una evaluación en la cual la Universidad Politécnica Salesiana participa con el rol de evaluador, mientras que el alumno tiene el rol de evaluado. La principal salida de este proceso es el quintil, una variable categórica en la cual el valor 1 representa el nivel socioeconómico más alto, mientras que el valor 5 representa el nivel más bajo. Tomando en cuenta que los registros se comenzaron a capturar en 2014, podemos entender que en la Tabla 4.14 el perfil socioeconómico se corresponda apenas con el 1,92% de las instancias del dominio de los perfiles.

Finalmente, en cuanto al perfil psicológico, como parte de la investigación planteada en esta tesis sobre algoritmos genéticos aplicados a la formación de grupos de trabajo colaborativo (más detalles en el Capítulo 6), realizamos un experimento piloto para recabar información de la personalidad de 183 alumnos. Sin entrar en más



Partiendo de los resultados de trazabilidad, pudimos constatar que existe una gran diferencia en el número de instancias de cada dimensión de los perfiles, lo cual no solo se debe a la complejidad de cada esquema, sino principalmente a los diferentes procesos de levantamiento de datos que afectan al alcance de la completitud sobre el total histórico de la institución. De un total de 101.943 estudiantes, se cuenta con un perfil académico de cada uno de ellos, con un perfil semi-completo (académico y socioeconómico) de 71.455, y con un perfil completo (académico, socioeconómico y psicológico) de sólo 183. Este hecho ilustra una de las grandes ventajas al momento de utilizar ontologías: que son altamente flexibles para trabajar con información incompleta de las instancias.

En la Figura 4.15 se visualiza el perfil completo de un estudiante. Podemos ver que el perfil académico está compuesto de varios nodos de cursos en los que participó y roles que tuvo en esos cursos, denotados de color morado y azul respectivamente. También vemos que el perfil personal está compuesto de tres nodos que corresponden a su documento de identidad, carta de individuo y la dirección de domicilio. En el perfil socioeconómico tenemos al nodo principal que simboliza el quintil en el que fue evaluado. Por último, en el perfil psicológico tenemos los resultados del test de personalidad que describiremos en el siguiente capítulo. Al contar con estas cuatro vertientes, se resalta que nuestra red ontológica OEA toma como centro a la persona, tal como se definió en los requerimientos de modelado de conocimiento.

## 4.9. Comprobación de inferencias *sameAs* en registros de estudiante y profesor

Dentro de los canales de población de instancias en la red ontológica para la Universidad Politécnica Salesiana surge un caso muy particular que provoca la duplicación de perfiles en la base de conocimiento. Se trata de las personas que, tras haberse formado en la UPS, se incorporan a la planta docente para participar de nuevos procesos académicos con el rol de profesor. Esto provoca que se duplique información y se pierda trazabilidad de una persona.

Como ejemplo del tipo de consolidaciones de conocimiento que habilita la tecnología semántica, podemos comprobar que se pueden unificar nodos afectados por este problema mediante la propiedad de datos *owl:sameAs*. Así, la Figura 4.16 muestra un fragmento de código SPARQL que construye un subgrafo que se ejecuta en el repositorio de GraphDB, generando sentencias con la relación *owl:sameAs* entre los nodos de la clase *foaf:Person* que tengan la misma cédula de identidad, denotada por la propiedad de datos *aeo:identifier*.

El subgrafo generado se carga al repositorio de conocimiento y, como se presenta en la Figura 4.17, dos grafos iniciales —correspondientes a una misma persona primero en rol de estudiante y luego en rol de docente— se fusionan en un único perfil. Por este procedimiento se consolidaron los registros de 444 profesores, equivalentes al 16% del plantel docente histórico de la UPS.



```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
CONSTRUCT {
    ?p1 <http://www.w3.org/2002/07/owl#sameAs> ?p2.
}
where {
    ?p1 rdf:type foaf:Person;
    <http://www.ups.edu.ec/academic/ontologies#RAC0000004> ?c1.
    ?p2 rdf:type foaf:Person;
    <http://www.ups.edu.ec/academic/ontologies#RAC0000004> ?c2.
    ?c1 <http://www.ups.edu.ec/academic/ontologies#identifier> ?ced1.
    ?c2 <http://www.ups.edu.ec/academic/ontologies#identifier> ?ced1.
    filter(?c1!=?c2)
    filter(?p1!=?p2)
}

```

Figura 4.16: Código SPARQL de inferencia de fusión de perfiles.

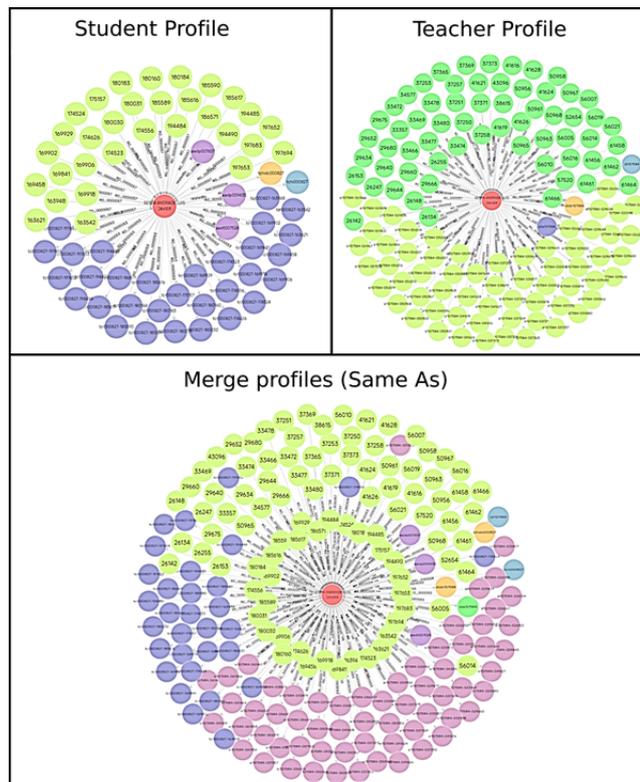


Figura 4.17: Visualización de la fusión de perfiles disjuntos de una persona que cumplió el rol de estudiante y profesor.

## 4.10. Sumario

El presente capítulo abordó la construcción de una red ontológica orientada al ámbito de las instituciones de Educación Superior, con capacidad de gestionar el conocimiento generado en las mismas de una manera más completa y eficiente en comparación con soluciones presentadas en el pasado, dando cuenta de forma explícita del desarrollo que los distintos actores del ecosistema toman a lo largo del tiempo y manejando perfiles que consideran información personal, académica, psicológica y socioeconómica.

En primer lugar, cabe destacar que los fundamentos de la metodología NeOn han permitido cubrir el ciclo de vida de nuestro modelo formal, aplicando como punto de referencia la técnica de entrevista en base a preguntas de competencia, con las cuales se identificaron los requerimientos funcionales que dieron pie a la conceptualización de requisitos, usuarios y usos previstos de la red ontológica.

La evaluación y reutilización de varios recursos ontológicos que cubren los requerimientos del dominio académico ha sido una constante. Así, se incorporaron recursos tales como vCard, FOAF y VIVO. Asimismo, la reingeniería de recursos no ontológicos, conjuntamente con el necesario proceso de alineación, ha permitido referir recursos adicionales bajo el empleo de criterios y métricas de selección, de una forma clara y eficiente.

En la migración de las bases de datos de la Universidad Politécnica Salesiana a los esquemas de la nueva red ontológica, hemos abordado algunas problemáticas de la información distribuida en varias bases de datos de sistemas transaccionales, en donde la interpretación de esa información se presta a la ambigüedad, por medio de procesos de mapeo. Estos procesos, además de permitirnos la coexistencia de información de diferentes fuentes, también nos ayudó a demostrar las posibilidades de consolidación de conocimiento, como ilustramos con la fusión de los dos perfiles que representaban a docentes que habían sido anteriormente estudiantes de la UPS. También dimos debida cuenta de aspectos de escalabilidad, a través de la especialización de un dialecto específico de OWL.

Todos los aportes teóricos de este capítulo, junto con el trabajo de creación de la base de conocimiento semántico para la UPS, han sustentado nuestra investigación en cuanto a la aplicación de técnicas de Inteligencia Artificial para mejorar la gestión del conocimiento en el entorno académico, como se detallará en los siguientes capítulos.



## Capítulo 5

# Aportes a la utilización de Inteligencia Artificial para formación de grupos de aprendizaje colaborativo y predicción de rendimiento académico

*En este capítulo desarrollamos las propuestas de aplicación de tecnologías de Inteligencia Artificial, sobre la red ontológica presentada en el capítulo anterior, a fin de contribuir al estado del arte en formación de grupos de aprendizaje colaborativo y a la predicción de rendimiento académico. Concretamente, se plantean nuevos enfoques de algoritmos genéticos y la utilización de técnicas de aprendizaje profundo, que se han validado en un entorno real con la participación de docentes y estudiantes universitarios. Se describen aquí los aspectos técnicos, mientras que la validación se tratará en el Capítulo 6.*

### 5.1. Introducción

La red ontológica presentada en el Capítulo 4 sienta unas bases sólidas para el desarrollo de numerosas aplicaciones de tecnologías de Inteligencia Artificial a la realidad específica de las instituciones de Educación Superior. En este capítulo desarrollamos nuestras propuestas en dos áreas diferentes, funcionando dentro de una misma arquitectura de servicios:

- Por una parte, presentamos un enfoque basado en algoritmos genéticos cuyo objetivo es formar grupos de aprendizaje colaborativo heterogéneos, procesando para ello información sobre los rasgos de personalidad y el rendimiento académico de los estudiantes. Para ello, en primer lugar, repasamos la literatura especializada en cuanto a la evidencia de que los grupos heterogéneos

potencian el rendimiento individual y grupal. Luego presentamos los detalles de nuestro algoritmo genético, diseñado para asegurar la heterogeneidad de los grupos formados.

- Por otra parte, presentamos un estudio sobre la utilización de redes neuronales en configuración de aprendizaje profundo para comprobar la viabilidad de hacer predicciones de rendimiento académico sobre la base de conocimiento resultante de incorporar a nuestra red ontológica los registros académicos acumulados durante más de una década en la Universidad Politécnica Salesiana. Nuestra exposición desarrollará los aspectos relativos al análisis de ese conocimiento, al proceso de normalización de los datos y de configuración de la red neuronal.

Previo a la descripción de estas propuestas, a continuación presentamos la arquitectura de servicios que da soporte a ambas, que desarrolla los esquemas del gestor de conocimiento para la red ontológica OEA (Sección 4.4) desde el punto de vista de su implementación modular, más que de sus capas y niveles de servicio como mostraba la Figura 4.3. Más adelante, en el Capítulo 6 reportamos los resultados de una serie de experimentos, llevados a cabo con la participación de docentes y estudiantes Universitarios, para validar las dos propuestas en diversos escenarios.

## 5.2. Arquitectura del sistema

La arquitectura de nuestro sistema híbrido se presenta en la Figura 5.1, conformada por 4 módulos principales, que se describen a continuación.

- **Interfaz de usuario.** Es el componente que permite registrar toda la información que se genera mediante la interacción de los diferentes usuarios de la plataforma: psicólogos, profesores, alumnos y analistas de datos.
- **Módulo de gestión del perfil del estudiante.** Cumple con la función de realizar un análisis completo del perfil del estudiante en base a tres criterios: perfil académico, perfil psicológico y perfil socio-económico.
- **Módulo de conocimiento.** Permite gestionar la información relacionada a los perfiles de cada estudiante, así como el levantamiento y almacenamiento de la información asociada a las tres partes antes mencionadas.
- **Sistema experto.** Contiene un conjunto de submódulos cuyo objetivo es proveer soporte para la toma de decisiones dentro del contexto de aprendizaje colaborativo, la planificación de actividades, la conformación de los grupos de aprendizaje y la realización de predicciones de rendimiento académico.

Como una estrategia de representación visual, en la Figura 5.1 se han empleado diferentes colores para denotar información relevante:

- El color verde se utiliza para indicar que un submódulo fue desarrollado, se encuentra operativo y ha sido parte del proceso de experimentación. Tal es el caso, en el módulo de conocimiento, del submódulo de rasgos de personalidad, del submódulo del perfil académico y el submódulo del perfil socio-económico

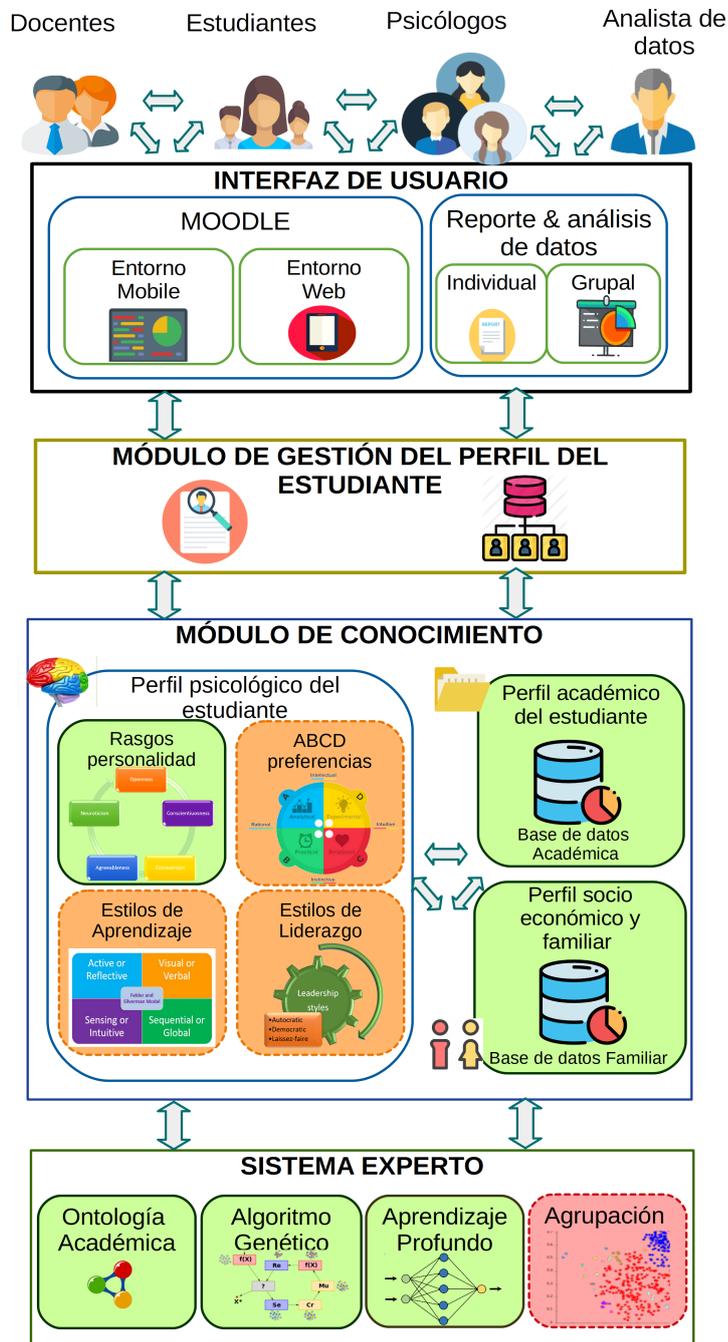


Figura 5.1: Arquitectura del sistema.

y familiar, y en el módulo del sistema experto, del submódulo de la ontología académica, del submódulo de los algoritmos genéticos y del submódulo de aprendizaje profundo.

- El color naranja indica que el submódulo fue desarrollado, se encuentra operativo y se dispone de datos para él, pero aún no se ha incorporado en él ninguna estrategia de Inteligencia Artificial, como sucede, en el módulo de gestión del conocimiento, con el submódulo de estilos de aprendizaje, el submódulo de estilos de liderazgo y el submódulo de dominancia cerebral.
- Finalmente, el color rojo representa submódulos para los cuales se ha definido un encaje en el marco global, pero cuyo desarrollo se ha dejado para trabajo futuro, considerando su relevancia para nuevos procesos de experimentación. Tal es el caso, en el módulo del sistema experto, del submódulo de agrupación (*clustering*).

A continuación describimos con algo más de detalle cada módulo.

### 5.2.1. Interfaz de usuario

El bloque de interfaces de usuario permite la interacción con los diferentes actores del sistema académico (psicólogos educativos, docentes, estudiantes y analistas de datos), a través de dispositivos electrónicos que soporten conectividad web y compatibilidad con la plataforma Moodle<sup>1</sup>.

Hemos empleado Moodle como componente básico de gestión para los procesos de trabajos, actividades, foros y aplicación de registro y recopilación de los resultados académicos. También lo utilizamos para el desarrollo y la aplicación de los cuestionarios psicológicos (véase captura del test NEO-PI-R en la Figura 5.2) y la visualización de predicciones de rendimiento.

Por otra parte, está el componente responsable del reporte y análisis de los datos. Este componente permite recuperar, analizar y generar informes sobre los datos de los estudiantes en base a los cuestionarios aplicados, ya sea de manera individual o grupal, y cuenta con diversas opciones de gestión visual de la información asociada a los aspectos psicológicos de los estudiantes. Por ejemplo, en la Figura 5.3, se pueden apreciar los resultados psicológicos de un estudiante en particular, mientras que la Figura 5.4 muestra la ponderación general del aspecto psicológico de un curso.

Finalmente, hemos creado también una interfaz como herramienta de apoyo para que el docente pueda generar grupos de aprendizaje colaborativo en función de un criterio seleccionado (véase Figura 5.5) y otra para que un analista de datos pueda visualizar estadísticas de rendimiento académico.

### 5.2.2. Gestión del perfil del estudiante

El módulo de gestión del perfil del estudiante, en primera instancia, se emplea como responsable de estandarizar de una manera fácil y eficiente la información que ingresa de las actividades realizadas por los estudiantes con los detalles necesarios (tests aplicados, actividades colaborativas, datos referentes a las cualidades de los

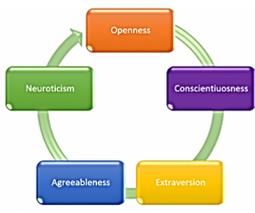
---

<sup>1</sup><https://moodle.org>

### Test de Personalidad

Continuar

Modo: Los nombres de los usuarios se mostrarán y registrarán con las respuestas



El propósito de esta encuesta es ayudarnos a conocer su rasgo de personalidad predominante y como potencializar su aprendizaje en función del mismo. Cada una de las 120 cuestiones siguientes le preguntará sobre su criterio relacionado aspectos de su comportamiento. No hay respuestas 'correctas' o 'erróneas': solo queremos su opinión. Le garantizamos que sus opiniones serán tratadas con el mayor grado de confidencialidad y no afectarán a su evaluación.

Se recomienda seleccionar *la primera idea que se le venga a la mente*, para que el resultado sea más acorde a su real proceder. Sus respuestas, nos ayudarán a mejorar el proceso de enseñanza académico.

Muchas gracias..!

Todas las preguntas son necesarias y deben ser contestadas

A continuación se presenta una escala informativa de ponderación:

**Ponderación**

1. Totalmente en desacuerdo    2. En desacuerdo    3. Ni de acuerdo; ni en desacuerdo    4. De acuerdo    5. Totalmente de acuerdo

Me preocupo por las cosas\*

1  2  3  4  5

Creo que hago amigos fácilmente\*

1  2  3  4  5

Figura 5.2: Captura de uno de los cuestionarios ofrecidos sobre Moodle.

grupos, etc.). Esta información será procesada y almacenada mediante la interacción entre el bloque de interfaz de usuario y el módulo de gestión del conocimiento.

En segundo instancia, este módulo se encarga de extraer e interpretar los resultados generados por la capa de conocimiento, procesarlos y enviarlos a formatos que el bloque de construcción de la interfaz puede reconocer. Finalmente, también es responsable de normalizar la información en esquemas comunes, permitiendo su uso posterior para la síntesis y análisis de datos.

### 5.2.3. Gestión del conocimiento

El módulo de gestión del conocimiento se encarga de la logística de integrar la información de un estudiante asociada a su comportamiento, rendimiento académico e información familiar. Está conformado por 3 submódulos:

- **El submódulo del perfil psicológico del estudiante** tiene el propósito de unificar la información proveniente de varios cuestionarios que evalúan la psique de un estudiante y formalizarlo por medio de una estructura, utilizando diferentes métricas establecidas en escalas específicas que se proponen en diferentes cuestionarios. La formalización de este conocimiento de aspectos psicológicos es genérica y se puede utilizar para múltiples propósitos.

A continuación describimos brevemente la información asociada a cada uno de los cuestionarios:

- **El test de rasgos de personalidad** contiene valoraciones de los rasgos de personalidad que define el denominado “*modelo de los 5 factores*” (*Big 5*) [218], que pasa por ser uno de los marcos más utilizados entre

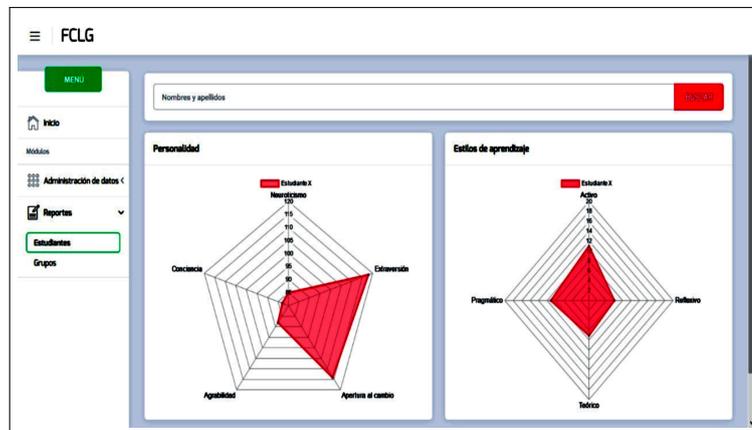


Figura 5.3: Visualización de los resultados psicológicos de un estudiante.

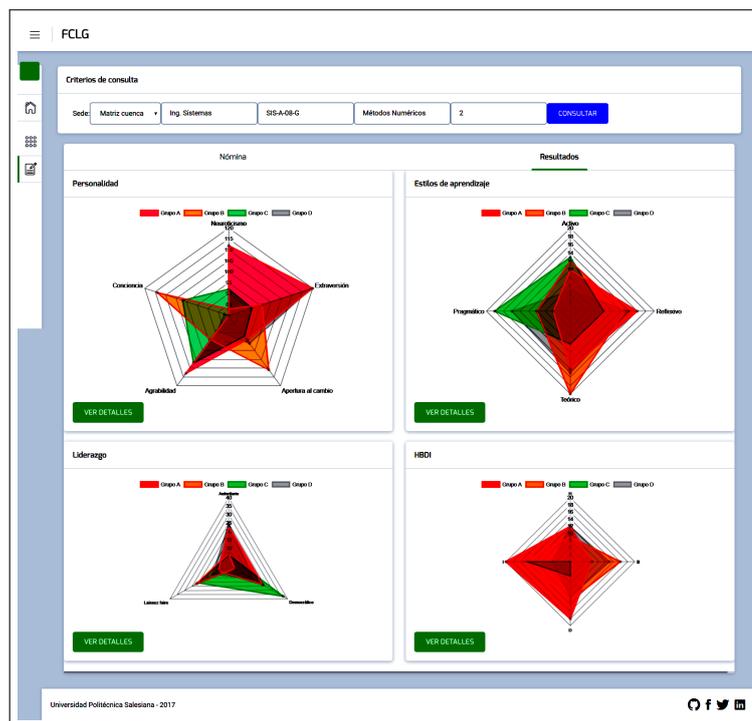


Figura 5.4: Visualización de ponderación de los aspectos psicológicos de un curso.

The screenshot shows the FCLG system interface. At the top, there are search criteria: 'Sede: Matriz cuenca', 'Ing. Sistemas', 'SIS-A-08-G', 'Métodos Numéricos', and '2'. A 'CONSULTAR' button is visible. Below the search criteria, there are two main sections: 'Nómina' and 'Resultados'.

**Nómina de estudiantes**

#	Nombres	Apellidos	Email
1	Rudhy Alejandra	Alarcón Murthy	rualea@yandes.ri
2	Paúl Esteban	Arivaldo Patiño	parevalo@gmail.com
3	Stephany Alejandra	Buastán Guerrero	staph@hotmail.com
4	Pedro Ignacio	Cabrera Zárate	pcabrera@gmail.com
5	José Estean	Calle Chuchuca	jcallec7@est.ups.edu.ec
6	Pablo Sebastian	Calle Ortega	pcallec@est.ups.edu.ec
7	Francis Mateo	Cordova Alvarez	fordova@est.ups.edu.ec
8	Verónica Alejandra	Crespo Camión	alecra@hotmail.com
9	Karen Estefanía	García Calle	kest@gmail.com
10	Daniel Paul	Hurdiales Fajardo	hdaniel@hotmail.com
11	Cristina Paola	Ingavelez Pintado	hdaniel@hotmail.com
12	Jonathan Oswaldo	Matute Curillo	jmatutec@est.ups.edu.ec
13	Domenica Alejandra	Merchan Garcia	dmerchang1@est.ups.edu.ec
14	Mark Daniel	Orellana Cargio	morellanac@est.ups.edu.ec
15	Luis Rene	Panjon Barreto	lpanjon@est.ups.edu.ec
16	Juan Fernando	Perez Sisalima	perez@gmail.com
17	Jorge Vinicio	Pizaro Romero	jpizaror@est.ups.edu.ec
18	Christian Geovanny	Rivera Loja	crivera2@est.ups.edu.ec
19	Yandry Daniel	Romero Carrion	yromeroc@est.ups.edu.ec
20	Paúl Andrés	Vasquez Arias	pvasequez@gmail.com

**Grupos conformados**

#	Nombres	Apellidos	Grupo
2	Paúl Esteban	Arivaldo Patiño	A
3	Stephany Alejandra	Buastán Guerrero	A
17	Jorge Vinicio	Pizaro Romero	A
4	Pedro Ignacio	Cabrera Zárate	A
20	Paúl Andrés	Vasquez Arias	A
5	José Estean	Calle Chuchuca	B
1	Rudhy Alejandra	Alarcón Murthy	B
13	Domenica Alejandra	Merchan Garcia	B
19	Yandry Daniel	Romero Carrion	B
8	Verónica Alejandra	Crespo Camión	B
9	Karen Estefanía	García Calle	C
15	Luis Rene	Panjon Barreto	C
10	Daniel Paul	Hurdiales Fajardo	C
12	Jonathan Oswaldo	Matute Curillo	C
6	Pablo Sebastian	Calle Ortega	C
7	Francis Mateo	Cordova Alvarez	D
14	Mark Daniel	Orellana Cargio	D
16	Juan Fernando	Perez Sisalima	D
11	Cristina Paola	Ingavelez Pintado	D
18	Christian Geovanny	Rivera Loja	D

Universidad Politécnica Salesiana - 2017

Figura 5.5: Organización de los grupos de aprendizaje colaborativo.

los psicólogos para representar a la psique humana en la medida de las dimensiones de apertura a la experiencia, conciencia, extraversión, amabilidad y neuroticismo. Específicamente, este test fue el empleado en el proceso de experimentación que se desarrolla más adelante en este capítulo, por lo cual realizaremos una descripción más minuciosa en la Subsección 5.4.2.

- El **test de estilos de aprendizaje** permite identificar y categorizar a cada estudiante conforme la manera en que recibe y procesa la información [60], en función de 4 dimensiones con 2 valores posibles cada una:
  - Procesamiento: activo/reflexivo.
  - Percepción: sensitivo/intuitivo.
  - Ingreso de información: visual/verbal.
  - Comprensión: secuencial/global.
- El **test de estilo de liderazgo** se enfoca a asociar conductas de comportamiento y dinámicas de interacción [154], permitiendo registrar el estilo de liderazgo predominante entre los tres tipos: autocrático, democrático o *laissez-faire*.
- Finalmente, el **test de preferencias o dominancia cerebral** se derivó del test ABCD [264], que ilustra y explica la forma en que un individuo prefiere pensar, aprender, comunicar y tomar decisiones. Nos permite entender por qué algunos elementos de la educación o el trabajo son

simples y satisfactorios, y por qué algunas materias académicas son más interesantes y atractivas.

- El **submódulo del perfil académico del estudiante** tiene el objetivo de manejar un historial académico del desempeño del estudiante en su proceso de formación. A través de este historial, es posible conocer y segregar el desempeño de cada estudiante tanto a nivel micro-curricular (desempeño por materia) como macro-curricular (desempeño de promedio general).
- El **submódulo del perfil personal, socio económico y familiar del estudiante** se enfoca en fusionar la información referente al perfil del contexto familiar y socio-económico en el que se desenvuelve cada estudiante. Este submódulo nos permite evaluar atributos externos al ecosistema académico, que condicionan la formación integral de cada alumno, tales como la situación económica familiar, el número de miembros del hogar, las fuentes de ingresos, la ubicación domiciliaria, la etnia o el estado civil, entre otros.

#### 5.2.4. Sistema experto

El bloque del sistema experto se centra en emplear diferentes técnicas de Inteligencia Artificial en conexión con los demás módulos. Se incluyen las siguientes contribuciones:

- El uso de ontologías para el modelado del conocimiento, ampliando lo explicado en el Capítulo 4 en relación con la inferencia de rasgos de personalidad (**submódulo ontológico**, Sección 5.3).
- El empleo de algoritmos genéticos para la conformación de los grupos de aprendizaje basados en los atributos de los estudiantes (**submódulo de algoritmos genéticos**, Sección 5.4).
- La aplicación de técnicas de aprendizaje profundo para realizar predicciones de rendimiento académico (**submódulo de aprendizaje profundo**, Sección 5.5).

### 5.3. Componente de la red ontológica OEA para la inferencia de rasgos de personalidad

El componente de la red ontológica OEA para rasgos de personalidad tiene el propósito de unificar la información proveniente de varias pruebas que miden los rasgos de personalidad y formalizarlo por medio de una estructura, desde una perspectiva holística. Este componente describe el comportamiento de un individuo basado en diferentes métricas establecidas en escalas específicas, que se proponen en diferentes pruebas.

Basándonos en la metodología NeOn, tomamos los resultados de la aplicación del cuestionario IPIP-NEO-120 [134] como un recurso base en la fase de obtención de requisitos mediante *preguntas de competencia* (QC). A partir de los requisitos resultantes —validados por especialistas en el área de psicología— buscamos las estrategias para la formalización del modelo.

Como se expuso en la Sección 4.2, el lenguaje que se eligió para transformar las diferentes conceptualizaciones de los perfiles de personalidad fue OWL, y se reutilizaron elementos de estándares pre-existentes, tales como FOAF (utilizado para describir personas y sus relaciones), vCard (para anotar información de individuos y organizaciones), SKOS (para representar taxonomías y esquemas de clasificación) y BFO (para aportar definiciones de roles y axiomas de consistencia).

### 5.3.1. Formalización del componente ontológico

El universo del discurso  $D$ , tal como se presenta en la Ecuación 5.3.1, contiene todos los elementos creados como instancias en el contexto de nuestra red ontológica presente en los perfiles de información general y rasgos de personalidad, así como los procesos encontrados en las evaluaciones de las diferentes cuestionarios que aplicaron a los estudiantes.

$$D = \{person_1, individual_1, Neuroticism, Anxiety, Facet, Domain, IPIP - NEO - 120, \dots\} \quad (5.3.1)$$

Las principales relaciones unitarias definidas en esta sección de nuestra red ontológica son las siguientes:

- **Persona:** es una representación de una persona, reutilizada de la ontología de FOAF.
- **Individuo:** es una representación de una persona o una entidad, reutilizada desde la ontología vCard.
- **Concepto:** representa uno de los conceptos de una taxonomía o esquema de de clasificación, reutilizado desde el esquema ontológico SKOS.
- **Esquema de concepto:** representa un esquema al que pertenecen varios conceptos relacionados, reutilizado desde la ontología SKOS.
- **Perfil de personalidad:** es una abstracción de un registro de múltiples valores relacionados con los rasgos de personalidad y facetas de un estudiante.
- **Cuestionario:** representa un documento independiente continuo que contiene parámetros para ser evaluados por un estudiante.
- **Actividad:** representa un conjunto de acciones desarrolladas por un individuo.
- **Acción:** describe interacciones realizadas entre individuos y otros elementos continuos que causan un resultado.
- **Ítem de actividad:** indica el nombre dado a una actividad.
- **Rol:** representa función realizada por una persona en una actividad, reutilizada desde la ontología BFO.
- **Respuesta:** representa una posible opción que pertenece a una pregunta.

- **Pregunta:** representa una pregunta que pertenece a una prueba específica.

Por otra parte, las relaciones binarias que fueron modeladas son las siguientes:

- **tiene perfil de personalidad:** indica la pertenencia de un perfil de personalidad a una persona.
- **prueba relacionada:** especifica que un objeto está relacionado con un cuestionario (en nuestra representación formal se usa para relacionar el perfil de personalidad con un cuestionario específico).
- **desarrolla:** desarrollo de una actividad por un rol inherente a un individuo.
- **tiene acción:** pertenencia de una acción a una actividad.
- **tiene puntuación:** relacionada con la puntuación de un objeto.
- **tiene escala de unidades de medida:** unidad de medida que pertenece a un puntaje.

Con base en las definiciones anteriores, nuestro conjunto de relaciones  $R$  se define en la Ecuación 5.3.2.

$$\begin{aligned}
 R = \{ & \text{Agent, Person, Individual, Role, Student - Role,} \\
 & \text{Teacher - Role, Activity, Action, Item - Activity,} \\
 & \text{Personality - Profile, Score, Partial - Score,} \\
 & \text{Total - Score, Test, Concept, Measure - Unit - Scale,} \\
 & \text{Answer, Question, Occurrent,} \\
 & \text{Informal - Content - Entity, Contact,} \\
 & \text{Individual, Personality - Profile,} \quad (5.3.2) \\
 & \text{has - contact - info, has - personality - profile, inheres - in,} \\
 & \text{bearer - of, develops, develops - by, has - action,} \\
 & \text{has - item - activity, item - activity - for, has - score, score - of,} \\
 & \text{related, related - test, has - measurement - unit - scale} \}
 \end{aligned}$$

En la Ecuación 5.3.3 llevamos a cabo la especificación de las subclases de las relaciones unarias:

$$\begin{aligned}
 O_0 = \{ & \text{Person}(x) \rightarrow \text{Agent}(x), \\
 & \text{Student - Role}(x) \rightarrow \text{Role}(x), \\
 & \text{Teacher - Role}(x) \rightarrow \text{Role}(x), \\
 & \text{Partial - Score}(x) \rightarrow \text{Score}(x), \\
 & \text{Total - Score}(x) \rightarrow \text{Score}(x), \quad (5.3.3) \\
 & \text{Answer}(x) \rightarrow \text{Informal - Content - Entity}(x), \\
 & \text{Question}(x) \rightarrow \text{Informal - Content - Entity}(x), \\
 & \text{Action}(x) \rightarrow \text{Occurrent}(x), \\
 & \text{Activity}(x) \rightarrow \text{Occurrent}(x), \\
 & \text{Personality - Profile}(x) \rightarrow \text{Contact}(x) \}
 \end{aligned}$$

Por otra parte, la Ecuación 5.3.4 nos permite realizar la especificación de dominios y rangos de relaciones binarias:

$$\begin{aligned}
O_1 = O_0 \cup \{ & \text{has-contact-info}(x, y) \rightarrow \text{Thing}(x) \wedge \\
& \text{Contact}(y), \text{has-personality-profile}(x, y) \rightarrow \text{Thing}(x) \wedge \\
& \text{Personality-Profile}(y), \text{has-activity}(x, y) \rightarrow \text{Thing}(x) \wedge \\
& \text{Activity}(y), \text{has-action}(x, y) \rightarrow \text{Thing}(x) \wedge \\
& \text{Action}(y), \text{has-score}(x, y) \rightarrow \text{Thing}(x) \wedge \text{Score}(y), \\
& \text{related}(x, y) \rightarrow \text{Thing}(x) \wedge \text{Concept}(y), \text{related} \\
& \text{Test}(x, y) \rightarrow \text{Thing}(x) \wedge \text{Test}(y), \text{has-measurement-unit} \\
& \text{scale}(x, y) \rightarrow \text{Thing}(x) \wedge \text{Measurement-Unit-Scale} \}
\end{aligned} \tag{5.3.4}$$

En la Ecuación 5.3.5 realizamos la especificación de las disyunciones de las relaciones unarias:

$$\begin{aligned}
O_1 = O_0 \cup \{ & \text{Partial-Score}(x) \rightarrow \neg \text{Total-Score}(x), \\
& \text{Answer}(x) \rightarrow \neg \text{Question}(x), \\
& \text{Action}(x) \rightarrow \neg \text{Activity}(x), \\
& \text{Personality-Profile}(x) \rightarrow \neg \text{Individual}(x)
\end{aligned} \tag{5.3.5}$$

Para la construcción de nuestra pieza ontológica asociada al aspecto psicológico del estudiante utilizamos la herramienta Protégé, definiendo las diferentes clases, relaciones y axiomas. Teniendo en cuenta la construcción modular general de nuestra red ontológica OEA, fueron separados los diferentes módulos que contienen las definiciones ya formalizadas. En la Figura 5.6, en el lado izquierdo podemos ver el módulo específico del cuestionario IPIP-NEO-120, que modela las preguntas del formulario, y en el lado derecho las acciones en que los actores interactúan con el cuestionario. Aquí podemos observar el módulo de rasgos de personalidad, que consiste en los componentes centrales, con los que el resto de los módulos están alineados consiguiendo generar una integridad en la red ontológica.

Finalmente, para demostrar la trazabilidad de un individuo, hemos probado nuestra red ontológica almacenándola en la base de triplas Allegro Graph, poblando las instancias mediante un proceso de mapeo basado en una hoja de cálculo con las evaluaciones de los estudiantes que respondieron al cuestionario IPIP-NEO-120. Mediante una consulta SPARQL como la de la Tabla 5.1 obtuvimos los resultados de trazabilidad de sus preguntas.

Tabla 5.1: Resultado de la consulta SPARQL asociada al cuestionario IPIP-NEO-120.

<b>Puntaje_Nodo</b>	<b>Etiqueta_NodoPregunta</b>	<b>Etiqueta_NodoFaceta</b>
5	“Me gusta llevar a cabo mis planes”	Autodisciplina
4	“Tomo el control de las cosas”	Asertividad
4	“Yo realizo mis actividades sin ningún problema”	Autodisciplina
3	“No entiendo a la gente que se emociona fácilmente”	Emocional

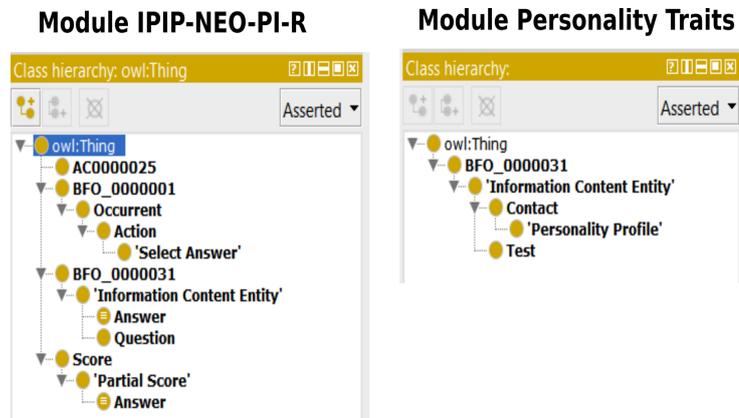


Figura 5.6: Módulos correspondientes al componente psicológico de la red ontológica.

Se pudo observar como funcionaba correctamente la coexistencia de diferentes contextos de los módulos referente al componente psicológico de nuestra red ontológica OEA basados en las triplas de Allegro Graph junto con la población de instancias en la fase de prueba. Esta información almacenada semánticamente será el insumo que alimente el algoritmo genético para la eficiente creación de grupos de trabajo colaborativo, como se expone en la siguiente sección.

#### 5.4. Algoritmo genético para formación de grupos de aprendizaje colaborativo basado en rasgos de personalidad

Como se explicó en el Capítulo 3, el aprendizaje colaborativo se define como el proceso educativo de organizar pequeños grupos para ayudar a los estudiantes a trabajar juntos y potenciar su propio aprendizaje, proporcionando así un medio para el desarrollo del pensamiento crítico y la resolución de problemas, a la vez que se fomentan las habilidades de relación interpersonal. Varios trabajos han permitido corroborar que los estudiantes que se preparan en grupos tienden a tener una mayor capacidad para resolver problemas, presentan un aprendizaje significativo de la materia que se está enseñando, e incluso lo retienen durante más tiempo que cuando se presenta el mismo contenido mediante otros mecanismos de instrucción (véase [49, 235, 258]). Otros trabajos también han demostrado que mediante esta metodología los estudiantes tienden a incrementar el pensamiento crítico, las habilidades sociales y las capacidades de aprendizaje meta-cognitivas (i.e. el *aprender a aprender*). [55, 161, 190]

Antes de desarrollar nuestra propuesta tecnológica para asistir en la formación de grupos, describiremos criterios y prácticas comunes, junto con las principales características del modelo concreto de representación utilizado en el submódulo ontológico descrito en la Sección 5.3.

### 5.4.1. Criterios para la formación de los grupos de aprendizaje colaborativo

En esta sección se aborda la categorización relacionada a la formación de los grupos de aprendizaje bajo tres criterios: (i) el tamaño de los grupos, (ii) la potenciación de la heterogeneidad intra-grupos y la homogeneidad inter-grupos y (iii) la estrategia de asignación de estudiantes a grupos.

#### 5.4.1.1. Sobre el tamaño de los grupos

Son varios los trabajos de investigación que han recomendado un tamaño ideal de grupo de 3 a 5 estudiantes. [59] Un tamaño de 4 se recomienda especialmente por varias razones, entre las que se incluyen las siguientes:

- Durante el trabajo en grupo, a los estudiantes les resulta más viable organizar reuniones, debido a que es más fácil coordinar los horarios.
- Los estudiantes se pueden distribuir parte del trabajo que deben realizar, con lo cual sienten que pueden hacer una contribución significativa al trabajo del grupo, y luego transferir e integrar su conocimiento al grupo.
- Los integrantes son más responsables y visibles entre sí, lo cual frecuentemente disminuye los inconvenientes relacionados con la retirada de esfuerzo.

Por su parte, Sharan [230] también realizó sugerencias sobre lo que denominó el tamaño óptimo de un grupo basado en los objetivos de aprendizaje:

- **Grupos de 2 estudiantes:** son ideales para resolver ejercicios de destrezas o habilidades, lo cual implica tareas a través de las cuales los estudiantes demuestran su comprensión del nuevo material. Los problemas de la tarea o las preguntas al final de un capítulo son ejemplos de ejercicios de destrezas. El trabajo en grupo se establece de tal manera que los estudiantes realizan algún grado de transferencia de conocimiento, pero los contextos no son completamente nuevos para los estudiantes.
- **Grupos de 3 estudiantes:** son apropiados para aplicarse en actividades que correspondan al aprendizaje guiado por descubrimiento. En este tipo de tareas se dispone un trabajo en grupo con la finalidad de que los estudiantes adquieran nuevos contenidos, mediante un conjunto de preguntas de consulta guiada y el uso de un modelo. Los estudiantes aprenden a través del descubrimiento en vez de recibir información directamente. No obstante, el método de descubrimiento no es completamente abierto (como en la investigación); la función del docente es proporcionar una base de recursos y guiar a los estudiantes a través del proceso.
- **Grupos de 4 estudiantes:** se emplean habitualmente para la resolución de problemas en clase, permite al instructor observar y evaluar las habilidades de resolución de problemas de los estudiantes. El trabajo en grupo está estructurado de tal manera que los estudiantes pueden practicar la solución de problemas junto con compañeros y el docente ofrece los comentarios oportunos para potencializar el desempeño.

- **Grupos de 5 estudiantes:** se aplican al establecer un proyecto de resolución de problemas a largo plazo. Este tipo de actividad suele consistir en un proyecto en el que el grupo dispondrá colectivamente entre 60 y 100 horas de trabajo. Tales proyectos requieren planificación al inicio, incorporación de muchos conceptos del curso y evaluación continua del desempeño por parte del equipo. El entregable del trabajo final más a menudo incluye un informe escrito detallado del desarrollo, así como una exposición o presentación.

#### 5.4.1.2. Sobre la heterogeneidad intra-grupos y la homogeneidad inter-grupos

Las sugerencias más ampliamente presentadas en la literatura sobre aprendizaje cooperativo son que la composición del grupo debe ser heterogénea intra-grupos y homogénea inter-grupos siempre que sea posible. En otras palabras, los estudiantes en grupos deben ser diversos en cuanto a antecedentes, ideas, personalidad, etnicidad y género. [160, 234, 238, 259]

Con el objetivo de incrementar la participación de los estudiantes y de propiciar la innovación, Romney [216], aplicó la estrategia de aprendizaje colaborativo para la traducción en un curso de Francés en Canadá. Acorde a su lógica, los grupos estuvieron conformados por 5 estudiantes, en base a los siguientes criterios:

- **Género:** aunque la mayoría de estudiantes fueron mujeres, en cada grupo hubo al menos un hombre.
- **Dominio del idioma:** cada grupo tenía por lo menos un miembro con habilidades nativas o casi nativas en francés, y uno cuyo primer idioma no fuese ninguno de los idiomas oficiales canadienses.
- **Personalidad:** no se asignó más de un estudiante extrovertido o introvertido en cada grupo de aprendizaje.
- **Factores individuales:** edad, trabajo y experiencia de vida.

Los grupos de aprendizaje organizados fueron tan heterogéneos intra-grupo y homogéneos inter-grupo como fue posible, permitiendo exhibir a los estudiantes a una amplia variedad de opiniones. La conclusión de Romney sobre los grupos explicita que, a nivel personal, los estudiantes estaban encantados de poder compartir sus dificultades con los demás. Adicionalmente, los estudiantes ganaron confianza al observar que si sus compañeros de equipo podían resolver problemas, también ellos podrían superarlos. El hecho de expresarse frente a un pequeño grupo, con el que estaban familiarizados, en lugar de estar frente a toda la clase, también les permitió sentirse menos estresados. Paralelamente esta experiencia les permitió establecer amistades cercanas con sus compañeros de grupo.

En relación con la agrupación basada en la capacidad, Srba [234] argumentó que la agrupación por capacidad era un tema complejo y, a menudo, divisivo en la educación. La agrupación heterogénea es necesaria para garantizar la igualdad de oportunidades para todos los estudiantes. Los estudiantes que se atascan en pistas de bajo nivel se ven privados de oportunidades para desarrollar habilidades de alto nivel. Bajo estas premisas, un conjunto de especialistas en agrupación de habilidades generó las siguientes recomendaciones sobre cómo agrupar a los estudiantes:

- Los grupos heterogéneos son idóneos cuando los estudiantes desarrollan tareas abiertas de resolución de problemas o actividades de investigación científica.
- También es apropiado que los estudiantes trabajen en grupos heterogéneos cuando han de discutir conceptos que son nuevos para todos ellos.
- Todos los estudiantes deben aprender las habilidades de trabajar juntos antes de que las actividades de aprendizaje colaborativo tengan éxito.
- Las estrategias de agrupación deben ser flexibles y los estudiantes deben poder trabajar independientemente, al menos ocasionalmente, de acuerdo a sus preferencias.
- Los estudiantes deben tener la oportunidad de seleccionar sus propios grupos en base a intereses.

#### 5.4.1.3. Sobre la asignación de miembros a grupos

Existen tres métodos comunes de asignación de miembros en grupos, a saber: asignación aleatoria, autoselección y criterios específicos. Los explicamos a continuación para sopesar las opciones para cada uno:

- **Asignación aleatoria.** Muchos docentes emplean algún método de asignación aleatoria para formar grupos. Dentro de los más comunes está el sistema denominado *call-off*: el profesor asigna a cada estudiante de la clase un número o una letra en una secuencia sistemática (por ejemplo: 1, 2, 3, 4, 5, ..., 1, 2, 3, 4, 5, ..., etc.), y los grupos se organizan al juntar todos los 1, 2, 3 etc.

El mecanismo de asignación aleatoria frecuentemente se utiliza debido a que se estipula que tiene varias ventajas como el no requerir ninguna preparación (lo que facilita su aplicación), el segregar los grupos de amistad (debido a que comúnmente los estudiantes se sientan con estudiantes que conocen, con lo cual generalmente caen en diferentes grupos), y el hacer posible que algunos estudiantes trabajen con otros que normalmente no escogerían. Sin embargo, el método posee algunos inconvenientes, empezando por que los estudiantes pueden no sentirse motivados (especialmente aquellos que conocen a otros compañeros en la clase y prefieren trabajar con ellos) o se preocupan por la posibilidad de ser asignados a un grupo con miembros con quienes tienen poca empatía. También pueden considerar que el docente ha utilizado la opción de formación más fácil disponible, llevándoles a pensar que no le interesa cómo se crean los grupos.

- **Autoselección.** En varias actividades, se les solicita a los estudiantes que formen grupos por sí mismos. Bajo estas condiciones, los estudiantes generalmente conocen a las personas en su clase y se decantan por trabajar con aquellos con quienes tienen amistad o con los que son más responsables. Por otra parte, aquellos estudiantes que no conocen a sus compañeros de la clase, tienden a formar grupos con los que están sentados cerca o con otros que pueden no conocer a nadie tampoco.

El método de autoselección es fácil de administrar y a los estudiantes les gusta la oportunidad de elegir a sus compañeros de grupo. Para muchos, es

la opción ideal. No obstante, puede ser difícil para los estudiantes que no conocen a nadie más en la clase y, a menudo, se considera que no es justo para todos.

- **Criterios específicos.** Este método busca específicamente organizar grupos heterogéneos. Se fundamenta en el paradigma de que los grupos funcionan mejor cuando los miembros están equilibrados. Algunos de los métodos más habituales pasan por emplear estilos de aprendizaje, estilos de liderazgo, roles funcionales o rasgos de personalidad. Estos sistemas involucran a los estudiantes, a quienes se les aplica un cuestionario. Los estudiantes con diferentes estilos son asignados a cada grupo para lograr el equilibrio deseado.

El método de criterios específicos dispone de varias ventajas, como que los estudiantes consideran que el método de selección es equitativo; se ven a sí mismos como valorados y están motivados para demostrar y ejercer sus destrezas; valoran las diferencias individuales y cómo la diversidad puede crear sinergias en el proceso de aprendizaje. Como inconveniente, el proceso puede ser costoso en tiempo, sobre todo si no se dispone de la información y una aplicación que lo automatice. El enfoque también puede dar como resultado grupos con características homogéneas, representación excesiva de algunas características o viceversa, lo cual puede causar problemas en la asignación.

#### 5.4.2. Descripción general del modelo *Big 5*

En este punto, es importante dar a conocer que en nuestra primera etapa del proceso de experimentación hemos aplicado los cuatro tests mencionados en la Subsección 5.2.3, pero por cuestiones de experimentación progresiva implementamos la conformación de los grupos utilizando únicamente los datos vinculados a los rasgos de personalidad. Para períodos posteriores se emplearán y se utilizarán los atributos relacionados a los otros aspectos. Por este motivo, incluimos a continuación una introducción al modelo de los cinco factores (*Big 5*), una taxonomía que describe la personalidad en base a cinco grandes *rasgos* o *dimensiones* [131,204,218]. Estos cinco factores a menudo se representan por los acrónimos OCEAN o CANOE [39]:

- O (“*openness*”): apertura a la experiencia.
- C (“*conscientiousness*”): consciencia o responsabilidad.
- E (“*extraversion*”): extraversión.
- A (“*agreeableness*”): amabilidad.
- N (“*neuroticism*”): neuroticismo o inestabilidad emocional.

Los cinco rasgos fueron recabados durante un estudio centrado en las descripciones que concebían unos individuos sobre la personalidad de otros [99], y han sido utilizados con profusión en multitud de áreas en el ámbito científico [61,134,189,226]. Como veremos a continuación, cada uno de los factores está constituido por un conjunto de facetas de personalidad más específicos [91,248]:

- El **factor O** clasifica al individuo como *creativo* y *curioso* frente a *consistente* y *cauteloso*. Alude también al apego por el arte, la emoción, las ideas inusuales,

la aventura, la curiosidad y la variedad de experiencias. La apertura evidencia el nivel de curiosidad intelectual, creatividad y una preferencia por la innovación y la variedad que posee una persona. Además, se describe como el grado en que una persona es creativa o independiente, presentando una predilección personal por una diversidad de actividades sobre una monotonía estricta. Un alto grado de apertura puede ser concebida como espontaneidad o falta de enfoque, y es más probable que se involucre en conductas de riesgo [12]. Los individuos que presentan un alto nivel de apertura tienden a inclinarse, en la ocupación y el pasatiempo, hacia las artes; son por lo general creativos y aprecian la importancia de las actividades intelectuales y artísticas [9]. Además, se dice que las personas con una gran apertura persiguen la autorrealización específicamente buscando experiencias intensas y eufóricas. A la inversa, aquellos con poca apertura buscan obtener satisfacción a través de la perseverancia y se caracterizan por ser pragmáticos e impulsados por los datos, a veces incluso percibidos como incuestionables y de mentalidad cerrada.

- El **factor C** hace alusión a ser *eficiente* y *organizado* frente a *despreocupado* y *descuidado*. Este rasgo está vinculado con la tendencia a ser organizado y confiable, a la muestra de autodisciplina, al actuar cautelosamente, enfocado al éxito y optar por una conducta planificada en lugar de espontánea. Un nivel alto de consciencia a menudo se percibe como ser obstinado y centrado. Por el contrario, la baja consciencia está asociada con la flexibilidad y la espontaneidad, pero también puede aparecer como falta de confianza, organización, compromiso y fiabilidad [248].
- El **factor E** clasifica como *sociable* y *enérgico* frente a *solitario* y *reservado*. Alude a la tendencia de buscar la estimulación en compañía de otros, con actitudes tales como el surgimiento, la asertividad, la sociabilidad y el alto grado de comunicación mediante el habla. Un alto grado de extraversión comúnmente se percibe como una búsqueda de atención e intención de mostrarse dominante. En cambio, la baja extraversión establece una personalidad reservada y reflexiva, que puede apreciarse como ser distante y frío..
- El **factor A** representa al *amigable* y *compasivo* en contraste con el *desafiante* y *desapegado*. Se refiere a la tendencia a ser compasivo y cooperativo en lugar de desconfiado y antagónico hacia los demás. También es una medida de la naturaleza confiada y útil de uno, y si una persona está generalmente de buen humor o no. La alta amabilidad se ve a menudo como ser ingenuo o sumiso; en cambio, las personalidades poco agradables suelen ser personas competitivas o desafiantes, que pueden considerarse argumentativas o poco fiables.
- El **factor N** define al *sensible* y *nervioso* frente al *seguro* y *confiado*. Observa la inclinación al estrés psicológico, así como la predisposición a percibir emociones desagradables fácilmente, como la ira, la ansiedad, la depresión y la vulnerabilidad. El factor de neuroticismo igualmente está relacionado con el nivel de estabilidad emocional y el autocontrol de los impulsos y, a menudo, se lo denomina por su extremo bajo, "*inestabilidad emocional*". El alto grado de estabilidad se manifiesta bajo una personalidad inalterable y tranquila, pero puede verse como poco inspiradora y despreocupada. La baja estabilidad se manifiesta como la personalidad reactiva y excitable que se encuentra a menudo en individuos dinámicos, pero puede percibirse como inestable o insegura.

Adicionalmente, se ha podido constatar que las personas con niveles altos de neuroticismo tienden a tener un peor bienestar psicológico [79].

De cara a la caracterización de la personalidad de un individuo, el modelo de los cinco factores aplica una técnica estadística denominada *análisis factorial* a los datos recopilados mediante una encuesta de personalidad, donde las palabras que se emplean para describir los aspectos de la personalidad permiten obtener los rasgos de cada individuo. Por lo tanto, esta teoría se basa en la relación entre las palabras seleccionadas, y no en experimentos neuropsicológicos.

### 5.4.3. Detalles del algoritmo propuesto

Para generar los grupos de aprendizaje, hemos implementado un algoritmo genético que realiza una consulta y emplea la información representada semánticamente por nuestro componente psicológico de la ontología OEA. En la versión cerrada en marzo de 2019, hacemos uso específico de la información vinculada al módulo *IPIP-NEO-120* asociado a los perfiles psicológicos de los estudiantes (expuesto en la Sección 5.3) y en base a ella determinamos las mejores combinaciones que lleguen a conformar grupos eficientes de trabajo colaborativo.

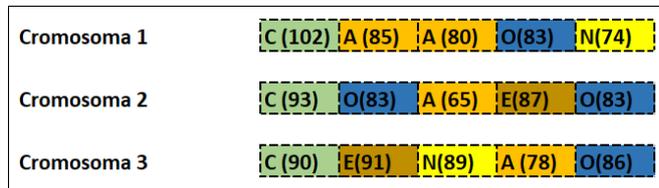
Los principales criterios considerados al experimentar con la composición del grupo fueron los que se detallan a continuación:

- Asignar los estudiantes con mayor puntaje en el rasgo de responsabilidad como coordinadores (cabeceras) de grupos (cromosomas) (Figura 5.7(a)).
- Evitar la composición de cromosomas en los cuales una mayoría de estudiantes (genes) tengan como predominante el rasgo de neuroticismo (Figura 5.7(b)).
- Evitar la composición de grupos en los cuales una mayoría de estudiantes tenga como predominante el rasgo de extroversión (Figura 5.7(c)).
- Evitar la composición de grupos de estudiantes en los cuales una mayoría de estudiantes tenga rasgos predominantes opuestos (Figura 5.7(d)).
- Intercambiar estudiantes con rasgos de personalidad predominante extroversión o neuroticismo, con rasgos de personalidad amables o de mente abierta (Figura 5.7(e)).
- En lo posible, integrar grupos de estudiantes totalmente heterogéneos en cuanto a sus rasgos de personalidad (Figura 5.7(f)).

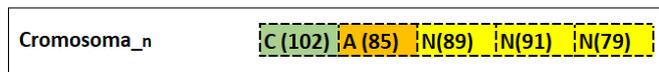
El vector de características que representa el perfil de cada estudiante se define a través de la Ecuación 5.4.1, donde  $id$  representa el identificador único del estudiante y  $O_{score}$ ,  $C_{score}$ ,  $E_{score}$ ,  $A_{score}$  y  $N_{score}$  representan los puntajes en los rasgos de apertura a la experiencia, responsabilidad o consciencia, extraversión, amabilidad y neuroticismo, respectivamente.<sup>2</sup>

$$\vec{P} = \{id, O_{score}, C_{score}, E_{score}, A_{score}, N_{score}\} \quad (5.4.1)$$

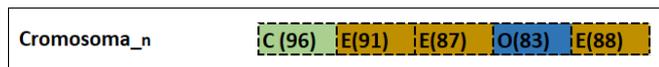
<sup>2</sup>De acuerdo con la prueba IPIP-NEO-120, el valor máximo para cada rasgo es 120.



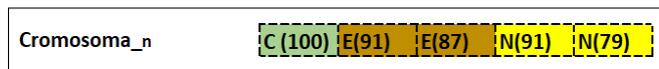
((a)) Estudiantes con rasgo de responsabilidad predominante son asignados como coordinador de grupo.



((b)) Cromosomas no aptos: genes con predominancia de neuroticismo.



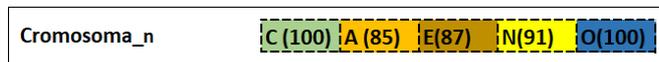
((c)) Cromosomas no aptos: genes con predominancia de extroversión.



((d)) Cromosomas no aptos: genes con predominancia de neuroticismo y/o extroversión.



((e)) Balance de cromosomas mediante intercambio de genes.



((f)) Cromosoma idóneo: combinación heterogénea de rasgos.

Figura 5.7: Ilustración de las reglas que rigen la composición de grupos.

Por otro lado, es importante mencionar que hemos definido una versión propia de los métodos de **inicialización** y **cruce** para el funcionamiento de nuestro algoritmo genético:

- Como método de inicialización, nuestro algoritmo ordena todos los perfiles de acuerdo con las puntuaciones C más altas (rasgo de responsabilidad o conciencia) y selecciona los  $N$  mejores perfiles principales, donde  $N$  representa el número requerido de grupos. Estos perfiles representan el gen que será el coordinador del cromosoma. Los genes requeridos para completar el cromosoma son agregados a través de una selección aleatoria de los perfiles restantes.
- El método de cruce, acorde a las reglas (duras y blandas) de la Figura 5.7, funciona como se describe a continuación:
  - El sistema identifica qué grupos están desequilibrados debido al predominio de rasgos opuestos, rasgo de neuroticismo o rasgos de extraversión. Este análisis se realiza a través de la Ecuación 5.4.2:

$$balance = N_{score} - (E_{score} + (K \cdot A_{score})) \quad (5.4.2)$$

El valor  $K$  es una constante para ponderar el rasgo de amabilidad, y se define en el rango  $[0, 1]$ . El equilibrio se determina sobre la base de que si un grupo tiene más predominio del rasgo de neuroticismo, es necesario encontrar el equilibrio a través de las personas que son más amables o agradables. De la misma manera, es fundamental considerar que el rasgo de extroversión se opone al neuroticismo.

- Una vez que se determina qué grupos están desequilibrados, el siguiente paso es determinar qué perfiles se pueden intercambiar por otro grupo con el objetivo de reducir la puntuación de equilibrio general. Esta puntuación se calcula utilizando una función de aptitud que evalúa en cada iteración si el balance ha mejorado o empeorado con respecto a la anterior.

## 5.5. Configuración de aprendizaje profundo para predicción de rendimiento académico

Encuadrado en el mismo marco que el proceso de generación de grupos de aprendizaje colaborativo, en esta sección describimos el módulo de aprendizaje profundo desarrollado para realizar predicciones sobre rendimiento académico. Concretamente, la propuesta pasa por realizar un estudio sobre la utilización del perceptrón multicapa (modelo de red neuronal descrito profusamente en la Capítulo 2) sobre el conjunto de datos socioeconómicos capturados en la base de conocimiento semántico de la Universidad Politécnica Salesiana, resultante del proceso de importación descrito en el Capítulo 4.

A continuación, se presenta un estudio analítico somero de los datos socioeconómicos proporcionados por el componente de conocimiento de la Figura 5.1. Posteriormente, se describen los procesos implementados en el módulo de aprendizaje profundo para normalizar y regularizar los datos para alimentar con ellos una red neuronal. Finalmente, describimos el procedimiento de estudio cualitativo

y cuantitativo de la parametrización de una red de perceptrones que busca inferir los resultados académicos de un alumno a partir de los datos acumulados en la monitorización de miles de alumnos en el pasado. Los experimentos realizados y sus resultados se presentan en el Capítulo 7.

### 5.5.1. Análisis de parámetros socioeconómicos de los estudiantes

La Universidad Politécnica Salesiana obtiene gran cantidad de datos de los alumnos en el momento de la matriculación. Como se expondrá en lo sucesivo, estos datos cubren un espectro muy amplio de las características del alumno, desde elementos básicos como su fecha de nacimiento y sexo hasta detalles como la operadora móvil que le brindaba servicio en el momento de matricularse. El objetivo de este estudio fue determinar los hiperparámetros óptimos de entrenamiento de un esquema ampliamente utilizado de aprendizaje máquina para predecir el rendimiento académico de los estudiantes en base a un subconjunto de tales parámetros. Tal estudio precedería la exploración de otro destinado a evaluar esquemas de aprendizaje profundo distintos del perceptrón multicapa, en pos de la solución que alcanzase el mayor grado de precisión.

Para cada entrada de registro se reciben un total de 64 parámetros de todo tipo, que se listan en la Tabla 5.2. Partimos de la hipótesis de que algunos de estos parámetros serían relevantes, mientras que otros no.

Tabla 5.2: Datos de los registros socioeconómicos de los estudiantes.

Propiedad	Tipo	Notas
Estado civil	Enumerable	Toma cinco valores: soltero, casado, divorciado, viudo, unión libre.
Género	Enumerable	Masculino/femenino.
Ciudad de nacimiento	Enumerable	Nombres de ciudades.
Fecha de nacimiento	Fecha <sup>3</sup>	Fechas entre 1970 y 1996.
Nacionalidad	Enumerable	Varios países, pero la mayor parte de las veces ecuatoriana.
Tipo de sangre	Enumerable	Incluye “desconocido”.
Operadora celular	Enumerable	Claro, CNT, Movistar, otro.
Raza	Enumerable	Blanco, indígena, mestizo, negro, afroecuatoriano, negro, desconocido.
Colegio del que proviene	Enumerable	Multitud de valores.
Tipo de colegio	Enumerable	Extranjero, fiscal, fisco-misional, particular.

*continúa en la página siguiente*

<sup>3</sup> Se consideró enumerable a efectos de determinados cálculos, que se explicarán posteriormente.

*sigue de la página anterior*

País colegio	Enumerable	Varios países, principalmente Ecuador.
Provincia colegio	Enumerable	Multitud de valores.
Ciudad colegio	Enumerable	Multitud de valores.
Título de bachillerato	Enumerable	Multitud de valores.
Año de graduación	Enumerable <sup>4</sup>	Multitud de valores.
Nota de bachillerato	Numérico	Valor entero entre 0 y 20, salvo los alumnos que provienen del extranjero.
Tiene otro título	Binario	Sí/No
Nivel del título	Enumerable	Varios valores, solo si el anterior es Sí.
Estudia otra carrera	Binario	Sí/No
País domicilio	Enumerable	Ecuador
Provincia domicilio	Enumerable	Varias provincias.
Ciudad domicilio	Enumerable	Varias ciudades.
Parroquia domicilio	Enumerable	Numerosas parroquias.
Tipo de parroquia	Enumerable	Rural, urbano, urbano marginal.
Barrio	Enumerable	Numerosos barrios.
Zona domicilio	Enumerable	Centro, norte, sur, valles, rural, suburbios.
País de procedencia	Enumerable	Ecuador/vacío
Provincia de procedencia	Enumerable	Varias provincias para procedentes de Ecuador.
Ciudad de procedencia	Enumerable	Ciudades
Parroquia de procedencia	Enumerable	Parroquias
Tipo de parroquia de procedencia	Enumerable	Rural, urbano, urbano marginal.
Barrio de procedencia	Enumerable	Barrios
Zona de procedencia	Enumerable	Centro, norte, sur, valles, rural, suburbios.
Cabeza de familia	Binario	Sí/No
Depende económicamente	Binario	Sí/No
Problema de salud	Binario	Sí/No
Discapacidad	Binario	Sí/No
Carné CONADIS	Enumerable	Valores diversos.
Discapacidad_1	Enumerable	Diversas discapacidades.
Vivienda	Enumerable	Propia, cedida, arrendada, ...
Tipo de vivienda	Enumerable	Casa, departamento, ...

*continúa en la página siguiente*

<sup>4</sup> El año de graduación se debe considerar enumerable a efectos de las redes neuronales, pues carece de ordinalidad relevante y falsearía los datos al otorgar mayor peso específico a unos años que a otros.

*sigue de la página anterior*

Estructura de la vivienda	Enumerable	Ladrillo, bloque, concreto.
Agua potable	Binario	Sí/No
Alcantarillado	Binario	Sí/No
Energía eléctrica	Binario	Sí/No
Teléfono fijo	Binario	Sí/No
Internet	Binario	Sí/No
TV por cable	Binario	Sí/No
Quién cubre los gastos	Enumerable	Varios valores (medios propios, padres, etc).
Número de propiedades extra.	Numérico	Valor numérico pequeño (0, 1, 2...).
Valor de las propiedades extra.	Numérico	Valor en dólares USD.
Número de vehículos familiares	Numérico	Valor numérico.
Valor de vehículos	Numérico	Valor en dólares USD.
Gasto mensual vivienda	Numérico	Valor en dólares USD.
Gasto mensual alimentación	Numérico	Valor en dólares USD.
Gasto mensual educación	Numérico	Valor en dólares USD.
Gasto mensual transporte	Numérico	Valor en dólares USD.
Gasto mensual salud	Numérico	Valor en dólares USD.
Gasto mensual servicios	Numérico	Valor en dólares USD.
Gasto mensual otro	Numérico	Valor en dólares USD.
Gasto mensual total	Numérico	Suma de los anteriores.

### 5.5.2. Selección de parámetros

El primer paso del estudio fue seleccionar los parámetros que queríamos tener en consideración para las entradas de la red neuronal. Si bien en principio se podrían tener todos en cuenta, eliminar los que, a priori, parecieran menos interesantes nos ayudaría a reducir la dimensionalidad de la red. Así, se hicieron las siguientes consideraciones:

- **Parámetros de baja dispersión.** Algunos parámetros, como la nacionalidad, debían ser tratados aparte, pues el número de registros con nacionalidad distinta a la ecuatoriana era marginal. Lo mismo se puede aplicar al país de domicilio y el país de nacimiento. Aunque pueda ser interesante estudiar las diferencias entre nacionalidades, la convergencia de la red neuronal no será igual de rápida ni buena sabiendo que, posteriormente, la mayor parte de los valores que deseamos estimar serán de nacionalidad ecuatoriana.

Igualmente, en principio no pareció relevante tener en cuenta la operadora móvil, aunque se puede intentar encontrar correlaciones entre el nivel económico y la operadora elegida.

- **Parámetros dependientes.** De nuevo con el objetivo de reducir la dimensionalidad se pueden considerar diferentes redes e intentar diferentes parametrizaciones para parámetros que son dependientes entre sí. En una primera aproximación se puede utilizar el gasto total mensual como aproximador de la capacidad económica de cada estudiante, dejando los detalles de cada uno de los apartados de gasto para sucesivas iteraciones.
- **Huecos y campos vacíos o escasos.** En las tablas proporcionadas observamos que parte de los valores de procedencia (provincia, ciudad, etc) tenían un número demasiado elevado de huecos. Esto no sucedía en los campos de domicilio, y solo parcialmente en el campo “*Barrio*”. No es deseable utilizar campos con muchos huecos, pues habrá que asignarles un valor durante la normalización. Cuando el número de registros sin valor para un campo en concreto es pequeño podemos descartar estos registros, mientras que si una gran parte de los registros presentan un campo vacío es mejor descartar ese campo.

El campo “*Nivel del título*”, que existe solo en aquellos registros con “*Tiene otro título*” positivo, fue otro valor no considerado, al igual que los campos referentes a la institución de la segunda carrera.

- **Fechas.** Como nota general, las variables temporales merecen un tratamiento especial antes de ser utilizadas como entrada de una red neuronal, pues presentan varios problemas; principalmente, que su magnitud y rango las hacen difícilmente normalizables. Un tratamiento incorrecto de las fechas puede llevar al *overfitting* y anular completamente la validez de este campo.

Consideramos más interesante normalizar la edad de la persona a fecha de la graduación del bachillerato, y en la búsqueda de los hiperparámetros más adecuados se propusieron redes con y sin este dato.

De todas formas, para estudiar las posibles diferencias entre graduados de diferentes épocas temporales se puede utilizar una categorización de fechas, pero el valor predictivo será bajo, dado que los nuevos registros, evidentemente, serán de fechas posteriores.

- **Otros parámetros no considerados.** El campo “Carné CONADIS” se presenta solo en aquellos registros con discapacidad, pero no parece aportar información relevante. El tipo de discapacidad, por estar presente en pocos registros y presentar una gran variabilidad, tampoco se consideró.
- **Parámetros de muy alta dimensionalidad.** Algunos parámetros, como el “*Colegio del que proviene*” tienen una dimensionalidad muy elevada (decenas o centenares). Se decidió eliminarlos de la primera fase del trabajo y estudiarlos de forma independiente en el futuro. Sin menospreciar la utilidad de este parámetro, entendimos que la parametrización de “tipo de colegio” más “provincia del colegio” proporcionaría ya una granularidad estudiable sobre la influencia del colegio en las predicciones.

El mismo razonamiento se aplica a parte de los campos de residencia y procedencia, como parroquia o ciudad, quedándonos con un parámetro de alta granularidad (provincia) para contener la dimensionalidad.

- **Infraestructuras.** Por último, una forma adicional de reducir la dimensionalidad fue tratar los campos referentes a infraestructuras con un valor numérico

acumulado, una suerte de marcador. Así, la existencia de agua potable, electricidad, alcantarillado, teléfono fijo, Internet y TV por cable supondrían cada una una parte proporcional de un nuevo parámetro llamado “*infraestructuras*” con un valor numérico normalizado entre 0 y 1. No se consideró esencial asignar distintos pesos a cada uno de estos parámetros (e.g. 0,175 a agua, alcantarillado y electricidad, 0,1 a teléfono, Internet y televisión).

Se utilizaron tanto los valores independientes como el valor numérico equivalente como variantes a estudiar en la convergencia y velocidad de la red.

En cuanto a la normalización de los datos, como se explicó en el Capítulo 2, la literatura es extensa en cuanto a qué métodos utilizar para cada caso particular, para cada combinación de parámetros, para cada tipo de datos de salida, etc. Para el estudio llevado a cabo en el marco de esta tesis se escogieron los métodos que se explican a continuación:

- **Variables enumerables.** En general, el método que añade menos dispersión topológica a las entradas de una red neuronal para variables enumerables es el uno-a-k. Como inconveniente, la dimensionalidad de la red crece de manera exponencial con cada valor posible. Por ello, se debe tener cuidado a la hora de qué campos elegir o si se deben agrupar algunos valores en subcategorías.
- **Variables binarias.** Se ha decidido codificar las variables binarias como si fueran enumerables de 2 valores, por simplicidad y por considerarlas de igual calidad que las enumerables.
- **Variables numéricas.** Se aplicó normalización gaussiana de los campos numéricos. Cabe destacar que se preprocesaron los datos para eliminar valores fuera de rango y falsos ceros que pudiesen alterar gravemente el resultado.

El resultado de la selección y normalización de parámetros queda recogido en la Tabla 5.3.

Tabla 5.3: Resultado de la selección y normalización de los datos.

Propiedad	Tipo	Normalización	Resultado
Estado civil	Enumerable	uno-a-k	5 dimensiones
Género	Enumerable	uno-a-k	2 dimensiones
Ciudad de nacimiento	Enumerable	uno-a-k	189 dimensiones
Tipo de sangre	Enumerable	uno-a-k	9 dimensiones
Raza	Enumerable	uno-a-k	6 dimensiones
Tipo de colegio	Enumerable	uno-a-k	4 dimensiones
Provincia colegio	Enumerable	uno-a-k	44 dimensiones
Título de bachillerato	Enumerable	uno-a-k	526 dimensiones

*continúa en la página siguiente*

*sigue de la página anterior*

Edad en el año de graduación	Numérico	gaussiana	Escalado a media 0, varianza 1
Nota de bachillerato	Numérico	gaussiana	Escalado a media 0, varianza 1
Tiene otro título	Binario	uno-a-k	2 dimensiones
Estudia otra carrera	Binario	uno-a-k	2 dimensiones
Provincia domicilio	Enumerable	uno-a-k	21 dimensiones
Tipo de parroquia	Enumerable	uno-a-k	3 dimensiones
Zona domicilio	Enumerable	uno-a-k	6 dimensiones
Zona de procedencia	Enumerable	uno-a-k	6 dimensiones
Cabeza de familia	Binario	uno-a-k	2 dimensiones
Depende económicamente	Binario	uno-a-k	2 dimensiones
Problema de salud	Binario	uno-a-k	2 dimensiones
Discapacidad	Binario	uno-a-k	2 dimensiones
Vivienda	Enumerable	uno-a-k	4 dimensiones
Tipo de vivienda	Enumerable	uno-a-k	5 dimensiones
Estructura de la vivienda	Enumerable	uno-a-k	6 dimensiones
Agua potable	Binario	uno-a-k	2 dimensiones
Alcantarillado	Binario	uno-a-k	2 dimensiones
Energía Eléctrica	Binario	uno-a-k	2 dimensiones
Teléfono Fijo	Binario	uno-a-k	2 dimensiones
Internet	Binario	uno-a-k	2 dimensiones
TV por cable	Binario	uno-a-k	2 dimensiones
Quién cubre los gastos	Enumerable	uno-a-k	7 dimensiones
Número de propiedades extra	Numérico	gaussiana (con prefiltrado)	Escalado a media 0, varianza 1
Valor de las propiedades extra	Numérico	gaussiana (con prefiltrado)	Escalado a media 0, varianza 1

*continúa en la página siguiente*

*sigue de la página anterior*

Número de vehículos familiares	Numérico	gaussiana (con prefiltrado)	Escalado a media 0, varianza 1
Valor de vehículos	Numérico	gaussiana (con prefiltrado)	Escalado a media 0, varianza 1
Gasto mensual total	Numérico	gaussiana (con prefiltrado)	Escalado a media 0, varianza 1

### 5.5.3. Valores de salida

Para cada registro de entrada se tiene una colección de datos de salida compuestos por:

- Identificador del periodo lectivo (numérico, realmente enumerable)
- Materias tomadas en dicho periodo lectivo (numérico)
- Promedio (numérico)
- Aprobadas (numérico)
- Reprobadas (numérico)

Para una primera aproximación, se propuso definir una métrica de predicción en función a estos parámetros, por ejemplo, la nota media total (sumatorio de los promedios dividido por el número de promedios lectivos cursados). Podrían considerarse otros estimadores que tuvieran en cuenta el número de periodos cursados.

### 5.5.4. Balance de datos y regularización

El análisis previo de los datos de entrada arrojaba un resultado: si tomábamos un clasificador que dividiese la métrica objetivo en aprobados y suspensos (dos clases) el 82,5% de los elementos caían en la clase “*aprobados*” mientras que el 17,5% caían en la clase “*suspensos*”.

Esta circunstancia es un problema para los modelos de aprendizaje automático, pues caen naturalmente del lado de *overfitting* sobre la clase más representada. Se pueden utilizar varias tácticas para balancear los datos. En nuestro estudio optamos por la más simple: multiplicar por copia simple las muestras de la clase menos representada.

Otra forma de luchar contra el *overfitting* es mediante regularización. Una de las características del *overfitting* es producir funciones demasiado complejas, de la forma en que una ecuación polinomial se ajusta perfectamente a los puntos que interpola, pero varía radicalmente si añadimos un punto intermedio. La regularización ayuda a “*aplanar*” el modelo añadiendo un factor de penalización. Hemos utilizado este factor como parte de los hiperparámetros, dependiendo del tamaño de la red.

### 5.5.5. Modelado e implementación de la red

Para el modelado de la red neuronal partimos de las siguientes hipótesis:

- Los diferentes datos de entrada, anteriormente expuestos, tienen diferentes influencias sobre la red de perceptrones.
- Es difícil modelar, a priori, la mejor configuración de hiperparámetros de la red neuronal.
- Existe un punto de equilibrio entre el número de datos de entrada, dimensionalidad de la red e hiperparámetros que permite un balance entre rendimiento y capacidad de predicción de la red.

El estudio planteó realizar un barrido en todas las variables para acercarse al punto de equilibrio descrito. Es importante destacar que una red neuronal no hace sino capturar la fórmula inherente, existente pero oculta o difícil de analizar, que relaciona las entradas con las salidas de nuestro problema. Era una hipótesis de trabajo que esta fórmula existía. Si las entradas y las salidas no estuvieran correlacionadas la red neuronal actuaría, sencillamente, como un banco de memoria que relacionaría las entradas y salidas de los datos utilizados como entrenamiento, pero no tendría capacidad de predicción sobre datos futuros.

Antes de decidir sobre otros hiperparámetros, debimos enfrentarnos a los dos problemas de dimensionamiento de la red: el número de capas ocultas y el número de perceptrones en cada una.

- **Número de capas ocultas.** Hasta el advenimiento de las redes neuronales profundas, a partir de 2006, las técnicas numéricas utilizadas en los problemas de redes neuronales solo permitían un máximo de una capa oculta intermedia completamente conectada. A partir de entonces, las técnicas de aprendizaje profundo y alimentación inversa de la red permiten un número arbitrario de capas, representando cada una diferentes funciones que ayudan a resolver problemas de aprendizaje mucho más complejos. A modo genérico, presentamos a continuación un resumen de los motivos para utilización de capas de neuronas ocultas, es decir, aquellas no conectadas ni a la entrada ni a la salida de la red:

En el problema que nos atañe, la complejidad de las variables no parecía justificar la utilización de más de 2 capas ocultas. De hecho, consideramos probable que con una sola capa oculta completamente interconectada se conseguirían resultados de precisión muy semejante a los que se podrían obtener con más. Así pues, se probaron ambas opciones, pero no se buscarán soluciones más complejas.

- **Dimensionamiento de las capas ocultas.** El número de neuronas en las capas extremas (entrada y salida) están determinados por la arquitectura del problema: necesitamos tantas neuronas en la entrada como variables recibamos y tantas neuronas en la salida como dimensiones tenga el problema a modelar. Las capas ocultas, en cambio, pueden variar, y su configuración marca el comportamiento de la red en términos muy importantes:
  - **Equilibrio.** En términos generales, un número alto de neuronas en las capas intermedias harán que la red tienda al *overfitting*, es decir, modele

Capas ocultas	Resultado	Uso
Ninguna	Solo se pueden modelar funciones linealmente separables	Resolución de problemas lineales donde, realmente, las ventajas de usar una red neuronal no son tantas.
1	Puede aproximar cualquier función continua	Resolución de funciones genéricas
2	Puede representar con precisión arbitraria cualquier función	Mapeado de cualquier problema
Más de 2	Suelen ser capas especializadas, no genéricas, no totalmente conectadas, convolucionales, etc	Optimización de problemas complejos (como visión artificial)

con precisión los datos de entrenamiento a costa de pérdida de generalidad ante datos desconocidos, lo que le resta utilidad.

En el otro extremo, un número demasiado bajo de neuronas restará precisión a la red, dado que no habrá suficientes elementos como para capturar la función inherente al problema que intentamos modelar.

- **Rendimiento.** Adicionalmente al problema del *overfitting*, un exceso de neuronas intermedias hace crecer el tiempo de entrenamiento requerido para la red, pudiendo llegar a ser inmanejable en situaciones cotidianas.

Para decidir el número de perceptrones en las capas intermedias es importante conocer las dimensiones del problema a resolver, pero también la cantidad de datos de entrenamiento. Resulta inútil hacer crecer de forma desmesurada el número de neuronas intermedias si no tenemos datos suficientes para entrenarlas todas. Las aproximaciones frecuentemente utilizadas para estimar estos números son:

- El número de neuronas de las capas ocultas debería estar entre el número de neuronas de entrada y el número de neuronas de la salida.
- El número de neuronas ocultas será alrededor de  $\frac{2}{3}$  del tamaño de las entradas más el tamaño de las salidas [109].
- El número de neuronas ocultas deberá ser menos que el doble de las neuronas de entrada.

De todas maneras, la optimización del número de neuronas para cada problema debe afinarse mediante métodos de prueba y error, es decir, iteraciones sobre diferentes rangos de tamaños buscando el equilibrio entre el *overfitting* y la captura del problema.

Como nota final sobre el submódulo de aprendizaje profundo, cabe comentar que, dado que la cantidad de datos a modelar no era excesivamente grande (decenas de miles de registros de alumnos), no hemos utilizado soluciones basadas en GPU por no considerarlo necesario. Nuestra implementación se ha basado en la utilización de la librería *scikit-learn* de aprendizaje automático en Python, que ofrecen gran variedad de funciones completamente parametrizables. La integración

del módulo de conocimiento con el submódulo de aprendizaje profundo —con la materialización de todos los criterios comentados en las subsecciones previas (normalización, regularización, etc.) se ha implementado como un conjunto de *scripts* en la herramienta Allegro Graph.

## 5.6. Sumario

En el presente capítulo hemos abordado la construcción de una plataforma híbrida orientada al ámbito de las instituciones de Educación Superior, sustentada enteramente por la red ontológica descrita en el Capítulo 4, e incorporando módulos de interfaz de usuario y de sistema experto para soportar las nuevas aplicaciones de Inteligencia Artificial con fines de (i) proporcionar asistencia en la elaboración de grupos de aprendizaje colaborativo y (ii) realizar predicciones del rendimiento académico de los alumnos.

La red ontológica se ha mostrado suficientemente versátil y capaz de dar soporte a las nuevas aplicaciones, con el añadido de una nueva capa de definiciones (clases, propiedades y axiomas) para modelar los parámetros de personalidad que rigen el funcionamiento del algoritmo genético creado para la elaboración de grupos de aprendizaje. Dicho algoritmo supone, en relación con el estado del arte, la primera propuesta completa, sistemática y ampliamente interoperable (gracias a la base semántica) para la gestión de perfiles de estudiantes al efecto de organizar a los estudiantes con criterios de heterogeneidad intra-grupo y homogeneidad inter-grupos. Todo ello, con arreglo a estándares tecnológicos y del dominio de la psicología de la personalidad (e.g. los tests de personalidad y el modelo de los cinco factores), lo que aumenta las capacidades de aplicación de la plataforma más allá del ámbito experimental que explicaremos en el Capítulo 7.

En cuanto a la propuesta de aplicación de técnicas de aprendizaje profundo a la predicción de rendimiento académico, cabe destacar nuevamente la posibilidad de sistematizar la integración de la base de conocimiento semántica con los esquemas esperados en una implementación de redes neuronales, dando cuenta de los procedimientos de normalización y regularización de los datos. Dicha integración en el marco de una arquitectura de servicios amplia para instituciones de Educación Superior constituye el principal aporte de este capítulo en lo referente a las técnicas de aprendizaje profundo. Si bien en un principio se consideró implementar un esquema todavía más general —en el que, por ejemplo, se pudiesen utilizar otras arquitecturas de red neuronal diferentes al perceptrón multicapa—, dicha opción quedó postergada por los resultados del barrido de hiperparámetros que se presentará también en el Capítulo 7.

## Capítulo 6

# Evaluación experimental de los aportes

*En este capítulo resumimos los experimentos llevados a cabo, con la participación de docentes y estudiantes universitarios, para validar las aplicaciones de Inteligencia Artificial a la gestión y explotación de conocimiento en instituciones de Educación Superior. Describimos en primer lugar la logística de experimentación, y luego desarrollamos los resultados de la evaluación del algoritmo genético propuesto para la generación de grupos de aprendizaje colaborativo, seguido de los resultados del barrido de hiperparámetros realizado sobre una red neuronal de aprendizaje profundo para la predicción de rendimiento académico.*

### 6.1. Logística de experimentación

Previo a inicializar el proceso de experimentación, se procedió a solicitar el permiso respectivo a las autoridades de la Universidad Politécnica Salesiana (Rector y Vicerrector Académico) para la utilización de los datos históricos de la institución conforme a la normativa vigente a nivel nacional y a las políticas internas. Igualmente, se pidió permiso a los estudiantes seleccionados para el proceso de experimentación, firmando un acuerdo en el que constaba su consentimiento de manera voluntaria para rendir diferentes pruebas enfocadas al ámbito psicológico, además de poder hacer uso de sus datos en el proceso de experimentación. En el acuerdo establecido con los estudiantes, los responsables del estudio firmamos asimismo un acuerdo de confidencialidad, por el cual los datos no podrían ser utilizados con ninguna otra finalidad que no fuese para el proceso de experimentación. Adicionalmente, se les indicó a los estudiantes que los resultados de los cuestionarios se les informarían de manera personal, con la finalidad de que puedan tener un mejor conocimiento sobre este componente psicológico vinculado a su persona. Finalmente, también se estipuló que al concluir el proceso de experimentación estos datos serían eliminados.

El proceso de experimentación de la formación de grupos de aprendizaje colaborativo se realizó en el período lectivo 52 de la Universidad Politécnica Salesiana, correspondiente al período marzo-julio de 2018. Para el proceso de experimenta-

ción fueron seleccionados 3 grupos de estudiantes (183 en total) con sus respectivos docentes, correspondientes a tres diferentes carreras:

- 78 estudiantes de Ciencias de Computación (en adelante, CMP).
- 60 estudiantes de Ingeniería Electrónica (en adelante, ELT).
- 45 estudiantes de Ingeniería en Telecomunicaciones (en adelante, TLC).

La asignatura seleccionada para el estudio fue “*Métodos Numéricos*”, común a las tres carreras. Previo a la selección de la actividad colaborativa a desarrollar, reunimos a un equipo de expertos (psicólogos e ingenieros expertos en el área) para establecer las actividades y las evaluaciones objetivas de las rúbricas. El denominado “*claustro docente*” estuvo formado por cinco docentes con amplia experiencia en la asignatura, y quienes a la vez serían los responsables de impartir la cátedra durante ese parcial. Entre las diversas actividades que realizaron, las más relevantes fueron las siguientes:

- Consolidar y validar los contenidos en una base de conocimiento unificada, con la finalidad de impartir los mismos conocimientos entre diferentes carreras, donde se dictaba la asignatura común.
- Compartir experiencias, seleccionar las actividades colaborativas, seleccionar los temas y establecer fechas según calendario académico. A través del claustro docente se establecieron los tres temas, el método de conformación de los grupos, así como las fechas de revisión para los trabajos colaborativos, como se presenta en la Tabla 6.1.

Tabla 6.1: Planificación de las actividades colaborativas

<b>Tema de trabajo colaborativo</b>	<b>Fecha de revisión</b>	<b>Método de Agrupación</b>
1) Regresión por mínimos cuadrados [44]	13 de abril del 2018	Aleatorio
2) Interpolación mediante trazadores [45]	18 de mayo del 2018	Afinidad
3) Método de Runge-Kutta [46]	23 de junio del 2018	Algoritmo genético

- Eliminar la subjetividad al momento de evaluar, manejando una metodología y rúbrica de evaluación común, de tal forma que cualquier docente del claustro pueda evaluar los mismos resultados de aprendizaje.

En este punto se llegó a un consenso a través de la interacción entre los docentes del claustro en tres sesiones, la herramienta de trabajo colaborativo seleccionada fue el desarrollo de un “*mapa mental*” por tema (frente a otras opciones, tales como el mapa argumental, el mentefacto o el mapa conceptual), debido a que es una herramienta que permite a los estudiantes interactuar entre sí, y descubrir relaciones sistemáticas entre conceptos, promoviendo un mayor nivel de abstracción en el proceso de aprendizaje. [70, 83]

El coordinador del claustro docente mantuvo varias reuniones con expertos psicólogos en el área de la Educación Superior. Estas reuniones tuvieron como objetivo identificar y seleccionar aquellos cuestionarios asociados al aspecto psicológico del ser humano, pero que tuviesen implicación en el rendimiento académico. El aporte de los expertos psicólogos también nos orientó para conocer las pautas a considerar en la conformación de los grupos de aprendizaje colaborativo (las mismas que se emplearon como reglas duras y blandas en el desarrollo de nuestro algoritmo genético, Sección 5.4). Además, nos brindaron asesoramiento y soporte para la interpretación de los datos.

En cuanto al estudio predictivo de rendimiento académico, se procedió a la importación de la información contenida en las antiguas bases de datos de la Universidad Politécnica Salesiana, para incluirla en la base de conocimiento semántica acorde a nuestra red ontológica (Capítulo 4) y acto seguido se procedió a ejecutar los *scripts* mencionados en la Subsección 5.5.5. La principal salvaguarda a efectos de protección de datos pasó por anonimizar todos los registros utilizados, de manera que no se pudiese mostrar (en ninguna de las interfaces ofrecidas a los analistas de datos) información que permitiese identificar a un individuo concreto. Para ello, se aleatorizaron los nombres y apellidos, al igual que se eliminaron datos de contacto y se alteraron los números de calle en los domicilios.

## 6.2. Evaluación del algoritmo genético en la generación de grupos de aprendizaje colaborativo

Para evaluar las prestaciones de nuestro algoritmo genético en la formación de grupos, se procedió a realizar el levantamiento de la información relacionada a los rasgos de personalidad mediante la aplicación del cuestionario NEO-PI-R de 120 cuestiones. Como resume la Tabla 6.2, se manejaron 120 ítems correspondientes a 120 factores asociados a la personalidad de cada estudiante. Cada ítem emplea una *escala de Likert* entre 1 y 5 para especificar el nivel de acuerdo o desacuerdo con una declaración. De estos 120 factores, existen 4 factores que se agrupan para establecer una faceta de la personalidad (véase Tabla 6.4). La ponderación para cada faceta se encuentra en una escala entre 4 y 20. Finalmente, existen 6 facetas que se agrupan para integrar un rasgo de personalidad (Tabla 6.3), con lo cual se obtienen los 5 rasgos de personalidad a nivel macro, cada uno con una ponderación entre 24 y 120 puntos.

Tabla 6.2: Resumen de factores, facetas y rasgos de personalidad.

Denominación	Cantidad	Escala
Factores	120	1-5
Facetas	30	4-20
Rasgos	5	24-120

Tabla 6.3: Resumen de las 6 facetas de cada rasgo.

<b>Apertura al cambio (O)</b>	<b>Responsab. (C)</b>	<b>Extraversión (E)</b>	<b>Amabilidad (A)</b>	<b>Neuroticismo (N)</b>
Imaginación	Autoeficacia	Afabilidad	Confianza	Ansiedad
Intereses artísticos	Orden	Gregarismo	Moralidad	Enfado
Emoción	Sentido del deber	Asertividad	Altruismo	Depresión
Aventura	Logro por esfuerzo	Activo	Cooperación	Autoconciencia
Intelecto	Autodisciplina	Busco emoción	Modestia	Inmoderación
Liberalismo	Cautela	Alegría	Simpatía	Vulnerabilidad

Tabla 6.4: Resumen de una faceta de cada uno de los 5 rasgos.

<b>Imaginación (O<sub>1</sub>)</b>	<b>Autoeficacia (C<sub>1</sub>)</b>	<b>Afabilidad (E<sub>1</sub>)</b>	<b>Confianza (A<sub>1</sub>)</b>	<b>Ansiedad (N<sub>1</sub>)</b>
Tengo una imaginación vívida	Completo las tareas correctamente	Creo que hago amigos fácilmente	Confío en los demás	Me preocupó por las cosas
Disfruto de sueños de fantasía	Soy muy bueno en lo que hago	Me siento cómodo con la gente	Creo que las personas tienen buenas intenciones	Temo que suceda lo peor
Soy una persona que a veces sueña despierta	Realizo mis tareas sin ningún problema	A menudo evito el contacto con los demás	Confío en lo que dicen las personas	Tengo miedo de muchas cosas
Me gusta perderme en mis pensamientos	Generalmente sé cómo hacer las cosas	Siempre mantengo cierta distancia con las personas	Desconfío de la gente	Tiendo a estresarme con facilidad

El conjunto de datos empleado en el estudio contiene información de tres ámbitos distintos para cada alumno:

▪ **Ámbito académico:**

- **Carrera:** Computación (CMP), Electrónica (ELT) o Telecomunicaciones (TLC).
- **Método de agrupamiento:** aleatorio, por afinidad o genético.
- **Nota del trabajo individual (NI).**
- **Nota del trabajo en grupo (NG).**
- **Nota final (NF).**
- **Promedio general (PG).**
- **Materias aprobadas (MA).**
- **Materias reprobadas (MR).**

- **Ámbito personal y familiar:**
  - **Fecha de nacimiento (FN).**
  - **Quintil por situación socioeconómica (QT).**
  - **Estado civil (EC).**
  - **Etnia.**
  - **Género (GN).**
- **Ámbito psicológico:** puntajes obtenidos en los 5 rasgos de personalidad: O, C, E, A y N.

En la Tabla 6.5 podemos observar una muestra de 25 registros, de los 183 totales. Cabe aclarar que el algoritmo genético trabajó únicamente con los datos de los rasgos de personalidad, mientras que los otros métodos de formación de grupos no consideraban información alguna de los registros. Los datos recogidos a mayores (quintil, etnia, género, ...) los incorporamos a nuestro análisis estadístico para comprobar si de ellos se derivaría alguna correlación relevante, que pudiera ser tenida en cuenta a la hora de aplicar cualquiera de los métodos de agrupamiento.

La aplicación de estadísticos básicos arroja los resultados de la Tabla 6.6, donde se aprecia que la utilización de nuestro algoritmo genético da lugar a valores promedio más altos y dispersiones menores en las calificaciones de los trabajos en grupo, lo cual se considera un primer refrendo del valor de la propuesta. No obstante, para verificar si es posible afirmar que nuestro método es mejor que los demás, en base a los datos disponibles, es necesario realizar un análisis con estadísticos más avanzados, para lo cual existen distintos métodos.

Tabla 6.6: Medias y desviaciones típicas de las notas grupales (NG), según distintos métodos de agrupamiento.

Método	Media NG	Desviación típica NG
Aleatorio	7,8443	1,5426
Afinidad	7,7705	1,8609
Genético	8,5328	1,2345

Un limitante fundamental a la hora de escoger el método de análisis se deriva de las distribuciones de los parámetros de personalidad, pues, como muestran la Figura 6.1 para los rasgos globales y especialmente la Figura 6.2 para los rasgos por carrera, en ocasiones distan mucho de ser aproximables por distribuciones normales. Por otra parte, como muestra la Figura 6.3, las distribuciones del rendimiento académico en cada carrera sí podrían aproximarse por variables normales, aunque de medias y varianzas sensiblemente distintas. Estas observaciones hacen inaplicables en nuestro estudio los modelos de la regresión lineal ordinaria para establecer dependencias entre parámetros, y análisis de varianza tales como ANOVA para valorar la significación estadística de las afirmaciones, en tanto que no se cumplen los criterios básicos de linealidad y normalidad. En consecuencia, optamos por realizar el estudio según el **modelo lineal generalizado (GLM)** [163], una revisión flexible de la regresión lineal ordinaria que sí permite variables con modelos de distribución y de errores distintos a los de una distribución normal.

Tabla 6.5: Registros extraídos de la base de datos experimental.

Carrera	Método	NG	NI	NF	PG	MA	MR	FN	QT	GR	Etnia	EC	N	E	O	A	C
CMP	Aleatorio	10	7	8,5	78,20	15	1	1999-2-6T00:00Z	2	F	Mestizo	Soltero	93	93	95	67	82
CMP	Aleatorio	7	10	8,5	88,38	16	0	1997-5-12T00:00Z	3	M	Mestizo	Soltero	83	78	82	75	71
CMP	Afinidad	9	5,5	7,25	77,69	16	0	1998-4-17T00:00Z	3	M	Mestizo	Soltero	77	78	69	79	86
ELT	Genético	8	8,5	8,25	76,33	15	0	1995-12-24T00:00Z	3	M	Mestizo	Soltero	69	91	82	79	74
TLC	Genético	6,5	8	7,25	74,64	25	22	1994-6-11T00:00Z	2	F	Mestizo	Casada	77	90	87	91	106
CMP	Genético	10	6,5	8,25	80,06	16	0	1999-1-11T00:00Z	4	M	Mestizo	Soltero	71	84	81	78	77
ELT	Genético	8,5	7	7,75	75,42	12	2	1995-8-24T00:00Z	3	M	Mestizo	Soltero	84	96	86	81	68
ELT	Genético	9	10	9,5	78,40	15	0	1998-6-17T00:00Z	3	M	Mestizo	Soltero	69	79	75	86	88
ELT	Afinidad	5,5	7	6,25	79,56	16	5	1997-10-28T00:00Z	4	M	Mestizo	Soltero	89	56	76	78	64
CMP	Afinidad	9	10	9,5	88,38	16	0	1997-5-12T00:00Z	3	M	Mestizo	Soltero	83	78	82	75	71
CMP	Afinidad	9	5,5	7,25	76,81	16	0	1997-9-22T00:00Z	3	M	Mestizo	Soltero	81	76	82	83	90
ELT	Afinidad	10	10	10	84,27	15	0	1994-10-3T00:00Z	3	M	Mestizo	Soltero	59	89	68	84	89
TLC	Genético	7	7	7	74,57	23	13	1995-7-22T00:00Z	2	F	Mestizo	Soltero	81	79	84	88	99
CMP	Aleatorio	10	9	9,5	86,81	16	0	1998-5-5T00:00Z	3	M	Mestizo	Soltero	73	85	95	96	93
ELT	Aleatorio	8,5	8	8,25	79,60	15	0	1997-5-13T00:00Z	3	F	Mestizo	Soltero	86	70	78	83	73
ELT	Aleatorio	8,5	5,5	7	76,33	15	0	1995-12-24T00:00Z	3	M	Mestizo	Soltero	69	91	82	79	74
TLC	Aleatorio	5,5	6	5,75	76,65	20	14	1992-8-26T00:00Z	3	M	Mestizo	Soltero	74	78	80	81	75
TLC	Aleatorio	5,5	5,5	5,5	74,33	24	11	1994-7-11T00:00Z	3	M	Mestizo	Soltero	63	94	75	76	80
TLC	Genético	7	5	6	73,42	24	28	1994-7-9T00:00Z	3	M	Mestizo	Soltero	84	75	69	70	66
CMP	Aleatorio	7,5	8	7,75	84,94	16	0	1997-6-30T00:00Z	4	M	Mestizo	Soltero	61	77	83	79	98
CMP	Aleatorio	10	5,5	7,75	84,44	16	0	1998-1-13T00:00Z	3	M	Mestizo	Soltero	64	98	73	86	78
CMP	Afinidad	9	5	7	82,63	16	0	1999-4-11T00:00Z	4	M	Mestizo	Soltero	63	76	75	83	76
TLC	Aleatorio	8	6	7	74,57	23	13	1995-7-22T00:00Z	2	F	Mestizo	Soltero	81	79	84	88	99
CMP	Genético	8	8	8	86,06	16	0	1999-2-7T00:00Z	3	M	Mestizo	Soltero	73	94	78	81	90

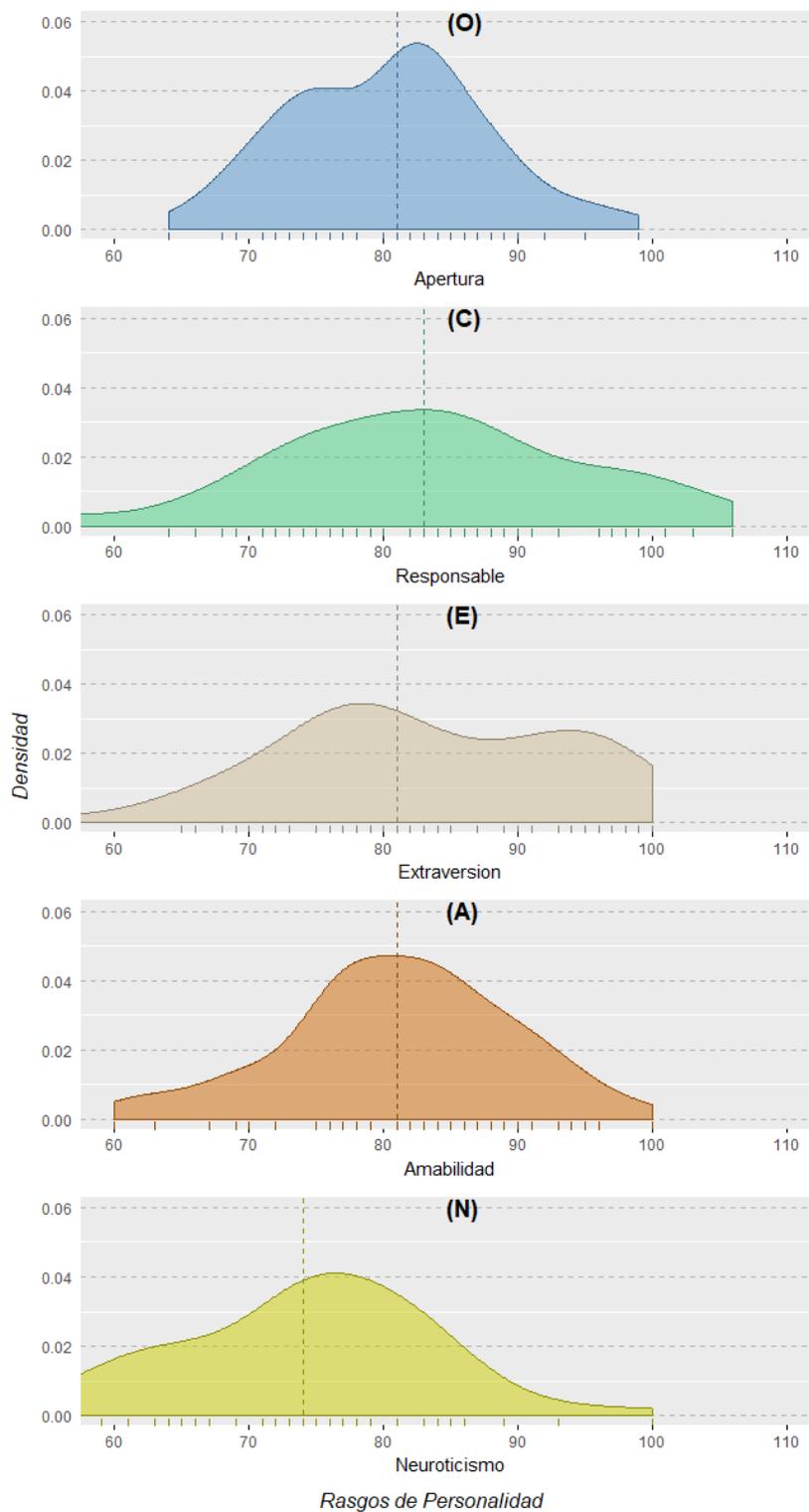


Figura 6.1: Distribución de rasgos de personalidad globales.

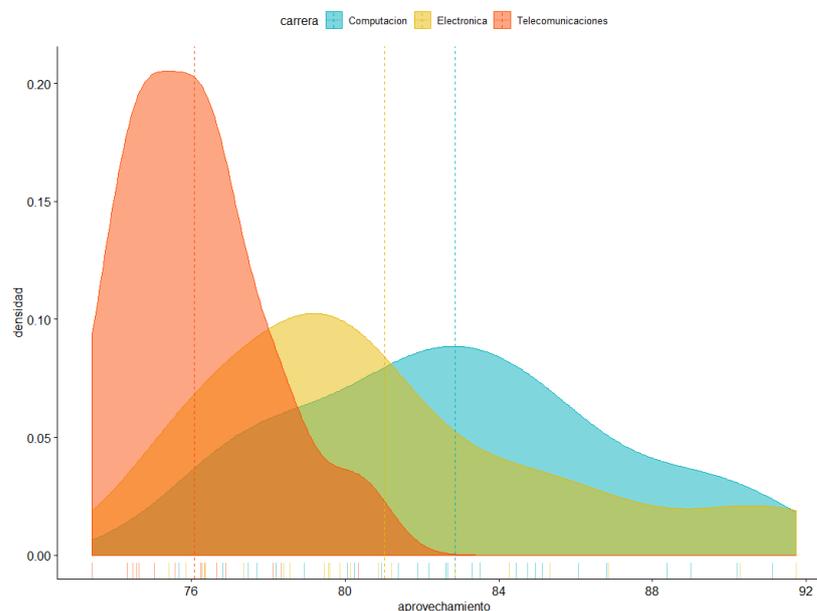


Figura 6.3: Distribución de rendimiento académico por carrera.

A modo de ejemplo de las varias dependencias estudiadas, mostramos a continuación (véase Tabla 6.7) el resultado de aplicar el modelo GLM para determinar la relación entre la nota del trabajo grupal y un predictor lineal conformado por el conjunto de las variables O, C, E, A y N, más el método empleado para la formación de los grupos y la carrera cursada. Con estos valores, se puede dar solidez a las siguientes afirmaciones:

- El uso de nuestro algoritmo genético tiene un reflejo directo en la nota grupal, porque tiene un p-valor estadísticamente significativo ( $0,02853 < 0.05$ ).
- El rendimiento en la carrera de Telecomunicaciones (en la materia de “*Métodos Numéricos*”, claro está) es sensible y significativamente inferior al de las otras.
- La presencia de personalidades marcadamente amables (A) y responsables (C) en los grupos repercute positivamente en la nota del trabajo grupal.

Bajo procedimientos análogos realizados con otros muchos predictores lineales, se pueden afirmar los siguientes puntos, siempre sustentados en valores adecuados de los indicadores de validez que ofrecen los residuos y el criterio de información de Akaike (ejemplo en la Tabla 6.7):

- El uso de nuestro algoritmo genético reporta rendimientos de grupo sistemáticamente mejores que los otros dos métodos de agrupamiento (ya sea en estadísticas agregadas o separadas por carrera), demostrando que las reglas duras y blandas que incorporamos en la lógica del mismo dan los resultados pretendidos.
- El rendimiento de los grupos de trabajo colaborativo (NG) no se ve afectado significativamente por el rendimiento general individual de sus miembros (pro-

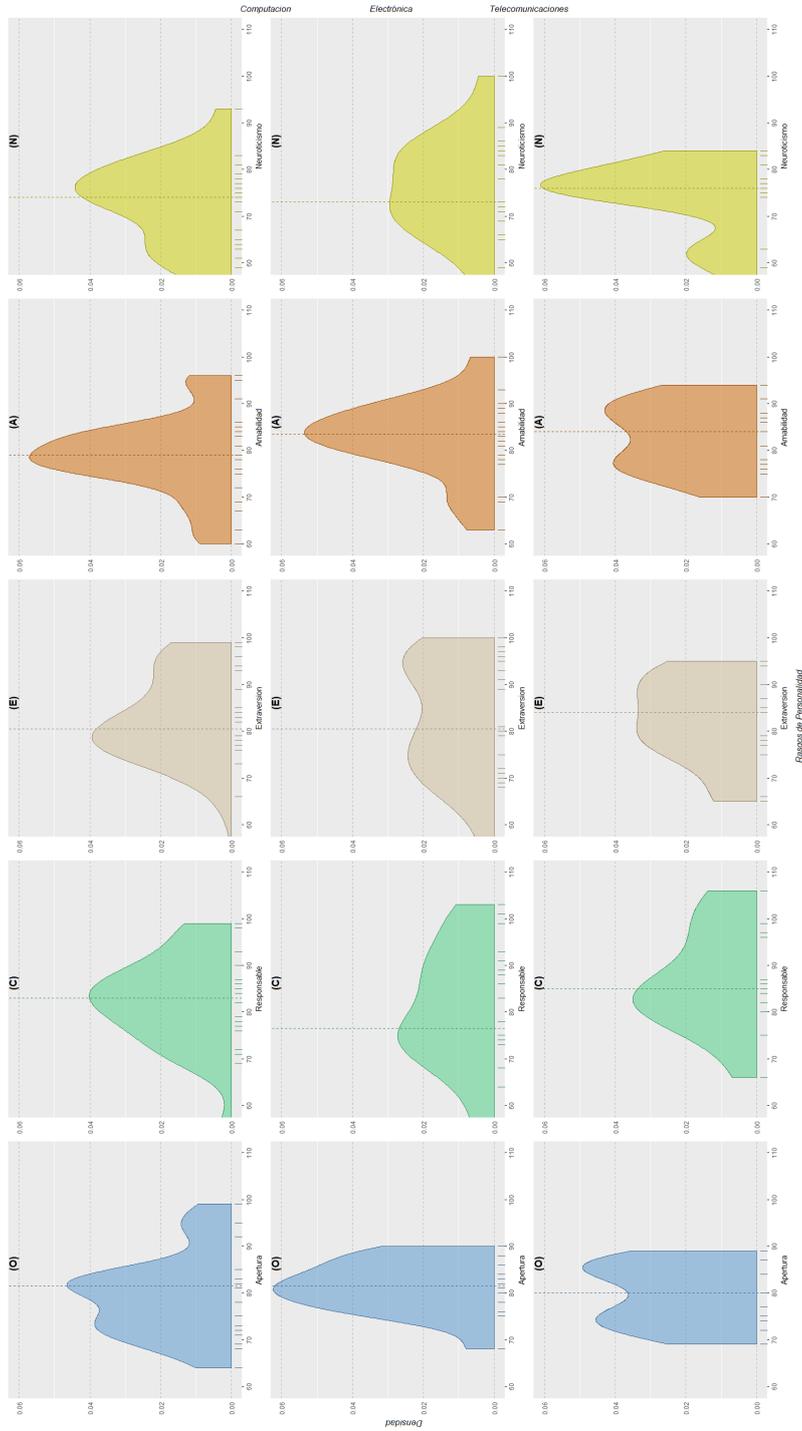


Figura 6.2: Distribución de rasgos de personalidad por carrera.

Tabla 6.7: Aplicación del modelo GLM a la nota del trabajo grupal.

	$P(\geq  t )$
O	0,27540
<b>C</b>	<b>0,00126</b>
E	0,32785
<b>A</b>	<b>0,01921</b>
N	0,29441
Aleatorio	0,21752
Afinidad	0,28722
<b>Genético</b>	<b>0,02853</b>
CMP	0,42306
ELC	0,38640
<b>TLC</b>	<b>0,00724</b>
(Intercept: 4,21e-6)	
(Desviación nula: 692,72 sobre 182 grados de libertad)	
(Desviación residual: 605,46 sobre 172 grados de libertad)	
(Criterio de Akaike (AIC): 758,29)	

mediando sus respectivos números de materias aprobadas, MA, y reprobadas, MR).

- El rendimiento de los grupos de trabajo colaborativo no se ve afectado por la información del quintil socioeconómico (QT), que tampoco correla con un mayor o menor rendimiento académico individual en ningún caso (nuevamente, tanto en estadísticas agregadas como separadas por carrera).
- No hay base empírica suficiente para discutir efectos relacionados con estado civil (EC), etnia o género (GN), dado que determinados valores de dichos parámetros (e.g. alumnos/as casados/as, blancos y mujeres) están insuficientemente representados y, por tanto, deparan varianzas excesivamente altas.

### 6.3. Evaluación del perceptrón multicapa en la predicción de rendimiento académico

Como se explicó en el Capítulo 5, nuestro experimento en lo tocante a la predicción de rendimiento académico consistió en realizar un barrido de hiperparámetros para un perceptrón multicapa, probando un gran número de configuraciones en pos de la que proporcionase la máxima precisión, contando con los datos acumulados en los registros históricos de la Universidad Politécnica Salesiana, convenientemente importados en las estructuras de nuestra red ontológica OEA.

El barrido de hiperparámetros consideró los parámetros y los rangos de valores que se indican en la Tabla 6.8.

A mayores de lo explicado sobre la implementación del submódulo de aprendizaje profundo en Sección 5.5, en nuestro experimento supervisamos la convergencia de dos parámetros a fin de optimizar la búsqueda de la mejor configuración del perceptrón multicapa. Esos parámetros son:

Tabla 6.8: Hiperparámetros y rangos de valores.

Parámetro	Significado	Valores a barrer
Momento	Estabiliza la red, es decir, marca el peso de la realimentación de la red. Si es muy bajo los pesos de la red varían mucho y se alarga el proceso de convergencia. Si es demasiado alto la red puede converger demasiado lejos del mínimo.	0.05 - 0.9
Alfa	Valor de regularización. Trata de evitar el sobreajuste cuando determinados pesos crecen demasiado en magnitud.	0,001 - 125
Neuronas ocultas	Arquitectura general de la red: número de neuronas en las capas ocultas y número de capas ocultas.	1 o 2 capas ocultas, con tantas neuronas como 0,7 o 0,8 veces la dimensión de la entradas.
Funciones de activación	Forma de la función de activación de las conexiones entre perceptrones.	{ReLU, cReLU, tanh}
Optimizadores	Funciones que deciden dirección y tamaño del gradiente para minimizar la métrica de error elegida.	{BFGS, Adam, Adadelta, SGD}
Tasa de aprendizaje	Tasa de cambio de los coeficientes. Los coeficientes (pesos) de la red neuronal pueden ir cambiándose más rápidamente o más lentamente, o dinámicamente.	{Constante, inversa, adaptativa}

- La **precisión de entrenamiento** monitoriza la velocidad a la cual la red neuronal se ajusta a los datos de entrenamiento; el valor debe tender a 1.
- La **precisión de prueba** monitoriza la velocidad a la cual la red neuronal aprende a predecir salidas correctas un conjunto de datos de entrada para los cuales no ha sido entrenada.

Si el valor de la precisión de entrenamiento es mayor que el de la precisión de prueba, entonces se puede asumir que la red ha caído en sobreajuste. Cualquier red que sea suficientemente grande y haya sido entrenada suficientemente debe converger a precisión total sobre los datos de entrenamiento, mientras que para los datos de prueba sólo se acercará a 1 si existe una relación matemática entre entradas y salidas. Por tanto, nuestro algoritmo ajusta los valores de regularización y el número de neuronas automáticamente cuando detecta signos de sobreajuste.

En total, el barrido probó más de 12.000 configuraciones. La Figura 6.4 muestra, para un subconjunto de ellas, que el perceptrón converge rápidamente a un nivel de precisión en torno al 80% después de proporcionar sucesivos lotes de datos de entrenamiento a la entrada, en particular para la métrica de si un nuevo estudiante aprobará o reprobará una determinada materia. Todas las demás métricas arrojan gráficas similares, si descartamos las configuraciones que no consiguen converger o que caen en el sobreajuste.

Un nivel de precisión en torno al 80%, a priori, podría parecer muy bueno a la hora de realizar predicciones de rendimiento académico. No obstante, como comen-

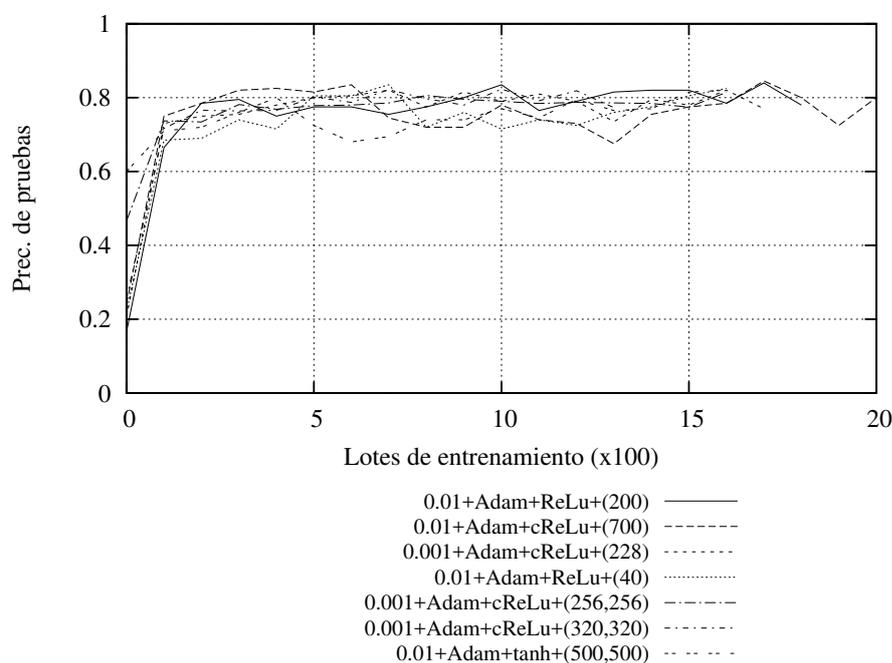


Figura 6.4: Convergencia de algunas configuraciones de ejemplo del perceptrón multicapa a la hora de predecir aprobados/reprobados en asignaturas individuales.

tamos en la Subsección 5.5.4, en realidad es posible acertar el resultado de aprobar o reprobar una materia para un nuevo estudiante con una precisión del 81,5% simplemente escogiendo siempre que aprobará, independientemente de cualesquiera otros datos del alumno. Por lo tanto, la Figura 6.4 no representa ninguna mejora sobre la *hipótesis nula*. De manera contraria a las expectativas generadas por los trabajos previos en la literatura (Sección 3.6), este mismo resultado lo observamos para todas las demás métricas que consideramos y los miles de configuraciones del perceptrón multicapa que probamos: como mucho, llegábamos a lo que muchos autores denominan el *punto natural* o *punto nulo* de los datos, ya que era posible adivinar los valores de las métricas de interés, con la máxima precisión posible, simplemente ponderando los valores de salida, sin importar los de entrada.

Cabe comentar que en muchos casos se obtuvieron picos de precisión por encima del punto nulo (e.g. en las configuraciones que lista la Tabla 6.9), pero estos se debían a oscilaciones de los pesos de la red neuronal, lo cual a su vez provocaba oscilaciones a la salida. Las medias a largo plazo permanecieron siempre por debajo de dicho punto, siendo este el motivo por el cual no continuamos el experimento en la línea planteada inicialmente de sustituir el perceptrón multicapa por otros diseños de red neuronal.

Cabe destacar que la hipótesis nula nos da una orientación sobre la calidad de nuestra red neuronal. Una red divergente o mal diseñada no se ajustaría a la hipótesis nula; más bien, su precisión tendría oscilaciones significativas o alcanzaría valores sistemáticamente más bajos de lo esperado a través del simple análisis de las salidas.

Tabla 6.9: Detalles de las configuraciones del perceptrón multicapa representadas en la Figura 6.4.

Tasa de aprendizaje	Alfa	Optimizador	Función de activación	Niveles ocultos	Precisión de entrenamiento	Precisión de prueba
0.01	0.48	Adam	cReLu	(560)	1	0.825
0.01	0.12	Adam	ReLu	(160)	1	0.815
0.01	0.576	Adam	ReLu	(128)	0.965	0.83
0.1	0.01	SGD	tanh	(200)	0.785	0.81
0.1	0.1	SGD	tanh	(200)	0.734	0.82
0.001	0.01	Adadelta	ReLu	(200,200)	0.5225	0.832
0.01	0.21	Adam	cReLu	(81,81)	1	0.81
0.01	0.4	SGD	cReLu	(320)	0.87	0.82
0.1	0.025	SGD	cReLu	(104)	0.955	0.825
0.001	0.358	Adam	ReLu	(104)	0.99	0.825
0.01	0.0358	Adam	ReLu	(104)	1	0.81
0.1	0.25	SGD	ReLu	(163,163)	1	0.82

En relación con los trabajos precedentes, hay que destacar que este experimento incluye el conjunto de datos más extenso y detallado de los registros de los estudiantes (30 veces el promedio de los trabajos mencionados en el Capítulo 2). Además, mientras que muchos otros se han centrado en comparar diferentes algoritmos y técnicas con configuraciones únicas o, a lo sumo, algunas decenas de combinaciones de hiperparámetros, hemos sido —hasta donde alcanza nuestra revisión del estado del arte— los primeros en realizar una exploración extensa (más de 12,000 configuraciones) para aprovechar al máximo una técnica muy bien estudiada en la literatura (a saber, el perceptrón multicapa). Sin una exploración de este tipo, y ante datos de entrenamiento limitados, podemos preguntarnos si los resultados positivos afirmados por los estudios anteriores pueden tomarse como evidencia de que las técnicas avanzadas de Inteligencia Artificial tienen realmente un papel que desempeñar en la gestión proactiva de las instituciones de Educación Superior, y si las configuraciones de las técnicas que resultaron más ventajosas funcionarían igual de bien en otros contextos y con otros conjuntos de datos.

Finalmente, es notable que la mayoría de los trabajos previos no comprobaron la hipótesis nula; es decir, no abordaron la cuestión esencial de si las técnicas de predicción en realidad proporcionaban información valiosa que no podría obtenerse de manera mucho más sencilla.

## 6.4. Sumario

En este capítulo hemos presentado los experimentos llevados a cabo, en el marco de la Universidad Politécnica Salesiana, para evaluar la utilidad de nuestro algoritmo genético para la generación de grupos de aprendizaje colaborativo, y para realizar un barrido de hiperparámetros con el que analizar la viabilidad de realizar predicciones de rendimiento académico en base a parámetros socioeconómicos.

Por una parte, si bien la masa muestral del experimento de formación de grupos de trabajo es limitada —con todo, mayor que la inmensa mayoría de los trabajos previos que se encuentran en la literatura relacionada—, podemos afirmar con apoyo en el análisis estadístico que nuestro algoritmo genético basado en características de personalidad representa una solución eficaz para asistir en tal labor, rindiendo resultados mucho mejores que procedimientos aleatorios o basados únicamente en las preferencias y afinidades de los alumnos. El estudio permitió asimismo refutar que el rendimiento académico individual o grupal se vea afectado por parámetros socioeconómicos, mientras que la evidencia experimental no alcanzó para hacer afirmaciones en cuanto a dependencias con factores de género o etnicidad.

En cuanto a la predicción de rendimiento académico, a pesar de lo reportado por numerosos autores en el pasado, nuestro barrido exhaustivo de configuraciones para una red neuronal (incluyendo miles de escenarios de aprendizaje profundo) terminó por no encontrar correlación alguna entre los muchos datos de entrada sobre los alumnos y cualquiera de las métricas elegidas de rendimiento académico. El experimento, por lo tanto, alumbró la conclusión de que no hay razón para desarrollar procedimientos tan complejos y costosos con objeto de realizar predicciones de rendimiento —al menos, con el nivel de granularidad que hemos considerado y centrándonos sólo en variables socioeconómicas. Asimismo, pone en tela de juicio la relevancia estadística de los experimentos anteriores, llevados a cabo sobre poblaciones muestrales mucho más reducidas y, sorprendentemente, sin verificar la

---

hipótesis nula de si los mecanismos de predicción realmente acertaban más allá de lo que cabría esperar por simple observación de los datos —en cuyo caso la toma de decisiones en la Universidad se puede basar en estadísticas básicas. No obstante, hay observaciones de carácter general pendientes a la hora de matizar el alcance de estas afirmaciones, como se explicará en el siguiente capítulo.



**Parte IV**

**Conclusiones**



## Capítulo 7

# Conclusiones y trabajos futuros

*En este capítulo se reseñan las principales conclusiones del presente trabajo de tesis, con la intención de dejar por sentadas las contribuciones en el ámbito de la utilización de tecnologías de Inteligencia Artificial para mejorar la administración y gestión de la información que se genera entorno al ámbito de la Educación Superior. Se plantean también líneas de trabajo futuro, con vistas a consolidar los aportes a través de nuevos planteamientos, desarrollos y experimentos.*

### 7.1. Conclusiones

Como se explicó en el Capítulo 1, esta tesis se planteó con el objetivo general de forjar las bases de un completo ecosistema de apoyo a la administración y gestión de la información que se genera en el ámbito de la Educación Superior, desde una perspectiva fundamentada en el modelado semántico del conocimiento y el empleo de innovadoras técnicas de Inteligencia Artificial, brindando un diseño altamente flexible e interoperable. Hay que destacar que la experimentación —al igual que la implementación de los aportes de Inteligencia Artificial— se ha visto facilitada en gran medida por la creación previa de una plataforma de estas características, con procedimientos uniformes para la consulta y adaptación de datos.

Las contribuciones de este trabajo de investigación se enmarcan en tres grandes áreas, a saber: (i) el modelado formal del conocimiento del ecosistema académico de la Educación Superior, (ii) la asistencia a la formación de grupos de aprendizaje colaborativo, y (iii) la predicción del rendimiento académico de los estudiantes.

#### 7.1.1. Sobre el modelado formal del conocimiento

Atendiendo al primer punto, se ha creado e implantado la red ontológica OEA para integrar la información generada en las distintas áreas del entorno académico, dando cuenta de las cuestiones de modelado del conocimiento e interoperabilidad. El desarrollo fue llevado a cabo según los lineamientos de la metodología NeOn, por los cuales comenzamos realizando entrevistas en base a preguntas de competencia para identificar los requerimientos funcionales que, posteriormente, dieron pie a

la conceptualización de requisitos, usuarios y usos previstos de la red ontológica. Asimismo, se han incorporado recursos ontológicos y no ontológicos preexistentes allí donde fue posible y se estimó conveniente, siempre con el objetivo claro de poder capturar de forma explícita el desarrollo que los distintos actores del ecosistema toman a lo largo del tiempo.

El trabajo a nivel de representación de conocimiento también nos permitió abordar algunas problemáticas de la información distribuida en varias bases de datos de sistemas transaccionales, en donde la interpretación de esa información se presta a la ambigüedad. En la aplicación al entorno de la Universidad Politécnica Salesiana —que sentó las bases para los demás aportes de la tesis— logramos integrar y unificar más de 150.000 registros en un conjunto consistente de instancias que poblaron nuestra red ontológica con 2 millones de triplas semánticas, con capacidad demostrada para inferir nueva información y consolidar conocimiento.

### **7.1.2. Sobre la formación de grupos de aprendizaje colaborativo**

En lo tocante a la formación de grupos de aprendizaje colaborativo, aprovechando el modelado formal de los estudiantes y de los planos de estudios a nivel micro y macroscópico, hemos implementado un sistema de algoritmos genéticos que tiene en cuenta las características de personalidad de los alumnos como elemento clave, realizando plenamente lo que en propuestas anteriores se había realizado de manera superficial y no sistemática. En nuestros experimentos con docentes y estudiantes de distintas carreras —aunque observando siempre las mismas materias— se ha demostrado que nuestra propuesta consigue un rendimiento muy superior al de técnicas clásicas de formación de grupos, mejorando todos los estadísticos relevantes sin que un análisis de varianzas arroje dudas al respecto. Asimismo, este estudio apuntaba ya en la dirección de refutar que el rendimiento académico individual o grupal se viera afectado por parámetros de nivel económico, mientras que determinó como insuficiente la evidencia experimental a propósito de valorar dependencias con otros parámetros tales como el género o la etnicidad. Dada la significación estadística alcanzada, sumada al hecho de que nuestra masa muestral supera a la inmensa mayoría de los trabajos previos que se encuentran en la literatura relacionada, creemos justificado afirmar que nuestra propuesta representa un aporte significativo al estado del arte, configurándose como una herramienta aplicable en otros entornos cualesquiera de Educación Superior que gestionen perfiles psicológicos de sus estudiantes.

### **7.1.3. Sobre la predicción del rendimiento académico**

En lo tocante a la predicción del rendimiento académico, hemos ejecutado un experimento sobre el uso potencial de las redes neuronales para la detección de correlaciones entre los diversos campos de datos almacenados en bases de datos con miles de registros de estudiantes, centrado específicamente en si las variables socioeconómicas, las condiciones familiares y relacionadas con la salud, y los lugares de origen o residencia podrían influir en los indicadores del rendimiento académico en una medida estadísticamente significativa. El contexto de la Universidad Politécnica Salesiana se consideró propicio porque, de acuerdo con la práctica común en Latinoamérica, los registros contienen campos que plantearían inquietudes bajo el escrutinio de las regulaciones de protección de datos de la Unión Europea.

Nuestro experimento se configuró de una manera que traería a la luz, al menos, las correlaciones más notables entre una selección de variables de entrada y las métricas elegidas, que consideraron aspectos de rendimiento académico con granularidad gruesa. A partir de esos hallazgos, pasaríamos a diseños de redes neuronales refinados y métricas más detalladas. Sin embargo, al contrario de las expectativas iniciales (alumbradas por la revisión del estado del arte en la materia), el escaneo de los hiperparámetros para la red de perceptrones multicapa mostró que no había una correlación entre las entradas y las salidas, sugiriendo que no se puede derivar nuevo conocimiento por la inexistencia de correlaciones significativas.

Antes de apresurarnos a lanzar la conclusión de que esta incapacidad de realizar predicciones pudiera ser el caso para todas las universidades de la región —lo que extrapolaría casi directamente a los países occidentales, donde las universidades no recopilan datos tan extensos sobre sus estudiantes— debemos considerar la hipótesis de que podríamos haber enfrentado un filtrado implícito. Dicho filtrado estaría relacionado con el hecho de que Universidad Politécnica Salesiana es una universidad privada con un costo de inscripción, matrícula y tarifas significativo para las familias del Ecuador, lo cual puede tener el efecto de que los registros de los estudiantes en las bases de datos de esta universidad conforman una población uniforme en términos socioeconómicos.

Por supuesto, nuestros hallazgos no implican que no haya propósito alguno en la recopilación de datos extensos sobre los estudiantes, ya que el Departamento de Bienestar Estudiantil de la Universidad Politécnica Salesiana los utiliza para impulsar sus muchas actividades en la promoción de la equidad, el bienestar psicológico y la salud, y la empleabilidad. Además, los sistemas de Inteligencia Artificial aún pueden utilizar los datos para asesorar a los estudiantes actuales a través de recomendaciones de cursos, opciones de trayectoria profesional, y demás usos explicados en el Capítulo 3.

#### 7.1.4. Perspectiva general

Creemos que con la investigación y los desarrollos llevados a cabo, esta tesis doctoral —con la debida labor de transferencia, a completar en años venideros— puede contribuir de manera efectiva a la denodada labor que realizan las instituciones de Educación Superior, de modo que docentes, estudiantes y personal administrativo puedan acometer de una forma más eficiente sus múltiples actividades y tareas. Las técnicas y los resultados se consideran de aplicación general no sólo a otras universidades del ámbito de Latinoamérica —que ofrecen grandes similitudes con los modelos estructurales y organizativos de la Universidad Politécnica Salesiana— sino también en el contexto internacional más amplio, pues los fundamentos semánticos de las distintas propuestas las hacen fácilmente trasladables a la realidad de las instituciones de Educación Superior en otras partes del mundo.

## 7.2. Trabajo futuro

Las principales líneas de trabajo futuro que nos planteamos tras el desarrollo de esta tesis siguen las líneas de los aportes principales:

- En lo tocante al modelado formal de conocimiento del ámbito de la Educación Superior, consideramos que la red ontológica OEA sienta unas bases muy adecuadas, que sería conveniente potenciar y transferir a nuevos entornos, en los cuales muy probablemente surgirían necesidades nuevas (e.g. de integración de otros tipos de bases de datos legadas o de consolidación de conocimiento) de las que habría que dar debida cuenta según lo ya previsto por la metodología NeOn. Se trataría, previsiblemente, de un trabajo con una componente menor de investigación, pero un esfuerzo de desarrollo notable.
- En la aplicación de algoritmos genéticos para la generación de grupos de aprendizaje, consideramos oportuno dar continuidad a la experimentación llevada a cabo y ampliarla a nuevas asignaturas a fin de acumular un registro de evidencias extensivo, que pueda brindarse abiertamente en Internet —con las salvaguardas requeridas por las normativas de protección de datos— para potenciar la investigación en esta línea de manos de terceros. Igualmente, tendría sentido crear un banco de pruebas que permitiese la comparación de distintos métodos de generación de grupos, donde la literatura especializada no contiene suficientes detalles como para replicar propuestas previas. Sería esta una piedra de toque para cuantificar definitivamente los aportes de nuestra propuesta, y evaluar la conveniencia de incorporar nuevos datos o consideraciones en la lógica de los procedimientos de selección, cruce y mutación.
- En la línea de la predicción de rendimiento académico a través de técnicas de aprendizaje profundo, hemos iniciado negociaciones para verificar la hipótesis del prefiltrado implícito en la UPS mediante la realización de un estudio similar en el contexto de una universidad pública ecuatoriana. Si obtuviéramos los mismos resultados, entonces eso sería al menos una evidencia de que la adopción generalizada de las normativas en materia de protección de datos no está obstaculizando los usos potenciales de la Inteligencia Artificial para mejorar los sistemas de Educación Superior, sino que solo evita usos indebido de los datos personales, como está en el espíritu de las normas. En cambio, si los resultados fuesen distintos y en el nuevo contexto sí fuera posible realizar predicciones más allá de las estadísticas básicas, entonces —además del hallazgo en sí de la disparidad entre instituciones públicas y privadas— se haría pertinente un estudio entre múltiples tipos de redes neuronales para identificar aquel que rindiese los mejores resultados, cubriendo redes de creencia profunda, redes recurrentes, redes convolucionales, máquinas profundas de Boltzmann, etc.

Por otra parte, planteamos la hipótesis de que los datos que actualmente maneja la UPS podrían mejorarse con datos detallados sobre la historia de los estudiantes en educación secundaria. Igualmente, creemos que podría ser valioso cruzar campos de datos detallados sobre temas relacionados (por ejemplo, en diferentes ramas de la ciencia) o incluso temas específicos dentro de ellos, toda vez que, por ejemplo, a un estudiante que era hábil con la trigonometría pero tenía dificultades con los derivados es probable que rinda mejor en materias relacionadas con el álgebra que con la física. Esta idea enfrenta tres desafíos importantes: (i) la necesidad de combinar múltiples fuentes de datos, con diferentes formatos y niveles de granularidad, (ii) el hecho de que estaríamos dividiendo los datos disponibles para la capacitación, y (iii) el he-

---

cho de que las regulaciones de protección de datos imponen severos límites a la fusión de bases de datos en manos de diferentes instituciones.

Finalmente, cabe comentar que nuestra red ontológica allana el camino para la exploración de nuevas aplicaciones de Inteligencia Artificial en el ámbito de la Educación Superior, tocando las demás ideas mencionadas en los primeros capítulos de esta tesis y muchas más: generación de contenido adaptado a los perfiles individuales/grupales de los estudiantes, sistemas de tutoría inteligentes, planificación de espacios y equipamientos, identificación de temas de interés, reutilización de contenidos, programas de apoyo a alumnos noveles por parte de veteranos, etc. Creemos que el recorrido potencial es inmenso, una vez que los datos existen y se pueden acomodar de manera integrada y estandarizada.



**Parte V**

**Bibliografía**



# Bibliografía

- [1] Abdellatif, M., and Asma, K. A functional approach of knowledge management system applied to institutions of higher education. In *4th International Symposium ISKO-Maghreb: Concepts and Tools for knowledge Management (ISKO-Maghreb)* (2014).
- [2] Abelló, A., Romero, O., Pedersen, T. B., Berlanga, R., Nebot, V., Aramburu, M.J., and Simitsis, A. Using Semantic Web technologies for exploratory OLAP: A survey. *IEEE Transactions on Knowledge and Data Engineering* 27, 2 (2015), 571–588.
- [3] Adeli, H. *Expert systems in construction and structural engineering*. CRC Press, 2014.
- [4] Aguilera-Luque, A. M., Zurutuza, U., Arrazola, P. J., Igartua-Lopez, J. I., Galdos, L., de Argandoña, E. S., Mendiguren, J., Uribeetxeberria, R., Aurrekoetxea, J., and Serrano-Jiménez, D. La gestión del conocimiento organizacional. *International Journal of Innovation Management* 22, 2 (2018).
- [5] Ahuja, R. K., Orlin, J. B., and Tiwari, A. A greedy genetic algorithm for the quadratic assignment problem. *Computers & Operations Research* 27, 10 (2000), 917–934.
- [6] Alan, Y., Alparslan, A., and Dittmann, L. Werkzeuge zur Sicherstellung der Adaptibilität des KOWIEN-Vorgehensmodells. *Projektbericht 6* (2003).
- [7] Alberola, J. M., del Val, E., Sanchez-Anguix, V., and Julian, V. Simulating a collective intelligence approach to student team formation. In *International Conference on Hybrid Artificial Intelligence Systems* (2013), Springer, pp. 161–170.
- [8] Alberola, J. M., Del Val, E., Sanchez-Anguix, V., Palomares, A., and Teruel, M. D. An artificial intelligence tool for heterogeneous team formation in the classroom. *Knowledge-Based Systems* 101 (2016), 1–14.
- [9] Allemang, D., and Hendler, J. *Semantic Web for the working ontologist: Effective modeling in RDFS and OWL*. Elsevier, 2011.
- [10] Almarabeh, H. Analysis of students’ performance by using different data mining classifiers. *International Journal of Modern Education and Computer Science* 9, 8 (2017), 9.

- [11] Alsheddy, A., and Habib, M. On the application of data mining algorithms for predicting student performance: A case study. *International Journal of Computer Science and Network Security* 17, 10 (2017), 189–197.
- [12] Ambridge, B. *Psy-Q: You know your IQ, now test your psychological intelligence*. Profile Books, 2014.
- [13] Amrieh, E. A., Hamtini, T., and Aljarah, I. Mining educational data to predict student's academic performance using ensemble methods. *International Journal of Database Theory and Application* 9, 8 (2016), 119–136.
- [14] Antoniou, G., and Van Harmelen, F. *A semantic web primer*. MIT Press, 2004.
- [15] Artz, D., and Gil, Y. A survey of trust in computer science and the Semantic Web. *Journal of Web Semantics* 5, 2 (2007), 58–71.
- [16] Back, T. *Evolutionary algorithms in theory and practice: Evolution strategies, evolutionary programming, genetic algorithms*. Oxford University Press, 1996.
- [17] Balmaceda, J. M., Schiaffino, S. N., and Pace, J. A. D. Using constraint satisfaction to aid group formation in CSCL. *Revista Iberoamericana de Inteligencia Artificial* 17, 53 (2014), 35–45.
- [18] Bartfai, G. Hierarchical clustering with ART neural networks. In *IEEE World Congress on Computational Intelligence* (1994).
- [19] Bekerman, D. G., and Dankner, L. A. La pareja pedagógica en el ámbito universitario, un aporte a la didáctica colaborativa. *Formación universitaria* 3, 6 (2010), 3–8.
- [20] Belbin, R. M. *Management teams*. Routledge, 2010.
- [21] Belbin, R. M. *Team roles at work*. Routledge, 2012.
- [22] Bench-Capon, T. J. *Knowledge representation: An approach to artificial intelligence*. Academic Press Professional, 1990.
- [23] Benítez, R., Escudero, G., Kanaan, S., and Rodó, D. M. *Inteligencia artificial avanzada*. Editorial UOC, 2014.
- [24] Bergey, P., and King, M. Team machine: A decision support system for team formation. *Decision Sciences Journal of Innovative Education* 12, 2 (2014), 109–130.
- [25] Berners-Lee, T. Semantic Web on XML. <http://www.w3.org/2000/Talks/1206-xml2k-tb1/Overview.html>, 2000.
- [26] Bernus, P., Schmidt, G., and Shaw, M. *International handbook on information systems*. Springer, 1998.
- [27] Bimonte, S., Sautot, L., Journaux, L., and Faivre, B. Multidimensional model design using data mining: A rapid prototyping methodology. *International Journal of Data Warehousing and Mining* 13, 1 (2017), 1–35.
- [28] Blauberger, I. V. *Diccionario marxista de filosofía*. Ediciones de Cultura popular, 1972.

- [29] Block, H.-D. The perceptron: A model for brain functioning. *Reviews of Modern Physics* 34, 1 (1962), 123.
- [30] Boella, G., Di Caro, L., Humphreys, L., Robaldo, L., Rossi, P., and van der Torre, L. Eunomos: A legal document and knowledge management system for the web to provide relevant, reliable and up-to-date information on the law. *Artificial Intelligence and Law* 24, 3 (2016), 245–283.
- [31] Boley, H., Tabet, S., and Wagner, G. Design rationale of RuleML: A markup language for semantic web rules. In *1st International Conference on Semantic Web Working (SWWS)* (2001).
- [32] Borst, W. N. *Construction of engineering ontologies for knowledge sharing and reuse*. Universiteit Twente, 1997.
- [33] Bote, V. P. G., and Tello, A. J. L. Vínculos entre las ontologías y la biblioteconomía y documentación. In *IV Congreso ISKO-España EOCONSID* (1999).
- [34] Bowker, G. How to be universal: Some cybernetic strategies, 1943-70. *Social Studies of Science* 23, 1 (1993), 107–127.
- [35] Bozu, Z., and Canto, P. J. El profesorado universitario en la sociedad del conocimiento: Competencias profesionales docentes. *Revista de Formación e Innovación Educativa Universitaria* 2, 2 (2009), 87–97.
- [36] Bramlette, M. F., and Bouchard, E. E. *Handbook of genetic algorithms*. Van Nostrand Reinhold, 1991, ch. Genetic algorithms in parametric design of aircraft, pp. 109–123.
- [37] Calabor, M. S., Mora, A., and Moya, S. Adquisición de competencias a través de juegos serios en el área contable: Un análisis empírico. *Revista de Contabilidad* 21, 1 (2018), 38–47.
- [38] Cannady, J. Artificial neural networks for misuse detection. In *National Information Systems Security Conference* (1998).
- [39] Carroll, J. B. *The role of constructs in psychological and educational measurement*. Lawrence Erlbaum Associates Publishers, 2002, ch. The Five-Factor personality model: How complete and satisfactory is it?, pp. 97–126.
- [40] Castells, M., Gimeno, C. M., and Alborés, J. *La sociedad red*. Alianza, 2005.
- [41] Cen, L., Ruta, D., Powell, L., and Ng, J. Interaction driven composition of student groups for optimal groupwork learning performance. In *IEEE Frontiers in Education Conference (FIE)* (2015).
- [42] Ceruzzi, P. Crossing the divide: Architectural issues and the emergence of the stored program computer. *IEEE Annals of the History of Computing* 19, 1 (1997), 5–12.
- [43] Chapman, A. Análisis DOFA y análisis PEST. <https://catedramacagnot3.files.wordpress.com/2012/03/anc3a1lisis-dofa-y-pest.pdf>, 2004.
- [44] Chapra, S. C., and Canale, R. P. *Métodos numéricos para ingenieros*, 5 ed. Regresión por mínimos cuadrados, 2007, ch. 17, pp. 466–486.

- [45] Chapra, S. C., and Canale, R. P. *Métodos numéricos para ingenieros*, 5 ed. Interpolación mediante trazadores, 2007, ch. 18, pp. 525–535.
- [46] Chapra, S. C., and Canale, R. P. *Métodos numéricos para ingenieros*, 5 ed. Métodos de Runge-Kutta, 2007, ch. 25, pp. 740–756.
- [47] Chen, H., Chiang, R. H., and Storey, V. C. Business intelligence and analytics: from big data to big impact. *MIS quarterly* (2012), 1165–1188.
- [48] Chen, H., Perich, F., Finin, T., and Joshi, A. Soupa: Standard ontology for ubiquitous and pervasive applications. In *1st Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MOBIQUITOUS)* (2004).
- [49] Cherif, A. H., Jedlicka, D., Verma, S., Uddin, K., and Movahedzadeh, F. Brain talking: Classroom activity to engage students in deep and meaningful learning. *Brain* 8, 32 (2017).
- [50] Chowdhury, G. G. Natural language processing. *Annual Review of Information Science and Technology* 37, 1 (2003), 51–89.
- [51] Christodouloupoulos, C. E., and Papanikolaou, K. A. A group formation tool in an e-learning context. In *19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI)* (2007).
- [52] Chung, H., and Kim, J. An ontological approach for semantic modeling of curriculum and syllabus in higher education. *International Journal of Information and Education Technology* 6, 5 (2016), 365–369.
- [53] Cocea, M., and Magoulas, G. D. User behaviour-driven group formation through case-based reasoning and clustering. *Expert Systems with Applications* 39, 10 (2012), 8756–8768.
- [54] Cochocki, A., and Unbehauen, R. *Neural networks for optimization and signal processing*. John Wiley & Sons, 1993.
- [55] Cohen, E. G., and Lotan, R. A. *Designing Groupwork: Strategies for the Heterogeneous Classroom Third Edition*. Teachers College Press, 2014.
- [56] Cohen, P. R., and Feigenbaum, E. A. *The handbook of Artificial Intelligence*. Butterworth-Heinemann, 2014.
- [57] Colmenares, A. M., and Piñero, M. L. La investigación acción: Una herramienta metodológica heurística para la comprensión y transformación de realidades y prácticas socio-educativas. *Laurus* 14, 27 (2008), 96–114.
- [58] Coro Montanet, G., Suárez García, A., Gómez Sánchez, M., and Gómez Polo, F. Didáctica de la introducción y uso de simuladores hápticos con entornos 3D en la docencia odontológica. In *XII Jornadas Internacionales de Innovación Universitaria* (2015).
- [59] Costaguta, R. Algorithms and machine learning techniques in collaborative group formation. *Lecture Notes in Computer Science* 9414 (2015), 249–258.
- [60] Costaguta, R., de los ángeles Menini, M., and Lescano, G. Applying data mining to discover successful collaborative groups styles. In *8th Euro-American Conference on Telematics and Information Systems (EATIS)* (2016).

- [61] Cupani, M., Pilatti, A., Urrizaga, A., Chincolla, A., and de Minzi, M. C. R. Inventario de personalidad IPIP-NEO: Estudios preliminares de adaptación al español en estudiantes argentinos. *Revista Mexicana de Investigación en Psicología* 6, 1 (2014), 55–73.
- [62] Cybenko, G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems* 2, 4 (1989), 303–314.
- [63] Cárdenas, C., Ramiro, L., Cajamarca Criollo, O. A., and Mantilla Crespo, X. A. Impacto del uso de simuladores en la enseñanza de la administración financiera. *Innovación Educativa* 17, 75 (2017), 103–122.
- [64] da Silva, I. N. *Artificial Neural Networks: A Practical Course*. Springer, 2016.
- [65] Daafouz, J., Riedinger, P., and Jung, C. Stability analysis and control synthesis for switched systems: A switched Lyapunov function approach. *IEEE Transactions on Automatic Control* 47, 11 (2002), 1883–1887.
- [66] Dalkir, K. *Knowledge Management: Converting Theory into Practice*. Butterworth-Heinemann, 2005.
- [67] Dattakumar, A., Chong, G., Malone, L., Sharma, R. S., and Valenzuela, J. F. Knowledge societies and their role in sustainable development. In *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (2016).
- [68] Davenport, T. H., Prusak, L., et al. *Information ecology: Mastering the information and knowledge environment*. Oxford University Press on Demand, 1997.
- [69] Davenport, T. H., Prusak, L., et al. *Working knowledge: How organizations manage what they know*. Harvard Business Press, 1998.
- [70] Davies, M. Concept mapping, mind mapping and argument mapping: What are the differences and do they matter? *Higher Education* 62, 3 (2011), 279–301.
- [71] Davis, L. *Handbook of genetic algorithms*. Van Nostrand Reinhold, 1991.
- [72] De Jong, K. *An analysis of the behavior of a class of genetic adaptive systems*. PhD thesis, University of Michigan, 1975.
- [73] Deibel, K. Team formation methods for increasing interaction during in-class group work. *ACM SIGCSE Bulletin* 37 (2005), 291–295.
- [74] Diestel, R. *Graph Theory*. Springer-Verlag, 2000.
- [75] Dillenbourg, P., and Schneider, D. Collaborative learning and the Internet. In *International Conference on Computer Assisted Instructions (ICCAI)* (1995).
- [76] Drucker, P. F. *Las nuevas realidades en el estado y la política, en la economía y los negocios, en la sociedad y en la imagen del mundo*. Edhasa, Ciencias de la Dirección, 1989.
- [77] Duque, R., Gómez-Pérez, D., Nieto-Reyes, A., and Bravo, C. Analyzing collaboration and interaction in learning environments to form learner groups. *Computers in Human Behavior* 47 (2015), 42–49.

- [78] Dutta, S., and De Meyer, A. *Knowledge Management and Business Model Innovation*. IGI Publishing, 2001, ch. Knowledge management at Arthur Andersen (Denmark): Building assets in real time and in virtual space, pp. 284–401.
- [79] Dwan, T., and Ownsworth, T. The Big Five personality factors and psychological well-being following stroke: A systematic review. *Disability and Rehabilitation* 41, 10 (2019), 1119–1130.
- [80] Díaz, F. J., Queiruga, C. A., and Fava, L. A. Juegos serios y educación. In *XVII Workshop de Investigadores en Ciencias de la Computación* (2015).
- [81] Díaz-Barriga Arceo, F., and Hernández Rojas, G. *Estrategias docentes para un aprendizaje significativo. Una interpretación constructivista*. McGraw Hill, 2002.
- [82] Eom, S. B. A survey of operational expert systems in business (1980–1993). *Interfaces* 26, 5 (1996), 50–70.
- [83] Eppler, M. J. A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. *Information Visualization* 5, 3 (2006), 202–210.
- [84] Ertel, W. *Introduction to Artificial Intelligence*. Springer, 2018.
- [85] Eshelman, L. J. *Foundations of genetic algorithms*, vol. 1. Elsevier, 1991, ch. The CHC adaptive search algorithm: How to have safe search when engaging in non-traditional genetic recombination, pp. 265–283.
- [86] Fernández, B. La comunicación para el desarrollo en la sociedad del conocimiento. *Comunicación: Estudios venezolanos de comunicación* 40 (2009), 171–172.
- [87] Fiad, S. B., and Galarza, O. D. El laboratorio virtual como estrategia para el proceso de enseñanza-aprendizaje del concepto de mol. *Formación Universitaria* 8, 4 (2015), 3–14.
- [88] Forero-Páez, Y., and Giraldo, J. A. Simulación de un proceso de fabricación de bicicletas: Aplicación didáctica en la enseñanza de la ingeniería industrial. *Formación Universitaria* 9, 3 (2016), 39–50.
- [89] Fowler, J. W., Wirojanagud, P., and Gel, E. S. Heuristics for workforce planning with worker differences. *European Journal of Operational Research* 190, 3 (2008), 724–740.
- [90] Fox, B., and McMahon, M. *Foundations of genetic algorithms*, vol. 1. Elsevier, 1991, ch. Genetic operators for sequencing problems, pp. 284–300.
- [91] Friedman, H. S., and Schustack, M. W. *Personality: Classic theories and modern research*. Allyn & Bacon, 1999.
- [92] Frohman, R. G. Collaborative efforts work! reflections on a two-year relationship between faculty of health and international student services-language and learning unit. *Journal of Academic Language and Learning* 6, 3 (2012), 48–58.
- [93] Fukushima, K. Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics* 20, 3-4 (1975), 121–136.

- [94] García, J. A. P., and Ayuga, J. M. D. La integración de las TIC y la gamificación en la enseñanza de español como LE/L2: Propuesta de adaptación para las universidades japonesas. *Cuadernos CANELA 28* (2017), 79–101.
- [95] García-Vélez, R., Galán-Mena, J., Robles-Bykbaev, V. E., López-Nores, M., Pesántez-Avilés, F., and Calle-López, D. An academic system based on ontological networks to support the inference of new knowledge in micro and macro curriculum in higher education. In *International Conference on Applied Human Factors and Ergonomics* (2018).
- [96] Genesereth, M. R., and Nilsson, N. J. *Logical foundations of artificial intelligence*. Morgan Kaufmann, 1987.
- [97] Goh, A. T. C. Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering 9*, 3 (1995), 143–151.
- [98] Goldberg, D. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.
- [99] Goldberg, L. R. The structure of phenotypic personality traits. *American Psychologist 48*, 1 (1993), 26–34.
- [100] Good, T. L., and Brophy, J. E. *Educational psychology: A realistic approach*. Longman/Addison Wesley Longman, 1990.
- [101] Gruber, T. R. A translation approach to portable ontology specifications. *Knowledge Acquisition 5*, 2 (1993), 199–220.
- [102] Grusec, J. E. Social learning theory and developmental psychology: The legacies of Robert Sears and Albert Bandura. *Developmental Psychology 28*, 5 (1992), 776–786.
- [103] Guangzuo, C., Fei, C., Hu, C., and Shufang, L. OntoEdu: a case study of ontology-based education grid system for e-learning. In *GCCCE International Conference* (2004).
- [104] Guarino, N., Oberle, D., and Staab, S. *Handbook on ontologies*. Springer, 2009, ch. What is an ontology?, pp. 1–17.
- [105] Guralnick, D. Putting the education into educational simulations: Pedagogical structures, guidance and feedback. *International Journal of Advanced Corporate Learning 2*, 1 (2009), 10–15.
- [106] Gutiérrez Hernández, R., Muñoz-Arteaga, J., Cruz Flores, R., and Alonso Lavernia, M. Objetos de aprendizaje multiculturales con juegos serios. In *V Conferencia Conjunta Iberoamericana sobre Tecnologías y Aprendizaje* (2013).
- [107] Hampshire, J. B., and Pearlmutter, B. *Connectionist Models*. Morgan Kaufmann, 1991, ch. Equivalence proofs for multi-layer perceptron classifiers and the Bayesian discriminant function.
- [108] Haykin, S. *Neural networks: A comprehensive foundation*. Prentice Hall PTR, 1994.

- [109] Heaton, J. The number of hidden layers. <https://www.heatonresearch.com/2017/06/01/hidden-layers.html>, 2017.
- [110] Henning, P. A., Heberle, F., Streicher, A., Zielinski, A., Swertz, C., Bock, J., and Zander, S. Personalized web learning: Merging open educational resources into adaptive courses for higher education. In *International Workshop on Personalization Approaches in Learning Environments (PALE)* (2014).
- [111] Hilerá González, J. R. *Redes neuronales artificiales: Fundamentos, modelos y aplicaciones*. RA-MA S.A. Editorial y Publicaciones, 1995.
- [112] Hinton, G. E., Sejnowski, T. J., et al. Learning and relearning in Boltzmann machines. *Parallel Distributed Processing 1* (1986), 282–317.
- [113] Hirschberg, J., and Manning, C. D. Advances in natural language processing. *Science 349*, 6245 (2015), 261–266.
- [114] Hislop, D., Bosua, R., and Helms, R. *Knowledge management in organizations: A critical introduction*. Oxford University Press, 2018.
- [115] Hodges, A. *Alan Turing: The Enigma*. Random House, 2012.
- [116] Holland, J. H. Adaptation in artificial and natural systems. *The University of Michigan Press* (1975).
- [117] Holland, J. H. *Adaption in Natural and Artificial Systems*. MIT Press, 1992.
- [118] Holland, J. H. *Hidden order: How adaptation builds complexity*. Basic Books, 1995.
- [119] Holland, J. H., and Goldberg, D. Genetic algorithms in search, optimization and machine learning. *Addison-Wesley* (1989).
- [120] Hooshyar, D., Binti Ahmad, R., Wang, M., Yousefi, M., Fathi, M., and Lim, H. Development and evaluation of a game-based Bayesian intelligent tutoring system for teaching programming. *Journal of Educational Computing Research 56*, 6 (2017), 775–801.
- [121] Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences 79*, 8 (1982), 2554–2558.
- [122] Hopfield, J. J. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences 81*, 10 (1984), 3088–3092.
- [123] Hornik, K., Stinchcombe, M., and White, H. Multilayer feedforward networks are universal approximators. *Neural networks 2*, 5 (1989), 359–366.
- [124] Horrocks, I., Patel-Schneider, P. F., Boley, H., Tabet, S., Grosz, B., Dean, M., et al. SWRL: A Semantic Web rule language combining OWL and RuleML. <https://www.w3.org/Submission/SWRL/>, 2004.
- [125] Hutter, M., and Legg, S. Fitness uniform optimization. *IEEE Transactions on Evolutionary Computation 10*, 5 (2006), 568–589.

- [126] Ichikawa, Y., and Ishii, Y. Retaining diversity of genetic algorithms for multivariable optimization and neural network learning. In *IEEE International Conference on Neural Networks* (1993).
- [127] Inhelder, B., Chipman, H., and Zwingmann, C., Eds. *Piaget and his school - A Reader in Developmental Psychology*. Springer, 1976.
- [128] Jackson, P. *Introduction to expert systems*. Addison-Wesley Longman Publishing, 1998.
- [129] Jain, A., Kulkarni, G., and Shah, V. Natural Language Processing. *International Journal of Computer Sciences and Engineering* 6, 1 (2018), 161-167.
- [130] Jensen, F. V., and Nielsen, T. D. *Bayesian networks and decision graphs*, 2001.
- [131] John, O. P., and Srivastava, S. *Handbook of personality: Theory and research*, vol. 2. Guilford, 1999, ch. The Big Five traits taxonomy: History, measurement, and theoretical perspectives, pp. 102-138.
- [132] Johnson, D., and Johnson, R. *Learning together and alone: Cooperative, competitive, and individualistic learning*. Pearson, 1999.
- [133] Johnson, D. W., and Johnson, F. P. *Joining together: Group theory and group skills*. Allyn & Bacon, 2013.
- [134] Johnson, J. A. Measuring thirty facets of the Five Factor model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality* 51 (2014), 78-89.
- [135] Juguera Rodríguez, L., Díaz Agea, J. L., Pérez Lapuente, M., Leal Costa, C., Rojo Rojo, A., and Echevarría Pérez, P. La simulación clínica como herramienta pedagógica: Percepción de los alumnos de Grado en Enfermería en la Universidad Católica San Antonio de Murcia. *Enfermería Global* 13, 33 (2014), 175-190.
- [136] Jurado, J. L., and Bustamante, H. M. Método de especificación de patrones colaborativos para plataformas de ciencia, un enfoque desde la gestión de conocimiento. *Campus Virtuales* 6, 1 (2017), 23-37.
- [137] Kagan, S., and Kagan, M. *Kagan Cooperative Learning*. Kagan Publishing, 2015.
- [138] Karan, E. P., Irizarry, J., and Haymaker, J. BIM and GIS integration and interoperability based on Semantic Web technology. *Journal of Computing in Civil Engineering* 30, 3 (2016).
- [139] Kardan, A. A., and Sadeghi, H. An efficacious dynamic mathematical modelling approach for creation of best collaborative groups. *Mathematical and Computer Modelling of Dynamical Systems* 22, 1 (2016), 39-53.
- [140] King, W. R. Knowledge management and organizational learning. In *Knowledge management and organizational learning*. Springer, 2009, pp. 3-13.
- [141] Knight, K. Connectionist ideas and algorithms. *Communications of the ACM* 33, 11 (1990), 58-74.

- [142] Knublauch, H., Hendler, J. A., and Idehen, K. SPIN: Overview and motivation. <https://www.w3.org/Submission/spin-overview/>, 2011.
- [143] Kohonen, T., and Honkela, T. Kohonen network. *Scholarpedia* 2, 1 (2007), 1568.
- [144] Komarraju, M., Karau, S. J., Schmeck, R. R., and Avdic, A. The Big Five personality traits, learning styles, and academic achievement. *Personality and Individual Differences* 51, 4 (2011), 472-477.
- [145] Kosko, B. Bidirectional associative memories. *IEEE Transactions on Systems, man, and Cybernetics* 18, 1 (1988), 49-60.
- [146] Koza, J. R. *Genetic programming II - Automatic discovery of reusable subprograms*. MIT Press, 1992.
- [147] Kuo, Y.-C., Chu, H.-C., and Huang, C.-H. A learning style-based grouping collaborative learning approach to improve EFL students' performance in English courses. *Journal of Educational Technology & Society* 18, 2 (2015).
- [148] Kurzweil, R., Richter, R., and Schneider, M. L. *The age of intelligent machines*. MIT Press, 1990.
- [149] Laoufi, A., Mouhim, S., Cherkaoui, C., et al. An ontology based architecture to support the knowledge management in higher education. In *International Conference on Multimedia Computing and Systems (ICMCS)* (2011).
- [150] Larrañaga, P. *Estimation of distribution algorithms*. Springer, 2002, ch. A review on estimation of distribution algorithms, pp. 57-100.
- [151] Latham, A., Crockett, K., and McLean, D. An adaptation algorithm for an intelligent natural language tutoring system. *Computers & Education* 71 (2014), 97-110.
- [152] Lee, T., Lee, I.-h., Lee, S., Lee, S.-g., Kim, D., Chun, J., Lee, H., and Shim, J. Building an operational product ontology system. *Electronic Commerce Research and Applications* 5, 1 (2006), 16-28.
- [153] Lescano, G., Costaguta, R., and Amandi, A. Genetic algorithm for automatic group formation considering student's learning styles. In *8th Euro-American Conference on Telematics and Information Systems (EATIS)* (2016).
- [154] Lewin, K., Lippitt, R., and White, R. K. Patterns of aggressive behavior in experimentally created social climates. *Journal of Social Psychology* 10, 2 (1939), 269-299.
- [155] León-Parra, E., and Cañas-Coto, F. Modelos de simulación en la Escuela de Administración de Negocios, Universidad de Costa Rica. *InterSedes: Revista de las Sedes Regionales* 15, 31 (2014).
- [156] Li, G., and Xu, Z. A theoretical analysis on mutation operator of standard genetic algorithm. *Wuhan University Journal of Natural Sciences* 1, 3-4 (1996), 599-604.

- [157] Li, J., Zhang, Y., Du, D., and Liu, Z. Improvements in the decision making for cleaner production by data mining: Case study of vanadium extraction industry using weak acid leaching process. *Journal of Cleaner Production* 143 (2017), 582–597.
- [158] Lie, A. *Cooperative learning*. Grasindo, 2002.
- [159] Liepins, G. E., and Vose, M. D. Characterizing crossover in genetic algorithms. *Annals of Mathematics and Artificial Intelligence* 5, 1 (1992), 27–34.
- [160] Lin, Y.-T., Huang, Y.-M., and Cheng, S.-C. An automatic group composition system for composing collaborative learning groups using enhanced particle swarm optimization. *Computers & Education* 55, 4 (2010), 1483–1493.
- [161] Loes, C. N., and Pascarella, E. T. Collaborative learning and critical thinking: Testing the link. *The Journal of Higher Education* 88, 5 (2017), 726–753.
- [162] Luckin, R., Holmes, W., Griffiths, M., and Forcier, L. B. *Intelligence unleashed: An argument for AI in education*. Pearson Education, 2016.
- [163] López-González, E., and Ruiz-Soler, M. Análisis de datos con el Modelo Lineal Generalizado. una aplicación con R. *Revista Española de Pedagogía* LXIX, 248 (2011), 59–80.
- [164] Maimon, O., and Rokach, L. *Introduction to Knowledge Discovery and Data Mining*. Springer, 2010, pp. 1–15.
- [165] Martín-Smith, P., Pelayo, F. J., Díaz, A., Ortega, J., and Prieto, A. A learning algorithm to obtain self-organizing maps using fixed neighbourhood Kohonen networks. In *International Workshop on Artificial Neural Networks* (1993).
- [166] Martínez, C. G. *Algoritmos genéticos locales*. Universidad de Granada, 2008.
- [167] Martínez Miguélez, M. *Ciencia y arte en la metodología cualitativa*. MAD, 2006.
- [168] Martínez-Sánchez, A., Hernández-Arzola, L. I., and Jiménez, C. Uso de simuladores clínicos con estudiantes de enfermería en la Universidad de la Sierra Sur de Oaxaca. *Revista de Enfermería del Instituto Mexicano del Seguro Social* 24, 3 (2016), 223–228.
- [169] Medina Garrido, J. Elementos en la gestión del conocimiento. una visión de teoría basada en recursos. In *XIV Congreso Nacional y X Congreso hispano-francés AEDEM* (2000).
- [170] Mendoza Vargas, F. D. Uso de simuladores de negocio como estrategia de aprendizaje adaptativo: Una experiencia en el aula. *Virtu@lmente* 5, 2 (2018).
- [171] Mesulam, M.-M. Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology* 28, 5 (1990), 597–613.
- [172] Michalewics, Z. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, 1992.
- [173] Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. *Machine learning: An artificial intelligence approach*. Springer Science & Business Media, 2013.

- [174] Minsky, M. L. A framework for representing knowledge. Tech. rep., Massachusetts Institute of Technology, 1974.
- [175] Minsky, M. L., and Papert, S. A. *Perceptrons: An introduction to computational geometry*. MIT Press, 1969.
- [176] Mira, J. La inteligencia artificial como ciencia y como ingeniería. In *50 Años de la Inteligencia Artificial – XVI Escuela de Verano de Informática* (2006).
- [177] Mishra, T., Kumar, D., and Gupta, S. Mining students’ data for performance prediction. In *International Conference on Advanced Computing & Communication Technologies* (2014).
- [178] Mitrovic, A., Ohlsson, S., and Barrow, D. K. The effect of positive feedback in a constraint-based intelligent tutoring system. *Computers & Education* 60, 1 (2013), 264–272.
- [179] Moody, J., and Darken, C. J. Fast learning in networks of locally-tuned processing units. *Neural Computation* 1, 2 (1989), 281–294.
- [180] Mora-Arciniegas, M.-B., Piedra, N., and Tenesaca-Luna, G.-A. Semantic representation of teaching planning, pilot experience at UTPL. In *IEEE 37th Central America and Panama Convention (CONCAPAN)* (2017).
- [181] Moreno, J., Ovalle, D. A., and Vicari, R. M. A genetic algorithm approach for group formation in collaborative learning considering multiple student characteristics. *Computers & Education* 58, 1 (2012), 560–569.
- [182] Moscoso-Zea, O., Luján-Mora, S., et al. Datawarehouse design for educational data mining. In *15th International Conference on Information Technology-based Higher Education and Training (ITHET)* (2016).
- [183] Mumma, F. S. *Team-Work & Team-Roles*. Hrdq, 1992.
- [184] Muñoz, A., López, V., and Díaz, V. Modelo de gestión de conocimiento para el Núcleo Universitario Santa Lucía. *KnE Engineering* 3, 1 (2018), 192.
- [185] Najadat, H., Al-Bdarneh, A., and Qawasmi, H. Accreditation software for higher education programs. In *15th International Conference on Information Technology Based Higher Education and Training (ITHET)* (2016).
- [186] Natek, S., and Zwilling, M. Student data mining solution–knowledge management system related to higher education institutions. *Expert Systems with Applications* 41, 14 (2014), 6400–6407.
- [187] Newell, A. Report on a general problem solving program. Tech. rep., Carnegie Institute of Technology, 1959.
- [188] Newman, A. Learning to adapt: A case for accelerating adaptive learning in higher education. <http://tytonpartners.com/library/accelerating-adaptive-learning-in-higher-education/>, 2013.
- [189] Niño de Guzmán, I., Calderón, A., and Cassaretto, M. Personalidad y rendimiento académico en estudiantes universitarios. *Revista de Psicología de la Pontificia Universidad Católica del Perú* XXI, 1 (2003), 119–143.

- [190] Nokes-Malach, T. J., Richey, J. E., and Gadgil, S. When is it better to learn together? insights from research on collaborative learning. *Educational Psychology Review* 27, 4 (2015), 645–656.
- [191] Nonaka, I., and Takeuchi, H. La empresa creadora de conocimiento. *Gestión del conocimiento* (2000), 1–9.
- [192] Novikoff, A. B. On convergence proofs for perceptrons. Tech. rep., Stanford Research Institute at Menlo Park, 1963.
- [193] Ochirbat, A., Namsraidorj, M., and Hwang, W. Y. Small k-teams recommendation in social learning networks. In *7th International Conference on Ubi-Media Computing and Workshops (U-MEDIA)* (2014).
- [194] Olivar, A., and Daza, A. Las tecnologías de la información y comunicación (TIC) y su impacto en la educación del siglo XXI. *Negotium: Revista de ciencias gerenciales* 3, 7 (2007).
- [195] Ormrod, J. E. *Human learning*. Pearson Higher Education, 2016.
- [196] Ortiz, A. Didáctica problematizadora y aprendizaje basado en problemas. *Barranquilla: Ediciones Litoral* (2009).
- [197] Ortiz-Colón, A. M., Jordán, J., and Agredal, M. Gamificación en educación: una panorámica sobre el estado de la cuestión. *Educação e Pesquisa* 44 (2018).
- [198] Osmanbegović, E., and Suljić, M. Data mining approach for predicting student performance. *Economic Review* 10, 1 (2012), 3–12.
- [199] Osório, F. S., and Vieira, R. Sistemas híbridos inteligentes. In *Encontro Nacional de Inteligência Artificial (ENIA)* (1999).
- [200] Paik, J. K., and Katsaggelos, A. K. Image restoration using a modified Hopfield network. *IEEE Transactions on Image Processing* 1, 1 (1992), 49–63.
- [201] PanduRanga Vital, T., Lakshmi, B., Swapna Rekha, H., and DhanaLakshmi, M. *Soft Computing in Data Analytics*. Springer, 2019, ch. Student Performance Analysis with Using Statistical and Cluster Studies, pp. 743–757.
- [202] Pantoja, C. T., Placencia, J. P. A., Castro, M. S., Mora, E. C., and Valenzuela, J. V. Diseño de un simulador de ficha clínica electrónica para estudiantes de Nutrición y Dietética. *Educación Médica* 19, 3 (2018), 238–245.
- [203] Park, J. H., and Choi, M. Generation of an optimal gait trajectory for biped robots using a genetic algorithm. *JSME International Journal Series C – Mechanical Systems, Machine Elements and Manufacturing* 47, 2 (2004), 715–721.
- [204] Paunonen, S. V. Big Five factors of personality and replicated predictions of behavior. *Journal of Personality and Social Psychology* 84, 2 (2003), 411.
- [205] Pease, A., Niles, I., and Li, J. The Suggested Upper Merged Ontology: A large ontology for the Semantic Web and its applications. In *AAAI Workshop on Ontologies and the Semantic Web* (2002).

- [206] Petri, G., Gresse von Wangenheim, C., and Borgatto, A. *Encyclopedia of Computer Graphics and Games*. Springer, 2018, ch. MEEGA+, Systematic Model to Evaluate Educational Games.
- [207] Pinninghoff, M. A., Ramírez, M., Arriagada, R. C., and Lagos, P. S. Collaborative group formation using genetic algorithms. In *International Work-Conference on the Interplay Between Natural and Artificial Computation (IWINAC)* (2015).
- [208] Pondel, J., and Pondel, M. The concept of project management platform using BI and Big Data technology. In *18th International Conference on Enterprise Information Systems* (2016).
- [209] Poy-Castro, R., Mendaña-Cuervo, C., and González, B. Diseño y evaluación de un juego serio para la formación de estudiantes universitarios en habilidades de trabajo en equipo. *Revista Ibérica de Sistemas e Tecnologías de Informação, SPE3* (2015), 71-83.
- [210] Provost, F., and Fawcett, T. *Data Science for Business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media, Inc., 2013.
- [211] Randell, B. *The origins of digital computers: selected papers*. Springer, 2013.
- [212] Reza Hejazi, S., and Saghafian, S. Flowshop-scheduling problems with makespan criterion: A review. *International Journal of Production Research* 43, 14 (2005), 2895-2929.
- [213] Rodríguez Hernández, A. A., and Avella Forero, F. La gestión de TI en la educación superior, caso incorporación de la simulación en la educación superior. In *3a Conferencia de Directores de Tecnología de Información (TICAL)* (2013).
- [214] Roiger, R. J. *Data mining: A tutorial-based primer*. Chapman and Hall/CRC, 2017.
- [215] Romero, C., Zafra, A., Gibaja, E., Luque, M., and Ventura, S. Predicción del rendimiento académico en las nuevas titulaciones de grado de la EPS de la Universidad de Córdoba. In *Jornadas de Enseñanza de la Informática* (2012).
- [216] Romney, J. Collaborative learning in a translation course. *Canadian Modern Language Review* 54, 1 (1997), 48-67.
- [217] Rosenblatt, F. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Spartan Books, 1962.
- [218] Rothmann, S., and Coetzer, E. P. The big five personality dimensions and job performance. *SA Journal of Industrial Psychology* 29, 1 (2003), 68-74.
- [219] Ruby, J., and David, K. Predicting the performance of students in higher education using data mining classification algorithms - A case study. *International Journal for Research in Applied Science & Engineering Technology* 2, XI (2014), 173-180.
- [220] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning internal representations by error propagation. Tech. rep., University of California San Diego, 1985.

- [221] Sadeghi, H., and Kardan, A. A. Toward effective group formation in computer-supported collaborative learning. *Interactive Learning Environments* 24, 3 (2016), 382–395.
- [222] Sahami, M. Learning non-linearly separable Boolean functions with linear threshold unit trees and Madaline-style networks. In *11th National Conference on Artificial Intelligence (AAAI)* (1993).
- [223] Salinas, J. Comunidades virtuales y aprendizaje digital. In *VI Congreso Internacional de Tecnología Educativa y NNNT aplicadas a la educación (EDUTEC)* (2003).
- [224] Samuel, A. L. Some studies in Machine Learning using the game of checkers. *IBM Journal of Research and Development* 3, 3 (1959), 210–229.
- [225] Sanin, C., and Szczerbicki, E. Towards the construction of decisional DNA: A set of experience knowledge structure Java class within an ontology system. *Cybernetics and Systems* 38, 8 (2007), 859–878.
- [226] Sanz, J., and García-Vera, M. P. Nuevos baremos para la adaptación española del Inventario de Personalidad NEO Revisado (NEO PI-R): Fiabilidad y datos normativos en voluntarios de la población general. *Clínica y Salud* 20, 2 (2009), 131–144.
- [227] Sareni, B., and Krahenbuhl, L. Fitness sharing and niching methods revisited. *IEEE Transactions on Evolutionary Computation* 2, 3 (1998), 97–106.
- [228] Serrano, A. J., Soria, E., and Martín, J. Redes neuronales artificiales. *Universidad de Valencia (Escuela Técnica Superior Ingeniería, Departamento de Ingeniería Electrónica): Valencia, España* (2009).
- [229] Shannon, E. Simulación de sistemas, diseño, desarrollo e implementación. *Editorial Trillas* (1988).
- [230] Sharan, S. The handbook of cooperative learning. *Praeger* (2012).
- [231] Slavin, R. E. Cooperative learning in elementary schools. *Education 3-13* 43, 1 (2015), 5–14.
- [232] Smullyan, R. M. *Gödel's incompleteness theorems*. Oxford University Press, 1992.
- [233] Son, L. H., and Fujita, H. Neural-fuzzy with representative sets for prediction of student performance. *Applied Intelligence* 49, 1 (2018), 1–16.
- [234] Srba, I., and Bielikova, M. Dynamic group formation as an approach to collaborative learning support. *IEEE Transactions on Learning Technologies* 8, 2 (2015), 173–186.
- [235] Strijbos, J.-W. *Handbook of human and social conditions in assessment*. Routledge, 2016, ch. Assessment of collaborative learning, pp. 302–318.
- [236] Studer, R., Benjamins, V. R., and Fensel, D. Knowledge engineering: principles and methods. *Data & Knowledge Engineering* 25, 1-2 (1998), 161–197.

- [237] Suganthi, L., Iniyan, S., and Samuel, A. A. Applications of fuzzy logic in renewable energy systems – A review. *Renewable and Sustainable Energy Reviews* 48 (2015), 585–607.
- [238] Sukstrienwong, A. A genetic algorithm approach for forming heterogeneous groups of students. *International Journal of Applied Engineering Research* 9, 3 (2014), 297–311.
- [239] Suárez-Figueroa, M. C., Gómez-Pérez, A., and Fernández-López, M. *Ontology engineering in a networked world*. Springer, 2012, ch. The NeOn methodology for ontology engineering, pp. 9–34.
- [240] Suárez-Figueroa, M. C., Gómez-Pérez, A., Motta, E., and Gangemi, A. *Ontology engineering in a networked world*. Springer, 2012, ch. Introduction to ontology engineering, pp. 1–6.
- [241] Takefuji, Y., and Szu, H. Design of parallel distributed Cauchy machines. In *IEEE International Joint Conference on Neural Networks (IJCNN)* (1989).
- [242] Tarifa, E. E. *Teoría de modelos y simulación*. Universidad de Jujuy, 2001.
- [243] Tejedor & Aguirre. Modelo de gestión del conocimiento de KPMG Consulting. <https://luisbringas.wordpress.com/2011/08/30/>, 1998.
- [244] Tello, A. L. Ontologías en la Web Semántica. In *I Jornadas de Ingeniería Web* (2001).
- [245] ter Horst, H. Combining RDF and part of OWL with rules: Semantics, decidability, complexity. *Lecture Notes in Computer Science* 3729 (2005), 668–684.
- [246] Tien, H.-W., Lin, Y.-S., Chang, Y.-C., and Chu, C.-P. A genetic algorithm-based multiple characteristics grouping strategy for collaborative learning. In *Workshops of 12th International Conference on Web-based Learning (ICWL)* (2015).
- [247] Tiwana, A. *The Knowledge Management Toolkit: practical techniques for building a knowledge management system*. Prentice Hall PTR, 2000.
- [248] Toegel, G., and Barsoux, J.-L. How to become a better leader. *MIT Sloan Management Review* 53, 3 (2012), 51–60.
- [249] Tohka, J., Krestyannikov, E., Dinov, I. D., Graham, A. M., Shattuck, D. W., Ruotsalainen, U., and Toga, A. W. Genetic algorithms for finite mixture model based voxel classification in neuroimaging. *IEEE Transactions on Medical Imaging* 26, 5 (2007), 696–711.
- [250] Trevor, H., Robert, T., and JH, F. *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009.
- [251] Tubella, I., and Requena, J. V. *Sociedad del conocimiento*. Editorial UOC, 2005.
- [252] Tünnermann Bernheim, C. El constructivismo y el aprendizaje de los estudiantes. *Universidades* 61, 48 (2011).
- [253] Turing, A. M. Computing machinery and intelligence. *Mind* 59, 236 (1950), 433–460.

- [254] Urbani, J., van Harmelen, F., Schlobach, S., and Bal, H. QueryPIE: Backward reasoning for OWL Horst over very large knowledge bases. *Lecture Notes in Computer Science* 7031 (2011), 730–745.
- [255] Valencia, D., Vizcaíno, A., Piattini, M., and Soto, J. P. Un juego serio para potenciar las habilidades de los estudiantes en el desarrollo global del software. In *XXII Jornadas de Enseñanza Universitaria de la Informática (JENUJ)* (2016).
- [256] van Gerven, M., and Bohte, S. *Artificial neural networks as models of neural information processing*. Frontiers Media SA, 2018.
- [257] Vandenbussche, P.-Y., Atemezing, G. A., Poveda-Villalón, M., and Vatan, B. Linked open vocabularies (LOV): A gateway to reusable semantic vocabularies on the web. *Semantic Web* 8, 3 (2017), 437–452.
- [258] Vygotsky, L. *Collaboration, Communication, and Critical Thinking: A STEM-inspired Path Across the Curriculum*. Rowman & Littlefield Publishers, 2019.
- [259] Wang, D.-Y., Lin, S. S., and Sun, C.-T. Diana: A computer-supported heterogeneous grouping system for teachers to conduct successful small learning groups. *Computers in Human Behavior* 23, 4 (2007), 1997–2010.
- [260] We Are Social Ltd. We Are Social - Digital Report 2018. <https://digitalreport.wearesocial.com/>, 2018.
- [261] White, B. A., and Elmasry, M. I. The digi-neocognitron: A digital neocognitron neural network model for VLSI. *IEEE Transactions on Neural Networks* 3, 1 (1992), 73–85.
- [262] Whitley, D. A genetic algorithm tutorial. *Statistics and computing* 4, 2 (1994), 65–85.
- [263] Widrow, B., and Hoff, M. E. Adaptive switching circuits. Tech. rep., Stanford University - Stanford Electronics Labs, 1960.
- [264] Wilt, J., Oehlberg, K., and Revelle, W. Anxiety in personality. *Personality and Individual Differences* 50, 7 (2011), 987–993.
- [265] Wong, D. M. L. Knowledge management catalyst for sustainable development. In *International Symposium on Information Technology (ITSim)* (2010), vol. 3, pp. 1444–1449.
- [266] Wooldridge, M., and Jennings, N. R. Agent theories, architectures, and languages: a survey. In *International Workshop on Agent Theories, Architectures, and Languages* (1994), Springer, pp. 1–39.
- [267] Yan, H., Zheng, J., Jiang, Y., Peng, C., and Xiao, S. Selecting critical clinical features for heart diseases diagnosis with a real-coded genetic algorithm. *Applied soft computing* 8, 2 (2008), 1105–1111.
- [268] Yang, F., and Li, F. W. Study on student performance estimation, student progress analysis, and student potential prediction based on data mining. *Computers & Education* 123 (2018), 97–108.

- [269] Yannibelli, V., and Amandi, A. A deterministic crowding evolutionary algorithm to form learning teams in a collaborative learning context. *Expert Systems with Applications* 39, 10 (2012), 8584–8592.
- [270] Yegnanarayana, B. *Artificial neural networks*. PHI Learning Pvt. Ltd., 2009.
- [271] Yeguas, E. *Un modelo de rendimiento de algoritmos evolutivos aplicados a la selección de la solución deseada*. PhD thesis, Universidad de Granada, 2009.
- [272] Yu, G., Wang, L.-G., Yan, G.-R., and He, Q.-Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* 31, 4 (2014), 608–609.
- [273] Zahedi, F., and Zahedi, Z. A review of neuro-fuzzy systems based on intelligent control. *Journal of Electrical and Electronic Engineering* 3, 2-1 (2015), 58–61.
- [274] Zheng, Y., Li, C., Liu, S., and Lu, W. An improved genetic approach for composing optimal collaborative learning groups. *Knowledge-Based Systems* 139 (2018), 214 – 225.
- [275] Zheng, Z. A dynamic group composition method to refine collaborative learning group formation. In *6th International Conference on Educational Data Mining (EDM)* (2013).