



UNIVERSIDAD NACIONAL DE COLOMBIA

Modelo de aprendizaje ontológico en el dominio de *e-recruitment* asociado a perfiles profesionales en TI y apoyado por redes sociales profesionales

Andrea Uribe Rendón

Universidad Nacional de Colombia
Facultad de Minas, Departamento de Ciencias de la Computación y la Decisión
Medellín, Colombia
2023

Modelo de aprendizaje ontológico en el dominio de *e-recruitment* asociado a perfiles profesionales en TI y apoyado por redes sociales profesionales

Andrea Uribe Rendón

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:
Magíster en Ingeniería Analítica

Director(a):
Ph.D., Jaime Alberto Guzmán Luna

Línea de Investigación:
Web Semántica
Grupo de Investigación:
Sistemas Inteligentes Web - SINTELWEB

Universidad Nacional de Colombia
Facultad de Minas, Departamento de Ciencias de la Computación y la Decisión
Medellí, Colombia
2023

(Dedicatoria)

A Dios, mi esposo y mi hija Emma que son el motor de mi vida.

La voluntad de vivir es la voluntad de poder.

Friedrich Nietzsche

Agradecimientos

Agradezco especialmente al Profesor Jaime Alberto Guzmán Luna, director de la tesis, quien orientó de manera oportuna y pertinente el desarrollo de este trabajo, inyectando toda su experiencia en el área de conocimiento y la investigación.

Resumen

El aprendizaje ontológico permite la creación automática o semiautomática de ontologías de cierto dominio, identificando clases, jerarquía de clases y restricciones a través de técnicas de recuperación de información, aprendizaje automático, aprendizaje profundo y procesamiento de lenguaje natural (PLN). En este trabajo se construye un modelo de aprendizaje ontológico que utiliza modelos, técnicas y algoritmos combinados (Transformadores, Incrustación de Grafos de Conocimiento (KGE) y Reglas de asociación) para elaborar un lenguaje común de competencias y ocupaciones en el área de *e-recruitment*, específicamente, profesionales en Tecnologías de la Información (TI). La fuente de extracción de información es una red social profesional seleccionada como caso de uso. Se define la ontología base a partir de la cual se inicia el aprendizaje, se diseña un modelo conceptual de aprendizaje ontológico que permite extraer clases, jerarquía de clases y restricciones para la ontología base. Seguido a esto, se implementa este modelo conceptual para obtener resultados de extensión de la ontología base, se evalúa la ontología aprendida a través de una *golden ontology* y se valida la ontología obtenida por medio de una aplicación para el dominio de *e-recruitment*.

Palabras clave: Aprendizaje ontológico, Transformadores, Incrustación de grafos de conocimiento, Reglas de asociación, *E-recruitment*, Tecnologías de la información, Redes profesionales.

Abstract

Ontological learning model in the e-recruitment domain associated with IT professional profiles and supported by professional social networks

Ontological learning allows the automatic or semi-automatic creation of ontologies in a domain, identifying classes, class hierarchy and restrictions through information retrieval techniques, machine learning and natural language processing (NLP). The aim of this document is built an ontological learning model that uses combined models, techniques and algorithms (*Transformers*, *Knowledge Graph Embedding*(KGE) and Association Rules) to develop a common language of skills and occupations in the e-recruitment area, specifically, Information Technology (IT) professionals. The source of information extraction is a selected professional social network as a use case. The base ontology from which learning begins is defined, a conceptual model of ontological learning is designed that allows information extract information for classes, hierarchy classes and the restrictions for the base ontology. Following this, this conceptual model is applied to obtain extension results of the base ontology, the learned ontology is evaluated through a *golden ontology* and the obtained ontology is validated through an application for the e-recruitment domain.

Keywords: Ontological learning, Transformers, Knowledge Graph Embedding, Association rules, E-recruitment, Information Technology (TI), Professional social networks

Contenido

Agradecimientos	vii
Resumen	ix
1. Capítulo 1: Introducción	1
1.1. Motivación	1
1.2. Objetivos de la tesis	3
1.2.1. Objetivo general	3
1.2.2. Objetivos específicos	3
1.3. Metodología	4
1.3.1. Etapa 1: Definición de la ontología base del modelo (enfoque teórico y aplicado)	4
1.3.2. Etapa 2: Definición del modelo conceptual de aprendizaje ontológico (enfoque teórico y aplicado)	4
1.3.3. Etapa 3: Implementación del modelo conceptual de aprendizaje ontológico (enfoque aplicado)	4
1.3.4. Etapa 4: Evaluación del modelo de aprendizaje ontológico (enfoque aplicado)	5
1.3.5. Etapa 5: Construcción de prototipo de validación para la ontología producto de la aplicación del modelo de aprendizaje de ontologías en el dominio de <i>e-recruitment</i> (enfoque aplicado)	5
1.4. Contribuciones	6
1.5. Distribución del documento	6
2. Capítulo 2: Marco teórico y Estado del arte	8
2.1. <i>E-recruitment</i>	8
2.2. Tecnologías de la información y la comunicación (TIC)	10
2.3. Web semántica	11
2.4. Ontologías	12
2.5. OWL	13
2.6. SPARQL	15
2.7. Procesamiento de Leguaje Natural(PLN)	15
2.8. Grafos de conocimiento (GC)	18
2.9. Aprendizaje ontológico	21

2.10. Análisis y conclusiones del capítulo	28
3. Capítulo 3: Ontología base	30
3.1. Adopción de ontologías en soluciones tecnológicas para <i>e-recruitment</i>	31
3.2. Ontología ESCO	33
3.2.1. ¿Por qué usar la ESCO?	34
3.2.2. ¿Cómo se puede utilizar la ESCO?	34
3.2.3. Clasificación de los pilares de la ESCO	35
3.2.4. Taxonomía de la ESCO	41
3.3. Ocupaciones y habilidades de los profesionales en TI	46
3.4. Proceso de diseño y creación de la ontología base	48
3.4.1. Definición del dominio y la aplicación de la ontología	51
3.4.2. Uso de estándares y clasificaciones existentes	52
3.4.3. Diseño de las categorías ontológicas	53
3.5. Análisis y conclusiones del capítulo	78
4. Capítulo 4: Diseño de un modelo conceptual de técnicas y algoritmos para el aprendizaje ontológico	80
4.1. Preparación de los datos	83
4.1.1. Descripción de los datos de la ontología base	83
4.1.2. Descripción de los datos del Web Scraping	83
4.1.3. Preprocesamiento de los datos	93
4.2. Análisis para aprendizaje de clases y jerarquía de clases	105
4.2.1. Análisis de técnicas de KGE	105
4.2.2. Propuesta de enriquecimiento de GC utilizado por KGE	108
4.3. Análisis para aprendizaje de restricciones	110
4.4. Modelo Conceptual general para el aprendizaje ontológico en clases, jerarquía de clases y restricciones	112
4.5. Análisis y conclusiones del capítulo	113
5. Capítulo 5: Implementación de portafolio de técnicas de aprendizaje de ontologías y obtención de nueva ontología	115
5.1. Preparación de los datos	116
5.2. Preprocesamiento	118
5.2.1. Implementación de frecuencia de frases	118
5.2.2. Resultado de análisis exploratorio de datos	120
5.2.3. Implementación de correspondencia exacta	121
5.2.4. Implementación de similaridad semántica con <i>Transformers</i> para definición de etiquetas alternativas	122
5.2.5. Implementación de <i>Transformers</i> para clasificación de competencias	126

5.3. Implementación de modelo KGE para aprendizaje de clases y jerarquía de clases	130
5.3.1. Implementación de tripletas para el grafo inicial basado en la ontología base	131
5.3.2. Implementación de <i>Transformers</i> para encontrar relaciones semánticas que enriquezcan el grafo inicial	132
5.4. Resultados de la implementación del modelo KGE diseñado para aprendizaje de clases y jerarquía de clases	134
5.5. Resultados de la implementación de reglas de asociación para aprendizaje restricciones	136
5.6. Análisis y conclusiones del capítulo	140
6. Capítulo 6: Evaluación de la ontología	141
6.1. Evaluación del criterio 1: Lenguaje	144
6.1.1. Extracción del dominio y codificación	145
6.1.2. Evaluación de las palabras en el lenguaje español	146
6.2. Evaluación del criterio 2: Vocabulario	146
6.3. Evaluación del criterio 3: Taxonomía	150
6.3.1. Exactitud de la extracción taxonómica	150
6.4. Evaluación del criterio 4: Aplicación	150
6.5. Análisis y conclusiones del capítulo	151
7. Capítulo 7: Aplicación de la ontología en el dominio de e-recruitment a través de casos de uso	152
7.1. Caso de uso 1	154
7.2. Caso de uso 2	160
7.3. Caso de uso 3	164
7.4. Análisis y conclusiones del capítulo	168
8. Capítulo 8: Conclusiones y recomendaciones	170
8.1. Conclusiones	170
8.2. Recomendaciones o trabajo futuro	173
A. Anexo: Definición de terminología de la ontología base	175
B. Anexo: Código en Python usando la librería Owlready2 para construcción de la ontología base	201
C. Anexo: Extracción Código en Python para web scraping de LinkedIn	205
Bibliografía	207

Lista de Figuras

1-1. Diagrama general de la propuesta general. Elaboración propia	5
1-2. Diagrama del desarrollo ordenado por etapas. Elaboración propia	6
2-1. Arquitectura de un <i>Transformer</i> . Tomado de [Vaswani et al., 2017]	18
2-2. Arquitectura de KGE. Tomado de [Costabello et al., 2023]	20
3-1. Metodología de desarrollo objetivo 1. Elaboración propia	30
3-2. Representación del panorama actual de la ontología ESCO, tomado de [European Commission website, 2022]	32
3-3. Evolución de la ESCO, tomado de [European Commission website, 2022]	33
3-4. Jerarquías de ocupación en la ESCO usando la ISCO-08, tomado de [Chala, 2018]	36
3-5. Modelo de la ontología ESCO, tomado de [European Commission, 2022]	42
3-6. Relaciones importantes de la ontología ESCO, tomado de [European Commission, 2022]	42
3-7. Ocupaciones relacionadas con TI categorizadas con el nivel de cada categoría, tomado de [OECD, 2004]	48
3-8. Clasificación de las Ocupaciones relacionadas con TI, tomado de [Arnal et al., 2001]	49
3-9. Modelo de las competencias en TI, tomado de [Lopez-Bassols, 2002]	50
3-10. Clasificación de competencias adoptado por el Ministerio de Educación en Colombia, tomado de https://www.mineducacion.gov.co/1621/w3-printer-299611.html	54
3-11.a) modelo de pilares de la ontología OPTI vs b) modelo de pilares de la ontología OPTI. Elaboración propia	55
3-12. Definición de superclases de la ontología base. Elaboración propia	63
3-13. Relación taxonómica entre clases	63
3-14. Relaciones entre conceptos	70
3-15. Relaciones entre conceptos y tipos de datos	71
3-16. IRI de la ontología base, extraído de Protégé	74
3-17. Clases y jerarquía de clase de la ontología base, extraído de Protégé	75
3-18. Propiedades de objeto de la ontología base, extraído de Protégé	75
3-19. Propiedades de los datos de la ontología base, extraído de Protégé	76
3-20. Instancias definidas para la ontología base	76

3-21.	Instancias de clase de la ontología base	77
3-22.	Representación gráfica de la ontología base OPTI	77
4-1.	Representación gráfica del objetivo del capítulo 4	81
4-2.	Metodología de desarrollo objetivo 2	81
4-3.	Definición de tareas de aprendizaje con base en la torta de capas de aprendizaje de una ontología, adaptado de [Asim et al., 2018b]	82
4-4.	Campos generales de una oferta laboral online	87
4-5.	Campos requeridos para extraer de la red social profesional y utilizarlos para el aprendizaje de la ontología final	88
4-6.	Arquitectura del <i>web scraping</i> . Tomado de [Kasereka, 2020]	91
4-7.	Visualización de oferta laboral filtrada por TI en  LinkedIn	91
4-8.	Arquitectura de <i>web scraping</i> para este modelo	92
4-9.	Campos que se pueden extraer con <i>linkedin-jobs-scrapers</i> de una oferta laboral en LinkedIn	92
4-10.	Etapas 1 del modelo conceptual de aprendizaje ontológico	93
4-11.	Nube de palabras de los datos estructurados de la ontología base para ocupaciones	95
4-12.	Nube de palabras de los datos estructurados de la ontología base para competencias	95
4-13.	Ejemplo de extracción de competencias del campo <i>description</i> del <i>web scraping</i>	99
4-14.	Nube de palabras de los datos estructurados del <i>web scraping</i> para ocupaciones	99
4-15.	Nube de palabras de los datos estructurados del <i>web scraping</i> para competencias	99
4-16.	Arquitectura SBERT para similaridad semántica. Tomado de [Reimers and Gurevych, 2019]	103
4-17.	Arquitectura SBERT para clasificación de competencias. Elaboración propia	104
4-18.	Etapas 2 del modelo conceptual de aprendizaje ontológico	105
4-19.	Etapas 3 del modelo conceptual de aprendizaje ontológico	108
4-20.	Etapas 4 del modelo conceptual de aprendizaje ontológico	110
4-21.	Propuesta de aplicación del algoritmo apriori para aprender las restricciones de dominio y rango sobre una propiedad	111
4-22.	Modelo conceptual de aprendizaje ontológico para clases, jerarquía de clases y restricciones con aplicación al dominio de <i>e-recruitment</i> para profesionales asociados a TI. Elaboración propia	112
5-1.	Representación gráfica del objetivo del capítulo 5	115
5-2.	Metodología de desarrollo objetivo 3	116
5-3.	Diagrama resultado de análisis exploratorio de los datos	120
5-4.	Representación gráfica en 2D para algunos vectores con similaridad semántica para las ocupaciones	123

5-5.	Representación gráfica en 2D para algunos vectores con similaridad semántica para las competencias	124
5-6.	Representación vectorial de las ocupaciones	126
5-7.	Representación vectorial de las competencias	126
5-8.	Superclase competencias antes y después de la clasificación	128
5-9.	Distribución espacial de los vectores de las competencias de la ontología base clasificados	130
5-10.	Distribución espacial de los vectores de las competencias de la del <i>web scraping</i> clasificados	130
5-11.	Representación parcial del grafo de conocimiento asociado a las tripletas extraídas de la ontología base	132
5-12.	Base de inferencia de nuevas clases y jerarquía de clases	134
6-1.	Representación gráfica del objetivo del capítulo 6	141
6-2.	Metodología de desarrollo objetivo 5	142
6-3.	Criterios de evaluación ontológica, tomado de [Ramos et al., 2009]	143
6-4.	Descripción general de los enfoques de evaluación de ontologías, tomado de [Asim et al., 2018a]	143
6-5.	Validadores de ontologías en el dominio de IOT, tomado de [Coggle, 2021]	144
6-6.	Resultado del validador OWL	145
7-1.	Matriz de confusión de la búsqueda de perfiles TI en la ontología	163

Lista de Tablas

2-1. Herramientas profesionales para <i>e-recruitment</i> y sus características	10
2-2. Comparación de características en trabajos de aprendizaje de ontologías. Elaboración propia	27
3-1. Revisión de ontologías adoptadas en soluciones tecnológicas en el dominio de <i>e-recruitment</i> . Elaboración propia	31
3-2. Clasificación de ocupaciones según la ISCO-08	38
3-3. Clasificación de competencias según la ESCO	40
3-4. Categorías de competencias TIC. Elaboración propia	47
3-5. Habilidades/competencias de TI buscadas por las empresas. Elaboración propia	48
3-6. Definición de propiedades del objeto y sus características. F: Funcional, S: Simétrica, FI: Funcional Inversa, T: Transitiva, AS: Asimétrica, R: Reflexiva, IR: Irreflexiva, IV: Inversa	71
3-7. Definición de propiedades de los datos y sus características. F: Funcional	72
4-1. Campos 1 de la ontología base inicial	84
4-2. Campos 2 de la ontología base inicial	85
4-3. Campos 3 de la ontología base inicial	86
4-4. Campos extraídos con el web scraping	94
4-5. Datos estructurados de la ontología base	96
4-6. Extracto de resultado de análisis de frecuencia de sentencias de las ocupaciones de la extracción del <i>web scraping</i>	98
4-7. Datos estructurados del web scraping	101
4-8. Hiperparámetros modelos KGE	107
4-9. Comparación de las métricas de los modelos KGE	107
4-10. Comparación del modelo KGE con y sin el GC enriquecido con relaciones semánticas	109
4-11. Análisis de la literatura vs la propuesta	109
5-1. Títulos y frecuencia	120
5-2. Resultado de correspondencia exacta	121
5-3. Similaridad semántica entre ocupaciones existentes y posibles nuevas	122
5-4. Similaridad entre competencias existentes y posibles nuevas	123
5-5. Nuevas etiquetas alternativas	125

5-6. Clasificación de competencias No clasificadas de la ontología base	127
5-7. Clasificación de las posibles competencias del <i>scraping</i>	127
5-8. Extracción de clasificación de competencias No clasificadas de la ontología base	128
5-9. Extracción de clasificación de competencias del <i>scraping</i>	129
5-10. Relación semántica entre <i>items</i>	133
5-11. Resultados de aprendizaje de clases y jerarquía de clases	135
5-12. Ejemplo de reglas de propiedad que relaciona una instancia con dominio Ocupación y rango Competencia	138
5-13. Ejemplo de reglas propiedad que relaciona una instancia con dominio Competencia y rango Competencia	139
6-1. Porción o extracción de resultado de las ejecuciones de <i>LogMap</i>	148
6-2. Cálculo de medidas de calidad ontología aprendida vs ESCO (ocupaciones) .	149
6-3. Cálculo de medidas de calidad ontología aprendida vs ESCO (competencias)	149
7-1. Porción o extracción del resultado de la consulta	157
7-2. Porción o extracción de hojas de vida de personas que pertenecen a perfiles relacionados con TI y personas con perfiles relacionados a Cocina	160
7-3. Resultado de búsqueda de la ontología para personas que pertenecen a perfiles relacionados con TI	162
7-4. Porción o extracción de resultado de búsqueda de competencias de las personas que pertenecen a perfiles relacionados con TI	168

1. Capítulo 1: Introducción

1.1. Motivación

El *e-recruitment* está aportando beneficios atractivos para negocios y empresas en Colombia y el mundo y, por lo tanto, representa una oportunidad importante de investigación en el país. El proceso de contratación a través de internet simplifica los procesos de selección, reduce los costos asociados a ello, aumenta la capacidad predictiva de mejores candidatos para una vacante y disminuye costos asociados a errores de selección y rotación de personal muy arraigada en el área de TI. Sin embargo, las exigencias, competencias y habilidades de los profesionales están siendo modificadas en el contexto también cambiante de la industria 4.0.

En el área *e-recruitment* basada en Datos Abiertos Enlazados (LOD), es de gran importancia la calidad de la base de conocimiento usada, es decir, que los términos, conceptos, relaciones y nuevas restricciones que se definen a través del modelamiento ontológico sean pertinentes, y precisos para responder a las necesidades en un dominio específico. Ahora bien, el modelo ontológico utilizado para ser usado en algún sistema de recomendación de profesionales en el área profesional TIC no debe ser estático, sino que debe ir aprendiendo de forma inteligentemente estructuras nuevas según las exigencias del empleador, las nuevas conexiones en el mundo, la dirección tecnológica industrial 4.0 y los intereses del profesional. Las redes sociales y profesionales están actualmente realizando grandes aportes en la identificación de nuevas habilidades y experticias de los profesionales.

De acuerdo con el panorama anterior, surgen las siguientes preguntas de investigación ¿Es posible mantener actualizados los modelos ontológicos usados en el marco de *e-recruitment* para profesionales en TI y así, diseñar un modelo de aprendizaje ontológico basado en la extracción de conocimiento de redes sociales y profesionales para que aprenda o actualice nuevos términos, clases, jerarquías de clases y restricciones que pueden modificar el dominio, a través de la aplicación de diferentes técnicas de aprendizaje combinadas?, ¿La combinación de técnicas de aprendizaje de ontologías puede mejorar los resultados de evaluación de la ontología aprendida?, ¿Las redes sociales profesionales pueden servir de base para el aprendizaje de la ontología?.

El *e-recruitment* se fortalece con el uso del concepto de la web semántica [BERNERS-LEE

et al., 2001] y los datos abiertos enlazados (LOD), que a través de su columna vertebral: las ontologías, permiten modelar la información dentro del área particular de estudio y las relaciones alrededor de ella [Harzallah et al., 2002]. Así pues, permite que el contratista y el solicitante de empleo tengan un entendimiento común de las competencias y las tareas transversales de una profesión y compartan el mismo vocabulario para denotar dichas nociones. Existen diferentes modelos ontológicos que representan en alguna medida las bases necesarias para representar las relaciones en un proceso de contratación online [García-Sánchez et al., 2006], [Tong, 2009], [Varen Caballero et al., 2019], [Radevski and Trichet, 2006] y [Jun, 2011], sin embargo, es claro que con las exigencias de la industria 4.0 la relación de habilidades requeridas y deseadas en profesionales del área de tecnología se van modificando rápidamente.

Existen diferentes modelos para generar aprendizaje ontológico, dentro de dichos modelos se puede evidenciar la extracción de términos, clases, relaciones y axiomas desde algunas fuentes de información definidas, desde texto plano hasta las redes sociales. Abordar el tema de aprendizaje ontológico desde el punto de vista de los axiomas y restricciones es complejo pero algunos autores lo han logrado a través de la aplicación de técnicas basadas en lógica. En términos generales un modelo de aprendizaje ontológico debería contener procesos de recuperación de información, extracción de conceptos, refinamiento de información, implementación y evaluación dentro del proceso.

Por su parte, las redes sociales son plataformas de comunicación on-line para personas que están conectadas de alguna manera y se están convirtiendo cada vez más en una fuente clave para determinar el comportamiento de los individuos en función de sus conexiones, contribuciones y reacciones. Aunque las redes sociales no están integradas en los sistemas de *e-recruitment* [Soni and Swaminarayan, 2017], están siendo usadas por empleadores como herramienta principal para buscar sus posibles empleados, debido a que proporcionan información de respaldo de habilidades de una persona y se generan recomendaciones y evaluaciones de usuarios que colaboraron. La analítica predictiva y las técnicas de aprendizaje automático pueden aprovechar esa información disponible como base para robustecer y mejorar la base de conocimiento del área de profesionales de TI para que pueda ser utilizada mediante técnicas de software en aplicaciones reales de reclutamiento on-line. Es importante validar que la aplicación de la combinación de varias técnicas de aprendizaje ontológico, junto con la extracción de información de varios campos de los currículos obtenidos de una red social logra mejorar la evaluación de la ontología obtenida y permite recuperar términos, conceptos, relaciones y axiomas. Finalmente, la evaluación con técnicas automáticas como la Golden standard han permitido la objetividad y agilidad para la obtención de resultados.

En conclusión y teniendo en cuenta el potencial e impacto de los procesos de contratación en línea hoy en día, la fuerza de las redes profesionales, la necesidad de gestionar adecuadamente el conocimiento de la evolución de perfiles y habilidades de profesionales de las tecnologías

de la información (TI) y la exigencia creciente de este tipo de profesionales en las empresas, se requiere crear un modelo de aprendizaje automático de ontologías que utilice técnicas y algoritmos combinados para mantener actualizado este dominio particular de profesionales y contribuir con su aplicación posterior en *e-recruitment*.

1.2. Objetivos de la tesis

1.2.1. Objetivo general

Proponer un modelo de aprendizaje ontológico en el dominio de *e-recruitment* asociado a perfiles tecnológicos profesionales en TI y apoyado redes sociales profesionales.

1.2.2. Objetivos específicos

1. Definir los conceptos básicos o la estructura ontológica para el dominio de *e-recruitment* asociado a perfiles profesionales en TI que sean relevantes para la investigación a partir de las fuentes de información disponibles, con el fin de definir la ontología base del modelo.
2. Diseñar un modelo conceptual de aprendizaje ontológico con técnicas combinadas de aprendizaje que permita extraer desde los diferentes campos de la red social profesional los términos, clases, jerarquías de clases y restricciones con el fin de extender y refinar la ontología base.
3. Implementar el modelo conceptual de aprendizaje ontológico con técnicas combinadas de aprendizaje que permita extraer desde los diferentes campos de la red social profesional los términos, clases, jerarquías de clases y restricciones, con el fin validar la utilidad del modelo de aprendizaje ontológico propuesto.
4. Verificar la utilidad del modelo de aprendizaje ontológico implementado por medio de la construcción de un prototipo para el dominio de *e-recruitment*.
5. Evaluar los resultados de la implementación del modelo de aprendizaje ontológico por medio de técnicas o métricas de evaluación de aprendizaje de ontologías, con el fin verificar los cambios sobre la ontología base.

1.3. Metodología

La metodología propuesta para la investigación está basada en el seguimiento de 5 etapas de forma ordenada, las cuales representan el hilo conductor desde la definición de la ontología base del modelo, hasta la evaluación final del modelo. La metodología sigue dos enfoques: el teórico y el aplicado. En el transcurso de los capítulos se verán aplicados estos enfoques, los cuales aportan soporte, propuesta y desarrollo de los mismos.

1.3.1. Etapa 1: Definición de la ontología base del modelo (enfoque teórico y aplicado)

- Actividad 1: Revisión de fuentes de información disponibles para definir las ontologías más usadas en *e-recruitment*.
- Actividad 2: Revisión de la estructura de la mejor ontología disponible en el sector de *e-recruitment* y recursos humanos.
- Actividad 3: Definición la ontología base para el modelo de aprendizaje de ontologías.

1.3.2. Etapa 2: Definición del modelo conceptual de aprendizaje ontológico (enfoque teórico y aplicado)

- Actividad 1: Revisión bibliográfica de las técnicas usadas para el aprendizaje de las capas ontológicas.
- Actividad 2: Definición de la fuente de extracción de información y los métodos de extracción de información desde las redes sociales profesionales.
- Actividad 3: Diseño del portafolio de técnicas de aprendizaje.

1.3.3. Etapa 3: Implementación del modelo conceptual de aprendizaje ontológico (enfoque aplicado)

- Actividad 1: Implementación de los métodos de extracción de campos de información desde las redes sociales profesionales.
- Actividad 2: Implementación de los métodos y técnicas de aprendizaje ontológico a utilizar para aportar a las nuevas clases, jerarquía de clases y restricciones en la ontología.

1.3.4. Etapa 4: Evaluación del modelo de aprendizaje ontológico (enfoque aplicado)

- Actividad 1: Implementación de las técnicas y métricas definidas de evaluación automática de aprendizaje ontológico.
- Actividad 2: Análisis de los resultados de la evaluación de la ontología final aprendida.

1.3.5. Etapa 5: Construcción de prototipo de validación para la ontología producto de la aplicación del modelo de aprendizaje de ontologías en el dominio de e-recruitment (enfoque aplicado)

- Actividad 1: Definición del caso de uso asociado al proceso de *e-recruitment*.
- Actividad 2: Implementación de las consultas a la ontología aprendida, asociadas al caso de uso.
- Actividad 3: Presentación de resultados del caso de uso en *e-recruitment* empleando la ontología final aprendida.

Luego de definir los objetivos y la metodología de este trabajo de investigación, en la Figura 1-1 se muestra un diagrama general de la propuesta y en la Figura 1-2 se especifica el diagrama del desarrollo ordenado por etapas de este proyecto de tesis.

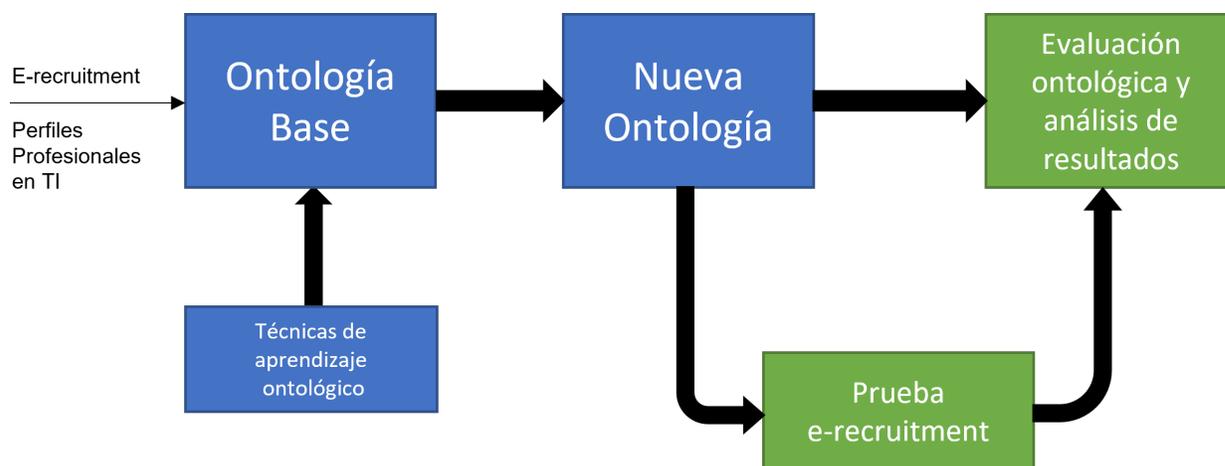


Figura 1-1.: Diagrama general de la propuesta general. Elaboración propia

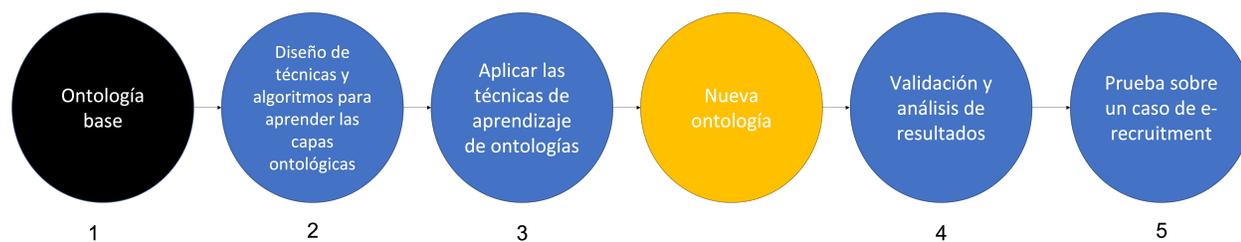


Figura 1-2.: Diagrama del desarrollo ordenado por etapas. Elaboración propia

1.4. Contribuciones

- Esta tesis representa una iniciativa importante, que se espera pueda contribuir a la actualización y descubrimiento de nuevas ocupaciones y competencias del nicho de los profesionales de las tecnologías de la información y la comunicación. Categoría de profesionales que, según las estadísticas, representan un sector con alta rotación de personal debido a que las empresas contratantes aún desconocen la evolución y transformación de las ocupaciones y competencias para estas vacantes.
- Los resultados de este trabajo permite a los desarrolladores ontológicos robustecer a través de la de información que ofrece las redes sociales profesionales y el aprendizaje ontológico, las ontologías diseñadas para el área de *e-recruitment*.
- El modelo consolidado de aprendizaje ontológico de este trabajo proporciona una herramienta que servirá como habilitador para el desarrollo de ontologías especializadas en cualquier profesión a través de la plataforma LinkedIn. Dichas ontologías tendrán un nivel de calidad adecuado para usarlas en procesos de contratación en línea.
- Este trabajo servirá como base para la exploración de nuevas técnicas de aprendizaje de máquina, procesamiento de lenguaje natural, minería de datos y demás que puedan ser combinadas de manera asertiva para aprender las diferentes capas de una ontología destinada a *e-recruitment* y basada en dos grandes clases: ocupaciones y competencias.

1.5. Distribución del documento

Este documento de tesis está organizado de la siguiente manera:

En el capítulo 2 se describe de forma detallada el estado del arte y el marco teórico de la investigación, proporcionando una visión general de la literatura relacionada con las ontologías y el aprendizaje de ontologías, el *e-recruitment* o servicios de contratación en línea, las tecnologías de la información y de la comunicación (TIC), la web semántica y los lenguajes de programación y consulta de ontologías. Finalmente, se expone una tabla resumen

para comparar las características más relevantes encontradas en trabajos de aprendizaje de ontologías a través de la revisión sistemática de la literatura.

En el capítulo 3 se presenta el estudio bibliográfico acerca de las ontologías más usadas para la generación de servicios web para recursos humanos, en particular el reclutamiento (*e-recruitment*), se estudia de forma detallada la ontología más relevante con el fin de conocer los puntos importantes, se explica el papel del ente regulador internacional encontrado para la clasificación de ocupaciones hasta llegar a la definición de las ocupaciones específicas para las Tecnologías de la Información (TI). Finalmente, se condensa todo este estudio en la propuesta de la ontología base en el dominio de *e-recruitment* asociado a perfiles profesionales en TI, definiendo detalladamente sus partes y su construcción.

En el capítulo 4 se detalla la propuesta del modelo teórico o portafolio de aprendizaje de ontologías a través de una búsqueda conceptual de las técnicas que se usan para aprender clases, jerarquía de clases y restricciones. El modelo o portafolio de técnicas de aprendizaje ontológico es dedicado a cada capa ontológica. Se define la fuente de extracción y se estructuran los datos, esta extracción se realiza a través de un wrapper especificado.

En el capítulo 5 se implementa el portafolio de aprendizaje de ontologías conceptual diseñado en el capítulo anterior. Se muestra de manera detallada los resultados de aplicación de las técnicas, modelos y algoritmos ejecutados para extraer la información de la red profesional seleccionada y para realizar el aprendizaje de clases, jerarquía de clases y restricciones de la ontología. Se muestra la aplicación y los resultados obtenidos del *web scraping* y el aprendizaje ontológico.

En el capítulo 6 se evalúa de forma cuantitativa la ontología aprendida teniendo en cuenta los criterios básicos para la evaluación ontológica: taxonomía, lenguaje, vocabulario y aplicación. Se extraen las métricas para cada uno de los criterios y se explica de forma detallada los resultados de la evaluación.

En el capítulo 7 se determina un caso de uso de la ontología aprendida en el dominio del *e-recruitment* con el fin de validar la calidad y utilidad de dicha ontología en cuanto a las respuestas de los requerimientos en la aplicación desarrollada.

En el capítulo 8 se presenta las conclusiones generales y la propuesta de trabajo futuro.

2. Capítulo 2: Marco teórico y Estado del arte

A continuación, se suministra el marco teórico y se revisa el estado del arte de las áreas contenidas en esta investigación. En principio se presentarán los temas básicos asociados a las ontologías, el dominio de *e-recruitment*, las TIC y el aprendizaje ontológico. Luego, se analizarán varias aproximaciones para la generación de aprendizaje de ontologías que será el objetivo principal de esta tesis.

2.1. E-recruitment

El proceso de contratación conocido en la literatura como *recruitment*, la gestión de conocimiento está orientada a la selección adecuada de cierto capital humano para que ocupe un puesto de trabajo dentro de una organización, lo anterior, no solamente influye en atraer la atención de potenciales candidatos, si no en retener y mantener el interés del individuo en la oferta realizada [Chow and Chapman, 2013]. Todas las organizaciones, indistintamente de su función, se enfrentan al desafío de atraer y retener empleados cuyas habilidades y competencias estén enmarcadas en las necesidades de la organización. El reclutamiento o contratación llevado a cabo por recursos humanos debe ser flexible, diferenciar individuos entre sus valores, aptitudes y capacidades y seguir nuevas tendencias para hacerlo correctamente [Fajčíková et al., 2018].

Por su parte, Wong et. Al [Wong et al., 2017] afirma que es necesario ir remodelando los procesos para seleccionar y retener el personal talentoso dentro de una organización, debido a que constantemente está cambiando la forma en que conciben las nuevas generaciones los trabajos y las exigencias de habilidades y experticias requeridas por los empleadores. Así, las compañías se están sumando a este reto con soluciones basadas en software como servicio y computación en la nube y el análisis de datos e inteligencia artificial para reclutar y administrar talentos [CB Insights Research, 2020], esto posibilita la introducción de un nuevo concepto que es el *e-recruitment* o contratación usando como medio el internet, que según [Lievens et al., 2007] mejora la marca del empleador atrayendo y reteniendo solicitantes potenciales.

A continuación se mencionan algunos trabajos adicionales encontrados en la literatura que acercan al lector a la comprensión del término *e-recruitment* en contexto de las aplicaciones ontológicas. Algunos trabajos ya fueron mencionados en la motivación de este trabajo y otros serán descritos más adelante en este capítulo en el literal sobre **aprendizaje ontológico**.

[Ilieva, 2013] en su artículo *E-recruitment public services ontological model* argumenta que la aplicación de ontologías para el reclutamiento electrónico se está convirtiendo en una tarea importante para hacer coincidir semánticamente las ofertas de trabajo y los solicitantes en una tecnología de web semántica que utiliza ontologías. Este documento tiene como objetivo optimizar y mejorar los procesos de contratación electrónica en el ámbito de los servicios de contratación de personal de TI, y especialmente los procesos de contratación electrónica que utilizan plataformas web como medio para buscar candidatos. Se propone un modelo como respuesta a una serie de categorías de requisitos, que van desde la semántica hasta el rendimiento del software en tiempo de ejecución. Se desarrolla una clasificación de los tipos de información de contratación electrónica y luego se examina un esquema de ontología jerárquica, que incluye la ontología de dominio y el conjunto de ontologías de aplicaciones y tareas.

[ENACHESCU, 2016] desarrolló un prototipo de plataforma de contratación electrónica que utiliza tecnologías de web semántica que se centra en el uso de ontologías en el desarrollo de un sistema de recomendación de empleo, que ayuda a hacer coincidir automáticamente las ofertas de trabajo con los perfiles de los candidatos y a la inversa. El diseño de la ontología IT *e-recruitment* recopila una lista de todas las características que una plataforma de este tipo debe ofrecer, tanto para el buscador de empleo como para el reclutador. También propone una arquitectura basada en Java para implementar la plataforma *e-recruitment*. Para convertir la entrada de los usuarios en una descripción RDF, se emplean las API RDF2Go y RDFBeans. El almacenamiento y la recuperación de datos utilizan Jena Framework, que proporciona interfaces dedicadas para acceder al servidor Fuseki2 a través de HTTP.

[Javed et al., 2019] implementaron la gestión del conocimiento en el proceso de desarrollo de una plataforma de contratación electrónica utilizando un enfoque basado en ontología. El enfoque facilita a los empleadores la contratación y el reclutamiento según los requisitos deseados de la organización con una utilización eficaz de los datos de los MOOC (*Masive Open Online Courses*). La ontología desarrollada puede recomendar candidatos al reclutador electrónico en función de las funciones requeridas del puesto de trabajo y, de manera similar, los estudiantes o candidatos elegibles que hayan completado el MOOC podrán ver los puestos vacantes en función de su campo de interés. El modelo propuesto también incorpora los datos de los estudiantes, incluida la información requerida sobre la matrícula, las calificaciones y las habilidades obtenidas en varios MOOC.

Por su parte, se mencionan algunas herramientas profesionales utilizadas en el ámbito de la gestión de recursos humanos y el reclutamiento de personal en línea. Cada una de ellas se enfoca en varias características para facilitar diferentes aspectos del proceso de contratación y gestión de candidatos.

herramienta	enlace	características
Manatal	https://www.manatal.com/	Sugiere los mejores candidatos para un trabajo determinado mientras automatiza tareas redundantes, califica los perfiles de los candidatos según los requisitos laborales para facilitar su proceso de selección, enriquece los perfiles de los candidatos con LinkedIn y otros datos de redes sociales para obtener mejores recomendaciones
Workable	https://www.workable.com/	Recomendaciones de candidatos impulsadas por IA, integración con LinkedIn, evaluación de candidatos con entrevistas de video
Breezy	https://breezy.hr/	Preselección sin contacto estableciendo coincidencia entre el contacto y el perfil, puntuación de coincidencia de candidatos impulsada por IA
Lever	https://www.lever.co/	Contratación de alto volumen, evaluación rápida de candidatos en una sola pantalla, realimentación automática del proceso de contratación
Hired	https://hired.com/	Utiliza algoritmos de IA para conectar a empresas con candidatos técnicos, ayuda a las empresas a encontrar candidatos que se ajusten a sus necesidades específicas

Tabla 2-1.: Herramientas profesionales para *e-recruitment* y sus características

2.2. Tecnologías de la información y la comunicación (TIC)

La creciente importancia y el alcance global de la tecnología informática en todos los aspectos de la sociedad actual fue la base para el surgimiento de la disciplina de la tecnología de

la información. La disciplina de tecnología de la información, la más joven entre las cinco disciplinas informáticas actuales, nació en 2008 con el lanzamiento del primer informe ACM / IEEE-CS sobre las directrices curriculares para los programas académicos en TI. Desde entonces, sus disciplinas hermanas ya han sido actualizadas y revisadas. Los sistemas de información actualizaron su plan de estudios en 2010. La informática completó sus revisiones en 2013, seguida de la ingeniería de software en 2014. La ingeniería informática publicó sus directrices curriculares en 2016 [on Information Technology Curricula, 2017].

Según iT2017 [on Information Technology Curricula, 2017] la tecnología de la información es definida como “el estudio de enfoques sistémicos para seleccionar, desarrollar, aplicar, integrar y administrar tecnologías informáticas seguras que permitan a los usuarios lograr sus objetivos personales, organizativos y sociales”.

En todo el mundo, el término “tecnología de la información” generalmente se refiere a todos los aspectos de la informática y su integración en todos los aspectos de la sociedad actual y la economía de la plataforma digital. Las organizaciones de todo tipo dependen de la tecnología de la información y los sistemas informáticos que deben funcionar de manera adecuada y eficiente, ser seguros y ajustarse a los objetivos de la organización y las necesidades del cliente. Los profesionales de TI seleccionan productos y servicios informáticos, los integran para mejorar entornos compatibles y desarrollar, adaptar y administrar tecnologías informáticas para cumplir con las metas y los objetivos comerciales de la organización. Las innovaciones TIC son objeto de estudio bajo capacidades de aplicaciones móviles, plataformas sociales, experiencia de usuario, internet de las cosas, ciberseguridad, automatización, dispositivos inteligentes, entre otros. Estas innovaciones enmarcan las preguntas que plantean los profesionales e investigadores de TIC e informan los métodos y prácticas mediante los cuales se resuelven los problemas complejos de TIC y se hacen nuevos descubrimientos [OECD, 2010].

2.3. Web semántica

El Consorcio de la *World Wide Web (W3C)* ha introducido estándares de la Web Semántica, es decir, ha dispuesto una serie de tecnologías para soportar la red de datos vinculados, permitiendo a las computadoras realizar tareas más útiles dentro de una interacción de red confiable. Las tecnologías web semánticas (SWT) admiten la gestión de datos, la creación de ontologías y las reglas estatales para manejar los datos [World Wide Web Consortium, “Semantic Web - W3C,” W3.org, 2012]. Algunas características que propone la web semántica y desarrollan las SWT son los datos vinculados, los datos y relaciones entre los datos son disponibles en la web en un formato estándar; las ontologías, que define conceptos y relaciones a través de clase y propiedades; las consultas y las inferencias, como proceso de razonamiento de acuerdo con una base de reglas.

2.4. Ontologías

Al inicio de los años noventa, el concepto de ontología comenzó a tomar fuerza dentro de los temas de interés para algunos grupos de investigación asociados a las líneas de inteligencia artificial, ingeniería del conocimiento, procesamiento de lenguaje natural y representación de conocimiento. Actualmente, dicho concepto se ha extendido a otras áreas como la gestión de conocimiento e integración inteligente de información desde orígenes heterogéneos [López Bonilla et al., 2012]. Una ontología es una característica propuesta por la web semántica con el fin de que la estructuración de los contenidos y las relaciones de la web puedan ser interpretadas tanto por humanos como entidades de software [World Wide Web Consortium, “Semantic Web - W3C,” W3.org, 2012].

Desde el punto de vista informático el término ontología se ha definida de varias formas, pero en general, una ontología se considera como un vocabulario relativo a cierto dominio el cual define entidades, clases, propiedades, funciones y relaciones de este. Dentro del estado del arte se encuentran las siguientes definiciones formales del concepto de la ontología:

- Según [Neches et al., 1991] “Una ontología define las condiciones básicas y relaciones que comprenden el vocabulario de un área del tema, así como las reglas para combinar condiciones y las relaciones para definir extensiones del vocabulario”. Dentro de esta definición se puede abstraer que una ontología incluye todas las condiciones explícitas establecidas y, además, todo el conocimiento implícito que puede inferirse. Esta fue la primera definición para el término ontología.
- Dentro de las definiciones más populares en el mundo científico, también se encuentra la de Gruber [Gruber, 1993] “Una ontología es una especificación explícita de una conceptualización.” En esta definición en particular se indica que el conocimiento, condiciones y restricciones deben ser generados de forma explícita.
- Unos años después la definición de Gruber fue refinada por Borst [Borst, 1999], el que indica que “Una ontología es una especificación formal de una conceptualización compartida”. El concepto ‘formal’ abarca la necesidad de tener ontologías disponibles que sean comprensibles por aplicaciones de software y el concepto ‘compartida’ hace alusión a que el conocimiento representado por la ontología debe ser un vocabulario común, consensado y público.
- Otra definición importante de este concepto la genera Guarino [Guarino, 1997], que establece que “La ontología describe una cierta realidad con un vocabulario específico, usando un conjunto de premisas de acuerdo con un sentido intencional de palabras del vocabulario”.

- Finalmente, el W3C [World Wide Web Consortium, “Semantic Web - W3C,” W3.org, 2012] que introduce los estándares de la Web Semántica indica que “Una ontología define los términos a utilizar para describir y representar un área de conocimiento. Las ontologías son utilizadas por las personas, las bases de datos, y las aplicaciones que necesitan compartir un dominio de información (un dominio es simplemente un área de temática específica o un área de conocimiento, tales como medicina, fabricación de herramientas, bienes inmuebles, reparación automovilística, gestión financiera, entre otras). Las ontologías incluyen definiciones de conceptos básicos del dominio, y las relaciones entre ellos, que son útiles para los computadores. Codifican el conocimiento de un dominio y también el conocimiento que extiende los dominios. En este sentido, hace el conocimiento reutilizable.

Para lograr sus objetivos, las ontologías requieren de varios componentes que permiten representar el conocimiento de algún dominio:

Conceptos: son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento.

Relaciones: representan la interacción entre los conceptos del dominio. Suelen formar la taxonomía del dominio. Por ejemplo: subclase-de, parte-de, parte-exhaustiva-de, conectado-a.

Funciones: son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología. Por ejemplo, pueden aparecer funciones como categorizar-clase, asignar fecha.

Instancias: se utilizan para representar objetos determinados de un concepto, o ejemplos de este.

Axiomas: son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología.

2.5. OWL

Luego de definir las ontologías como tecnología estándar para la representación de conocimiento en la Web Semántica, es importante conocer que se han usado varios lenguajes para la representación de las ontologías. Entre estos se destacan SHOE, DAML-ONT [McGuinness et al., 2003], OIL [Fensel et al., 2001], DAML+OIL [McGuinness et al., 2002]. Todos estos lenguajes siguen la estructura presentada por Berners-Lee [BERNERS-LEE et al., 2001] basada en la W3C. Sin embargo, el lenguaje estándar actual para la representación de ontologías es el OWL (Ontology Web Language). Los documentos OWL representan ontologías

utilizando el formato estandarizado OWL y los archivos .owl. La sintaxis está basada en etiquetas y es similar a la estructura XML. Hoy en día, OWL tiene la versión 2 aprobada a mediados del 2009 y tiene tres sublenguajes: OWL Full, OWL DL y OWL Lite. Los elementos no técnicos o de dominio específicos más importantes definidos en el OWL son las clases, las propiedades, los individuos y los tipos de datos [Bechhofer et al., 2004]. La versión OWL 1 contiene las siguientes funcionalidades [Fensel et al., 2001]:

- Definición de clases mediante restricciones sobre propiedades, valores o cardinalidad
- Definición de clases mediante operaciones booleanas sobre otras clases: intersección, unión y complemento
- Relaciones entre clases (p.ej. inclusión, disyunción, equivalencia)
- Propiedades de las relaciones (p.ej. inversa, simétrica, transitiva)
- Cardinalidad (p.ej. “únicamente una”)
- Igualdad y desigualdad de clases
- Igualdad y desigualdad de instancias
- Clases enumeradas

El OWL 2 incorpora las siguientes nuevas funcionalidades:

- Claves
- Cadenas de propiedades
- Tipos de datos y rangos de datos más ricos
- Restricciones de cardinalidad
- Propiedades asimétrica, reflexiva y disyuntiva
- Capacidades de anotación mejoradas
- Nuevos sublenguajes

Las nuevas incorporaciones acercan la sintaxis de OWL a un grafo RDF. Además, permiten un mapeo de un grafo RDF a OWL, lo cual no era posible con la versión OWL 1 [Fensel et al., 2001].

OWL Lite usa solo algunas de las funciones del lenguaje OWL y tiene más limitaciones en el uso de las funciones que OWL DL u OWL Full. Por ejemplo, en OWL Lite las clases solo se pueden definir en términos de superclases con nombre (las superclases no pueden ser expresiones arbitrarias), y solo se pueden usar ciertos tipos de restricciones de clase. La equivalencia entre clases y las relaciones de subclase entre clases también solo se permiten entre clases con nombre, y no entre expresiones de clase arbitrarias. Del mismo modo, las restricciones en OWL Lite usan solo clases con nombre. OWL Lite también tiene una noción limitada de cardinalidad: las únicas cardinalidades que se pueden establecer explícitamente son 0 o 1 [McGuinness et al., 2004].

2.6. SPARQL

El Lenguaje de consultas SPARQL para RDF es un formato de datos de gráfico etiquetado dirigido para representar información en la Web. SPARQL se puede usar para expresar consultas en diversas fuentes de datos, ya sea que los datos se almacenen de forma nativa como RDF o se vean como RDF a través de middleware. SPARQL contiene capacidades para consultar patrones gráficos requeridos y opcionales junto con sus conjunciones y disyunciones. Además, SPARQL admite pruebas de valor extensible y consultas de restricción por el gráfico RDF de origen. Los resultados de las consultas SPARQL pueden ser conjuntos de resultados o gráficos RDF [Prud'hommeaux and Seaborne, 2007]. EL RDF es un lenguaje basado en el XML de aserciones simples que está diseñado para representar la información en forma de tripletas, es decir, sentencias que contienen un sujeto, un predicado y un objeto.

2.7. Procesamiento de Leguaje Natural(PLN)

PLN es un área de la inteligencia artificial y la lingüística computacional que se enfoca en la interacción entre los seres humanos y las máquinas a través del lenguaje humano. Su objetivo principal es permitir que las computadoras comprendan, interpreten y generen lenguaje natural de manera similar a como lo hacen los humanos.

El lenguaje natural es la forma en que las personas se comunican, ya sea hablando, escribiendo o leyendo. Sin embargo, el lenguaje natural es altamente complejo y ambiguo, lo que hace que el procesamiento y comprensión de este lenguaje por parte de las máquinas sea un desafío significativo.

El PLN se aplica en una amplia variedad de tareas y aplicaciones, algunas de las cuales incluyen:

- Reconocimiento y generación de lenguaje: Incluye el reconocimiento de voz (como los asistentes de voz) [Hsu et al., 2021] la generación de texto (como *chatbots* o sistemas

de recomendación) [Puduppully and Lapata, 2021].

- Análisis de sentimientos: Determinar si un texto tiene una connotación positiva, negativa o neutra.
- Extracción de información: Identificar información clave y estructurada en un texto no estructurado, como nombres de personas, lugares, fechas, etc. [Vashishth et al., 2018].
- Traducción automática: Traducir texto de un idioma a otro de manera automática.
- Resumen de texto: Generar un resumen breve de un texto más largo [Bui et al., 2021].
- Clasificación de texto: Asignar una categoría o etiqueta a un texto según su contenido [Howard and Ruder, 2018].
- Análisis de opiniones y reseñas: Identificar opiniones y comentarios específicos en textos, como reseñas de productos o servicios.
- Búsqueda semántica: Mejorar la precisión y relevancia de los resultados de búsqueda en motores de búsqueda [Zhou et al., 2021].

Para abordar estas tareas, se utilizan diferentes técnicas y enfoques en el PLN, que van desde modelos basados en reglas y enfoques estadísticos hasta modelos de aprendizaje automático avanzado, como redes neuronales y *Transformers*.

En el campo del Procesamiento del Lenguaje Natural (PLN), las redes neuronales han demostrado ser extremadamente efectivas para una amplia gama de tareas, gracias a su capacidad para aprender patrones complejos y representar características semánticas del lenguaje. Algunas arquitecturas de redes neuronales utilizadas en PLN se describen a continuación.

Las Redes Neuronales Convolucionales (CNN) son originalmente desarrolladas para tareas de visión por computadora, las CNN también se han aplicado con éxito en PLN. Las capas convolucionales pueden extraer características importantes de secuencias de texto, como palabras y n-gramas, para tareas como clasificación de texto, análisis de sentimiento y etiquetado de entidades [Rohman et al., 2023].

Las Redes Neuronales Recurrentes (RNN) son ideales para modelar secuencias de texto porque pueden mantener una memoria interna a través del tiempo. Son especialmente útiles en tareas que implican comprensión de contexto, como generación de texto, traducción automática y resumen de texto [Cordero Mena and Yunga Tucto, 2022].

Long Short-Term Memory (LSTM) y *Gated Recurrent Units* (GRU) son variantes de las RNN que abordan el problema del desvanecimiento y la explosión del gradiente, lo que les

permite aprender dependencias a largo plazo en secuencias de texto [Lawi et al., 2022].

Las Redes Neuronales Transformadoras (*Transformers*) [Vaswani et al., 2017] son una arquitectura revolucionaria que ha tenido un gran impacto en PLN. Permiten capturar el contexto bidireccional en un texto a través de mecanismos de atención, como se ve en el modelo BERT.

La estructura de un *Transformer* contiene un sistema de *encoding* y *decoding* por capas, y si bien muchas veces son acompañados de las Redes Neuronales Recurrentes, las capas de recurrencia son reemplazadas por un modelo de *self-attention*. A diferencia de las redes neuronales recurrentes (RNN), los *Transformers* no requieren que los datos estén ordenados. Por ejemplo, si el input del modelo es una oración en español, el *Transformer* no necesita partir desde la primera palabra continuando con la siguiente, y así de forma sucesiva para hacer finalmente la predicción. Esto le permite a los *Transformers* una mayor paralelización que las RNN lo que se traduce en menores tiempos de entrenamiento [Vaswani et al., 2017].

A continuación se muestra la arquitectura de un *Transformer* (ver Figura 2-1). Internamente un *Transformer* tiene 6 encoders y 6 decoders. El *Transformer* recibe una secuencia de entrada y la convierte en dos secuencias: una secuencia de vectores de palabras y una secuencia de codificaciones posicionales. Ambos vectores son escritos usando representaciones numéricas del texto para que la red neuronal pueda procesarlas. Cada palabra del diccionario se representa como un vector. Las codificaciones posicionales son una representación vectorial de la posición de la palabra en la oración original. Ahora, el *Transformer* une las dos secuencias y pasa el resultado a través de una serie de codificadores, seguidos de una serie de decodificadores. A diferencia de las RNN, el input no es alimentado en la red de forma secuencial, sino que se pasa todo de una vez.

Cada uno de los codificadores convierte su entrada en otra secuencia de vectores llamados codificaciones. Los decodificadores hacen lo contrario, vuelven a convertir las codificaciones en una secuencia de probabilidades de diferentes palabras de salida. Cada codificador y decodificador contiene un componente llamado mecanismo de atención, que le permite generar un contexto a cada palabra que procesa, y cómo los mismos son colocados en forma paralela, la red tiene la capacidad de procesar todas las palabras en simultáneo, lo que le brinda su ventaja. Finalmente, las probabilidades numéricas en la capa de salida se pueden convertir en otra oración en lenguaje natural usando la función *softmax*.

Encoder-Decoder Architectures como el modelo *Sequence-to-Sequence*, son utilizadas para tareas de traducción automática y generación de texto. Consisten en dos partes principales: un codificador que procesa la entrada y un decodificador que genera la salida [Park et al., 2018].

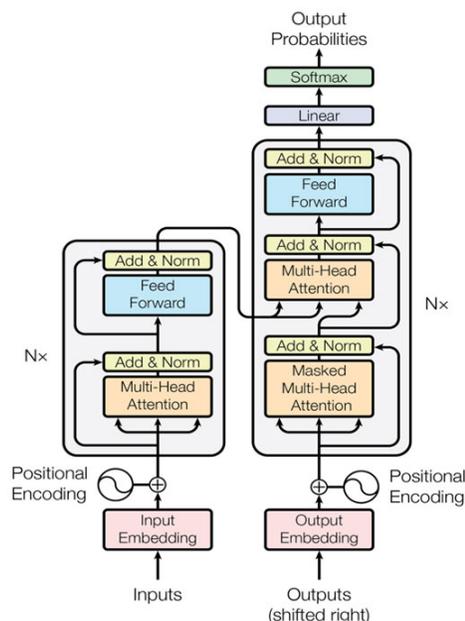


Figura 2-1.: Arquitectura de un *Transformer*. Tomado de [Vaswani et al., 2017]

Las Redes Neuronales Preentrenadas o Modelos preentrenados como BERT, GPT (*Generative Pre-trained Transformer*) y otros, han logrado resultados sobresalientes en diversas tareas de PLN al aprender representaciones semánticas profundas del lenguaje en grandes corpus de texto antes de ser ajustados finalmente para tareas específicas.

2.8. Grafos de conocimiento (GC)

Los GC (*Knowledge Graphs*) son una representación estructurada y semántica del conocimiento que captura relaciones entre entidades en un dominio específico. Estos grafos son una forma de organizar y almacenar información, donde las entidades son los nodos y las relaciones entre ellas son los bordes.

En un GC, las entidades pueden ser cualquier cosa que se quiera representar, como personas, lugares, eventos, conceptos, productos, etc. Los bordes o relaciones definen cómo se conectan las entidades entre sí y pueden tener diferentes tipos y direcciones [Hogan et al., 2021].

Los GC son utilizados para modelar el conocimiento del mundo real y representar datos de manera que se pueda acceder y comprender de forma más eficiente y significativa. Estos grafos son una forma de estructurar datos de manera que las máquinas puedan entender y procesar la información de manera más contextual y relacionada, lo que permite realizar consultas complejas y ricas en conocimiento [Fensel et al., 2020].

Un ejemplo famoso de GC es el *Knowledge Graph* de Google, que se utiliza para mejorar la precisión y relevancia de los resultados de búsqueda en su motor de búsqueda. Otros ejemplos incluyen Wikidata, DBpedia y Freebase, que son fuentes públicas de conocimiento utilizadas en diversos proyectos y aplicaciones.

Los GC desempeñan un papel fundamental en diversas áreas, como el Procesamiento del Lenguaje Natural (PLN), la Inteligencia Artificial, la recuperación de información, la recomendación y la búsqueda semántica. Al representar el conocimiento de manera estructurada y conectada, los Grafos de Conocimiento son una herramienta poderosa para mejorar la comprensión y el análisis de datos, así como para potenciar diversas aplicaciones inteligentes [Fensel et al., 2020].

Los GC y las Ontologías comparten el objetivo de representar el conocimiento de manera estructurada y semántica, pero difieren en la forma en que se organizan y se aplican. Los Grafos de Conocimiento y las Ontologías se complementan entre sí en la representación del conocimiento. Los Grafos de Conocimiento son adecuados para representar relaciones complejas y contextuales entre entidades, mientras que las Ontologías proporcionan una base formal y estructurada para definir conceptos y relaciones en dominios específicos [Ehrlinger and Wöß, 2016].

Existen técnicas de aprendizaje automático utilizadas para representar entidades y relaciones en un grafo de conocimiento, específicamente los *Knowledge Graph Embeddings (KGE)*, que generan dichas representaciones en un espacio vectorial de números reales de dimensionalidad reducida. Las KGE han demostrado ser valiosas en la recuperación de información y en tareas de aprendizaje en grafos, como la predicción de relaciones, la recomendación y el análisis de enlaces. Al mapear entidades y relaciones en un espacio vectorial, las KGE permiten realizar cálculos más eficientes y flexibles para inferir nuevas relaciones y patrones en el grafo. El objetivo del KGE es transformar las entidades y relaciones del grafo de conocimiento en representaciones vectoriales continuas y densas, también conocidas como *embeddings* o incrustaciones. Estos *embedding* permiten capturar la semántica y las relaciones implícitas presentes en el grafo, lo que facilita el razonamiento, la búsqueda semántica, la clasificación y otras tareas relacionadas con el conocimiento [Costabello et al., 2023]. En la Figura 2-2 se muestra la arquitectura de un KGE.

La arquitectura de un *Knowledge Graph Embedding (KGE)* puede variar según el enfoque y el modelo específico utilizado. Sin embargo, en general, un KGE consta de tres componentes principales:

La capa de entrada (*Lookup Layer*) que es responsable de procesar las entidades y relaciones

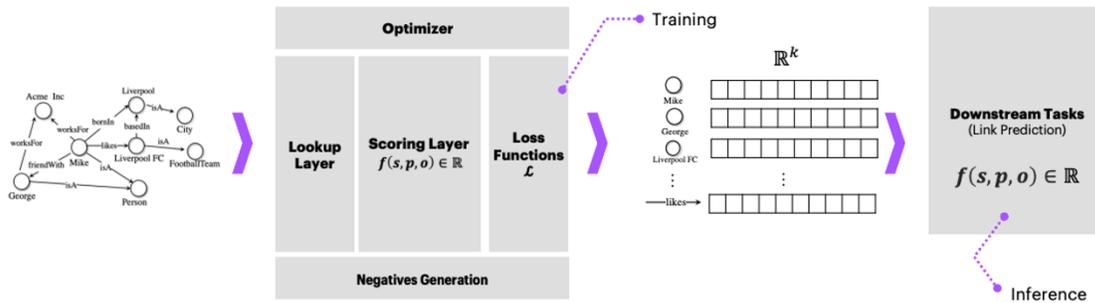


Figura 2-2.: Arquitectura de KGE. Tomado de [Costabello et al., 2023]

del grafo de conocimiento y convertirlas en representaciones vectoriales que puedan ser utilizadas por el modelo. Las entidades y relaciones se suelen representar mediante números enteros o códigos, y estas representaciones se convierten en *embeddings* o incrustaciones (vectores de números reales) en una matriz de *embeddings*. La capa de Incorporación (Scoring Layer) donde las entidades y relaciones se mapean a vectores de baja dimensión (*embeddings*) en un espacio vectorial continuo y denso. Los *embeddings* son parámetros entrenables del modelo y se actualizan durante el proceso de entrenamiento para capturar la semántica y las relaciones entre las entidades y relaciones en el grafo de conocimiento. Existen diferentes enfoques para la generación de *embeddings*, como TransE, TransH, DistMult, ComplEx, etc., y finalmente, la función de Pérdida (Loss Function) que es fundamental para el proceso de entrenamiento del modelo KGE. Esta función compara las similitudes calculadas entre entidades y relaciones en el espacio vectorial con las similitudes reales presentes en el grafo de conocimiento. El objetivo es minimizar la diferencia entre las similitudes calculadas y las reales para mejorar la calidad de los *embeddings* y, por lo tanto, la representación del conocimiento [Costabello et al., 2023].

Dentro de la literatura se han visto avances importantes en este tema, tales como:

Uso de KGE para el aprendizaje de mejores representaciones para captar el *Know-how* de un proceso particular a través del aprendizaje de grafos de conocimiento procedimentales basado en *sum-based Embedding Agregation* [Nordsieck et al., 2022].

KGE para generación de representaciones de grafos de conocimiento para tareas de *Data Aumentation* en dominios poco estructurados [Chiatti and Daga, 2022], [Erd et al., 2022].

Uso de KGE para predecir relaciones a través de los modelos existentes Dismult, ComplEx y TransE que ayudan a predecir relaciones semánticamente válidas [Hubert et al., 2022].

Predicción de relaciones causales eventos a través del uso de los KGE para reforzar GC de

temáticas como la Economía y el Cambio Climático [Khatiwada et al., 2022].

Uso de BERT (*Bidirectional Encoder Representations from Transformers*) y GC para resolver problemas de etiquetas múltiples con datos de entrenamiento limitados [Ebiele et al., 2022].

2.9. Aprendizaje ontológico

El aprendizaje de ontologías es entendido dentro de la literatura como la creación automática o semiautomática de ontologías, incluyendo la extracción de términos de dominio de una fuente específica de extracción (Página HTML, documento de lenguaje natural, Documento XML, red social, etc.). El proceso de construcción manual de ontologías tiene cierto grado de complejidad y, además, consume mucho tiempo, por ello se utilizan técnicas de aprendizaje automático y analítica predictiva para generar los procesos de aprendizaje de ontologías [Maedche and Staab, 2001], [Wong et al., 2017]. Los esfuerzos de investigación se dirigen gradualmente hacia el aprovechamiento de la inteligencia colectiva en la web, sin desconocer las complejidades reales involucradas en el descubrimiento automático de conceptos, relaciones e incluso axiomas. El ciclo de ingeniería para el aprendizaje ontológico generalmente contiene varias etapas: fusión (importar y reusar), extracción, recorte, refinamiento, aplicación y validación.

El aprendizaje de ontologías a partir de las diferentes fuentes de extracción permite entonces la identificación de términos, conceptos, relaciones y axiomas, donde es fundamental la aplicación de técnicas de recuperación de información y la extracción de datos.

Diversos autores han utilizado diferentes algoritmos para el aprendizaje de ontologías, generalmente se pueden clasificar en una de las siguientes categorías: técnicas de aprendizaje basadas en estadística, técnicas de aprendizaje basadas en reglas o técnicas de aprendizaje híbridas [Zhou, 2007]. Dentro de las técnicas basadas en la estadística se derivan propuestas de recuperación de información o minería de datos, para extracción de términos y jerarquías [NERI, 2005], [Turney, 2002]. Las técnicas basadas en lingüística se usan para el procesamiento de lenguaje natural muy usadas en el aprendizaje ontológico y finalmente, las técnicas basadas en lógica son menos usadas por que abordan temas más complejos como identificación de relaciones o axiomas.

Otro punto que se encuentra fundamental es el tema de la evaluación del modelo aprendido con el fin de asegurarse que el modelo ontológico concuerda con el conocimiento que se desea representar. Existen varias capas que se deben evaluar: 1) la capa léxica, en la que se revisan los términos y conceptos que forman el modelo; 2) la capa taxonómica, en las que se observan si las relaciones de clasificación aprendidas son las correctas y 3) la capa no-taxonómica en la que se evalúan otro tipo de relaciones [Fraga, 2016].

Existen métodos de evaluación cuantitativos y cualitativos, siendo el cuantitativo preferido por disminuir el coste de la evaluación, generar un proceso con capacidad de ser automatizado y obtener resultados similares a los que involucraría un proceso que involucre expertos humanos. Dentro de los métodos cuantitativos se destaca la evaluación a partir de un Gold Standard a través de medidas como el *recall* (determina en qué medida el conocimiento identificado en el modelo coincide con el conjunto del conocimiento que se debería haber identificado a partir del corpus), la precisión (determina en qué medida el conocimiento reflejado en la ontología se ha identificado correctamente, es decir, de todos los conceptos que forman parte del modelo, cuales tienen relación con el conocimiento que se quiere presentar) y el *F-Measure* (relaciona los resultados del *recall* y la precisión) [Radziwon et al., 2014], [Porzel and Malaka, 2004]. Otro tipo de métodos cuantitativos de evaluación son la evaluación basada en tareas, la evaluación basada en corpus [Brewster et al., 2004] y la evaluación basada en criterios [Vrandečić, 2009].

Dentro de la literatura han intentado modelar ontologías asociadas a los recursos humanos asociado al *e-recruitment* y algunos otros autores han intentado modelar un aprendizaje ontológico desde diferentes fuentes de extracción y utilizando diferentes técnicas de aprendizaje y evaluación:

Singto y Mingkhwan [Singto and Mingkhwan, 2013] proponen crear conceptos de carreras de TI de búsqueda semántica basados en ontología. Definen una estructura de la ontología *IT Careers Ontology (ITCO)* la cual consta de tres partes principales: Categoría de carrera de TI, Habilidad de TI y Educación de TI que se describen al anunciar el reclutamiento en anuncios de trabajos web que son ejemplos de modelos. Para extraer los términos de habilidades en TI se basan en el ACM / IEEE-CS (*IT fundamentals*) y la categoría de la carrera definida por ISCO-08 (títulos ocupacionales). La aplicación está enfocada a la búsqueda del ofertante laboral. Dentro de su metodología usan OWL como lenguaje principal para definir la ontología y SPARQL como lenguaje de consultas deductivas sobre la ontología, y finalmente el algoritmo TF/IDF como factor de ponderación para la recuperación de información. Los resultados experimentales muestran que la búsqueda semántica del ITC puede superar las limitaciones de la búsqueda en comparación con el modo tradicional de búsqueda de palabras clave y lograr una mayor tasa de *recall* y una relación de precisión.

Balachander y Moh [Balachander and Moh, 2018] en su estudio “similitud basada en ontología para habilidades de tecnología de la información” construye una ontología de habilidades personalizadas con taxonomía requerida rastreando DBpedia y se utiliza como base de conocimiento para comprender la relación entre habilidades. Las medidas de similitud semántica basadas en características se utilizan para derivar un puntaje de similitud entre las habilidades en ontología. Además de comprender la relación entre habilidades, el enfoque

también tiene como objetivo comprender las representaciones múltiples de la misma habilidad para derivar la puntuación. DBpedia se elige como fuente de información, ya que cubre una amplia gama de recursos en el dominio de la tecnología de software.

Hermelo [Alfonso-Hermelo et al., 2019a] propone un modelo de aprendizaje automático de una ontología de recursos humanos a partir de datos profesionales de redes sociales, permitiendo a los profesionales de reclutamiento generar y actualizar fácilmente esta ontología prácticamente sin intervención humana. Este autor dispone de un procedimiento para crear una ontología de recursos humanos automáticamente, utilizando una selección de la información pública semiestructurada que se encuentra en LinkedIn. Aunque todavía hay mucho por mejorar, este es un primer paso hacia un sistema que se asemeja a las ontologías de recursos humanos creadas por el hombre. Bajo la ontología ESCO jugando el papel de Golden Standard codifica jerarquías y relaciones entre ocupaciones y habilidades, competencias, calificaciones y ocupaciones. Utiliza el algoritmo Lovaina para agrupar nodos de acuerdo con sus similitudes y jerarquías, además los puntajes de calidad OntoQA para la evaluación de la ontología.

Harzallah y otros [Harzallah et al., 2002] desarrolla un modelado de competencias subyacentes a un *curriculum vitae* por medio de la construcción de una ontología para recursos humanos en el sector bancario. Este autor aborda las problemáticas del reclutamiento electrónico considerando un nuevo enfoque basado en la gestión de competencias con sus categorías y subcategorías (*Knowledge, know-how*, comportamientos, competencias fuertes y competencias suaves). El objetivo de la creación del CommOnCV es proporcionar a los usuarios finales de sitios web dedicados al reclutamiento electrónico nuevos servicios de búsqueda de empleo basados en la gestión de competencias.

Kumaran y Sankar [Kumaran and Sankar, 2013] generaron un sistema llamado EXPERT el cual pretende automatizar la selección inteligente de candidatos para el reclutamiento mediante mapeo de ontologías. El autor reconoce que, desde la participación en sitios de redes sociales, es posible encontrar el comportamiento social y características personales del candidato, es decir, la experiencia. Los resultados del experimento muestran que este modelo mejora la precisión de la correspondencia de candidatos con los requisitos de trabajo.

Soni y otros [Soni and Swaminarayan, 2017] realizan una revisión de diferentes ontologías que se utilizan para el desarrollo de sistemas de reclutamiento electrónico en línea y definen su arquitectura junto con diferentes tecnologías (Protégé 3.4, Isa Viz, Apollo, SWOOP). Definen 4 etapas de reclutamiento: análisis de requerimientos, publicación de oferta de empleo, recepción y preselección de las aplicaciones y la decisión final de reclutamiento. Dentro de sus contribuciones destaca la investigación que muestra cómo recopilar datos de los sitios de redes sociales, filtrarlos y conectarlos con los sitios de reclutamiento electrónico que

permiten obtener datos fácilmente y mejorar las recomendaciones de trabajo.

Por su parte Manzano y otros [Manzano-Macho et al., 2008] aportan de forma sustancial al aprendizaje no supervisado (mínima intervención del usuario) de ontologías de dominio, combinando fuentes heterogéneas de evidencia. En su trabajo se combinan entonces varias fuentes de información y mecanismos de extracción para construir una taxonomía de conceptos con mayor precisión. Su propuesta permite la integración de diferentes análisis estadísticos, sintácticos, semánticos y de diseño visual de documentos HTML para construir una taxonomía de conceptos mostrando la obtención de mejores resultados que cuando se utilizan individualmente. Específicamente en el Aprendizaje ontológico usan la extracción basada en patrones y la evaluación se realiza utilizando el método basado en la Golden Standard.

Existen otros trabajos encontrados relacionados al *e-recruitment* abordado desde el modelamiento ontológico [Faliagka et al., 2012], [Maree et al., 2018], [Salazar et al., 2015].

Otros autores han intentado implementar técnicas y algoritmos de aprendizaje ontológico para ontologías de diferentes dominios a los recursos humanos:

[Raza et al., 2020] presenta un modelo de aprendizaje de ontologías (OL) para crear automáticamente ontologías de dominio a partir de un conjunto de páginas HTML. La idea clave de esta investigación es combinar la estructura de la lista y los encabezados de las páginas HTML para reconocer el vocabulario de la ontología. El enfoque también incorpora relaciones de sinónimos con ontología y permite la interpretación semántica de conceptos de ontología. Implementamos el enfoque OL propuesto para construir una ontología deportiva a partir de una colección de documentos HTML de dominio deportivo. La nueva ontología deportiva se prueba con el razonador FaCT++; los resultados no muestran inconsistencia en la ontología.

[Dong et al., 2020] proponen un método de aprendizaje supervisado para descubrir relaciones de subsunción a partir de etiquetas. La clave de ese método es cuantificar la asociación probabilística entre etiquetas para caracterizar mejor sus relaciones. Además, desarrollan un algoritmo para organizar las etiquetas en jerarquías basadas en las relaciones aprendidas. Los experimentos se realizaron utilizando un gran conjunto de datos disponible públicamente, Bibsonomy, y tres bases de conocimiento populares, diseñadas por humanos o basadas en datos: DBpedia, *Microsoft Concept Graph* y *ACM Computing Classification System*.

[Saeeda et al., 2020] presentan un trabajo enfocado en la vinculación de entidades y la mejora de la ontología con relaciones extraídas entre conceptos. En este trabajo se discuten los beneficios de patrones léxico-semánticos adecuadamente diseñados en el aprendizaje de

ontologías. Y finalmente, se propone un conjunto preliminar de patrones léxico-semánticos diseñados para que el idioma checo aprenda nuevas relaciones entre conceptos en la ontología de dominio relacionado en un enfoque semisupervisado.

[Chen and Gu, 2021] expresan que la mayoría de las ontologías se desarrollan y amplían manualmente, lo que requiere que los desarrolladores posean ciertos conocimientos profesionales y requiere mucho tiempo. Y por lo anterior proponen un marco para construir y ampliar Ontología Educativa automáticamente. El marco de aprendizaje de ontología propuesto se llama 'ADOL', y puede convertir libros de texto de dominio en una ontología correspondiente de manera automática y eficiente.

[Lakzaei and Shamsfard, 2021] manifiestan que la construcción manual de ontologías es una tarea muy costosa, lenta y propensa a errores, la automatización del proceso de construcción de ontologías o, en otras palabras, el aprendizaje de ontologías a partir de los recursos existentes, es una buena opción. Además, expresan que una gran cantidad de datos en la web se almacenan en bases de datos relacionales, pero las bases de datos no se pueden utilizar directamente en la web semántica. Y por lo anterior, proponen un nuevo enfoque para crear automáticamente una ontología OWL a partir de una base de datos relacional. Definen un conjunto de reglas para analizar todos los componentes de la base de datos y convertirlos en componentes de ontología correspondientes. Su contribución más importante es el conjunto de reglas que pueden analizar y extraer elementos de ontología de procedimientos almacenados, funciones definidas por el usuario, vistas, herencia múltiple, la representación específica de herencia única, atributos comunes y las restricciones en las tablas y sus columnas.

[Mhammedi et al., 2021] proponen un enfoque para aprender la ontología OWL a partir de datos en la base de datos *Couchbase* mediante la aplicación de seis reglas de mapeo, usando el razonador Ontop para evaluar la consistencia de la ontología extraída. Mencionan que Big Data son colecciones de grandes conjuntos de datos de datos estructurados y no estructurados caracterizados por volumen, variedad y velocidad. Estas características acentúan la heterogeneidad y complejidad de los datos, que excede la capacidad de los sistemas tradicionales para hacerles frente. Por las características antes mencionadas, es crucial tener una visión conceptual unificada de los datos, así como una representación eficiente del conocimiento para el manejo de *big data*.

En enfoques modernos para la extracción de conceptos y relaciones para el aprendizaje de ontología [Katyshev and Anikin, 2023] y las últimas investigaciones sobre aprendizaje de máquinas para procesamiento de lenguaje natural y recuperación de información se implementa el uso de un modelo bidireccional con una arquitectura transformadora *Transformers* con la capacidad de codificar relaciones entre pares de entidades para extraer relaciones entre conceptos, así como utilizar un enfoque más rápido basado en el mecanismo de atención

Attention (técnica basada en redes neuronales que pretende imitar la atención cognitiva) para extraer directamente los conceptos a partir de textos en lenguaje natural para construir ontologías temáticas.

Ahora bien, por medio de la Tabla **2-2** se resume los resultados sobre los enfoques actuales en dicho tema, analizando características importantes a tener en cuenta en el tema de aprendizaje ontológico [Alfonso-Hermelo et al., 2019b], [Manzano-Macho et al., 2008], [Balachander and Moh, 2018], [Raza et al., 2020], [Xu et al., 2020], [Dong et al., 2020], [Saeeda et al., 2020], [Gao and Chen, 2020], [Chen and Gu, 2021], [Lakzaei and Shamsfard, 2021], [Mhammedi et al., 2021], [Vrolijk et al., 2022] y [Katyshev and Anikin, 2023]. En dicha tabla se condensa la información de los artículos citados.

Se realiza la revisión sistemática de la literatura basada en la Metodología de Kitchenham [Tebe et al., 2019] que consta de tres etapas: planificar la revisión, conducir la revisión y documentar la revisión.

En la etapa 1 se identifica la necesidad de revisión, es decir, resumir la evidencia existente respecto de cómo se ha abordado el tema del aprendizaje ontológico y el tema de las habilidades de TI para fortalecer o apoyar los procesos de e-recruitment. Se definen también preguntas de investigación alrededor de la temática y se toma como principal fuente de información SCOPUS® (base de datos bibliográfica de resúmenes y citas de artículos de revistas científicas). Adicionalmente, se define el protocolo de revisión el cual consiste en seleccionar de forma manual y a través del diccionario Linguee (<https://www.linguee.com/>) los términos, las expresiones de búsqueda y los algoritmos de búsqueda a partir de un grupo de términos candidatos que respondían al objetivo de la necesidad de revisión. Se definen tres términos: aprendizaje ontológico (variable 1), habilidades TI (variable 2) y *e-recruitment* (variable 3). Finalmente, se definen 7 criterios de inclusión/exclusión.

En la etapa 2 se obtienen los resultados de búsqueda de información (cantidad de publicaciones) en SCOPUS asociados a las expresiones o palabras claves individuales. Para todas las expresiones definidas en la variable 1 se encontraron 7.570 artículos, para las expresiones de la variable 2 se encontraron 2.627 artículos y para las expresiones de la variable 3 se encontraron 766 artículos. Luego, se obtienen los resultados de los algoritmos de búsqueda combinando las tres variables definidas. En total se obtienen los resultados de 9 algoritmos de búsqueda. Para la búsqueda de estudios primarios se usa cada uno de los algoritmos de búsqueda. Para la variable 1, 2 y 3 se seleccionan los primeros 40 artículos filtrados por “más citados” y los primeros 40 artículos filtrados por “Relevancia”. Para los otros algoritmos, se seleccionan todos los artículos encontrados. Se obtuvo inicialmente para cada una de las 3 variables 80 papers de estudios primarios y 2 papers para la mezcla de variables. Lo anterior generó un total de 242 artículos científicos o estudios primarios para comenzar a aplicarles

los criterios de inclusión/exclusión. Luego de aplicar dichos criterios se obtuvo 13 artículos para la variable 1, 29 artículos para la variable 2, 22 artículos para la variable 3 y 2 artículos para la variable 1+variable 2+ variable 3.

	Aprendizaje de términos	Aprendizaje de conceptos	Aprendizaje de relaciones	Aprendizaje de axiomas	Técnica de aprendizaje ontológico	Métrica o técnica de evaluación	Aprendizaje ontológico
[Alfonso-Hermelo et al., 2019]	SI	SI	NO	NO	Minería y clustering	Cuantitativa: Puntajes de calidad OntoQA, Golden Standard	Semisupervisado
[Manzano-Macho et al., 2008]	NO	SI	SI	NO	Técnica de aprendizaje híbridas (estadísticos, sintácticos, semánticos y de diseño visual de documentos HTML)	Cuantitativa: Golden Standard	No supervisado
[Balachander & Moh, 2018]	SI	SI	SI	NO	Estadístico	Cualitativa: Normalized Discounted Cumulative Gain (NDCG)- 10 expertos	Supervisado
[Raza et al., 2020]	NO	SI	SI	NO	Extracción encabezados HTML	Razonadores semánticos, precisión	Semisupervisado
[Xu et al., 2020]	SI	SI	SI	NO	Selección de características, reducción de dimensionalidad (K-means, K-medoids, Affinity propagation y algoritmos de agrupación)	No se menciona	No supervisado
[Dong et al., 2020]	NO	SI	SI	NO	Limpieza de datos, representación (probabilístico), generación de características (probabilístico), clasificación y enriquecimiento del conocimiento	Cuantitativa: Golden Standard Cualitativa: Expertos humanos	Semisupervisado
[Saeeda et al., 2020]	NO	SI	SI	NO	Patrones léxicos semánticos	Presición, recall, Fi-score	Semisupervisado
[Gao and Chen, 2020]	NO	SI	SI	NO	Algoritmo multivisión, estructuras de árboles	No se menciona	Supervisado
[Chen and Gu, 2021]	SI	SI	SI	NO	Riqueza de relaciones, Riqueza de atributos y Riqueza de enlaces. Revisa riqueza en profundidad, amplitud y relevancia	Riqueza de relaciones, riqueza de atributos y riqueza de enlaces. revisa riqueza en profundidad, amplitud y relevancia	Semisupervisado
[Lakzaei and Shamsfard, 2021]	NO	SI	SI	SI	No se encuentra	Presición, recall, F1-score	Semisupervisado
[Mhammedi et al., 2021]	NO	SI	SI	SI	Reglas de mapeo	Razonador Ontop (consistencia)	No se describe
[Vrolijk et al., 2022]	SI	NO	SI	NO	Ontojob (propio)	Presición, recall, F1-score	No supervisado
[Katyshev and Anikin, 2023]	NO	SI	SI	NO	Transformer (modelo de deep learning)	No se menciona	Supervisado

Tabla 2-2.: Comparación de características en trabajos de aprendizaje de ontologías. Elaboración propia

En las primeras cuatro columnas de la Tabla 2-2 se revisa que se está aprendiendo en la ontología: **términos, conceptos, relaciones o axiomas** (si se aprende la característica se pone SI y si no se aprende se pone NO). En la columna de técnicas de **aprendizaje ontológico**, se mencionan los algoritmos o métodos que se usan para el aprendizaje ontológico, en la columna **métrica o técnica de evaluación** se mencionan las técnicas o métricas

de evaluación que usa el autor en su trabajo. Finalmente, en la columna **aprendizaje ontológico** se describe si las técnicas utilizadas en el trabajo para aprender la ontología usan mayormente aprendizaje supervisado, no supervisado o semisupervisado.

2.10. Análisis y conclusiones del capítulo

Como se evidencia, la mayoría de los enfoques acerca del aprendizaje ontológico afirman que el proceso de construcción manual de ontologías es un proceso laborioso, complejo, que requiere conocimiento, esfuerzo y además, puede estar propenso a errores. Por lo anterior, el gremio científico está aunando esfuerzos en automatizar el proceso de creación de ontologías teniendo en cuenta que estas deben extenderse y evolucionar en el tiempo con el fin de mejorar la gestión del conocimiento en la misma.

Así entonces, los autores intentan incorporar métodos, algoritmos o tecnologías que les permitan aprender términos/conceptos, relaciones y axiomas, manteniendo la calidad sintáctica y semántica de la ontología. Un desafío generalizado en el aprendizaje ontológico es claramente el aprendizaje de axiomas y restricciones, debido a la complejidad de generalizar o deducir un gran número de relaciones conocidas que satisfagan ciertos criterios para formar dichos axiomas. En la revisión bibliográfica se observa que no todos los autores realizan un trabajo de aprender todas las capas de la torta de aprendizaje, unos aprenden sólo términos y conceptos, otros términos, conceptos y relaciones, y solo algunos llegan a incursionar en el aprendizaje de axiomas y restricciones. Es importante resaltar que el proceso de aprendizaje de ontologías es un proceso sistemático y debe aprenderse en el orden establecido términos/-conceptos, relaciones y axiomas/restricciones. El punto es que los esfuerzos de investigación actuales se encuentran en las etapas de mejora de las técnicas de reconocimiento de términos existentes o avanzando hacia la fase de descubrimiento de relaciones. Sigue siendo una tendencia que la investigación sobre el aprendizaje de axiomas es escasa.

La tendencia para mejorar el rendimiento de la extracción de términos es construir corpus de texto muy grandes a partir de la Web. Las técnicas para construir corpus de texto varían desde el simple enfoque de consulta y descarga hasta otras más complicadas que requieren el análisis del contenido de la página web. Las redes sociales profesionales pueden servir de insumo de la web para ello. Por su parte, en cuanto al descubrimiento de relaciones taxonómicas y no taxonómicas, estamos siendo testigos de la creciente aplicación de patrones léxico-sintácticos, minería de reglas de asociación y reglas basadas en dependencias sintácticas en conjuntos de datos muy grandes de la Web. En esencia se visualiza un aumento potencial en la relevancia de las técnicas basadas en la lógica para el proceso de aprendizaje de ontologías.

Ahora, en el proceso de evaluación ontológica existe mucho camino por recorrer, debido a que no existe una forma de evaluación que determine si una ontología es de calidad en cuanto a su sintaxis y su semántica asociada al dominio en el cual se desarrolla. La evaluación es un aspecto importante del aprendizaje de la ontología porque permite validar y verificar las ontologías resultantes y posiblemente guiar y refinar el proceso de aprendizaje. Uno de los desafíos actuales es incorporar herramientas o algoritmos para evaluar la capa taxonómica (relaciones) y la capa no-taxonómica (axiomas y restricciones).

3. Capítulo 3: Ontología base

En este capítulo se definirán los conceptos básicos para el dominio de *e-recruitment* asociado a perfiles profesionales en TI, con el fin de definir la ontología base del modelo. La ontología base se entenderá como aquella a partir de la cual se va a iniciar el aprendizaje de las capas de la ontología.

Crear una ontología desde cero requiere mucha experticia en el dominio y mucho esfuerzo. La premisa fundamental de las ontologías es que puedan reutilizar el conocimiento existente en el dominio, por esto se busca inicialmente ontologías realizadas con el objetivo de utilizarse en procesos de reclutamiento en línea o recursos humanos. Así mismo, se buscan entidades regulatorias que permitan tener un conocimiento base sobre las ocupaciones en TI.

Dentro de algunos artículos relacionados con soluciones tecnológicas para el *e-recruitment* se mencionan algunas de las ontologías adoptadas en este dominio. Se realiza un cuadro comparativo para explorar la adopción de alguna ontología en particular. Se define la ontología más adoptada en las aplicaciones de *e-recruitment*. Se estudia la estructura de la ontología. Y finalmente, se propone la ontología base de este trabajo. En la Figura 3-1 se muestra el diagrama de la metodología para el desarrollo de este capítulo.



Figura 3-1.: Metodología de desarrollo objetivo 1. Elaboración propia

3.1. Adopción de ontologías en soluciones tecnológicas para e-recruitment

Artículo	Ontología Usada
Job-Candidate Matching using ESCO Ontology (Shakya & Paudel, 2019)	ESCO
An Ontology Driven ESCO LOD Quality Enhancement (Kahlawi, 2020)	ESCO
Automatically Learning a Human-Resource Ontology from Professional Social-Network Data (Alfonso-Hermelo et al., 2019)	ESCO
Leveraging the Inherent Hierarchy of Vacancy Titles for Automated Job Ontology Expansion	ESCO
Ontology-guided Job Market Demand Analysis (Sibarani et al., 2017)	SARO
Skills-Matching and Skills Intelligence through curated and data-driven ontologies (Rentzsch, 2020)	ESCO, O*NET, Actonomy
Towards smart statistics in labour market domain (Novalija, 2020)	ESCO, SARO
Brecha de habilidades de los jóvenes en el mercado laboral colombiano (Díaz & Correa, 2020)	ESCO
ESCO : Boosting Job Matching in Europe with Semantic Interoperability (Vrang et al., 2014)	ESCO
Towards an Enterprise Software Component Ontology (Seedorf & Schader, 2011)	ESCO

Tabla 3-1.: Revisión de ontologías adoptadas en soluciones tecnológicas en el dominio de *e-recruitment*. Elaboración propia

En la Tabla **3-1** se condensa la búsqueda de los artículos relacionados con soluciones tecnológicas para el dominio de *e-recruitment* ([Shakya and Paudel, 2019], [Kahlawi, 2020], [Alfonso-Hermelo et al., 2019c], [Van Hautte et al., 2020], [Sibarani et al., 2017], [Rentzsch and Staneva, 2020], [Novalija and Grobelnik, 2018], [Salas and Salas, 2020], [le Vrang et al., 2014], [Seedorf and Schader, 2011]). Se revisaron 10 de los más recientes y principales artículos internacionales, que se acercan al dominio y de ellos se revisó la adopción de la ontología y como se define en la columna ontología usada, la que más aparece es la ontología **ESCO**.

Dentro de las otras ontologías que aparecen en esta revisión que son adoptadas para aplicacio-

nes en el dominio de *e-recruitment*, se encuentra también la **O*NET**, **SARO** y **Actonomy**. Sin embargo, los autores sugieren que la ontología ESCO se caracteriza por tener mejores características de completitud, aplicabilidad, disponibilidad y documentación en la web.

En la Figura 3-2 se puede ver el panorama actual de la ontología ESCO.



Figura 3-2.: Representación del panorama actual de la ontología ESCO, tomado de [European Commission website, 2022]

A continuación se mencionan las características más relevantes de la ontología ESCO:

- ESCO proporciona descripciones de 2942 ocupaciones y 13.485 habilidades relacionadas con estas ocupaciones, traducidas a 27 idiomas
- Describe la relación jerárquica de las habilidades y ocupaciones
- Clasifica todas las habilidades relacionadas con las ocupaciones
- Está diseñada para permitir el cálculo de las distancias que existen entre las habilidades
- Es una ontología multilingüe
- Marco estándar entre ocupaciones y habilidades para la Unión Europea
- Es publicada como Datos Abiertos Enlazados (LOD) para que pueda reutilizarse fácilmente y vincularse a otra fuente de datos, como clasificaciones de ocupaciones nacionales.

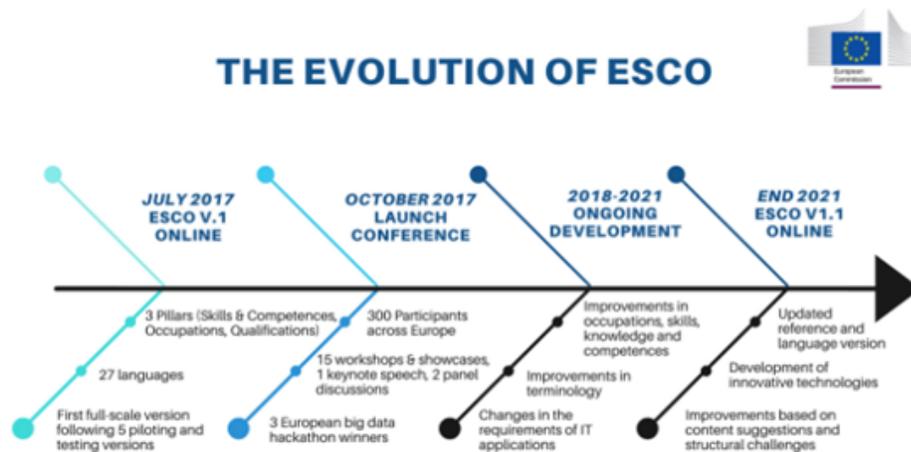


Figura 3-3.: Evolución de la ESCO, tomado de [European Commission website, 2022]

3.2. Ontología ESCO

ESCO es la clasificación europea multilingüe de Habilidades, Competencias y Ocupaciones, la cual funciona como un diccionario, describiendo, identificando y clasificando ocupaciones profesionales y habilidades relevantes para el mercado laboral de la Unión Europea (UE) y la Educación y Formación. Los conceptos y las relaciones entre ellos pueden entenderse mediante sistemas electrónicos, lo que permite que diferentes plataformas en línea utilicen la ESCO para servicios como vincular a los solicitantes de empleo con puestos de trabajo en función de sus habilidades, sugerir capacitaciones a las personas que desean volver a capacitarse o mejorar, entre otras [European Commission website, 2022].

El objetivo de ESCO es apoyar la movilidad laboral en Europa y, por lo tanto, un mercado laboral más integrado y eficiente, ofreciendo un “lenguaje común” sobre ocupaciones y habilidades que puedan ser utilizadas por diferentes partes interesadas en temas de empleo y educación y formación.

La comisión europea proporciona acceso a la clasificación ESCO a través de una interfaz de programación API, que permite el acceso a datos de la clasificación de habilidades, competencias, calificaciones y ocupaciones europeas.

La ESCO tiene actualmente 7 versiones: V0, V0.1, V0.2, V0.8, V0.9, V1 y V1.1. En la Figura 3-3 se puede evidenciar el proceso evolutivo y de mejora de esta ontología la cual inicia con los 3 pilares y la primera versión completa de la ESCO y finaliza con una versión en diferentes idiomas y desarrollada para el desarrollo de tecnologías innovadoras.

ESCO debe mejorarse continuamente para seguir siendo un valor agregado para el mercado

laboral y los sistemas de educación y formación, ya que los empleadores solicitan regularmente nuevas ocupaciones y habilidades emergentes y también se introducen regularmente cambios en los planes de estudio y en la terminología en los programas de educación y formación. Para abordar estos cambios, es importante compartir con frecuencia comentarios, sugerencias y propuestas sobre cómo mejorar el contenido y la gestión de la clasificación con las organizaciones que utilizan la ESCO y otras partes interesadas de la ESCO.

3.2.1. ¿Por qué usar la ESCO?

Las personas cambian de trabajo y de empleadores con más frecuencia que en el pasado, se necesitan regularmente nuevas competencias y la movilidad geográfica y ocupacional está aumentando. Las plataformas de talento en línea, como las bolsas de trabajo y las redes sociales, están transformando la forma en que se lleva a cabo la contratación. Los empleadores y los solicitantes de empleo utilizan cada vez más las herramientas digitales para publicar y solicitar ofertas de trabajo o para buscar y ofrecer oportunidades de formación. Las empresas y los proveedores de educación y formación necesitan información clara y actualizada sobre habilidades y calificaciones para gestionar mejor el talento y abordar las brechas de habilidades en los programas de educación y formación [European Commission website, 2022].

Los conceptos y descripciones de la ESCO pueden ayudar a las personas a comprender:

- qué conocimientos y habilidades se requieren generalmente cuando se trabaja en una ocupación específica [Shakya and Paudel, 2019];
- qué conocimientos, habilidades y competencias se obtienen como resultado de una calificación específica [le Vrang et al., 2014];
- qué calificaciones exigen o solicitan con frecuencia los empleadores a quienes buscan trabajo en una ocupación específica [Novalija and Grobelnik, 2018].

El uso de los conceptos de ESCO también puede facilitar la transición a este mercado laboral digital en constante crecimiento, al ofrecer un lenguaje común sobre ocupaciones y competencias que hace que estas herramientas digitales se comuniquen y funcionen mejor juntas [European Commission website, 2022].

3.2.2. ¿Cómo se puede utilizar la ESCO?

ESCO se utiliza en aplicaciones y plataformas en línea

Los conceptos de ESCO y las relaciones entre ellos pueden entenderse mediante sistemas electrónicos. Esto permite que diferentes sistemas y plataformas utilicen ESCO para sugerir

los trabajos más relevantes a los solicitantes de empleo en función de sus habilidades, competencias o las capacitaciones más relevantes para las personas que desean volver a capacitarse o mejorar sus habilidades [European Commission website, 2022].

ESCO conecta personas y trabajos

ESCO contiene descripciones precisas de ocupaciones y competencias que se pueden utilizar tanto en la redacción de ofertas de trabajo como de CV, lo que facilita la comparabilidad y el emparejamiento. Esto ayudará tanto a los empleadores a encontrar a las personas adecuadas para sus vacantes como a los solicitantes de empleo a encontrar los trabajos adecuados para sus habilidades [European Commission website, 2022].

ESCO conecta el empleo con la educación

El uso de ESCO en ofertas de trabajo, investigación, análisis de *big data*, etc., ayuda a los proveedores de educación y formación a comprender qué habilidades o competencias necesitan los mercados laborales. Luego, pueden adaptar sus planes de estudio en consecuencia para preparar mejor a sus estudiantes para los mercados laborales del mañana. La ESCO también ayuda a los posibles empleadores a comprender mejor lo que han aprendido los estudiantes [European Commission website, 2022].

ESCO conecta los mercados laborales y promueve la movilidad

ESCO está disponible en 27 idiomas, lo que permite a los solicitantes de empleo y empleadores comunicarse mejor sobre habilidades o competencias, formación y puestos de trabajo en cualquier idioma europeo elegido. Mediante su uso en el portal EURES y en muchas otras plataformas en línea, ayuda a los servicios de empleo públicos y privados europeos a ofrecer sus servicios a través de fronteras, idiomas y sistemas electrónicos, para que puedan ayudar mejor a las personas que desean trabajar en otro país europeo [European Commission website, 2022].

3.2.3. Clasificación de los pilares de la ESCO

La ontología ESCO considera los 3 siguientes pilares:

Ocupaciones:

Utiliza relaciones jerárquicas entre ellos, metadatos y asignaciones a la Clasificación Internacional Uniforme de Ocupaciones (ISCO) para estructurar las ocupaciones. Cada concepto de ocupación contiene un término preferido y cualquier número de términos no preferidos y términos ocultos en cada uno de los idiomas de la ESCO. También incluye información sobre profesiones reguladas que son relevantes en el contexto de esta ocupación. Cada ocupación también viene con un perfil ocupacional. Los perfiles contienen una explicación de la ocupación en forma de descripción, nota de alcance y definición.

La Clasificación Internacional Uniforme de Ocupaciones (ISCO-08) sirve como estructura jerárquica para el pilar de ocupaciones. Cada ocupación de la ESCO se asigna exactamente a un grupo de unidades de la ISCO-08 (nivel 4). Por lo tanto, la ISCO-08 proporciona los cuatro niveles superiores del pilar de ocupaciones. ESCO proporciona el quinto y los niveles inferiores de la estructura jerárquica con su lista de ocupaciones (ver Figura 3-4).

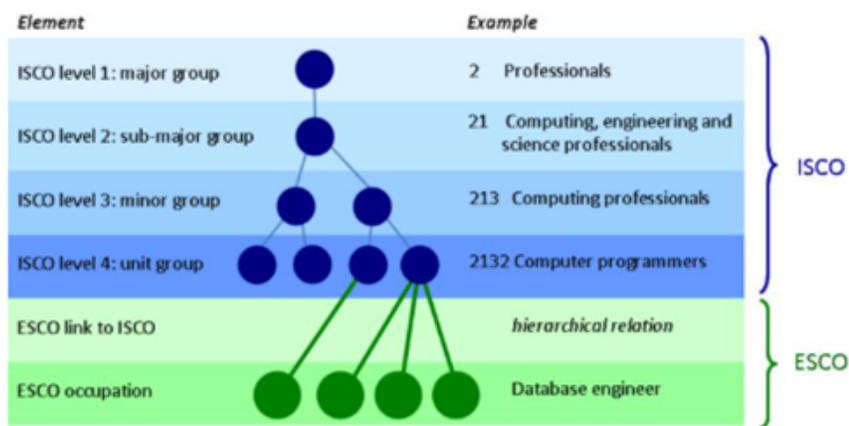


Figura 3-4.: Jerarquías de ocupación en la ESCO usando la ISCO-08, tomado de [Chala, 2018]

Además, ESCO contiene relaciones jerárquicas entre las ocupaciones de ESCO. De esta manera, la ESCO muestra que las especialidades tienen un alcance más limitado que una ocupación más genérica. Por ejemplo, ESCO contiene una ocupación *bartender* y dos especialidades *cocktail bartender* y *barista*, que están jerárquicamente vinculadas a ella.

En la Tabla 3-2 se muestra la clasificación de las ocupaciones en 10 actividades. Así mismo, estas 10 actividades u ocupaciones del gran grupo (*Major Group*) tienen diferentes jerarquías agrupadas en un subgrupo principal (*Sub Major Group*), los grupos menores (*Minor Group*) y grupos de unidades (*Unit Group*) [Organización internacional del trabajo, 2005] .

ID	Ocupaciones (Gran grupo)	Jerarquía 1 ISCO-08 (Subgrupo principal)
0	Ocupaciones militares	01 Oficiales de las fuerzas armadas 02 Suboficiales de las fuerzas armadas 03 Otros miembros de las fuerzas armadas
1	Directores y gerentes	11 Directores ejecutivos, personal directivo de la administración pública y miembros del poder ejecutivo y de los cuerpos legislativos 12 Directores administradores y comerciales 13 Directores y gerentes de producción y operaciones 14 Gerentes de hoteles, restaurantes, comercios y otros servicios
2	Profesionales científicos e intelectuales	21 Profesionales de las ciencias y de la ingeniería 22 Profesionales de la salud 23 Profesionales de la enseñanza 24 Especialistas en organización de la administración pública y de empresas 25 Profesionales de tecnología de la información y las comunicaciones 26 Profesionales en derecho, en ciencias sociales y culturales
3	Técnicos y profesionales de nivel medio	31 Profesionales de las ciencias y la ingeniería de nivel medio 32 Profesionales de nivel medio de la salud 33 Profesionales de nivel medio en operaciones financieras y administrativas 34 Profesionales de nivel medio de servicios jurídicos, sociales, culturales y afines 35 Técnicos de la tecnología de la información y las comunicaciones
4	Personal de apoyo administrativo	41 Oficinistas 42 Empleados en trato directo con el público 43 Empleados contables y encargados del registro de materiales 44 Otro personal de apoyo administrativo

Sigue en la página siguiente.

ID	Ocupaciones (Gran grupo)	Jerarquía 1 ISCO-08 (Subgrupo principal)
5	Trabajadores de los servicios y vendedores de comercios y mercados	51 Trabajadores de los servicios personales 52 Vendedores 53 Trabajadores de los cuidados personales 54 Personal de los servicios de protección
6	Agricultores y trabajadores calificados agropecuarios, forestales y pesqueros	61 Agricultores y trabajadores calificados de explotaciones agropecuarias con destino al mercado 62 Trabajadores forestales calificados, pescadores y cazadores 63 Trabajadores agropecuarios, pescadores, cazadores y recolectores de subsistencia
7	Oficiales, operarios y artesanos de artes mecánicas y de otros oficios	71 Oficiales y operarios de la construcción excluyendo electricistas 72 Oficiales y operarios de la metalurgia, la construcción mecánica y afines 73 Artesanos y operarios de las artes gráficas 74 Trabajadores especializados en electricidad y la electrotecnología 75 Operarios y oficiales de procesamiento de alimentos, de la confección, ebanistas, otros artesanos y afines
8	Operadores de instalaciones y máquinas y ensambladores	81 Operadores de instalaciones fijas y máquinas 82 Ensambladores 83 Conductores de vehículos y operadores de equipos pesados móviles
9	Ocupaciones elementales	91 Limpiadores y asistentes 92 Peones agropecuarios, pesqueros y forestales 93 Peones de la minería, la construcción, la industria manufacturera y el transporte 94 Ayudantes de preparación de alimentos 95 Vendedores ambulantes de servicios y afines 96 Recolectores de desechos y otras ocupaciones elementales

Tabla 3-2.: Clasificación de ocupaciones según la ISCO-08

Competencias:

El pilar de habilidades de la ESCO, que a veces se denomina pilar de habilidades y competencias, aplica una definición amplia de habilidades. Contiene no solo habilidades, sino también conocimientos, habilidades y competencias. Este pilar hace una distinción entre Competencias/Habilidades y Conocimientos.

La jerarquía ESCO de habilidades y competencias contiene las siguientes subclasificaciones: conocimientos, competencias, competencias transversales y habilidades de lenguaje. Esta jerarquía está siempre en un proceso de continua mejora. En la Tabla **3-3** se observa la jerarquización principal de este pilar. Cada ítem de este pilar puede contener elementos tales como, descripción, definición, notas explicativas, etiquetas alternativas, nivel de reutilización de la capacidad, URI del concepto, entre otras [European Commission website, 2022].

ID	Competencias	Capacidades/competencias más especializadas
K	Conocimientos	agricultura, silvicultura, pesca y veterinaria administración de empresas y derecho educación ingeniería, industria y construcción campo desconocido programas y certificaciones genéricos salud y bienestar tecnologías de la información y la comunicación (tic) ciencias naturales, matemáticas y estadística servicios ciencias sociales, periodismo e información
L	Competencias Lingüísticas y conocimiento de lenguas	L1 -Lenguas L2 - lenguas clásicas
S	Competencias	S1 - comunicación, colaboración y creatividad S2 - competencias en materia de información S3 - prestar asistencia y cuidados S4 - competencias de gestión S5 - trabajar con ordenadores S6 - manipular y mover S7 - construir S8 - trabajar con maquinaria y equipo especializado

Sigue en la página siguiente.

ID	Competencias	Capacidades/competencias más especializadas
T	Competencias Transversales	T1 - capacidades y competencias básicas T2 - capacidades y competencias de razonamiento T3 - capacidades y competencias de autogestión T4 - capacidades y competencias sociales y de comunicación T5 - capacidades y competencias físicas y manuales T6 - capacidades y competencias para la vida

Tabla 3-3.: Clasificación de competencias según la ESCO

Cualificaciones:

El pilar de las cualificaciones apoya la comprensión de las cualificaciones individuales que necesitan los empleadores, los servicios de empleo públicos y privados, los estudiantes, los trabajadores, los solicitantes de empleo, los proveedores de educación y formación y otros agentes. Esta información debe ser lo más completa y transparente posible para satisfacer las necesidades de información de estas partes interesadas. Las cualificaciones son el resultado formal de un proceso de evaluación y validación, obtenido cuando una autoridad competente establece que una persona ha obtenido los resultados de aprendizaje correspondientes a una norma dada. Por lo tanto, solo se muestran en Europass los datos de cualificaciones que incluyen la siguiente información básica:

Título: Título exacto de la calificación, Campo: Basado en Campos de Educación y Formación ISCED 2013, País / Región, Marco europeo de cualificaciones (EQF): solo es relevante para las cualificaciones que ya tienen un nivel EQF asignado a través del proceso de referencia de los marcos nacionales de cualificaciones al EQF, Organismo adjudicador o autoridad competente, Descripción de la calificación expresada en resultados de aprendizaje. El nivel EQF es el marco común de referencia que relaciona entre sí los sistemas de cualificaciones de los países y sirve de mecanismo de conversión para mejorar la interpretación y comprensión de las cualificaciones de diferentes países y sistemas de Europa.

La información sobre las cualificaciones a escala europea figura en Europass [europass Union Europea, 2022] y procede de bases de datos de cualificaciones nacionales que reflejan los marcos nacionales de cualificaciones que son propiedad de los Estados miembros europeos que los gestionan. Europass ofrece el repertorio de datos de alta calidad mejor surtido y más actualizado sobre las cualificaciones, los marcos nacionales de cualificaciones y las oportunidades de aprendizaje en Europa, lo que ayuda a los aprendientes a encontrar un curso en otro país y a los empleadores, a comprender el valor de una calificación de otro Estado miembro de la UE.

El Marco Europeo de Cualificaciones (MEC) es un marco de ocho niveles para todo tipo de cualificaciones, basado en los resultados de aprendizaje, que sirve como instrumento de conversión entre los diferentes marcos nacionales de cualificaciones. Este marco contribuye a mejorar la transparencia, la comparabilidad y la transferibilidad de las cualificaciones de las personas y permite comparar las cualificaciones de distintos países e instituciones [Europass, 2017].

El MEC cubre todos los tipos y niveles de cualificación y, gracias al uso de los resultados de aprendizaje, pone claramente de manifiesto lo que una persona sabe, comprende y es capaz de hacer. Presenta una estructura de niveles ascendentes de competencia (desde el 1 hasta el 8) y, lo que es más importante, está estrechamente vinculado a los marcos nacionales de cualificaciones. De ese modo, puede proporcionar una cartografía completa de todos los tipos y niveles de cualificación existentes en Europa, a los que cada vez es más fácil acceder a través de bases de datos [Europass, 2017].

Los 8 niveles progresivos o EQF definidos por el MEC son:

- Nivel 1: conocimientos básicos
- Nivel 2: conocimientos fácticos básicos en un campo concreto
- Nivel 3: conocimientos de hecho, principios, procesos y conceptos en un campo concreto
- Nivel 4: conocimientos fácticos y teóricos en contextos amplios en un campo concreto
- Nivel 5: amplios conocimientos especializados, fácticos y teóricos en un campo concreto, siendo consciente de los límites de conocimiento (Técnicos y Tecnólogos)
- Nivel 6: conocimientos avanzados en un campo que requiera comprensión crítica de teorías y principios (Profesional y especialización)
- Nivel 7: conocimientos altamente especializados, algunos a la vanguardia en un campo en concreto, que sienten bases de una investigación original (Maestría)
- Nivel 8: conocimientos avanzados de un campo concreto con inculcación a otros (Doctorado)

3.2.4. Taxonomía de la ESCO

La ontología ESCO usa para su modelado lo siguiente:

- Propiedades de mapeo de SKOS para relacionar los conceptos del pilar ESCO con conceptos en otras taxonomías (externas) (por ejemplo, CINE-FoET e ISCO-08. Se pueden agregar más mapeos en el futuro)

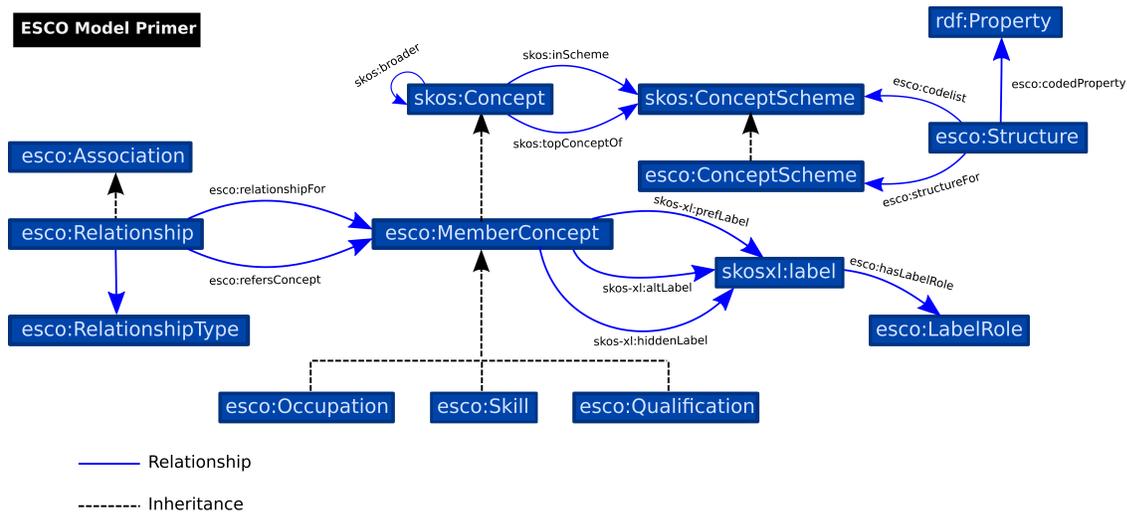


Figura 3-5.: Modelo de la ontología ESCO, tomado de [European Commission, 2022]

- Etiquetado de conceptos de pilar ESCO por otras taxonomías (externas) (NUTS, EQF, NACE, ...), estos taggings se definen utilizando estructuras ESCO
- Captura detalles específicos de género en las etiquetas de los conceptos del pilar ESCO
- Relaciones ricas de conceptos de ESCO que contienen una descripción y otras características específicas de la relación entre dos conceptos pilares de ESCO

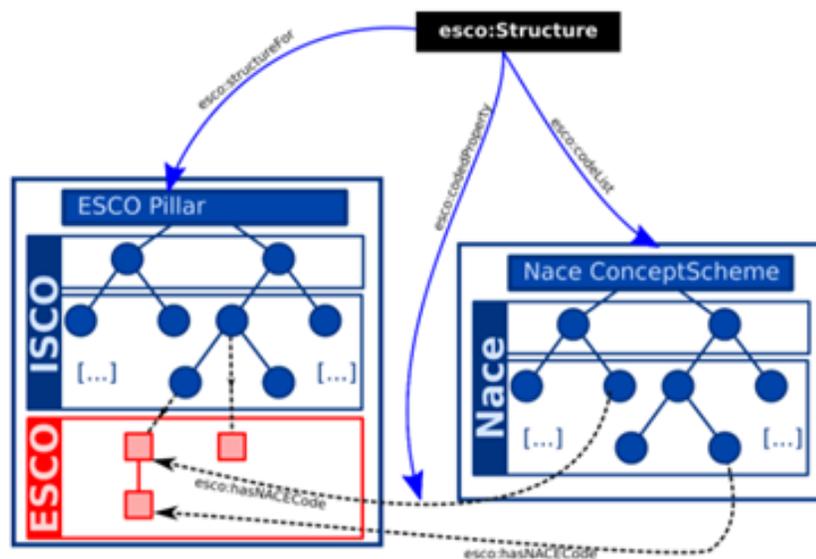


Figura 3-6.: Relaciones importantes de la ontología ESCO, tomado de [European Commission, 2022]

Además, mantiene 2 registros adicionales: Un organismo adjudicador (control y vigilancia) y los organismos adjudicadores del contexto laboral suelen estar referenciados por las calificaciones de la ESCO. Las ocupaciones pueden tener uno o más contextos laborales. En la Figura **3-5** se puede ver el modelo de la ontología ESCO.

La estructura define una relación de etiquetado entre uno o más esquemas de conceptos y un pilar ESCO. La estructura identifica el predicado utilizado para la relación entre un concepto esco: MemberConcept y un concepto skos: que forma parte de los esquemas de Conceptscheme (ver Figura **3-6**).

En la siguiente lista se pueden ver las *Classes* disponibles en la ontología ESCO.

CLASES ESCO:

- Accreditation
- Association object
- Awarding activity
- Awarding body
- Competence sub-type
- Concept scheme (ESCO pillar)
- entry requirement
- Esco extension
- ESCO Structure
- Label role
- Member concept
- node literal
- Occupation
- Qualification
- Recognition
- Skill
- Work context

A continuación, se enuncian también los *Object properties*, los *Data properties* y finalmente, el *Named individuals* y los *Annotation properties*.

OBJECT PROPERTIES:

- Additional note
- awarding body (descriptive)
- has code list
- has coded property
- has accreditation
- has association
- has awarding activity
- has awarding body
- has entry requirement
- ISCED FoET Code
- has label role
- has NACE code
- has poly-hierarchy
- can have work context
- association from
- is essential skill for
- is indirectly included
- is optional skill for
- location belongs to ISCO
- Group with code
- to ESCO concept
- has essential skill

- has optional skill
- replaced by
- replaces
- requirement level
- skill type
- defines structure for
- supported language
- association to
- target framework
- way to acquire
- has version record

DATA PROPERTIES:

- editorial status
- qualification expiration period
- expiry period
- ECTS credit points
- has poly-hierarchy
- is indirectly included
- is partial qualification
- language
- review date
- target notation
- volume of learning
- status

NAMED INDIVIDUALS:

- skill reuse level

ANNOTATIONS PROPERTIES

- model axiom or term status
- original label
- regulated profession note

3.3. Ocupaciones y habilidades de los profesionales en TI

La Organización para la Cooperación y el Desarrollo Económicos (OCDE) es una organización internacional que en colaboración con otros gobiernos diseña y promueve políticas públicas para establecer estándares internacionales y proponer soluciones basadas en datos a diversos retos sociales. Este ente regulador emitió en el 2010 el documento *OECD Information Technology Outlook 2010* ([OECD, 2010]) que proporciona una amplia visión general de las tendencias y perspectivas a corto plazo en la industria de la tecnología de la información (TI), análisis del impacto creciente de la TI en la economía y la sociedad, desarrollos y aplicaciones emergentes en áreas seleccionadas de tecnología de la información y una revisión de las políticas de TI. La edición de 2010 se basa en ediciones anteriores para ampliar aún más el análisis económico y de políticas. Esta edición se centra ampliamente en la crisis económica y la recuperación y sus impactos en el lado de la oferta de las tecnologías de la Información y la Comunicación. Se tomará como base referente la información de este ente regulador, debido a que es el más confiable y experto en el área de las ocupaciones y específicamente en el área de TI.

La OCDE define entonces 3 categorías de competencias TIC (especialistas en TIC, usuarios avanzados y usuarios básicos), como se muestra en la Tabla **3-4**. El empleo relacionado con las tecnologías de la información y la comunicación (TIC) es una proporción relativamente grande del empleo total, y contribuye tanto al empleo agregado como cambios estructurales en los perfiles de habilidades [OECD, 2010]. Básicamente, la primera categoría cubre las personas quienes brindan los servicios o herramientas y las otras dos categorías quienes las utilizan.

Se relaciona la categoría de empleo de especialista ICT con el nivel Altamente cualificado (*High Skilled*). Así mismo la categoría de Usuario avanzado con el nivel Medianamente cualificado (*Medium Skilled*) y la categoría de Usuarios básicos con el nivel Cualificación baja (*Low Skilled*). En la Figura **3-7** se muestra como las ocupaciones relacionadas con TI se

#	Empleo TIC	Definición
1	Especialistas en TIC	Aquellos quienes tienen la capacidad de desarrollar, operar y mantener sistemas de TIC. Las TIC constituyen la parte principal de su trabajo
2	Usuarios avanzados	Aquellos usuarios competentes en herramientas de software avanzadas, y a menudo específicas del sector. Las TIC no son el trabajo principal sino una herramienta
3	Usuarios básicos	Aquellos usuarios competentes en herramientas genéricas (por ejemplo, Word, Excel, Outlook, PowerPoint) necesarias para la sociedad de la información, el gobierno electrónico y la vida laboral. Aquí también, las TIC no son el trabajo principal sino una herramienta

Tabla 3-4.: Categorías de competencias TIC. Elaboración propia

distribuyen con el nivel de cada categoría mencionada anteriormente.

Así entonces, la OCDE también realiza un estudio particular para establecer la clasificación de ocupaciones (CIUO o ISCO) como se mostró en la Tabla **3-2**, la cual pretende organizar los empleos en una serie de grupos definidos en función de las tareas de cada empleo, con el fin de ayudar a definir claramente soluciones para la contratación de trabajadores, la gestión de migración de trabajadores y el desarrollo de programas de formación. Esta clasificación ha sido mejorada durante el transcurso del tiempo, la primera versión fue realizada en 1957 conocida como la ISCO-58, la segunda versión de dio en 1966 llamada ISCO-68 y en 1987 se sacó la versión ISCO-88 la cual fue mejorada y ajustada finalmente en 2007 para dar lugar a la última versión ISCO-08.

El autor [Arnal et al., 2001] a partir de la clasificación ISCO-08, propone una nueva clasificación de las ocupaciones relacionadas con TI como se muestra en la Figura **3-8** donde las numeraciones en negrilla representan las ocupaciones y jerarquías dedicadas específicamente a las labores de TI.

Se espera cada vez más que los especialistas en TIC tengan habilidades adicionales y esto es porque las empresas de TI tienden a buscar una combinación de tres tipos principales de habilidades y competencias ([Lopez-Bassols, 2002]) como se muestra en la Tabla **3-5**.

Las habilidades no se pueden medir fácilmente, pero si se pueden adquirir a través de varias vías. Se puede capturar de características observables como las habilidades naturales, la educación formal, la experiencia laboral, la capacitación laboral interna o externa y la

Table C.6.19. IT-related occupations: US Department of Commerce

IT-related occupations	
High skill level	
Engineering managers	Computer and information systems managers
Computer and information scientists, research	Computer programmers
Computer software engineers, applications	Computer software engineers, systems software
Computer support specialists	Computer systems analysts
Database administrators	Network and computer systems administrators
Network systems and data communication analysts	Computer hardware engineers
Electrical engineers	
Electrical and electronic engineering technicians	Electronics engineers, except computer
Moderate skill level	
Computer, automated teller, and office machine repairers	Data entry keyers
Electromechanical equipment assemblers	Electrical power line installers and repairers
Electrical power line installers and repairers	Semiconductor processors
Telecommunications line installers and repairers	Telecommunications equipment installers and repairers, except line installers
Electrical and electronics repairers, commercial and industrial	
Low skill level	
Communications equipment operators	Computer operators
Billing and posting clerks and machine operators	Other office machine operators

Source: US Department of Commerce, 2003a.

StatLink: <http://dx.doi.org/10.1787/722742184887>

Figura 3-7.: Ocupaciones relacionadas con TI categorizadas con el nivel de cada categoría, tomado de [OECD, 2004]

#	Habilidad TIC	Definición
1	Técnica	Principalmente en temas asociados a TI, pero también análisis cuantitativo/modelado de datos, medios digitales, redacción técnica, etc.
2	Negocios y Administración	En particular marketing, estrategia y redacción comercial
3	Personal	Comunicación, liderazgo, trabajo en equipo, habilidad para resolver problemas

Tabla 3-5.: Habilidades/competencias de TI buscadas por las empresas. Elaboración propia

capacitación personal o aprendizaje no formal. La combinación de todo lo anterior permite deducir que COMPETENCIA TI = (CONOCIMIENTO + HABILIDAD + DISPOSICIÓN PERSONAL) en el contexto como se muestra en la Figura 3-10.

3.4. Proceso de diseño y creación de la ontología base

De acuerdo con la literatura, no hay una forma correcta de modelar un dominio debido a que siempre existen alternativas viables. Se dice que la mejor solución casi siempre depende de la aplicación para la cual se va a usar y sus propósitos futuros.

Table C.6.21. Classification of IT occupations

3-digit ISCO88	
Knowledge workers Engineers and applied and social scientists	<ul style="list-style-type: none"> • Physical, mathematical and engineering science professionals (211, 212, 214) • Life science and health professionals (221, 222, 223) • Teaching professionals (231) • Other professionals (241, 242, 244, 247) • Physical and engineering science associate professionals (311, 313, 314, 315) • Life science and health associate professionals (321, 322, 323) • Other associate professionals (341, 342)
Computer specialists	<ul style="list-style-type: none"> • Computing professionals (213) • Computer associate professionals (312)
Managerial workers	<ul style="list-style-type: none"> • Legislators and senior officials (111, 114) • Corporate managers (121, 122, 123) • Manager of small enterprises (131)
Data workers	<ul style="list-style-type: none"> • Teaching professionals (232, 233, 234, 235) • Other professional (243) • Teaching associate professionals (331, 332, 333, 334) • Other associate professionals (343, 344) • Office clerks (411, 412, 413, 414, 419) • Customer service clerks (421, 422)
Service workers	<ul style="list-style-type: none"> • Other professionals (245, 246) • Other associate professionals (345, 346, 347, 348) • Personal and protective service workers (511, 512, 513, 514, 515, 516) • Models, sales persons and demonstrators (521, 522) • Sales and services elementary occupations (911, 912, 913, 914, 915, 916)
Goods producing workers	<ul style="list-style-type: none"> • Skilled agricultural and fishery workers (611, 612, 613, 614, 615) • Extraction and building trade workers (711, 712, 713, 714) • Metal, machinery and related trades workers (721, 722, 723, 724) • Precision, handicraft, printing and related trades workers (731, 732, 733, 734) • Other craft and related trades workers (741, 742, 743, 744) • Stationary plant, and related operators (811, 812, 813, 814, 815, 816, 817) • Machine operators and assemblers (821, 822, 823, 824, 825, 826, 827, 828, 829) • Drivers and mobile plant operators (831, 832, 833, 834) • Agricultural, fishery and related labourers (921) • Labourers in mining, construction, manufacturing and transport (931, 932, 933)

Note: Occupations in bold are included in this chapter's broad definition of ICT-skilled employment; the main overlap occurs within the first three categories of workers.

Source: Arnal et al. (2001), based on Lavoie and Roy (1998).

Figura 3-8.: Clasificación de las Ocupaciones relacionadas con TI, tomado de [Arnal et al., 2001]

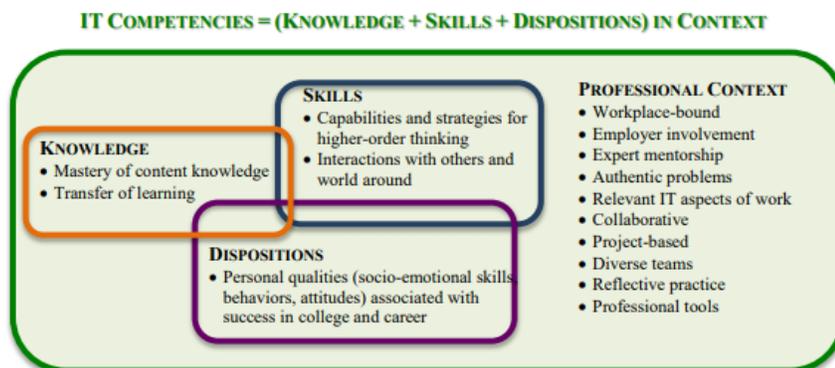


Figure 4.2 IT competency model

Figura 3-9.: Modelo de las competencias en TI, tomado de [Lopez-Bassols, 2002]

Existen diferentes metodologías descritas por autores que permiten abordar el proceso de creación de ontologías. Para el desarrollo de la ontología base de esta tesis, se tendrá en cuenta la metodología propuesta por Bernaras, Laresgoiti y Corera bajo el proyecto Espirit KACTUS en el dominio de redes eléctricas [Bernaras et al., 1996] y la metodología propuesta por la Universidad de Stanford EEUU para la creación de ontologías *Ontology Development 101* [Noy and McGuinness, 2001]. Es oportuno mencionar que dichas metodologías son para el diseño de ontologías utilizando OWL pero pueden extenderse otros lenguajes.

La metodología de Bernaras et. al define los siguientes pasos:

- (1) Especificación de la aplicación.
- (2) Diseño preliminar basado en categorías ontológicas relevantes de primer nivel.
- (3) Refinamiento y estructuración de la ontología.

La metodología de la universidad de Stanford EEUU define los siguientes pasos:

- (1) Determinar el dominio y ámbito de la ontología.
- (2) Determinar la intención de uso de la ontología.
- (3) Reutilizar ontologías o vocabularios controlados existentes.
- (4) Enumerar los términos importantes del dominio.
- (5) Definir jerarquía de clases.
- (6) Crear las instancias.

Estas dos metodologías son adecuadas para ontologías de dominio prácticas y por eso la metodología propuesta en este trabajo está basada en ellas. Se quiere lograr una correcta estructuración ontológica, pero con un enfoque práctico para el dominio de *e-recruitment*. Con

inspiración de las metodologías mencionadas se extrae una propia, que consta de:

- (1) Definición del dominio y la aplicación de la ontología.
- (2) Uso de estándares y clasificaciones existentes.
- (3) Diseño de categorías las ontológicas (definición de terminología, definición de clases y relaciones taxonómicas entre clases, propiedades del objeto y propiedades de los datos, definición de instancias).

3.4.1. Definición del dominio y la aplicación de la ontología

La ontología base se diseña en este capítulo, servirá como referencia para implementar el modelo de aprendizaje ontológico que se propondrá más adelante. La ontología base está fundamentada en la construcción del conocimiento alrededor de las profesiones y/o profesionales de las Tecnologías de la Información y la Comunicación (TIC), para facilitar y potencializar algunas necesidades de las áreas de recursos humanos que contratan profesionales de este oficio a través de una plataforma en línea. El *e-recruitment* o reclutamiento digital se hace cargo de manera automática de las tareas más tediosas de un proceso de selección, como por ejemplo comparar currículos por experiencia, formación, habilidades o competencias.

El *e-recruitment* es un fenómeno que está en aumento constante y consiste en el proceso de ofrecer y encontrar empleados utilizando internet. Esta modalidad tiene un gran éxito debido a que el internet ofrece un bajo costo y una gran velocidad de distribución de la oferta y también una amplitud a nivel global que lo hace el lugar perfecto para buscar habilidades en TI. Diferentes estudios de la OCDE demuestran que en este momento son el medio más utilizado a nivel mundial para encontrar trabajo.

El reto en el *e-recruitment* consiste en poder evaluar y confirmar que las habilidades del aspirante son las habilidades necesarias para el cargo. Cuando se trata de perfilar un aspirante se hace apertura de la necesidad de un dominio ontológico que permita realizarlo de manera clara abordando todos los aspectos indispensables.

En el caso del desarrollo de esta tesis se abordará la ontología base y el proceso de aprendizaje ontológico a partir de las siguientes premisas:

1. La ontología servirá como lenguaje común sobre ocupaciones y habilidades para TI.
2. Se trabajará en la categoría de Especialistas TIC, Usuarios avanzados y Usuarios básicos
3. Se tomarán en cuenta las habilidades duras y blandas de los profesionales de TI.
5. La ontología será en español.

Preguntas de competencia

¿Cómo diseñar una ontología que pueda servir de base para hacer un proceso de *e-recruitment* básico para profesionales de las tecnologías de información y comunicación considerando la definición formal de este tipo de modelos y representarlos en un enfoque de modelado semántico?

¿Qué clases, relaciones e individuos se deben considerar para asegurar la cobertura de las características más importantes para un sistema de *e-recruitment* de profesionales en el área de TI?

¿Existe alguna entidad que ya estandariza las ocupaciones en el área de TI que pueda enriquecer o servir como base del modelo de la ontología base?

¿Existe alguna entidad que ya estandariza las competencias que pueda enriquecer o servir como base del modelo de la ontología base?

¿Cómo deben definirse las clases, jerarquías de clase e instancias para que la ontología base sirva de base para luego ser extendida a través del aprendizaje ontológico?

3.4.2. Uso de estándares y clasificaciones existentes

La Organización Internacional del Trabajo (OIT) diseñó la Clasificación Internacional Uniforme de Ocupaciones (CIUO), también conocida en inglés como la (ISCO).

La Clasificación Internacional Uniforme de Ocupaciones es una herramienta para organizar los empleos en una serie de grupos definidos claramente en función de las tareas que comporta cada empleo. Está destinada tanto a los usuarios del campo de la estadística como a los usuarios orientados al cliente. Las principales aplicaciones orientadas al cliente las encontramos en la contratación de trabajadores a través de oficinas de empleo, en la gestión de migraciones de trabajadores entre países de duración corta o larga, así como en el desarrollo de programas de formación y orientación profesionales [OIT, 2022].

La primera versión de la Clasificación Internacional Uniforme de Ocupaciones fue adoptada en 1957 por la Conferencia Internacional de Estadísticos del Trabajo. Es conocida como la CIUO-58. Esta versión fue reemplazada rápidamente por la CIUO-68, que fue adoptada por la Undécima Conferencia Internacional de Estadísticos del Trabajo en 1966. La tercera versión, que es la actual, de la Clasificación Internacional de Estadísticos del Trabajo, la CIUO-88, fue adoptada por la Decimocuarta Conferencia Internacional de Estadísticos del Trabajo en 1987. Muchas clasificaciones nacionales de ocupaciones se basan en una de

estas versiones de la CIUO, dependiendo de la fecha en que fueron desarrolladas [OIT, 2022].

La CIUO ha sido recientemente actualizada para tomar en cuenta el desarrollo en el mundo del trabajo desde 1988 y para mejorarla teniendo en cuenta las experiencias adquiridas en la aplicación de la CIUO-88. La actualización no afectó los principios básicos ni la estructura de la clasificación, pero se efectuaron perfeccionamientos en ciertas áreas. La clasificación actualizada fue adoptada en diciembre del 2007 y es conocida como la CIUO-08 [OIT, 2022].

Para el desarrollo específico de la ontología base de esta tesis se usará la clasificación más reciente CIUO-08. Esta clasificación se encuentra en <https://www.ilo.org/public/spanish/bureau/stat/isco/isco08/index.htm>, en PDF, word o excel. Esta clasificación maneja una codificación de hasta 4 dígitos. Las de 1 dígito son los grandes grupos, las de 2 dígitos son subgrupos principales, las de 3 dígitos son subgrupos y finalmente, las de 4 dígitos son grupos primarios. Los empleos se clasifican por ocupación con respecto al tipo de trabajo realizado o que se ha de realizar. Los criterios básicos utilizados para definir el sistema de grandes grupos, subgrupos principales, subgrupos y grupos primarios son el “nivel de competencias” y la “especialización de las competencias” requeridos para efectuar eficazmente las tareas y cometidos de las ocupaciones.

Para delimitar esta clasificación de la OIT en el área de interés de las tecnologías de la información TI, se utiliza la nueva clasificación de las ocupaciones relacionadas con TI a partir de la clasificación CIUO-08 propuesta por [Arnal et al., 2001] como se muestra en la Figura 3-8. Las numeraciones en negrilla representan las ocupaciones y jerarquías dedicadas específicamente a las labores de TI dentro de la clasificación ICIUO-08.

Por otro lado, el Ministerio de Educación Nacional de Colombia define, articula y clasifica el concepto de Competencia teniendo en cuenta la definición de integralidad que recoge de manera multidisciplinar los contenidos requeridos para el desarrollo de las capacidades, habilidades, destrezas y actitudes que posibilitan desempeño laboral. La clasificación de las competencias definidas por el Ministerio se encuentra en <https://www.mineducacion.gov.co/1621/w3-printer-299611.html>. En la figura se muestra la clasificación inicial de competencias que se tomarán para la construcción de la ontología base.

3.4.3. Diseño de las categorías ontológicas

El proceso para la definición de las categorías ontológicas base se describen a continuación. La ontología de este trabajo llamará OPTI. En la Figura 3-11 se muestra un paralelo comparativo entre el modelo de pilares básicos de OPTI y el modelo de la ontología ESCO estudiada anteriormente.

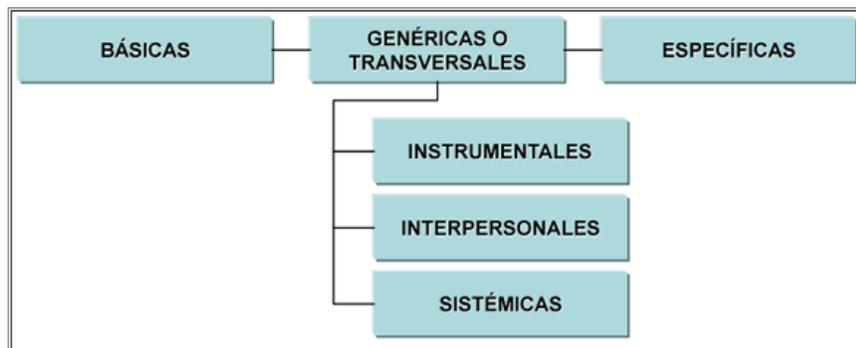


Figura 3-10.: Clasificación de competencias adoptado por el Ministerio de Educación en Colombia, tomado de <https://www.mineducacion.gov.co/1621/w3-printer-299611.html>

La ontología OPTI a diferencia de la ontología ESCO se basará en 4 pilares: Ocupaciones, Competencias, Experiencia y Nivel de estudio. Los pilares Ocupaciones y Competencias están basados en los 4 niveles de la clasificación CIUO-08 dada por la Organización Internacional del Trabajo (OIT) y la clasificación del Ministerio de Educación Nacional de Colombia respectivamente. Los pilares Experiencia y Nivel de estudio es una propuesta propia basada en las necesidades puntuales de los procesos de reclutamiento de personal.

Definición de la terminología

La terminología para definir la ontología de perfiles profesionales e TI para el dominio de *e-recruitment* se basa en aspectos importantes a la hora de definir un perfil para una vacante laboral: Ocupaciones, Competencias, Experiencia y Nivel de estudio. El propósito es cubrir los aspectos principales para brindar una ontología base en el dominio y que se adapte a procesos de aprendizaje ontológico.

A continuación, se listan los 199 términos relacionados con el dominio de la ontología:

- Ocupaciones
- Directores y gerentes
- Directores ejecutivos personal directivo de la administración pública y miembros del poder ejecutivo y de los cuerpos legislativos
- Directores generales y gerentes generales
- Directores administradores y comerciales
- Directores de administración y servicios

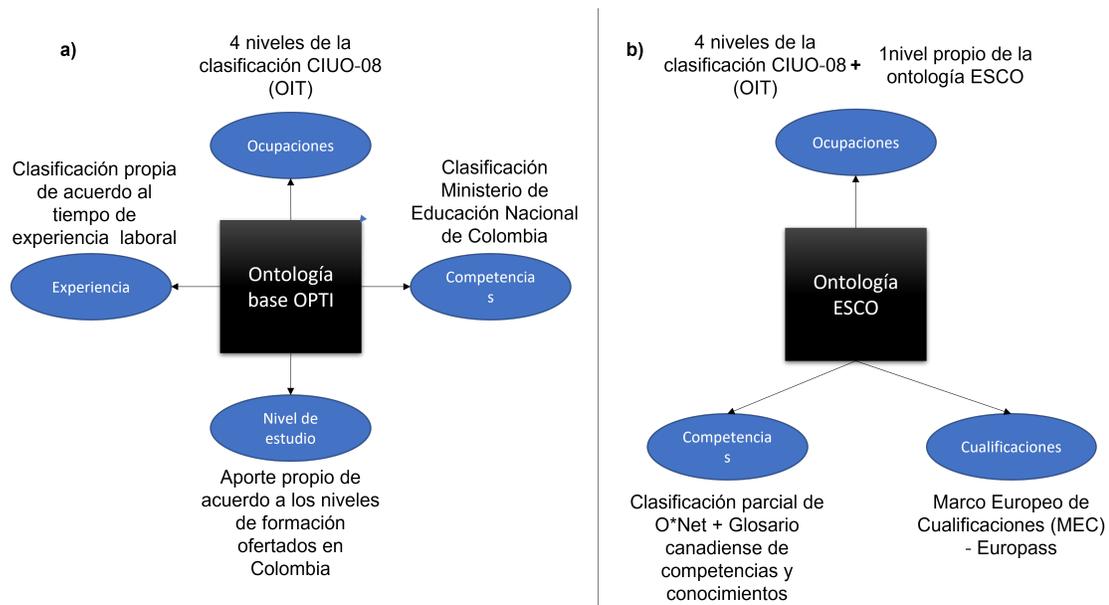


Figura 3-11.: a) modelo de pilares de la ontología OPTI vs b) modelo de pilares de la ontología OPTI. Elaboración propia

- Directores de ventas comercialización y desarrollo
- Directores y gerentes de producción y operaciones
- Directores de producción agropecuaria silvicultura y pesca
- Directores de industrias manufactureras de minería construcción y distribución
- Directores de servicios de tecnología de la información y las comunicaciones
- Directores y gerentes de servicios profesionales
- Gerentes de hoteles restaurantes comercios y otros servicios
- Gerentes de hoteles y restaurantes
- Gerentes de comercios al por mayor y al por menor
- Otros gerentes de servicios
- Profesionales científicos e intelectuales
- Profesionales de las ciencias y de la ingeniería
- Físicos químicos y afines
- Matemáticos actuarios y estadísticos

- Ingenieros excluyendo electrotecnólogos
- Ingenieros en electrotecnología
- Arquitectos urbanistas agrimensores y diseñadores
- Profesionales de la salud
- Otros profesionales de la salud
- Especialistas en organización de la administración pública y de empresas
- Especialistas en finanzas
- Especialistas en organización de administración
- Profesionales de las ventas la comercialización y las relaciones públicas
- Profesionales de tecnología de la información y las comunicaciones
- Desarrolladores y analistas de software y multimedia
- Especialistas en bases de datos y en redes de computadores
- Profesionales en derecho en ciencias sociales y culturales
- Profesionales en derecho
- Archivistas bibliotecarios curadores y afines
- Artistas creativos e interpretativos
- Técnicos y profesionales de nivel medio
- Profesionales de las ciencias y la ingeniería de nivel medio
- Supervisores en ingeniería de minas de industrias manufactureras y de la construcción
- Técnicos en control de procesos
- Profesionales de nivel medio de la salud
- Técnicos médicos y farmacéuticos
- Profesionales de nivel medio en operaciones financieras y administrativas
- Profesionales de nivel medio en finanzas y matemáticas
- Agentes comerciales y corredores

- Agentes de servicios comerciales
- Secretarios administrativos y especializados
- Profesionales de nivel medio de servicios jurídicos sociales culturales y afines
- Profesionales de nivel medio de servicios jurídicos sociales y religiosos
- Profesionales de nivel medio en actividades culturales artísticas y culinarias
- Técnicos de la tecnología de la información y las comunicaciones
- Técnicos en operaciones de tecnología de la información y las comunicaciones y asistencia al usuario
- Técnicos en telecomunicaciones y radiodifusión
- Personal de apoyo administrativo
- Oficinistas
- Secretarios general
- Operadores de máquinas de oficina
- Empleados en trato directo con el público
- Empleados de servicios de información al cliente
- Empleados contables y encargados del registro de materiales
- Auxiliares contables y financieros
- Oficiales operarios y artesanos de artes mecánicas y de otros oficios
- Trabajadores especializados en electricidad y la electrotecnología
- Instaladores y reparadores de equipos eléctricos
- Oficiales y operarios del tratamiento de la madera ebanistas y afines
- Directores financieros
- Directores de recursos humanos
- Directores de políticas y planificación
- Directores de administración y servicios no clasificados bajo otros epígrafes

- Directores de ventas y comercialización
- Directores de publicidad y relaciones públicas
- Directores de investigación y desarrollo
- Directores de producción agropecuaria y silvicultura
- Directores de producción de piscicultura y pesca
- Directores de industrias manufactureras
- Directores de explotaciones de minería
- Directores de empresas de construcción
- Directores de empresas de abastecimiento, distribución y afines
- Directores de servicios de tecnología de la información y las comunicaciones
- Directores de servicios de cuidados infantiles
- Directores de servicios de salud
- Directores de servicios de cuidado de las personas de edad
- Directores de servicios de bienestar social
- Directores de servicios de educación
- Gerentes de sucursales de bancos, de servicios financieros y de seguros
- Directores y gerentes de servicios profesionales no clasificados bajo otros epígrafes
- Gerentes de hoteles
- Gerentes de restaurantes
- Gerentes de servicios no clasificados bajo otros epígrafes
- Físicos y astrónomos
- Meteorólogos
- Químicos
- Geólogos y geofísicos
- Matemáticos, actuarios y estadísticos

-
- Ingenieros industriales y de producción
 - Ingenieros civiles
 - Ingenieros medioambientales
 - Ingenieros mecánicos
 - Ingenieros químicos
 - Ingenieros de minas, metalúrgicos y afines
 - Ingenieros no clasificados bajo otros epígrafes
 - Ingenieros electricistas
 - Ingenieros electrónicos Ingenieros en telecomunicaciones
 - Arquitectos
 - Arquitectos paisajistas
 - Urbanistas e ingenieros de tránsito
 - Cartógrafos y agrimensores
 - Farmacéuticos
 - Contables
 - Asesores financieros y en inversiones
 - Analistas financieros
 - Analistas de gestión y organización
 - Especialistas en políticas de administración
 - Especialistas en políticas y servicios de personal y afines
 - Especialistas en formación del personal
 - Profesionales de la publicidad y la comercialización Profesionales de relaciones públicas
 - Profesionales de ventas técnicas y médicas (excluyendo la TIC)
 - Profesionales de ventas de tecnología de la información y las comunicaciones
 - Analistas de sistemas

- Desarrolladores de software
- Desarrolladores Web y multimedia
- Programadores de aplicaciones
- Desarrolladores y analistas de software y multimedia y analistas no clasificados bajo otros epígrafes
- Diseñadores y administradores de bases de datos
- Administradores de sistemas
- Profesionales en redes de computadores
- Especialistas en bases de datos y en redes de computadores no clasificados bajo otros epígrafes
- Abogados
- Jueces
- Profesionales en derecho no clasificados bajo otros epígrafes
- Archivistas y curadores de museos
- Bibliotecarios, documentalistas y afines
- Directores de cine, de teatro y afines
- Supervisores de la construcción
- Técnicos en control de procesos no clasificados bajo otros epígrafes
- Técnicos en aparatos de diagnóstico y tratamiento médico
- Agentes de bolsa, cambio y otros servicios financieros
- Oficiales de préstamos y créditos
- Tenedores de libros
- Profesionales de nivel medio de servicios estadísticos, matemáticos y afines
- Tasadores
- Agentes de seguros
- Representantes comerciales

-
- Agentes de compras
 - Agentes de compras y consignatarios
 - Declarantes o gestores de aduana
 - Organizadores de conferencias y eventos
 - Agentes de empleo y contratistas de mano de obra
 - Agentes inmobiliarios
 - Agentes de servicios comerciales no clasificados bajo otros epígrafes
 - Supervisores de secretaría
 - Secretarios jurídicos
 - Secretarios administrativos y ejecutivos
 - Secretarios médicos
 - Profesionales de nivel medio del derecho y servicios legales y afines
 - Fotógrafos
 - Técnicos en operaciones de tecnología de la información y las comunicaciones
 - Técnicos en asistencia al usuario de tecnología de la información y las comunicaciones
 - Técnicos en redes y sistemas de computadores
 - Técnicos de la Web
 - Técnicos de radiodifusión y grabación audio visual
 - Técnicos de ingeniería de las telecomunicaciones
 - Operadores de máquinas de procesamiento de texto y mecanógrafos
 - Grabadores de datos
 - Empleados de agencias de viajes
 - Empleados de contabilidad y cálculo de costos
 - Empleados de servicios estadísticos, financieros y de seguros
 - Empleados encargados de las nóminas

- Instaladores y reparadores de líneas eléctricas
- Mecánicos y reparadores en electrónica
- Instaladores y reparadores en tecnología de la información y las comunicaciones
- Competencias
 - Competencias básicas
 - Competencias transversales
 - Competencias instrumentales
 - Competencias interpersonales
 - Competencias sistémicas
 - Competencias específicas
- Nivel de estudio
 - Técnico
 - Tecnólogo
 - Profesional
 - Especialista
 - Magíster
 - Doctor/Doctora
- Experiencia
 - Junior
 - Máster
 - Senior

En el Anexo A se puede observar las definiciones de la terminología a detalle.

Ahora bien, es necesario definir las clases que son esenciales para representar de una forma muy básica el perfil profesional de TI sin ambigüedades y agrupando lo más posible los términos relacionados.

Definición de clases y relaciones taxonómicas entre clases

Se define para la ontología base 4 superclases de las cuales se iniciará el proceso de relacionamiento taxonómico de la misma (ver Figura 3-12).



Figura 3-12.: Definición de superclases de la ontología base. Elaboración propia

Así mismo, se define la relación taxonómica entre clases teniendo en cuenta la Figura 3-13.

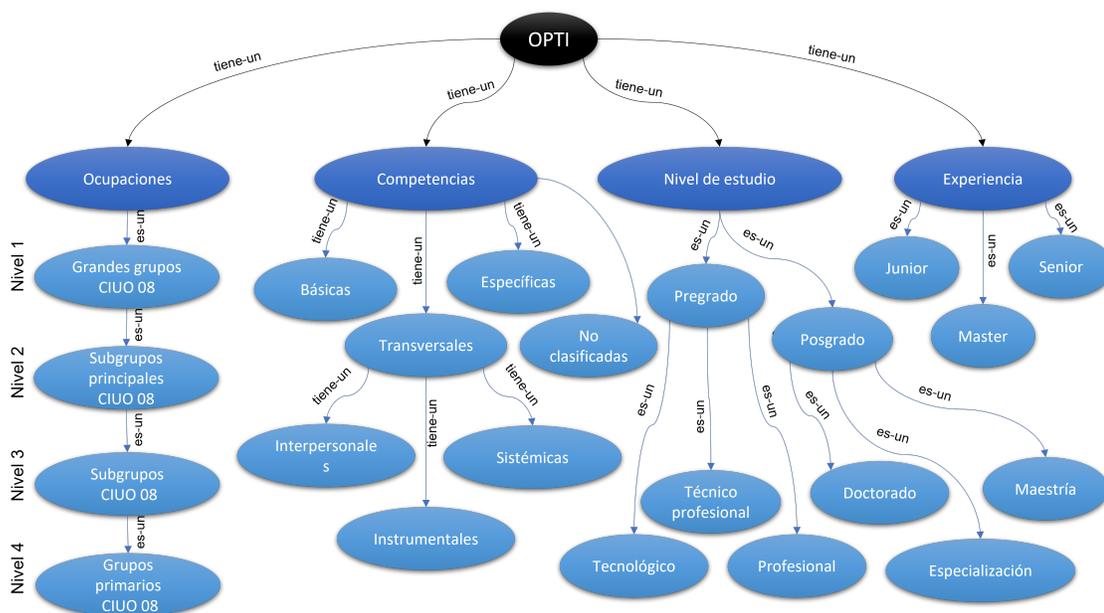


Figura 3-13.: Relación taxonómica entre clases

Las clases definitivas quedan jerarquizadas de la siguiente forma:

■ Ocupaciones

● Directores y gerentes

- Directores ejecutivos personal directivo de la administración pública y miembros del poder ejecutivo y de los cuerpos legislativos

- ◇ Directores generales y gerentes generales
- Directores administradores y comerciales
 - ◇ Directores de administración y servicios
 - *Directores financieros
 - *Directores de recursos humanos
 - *Directores de políticas y planificación
 - *Directores de administración y servicios no clasificados bajo otros epígrafes
 - ◇ Directores de ventas comercialización y desarrollo
 - *Directores de ventas y comercialización
 - *Directores de publicidad y relaciones públicas
 - *Directores de investigación y desarrollo
- Directores y gerentes de producción y operaciones
 - ◇ Directores de producción agropecuaria silvicultura y pesca
 - *Directores de producción agropecuaria y silvicultura
 - *Directores de producción de piscicultura y pesca
 - ◇ Directores de industrias manufactureras de minería construcción y distribución
 - *Directores de industrias manufactureras
 - *Directores de explotaciones de minería
 - *Directores de empresas de construcción
 - *Directores de empresas de abastecimiento, distribución y afines
 - ◇ Directores de servicios de tecnología de la información y las comunicaciones
 - *Directores de servicios de tecnología de la información y las comunicaciones
 - ◇ Directores y gerentes de servicios profesionales
 - *Directores de servicios de cuidados infantiles
 - *Directores de servicios de salud
 - *Directores de servicios de cuidado de las personas de edad
 - *Directores de servicios de bienestar social
 - *Directores de servicios de educación
 - *Gerentes de sucursales de bancos, de servicios financieros y de seguros *Directores y gerentes de servicios profesionales no clasificados bajo otros epígrafes
- Gerentes de hoteles restaurantes comercios y otros servicios

- ◇ Gerentes de hoteles y restaurantes
 - *Gerentes de hoteles
 - *Gerentes de restaurantes
- ◇ Gerentes de comercios al por mayor y al por menor
- ◇ Otros gerentes de servicios
 - *Gerentes de servicios no clasificados bajo otros epígrafes
- Profesionales científicos e intelectuales
 - Profesionales de las ciencias y de la ingeniería
 - ◇ Físicos químicos y afines
 - *Físicos y astrónomos
 - *Meteorólogos
 - *Químicos
 - *Geólogos y geofísicos
 - ◇ Matemáticos actuarios y estadísticos
 - *Matemáticos, actuarios y estadísticos
 - ◇ Ingenieros excluyendo electrotecnólogos
 - *Ingenieros industriales y de producción
 - *Ingenieros civiles
 - *Ingenieros medioambientales
 - *Ingenieros mecánicos
 - *Ingenieros químicos
 - *Ingenieros de minas, metalúrgicos y afines
 - *Ingenieros no clasificados bajo otros epígrafes
 - ◇ Ingenieros en electrotecnología
 - *Ingenieros electricistas
 - *Ingenieros electrónicos
 - *Ingenieros en telecomunicaciones
 - ◇ Arquitectos urbanistas agrimensores y diseñadores
 - *Arquitectos
 - *Arquitectos paisajistas
 - *Urbanistas e ingenieros de tránsito
 - *Cartógrafos y agrimensores
 - Profesionales de la salud
 - ◇ Otros profesionales de la salud
 - *Farmacéuticos
 - Especialistas en organización de la administración pública y de empresas

- ◇ Especialistas en finanzas
 - *Contables
 - *Asesores financieros y en inversiones
 - *Analistas financieros
- ◇ Especialistas en organización de administración
 - *Analistas de gestión y organización
 - *Especialistas en políticas de administración
 - *Especialistas en políticas y servicios de personal y afines
 - *Especialistas en formación del personal
- ◇ Profesionales de las ventas la comercialización y las relaciones públicas
 - *Profesionales de la publicidad y la comercialización
 - *Profesionales de relaciones públicas
 - *Profesionales de ventas técnicas y médicas (excluyendo la TIC)
 - *Profesionales de ventas de tecnología de la información y las comunicaciones
- Profesionales de tecnología de la información y las comunicaciones
 - ◇ Desarrolladores y analistas de software y multimedia
 - *Analistas de sistemas
 - *Desarrolladores de software
 - *Desarrolladores Web y multimedia
 - *Programadores de aplicaciones
 - *Desarrolladores y analistas de software y multimedia y analistas no clasificados bajo otros epígrafes
 - ◇ Especialistas en bases de datos y en redes de computadores
 - *Diseñadores y administradores de bases de datos
 - *Administradores de sistemas *Profesionales en redes de computadores
 - *Especialistas en bases de datos y en redes de computadores no clasificados bajo otros epígrafes
- Profesionales en derecho en ciencias sociales y culturales
 - ◇ Profesionales en derecho
 - *Abogados
 - *Jueces
 - *Profesionales en derecho no clasificados bajo otros epígrafes
 - ◇ Archivistas bibliotecarios curadores y afines
 - *Archivistas y curadores de museos
 - *Bibliotecarios, documentalistas y afines
 - ◇ Artistas creativos e interpretativos
 - *Directores de cine, de teatro y afines

- Técnicos y profesionales de nivel medio
 - Profesionales de las ciencias y la ingeniería de nivel medio
 - ◇ Supervisores en ingeniería de minas de industrias manufactureras y de la construcción
 - *Supervisores de la construcción
 - ◇ Técnicos en control de procesos
 - *Técnicos en control de procesos no clasificados bajo otros epígrafes
 - Profesionales de nivel medio de la salud
 - ◇ Técnicos médicos y farmacéuticos
 - *Técnicos en aparatos de diagnóstico y tratamiento médico
 - Profesionales de nivel medio en operaciones financieras y administrativas
 - ◇ Profesionales de nivel medio en finanzas y matemáticas
 - *Agentes de bolsa, cambio y otros servicios financieros
 - *Oficiales de préstamos y créditos
 - *Tenedores de libros
 - *Profesionales de nivel medio de servicios estadísticos, matemáticos y afines
 - *Tasadores
 - ◇ Agentes comerciales y corredores
 - *Agentes de seguros
 - *Representantes comerciales
 - *Agentes de compras
 - *Agentes de compras y consignatarios
 - ◇ Agentes de servicios comerciales
 - *Declarantes o gestores de aduana
 - *Organizadores de conferencias y eventos
 - *Agentes de empleo y contratistas de mano de obra
 - *Agentes inmobiliarios
 - *Agentes de servicios comerciales no clasificados bajo otros epígrafes
 - ◇ Secretarios administrativos y especializados
 - *Supervisores de secretaría
 - *Secretarios jurídicos
 - *Secretarios administrativos y ejecutivos
 - *Secretarios médicos
 - Profesionales de nivel medio de servicios jurídicos sociales culturales y afines
 - ◇ Profesionales de nivel medio de servicios jurídicos sociales y religiosos
 - *Profesionales de nivel medio del derecho y servicios legales y afines

- ◇ Profesionales de nivel medio en actividades culturales artísticas y culinarias
 - *Fotógrafos
- Técnicos de la tecnología de la información y las comunicaciones
 - ◇ Técnicos en operaciones de tecnología de la información y las comunicaciones y asistencia al usuario
 - *Técnicos en operaciones de tecnología de la información y las comunicaciones
 - *Técnicos en asistencia al usuario de tecnología de la información y las comunicaciones
 - *Técnicos en redes y sistemas de computadores
 - *Técnicos de la Web
 - ◇ Técnicos en telecomunicaciones y radiodifusión
 - *Técnicos de radiodifusión y grabación audio visual
 - *Técnicos de ingeniería de las telecomunicaciones
- Personal de apoyo administrativo
 - Oficinistas
 - ◇ Secretarios general
 - ◇ Operadores de máquinas de oficina
 - *Operadores de máquinas de procesamiento de texto y mecanógrafos
 - *Grabadores de datos
 - Empleados en trato directo con el público
 - ◇ Empleados de servicios de información al cliente
 - *Empleados de agencias de viajes
 - Empleados contables y encargados del registro de materiales
 - ◇ Auxiliares contables y financieros
 - *Empleados de contabilidad y cálculo de costos
 - *Empleados de servicios estadísticos, financieros y de seguros
 - *Empleados encargados de las nóminas
- Oficiales operarios y artesanos de artes mecánicas y de otros oficios
 - Trabajadores especializados en electricidad y la electrotecnología
 - ◇ Instaladores y reparadores de equipos eléctricos
 - *Instaladores y reparadores de líneas eléctricas
 - ◇ Instaladores y reparadores de equipos electrónicos y de telecomunicaciones

*Mecánicos y reparadores en electrónica

*Instaladores y reparadores en tecnología de la información y las comunicaciones

■ **Competencias**

- Básicas
- Transversales
 - Interpersonales
 - Instrumentales
 - Sistémicas
- Específicas
- No clasificadas

■ **Nivel de estudio**

- Técnico
- Tecnólogo
- Profesional
- Especialista
- Magister
- Doctor

■ **Experiencia**

- Junior
- Senior
- Master

De la página web de la ISCO [Organization, 2022] <https://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm> se extrae así toda la información valiosa que contiene sobre la clasificación de ocupaciones: el id de cada ocupación, las ocupaciones, la descripción de la ocupación y las competencias relacionadas a la ocupación (éstas quedan en la super clase competencias como No clasificadas). Es importante resaltar que la ISCO-O8 contiene todas las ocupaciones, pero se realizó un trabajo de seleccionar las ocupaciones y competencias que realmente pertenecen a un perfil profesional en TI, basados en el artículo de la **3-8** y algunos adicionales que se consideraron importante incluir. Se tuvo en cuenta los profesionales de TI bajo las 3 categorías mencionadas en la Tabla **3-7**.

Se definen 4 superclases: Ocupaciones, Competencias, Experiencia y Nivel de estudio que pretenden abarcar lo necesario para evaluar de forma consistente un perfil profesional para una vacante de trabajo en países hispanohablantes.

Propiedades del objeto y Propiedades de los datos

Las propiedades del objeto (*Object properties*) admiten la representación de la relación entre entidades de dos clases diferentes y las propiedades de los datos (*Data properties*) ayudan a definir la relación entre los objetos y un valor determinado, posibilitando la caracterización de variables. Cada diseño de ontología describe sus propias propiedades de objeto y de datos.

Se define para la ontología OPTI las propiedades del objeto basados en la Figura 3-14.

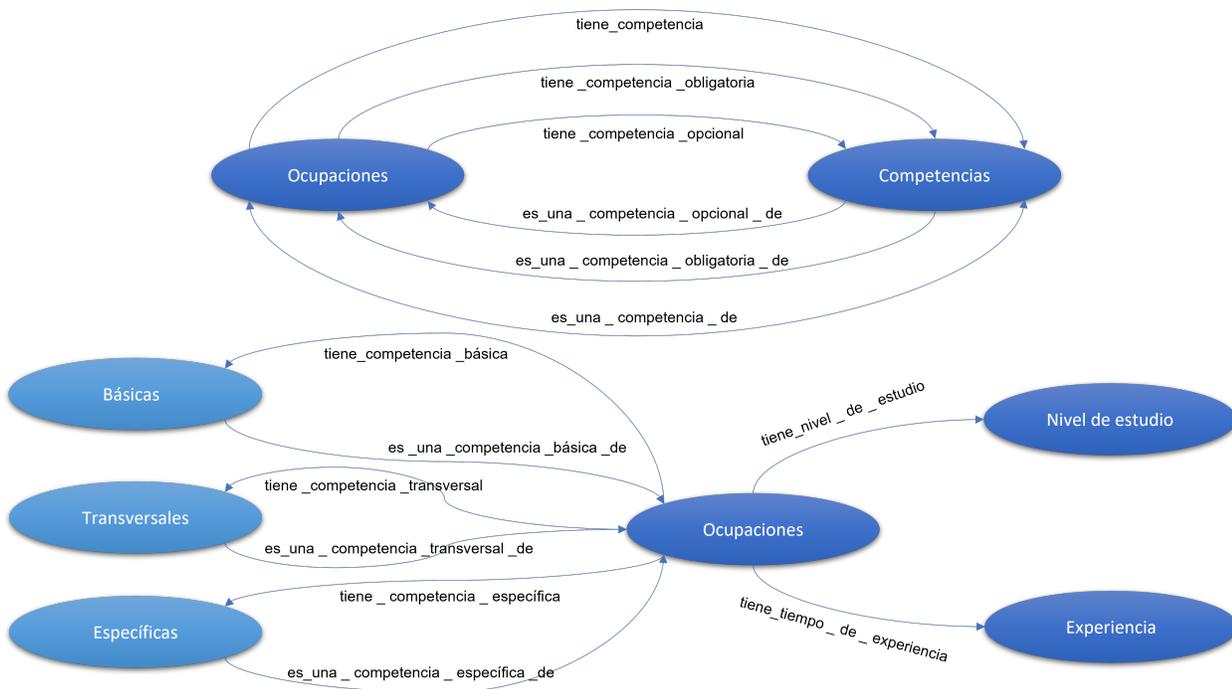


Figura 3-14.: Relaciones entre conceptos

Y las propiedades de los datos basados en la Figura 3-15.

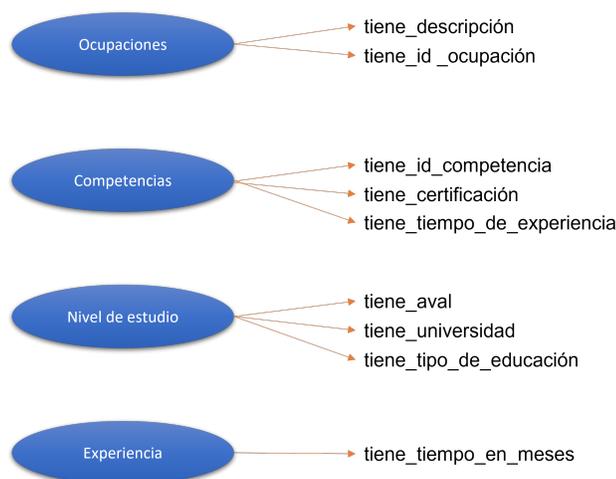


Figura 3-15.: Relaciones entre conceptos y tipos de datos

Las relaciones se diseñaron pensando en una ontología básica lista para un proceso de aprendizaje ontológico y que pudiera relacionar los elementos asociados a las 4 superclases propuestas. Estas propiedades se proponen de la manera más adecuada posible para proporcionar un conocimiento suficiente a la ontología y dar más sentido para ser implementada en sistemas de *e-recruitment*. La Tabla 3-6 proporciona las definiciones de propiedades de objeto y la Tabla 3-7 proporciona las definiciones de propiedades de datos.

Propiedad de objeto	Dominio	Rango	F	S	FI	T	AS	R	IR	IV
tiene_competencia	Ocupaciones	Competencias					X		X	es_una_competencia_de
es_una_competencia_de	Competencias	Ocupaciones					X		X	tiene_competencia
tiene_competencia_obligatoria	Ocupaciones	Competencias					X		X	es_una_competencia_obligatoria
es_una_competencia_obligatoria_de	Competencias	Ocupaciones					X		X	tiene_competencia_obligatoria
tiene_competencia_opcional	Ocupaciones	Competencias					X		X	es_una_competencia_opcional_de
es_una_competencia_opcional_de	Competencias	Ocupaciones					X		X	tiene_competencia_opcional
tiene_nivel_de_estudio	Ocupaciones	Nivel de estudio					X		X	
tiene_tiempo_de_experiencia	Ocupaciones	Experiencia	X						X	
tiene_competencia_básica	Ocupaciones	Básicas					X		X	es_una_competencia_básica_de
es_una_competencia_básica_de	Básicas	Ocupaciones					X		X	tiene_competencia_básica
tiene_competencia_transversal	Ocupaciones	Transversales					X		X	es_una_competencia_transversal_de
es_una_competencia_transversal_de	Transversales	Ocupaciones					X		X	tiene_competencia_transversal
tiene_competencia_específica	Ocupaciones	Competencias					X		X	es_una_competencia_específica_de
es_una_competencia_específica_de	Ocupaciones	Competencias					X		X	tiene_competencia_específica

Tabla 3-6.: Definición de propiedades del objeto y sus características. F: Funcional, S: Simétrica, FI: Funcional Inversa, T: Transitiva, AS: Asimétrica, R: Reflexiva, IR: Irreflexiva, IV: Inversa

Propiedad del dato	Dominio	Rango	F
tiene_descripción	Ocupaciones	xsd:string	X
tiene_id_ocupación	Ocupaciones	xsd:integer	X
tiene_id_competencia	Competencias	xsd:integer	X
tiene_certificación	Competencias	xsd:string	
tiene_tiempo_de_experiencia	Competencias	xsd:integer	X
tiene_aval	Nivel de estudio	xsd:Boolean	
tiene_universidad	Nivel de estudio	xsd:string	
tiene_tipo_de_educación	Nivel de estudio	xsd:string	
tiene_tiempo_en_meses	Experiencia	xsd:integer	X

Tabla 3-7.: Definición de propiedades de los datos y sus características. F: Funcional

Definición de instancias

Ahora bien, es necesario comenzar a definir algunos individuos de las clases, los cuales representan instancias que son los conceptos más específicos representados en una ontología. Se proponen instancias para las diferentes clasificaciones de Experiencia y de competencias. Es fundamental tener en cuenta que si estos conceptos específicos definidos aquí, llegaran a formar una jerarquía natural, podríamos representarlos como clases más adelante.

A continuación, se listan las instancias definidas en las clases seleccionadas.

SUPERCLASE TRANSVERSALES

Instrumentales:

- Capacidad de análisis y síntesis
- Capacidad de organización y planificación
- Comunicación oral y escrita
- Conocimiento de una lengua extranjera
- Conocimientos de informática relativos al ámbito de estudio
- Capacidad de gestión de la información
- Resolución de problemas
- Toma de decisiones

Interpersonales:

- Trabajo en equipo
- Trabajo en un equipo de carácter interdisciplinario
- Trabajo en un contexto internacional
- Habilidades en las relaciones interpersonales
- Reconocimiento a la diversidad y la multiculturalidad
- Razonamiento crítico

- Compromiso ético

Sistémicas:

- Aprendizaje autónomo
- Adaptación a nuevas situaciones
- Creatividad
- Liderazgo
- Conocimiento de otras culturas y costumbres
- Iniciativa y espíritu emprendedor
- Motivación por la calidad
- Sensibilidad hacia temas medioambientales

No clasificadas: lista de competencias de cada una de la jerarquía de ocupaciones tomada de la CIUO-08

SUPERCLASE EXPERIENCIA**Junior:**

- 1 año
- 2 años

Senior:

- 3 años
- 4 años

Master:

- 5 años
- 6 años
- 7 años
- 8 años
- 9 años
- 10 años

Construcción de la ontología

Luego de mostrar el diseño de la ontología se procede a su construcción con una herramienta de software. En este caso, podría utilizarse un software editor de ontologías como Protégé para construir manualmente la ontología diseñada. Sin embargo, teniendo en cuenta en que el proceso de aprendizaje ontológico busca automatizar la construcción de las ontologías en la medida que se vaya aprendiendo, se utiliza desde este momento los algoritmos de la librería para Python Owlready2 (<https://owlready2.readthedocs.io/en/v0.37/>). Esta librería permite crear una ontología a través de objetos en Python, además permite modificar, guardar, realizar consultas y realizar razonamiento en la ontología. El manejo de esta librería facilitará en un futuro la adición y modificación de la ontología respecto a los aprendizajes

obtenidos.

Así entonces, veremos a continuación como se realiza la construcción de la ontología base diseñada con Owlready2 y se mostrarán algunas imágenes de como se ve la ontología construida usando Protégé.

En principio se construye un archivo .csv que contiene todas las clases, el id de clase que determina la jerarquía y las instancias que las componen. De allí, se importa este archivo para por medio de Python realizar una limpieza del texto de las ocupaciones, donde se le extraen las comas, los paréntesis y se sustituye el carácter espacio por guiones, esto con el fin de construir una ontología limpia. Cabe resaltar que las tildes se dejan en el texto. Se dispone de la información del archivo .csv construyendo un diccionario donde la clave es el id y el valor es la ocupación. Y un diccionario donde la clave es el id y el valor es la definición de la ocupación.

Luego, se realiza un proceso de limpieza para las ocupaciones, donde se eliminan los caracteres espaciales que no son necesarios y también se dividen las competencias de cada ocupación. Finalmente, se crea un diccionario donde la clave es el id y el valor son las competencias.

Para construir la ontología se realiza un código en Python usando la librería Owlready2 que permite la definición de IRI, la creación de clases y sus jerarquías, la creación de propiedades de objeto y de datos con sus características y finalmente, guardar la ontología en un archivo OWL. En el Anexo B se puede ver una extracción del código implementado para la construcción de la ontología base con ésta librería. Este archivo OWL se abre en Protégé para verificar que todo haya quedado consistente.

Se define el IRI, <http://OPTI2023-04-1305:30:49.882527.org/onto>. Después se le agregan algunos comentarios generales. En la Figura 3-16 se puede ver el encabezado de la ontología.

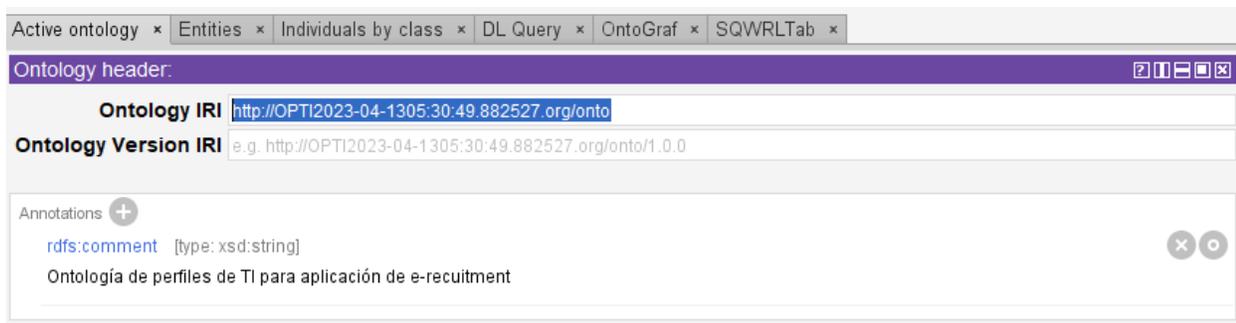


Figura 3-16.: IRI de la ontología base, extraído de Protégé

En la Figura 3-17 se pueden ver las clases y jerarquía de clases definidas para la ontología base.

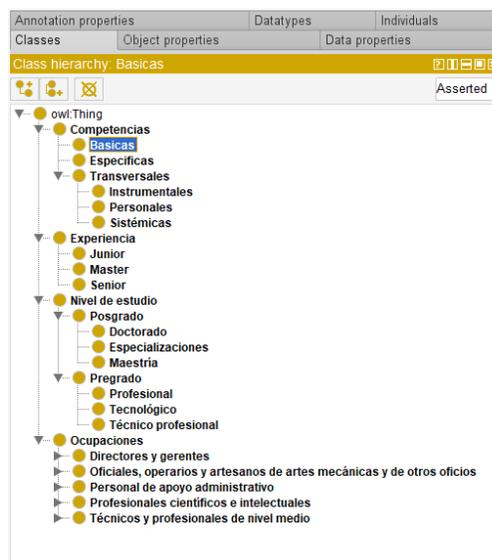


Figura 3-17.: Clases y jerarquía de clase de la ontología base, extraído de Protégé

En la Figura 3-18) se muestra la propiedad de objeto inicial definida para la ontología base con la definición de sus características, su dominio y su rango. También se definen dos propiedades de datos, id y descripción con su dominio y rango preestablecido (ver Figura 3-19).

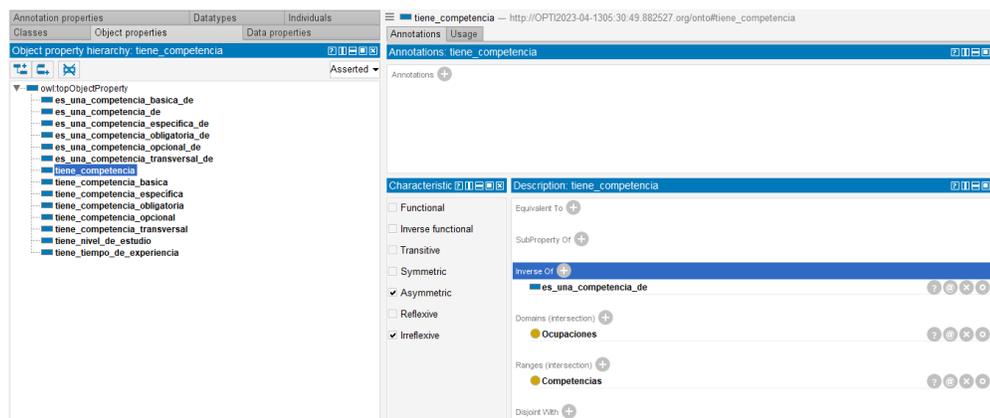


Figura 3-18.: Propiedades de objeto de la ontología base, extraído de Protégé

Ahora, en la Figura 3-20 se muestran las instancias definidas para el último nivel de jerarquía de las clases.

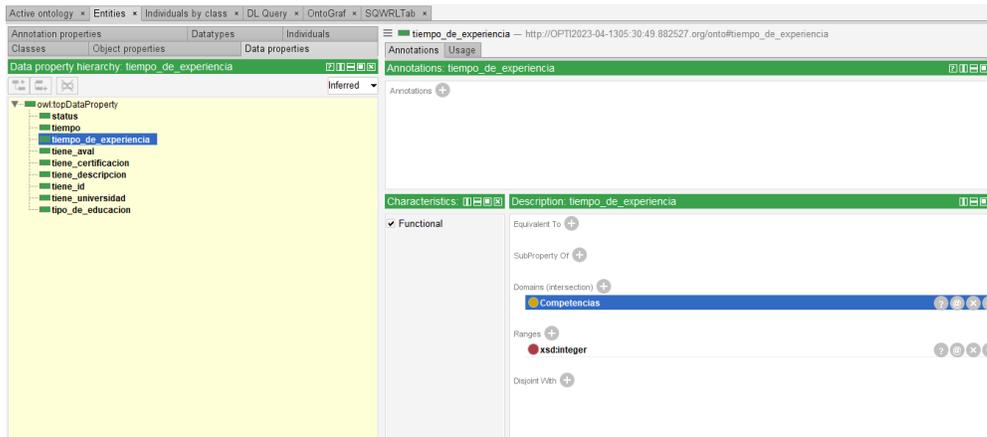


Figura 3-19.: Propiedades de los datos de la ontología base, extraído de Protégé

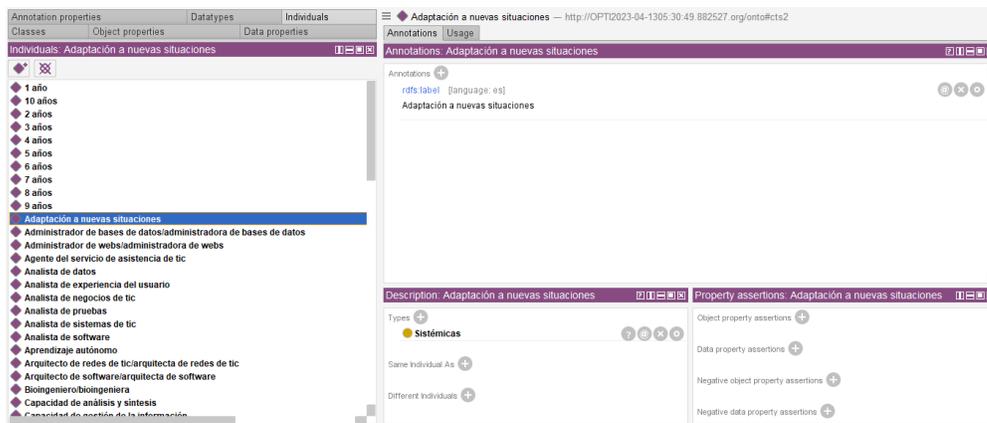


Figura 3-20.: Instancias definidas para la ontología base

Así entonces, la ontología teniendo una construcción semántica y relación entre sus clases. En la Figura 3-21 se puede ver una vista puntual de la ocupación **Analistas de sistemas**, su jerarquía y sus instancias relacionadas.

En la Figura 3-22 se muestra una representación gráfica de la ontología base extraída con la herramienta OntoGraf del software Protégé.

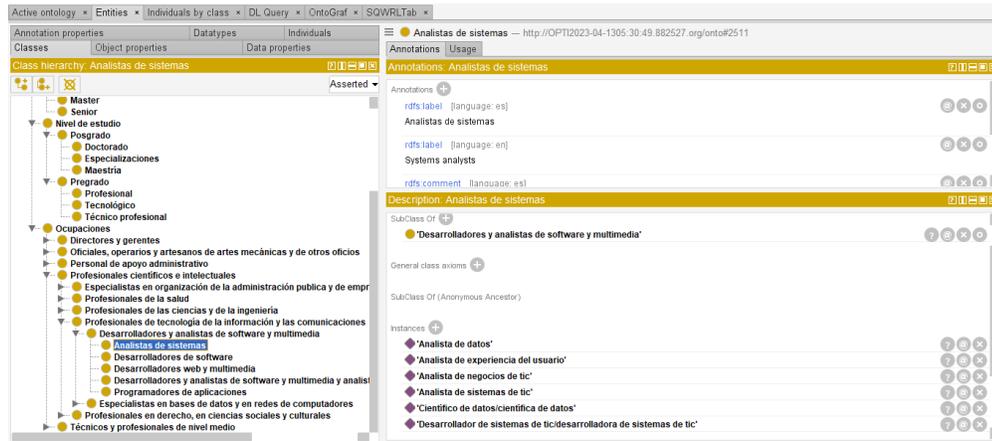


Figura 3-21.: Instancias de clase de la ontología base

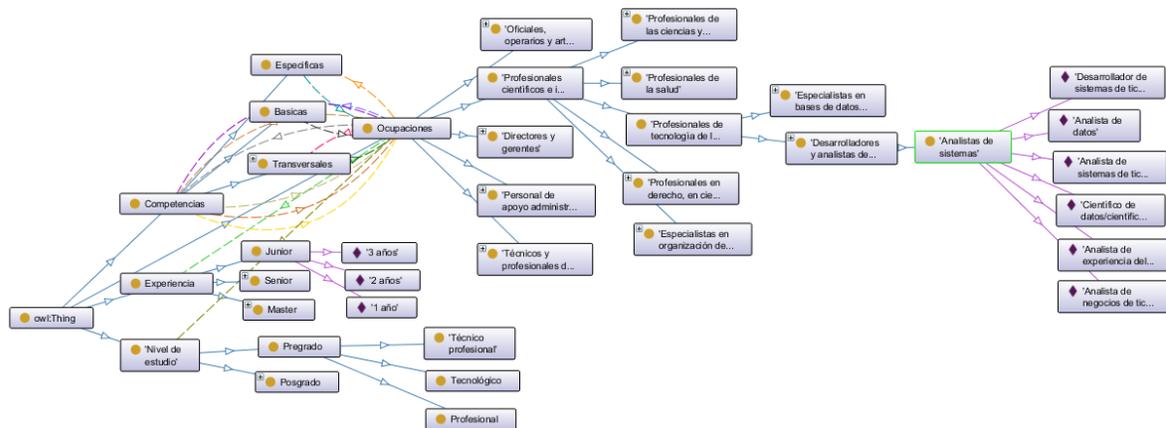


Figura 3-22.: Representación gráfica de la ontología base OPTI

3.5. Análisis y conclusiones del capítulo

El objetivo de este capítulo es definir una ontología base para el dominio de *e-recruitment* que comprenda los perfiles profesionales de TI. Para ello, se toma como inspiración la estandarización internacional encontrada sobre la clasificación de ocupaciones (ISCO-08), la clasificación de competencias dadas por el Ministerio de Educación Colombiano y finalmente, los conocimientos adquiridos sobre la necesidad del *e-recruitment* y la contratación de personal en general en países hispanohablantes.

Se toman las siguientes decisiones para definir la ontología base:

- La ontología base tendrá 4 superclases: Ocupaciones, Competencias, Experiencia y Nivel de estudio suficientes para modelar el conocimiento en un sistema básico de reclutamiento para cualquier profesión. Se parte de la premisa de que un perfil debe estar sintetizado en las **cualidades** más relevantes para el **puesto de trabajo** al que se aspira, las cualidades se condensan en competencias y el puesto de trabajo reposa en una ocupación como tal. Se toma en cuenta la propuesta para definir una competencia (Figura 3-10), siendo este término el que abarca los conocimientos, las habilidades y las cualidades personales (el cual se investiga en un examen psicológico en un proceso de contratación). El Nivel de estudio y la Experiencia, es una propuesta propia teniendo en cuenta las características más importantes que tienen en cuenta los reclutadores al examinar la hoja de vida de un perfil profesional.
- Las ocupaciones que se seleccionan para la ontología base se definen tomando como referencia la clasificación de ISCO-08. El trabajo aquí fue proponer cuales de las ocupaciones clasificadas es este estándar internacional pueden entenderse desde el punto de vista de los perfiles profesionales de TI y para ello, se tiene en cuenta la propuesta de la Figura 3-8 [Arnal et al., 2001], entender que un profesional puede tener diferentes categorías de competencias de TI (Especializadas, avanzadas o básicas) como se muestra en la Tabla 3-4. En total se extrajeron 181 clases de ocupación con su respectiva definición y competencias.
- La jerarquía de clases de ocupación se tomó directamente de la clasificación de la ISCO-08, que es precisamente un aporte importante de la utilización de normativa y estandarización en el dominio en el que se trabaja. Si se quiere lograr un aprendizaje ontológico deberá ser a partir de algo reutilizado o previamente preestablecido con el fin de realmente obtener aportes importantes. Por su lado, la ontología ESCO muestra que su aporte está en su propuesta después del último nivel de la clasificación ISCO-08 como se muestra en la Figura 3-4.
- Se definen 14 propiedades de objeto que relaciona la clase Ocupaciones con la clase Competencias, Nivel de estudio y Experiencia. Se declaran con características asimétri-

ca e irreflexiva. Solo hay una con característica funcional y se proporcionan las propiedades inversas en cada caso.

- Finalmente, se proponen 9 propiedades de los datos para cada una de las superclases y se define para cada una el dominio y rango. Algunas propiedades del dato definidas se declaran con característica funcional.
- El código en Python usando la librería Owlready2 implementado se usó como mecanismo de desarrollo para acercarse a la automatización de ontologías de manera programática. Incluir las nuevas entidades aprendidas en la ontología no debe ser un proceso manual y tedioso, al contrario, debe tender a la automatización como lo propone la literatura. Se pudo haber utilizado Protégé para hacerlo manual, pero ésta sólo se usó como mecanismo de visualización.

En este capítulo se cumplió el objetivo 1 de la tesis: Definir los conceptos básicos o la estructura ontológica para el dominio de *e-recruitment* asociado a perfiles profesionales en TI que sean relevantes para la investigación a partir de las fuentes de información disponibles, con el fin de definir la ontología base del modelo.

4. Capítulo 4: Diseño de un modelo conceptual de técnicas y algoritmos para el aprendizaje ontológico

En este capítulo se diseña un modelo conceptual de aprendizaje ontológico para la ontología base propuesta. Esto se traduce en la creación de un portafolio de técnicas y algoritmos que se aplicarán para el aprendizaje de clases, jerarquía de clases y las restricciones de la ontología. Así mismo, se justifica la red social seleccionada para este desarrollo particular y se describe la extracción de información de esta. Lo anterior, servirá como base para ejecutar los algoritmos y mostrar los resultados de aprendizaje en el capítulo siguiente.

El diseño conceptual del portafolio de algoritmos y técnicas a aplicar se realiza con base en el rastreo bibliográfico de artículos que realizan Aprendizaje de Máquinas y Aprendizaje Profundo. Se intenta seleccionar y ajustar los algoritmos y técnicas (o combinación de estas) idóneas, teniendo cuenta la fuente de extracción, la ontología base, el dominio y la aplicación final para la que se usará la ontología (*e-recruitment*).

En la Figura 4-1 se representa de forma gráfica lo que se pretende en este capítulo. A partir del conocimiento de técnicas y algoritmos usados en la literatura para el aprendizaje ontológico, se seleccionan estratégicamente algunas de ellas para conformar el modelo conceptual o portafolio de aprendizaje ontológico que se aplicará en el desarrollo de este trabajo. El modelo conceptual de aprendizaje ontológico tendrá como entrada la información obtenida de la fuente de extracción de datos, que en este caso será una red social profesional seleccionada y la ontología base construida en el capítulo anterior. Se desea diseñar un modelo conceptual que permita el aprendizaje de clases, jerarquía de clases y restricciones en la ontología base.

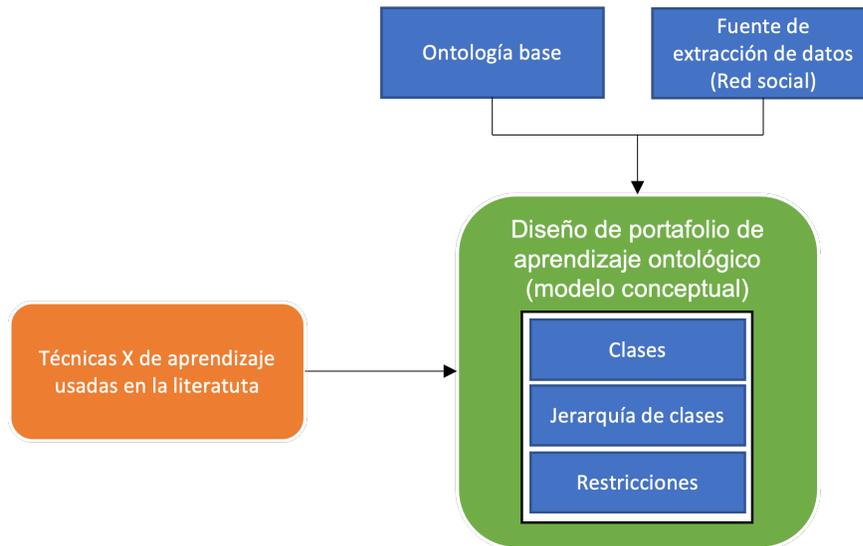


Figura 4-1.: Representación gráfica del objetivo del capítulo 4

Teniendo en cuenta lo anterior, en la Figura 4-2 se muestra la metodología para el desarrollo de este capítulo. Se define la red social profesional que se va a usar para extraer la información y además se define como se realizará la extracción de dicha información. Finalmente, se presenta un modelo conceptual o portafolio de técnicas de aprendizaje ontológico que se implementarán en el siguiente capítulo para aprender clases, jerarquía de clase y restricciones.

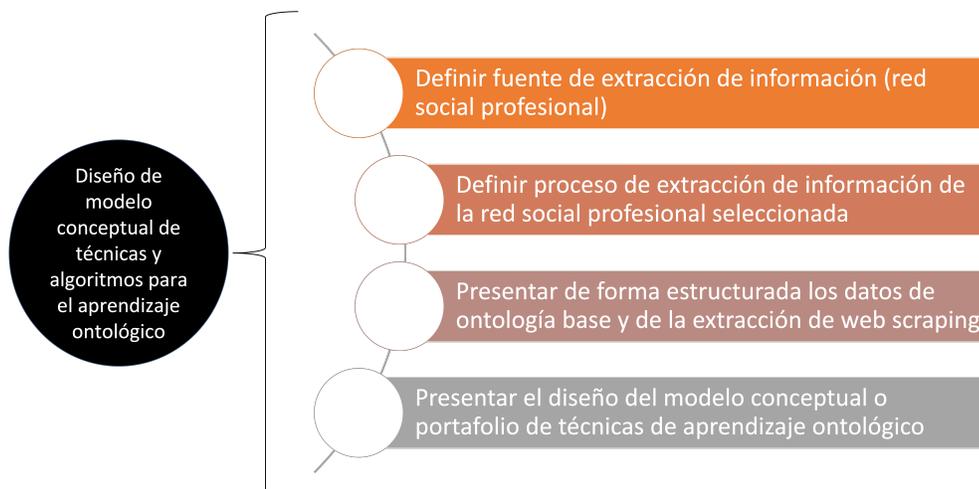


Figura 4-2.: Metodología de desarrollo objetivo 2

Para el marco de este trabajo, se toma como referencia la torta de capas de aprendizaje de ontologías para definir las tareas de aprendizaje de clases, jerarquía de clases y restricciones. (ver Figura 4-3).

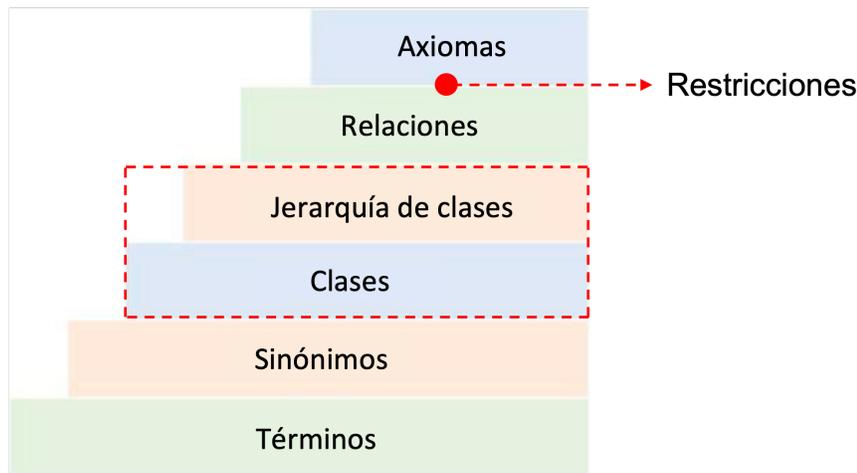


Figura 4-3.: Definición de tareas de aprendizaje con base en la torta de capas de aprendizaje de una ontología, adaptado de [Asim et al., 2018b]

- **Aprendizaje de clases:** Se refiere al proceso de identificar y definir las clases o conceptos en una ontología a partir de datos existentes o de conocimiento previo. Cuando se aprende una clase en una ontología, significa que se ha adquirido conocimiento sobre la existencia y las características de esa clase.

- **Aprendizaje de jerarquía de clases:** Implica establecer las relaciones jerárquicas entre las clases en la ontología. Esto incluye determinar qué clases son subclases de otras y cómo se organizan en una estructura taxonómica.

Si se aprende una clase que es una subclase de otra, también se ha aprendido implícitamente sobre la superclase a la que pertenece. Las subclases heredan las propiedades y características de sus superclases, lo que significa que las instancias que pertenecen a la subclase también comparten las propiedades y características de la superclase.

Se puede decir que cuando se aprende una clase en una ontología, especialmente si es una subclase de otra, también se está aprendiendo sobre la jerarquía de clases. Entonces, si se aprende una clase que es una subclase de otra, eso implica que se entiende cómo esa clase se relaciona con la superclase y se extrae conocimiento sobre la relación de especialización entre ambas clases [Revenko et al., 2022].

- **Aprendizaje de restricciones:** En general, las restricciones son condiciones o reglas que se aplican a las clases y propiedades para definir relaciones, características y comportamientos específicos de las entidades que pertenecen a esas clases o poseen esas propiedades. En este trabajo tomaremos en cuenta solo las restricciones sobre las propiedades, específicamente,

restricciones de dominio y rango. La restricción de dominio especifica que una propiedad solo puede ser aplicada a instancias de una clase o un conjunto de clases específicas (limita qué tipos de sujetos pueden tener esa propiedad) y la restricción de rango especifica que una propiedad solo puede tener instancias de una clase o un conjunto de clases específicas como objetos (limita qué tipos de objetos pueden estar relacionados con la propiedad).

4.1. Preparación de los datos

Esta etapa de diseño es crítica y fundamental en el flujo de trabajo de analítica de datos y aprendizaje automático y profundo, ya que la calidad de los resultados finales depende en gran medida de la calidad de los datos con los que se trabaja.

La preparación de datos implica varias tareas que permiten transformar, limpiar y estructurar los datos de manera adecuada para que puedan ser analizados y utilizados en modelos de aprendizaje automático u otras técnicas de análisis.

4.1.1. Descripción de los datos de la ontología base

Los primeros datos de los cuales se va a alimentar el modelo de aprendizaje será la ontología base que fue creada en el capítulo anterior. En este punto de diseño del Modelo conceptual de aprendizaje es importante rescatar que dicha ontología queda almacenada en una base de datos la cual contiene todos los elementos definidos para la ontología base: clases y subclases, definición de las clases y subclases, propiedades de los objetos, propiedades de los datos, características de las propiedades y código de cada una de las clases y subclases. Esta base de datos es un archivo adicional del OWL de la ontología.

En la Tabla 4-1, Tabla 4-2 y Tabla 4-3 hay una vista de los labels que describen las características o campos de la ontología base.

Estos campos serán almacenados en un espacio de datos crudos con acceso centralizado a los datos, el cual permite almacenar información de diferentes fuentes y formatos sin tener que estructurarlos previamente, para luego poder ser analizados y explorados.

4.1.2. Descripción de los datos del Web Scraping

Desde el planteamiento de los objetivos de esta tesis se considera como fuente de extracción de información las redes sociales profesionales que permitan encontrar información acerca de los perfiles y empleos ofertados, en este caso, asociados a las tecnologías de la información (TI). La extracción de información se diseña pensando en encontrar nuevos perfiles de TI de cara a las necesidades de las empresas, es decir, bajo el interés de contrastar lo que existe

Superclase	Tipo	Código	Título	Definición	Competencias
ocupaciones	clase	2	Profesionales de tecnología de la información y las comunicaciones	"Los profesionales de las tecnologías de la información y la comunicación investigan, planifican, diseñan, escriben, prueban, asesoran y mejoran sistemas de tecnologías de la información, hardware, software y conceptos relacionados para aplicaciones específicas..."	investigar el uso de las tecnologías de la información en las funciones empresariales; identificar áreas de mejora e investigar los aspectos teóricos y los métodos operativos para el uso de ordenadores; evaluar, planificar y diseñar configuraciones de hardware o software para aplicaciones específicas, incluyendo para Internet, Intranet y sistemas multimedia..."
ocupaciones	clase	251	Desarrolladores y analistas de software y multimedia	"Los desarrolladores y analistas de software y aplicaciones investigan, planifican, diseñan, escriben, prueban, asesoran y mejoran sistemas informáticos como hardware, software y otras aplicaciones para satisfacer requisitos específicos..."	^{a)} consultar con los usuarios para formular y documentar los requisitos y con la dirección para garantizar el acuerdo sobre los principios de los sistemas; b) identificar y analizar los procesos empresariales, los procedimientos y las prácticas de trabajo; c) identificar y evaluar las ineficiencias y recomendar prácticas empresariales y funcionalidades y comportamientos óptimos de los sistemas;...
ocupaciones	clase	2511	Analistas de sistemas	"Los analistas de sistemas investigan, analizan y evalúan los requisitos, procedimientos o problemas informáticos de los clientes, y elaboran y aplican propuestas, recomendaciones y planes para mejorar los sistemas de información actuales o futuros."	
ocupaciones	clase	2512	Desarrolladores de software	"Los desarrolladores de software investigan, analizan y evalúan los requisitos de aplicaciones de software y sistemas operativos nuevos o existentes, y diseñan, desarrollan, prueban y mantienen soluciones de software para satisfacer dichos requisitos."	^{a)} investigar, analizar y evaluar requisitos para aplicaciones de software y sistemas operativos; b) investigar, diseñar y desarrollar sistemas de software informático; c) consultar con personal de ingeniería para evaluar interfaces entre hardware y software;..."
ocupaciones	clase	2513	Desarrolladores Web y multimedia	"Los profesionales del desarrollo web y multimedia combinan conocimientos técnicos y de diseño para investigar, analizar, evaluar, diseñar, programar y modificar sitios web y aplicaciones que combinen texto, gráficos, animaciones, imágenes, audio y vídeo y otros medios interactivos."	^{a)} analizar, diseñar y desarrollar sitios de Internet aplicando una mezcla de arte y creatividad con lenguajes de programación de software y scripting e interfaces con entornos operativos; b) diseñar y desarrollar animaciones digitales, imágenes, presentaciones, juegos, clips de audio y vídeo y aplicaciones de Internet utilizando software multimedia, herramientas y utilidades; gráficos interactivos y lenguajes de programación;..."
ocupaciones	clase	2514	Programadores de aplicaciones	"Los programadores de aplicaciones escriben y mantienen código programable descrito en instrucciones y especificaciones técnicas para aplicaciones de software y sistemas operativos."	"(a) escribir y mantener el código del programa descrito en las instrucciones y especificaciones de acuerdo con las normas de calidad acreditadas; (b) revisar, reparar o ampliar los programas existentes para aumentar la eficiencia operativa o adaptarse a nuevos requisitos;..."
...

Tabla 4-1.: Campos 1 de la ontología base inicial

Superclase	Tipo	Código	Título	Definición
nivel_de_estudio	clase	ne1	Pregrado	Educación superior
nivel_de_estudio	clase	ne11	Técnico Profesional	Nivel écnico Profesional (relativo a programas Técnicos Profesionales).
nivel_de_estudio	clase	ne12	Tecnológico	Nivel Tecnológico (relativo a programas tecnológicos)
nivel_de_estudio	clase	ne13	Profesional	ivel Profesional (relativo a programas profesionales universitarios)
nivel_de_estudio	clase	ne2	Posgrado	TEducación superior
nivel_de_estudio	clase	ne21	Especializaciones	Especializaciones (relativas a programas de Especialización Técnica Profesional, Especialización Tecnológica y Especializaciones Profesionales)
nivel_de_estudio	clase	ne22	Maestría	Maestría
nivel_de_estudio	clase	ne23	Doctorado	Doctorado
experiencia	instancia	e11	1 año	*
experiencia	clase	e2	Senior	*
experiencia	instancia	e21	3 años	*
experiencia	clase	e3	Máster	*
experiencia	instancia	e31	5 años	*
experiencia	instancia	e32	6 años	*
competencias	clase	cb	Básicas	En el contexto internacional reciben distintos nombres tales como: core skills, key competences, basic skills, entre otros. Permiten el ingreso al trabajo o a la educación superior, por ello deben ser identificadas desde ambos sectores...
competencias	clase	ce	Específicas	Son las requeridas para el desempeño de una ocupación en concreto, están relacionadas más con funciones o puestos de trabajo. Aportan al estudiante o al trabajador los conocimientos, actitudes, habilidades y valores propios de cada profesión y actividad laboral...
competencias	clase	ct	Transversales	Son competencias requeridas en un amplio campo de profesiones y ocupaciones y aportan las herramientas requeridas por un trabajador profesional para analizar los problemas, evaluar las estrategias a utilizar y aportar soluciones pertinentes en situaciones nuevas...
competencias	clase	cti	Instrumentales	Reciben su nombre porque brindan las herramientas claves tanto para el aprendizaje como para el desempeño en el mundo del trabajo. Comprenden una serie de habilidades como las cognoscitivas que permiten comprender y procesar ideas ...
competencias	clase	ctp	Interpersonales	Son las que permiten mantener una buena relación social y un adecuado comportamiento ciudadano. Se relacionan con la capacidad de expresar los sentimientos, de hacer planteamientos críticos y de autocrítica, de participar en la vida política y de asumir los deberes y derechos ciudadanos en condiciones éticas...
competencias	clase	cts	Sistémicas	Están relacionadas con la visión de conjunto y la capacidad de gestionar integralmente los procesos organizacionales. Se logran mediante una combinación de comprensión, sensibilidad y conocimientos que permiten identificar las partes de un todo y las relaciones entre las partes que generan la estructura de totalidad...
...

Tabla 4-2.: Campos 2 de la ontología base inicial

Tipo	Código	Propiedad	Dominio	Rango	Funcional	Asimétrica	irreflexiva	Inversa
object_prop	op1	tiene competencia	ocupaciones	competencias		asimétrica	irreflexiva	op2
object_prop	op2	es una competencia de	competencias	ocupaciones		asimétrica	irreflexiva	op1
object_prop	op3	tiene competencia obligatoria	ocupaciones	competencias		asimétrica	irreflexiva	op4
object_prop	op4	es una competencia obligatoria de	competencias	ocupaciones		asimétrica	irreflexiva	op3
object_prop	op5	tiene competencia opcional	ocupaciones	competencias		asimétrica	irreflexiva	op6
object_prop	op6	es una competencia opcional de	competencias	ocupaciones		asimétrica	irreflexiva	op5
object_prop	op7	tiene nivel de estudio	ocupaciones	competencias	funcional		irreflexiva	
object_prop	op8	tiene tiempo de experiencia	ocupaciones	nivel_de_estudio		asimétrica	irreflexiva	
object_prop	op9	tiene competencia basica	ocupaciones	cb		asimétrica	irreflexiva	op10
object_prop	op10	es una competencia basica de	cb	ocupaciones		asimétrica	irreflexiva	op9
object_prop	op11	tiene competencia transversal	ocupaciones	ct		asimétrica	irreflexiva	op12
object_prop	op12	es una competencia transversal de	ct	ocupaciones		asimétrica	irreflexiva	op11
object_prop	op13	tiene competencia especifica	ocupaciones	competencias		asimétrica	irreflexiva	op14
object_prop	op14	es una competencia especifica de	ce	competencias		asimétrica	irreflexiva	op13
data_prop	dp1	tiene descripción	ocupaciones	str	funcional			
data_prop	dp2	tiene id	ocupación	ocupaciones	int	funcional		
data_prop	dp3	tiene id	competencia	competencias	int	funcional		
data_prop	dp4	tiene certificación	competencias	str				
...

Tabla 4-3.: Campos 3 de la ontología base inicial

en los perfiles profesionales reales académicos, con la demanda actual de las empresas en esta área profesional que está en constante construcción por la creciente evolución de la tecnología.

Definición de los campos requeridos en el Web Scraping

Ahora bien, se busca dentro de las redes sociales profesionales las ofertas de empleo orientadas a las profesiones de TI de las empresas, para extraer luego su información. En este caso, es indispensable comprender desde la literatura cuales son los campos requeridos para que una empresa genere una oferta laboral para un proceso de reclutamiento online [Brenčić and Norris, 2010], [Arendt et al., 2023], [Pouliakas, 2021], [Van der Velde, 2020] y [ul haq Dar and Dorn, 2018].

En la Figura 4-4 se muestra de forma gráfica una recopilación de los campos generales que debe tener una oferta laboral online y así mismo, una pequeña descripción de lo que contiene cada campo. Estos campos se analizarán con el fin de identificar los campos que deben ser buscados en las ofertas laborales online de una red social profesional.

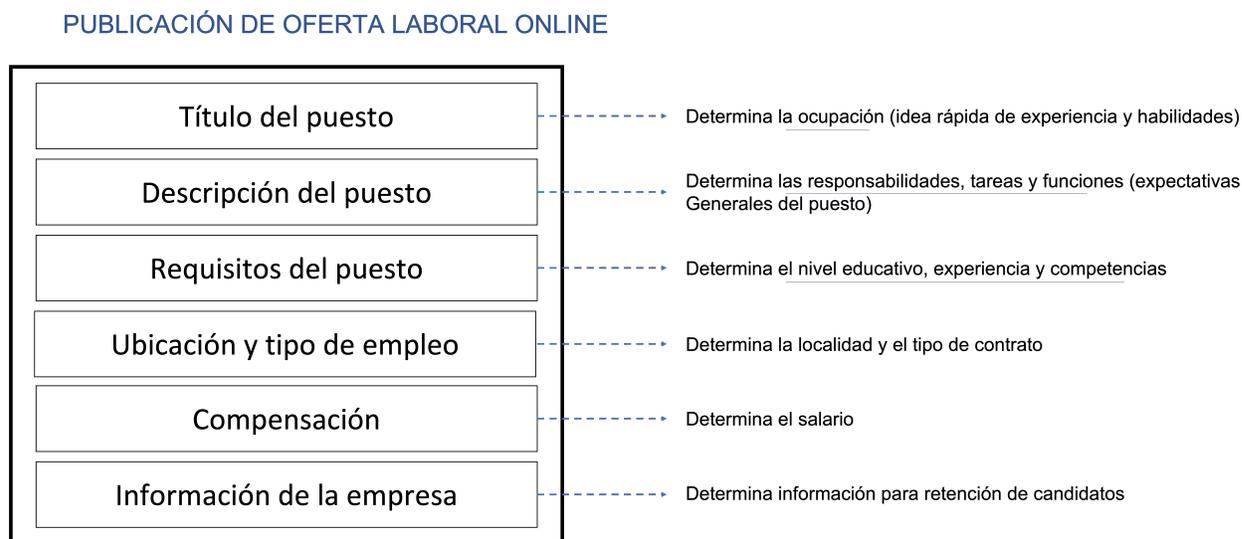


Figura 4-4.: Campos generales de una oferta laboral online

Se puede observar que los campos **Título del puesto**, **Descripción del puesto** y **Requisitos del puesto** puede contener información valiosa que se puede utilizar para alimentar las 4 superclases que se definieron en la ontología base: Ocupaciones, Competencias, Experiencia y Nivel de estudio. Por dicha razón se seleccionan estos tres (3) campos generales como los campos requeridos para ser extraídos de una red profesional seleccionada. En la Figura 4-5 se observa la selección de los campos requeridos para este modelo.

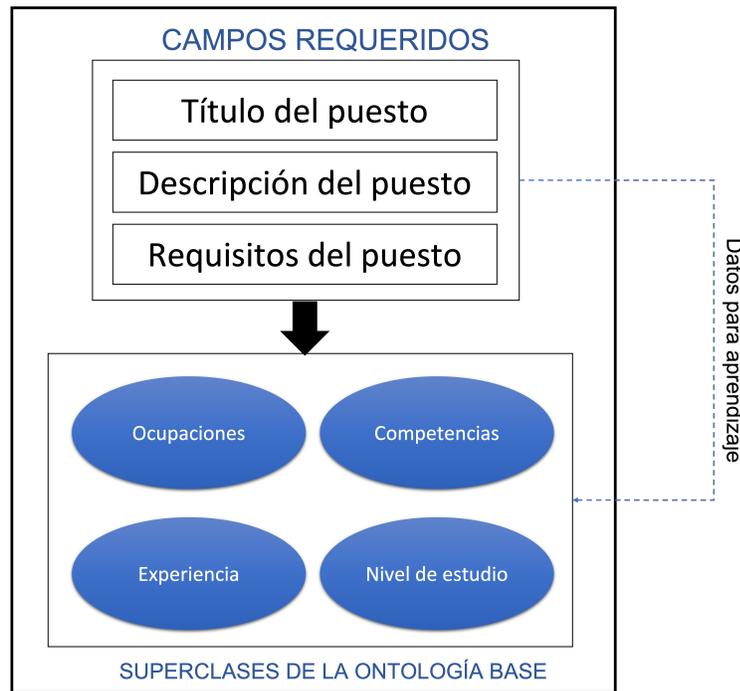


Figura 4-5.: Campos requeridos para extraer de la red social profesional y utilizarlos para el aprendizaje de la ontología final

El campo `textbfTítulo del puesto` contiene las ocupaciones, la **Descripción del puesto** contiene las responsabilidades, tareas y funciones que aunque no es compatible exactamente con una superclase de la ontología base puede llegar a contener información que al ser procesada genere datos importantes de competencias o la creación de nuevas superclases en la ontología y finalmente, los **Requisitos del puesto** contienen información de la experiencia, el nivel educativo y las competencias.

Selección y justificación de la red social que servirá como caso de uso en este trabajo

Las redes profesionales no solo van enfocadas a aumentar tus contactos personales, sino que lo hacen desde un punto de vista profesional y laboral. Las redes profesionales sirven para encontrar empleo, comunicarse con empresas y personas, y aumentar la red de contactos en el ámbito profesional (*networking*). Generalmente, se puede crear el perfil profesional, destacando las aptitudes, las ventajas competitivas y logros significativos en el campo laboral. Así mismo, si una empresa tiene una vacante laboral podría publicar sus requerimientos para así ser identificados como empleador y atraer recurso humano capacitado para su oferta [LinkedIn Corporation, 2022].

Ahora bien, como fuente de extracción de información de esta tesis se decide trabajar con **®LinkedIn** debido a que se considera la red profesional más grande del mundo en Inter-

net, conecta a empresas y empleados y en ella se puede encontrar el trabajo o la pasantía adecuada para cada perfil profesional. Esta red profesional, además, fortalece las relaciones profesionales y ayuda a aprender las habilidades que se necesitan para un área de trabajo profesional en particular.

ⓂLinkedIn permite específicamente:

- Encontrar y conectar con contactos sobre los campos de interés
- Ver dónde trabajan los colegas de la universidad
- Encontrar ofertas de empleo
- Participar en grupos especializados con otras personas que comparten afinidades
- Explorar y desarrollar nuevas habilidades de trabajo a través del módulo de aprendizaje de LinkedIn
- Utilizar la función LinkedIn Events para crear y unirse a eventos profesionales como talleres en línea, seminarios, eventos de ventas y *marketing*, eventos de *networking*, entre otros.
- Utilizar LinkedIn para escribir artículos, publicar fotos y videos, y más.

En el perfil de esta red social se encuentran varias opciones para movilizarse a través de las utilidades de esta. Las utilidades que se tendrán en cuenta para el desarrollo de este trabajo será la red de contactos y los empleos. En la primera, se encuentra a los profesionales con los que se tiene conexión, las personas que sigues, los grupos en los que estás, eventos, páginas y más. En la segunda, se puede buscar ofertas laborales, encuentra las vacantes guardadas, a las que se aplicó y las alertas de trabajo creadas por el propio perfil. Así mismo, dentro de las búsquedas se priorizarán los filtros de personas para buscar perfiles y empresas con características asociadas al dominio de TI.

ⓂLinkedIn es actualmente una de las redes profesionales más importante en todo el mundo, tiene 830 millones de miembros con más de 58 millones de empresas registradas.

“De acuerdo con el estudio de Uso de Redes Sociales en el 2021 de *Pew Research Center*, ⓂLinkedIn sigue siendo popular entre los estudiantes universitarios. Una investigación más profunda de estas estadísticas de ⓂLinkedIn ha revelado que más del 50 % de los adultos que tienen una licenciatura o un título avanzado en EE. UU. son usuarios de LinkedIn, mientras que el sitio solo atrae al 10 % de las personas cuya educación no supera la secundaria”. Lo anterior permite deducir que las ocupaciones y habilidades de los perfiles, además de las

ofertas de empleo requieren avanzado conocimiento en un área particular de conocimiento [Kinsta, Maddy Osman, 2022].

Existe otra razón por la cual los usuarios de  LinkedIn son activos en la plataforma y es por los esfuerzos de contratación de empleo. Más de 58 millones de empresas incluidas en el sitio y 50 millones de personas que buscan trabajo en LinkedIn cada semana, por lo que no es extraño que el 87 % de los reclutadores utilicen regularmente  LinkedIn [Kinsta, Maddy Osman, 2022].

Un estudio reveló que 122 millones de personas obtuvieron una entrevista a través de  LinkedIn y que 35,5 millones fueron contratadas por una persona con la que conectaron en el sitio. Lo anterior, propone que cada segundo se envían 95 solicitudes de empleo a través de LinkedIn y que cada minuto se contrata a 3 personas, según las estadísticas [Kinsta, Maddy Osman, 2022].

Además, un buen dato de  LinkedIn para los reclutadores es que los empleados reclutados a través del sitio tienen un 40 % menos de probabilidad de dejar la compañía en los primeros 6 meses [Kinsta, Maddy Osman, 2022].

Las estadísticas de esta red profesional y las funcionalidades de esta permiten ponerla como la opción número uno para la extracción de información para el aprendizaje de la ontología en el dominio de profesionales en TI. Los filtros de las búsquedas permiten adentrarse al mundo específico de perfiles y empleos de TI y los nuevos reclutadores siempre dan información acerca de las ocupaciones y competencias del perfil buscado.

Diseño del proceso de extracción de información de la red social profesional

Para poder obtener la información adecuada de la fuente de extracción, que en este caso es  LinkedIn, es necesario utilizar técnicas de *Web scraping*.

El *web scraping* es una técnica que sirve para extraer información de páginas web de forma automatizada. Con el *web scraping* se podrá conseguir la cantidad de información necesaria de la red profesional sin teclear una sola palabra mediante un programa de software, simulando la navegación de un humano en la dicha plataforma web. A través de los algoritmos de búsqueda podemos rastrear delimitar y extraer sólo aquella información que es útil [Khder, 2021]. En la Figura 4-6 se puede ver la arquitectura general del *web scraping*.

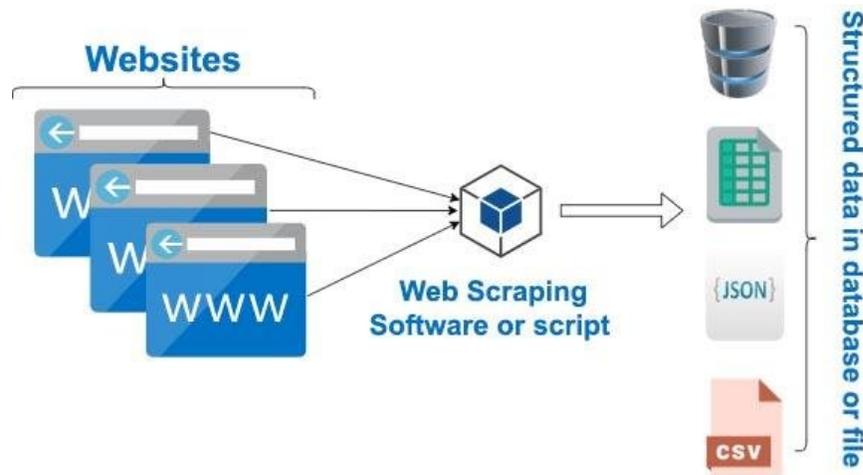


Figura 4-6.: Arquitectura del *web scraping*. Tomado de [Kasereka, 2020]

Así entonces, puede decirse que el *web scraping* está enfocado en la transformación de datos sin estructura en la web en datos estructurados que pueden ser almacenados y analizados en una base de datos central, en una hoja de cálculo o en alguna otra fuente de almacenamiento [Khder, 2021]. Debido a la necesidad de contenidos de calidad, el *web scraping* se ha convertido en los últimos años una técnica muy usada para generar grandes cantidades de datos a partir de la web.

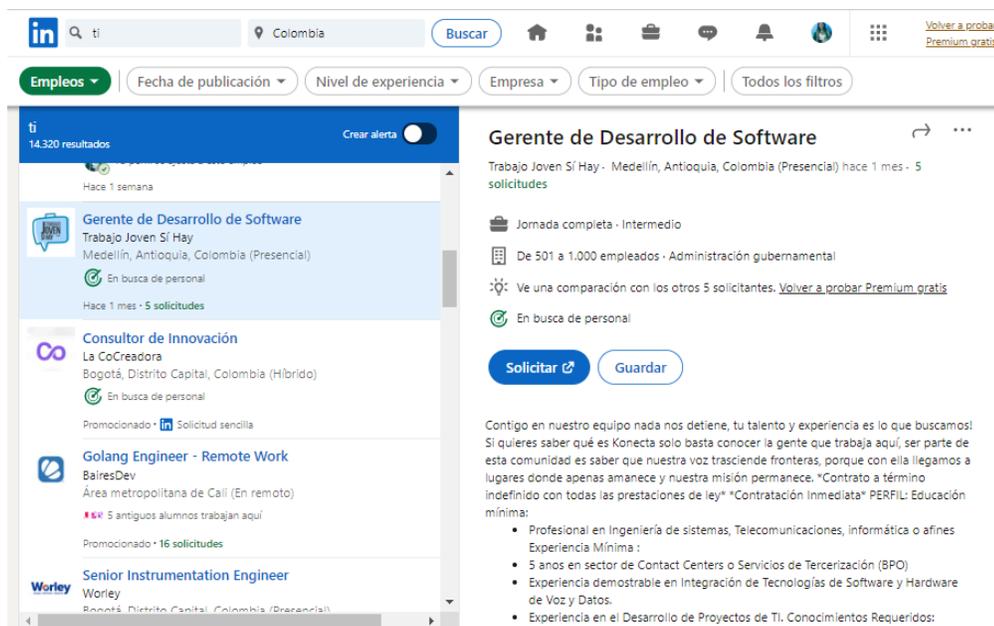


Figura 4-7.: Visualización de oferta laboral filtrada por TI en LinkedIn

El aprendizaje ontológico que se hará a partir de la ontología base propuesta en el capítulo

anterior utilizará la información dedicada a ocupaciones y competencias de profesionales de TI extraída de una red social, en este caso (R)LinkedIn como caso de uso. Una de las opciones es extraer información relevante de perfiles de personas en el área de TI encontrada en la red social, pero con el fin de simplificar la búsqueda y enfatizar el aprendizaje en las necesidades requeridas por la empresa en términos de ocupaciones y competencias, se extraerá la información del módulo de empleo (ver Figura 4-7). El módulo de empleo permite generar filtros sobre el tipo de ofertas que se están buscando, en este caso, TI.

Ahora bien, la arquitectura de *web scraping* que se propone para este trabajo se puede observar en la Figura 4-8.

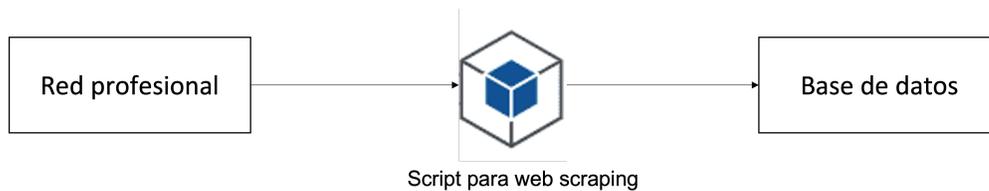


Figura 4-8.: Arquitectura de *web scraping* para este modelo

Para hacer web scraping de (R)LinkedIn se usa un script desarrollado usando la librería [Python Software Foundation, Ludovico Fabbri, 2022] de *Python linkedin-jobs-scrapers 1.15.4*. Esta librería permite extraer diferentes datos de los trabajos publicados en la plataforma web como se muestra en la Figura 4-9. Se identifica que *tittle* y *description* contienen los campos requeridos definidos, es decir, *tittle* contiene el **Título del puesto** y *description* contiene la **Descripción del puesto** y los **Requisitos del puesto**.

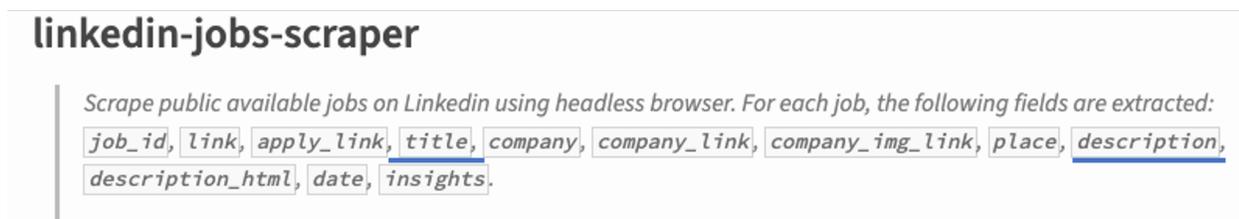


Figura 4-9.: Campos que se pueden extraer con *linkedin-jobs-scrapers* de una oferta laboral en LinkedIn

Los requerimientos de uso de esta librería de Python es Chrome or Chromium, Chromedriver: *latest version tested is 104.0.5112.79* y versión de Python mayor o igual a 3.6. En el Anexo C se muestra a detalle algunos a partes del script implementado para realizar el web scraping en (R)LinkedIn. Se extraen entonces los campos *tittle* y *description* y otros adicionales para revisar su contenido, tales como, *company* y *date*.

Finalmente, se obtiene una base de datos con la información extraída la cual será almacenada en un espacio de datos crudos con acceso centralizado a los datos, el cual permite almacenar información de diferentes fuentes y formatos sin tener que estructurarlos previamente, para luego poder ser analizados y explorados hacia la aplicación del modelo conceptual de aprendizaje ontológico en términos del dominio de este trabajo. En la Tabla 4-4 se muestran los campos extraídos con el *web scraping*.

En la Figura 4-10 se muestran las etapas del modelo conceptual que se comienzan a definir con esta descripción de los datos tanto de la ontología base, como los extraídos del web scraping.

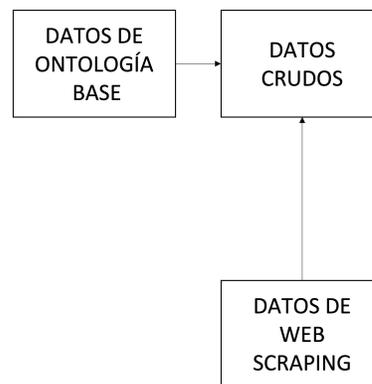


Figura 4-10.: Etapas 1 del modelo conceptual de aprendizaje ontológico

4.1.3. Preprocesamiento de los datos

En el preprocesamiento de los datos se quiere estructurar tanto los datos de la ontología base en crudo, como los datos del *web scraping* extraídos, con el fin tener dichos datos heterogéneos y de diferentes fuentes completamente organizados para trabajar los modelos de aprendizaje de la ontología.

Datos estructurados de la ontología base

Para estructurar los datos de la ontología base se procede a seleccionar las características que servirán de insumo de entrada para los modelos de aprendizaje. Se eligen los atributos relevantes que serán utilizados en el aprendizaje y se eliminan aquellas que no aportan información valiosa.

De todos los campos disponibles, se organizan los de la Tabla 4-1 alusivos a las ocupaciones y competencias. Se conserva el código de la ocupación, el título de la ocupación, se separan las competencias en una lista y finalmente, se le asigna un código a cada competencia y se hace una lista de códigos de competencias para cada ocupación. Esto me permite almacenar los

title	company	date	description
COORDINADOR DE TECNOLOGÍA DE LA INFORMACIÓN	Alfa Talent	20/07/22	Desarrollar, diseñar y mantener una arquitectura eficiente tanto en hardware como en software, implementar proyectos tecnológicos, políticas, procedimientos y capacidades internas con el fin de contribuir a ... Requisitos Estudios concluidos en Ingeniería de sistemas e Informáticas, Ingeniería en Telecomunicaciones o afines Experiencia total laboral mínimo de 5 años Experiencia en posiciones similares mínimo 3 años Tecnología Microsoft (Windows, Office, Dynamics GP, Windows Server, Active Directory, Group Policy, SMS/SCCM, MOM/SCOM, Exchange, SQL, DNS, IIS, ISA) Metodologías ...
Oficial de Riesgo de Tecnología de la Información (TI)	Konzerta Panamá	26/08/22	Descripción General del Puesto Contribuir al éxito general del Departamento de Riesgo Operativo garantizando que los objetivos planes e iniciativas individuales específicas se lleven a cabo y se cumplan en apoyo a las estrategias y objetivos del equipo. Garantizar que todas las actividades ... Requisitos Estudios universitarios de ingeniería industrial, sistemas computacionales, preferiblemente con estudios adicionales en finanzas o en administración de proyectos, riesgos. Mínimo 3 años de experiencia en áreas tecnológicas del sector bancario. Manipulación de base de datos para obtener análisis de riesgo, conocer los estándares ISO 27001, ISO 31000 e ISO 27001. Microsoft Office, Excel Avanzado. Beneficios...
ANALISTA DE CONTROL DE GESTIÓN TECNOLÓGICA DE LA INFORMACION	Saga Falabella	9/08/22	Misión Del Cargo Consolidar las fuentes de información de los sistemas de la compañía. Elaborar y automatizar reportes para la correcta toma de decisiones... Requisitos Ingeniería de sistemas (indispensable) Experiencia en BI, Inteligencia comercial, gestión de información, control de gestión o afines Inglés intermedio Ms Office Avanzado SQL Avanzado Visual Basic Avanzado Power BI Avanzado...
AUXILIAR DE TECNOLOGÍA DE LA INFORMACIÓN	Agustinos Perú	1/07/22	Funciones Principales Mantenimiento y reparación de computadoras / Laptops. Conocimiento sobre configuración de redes lan, wifi básico. Conocimiento de HTML ... Requisitos Grado de Instrucción Técnico en Computación Experiencia 2 años como técnico en computación deseable rubro educativo. Beneficios Ingreso a Planilla desde el primer día. Beneficios de ley vacaciones,cts, Convenios coporativos con las Instituciones Agustinas...
Especialista en Proyectos de Tecnología e Innovación	Sanitiana Guatemala	31/08/22	Objetivo General: planificar, organizar y dirigir los proyectos de tecnología e innovación
Analista de Tecnología de la información	Vinculos Estratégicos	2/08/22	Principales funciones Velar por el correcto funcionamiento de la plataforma de servidores de AES de acuerdo con los procedimientos establecidos para la administración de la misma... Optimizar soluciones tecnológicas en base a necesidades y oportunidades de mejora funcionales. Participar en proyectos tecnológicos con proveedores
Jefe de Informática	Konzerta Panamá	3/08/22	Funciones. Administrar el centro de cómputo. Instalar, configurar y administrar las redes de computadoras. Implementar soluciones integrales de cómputo de la organización. Definir propuestas de creación desarrollo y modificación de los sistemas de información... Requisitos Planificar, diseñar, ejecutar y monitorear la estrategia de tecnologías de información de la entidad. Proporcionar formación, apoyo y asesoramiento a los usuarios. Asesora a la alta dirección en la adquisición, implementación y desarrollo de toda la tecnología de la empresa...
...

Tabla 4-4.: Campos extraídos con el web scraping

código	título	lista.compe	lista.compe.cod
25	Profesionales de tecnología de la información y las comunicaciones	[Investigar el uso de las tecnologías de la información en las funciones empresariales; Identificar áreas de mejora e investigar los aspectos teóricos y los métodos operativos para el uso de ordenadores; Evaluar, planificar y diseñar configuraciones de hardware o software para aplicaciones específicas, incluyendo para internet, intranet y sistemas multimedia; Diseñar, redactar, probar y mantener programas informáticos; Desarrollar e implantar arquitecturas de bases de datos y sistemas de gestión de bases de datos; Desarrollar e implantar planes de seguridad y políticas de administración de datos, y administrar redes informáticas y entornos informáticos relacionados; Analizar, desarrollar, interpretar y evaluar especificaciones complejas de diseño y arquitectura de sistemas, modelos de datos y diagramas en el desarrollo, configuración e integración de sistemas informáticos]	[compe653; compe588; compe497; compe359; compe347; compe251; compe34]
251	Desarrolladores y analistas de software y multimedia	[Investigar el uso de las tecnologías de la información en las funciones empresariales e identificar las áreas en las que podrían introducirse mejoras para maximizar la eficacia y la eficiencia; Investigar los aspectos teóricos y los métodos operativos del uso de ordenadores; Evaluar, planificar y diseñar configuraciones de hardware o software para aplicaciones específicas; Diseñar, escribir, probar y mantener programas informáticos para requisitos específicos; Evaluar, planificar y diseñar sistemas de internet, intranet y multimedia]	[compe654; compe660; compe496; compe358; compe498]
2511	Analistas de sistemas	[Consultar con los usuarios para formular y documentar los requisitos y con la dirección para garantizar el acuerdo sobre los principios de los sistemas; Identificar y analizar los procesos empresariales, los procedimientos y las prácticas de trabajo; Identificar y evaluar las ineficiencias y recomendar prácticas empresariales y funcionalidades y comportamientos óptimos de los sistemas; Asumir la responsabilidad del despliegue de soluciones funcionales, como la creación, adopción y aplicación de planes de prueba del sistema; Desarrollar especificaciones funcionales para uso de los desarrolladores de sistemas; Ampliar o modificar los sistemas para mejorar el flujo de trabajo o servir a nuevos propósitos; Coordinar y vincular los sistemas informáticos dentro de una organización para aumentar la compatibilidad]	[compe181; compe581; compe587; compe107; compe254; compe21; compe230]
2514	Programadores de aplicaciones	[Escribir y mantener el código del programa descrito en las instrucciones y especificaciones de acuerdo con las normas de calidad acreditadas; Revisar, reparar o ampliar los programas existentes para aumentar la eficiencia operativa o adaptarse a nuevos requisitos; Realizar pruebas de funcionamiento de programas y aplicaciones de software para confirmar que producirán la información deseada; Recopilar y escribir la documentación del desarrollo del programa; Identificar y comunicar problemas técnicos, procesos y soluciones]	[compe406; compe1072; compe994; compe1010; compe582]
2519	Desarrolladores y analistas de software y multimedia y análisis no clasificados bajo otros epígrafes	[Desarrollar y documentar planes de pruebas de software; Instalar software y hardware y configurar el software del sistema operativo como preparación para las pruebas; Verificar que los programas funcionan de acuerdo con los requisitos del usuario y las directrices establecidas; Ejecutar, analizar y documentar los resultados de las pruebas de aplicaciones de software y de los sistemas de información y telecomunicaciones; Desarrollar e implementar políticas, procedimientos y guiones de pruebas de software y sistemas de información]	[compe266; compe26; compe1194; compe371; compe252]
...

Tabla 4-5.: Datos estructurados de la ontología base

responsabilidades, tareas, nivel de estudio, experiencia, competencias y datos adicionales. Es necesario entonces encontrar un patrón en los datos de entrada para extraer satisfactoriamente las competencias contenidas dentro de ese campo. Para el aprendizaje de ontologías de este trabajo se va a extraer solo ocupaciones y competencias, en un posible trabajo futuro se podrán extraer la información adicional contenida por los campos requeridos explicados en la Figura 4-5.

Para lograr esta estructuración de los datos se realiza un análisis de frecuencia de sentencias y una limpieza de texto.

- **Análisis de frecuencia de sentencias:** El análisis de frecuencia de sentencias, también conocido como análisis de frecuencia de oraciones o análisis de frecuencia de frases, se refiere al proceso de contar y evaluar la ocurrencia de frases o sentencias completas en un conjunto de documentos o corpus de texto. Es importante aclarar que esto no captura la complejidad semántica de las frases, pero si me permite identificar la ocurrencia de cada ocupación extraída del *web scraping* con el fin de conocer las ocupaciones que más se repiten o solicita la empresa en el ámbito de perfiles TI.

La importancia de este análisis es precisamente filtrar la información extraída con el fin de utilizar las ocupaciones de mayor ocurrencia. A mayor ocurrencia de una ocupación particular significa que está mayormente requerida por el mercado laboral y así mismo puede llegar a convertirse posiblemente en un dato nuevo para la ontología.

Ahora bien, sistemáticamente lo que se hace es crear un diccionario con las ocupaciones del campo *title* de los datos extraídos del *scraping*, luego se recorre el campo ocupación por ocupación y se incrementa el contador correspondiente para cada ocupación encontrada, se ordenan las ocupaciones por frecuencia en orden descendente y se establece un criterio para seleccionar las ocupaciones con las cuales se va a trabajar posteriormente.

En la Tabla 4-6 se puede ver un resultado parcial del análisis de frecuencia de frases. Es importante aclarar que en este punto dos ocupaciones que puedan ser sinónimos o etiquetas alternativas entre sí no son identificadas. Esto se realizará más adelante.

- **Limpieza de texto:** Este proceso tiene como objetivo limpiar y organizar los datos extraídos del *web scraping* con el fin de disponer la información con los mismos campos de los datos estructurados de la ontología base de la Tabla 4-5. La limpieza de texto es esencial para eliminar ruido, errores y elementos innecesarios que pueden dificultar el procesamiento y la extracción de información significativa de los datos de texto.

código nueva ocupación	ocupación	frecuencia
new_ocu1	Analista Junior de Seguridad - Oficina de Tecnología de la Información	35
new_ocu2	Analista de requerimientos Junior - Oficina de Tecnología de la Información	32
new_ocu3	Ingeniero de Datos - Oficina Tecnología de la Información	26
new_ocu4	Arquitecto de Datos - Oficina Tecnología de la Información	26
new_ocu5	Banco de Talento IT	7
...

Tabla 4-6.: Extracto de resultado de análisis de frecuencia de sentencias de las ocupaciones de la extracción del *web scraping*

En primer lugar, se cogen los campos *title* y *description* de la base de datos de la Tabla 4-4 y se dispone todo el texto en minúscula. Dicha conversión de todas las palabras del texto a minúsculas permite que las palabras en mayúsculas y minúsculas se consideren iguales y no se dupliquen en el análisis.

Luego, se realiza un proceso de eliminación de caracteres especiales, es decir, símbolos o emoticones que no sean palabras del texto. Estos caracteres no aportan información semántica relevante para el análisis de texto y pueden introducir ruido o distorsiones en los datos. Al eliminar los caracteres especiales, se obtiene un texto más limpio y adecuado para su procesamiento posterior, lo que mejora la calidad de los resultados en tareas de PLN.

Finalmente, se limpia el campo *description* con el fin de extraer la información de competencias requeridas para la ocupación ofertada. Para ello, se identifica que todo el campo tiene un patrón en la disposición de la información, donde al inicio describen las responsabilidades y tareas del cargo, luego se encuentra una etiqueta **Requisitos** donde a continuación se presentan las competencias y al final, se mencionan otros temas de experiencia, nivel de estudio, salario, etc. Por esto, se elimina automáticamente todo lo que esté antes de la etiqueta requisitos y todo lo que esté después de dicha etiqueta, con el fin de conservar en este caso las competencias (ver Figura 4-13). Se separa cada competencia por punto y luego se genera una lista de competencias por ocupación. A cada ocupación y competencia se le asigna un código y se realiza una lista de códigos de competencia por ocupación. En la Tabla 4-7 se observa un extracto de como quedan los datos estructurados del *web scraping* y en las Figuras 4-14, 4-15 se observa una nube de palabras para visualizar las palabras más importantes de estos datos.

En este punto, tanto los datos de la ontología base como los extraídos con el *web scraping* contienen las mismas características y están estructurados de forma homogénea para ser usados por un modelo de aprendizaje.

Ahora bien, es necesario generar otros tipos de preprocesamiento a los datos con el fin de eliminar datos innecesarios extraídos, definir ciertas coincidencias semánticas y clasificar competencias no clasificadas de la ontología base indicadas en la Figura **3-13**. A continuación, se explica a detalle estos procesos.

- **Análisis de correspondencia exacta:** El análisis de correspondencia exacta consiste en buscar e identificar ocurrencias exactas de una cadena o patrón de texto en un conjunto de datos. En otras palabras, se trata de encontrar todas las instancias en las que una secuencia específica de caracteres aparece exactamente igual en un texto o conjunto de textos [Gusfield, 1997], [Apostolico and Galil, 1997] y [Cormen et al., 2022].

El objetivo de realizar un análisis de correspondencia exacta aquí es descartar datos que no se consideren nuevo conocimiento, es decir, si una ocupación extraída del *web scraping* tiene correspondencia exacta con una ocupación de la ontología base se descarta para entrar en el proceso de aprendizaje ontológico porque no se considera datos nuevos. Por ejemplo, si en el *web scraping* se extrajo Ingeniero Desarrollador de Software y dentro de la ontología base ya existe esta ocupación exacta como clase, ese dato seguramente no me proporcionará ningún nuevo conocimiento del que el modelo pueda extraer aprendizaje.

- **Análisis de similitud semántica:** El análisis de similitud semántica es una técnica utilizada en el procesamiento de lenguaje natural (NLP) para medir la similitud entre palabras, frases, oraciones o documentos en función de su significado o semántica. El objetivo de esta técnica es cuantificar la similitud semántica entre unidades de texto con el fin de entender mejor su relación y extraer información relevante.

Existen varias formas de realizar un análisis de similitud semántica, algunas de las cuales incluyen modelos de incrustación de palabras *word embedding*, modelos de incrustación de palabras, medidas de similitud textual como la similitud del coseno, la distancia de Jaccard y la distancia de Levenshtein, etc., [Mikolov et al., 2013], [Pennington et al., 2014], [Mnih and Kavukcuoglu, 2013], [Cer et al., 2018].

Sin embargo, para este análisis se aplicará una arquitectura de modelos de lenguaje basados en el aprendizaje profundo que ha demostrado un gran éxito en tareas de procesamiento de lenguaje natural (NLP), incluido el análisis de similitud semántica. Esta arquitectura es llamada *Transformers*. Los *Transformers*, y en particular modelos como BERT (*Bidirectional Encoder Representations from Transformers*) y RoBERTa (*A Robustly Optimized*

código nueva ocupación	título	lista_compe_nuevas	lista_compe_nuevas_cod
new_ocu1	Analista Junior de Seguridad - Oficina de Tecnología de la Información	[Conocimiento en redes y telecomunicaciones, administración y mantenimiento de infraestructura de seguridad (firewall, vpn, antivirus, administración y mantenimiento de sistemas operativos windows y manejo de herramientas como kali linux, wireshark, nmap o similares', 'Contrato directo, a término indefinido y todas las prestaciones de ley']	['compe_new_59', 'compe_new_30', 'compe_new_36', 'compe_new_89']
new_ocu2	Analista de requerimientos Junior - Oficina de Tecnología de la Información	[Conocimiento en herramientas ofimáticas (word, excel, power point, outlook, internet', 'Conocimiento en procesamiento de datos, manejo de sql server para realización de consultas, notación de procesos bpmn, metodología scrum, herramientas de pruebas de web service', '2 años de experiencia en pruebas funcionales empresariales, preferiblemente en sector salud', 'Contrato indefinido y con todas las prestaciones de ley']	['compe_new_79', 'compe_new_27', 'compe_new_29', 'compe_new_3', 'compe_new_37', 'compe_new_89']
new_ocu3	Ingeniero de Datos - Oficina Tecnología de la Información	[Conocimiento y manejo de servicios cloud, buenas prácticas de uso, seguridad, y virtualización de datos, integración de datos con tecnologías etl y frameworks de orquestación, herramientas de calidad de datos (tales como master data services, mds, construcción de modelos físicos (t-sql, y lógicos (analysis services, power bi, otros', 'Conocimientos de controles centrados en datos usando iso 27001 y protección de datos personales', '2 años de experiencia como desarrollador en proyectos de inteligencia de negocio o análisis de datos en el ámbito empresarial, preferiblemente en sector salud', 'Contrato indefinido y todas las prestaciones de ley']	['compe_new_80', 'compe_new_20', 'compe_new_33', 'compe_new_35', 'compe_new_2', 'compe_new_38', 'compe_new_89']
new_ocu4	Arquitecto de Datos - Oficina Tecnología de la Información	[Certificado en manejo de bases de datos en productos microsoft power bi', 'Conocimiento general y técnico en el uso de los principales componentes de arquitecturas on-premise y nube, manejo de (sql database engine, ssas, ssrs, ssis, master data services, oracle, procesos de etl y calidad de datos (integración services, master data services, power-bi', 'Experiencia de 3 años en análisis y diseño en base de datos, calidad de datos, inteligencia de negocios, arquitectura de datos y gobierno de la información', 'Contrato indefinido y con todas las prestaciones de ley']	['compe_new_78', 'compe_new_21', 'compe_new_31', 'compe_new_53', 'compe_new_37', 'compe_new_89']
new_ocu5	Banco de Talento IT	[Sólida experiencia en el trabajo de campo con: programación, negocio y análisis de requisitos, data science, arquitectura, infraestructura, ingeniería de datos, rpa, proyectos de ti y funciones relacionadas del área de tecnología.]	['compe_new_12', 'compe_new_76', 'compe_new_23', 'compe_new_90', 'compe_new_99', 'compe_new_94', 'compe_new_65', 'compe_new_87', 'compe_new_92']
...

Tabla 4-7.: Datos estructurados del web scraping

BERT Pretraining Approach), han revolucionado el campo del PLN y se han convertido en referentes para el análisis de similitud semántica [Devlin et al., 2018], [Liu et al., 2019]. Los *Transformers* han demostrado ser altamente efectivos en el análisis de similitud semántica y han superado a enfoques anteriores en diversas tareas relacionadas con la comprensión del lenguaje natural. Su capacidad para capturar el contexto y el significado de las palabras y oraciones ha llevado a grandes avances en el campo del procesamiento de lenguaje natural [Sanh et al., 2019], [Feng et al., 2020].

Conociendo el potencial de los *Transformers* se propone realizar en este punto el análisis de similitud semántica entre las ocupaciones de la ontología base y las ocupaciones filtradas del *web scraping*.

Para la anterior se propone usar el modelo de lenguaje basado en *Transformers* SentenceBERT (SBERT), que es una modificación de la red BERT preentrenada que utiliza estructuras de red siamesa y triplete para derivar incrustaciones de oraciones semánticamente significativas que se pueden comparar mediante la similitud del coseno. Esto reduce el esfuerzo para encontrar el par más similar de 65 horas con BERT/RobERTa a unos 5 segundos con SBERT, manteniendo la precisión de BERT [Reimers and Gurevych, 2019].

El objetivo principal de SBERT es generar representaciones vectoriales para oraciones que capturen la similitud semántica entre ellas. Para lograr esto, SBERT utiliza dos encoders BERT idénticos (una arquitectura Siamesa) para procesar dos oraciones de entrada de manera paralela. Cada oración se pasa por su propio *encoder* BERT, y la salida de ambos *encoders* se utiliza para calcular la similitud semántica entre las dos oraciones (ver Figura 4-16).

El marco está basado en *PyTorch* y *Transformers* y ofrece una gran colección de modelos previamente entrenados ajustados para diversas tareas. El modelo preentrenado a usar es *hiiamsid/sentence-similarity-spanish-es* que asigna oraciones y párrafos a un espacio vectorial denso de 768 dimensiones. El tamaño máximo de secuencia (*max_seq_length*) es 512.

Para extraer la similitud semántica entre las incrustaciones de las ocupaciones y competencias de la base de datos del *web scraping* y las ocupaciones y competencias de la ontología base se calcula la similitud de coseno. La similitud del coseno se basa en el coseno del ángulo entre dos vectores en un espacio multidimensional.

Dadas dos incrustaciones de oraciones como vectores, la similitud del coseno se calcula de la siguiente manera (Ver Ecuación 4-1):

$$similitud_coseno(A, B) = (AB) / (||A|| * ||B||) \quad (4-1)$$

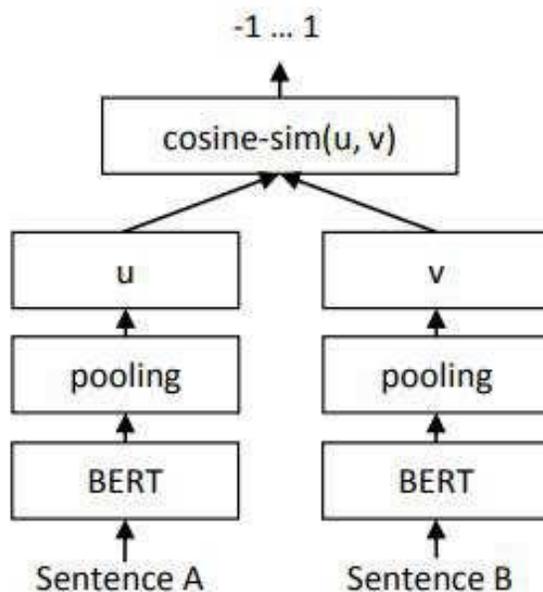


Figura 4-16.: Arquitectura SBERT para similaridad semántica. Tomado de [Reimers and Gurevych, 2019]

- A y B son las incrustaciones de oraciones (vectores) que se van a comparar.
- $(A \cdot B)$ representa el producto punto entre los vectores A y B.
- $\|A\|$ y $\|B\|$ representan las normas (magnitudes) de los vectores A y B, respectivamente.

La similitud del coseno oscila entre -1 y 1, pero se normaliza entre 0 y 1 para comprender mejor su resultado. Si el valor es 1 indica que los dos vectores están en la misma dirección y si es 0 indica que los dos vectores están en direcciones opuestas y son perfectamente diferentes.

En este caso, los elementos que tengan similaridad semántica por encima de 0.9 se excluirán de la lista de dato nuevo y se convertirán en una etiqueta alternativa posteriormente en la ontología final.

- **Clasificación de competencias:** Por su parte, es necesario retomar la idea de que en la ontología base diseñada se dejó una clase con todas las competencias No clasificadas como se muestra en la Figura 3-13. Además, a través del *web scraping* se tienen una lista de competencias que no están clasificadas dentro de las subclases definidas para la superclase competencias. En este paso se quiere clasificar todas estas competencias en las diferentes clases de la ontología.

Esto se realiza a través de la arquitectura de *Transformer* usada para el cálculo de la si-

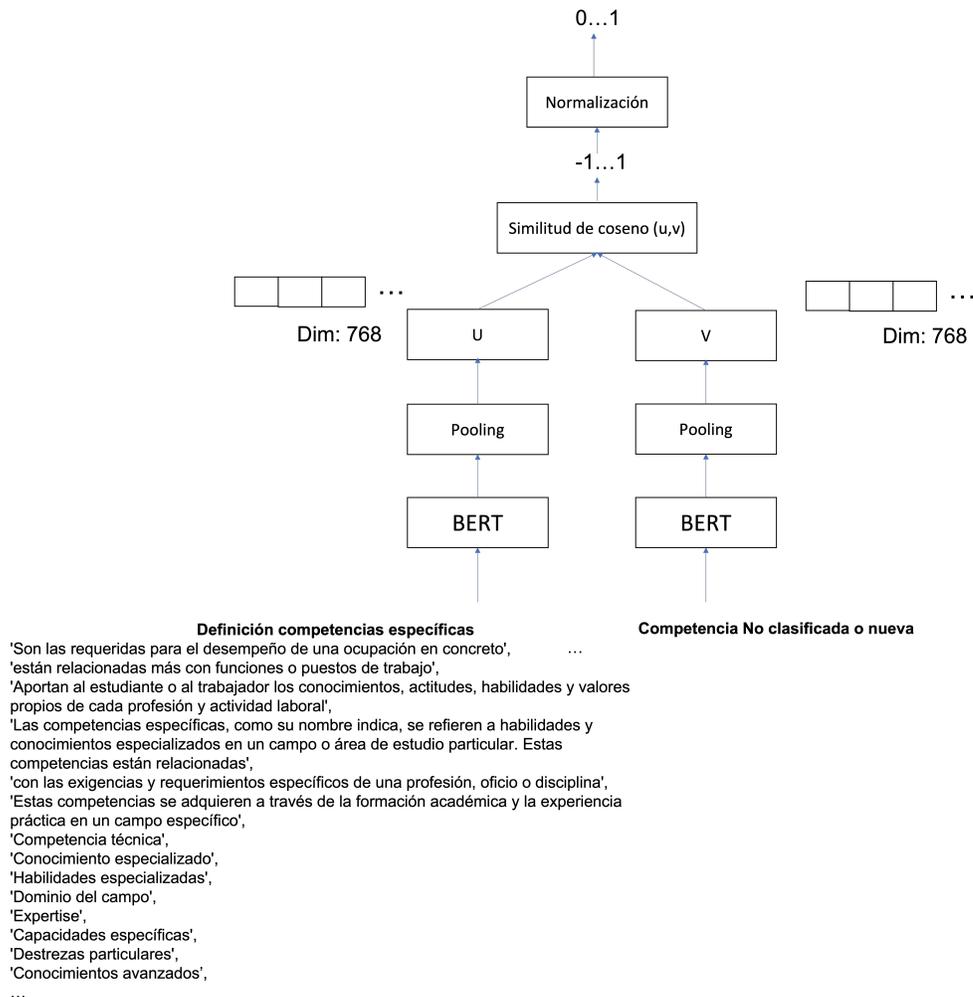


Figura 4-17.: Arquitectura SBERT para clasificación de competencias. Elaboración propia

militud semántica en el punto anterior, pero en este caso no se compara los *labels* de las competencias entre sí, si no la definición de cada una de las clases de competencias (Básicas, Transversales, Interpersonales, Instrumentales, Sistémicas y Específicas) con la competencia específica a la cual se le quiere medir la similitud semántica. La Figura 4-17 muestra un panorama general de entrada y salida de este paso del preprocesamiento.

En este caso, la competencia que tengan similaridad semántica con la definición de una clase de competencia por encima de 0.8 se clasifica como subclase dentro de dicha clase.

En la Figura 4-18 se muestra la etapa que se añade a la construcción del modelo conceptual hasta la descripción anterior.

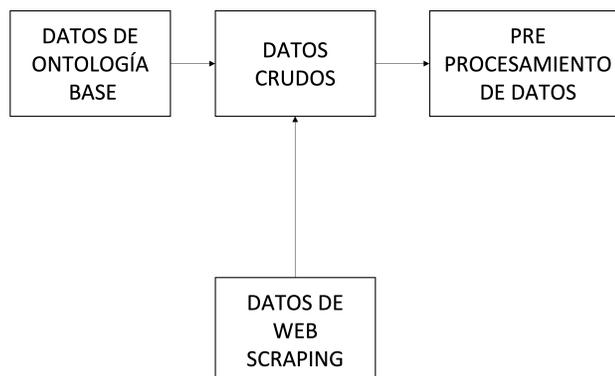


Figura 4-18.: Etapas 2 del modelo conceptual de aprendizaje ontológico

4.2. Análisis para aprendizaje de clases y jerarquía de clases

4.2.1. Análisis de técnicas de KGE

Un modelo KGE (*Knowledge Graph Embedding*) es un tipo de modelo de aprendizaje automático que se utiliza para generar *embeddings* o representaciones vectoriales de entidades y relaciones en un grafo de conocimiento. Estos modelos tienen como objetivo capturar la semántica y las relaciones implícitas presentes en el grafo de conocimiento, de manera que las entidades y las relaciones se representen de manera continua en un espacio vectorial. Se escogen las representaciones de grafos de conocimiento GC y específicamente los modelos KGE porque pueden ayudar en la inferencia y el razonamiento sobre la ontología. Al aprender las relaciones implícitas entre las clases, los modelos KGE pueden identificar conexiones indirectas y realizar inferencias basadas en las relaciones existentes en el grafo de conocimiento.

En esta sesión se describe el proceso de selección del modelo KGE que se aplicará posteriormente para el aprendizaje de clases y jerarquía de clases.

Con los datos de la ontología base debidamente estructurados, se realiza una carga de estos datos en el ambiente de pruebas. Luego de tener el conjunto de datos, se generan las tripletas que definen el grafo de conocimiento. Estas tripletas modelan las relaciones entre entidades.

```
[[ '11', 'sub_clase', '1'],  
 [ 'Compe141', 'es_competencia_de', '1120'],  
 [ 'e1', 'sub_clase', 'nivel_de_experiencia'], ... ]
```

Para evaluar el mejor modelo, se divide el conjunto de datos entre datos de entrenamiento, datos de prueba y datos de validación. Con los datos de entrenamiento se hace el entrena-

miento de todos los modelos y con los datos de prueba y validación se realizan las pruebas de los modelos. Esto permite comparar los resultados de los diferentes modelos para poder seleccionar el modelo que tiene mejor desempeño.

El número total de datos es de 5845. Los datos de entrenamiento son 4092, los datos de prueba son 1169 y los datos de validación son 584.

Para evaluar el mejor modelo, se hace una búsqueda sistemática, donde se definen los hiperparámetros del modelo que serán fijos y los cuales tendrán diferentes configuraciones. A continuación, se describen los hiperparámetros.

- Modelo: clase del modelo de KGE a evaluar. Los modelos de prueba son ComplEx, TransE y HolE.
- Batches: número de lotes en los que se divide el conjunto de entrenamiento durante el entrenamiento del modelo. Las opciones son 5, 10, 15, 20, 25, 30, 35, 40, 50, 100, 500, 1000.
- Epocas: iteraciones completas que el modelo realizará durante el entrenamiento. Las opciones son 20, 100, 500, 1000.
- Tamaño del vector: tamaño de las incrustaciones vectoriales aprendidas por el modelo. Las opciones son 500, 1000, 3000, 4000.
- eta: controla la esca de los gradientes utilizados para actualizar los parámetros del modelo durante el entrenamiento. Las opciones son 1, 5, 10.
- Función de pérdida: tipo de función de pérdida utilizada en el entrenamiento del modelo. Las opciones son pairwise, nll, self_adversarial.
- Regularización: estrategia de regularización usada por la función de pérdida. Las opciones son None y LP con lambda de 1e-4 o 1e-5.
- Optimizador: el optimizador utilizado para minimizar la función de pérdida. El optimizador es adam, con tasa de aprendizaje de lr 0.01 o 0.0001.

Se encuentran los mejores hiperparámetros para cada modelo y se obtienen las métricas de rendimiento del modelo. A continuación se muestran los hiperparámetros para el modelo TransE, ComplEx y HolE (ver Tabla 4-8).

Modelo	Batches	Épocas	K	eta	loss	loss margin	loss alpha	neg. corr. ent.	regul	optimiz.	opt. lr
TransE	10	100	500	5	pairwise	0.5	0.5	all	None	adam	0.01
ComplEx	10	100	500	5	pairwise	0.5	0.5	all	None	adam	0.01
Hole	10	100	500	5	pairwise	0.5	0.5	all	None	adam	0.01

Tabla 4-8.: Hiperparámetros modelos KGE

TransE, Representa relaciones como traslaciones vectoriales entre las incrustaciones de las entidades involucradas en la tripleta. ComplEx, representa las tripletas utilizando números complejos para capturar las relaciones simétricas y antisimétricas. Hole, utiliza operaciones holográficas para representar las tripletas capturando también relaciones simétricas y anti-simétricas de manera efectiva.

Con el fin de seleccionar el mejor modelo. Se evalúan diferentes modelos (ComplEx, TransE y Hole) con los hiperparámetros ya encontrados. En la tabla **4-9** se muestra la comparación de los resultados de los diferentes modelos evaluados y sus resultados.

Modelo	MR	Hits@1	Hits@3	Hits@10
TransE	47.315	0.382	0.752	0.825
ComplEx	83.785	0.777	0.782	0.802
Hole	75.941	0.730	0.840	0.870

Tabla 4-9.: Comparación de las métricas de los modelos KGE

Mean Rank (MR) es una métrica que evalúa qué tan bien el modelo puede clasificar las tripletas verdaderas (sujeto, relación, objeto) frente a las tripletas incorrectas. Se calcula como el promedio de los rangos de las tripletas verdaderas en el conjunto de prueba, donde el rango de una tripleta es el lugar que ocupa en la lista ordenada de todas las tripletas generadas, considerando una medida de similitud (por ejemplo, la distancia de coseno) entre las representaciones vectoriales de las tripletas. Generalmente, los valores de MR están en el rango de 1 a la longitud del conjunto de datos de prueba. Un valor bajo de MR indica un mejor rendimiento.

Hits@K mide la proporción de consultas para las cuales la entidad correcta se encuentra entre las primeras K entidades clasificadas por similitud. Los valores típicos para K son 1, 3, 5, 10, etc. Un valor más alto de Hits@K indica que el modelo tiene un mejor rendimiento.

De acuerdo con lo anterior, se define que el mejor modelo a usar es Hole ya que tiene un equilibrio razonable entre una buena clasificación en las primeras posiciones (Hits@1) y un MR aceptablemente bajo. Aunque el modelo TransE tiene un MR más bajo, su baja precisión en la primera posición (Hits@1) indica que tiene dificultades para identificar correctamente

los objetos en las principales predicciones.

En la Figura 4-19 se muestran las etapas que se añaden a la construcción del modelo conceptual hasta la descripción anterior.

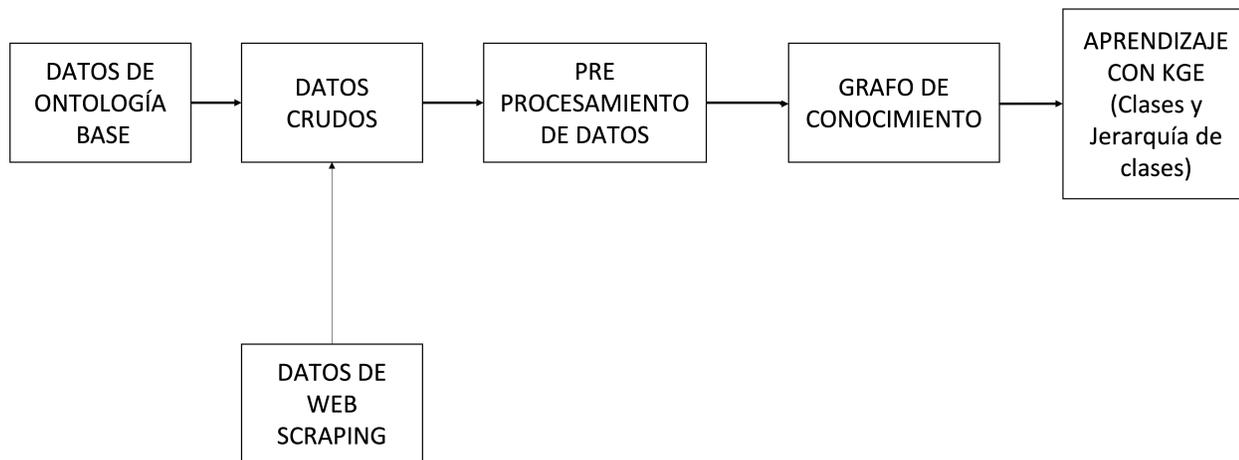


Figura 4-19.: Etapas 3 del modelo conceptual de aprendizaje ontológico

4.2.2. Propuesta de enriquecimiento de GC utilizado por KGE

Con el fin de usar la capacidad que tienen los *Transformers*, es decir, su capacidad de extraer la relación semántica, se propone agregar más relaciones entre los datos al grafo de conocimiento para ver si los resultados de las predicciones dadas por el modelo KGE aplicado mejora. Esto se realiza extrayendo la mayor probabilidad de relación semántica entre todos los conceptos de la ontología a través de la aplicación de la arquitectura de *Transformers* ya explicada.

Agregar estas nuevas relaciones trae ventajas como las de aumentar y enriquecer el GC y poder así tener más datos para entrenar, probar y validar el modelo. Además, el modelo KGE entrenado con estas relaciones semánticas demuestra tener mejor capacidad para realizar análisis en los datos nuevos.

Mediante los *Transformers*, se usa un modelo previamente entrenado y ajustado para comparar los conceptos por medio de similitud semántica, esto es comparando la distancia coseno entre los vectores que representan cada instancia.

Utilizando el *framework* de Python **SentenceTransformers** y de los posibles modelos preentrenados, se selecciona [hiiamsid/sentence_similarity_spanish_es](https://huggingface.co/hiiamsid/sentence_similarity_spanish_es) https://www.sbert.net/docs/pretrained_

`models.html`. Y con este modelo se analiza la similaridad semántica entre todos los elementos de la ontología base. Es decir, se compara cada elemento con todos los demás. El resultado son tripletas de relación semántica:

```
[['1', 'relacion_semantica', '11'],
 ['11', 'relacion_semantica', '3342'],
 ['1120', 'relacion_semantica', '1120'],
 ['1120', 'relacion_semantica', '1120'],
 ['12', 'relacion_semantica', '1'], ...]
```

Después de esto, el GC enriquecido queda con 10.126 tripletas. Para evaluar el modelo se divide en 8373 datos de entrenamiento, 584 datos de validación y 1169 datos de prueba.

En la Tabla 4-10 se muestra la comparación de los resultados del modelo de aprendizaje con KGE sin relación semántica y con relación semántica, es decir, con el GC inicial extraído de la ontología base y el GC enriquecido con las relaciones de similitud semántica extraídas con *Transformers*.

Modelo HoIE	MR	Hits@1	Hits@3	Hits@10
Sin relacion semántica	75.941	0.730	0.840	0.870
Con relacion semántica	25.79	0.851	0.874	0.900

Tabla 4-10.: Comparación del modelo KGE con y sin el GC enriquecido con relaciones semánticas

La propuesta de agregar la relación semántica entre las instancias genera un mejor resultado en las pruebas de definición del modelo. Lo que permite verificar que la propuesta genera un mejor resultado. En la Tabla 4-11 se puede ver el análisis de los trabajos más relacionados comparados con la propuesta.

Artículo	Descripción	Enfoque	Diferencia
Zengcan Xue, et al (2023). Learning knowledge graph embedding with multi-granularity relational augmentation network	Por medio de una red convolucional, extraen relación bidireccional implícita entre relaciones y entidades.	Relación entre instancias y entidades	Relación semántica entre todas las entidades
Cvetkov-Iliev, Alexis, et al (2023). Relational data embeddings for feature enrichment with background information	Extraer información numérica de los vectores para enriquecer el grafo.	Relación numérica	Relación semántica
Rezayi, Saed, et al (2021). EDGE: Enriching Knowledge Graph Embeddings with External Text	Enriquecer el grafo con textos externos.	Textos externos	La relación se hace desde los datos del grafo base
Aidan Hogan, et al (2021). KNOWLEDGE GRAPHS BOOK. Chapter 6: Creation and Enrichmen	Usando Wikipedia o TeX	Fuente externa	La relación se hace desde los datos del grafo base

Tabla 4-11.: Análisis de la literatura vs la propuesta

En la Figura 4-20 se muestran las etapas que se añaden a la construcción del modelo conceptual hasta la descripción anterior.

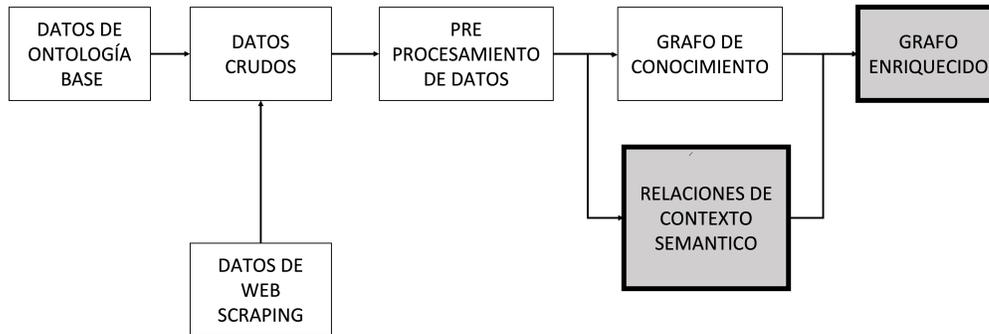


Figura 4-20.: Etapas 4 del modelo conceptual de aprendizaje ontológico

4.3. Análisis para aprendizaje de restricciones

Para el marco de esta tesis se analizan las restricciones en las propiedades definidas, más específicamente el dominio y el rango.

Para analizar las restricciones en su existencia de dominio se analizan las clases a las que pertenecen todas las instancias que se relacionan con la propiedad. El algoritmo apriori, es un algoritmo utilizado en minería de datos, sobre bases de datos transaccionales, que permite encontrar de forma eficiente **conjuntos de ítems frecuentes**. Estos conjuntos de elementos son conocidos como **itemsets** frecuentes. El algoritmo apriori se basa en el principio de que si, un conjunto de elementos es frecuente, entonces todos sus subconjuntos también lo son. Esto significa que si, un conjunto de elementos no cumple con el umbral de frecuencia mínimo, sus **supersets** no cumplirán con el umbral tampoco.

Entonces, para cada entidad, se visualiza toda la jerarquía de clases a la que pertenece a través de transacciones, por ejemplo, si la clase usada es 2151, se sabe que es subclase de 215, que 215 es subclase de 21, 21 es subclase de 2 y que 2 es subclase de Ocupaciones.

Ahora bien, al algoritmo apriori se le entrega una lista de transacciones de las clases que usan dicha propiedad, con la jerarquía de la clase correspondiente, así:

```

[[ '215 ', '21 ', '2 ', 'Ocupaciones ', 'tiene_competencia ' ],
  [ '13 ', '1 ', 'Ocupaciones ', 'tiene_competencia ' ]
  ... ]
  
```

En la Figura 4-21 se muestra la propuesta de aprendizaje de restricciones de dominio y

rango sobre las propiedades definidas en la ontología. Se inicia suponiendo que no se conoce el dominio y rango de la propiedad.

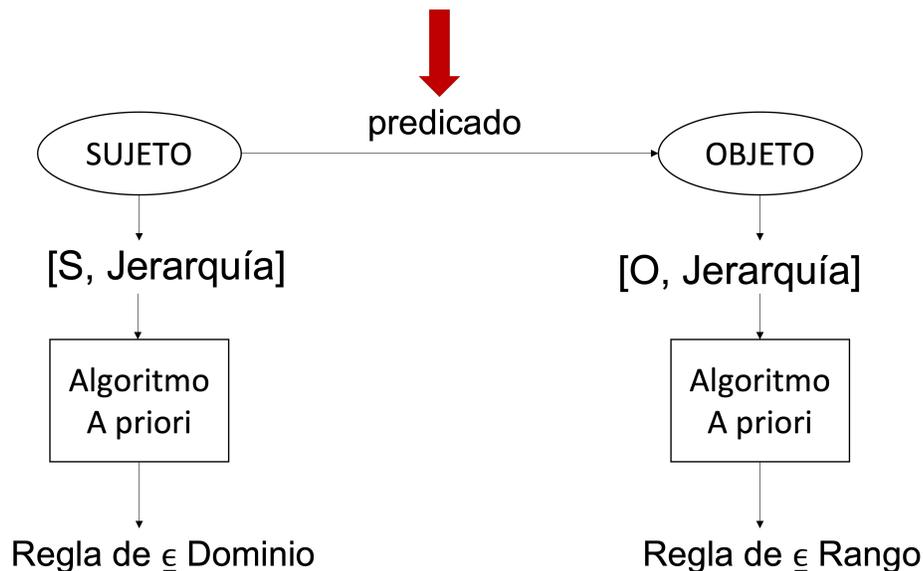


Figura 4-21.: Propuesta de aplicación del algoritmo apriori para aprender las restricciones de dominio y rango sobre una propiedad

Para esta lista se hace un análisis por reglas de asociación, esto porque las reglas de asociación permiten descubrir patrones en común dentro de los datos. El resultado es el siguiente:

Regla: RelacionX Asociado con clase
 Soporte: #
 Confianza: #
 Lift: #

Donde, soporte es la medida que define que tanto los datos contienen este conjunto y la confianza es el porcentaje de las reglas de los datos que contienen la pareja. Si la medida de soporte es de 1.0 podemos afirmar que todos los elementos de RelaciónX están asociadas a alguna clase, entonces el dominio de la propiedad RelaciónX es la clase.

El valor de soporte puede variar entre 0 y 1, un soporte más alto implica que el *itemset* es más frecuente en el conjunto de transacciones.

Por su parte, la confianza es una métrica que mide la fuerza de la asociación entre los *itemsets* y se utiliza para identificar reglas de asociación significativas y el *lift* es una métrica que compara la confianza de una regla de asociación con la probabilidad de ocurrencia

independiente de los *itemsets*. El valor de confianza puede variar entre 0 y 1, un valor de confianza más alto implica una relación más fuerte entre los *itemsets* X y Y. El valor de *lift* puede ser mayor o igual a 1. Un valor mayor que 1 indica una asociación interesante entre los *itemsets* X y Y.

4.4. Modelo Conceptual general para el aprendizaje ontológico en clases, jerarquía de clases y restricciones

Todo lo descrito en este capítulo se utiliza para validar la propuesta de modelo conceptual de aprendizaje ontológico para clases, jerarquía de clases y restricciones con aplicación al dominio de *e-recruitment*. A continuación, se proporciona el diseño del modelo conceptual (ver Figura 4-22).

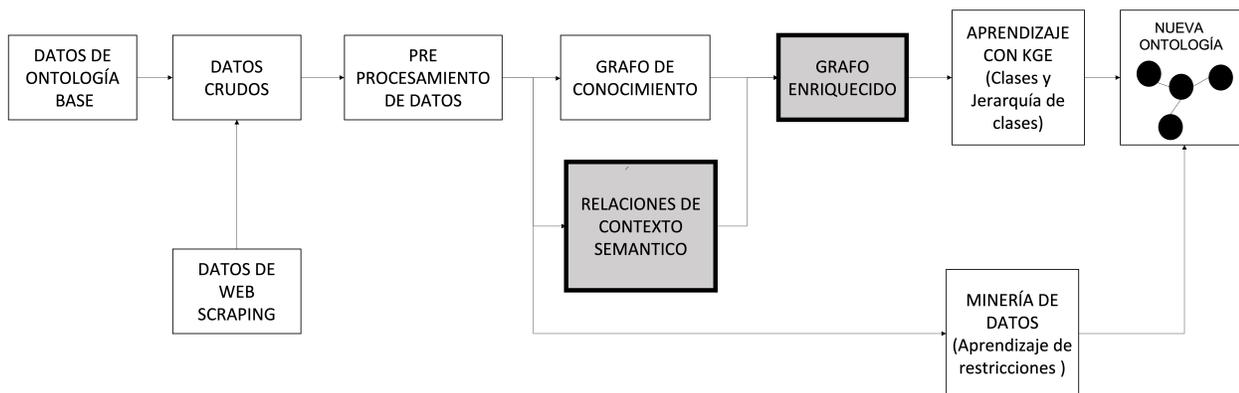


Figura 4-22.: Modelo conceptual de aprendizaje ontológico para clases, jerarquía de clases y restricciones con aplicación al dominio de *e-recruitment* para profesionales asociados a TI. Elaboración propia

4.5. Análisis y conclusiones del capítulo

El modelo conceptual de aprendizaje ontológico diseñado en este capítulo está formado por 10 módulos: datos de ontología base, datos de *web scraping*, espacio de datos crudos, preprocesamiento de datos, grafo de conocimiento, relaciones de contexto semántico, grafo enriquecido, aprendizaje con modelo KGE para clases y jerarquía de clase, minería de datos para aprendizaje de restricciones y finalmente, la nueva ontología. Este modelo de aprendizaje ontológico es general para una ontología base que tenga características parecidas a la construida. Además, permite trabajar adecuadamente el dominio de *e-recruitment* para diferentes perfiles profesionales.

La propuesta conceptual generada para este modelo se presenta teniendo en cuenta, primero, el dominio y las características de la fuente y la base de datos de información extraída de la misma; segundo, algunas anotaciones importantes que respalda la literatura, siendo en algunos casos técnicas, algoritmos o modelos bien alcanzados y aprobados por diferentes autores o en otros casos, técnicas, algoritmos o modelos que aún requieren profundizar su aplicación para verificar su real aporte. El capítulo 5 corresponderá a la implementación de este modelo conceptual.

De los aportes más importantes de este capítulo son:

- La definición de los campos requeridos para realizar un *web scraping*, ayudará a generalizar la búsqueda de la información de una oferta laboral en una red social profesional y así poder encontrar la información deseada para aprender sobre las clases y jerarquías de la ontología base. Se definen tres: Título del puesto, Descripción del puesto y Requisitos del puesto.
- Se estructuran los datos de la ontología base y los datos del *web scraping* con el fin de tener estas dos fuentes diferentes con los mismos campos: código, título, lista de competencias, lista de códigos de competencias. Estos datos estructurados son el insumo principal para un proceso posterior de aprendizaje.
- En el modelo conceptual diseñado, los transformadores de oraciones se usan para encontrar etiquetas alternativas, clasificación de competencias y enriquecimiento del grafo de conocimiento.
- Los SBERT de los *Transformers* utilizados en este trabajo permiten ingresar al *encoder* siamés 2 tipos de comparaciones entre sentencias, la primera, para comparar dos sentencias de competencias y la segunda, para comparar una definición de una clase de competencias con una competencia como tal, sin importar la dimensión de las sentencias.

- En este trabajo y para las características de los datos se encontró que el mejor modelo para KGE es HolE con MR 75.941 y Hits@1 0.730, con el cual se pretende hallar clases y jerarquías de clase.
- Con el grafo enriquecido se demuestra la mejora del rendimiento del modelo KGE con HolE a través de métricas MR 25.79 y Hits@1 0.851.
- Se define que entregando al algoritmo apriori transacciones de sujeto y sus jerarquías se puede hallar la restricción dominio sobre la propiedad. Así mismo, entregando al algoritmo apriori transacciones de objeto y sus jerarquías se puede hallar la restricción rango sobre la propiedad.

En este capítulo se cumplió el objetivo 2 de la tesis: Diseñar un modelo conceptual de aprendizaje ontológico con técnicas combinadas de aprendizaje que permita extraer desde los diferentes campos de la red social profesional las clases, jerarquía de clases y restricciones con el fin de extender y refinar la ontología base.

5. Capítulo 5: Implementación de portafolio de técnicas de aprendizaje de ontologías y obtención de nueva ontología

Este capítulo muestra la implementación del modelo conceptual o portafolio de técnicas y algoritmos de aprendizaje ontológico propuesto en la Figura 4-22 para extraer desde los diferentes campos tomados de una red social profesional las clases, jerarquía de clases y restricciones.

En la Figura 5-1 se representa de forma gráfica lo que se pretende en este capítulo. Se tomará como entrada la ontología base y la red profesional seleccionada (en este caso LinkedIn), se aplicará una técnica de *web scraping* con Python para extraer los datos para el aprendizaje ontológica y se implementará el portafolio de técnicas y algoritmos para aprender de las diferentes partes de la capa léxica, la capa taxonómica y las restricciones, para finalmente obtener una nueva ontología.

En la Figura 5-2 se muestra la metodología para el desarrollo de este capítulo.

A continuación, se mostrará sistemáticamente los resultados obtenidos en cada etapa del modelo conceptual de aprendizaje diseñado en el capítulo anterior.

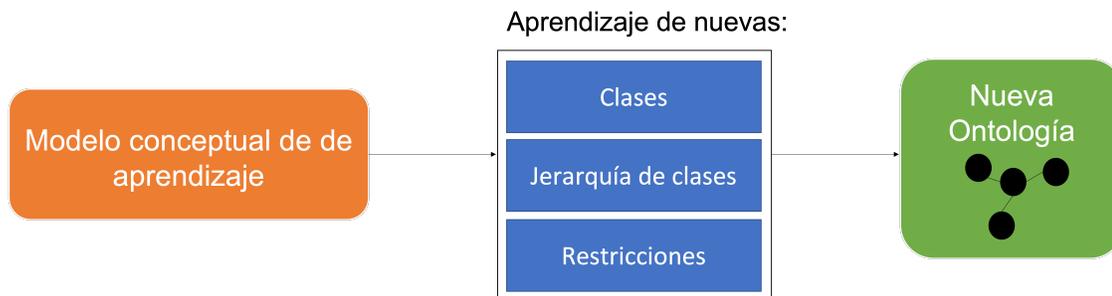


Figura 5-1.: Representación gráfica del objetivo del capítulo 5



Figura 5-2.: Metodología de desarrollo objetivo 3

5.1. Preparación de los datos

En este primer aparte se ampliarán y concluirán los resultados obtenidos de los bloques de las etapas 1 del modelo conceptual de aprendizaje ontológico: Datos de la ontología base, Datos del *web scraping* y Datos crudos (ver Figura 4-10).

Así entonces, se da claridad de que los datos crudos de la ontología base tiene diferentes *labels* que describen sus características: Superclase, Tipo, Código, Título, Definición y Competencias. Después de procesados, tiene los campos código, título, lista_compe y lista_compe_cod.

El número de ocupaciones es de 175. El número de clases de nivel de estudio es de 8. El número de clases asociadas a experiencia es 3 y las instancias 11. El número de competencias es de 1233.

Es importante remarcar que el procesamiento de datos se realiza con los códigos creados para los *labels* que definen las características de ocupaciones y competencias, para esto se organizan dos diccionarios, que son parejas ordenadas en (clave, valor). El primer diccionario tiene como clave el código y como valor el nombre y el segundo diccionario tiene como clave el nombre y como valor el código, como se muestra a continuación:

```
diccionario_codigo = {  
...  
'25': 'Profesionales de tecnología de la información y las comunicaciones',  
'251': 'Desarrolladores y analistas de software y multimedia',  
...  
'compe346': 'Diseñar y desarrollar algoritmos de procesamiento de
```

```
señales e implementarlos mediante la elección adecuada de hardware
y software',
...
}

diccionario_titulo = {
...
'Profesionales de tecnología de la información y las comunicaciones': '25',
'Desarrolladores y analistas de software y multimedia': '251',
...
'Diseñar y desarrollar algoritmos de procesamiento de señales e
implementarlos mediante la elección adecuada de hardware y software':
'compe346',
...
}
```

Por su parte, los datos crudos del web scraping tiene diferentes labels que describen sus características: *title*, *company*, *date* y *description*, Después de procesados, tiene los campos código, título, lista_compe y lista_compe_cod.

Al igual que con los datos de la ontología base, se definen dos diccionarios para los datos del *web scraping*, que son parejas ordenadas en (clave, valor). El primer diccionario tiene como clave el código y como valor el nombre y el segundo diccionario tiene como clave el nombre y como valor el código, como se muestra a continuación:

```
diccionario_n_codigo = {
...
'new_ocu1': 'Analista de Seguridad',
...
'compe_new_26': 'Conocimiento de aplicaciones de minería de procesos',
...
}

diccionario_n_titulo = {
...
'Analista de Seguridad': 'new_ocu1',
...
'Conocimiento de aplicaciones de minería de procesos': 'compe_new_26',
...
}
```

Esto me permite tener los datos estructurados de la misma forma, es decir, con los mismos *labels* que describen las características tanto para la ontología base, como para los datos extraídos del *web scraping*, para que puedan entrar en una etapa posterior de preprocesamiento y modelos de aprendizaje.

5.2. Preprocesamiento

En este segundo aparte se ampliarán y concluirán los resultados obtenidos en el bloque añadido a la etapa 2 del modelo conceptual de aprendizaje ontológico: Preprocesamiento (4-18).

5.2.1. Implementación de frecuencia de frases

En el contexto del análisis de frecuencia de frases que se implementa, es importante resaltar que la base de datos del *scraping* tiene 962 datos extraídos para cada uno de los *labels* que describen sus características. Para poder comparar el *label* de títulos, se estandarizan todos en minúscula y sin caracteres raros. Después se agrupan y se cuentan los datos por el *label* del título de la ocupación (la descripción de cada título se agrupa en un sólo texto). El resultado es un *dataframe* de 389 datos para el *label* de título.

Estos datos resultantes se organizan de mayor a menor. Esto para ordenar los títulos más frecuentes al inicio y los menos frecuentes al final. Por los valores observados de frecuencia se seleccionan los 50 primeros títulos para ser analizadas posteriormente como posibles nuevas ocupaciones dentro de la ontología.

En la Tabla 5-1 se pueden ver los títulos completos seleccionados y su frecuencia en los datos.

Número	Título	Frecuencia
1	Analista de seguridad	36
2	Analista de requerimientos	32
3	Analista de datos	26
4	Arquitecto de datos	26
5	Ingeniero de datos	26
6	Gerente de proyectos de ti	19
7	Analista funcional	18
8	Jefe de informática	16
9	Auxiliar de ingeniería	16
10	Gerente de tecnología encriptación	12

Sigue en la página siguiente.

Número	Título	Frecuencia
11	Administrador de proyectos	9
12	Director proyectos ti	9
13	Ingeniero de procesos	8
14	Analista de operaciones	8
15	Analista de procesos	8
16	Inteligencia de negocios	7
17	Banco de talento it	7
18	Jefe de tecnología de la información	7
19	Gerente de sistemas	6
20	Analista de proyectos	6
21	Analista programador	6
22	Ingeniero informático	6
23	Administrador de base de datos	5
24	Analista de sistemas	5
25	Gerente de servicios de tecnología e información	5
26	Analista de calidad de producto	4
27	Analista de base de datos	4
28	Coordinador experiencia del cliente	4
29	Analista asuntos regulatorios	4
30	Analista de gobierno de datos	4
31	Analista financiero	4
32	Subgerente requerimientos infraestructura	4
33	Gerente infraestructura y tecnología	4
34	Profesional de tecnología e innovación	4
35	Analista de sistemas funcional	4
36	Asistente de gestión de la calidad	4
37	Ingeniero de proyectos	4
38	Consultoría en soporte ti	4
39	Analista de planeamiento	4
40	Gerente de mejora continua	4
41	Auxiliar de bases de datos	4
42	Analista de negocio	3
43	Analista continuidad operacional junior	3
44	Consultor junior para tecnología	3
45	Auxiliar de tecnología de la información	3
46	Analista de tecnología de la información	3
47	Desarrollo software	3

Sigue en la página siguiente.

Número	Título	Frecuencia
48	Subgerente de aplicaciones ti	3
49	Asistente de auditoria de sistemas de información y tecnología	3
50	Analista de operaciones ti	3

Tabla 5-1.: Títulos y frecuencia

5.2.2. Resultado de análisis exploratorio de datos

El análisis exploratorio de los datos es una fase fundamental en cualquier proyecto de análisis de datos, donde el analista examina y explora los datos en bruto para obtener una comprensión inicial de su estructura, contenido y características.

Básicamente se visualizan los datos para identificar patrones y tendencias, y detectar posibles relaciones entre las variables. Todo lo anterior, tratando de encontrar correlaciones en los datos que puedan ser relevantes para la tarea o problema que se aborda que es precisamente el aprendizaje de clases y jerarquía de clases.

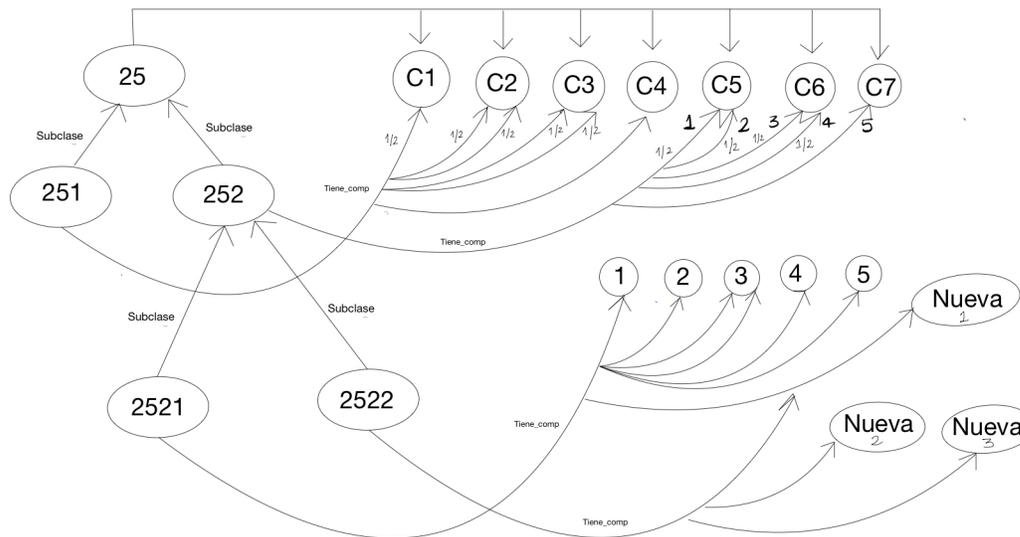


Figura 5-3.: Diagrama resultado de análisis exploratorio de los datos

En la Figura 5-3 se muestra resultado de análisis exploratorio de los datos. Los círculos de

la izquierda representan clases de la ontología base (25-Profesionales TIC, 251-Desarrollador y analista de software, 252-Especialista en base de datos redes) y los círculos a la derecha representan competencias asociadas a esas ocupaciones.

Del análisis exploratorio de los datos de la ontología base se pudo extraer que mientras más abajo en la jerarquía de clases de ocupaciones, más especialización tienen las competencias asociadas con respecto a la superclase principal, es decir, a través de toda la jerarquía creada por las clases se logra construir las competencias generales de la ocupación del primer nivel. Ahora bien, si tengo una ocupación de segundo nivel (25 en este caso), y esa ocupación tiene una jerarquía de subclases (251 y 251), estas últimas contienen de forma discretizada todas las competencias de su superclase. Si se baja a un tercer nivel de ocupaciones (2521 y 2522), sus competencias se especializan mucho más con respecto a la superclase que las contiene y adicionalmente, aparecen nuevas competencias.

Esta correlación entre las ocupaciones a través de la especialización de sus competencias alrededor de su jerarquía servirá como base para poder establecer la pregunta correspondiente al modelo KGE para poder extraer clases y jerarquía de clases.

5.2.3. Implementación de correspondencia exacta

En el análisis de correspondencia exacta se corrió un script que buscaba si había igualdad en todos los datos de los *labels* de título y los datos de los *labels* de competencias entre la ontología base y los datos del *web scraping*, con el fin de eliminar datos duplicados. En este caso se encontraron 2 elementos con correspondencia exacta, que se descartan inmediatamente como dato nuevo dentro de los elementos del *web scraping* (ver Tabla). Ahora bien, el resto de los datos del *web scraping* pueden ser posibles datos nuevos para la ontología y representan especializaciones de la jerarquía propuesta inicialmente.

título	código en onto base	código en web scraping
Trabajo en equipo	ctp1	new_compe_657
Liderazgo	cts4	new_compe_437

Tabla 5-2.: Resultado de correspondencia exacta

Aunque no se encuentren coincidencias exactas entre las características, este bloque se debe estar incluido siempre en el preprocesamiento porque una futura extracción de datos de una red social diferente puede variar en número de datos extraídos y en calidad. Y se hace entonces necesario descartar que haya correspondencia exacta.

5.2.4. Implementación de similaridad semántica con Transformers para definición de etiquetas alternativas

De acuerdo al diseño del capítulo anterior con la arquitectura de *Transformers* seleccionada, se compara cada una las posibles nuevas ocupaciones y competencias del *scraping* con todas las ocupaciones y competencias de la ontología base. Esto con el fin de encontrar cuales de las posibles ocupaciones tienen mayor similaridad semántica con las ocupaciones existentes en la ontología base. Si dos elementos tienen una alta similaridad semántica entre sí, se define que la posible característica es una etiqueta alternativa para la característica existente de la ontología base.

Número	Posible Nueva ocu	Posible nuevo código	Ocupación existente Similar	Código ocu similar	Medida
1	Analista de sistemas	new_ocu24	Analistas de sistemas	2511	0.9424
2	Analista financiero	new_ocu31	Analistas financieros	2413	0.9070
3	Analista de sistemas funcional	new_ocu35	Analistas de sistemas	2511	0.9019
4	Jefe de tecnología de la información	new_ocu18	Profesionales de tecnología de la información y las comunicaciones	25	0.8392
5	Analista de procesos	new_ocu15	Analistas de sistemas	2511	0.8351
6	Analista programador	new_ocu21	Analistas de sistemas	2511	0.8350
7	Auxiliar de tecnología de la información	new_ocu45	Técnicos en operaciones de tecnología de la información y las comunicaciones y asistencia al usuario	351	0.8265
8	Analista de operaciones	new_ocu14	Analistas de sistemas	2511	0.8222
9	Analista de operaciones ti	new_ocu50	Analistas de sistemas	2511	0.8097
10	Ingeniero de procesos	new_ocu13	Técnicos en control de procesos	313	0.8068

Tabla 5-3.: Similaridad semántica entre ocupaciones existentes y posibles nuevas

En la Tabla 5-3 se puede ver el resultado del análisis entre las ocupaciones de la ontología base y las posibles nuevas ocupaciones. La tabla está organizada de mayor similaridad a menor similaridad y sólo muestra las ocupaciones que tienen similaridad mayor a 0.8. Sin embargo, para considerar que una nueva ocupación es una etiqueta alternativa, se seleccionan sólo las que tienen similaridad mayor a 0.9. Esto quiere decir que Analista de sistemas y Analistas de sistemas se refieren a la misma ocupación, Analista financiero y Analistas financieros se refieren a la misma ocupación, y finalmente Analista de sistemas funcional y Analistas de sistemas también se refieren a la misma ocupación.

Con el objetivo de ver los resultados de forma gráfica, se toman los vectores de los conceptos y se realiza una reducción de la dimensionalidad PCA. Esto permite representar el vector en 2 dimensiones y poder así graficar. Cabe resaltar que, al hacer una reducción de la dimensionalidad, se pierde información en los datos. En este caso la varianza de los 2 ejes es de 0.3225 y 0.2207. Lo que suman 0.5433.

Por medio de la Figura 5-4 se puede hacer una interpretación espacial de los vectores.

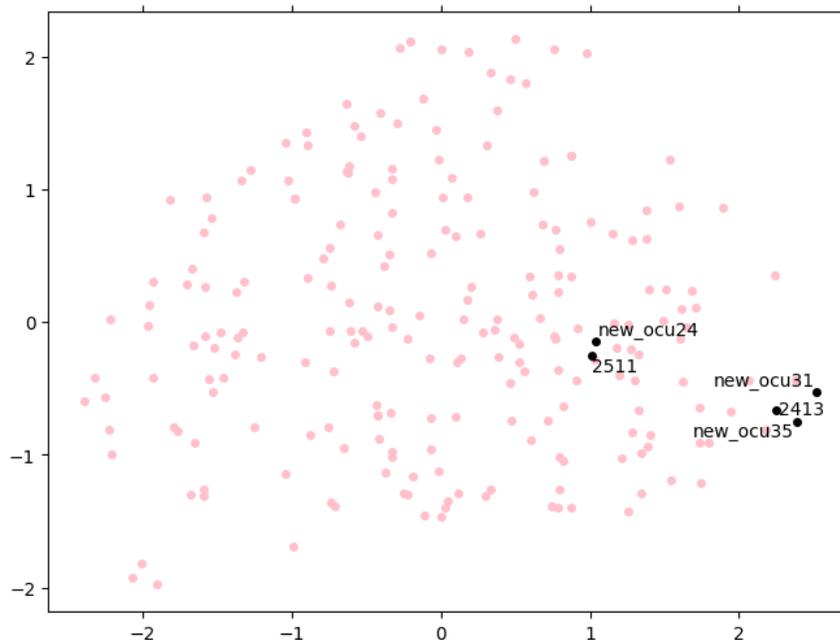


Figura 5-4.: Representación gráfica en 2D para algunos vectores con similitud semántica para las ocupaciones

Número	Nueva competencia	Código nueva compe	Competencia Similar	Código compe similar	Medida
1	Consultar con clientes de negocio para determinar las especificaciones funcionales del sistema y ofrecer respaldo operativo y a usuarios	new_compe.168	Consultar con los usuarios para formular y documentar los requisitos y con la dirección para garantizar el acuerdo sobre los principios de los sistemas	compe181	0.8463
2	Planificar, diseñar, ejecutar y monitorear la estrategia de tecnologías de información de la entidad	new_compe.520	Formular y dirigir estrategias, políticas y planes de tecnología de la información y la comunicación ti	compe538	0.8287
3	Buenas relaciones interpersonales	new_compe.101	Habilidades en las relaciones interpersonales	ctp4	0.8283
4	Desarrollar e implementar las políticas y protocolos de tecnología en la empresa	new_compe.206	Formular y dirigir estrategias, políticas y planes de tecnología de la información y la comunicación ti	compe538	0.8196
5	Consultoría especializada en tecnologías de la información y las comunicaciones	new_compe.170	Investigar el uso de las tecnologías de la información en las funciones empresariales	compe653	0.8181
6	Habilidades de comunicación oral y escrita	new_compe.374	Comunicación oral y escrita	cti3	0.8119
7	Adecuadas relaciones interpersonales	new_compe.38	Habilidades en las relaciones interpersonales	ctp4	0.8034
8	Contribuir al logro de los objetivos de la compañía a través del diseño, gestión y control de procesos que garanticen la calidad, simplicidad y eficiencia en su operación, alineados a la organización	new_compe.480	Ayudar y fomentar el desarrollo de objetivos, estrategias y planes destinados a lograr la satisfacción del cliente y el uso eficiente de los recursos de las organizaciones	compe128	0.7985
9	Garantizar que los procesos del departamento de tecnología e información funcionen sin contratiempos y eficientemente, procurando el correcto funcionamiento de los sistemas informáticos y la seguridad de la información	new_compe.481	Supervisar el funcionamiento diario de los sistemas informáticos y de comunicaciones	compe1103	0.7971

Tabla 5-4.: Similitud entre competencias existentes y posibles nuevas

En la Tabla 5-4 se puede ver el resultado del análisis entre las competencias de la ontología base y las posibles nuevas competencias del *scraping*. La tabla está organizada de mayor similitud a menor similitud. Para considerar que un elemento del *scraping* es una etiqueta alternativa de otro elemento de la ontología base, se seleccionan sólo las que tienen similitud mayor a 0.8.

Con el objetivo de graficar, se toman los vectores de los conceptos y se realiza una reducción de la dimensionalidad PCA. Esto permite representar el vector en 2 dimensiones y poder así graficar. Cabe resaltar que, al hacer una reducción de la dimensionalidad, se pierde información en los datos. En este caso la varianza de los 2 ejes es de 0.4193 y 0.1209. Lo que suman 0.5403.

Por medio de la Figura 5-5 se puede hacer una interpretación espacial de los vectores.

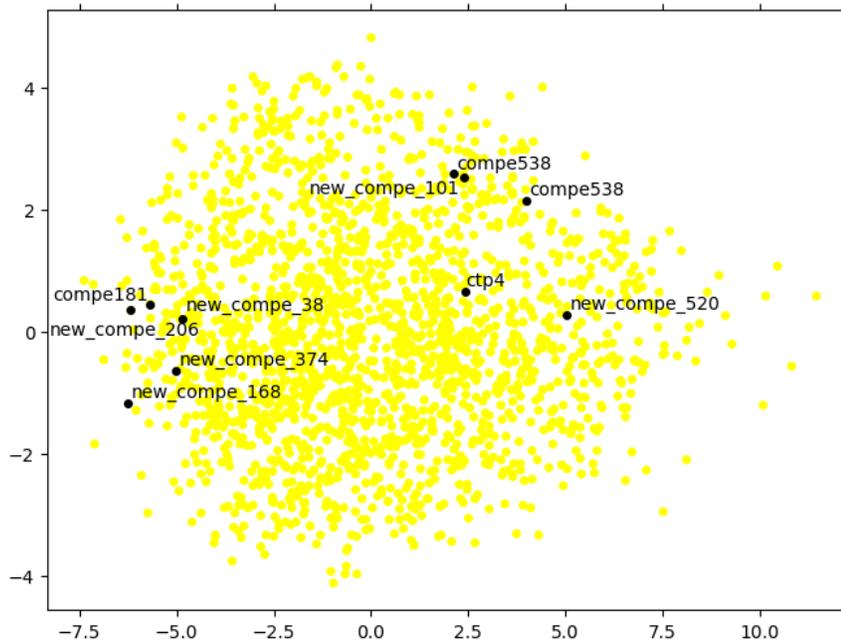


Figura 5-5.: Representación gráfica en 2D para algunos vectores con similitud semántica para las competencias

Ahora bien, se extraen 3 etiquetas alternativas para ocupaciones y 7 etiquetas alternativas para las competencias. En la Tabla 5-5 se puede ver la lista de nuevas etiquetas alternativas.

En la Figura 5-6 se ve la representación vectorial en 2D de los vectores de las ocupaciones y en la Figura 5-7 se ve la representación vectorial en 2D de los vectores de las competencias.

Número	Tipo	Elemento	Etiqueta alternativa
1	Ocupación	Analistas de sistemas	Analista de sistemas
2	Ocupación	Analistas financieros	Analista financiero
3	Ocupación	Analistas de sistemas	Analista de sistemas funcional
4	Competencia	Consultar con los usuarios para formular y documentar los requisitos y con la dirección para garantizar el acuerdo sobre los principios de los sistemas	Consultar con clientes de negocio para determinar las especificaciones funcionales del sistema y ofrecer respaldo operativo y a usuarios
5	Competencia	Formular y dirigir estrategias, políticas y planes de tecnología de la información y la comunicación ti	Planificar, diseñar, ejecutar y monitorear la estrategia de tecnologías de información de la entidad
6	Competencia	Habilidades en las relaciones interpersonales	Buenas relaciones interpersonales
7	Competencia	Formular y dirigir estrategias, políticas y planes de tecnología de la información y la comunicación ti	Desarrollar e implementar las políticas y protocolos de tecnología en la empresa
8	Competencia	Técnicos en operaciones de tecnología de la información y las comunicaciones y asistencia al usuario	· consultoría especializada en tecnologías de la información y las comunicaciones
9	Competencia	Comunicación oral y escrita	Habilidades de comunicación oral y escrita
10	Competencia	Habilidades en las relaciones interpersonales	Adecuadas relaciones interpersonales

Tabla 5-5.: Nuevas etiquetas alternativas

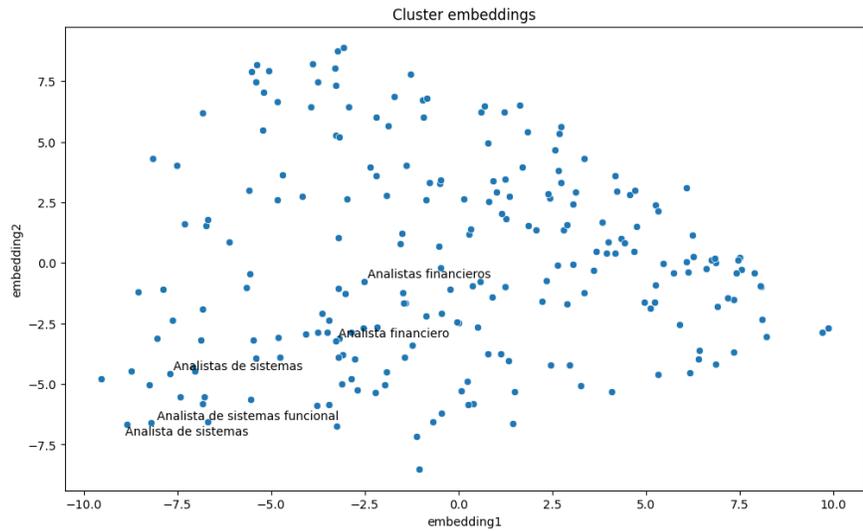


Figura 5-6.: Representación vectorial de las ocupaciones

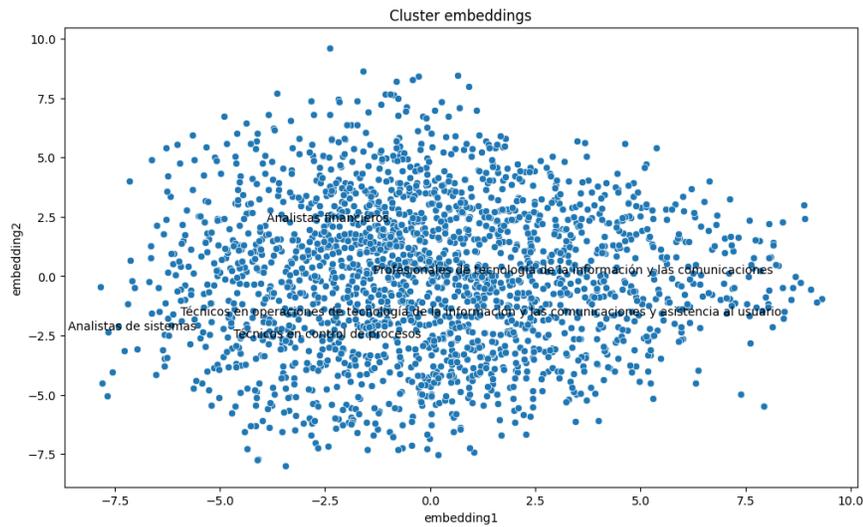


Figura 5-7.: Representación vectorial de las competencias

5.2.5. Implementación de Transformers para clasificación de competencias

Los datos de la ontología base tienen 1203 competencias y los datos del *scraping* tienen 676 posibles labels que describen la característica competencias. Ahora bien, como se mostró en el diseño de la arquitectura del *Transformer* se compara la medida de similitud semántica de la definición de cada una de las clases de competencias de la ontología base con: las competencias No clasificadas de la ontología base y las posibles competencias del *web scraping*

y se hace la clasificación.

En la tabla **5-6** se ve el resultado de la clasificación de las competencias No clasificadas de la ontología base.

Clase	Cantidad
Sistémicas	372
Específicas	368
Instrumentales	247
Interpersonales	175
Básicas	41

Tabla 5-6.: Clasificación de competencias No clasificadas de la ontología base

En la tabla **5-7** se ve el resultado de la clasificación de las posibles competencias de los datos del *scraping*.

Clase	Cantidad
Sistémicas	169
Específicas	182
Instrumentales	143
Interpersonales	75
Básicas	107

Tabla 5-7.: Clasificación de las posibles competencias del *scraping*

En la Tabla **5-8** se puede ver una porción de las competencias y posibles competencias clasificadas en una clase de competencias de la ontología bases porque tienen mayor similitud semántica con la definición de la clase.

En la Tabla **5-9** se puede ver una porción de las competencias del *scraping*, su clase con la que coincide mejor la definición y su puntaje.

En la Figura **5-8** se muestra el extracto de ontología para la superclase competencias antes y después de la clasificación.

Número	Competencia	Clase	Medida
0	Coordinar y vincular los sistemas informáticos dentro de una organización para aumentar la compatibilidad	Sistémicas	0.8175
1	Desarrollar especificaciones funcionales para uso de los desarrolladores de sistemas	Específicas	0.7763
2	Recopilar y escribir la documentación del desarrollo del programa	Instrumentales	0.7741
3	Aplicar, supervisar y evaluar las estrategias y políticas	Sistémicas	0.7497
4	Planificar e implementar la gestión informatizada de archivos y registros electrónicos	Instrumentales	0.7465
5	Supervisar el desarrollo y la aplicación de los sistemas de información de gestión	Instrumentales	0.746
6	Realizar investigaciones	Sistémicas	0.7301
7	Diseñar, redactar, probar y mantener programas informáticos	Específicas	0.7038
8	Modificar programas informáticos existentes para corregir errores, adaptarlos a nuevos equipos o actualizar interfaces y mejorar el rendimiento	Sistémicas	0.7031
9	Preparar artículos e informes científicos	Instrumentales	0.7001

Tabla 5-8.: Extracción de clasificación de competencias No clasificadas de la ontología base

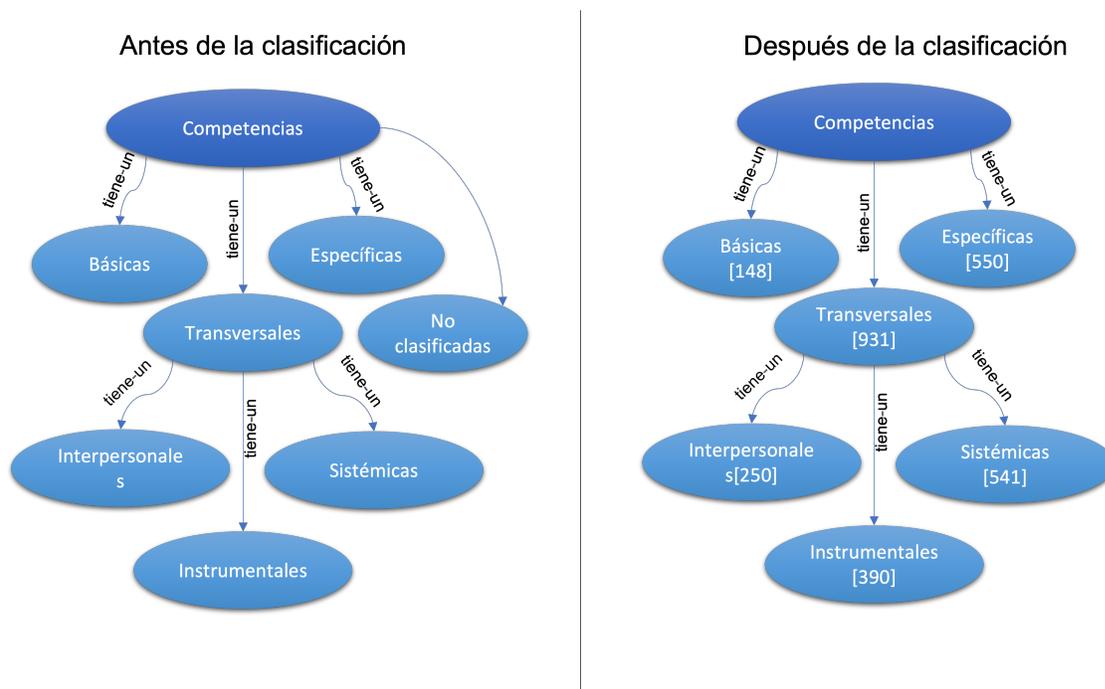


Figura 5-8.: Superclase competencias antes y después de la clasificación

	Nueva Competencia	Clase	Medida
1	Comprar maquinaria, equipos y suministros como tractores, semillas, fertilizantes y productos químicos	Instrumentales	0.7037
2	Dirigir y organizar exámenes de las poblaciones acuícolas o pesqueras para identificar enfermedades o parásitos	Instrumentales	0.6973
3	Diseñar y desarrollar animaciones digitales, imágenes, presentaciones, juegos, clips de audio y vídeo y aplicaciones de internet utilizando software multimedia, herramientas y utilidades, gráficos interactivos y lenguajes de programación	Interpersonales	0.6745
4	Investigar, desarrollar y ejecutar campañas de marketing y relaciones públicas	Interpersonales	0.6736
5	Evaluar y comparar la calidad relativa de diversos valores en una industria determinada	Instrumentales	0.6713
6	Instalar y realizar reparaciones menores de hardware, software o equipos periféricos, siguiendo las especificaciones de diseño o instalación	Instrumentales	0.6695
7	Determinar listas de precios, condiciones de descuento y entrega, presupuestos de promoción de ventas, métodos de venta, incentivos especiales y campañas	Específicas	0.6675
8	Aplicar conocimientos de principios y prácticas estadísticos, matemáticos, actuariales, contables y afines para identificar y resolver los problemas que surjan en el curso de su trabajo	Instrumentales	0.6672
9	Estudiar, mejorar y desarrollar teorías y técnicas matemáticas, actuariales y estadísticas	Básicas	0.6631

Tabla 5-9.: Extracción de clasificación de competencias del *scraping*

Por medio de la Figura 5-9 se puede hacer una interpretación de la distribución espacial de la clasificación de las competencias de la ontología base. La varianza de los 2 ejes es de 0.3698 y 0.1238, y suman 0.4937.

Por medio de la Figura 5-10 se puede hacer una interpretación de la distribución espacial de la clasificación de las competencias del *web scraping*. La varianza de los 2 ejes es de 0.3946 y 0.1325, y suman 0.5272.

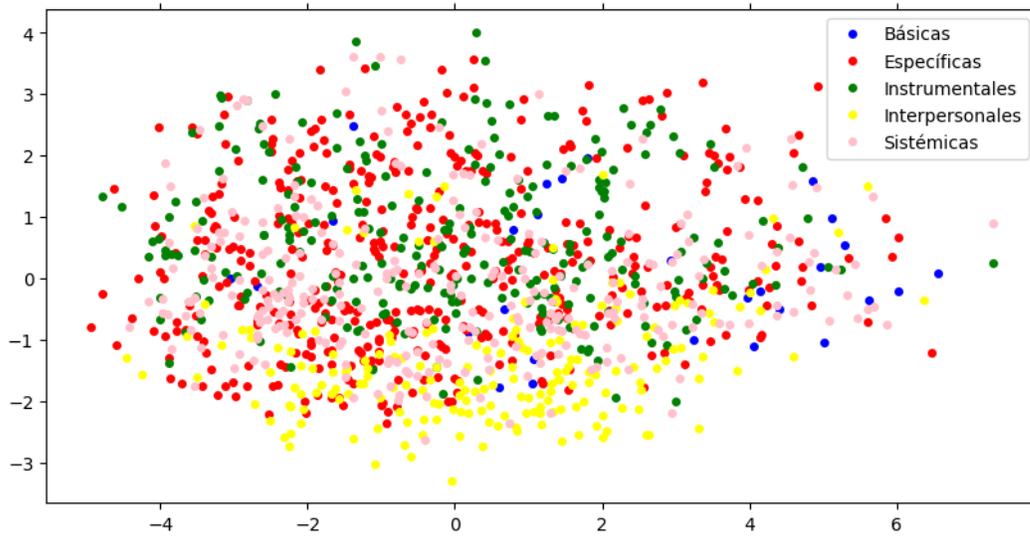


Figura 5-9.: Distribución espacial de los vectores de las competencias de la ontología base clasificados

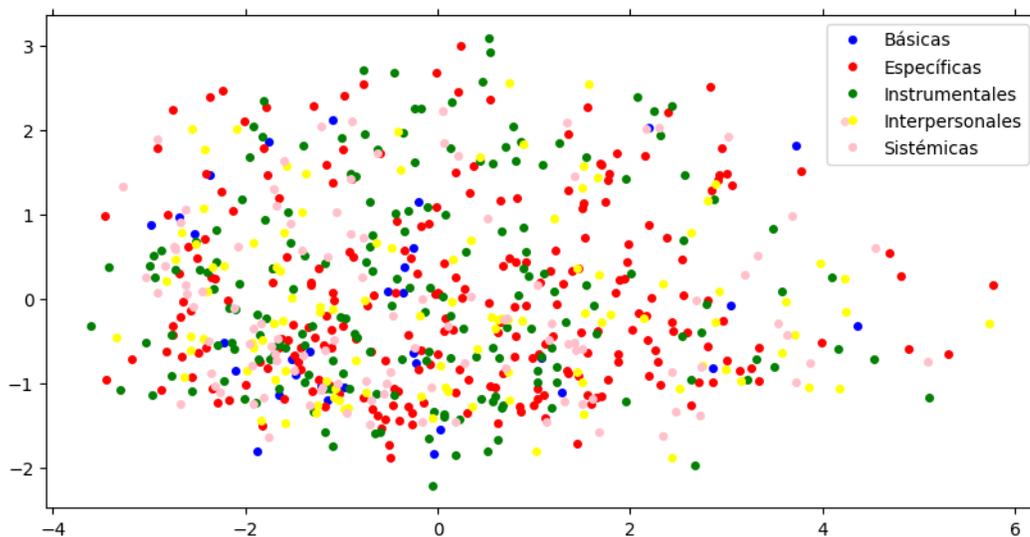


Figura 5-10.: Distribución espacial de los vectores de las competencias de la del *web scraping* clasificados

5.3. Implementación de modelo KGE para aprendizaje de clases y jerarquía de clases

En este primer aparte se ampliarán y concluirán los resultados obtenidos de los bloques de las etapas 3 (Figura 4-19) y etapas 4 (Figura (4-20) del modelo conceptual de aprendizaje

ontológico: Grafo de conocimiento, Relaciones de contexto semántico, Grafo enriquecido y Aprendizaje con KGE.

5.3.1. Implementación de tripletas para el grafo inicial basado en la ontología base

El grafo se construye con las diferentes relaciones que tiene la ontología base. La relación subclase representa la jerarquía de clases existente. La relación tiene_competencia y es_competencia_de relaciona las ocupaciones con cada una de las competencias que tiene asociadas.

Para las jerarquías (subclase_de), se tiene 175 tripletas de ocupaciones, 1203 de competencias y 52 para nivel de estudio y experiencia. Para tiene_competencia hay 1343 tripletas y para es_competencia_de hay 1343 tripletas. En total, el grafo inicial tiene 4116 tripletas.

A continuación, se amplía la vista de las tripletas construidas:

```
[['1', 'subClase', 'Ocupaciones'],
 ['11', 'subClase', '1'],
 ['112', 'subClase', '11'],
 ...,
 ['cb', 'subClase', 'Competencias'],
 ...,
 ['ne1', 'subClase', 'Nivel de Estudio'],
 ...,
 ['e1', 'subClase', 'Experiencia'],
 ...,
 ['compe1', 'subClase', 'cts'],
 ...,
 ['1120', 'tiene_competencia', 'compe846'],
 ['compe846', 'es_competencia_de', '1120'],
 ...,
]
```

En la Figura 5-11 se muestra el grafo inicial construido a través de las tripletas extraídas de la ontología base.

El modelo KGE implementado es el descrito en el capítulo anterior. En este caso, el modelo se entrena con el 100% de los datos, para poder tener un mejor modelo para la tarea de aprendizaje de clases y jerarquía de clases. Más adelante se describe la implementación del modelo KGE con el grafo inicial construido.

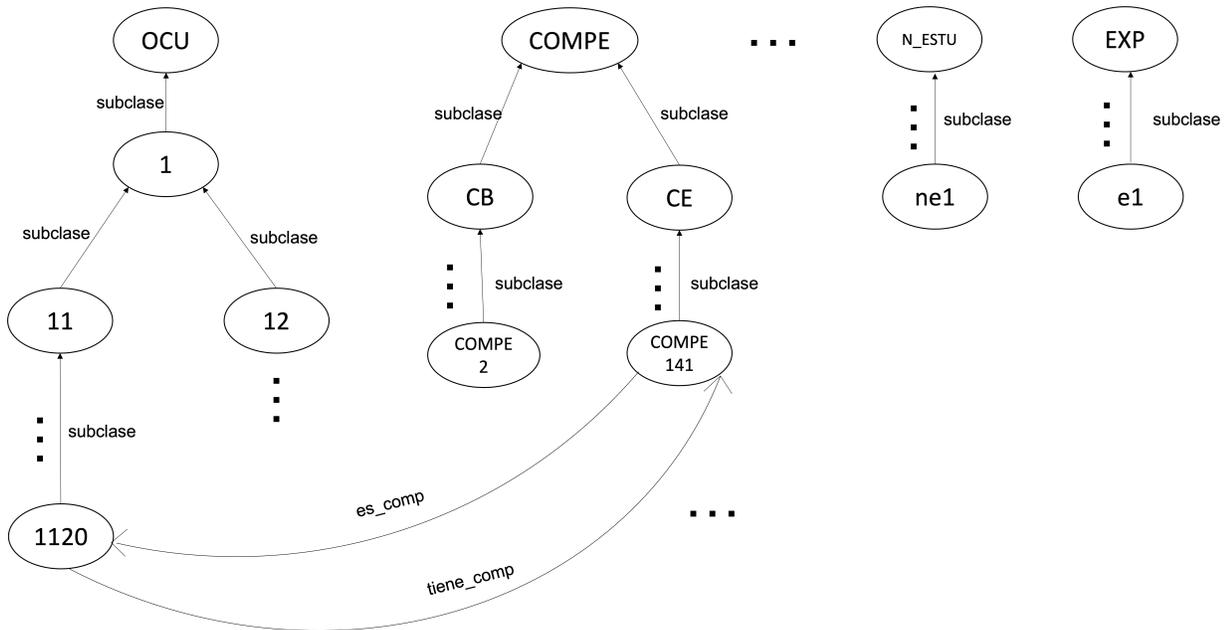


Figura 5-11.: Representación parcial del grafo de conocimiento asociado a las tripletas extraídas de la ontología base

5.3.2. Implementación de Transformers para encontrar relaciones semánticas que enriquezcan el grafo inicial

Se usa la arquitectura *Transformers* diseñada en el capítulo anterior para comparar cada uno de los elementos de la ontología con ellos mismos a través de un análisis de similitud semántica. Esto con el fin de enriquecer el grafo inicial con información valiosa implícita en la ontología, en este caso, la relación semántica entre sus elementos.

En la Tabla 5-10 se puede ver una extracción de la información. De los elementos que tienen entre sí mayor relación semántica.

	Item	Match	Medida
1	Directores de ventas, comercialización y desarrollo	Directores de ventas y comercialización	0.9469
2	Técnicos en asistencia al usuario de tecnología de la información y las comunicaciones	Técnicos en operaciones de tecnología de la información y las comunicaciones y asistencia al usuario	0.9459
3	Directores generales y gerentes generales	Directores y gerentes	0.9213
4	Técnicos en redes y sistemas de computadores	Profesionales en redes de computadores	0.9208
5	Profesionales en redes de computadores	Auxiliares contables y financieros	0.9208
6	Profesionales de las ciencias y de la ingeniería	Profesionales de las ciencias y la ingeniería de nivel medio	0.9151
7	Profesionales de las ciencias y la ingeniería de nivel medio	Directores de recursos humanos	0.9151
8	Desarrolladores y analistas de software y multimedia	Desarrolladores y analistas de software y multimedia y analistas no clasificados bajo otros epígrafes	0.9085
9	Instaladores y reparadores de líneas eléctricas	Instaladores y reparadores de equipos eléctricos	0.9057
10	Profesionales de tecnología de la información y las comunicaciones	Técnicos de la tecnología de la información y las comunicaciones	0.9085

Tabla 5-10.: Relación semántica entre *items*

Ahora, se pueden agregar al grafo las nuevas tripletas que contienen la información semántica extraída de la ontología base. Se muestra a continuación alguna de ellas:

```
[...,
 ['251', 'relacion_semantica', '2514'],
 ['2511', 'relacion_semantica', '2421'],
 ['2512', 'relacion_semantica', '251'],
 ['2513', 'relacion_semantica', '251'],
 ['2514', 'relacion_semantica', 'compe254'],
 ['2519', 'relacion_semantica', '251'],
 ['compe251', 'relacion_semantica', 'compe837'],
 ...,
 ]
```

El grafo se construye con las diferentes relaciones que se tienen. La relación subClase representa la jerarquía de clases existente. La relación tiene_competencia y es_competencia_de

relaciona las ocupaciones con cada una de las competencias que tiene asociadas. Además de las relaciones de semántica. La cantidad de tripletas del grafo enriquecido es de 10.126.

El modelo KGE diseñado se implementa también con este grafo enriquecido. En este caso, el modelo se entrena con el 100 % de los datos, para poder tener un mejor modelo de aprendizaje. Más adelante en el capítulo se comparan los resultados de la aplicación del mejor modelo KGE para aprender clases y jerarquía de clases con el grafo inicial y el grafo enriquecido.

5.4. Resultados de la implementación del modelo KGE diseñado para aprendizaje de clases y jerarquía de clases

Teniendo en cuenta la capacidad de la librería *Ampligraph* [Costabello et al., 2019] para aplicar modelos KGE que descubran nuevos conocimientos a partir de un GC existente y complete grandes gráficos de conocimiento con declaraciones faltantes se genera la siguiente base de inferencia para extraer nuevas clases y jerarquía de clases (ver Figura 5-12).

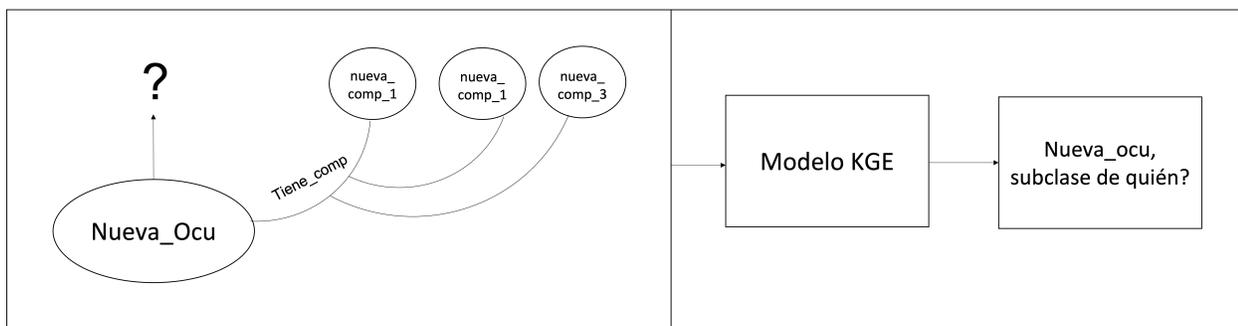


Figura 5-12.: Base de inferencia de nuevas clases y jerarquía de clases

En este punto del desarrollo las competencias del *scraping* fueron clasificadas en la ontología base en las superclases de competencias, esas competencias están asociadas a una posible ocupación que no está aún dentro de la ontología. Como *Ampligraph* ya conoce las relaciones *tiene_competencia* y *es_competencia_de*, se forman tripletas como las mostradas en la Figura 5-12. Si se ingresan esas tripletas al modelo KGE, conociendo ya la relación **subClase** y se pregunta esa nueva posible ocupación *Nueva_ocu* de quién es subclase se obtienen resultados de aprendizaje de clases dentro de una taxonomía establecida, entonces también de jerarquía de clases.

En general, se consulta por las nuevas relaciones que pueden tener los datos. Las siguientes

son las nuevas clases ubicadas en la jerarquía de clases existente, que se pueden sacar por medio del modelo (ver Tabla 5-11).

nueva clase	subclase de	código subc.	Prob.
Analista de negocio	Analistas financieros	2413	0.5572
Analista de procesos	Analistas de sistemas	2511	0.5549
Analista programador	Analistas de sistemas	2511	0.5543
Analista de base de datos	Directores de servicios de tecnología de la información y las comunicaciones	1330	0.5483
Inteligencia de negocios	Profesionales de ventas de tecnología de la información y las comunicaciones	2434	0.5469
Analista de gobierno de datos	Analistas financieros	2413	0.5459
Analista de seguridad	Analistas de sistemas	2511	0.5453
Administrador de base de datos	Analistas de sistemas	2511	0.5449
Gerente de tecnología encriptación	Administradores de sistemas	2522	0.5397
Jefe de tecnología de la información	Administradores de sistemas	2522	0.5390
Asistente de auditoria de sistemas de información y tecnología	Analistas de gestión y organización	2421	0.5389
Analista de operaciones	Directores de servicios de tecnología de la información y las comunicaciones	1330	0.5295
Auxiliar de bases de datos	Directores de servicios de tecnología de la información y las comunicaciones	1330	0.5278
Consultor junior para tecnología	Especialistas en bases de datos y en redes de computadores no clasificados bajo otros epígrafes	2529	0.5220
Ingeniero de proyectos	Directores de servicios de tecnología de la información y las comunicaciones	1330	0.5201

Tabla 5-11.: Resultados de aprendizaje de clases y jerarquía de clases

En total se extrajeron 15 nuevas clases sobre la jerarquía existente de la ontología.

5.5. Resultados de la implementación de reglas de asociación para aprendizaje restricciones

A continuación, se buscan reglas de asociación, en las que se relaciona una competencia con una ocupación. Y Reglas de asociación, donde se relacionan 2 competencias. Como en el diseño de esta etapa realizado en el capítulo anterior, se entregan sujetos con jerarquías como transacciones para encontrar un posible restricción de dominio de la relación y luego objetos con su jerarquía para encontrar posibles restricciones de rango de la relación.

Entonces, para cada elemento, se visualiza toda la jerarquía de clases a la que pertenece, es decir, si la clase usada es 2151, se sabe que es subclase de 215, que 215 es subclase de 21, que 21 es subclase de 2 y que 2 es subclase de Ocupaciones.

Finalmente, se tiene una lista las clases que usan dicha propiedad, con la jerarquía de la clase.

```
[[ '215', '21', '2', 'Ocupaciones', 'tiene_competencia'],  
  [ '13', '1', 'Ocupaciones', 'tiene_competencia']  
  ...]
```

Para esta lista se hace un análisis por reglas de asociación, esto porque las reglas de asociación permiten descubrir patrones en común dentro de los datos. El resultado es el siguiente:

```
Regla: tiene_competencia Asociado con Ocupaciones  
Soporte: 1.0  
Confianza: 1.0  
Lift: 1.0
```

Donde soporte es la medida que define que tanto los datos contienen este conjunto y la confianza es el porcentaje de las reglas de los datos que contienen la pareja. Como la medida de soporte es de 1.0 podemos afirmar que todos los elementos de tiene_competencia están asociados a alguna clase de ocupación, entonces el dominio de la propiedad tiene_competencia es la clase Ocupación.

Para analizar el rango, se analizan las clases a las que se asocian dichas propiedades.

```
[[ 'tiene_competencia', 'cti', 'ct', 'Competencias'],  
  [ 'tiene_competencia', 'ctp', 'ct', 'Competencias']]  
  ...]
```

Regla: tiene_competencia Asociado con Competencias
Support: 1.0
Confidence: 1.0
Lift: 1.0

Entonces el rango de la propiedad es la clase Competencias. Se asume que esto no se sabía, y con los resultados se puede evidenciar que la asociación de dominio y rango es correcta.

Con lo anterior podemos definir que para la propiedad tiene_competencia, su dominio son las clases Ocupaciones y su rango son las clases Competencias. Además, que la propiedad tiene_competencia tiene una restricción existencial (*Some*) asociada a las clases de Competencias.

Realizando el análisis anterior para la propiedad es_competencia_de se tiene que:

Dominio:
Rule: es_competencia_de with Competencias
Support: 1.0
Confidence: 1.0
Lift: 1.0

Rango:
Rule: es_competencia_de with Ocupaciones
Support: 1.0
Confidence: 1.0
Lift: 1.0

Como trabajo adicional se presentan unos avances preliminares de como aprender nuevas restricciones. Para realizar el aprendizaje de restricciones, también se pueden usar aprendizajes con reglas de asociación.

Después de ingresar las transacciones jerárquicas al algoritmo de reglas de asociación, se encontró que hay 2 tipos de relaciones.

La regla que relaciona un elemento con dominio Ocupación y rango Competencia. Esto podría significar que para esas ocupaciones, la competencia es una competencia muy común, o en su defecto una competencia esencial u obligatoria (ver Tabla 5-12).

regla	asociado con	soporte	confianza	lift
12 Directores administradores y comerciales	compe438 Establecer y dirigir los procedimientos operativos	0.1463	0.6000	3.0749
12 Directores administradores y comerciales	compe838 Planificar y dirigir las operaciones diarias	0.1463	0.6000	2.4600
14 Gerentes de hoteles, restaurantes, comercios y otros servicios	compe214 Controlar la selección, formación y supervisión del personal	0.1219	0.6250	5.1250
41 Oficinistas	compe1188 Usar varios paquetes de software de computadora, incluidas hojas de cálculo, para brindar apoyo administrativo	0.1333	0.3333	2.5000
compe568 Gestionar el correo entrante o saliente	41 Oficinistas	0.1333	0.3333	2.500
412 Secretarios (general)	compe1188 Usar varios paquetes de software de computadora, incluidas hojas de cálculo, para brindar apoyo administrativo	0.1333	1.000	7.5000
compe568 Gestionar el correo entrante o saliente	412 Secretarios (general)	0.1333	1.0000	7.5000
compe123 Ayudar con los registros y cálculos de contabilidad y teneduría de libros	43 Empleados contables y encargados del registro de materiales	0.13333	0.4000	3.0000
compe136 Calcular los costos unitarios de producción	43 Empleados contables y encargados del registro de materiales	0.1333	0.4000	3.0000
compe762 Obtención, compilación y computación de datos estadísticos o actuariales	43 Empleados contables y encargados del registro de materiales	0.1333	0.4000	3.0000
...

Tabla 5-12.: Ejemplo de reglas de propiedad que relaciona una instancia con dominio Ocupación y rango Competencia

La regla que relaciona un elemento con dominio Competencia y rango Competencia. Esto podría significar que son competencias que están estrechamente relacionadas y que es muy

común que si una ocupación tiene una de ellas, también tiene la otra competencia (ver Tabla 5-13).

regla	asociado con	soporte	confianza	lift
compe838 Planificar y dirigir las operaciones diarias	compe1139 Supervisar la selección, formación y rendimiento del personal	0.1219	0.7142	2.9285
compe568 Gestionar el correo entrante o saliente	compe1188 Usar varios paquetes de software de computadora, incluidas hojas de cálculo, para brindar apoyo administrativo	0.1333	1.0000	7.5000
compe123 Ayudar con los registros y cálculos de contabilidad y teneduría libros	compe136 Calcular los costos unitarios de producción	0.1333	1.0000	7.5000
compe762 Obtención, compilación y computación de datos estadísticos o actuariales	compe123 Ayudar con los registros y cálculos de contabilidad y teneduría de libros	0.1333	1.0000	7.5000
compe762 Obtención, compilación y computación de datos estadísticos o actuariales	compe136 Calcular los costos unitarios de producción	0.1333	1.0000	7.5000
...

Tabla 5-13.: Ejemplo de reglas propiedad que relaciona una instancia con dominio Competencia y rango Competencia

5.6. Análisis y conclusiones del capítulo

En este capítulo se cumplió el objetivo 3 de la tesis: Implementar el modelo conceptual de aprendizaje ontológico con técnicas combinadas de aprendizaje que permita extraer desde los diferentes campos de la red social profesional las clases, jerarquía de clases y restricciones, con el fin de validar la utilidad del modelo de aprendizaje ontológico propuesto.

En general, los hallazgos de la implementación del modelo conceptual de aprendizaje ontológico propuesto en el capítulo anterior fueron positivos. En la aplicación de cada etapa se observaron resultados coherentes a lo que se pretendía lograr o aprender. De 972 datos extraídos con el *web scraping* se agrupan por título obteniendo 389 posibles ocupaciones, de las cuales con análisis de frecuencia de frases se seleccionan las 50 más frecuentes para ingresar al modelo de aprendizaje. Se encontraron dos (2) correspondencias exactas para competencias. Se extrajeron 3 etiquetas alternativas para ocupaciones y 7 para competencias usando *Transformers*. Se clasificaron también con esta misma técnica 148 competencias en básicas, 550 competencias en específicas y 931 competencias entre las 3 clases de transversales. Se enriqueció el grafo de conocimiento de 4.116 tripletas a 10.126. Se encontraron nuevas 15 clases dentro de la jerarquía de clases de las ocupaciones. Y finalmente, se extrae el dominio y rango para las propiedades `es_competencia_de` y `es_competencia`.

Con los campos requeridos delimitados en el diseño, fue posible modelar el caso de uso particular de LinkedIn, demostrando que la información requerida por el modelo de aprendizaje se puede encontrar con los campos definidos en oferta laboral online de una red profesional.

La estructura *SentenceTransformers* de Python se usó para definir el *Transformer* SBERT para realizar similitud semántica en diferentes etapas de la implementación del modelo, permitió extraer 10 etiquetas alternativas, clasificar 1629 competencias y extraer 6.010 tripletas de relación semántica para enriquecer el grafo.

Con el modelo KGE con HolE se extrajeron nuevas 15 clases dentro de la jerarquía de clases existentes en la ontología.

Con el algoritmo apriori se encontró el dominio y rango de las propiedades `tiene_competencia` y `es_competencia_de`.

6. Capítulo 6: Evaluación de la ontología

Durante el desarrollo de una ontología, se espera tener un uso correcto del lenguaje, tener exactitud en la estructura taxonómica y que los conceptos implementados tengan un significado dentro del dominio.

En este capítulo se extraen las métricas importantes para evaluar la nueva ontología. En la Figura 6-1 se representa de forma gráfica y general lo que se pretende en este capítulo. Se tomará como entrada la ontología aprendida y se aplicarán algunos algoritmos, técnicas y métodos para evaluar la ontología y extraer las métricas correspondientes. Finalmente, la nueva ontología se utilizará para la validación del caso de uso en el siguiente capítulo.

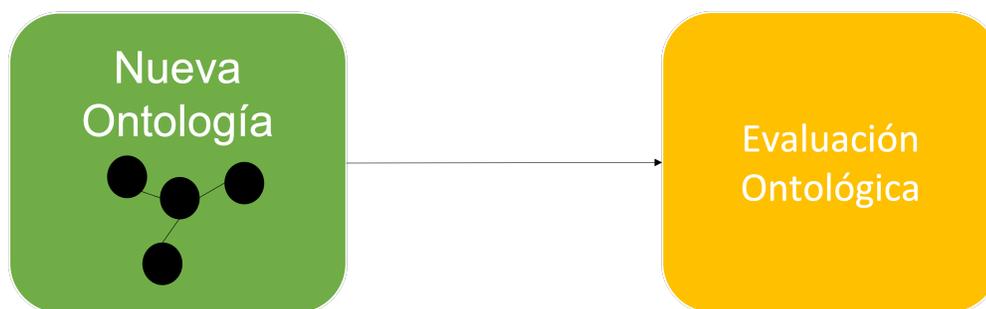


Figura 6-1.: Representación gráfica del objetivo del capítulo 6

En la Figura 6-2 se muestra la metodología para el desarrollo de este capítulo.

El proceso de evaluación de ontologías consiste en la presentación de un informe técnico del contenido de una ontología con respecto a un marco de referencia (requerimientos, preguntas de competencia, entre otros) en cada fase del ciclo de vida. La evaluación considera la verificación y validación de la ontología; en la primera se chequea la construcción correcta, es decir, que las definiciones implementen los requerimientos y den respuestas a las preguntas de competencia preestablecidas. La validación se refiere a que las definiciones de la ontología modelen lo más exactamente posible el dominio para el cual fueron creadas. Más específicamente, la evaluación de una ontología consiste en determinar si ésta satisface los criterios de diseño preestablecidos [Ramos et al., 2009].

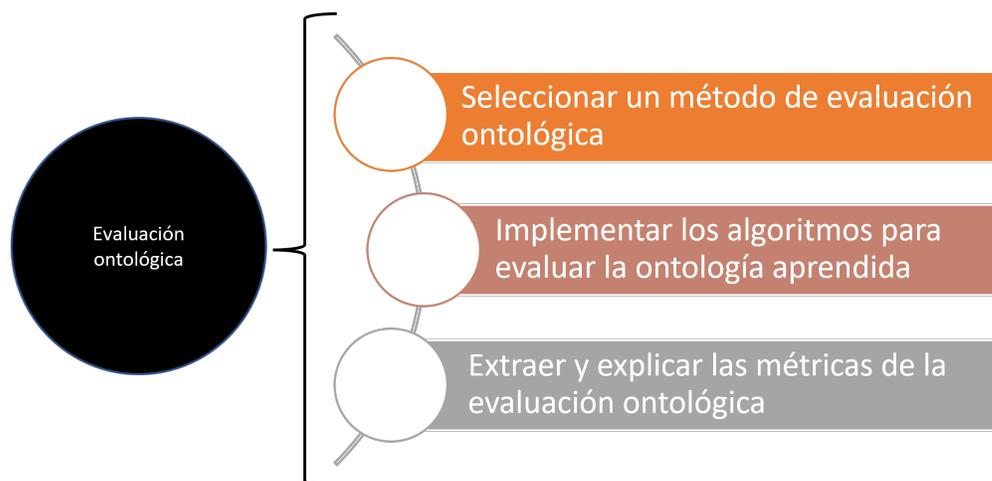


Figura 6-2.: Metodología de desarrollo objetivo 5

La mayoría de las propuestas para evaluar ontologías encontradas en la literatura coinciden en que se deben evaluar los siguientes criterios (ver Figura 6-3) :

- La rigurosidad taxonómica
- El lenguaje utilizado para la codificación
- El rendimiento de las aplicaciones o tareas que utilizan las ontologías
- El vocabulario utilizado para representar los conceptos y relaciones del dominio modelado

Muchos de estos criterios se evalúan sobre la base de la comparación con otras ontologías disponibles las cuales se usan como referencia o estándar de oro (*Golden standard*) [Ramos et al., 2009].

La evaluación del aprendizaje de ontologías permite refinar y remodelar todo el proceso de aprendizaje de ontologías en caso de ontologías resultantes inesperadas, que no se ajustan a los requisitos específicos de un usuario.

El aprendizaje de ontologías es un proceso de varios niveles, por lo que el proceso de evaluación de la extracción de ontologías es bastante difícil. Teniendo en cuenta la complejidad de evaluar ontologías de dominio, se han propuesto innumerables técnicas de evaluación en los últimos años y esta área aún se encuentra en continuo desarrollo [Asim et al., 2018a].

Según la literatura se proponen cuatro (4) formas principales de hacer evaluación ontológica: la evaluación basada en el estándar de oro (*Golden standard*) que se trata de evaluar

Criterios de evaluación de ontologías

Criterio Autor	Taxonomía	Lenguaje	Aplicación	Vocabulario	Arquitectura Requerimientos	Aceptación Social	Razonamiento Automático	Software
(1)	√	-	√	√	-	-	-	-
(2)	-	√	√	√	-	-	√	-
(3)	√	-	√	√	-	-	-	-
(4)	-	√	√	√	-	√	-	-
(5)	√	√	√	√	√	-	-	-
(6)	√	√	√	√	√	-	-	√

(1) Brewster y cols, 2004 (2) Obrst y cols, 2007 (3) Porzel y Malaka, 2004
 (4) Burton-Jones y cols, 2005 (5) Brank y cols, 2005 (6) Lozano-Tello, 2002 √ Considera - No considera

Figura 6-3.: Criterios de evaluación ontológica, tomado de [Ramos et al., 2009]

la ontología resultante con un punto de referencia predefinido o una ontología estándar; la evaluación basada en aplicaciones que es una evaluación orientada a tareas y aplicaciones, ya que evalúa una ontología determinada al explotarla en una aplicación específica para realizar alguna tarea; la evaluación basada en datos denominada basada en corpus que utiliza fuentes de conocimiento específicas del dominio existentes (generalmente corpus textuales) para evaluar el alcance de la cobertura por ontología específica en un dominio particular; y finalmente, la evaluación humana que se basa en la definición y formulación de varios criterios de decisión para la selección de la mejor ontología de un conjunto específico de ontologías candidatas [Asim et al., 2018a].

La Figura 6-4 presenta una descripción general de los enfoques de evaluación de ontologías frente a varios niveles de evaluación soportados del aprendizaje de ontologías.

Level	Golden standard	Application-based	Data-driven	Assessment by humans
Lexical, vocabulary, concept and data	x	x	x	x
Hierarchy and taxonomy	x	x	x	x
Other semantic relations	x	x	x	x
Context and application		x		x
Syntactic	x			x
Structure, architecture and design				x

Figura 6-4.: Descripción general de los enfoques de evaluación de ontologías, tomado de [Asim et al., 2018a]

Teniendo en cuenta lo anterior y buscando un método de evaluación automática se toma para esta tesis la aplicación de la evaluación basada en el estándar de oro (*Golden standard*). En el capítulo 3 de esta tesis se documentó sobre las ontologías más usadas en el dominio de *e-recruitment* concluyendo que la ontología ESCO es la más completa, actualizada y la más usada por otros autores para trabajos ontológicos en el tema de contratación de personal en línea. Por lo anterior, se toma la ontología ESCO como la (*Golden ontology*) para comparar la ontología aprendida.

Por su parte, en la Figura 6-5 se muestra un diagrama general sobre validadores relevantes de ontologías en el dominio de IOT encontrados en la literatura [Coggle, 2021]. Esto con el fin de validar alguno de sus usos en algunos de los criterios de evaluación.

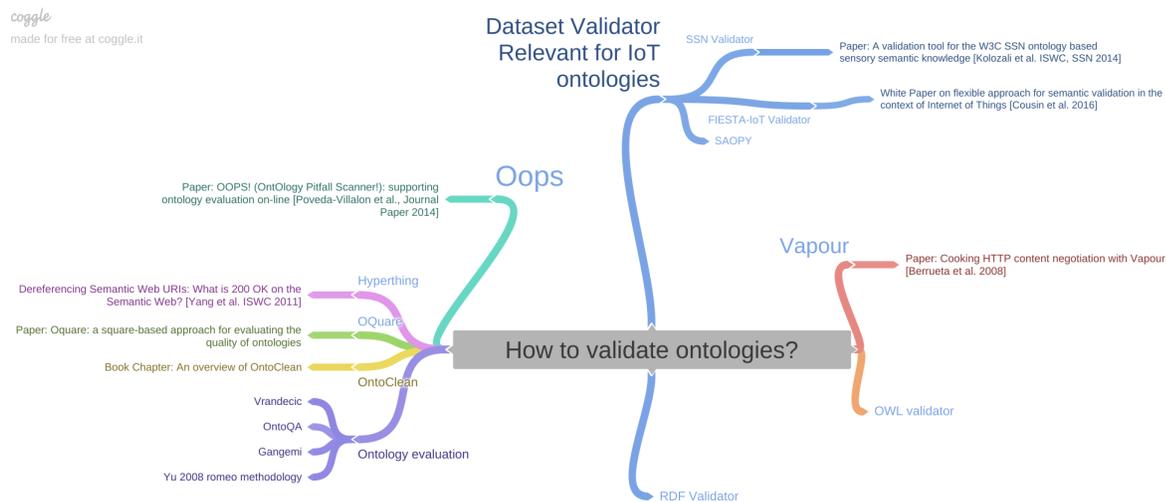


Figura 6-5.: Validadores de ontologías en el dominio de IOT, tomado de [Coggle, 2021]

6.1. Evaluación del criterio 1: Lenguaje

El uso adecuado del lenguaje permitirá evaluar, con base en las características y reglas de construcción del lenguaje usado, la codificación de la ontología. Además, evaluará las palabras en el lenguaje (en este caso español).

Se espera encontrar que los términos usados en la ontología estén libres de errores y defectos, es decir, que la extracción para el dominio haya sido exitosa.

6.1.1. Extracción del dominio y codificación

Para evaluar el lenguaje de las palabras con las que se construyó la ontología, es necesario hacer hincapié en que la extracción se realizó con base en la red profesional LinkedIn, con la consulta específica de trabajos en Tecnología de la Información. Esto nos permite afirmar con certeza que la extracción pertenece al dominio de TI.

Por su parte, para validar la codificación de la ontología final construida y aprendida, se usa el *OWL validator open source* en línea creado por la Universidad de Manchester en 2009 (<http://mowl-power.cs.man.ac.uk:8080/validator/>) (OWL API Versión 3.4.5-SNAPSHOT). En la Figura 6-6 se muestra el resultado arrojado una vez se ingresa el código OWL de la ontología en el validador.



Figura 6-6.: Resultado del validador OWL

El resultado permite verificar que la conformación sintáctica de la ontología aprendida es correcta en OWL2, OWL2 DL, OWL2 EL, OWL2 QL y OWL2 RL, bajo la Manchester OWL syntax (<https://www.w3.org/TR/owl2-manchester-syntax/>), DL syntax y Functional syntax (<https://www.w3.org/TR/owl2-syntax/>). Esto hace que la ontología construida sea una ontología reutilizable y escalable.

OWL2 DL, OWL2 EL, OWL2 QL y OWL2 RL son versiones reducidas de OWL 2 que intercambian algo de poder expresivo por la eficiencia del razonamiento. Cada uno de ellos logra la eficiencia de una manera diferente y es útil en diferentes escenarios de aplicación. El OWL

2 EL es particularmente útil en aplicaciones que emplean ontologías que contienen un gran número de propiedades y/o clases, los algoritmos de razonamiento dedicados para este perfil están disponibles y se ha demostrado que se pueden implementar de una manera altamente escalable. El OWL 2 QL está dirigido a aplicaciones que utilizan grandes volúmenes de datos de instancia y donde la respuesta a consultas es la tarea de razonamiento más importante. El OWL 2 RL está dirigido a aplicaciones que requieren un razonamiento escalable sin sacrificar demasiado poder expresivo. El OWL 2 DL se asigna a lógicas de descripción (DL) que son un subconjunto de la lógica de predicados para la cual es posible un soporte de razonamiento eficiente [W3C, 2012].

6.1.2. Evaluación de las palabras en el lenguaje español

Aquí se realiza una validación de si las palabras usadas en la ontología pertenecen al idioma español y si están escritas adecuadamente. Para lo anterior, se usa el modelo de *FastText embeddings from SUC*, donde se emplean diferentes corpus del idioma español para realizar una incrustación de palabras [Cañete et al., 2020].

Al listar todas las palabras que pertenecen a la ontología construida, encontramos que tiene 23.334 palabras. De estas palabras, 3.031 son palabras únicas. Se busca si estas 3.031 pertenecen o no al corpus. Se encuentra que hay 2.725 palabras que si pertenecen al corpus y 306 palabras que no pertenecen al mismo.

Algunas de las palabras encontradas que no pertenecen al corpus son: switches, ssis, hse, vsm, insumo-producto, world, programarlas, hq, scrum, owap, wco, boosting, vmwarevdi, datadriven, pytorch, tsql, etc. Muchas de estas palabras son palabras en otro idioma, son abreviaturas, son una metodología, etc.

Entonces se puede decir que, el **89.9 %** de las palabras de la ontología pertenecen al lenguaje español y el otro **10.1 %** restante no pertenece al idioma español. Sin embargo, la mayoría de las palabras del 10.1 % restante sí pertenecen al dominio, pues describen alguna metodología, técnica, sistema, etc., asociadas a perfiles profesionales de TI.

6.2. Evaluación del criterio 2: Vocabulario

Para evaluar el vocabulario utilizado para representar los conceptos y las relaciones del dominio modelado, se requiere obtener el significado de los términos y conceptos a partir del conocimiento de expertos, recopilaciones de textos o cualquier otra fuente de conocimiento disponible del dominio. En este caso, se realizará a través de la comparación con la *Golden*

Ontology (ESCO).

Se tendrán en consideración las medidas de calidad de resultados usadas en escenarios de recuperación de información (búsqueda de documentos), tales como la precisión, el *recall* (exhaustividad) y el *f1-score*. Los insumos para esta actividad será ontología aprendida y la ontología ESCO.

Precisión: porcentaje de los términos de la ontología aprendida que aparecen en la ESCO con relación a la cantidad total de términos de la ontología aprendida [Ramos et al., 2009]. En la Ecuación 6-1 se indica como se calcula la precisión.

$$Precisión = \frac{CO - C}{C_{onto}} \quad (6-1)$$

CO-C es la cantidad de términos que se solapan entre la ontología aprendida y la ESCO.

C_{onto} = Cantidad total de términos de la ontología aprendida.

Recall: porcentaje de términos de la ESCO que aparecen en la ontología aprendida con relación al total de términos en la ESCO [Ramos et al., 2009]. En la Ecuación 6-2 se indica como se calcula el *recall*.

$$Recall = \frac{CO - C}{CCorp} \quad (6-2)$$

CCorp es la cantidad total de términos en la ESCO.

En función de los valores obtenidos para la precisión y el recall, se establece una valoración cualitativa acerca de lo adecuado del vocabulario.

El valor **f1-score** combina la medida de precisión y recall en un sólo valor, lo que hace más fácil el poder comparar el rendimiento combinado de la precisión y la exhaustividad. En la Ecuación 6-3 se indica como se calcula el f1-score.

$$F1 - score = 2 \cdot \left(\frac{Precisión * Recall}{Precisión + Recall} \right) \quad (6-3)$$

Ahora bien, para extraer las medidas de calidad se realiza la comparación de ontología aprendida con la ontología oro ESCO a través de *LogMap*. *LogMap* es un sistema de comparación de ontologías altamente escalable que extrae información de comparación entre clases, propiedades e instancias. Este sistema permite realizar un emparejamiento semántico entre las

ontologías e incorpora técnicas de razonamiento y reparación de ontologías para minimizar el número de inconsistencias [Jiménez-Ruiz et al., 2012].

LogMap puede ser usado como un sistema de emparejamiento de ontologías, como un sistema de depuración de mapeo o como un sistema de alineación de ontologías. En este caso usamos una versión independiente o *StandAlone* con el fin de comparar la ontología construida con la ontología ESCO.

Este sistema de *LogMap* está desarrollado en java y se ejecuta con el comando MATCHER. Sus parámetros son la ontología ESCO y la ontología aprendida de TI, esto para realizar la comparación entre las ontologías. Debido a que no se puede acceder a la ontología ESCO en un sólo URI, este comando se debe ejecutar por cada una de las URIs disponibles para poder recorrer las clases e instancias de la ontología.

La comparación constó de 3.551 ejecuciones, con una duración de 452 minutos (7.5 horas). En la Tabla 6-1 se muestra la extracción de una porción del resultado de las ejecuciones.

file	esco	onto OPTI	value	type
f9	http://data.europa.eu/esco/isco/C11	http://ti-2023-06-2418:12:32.586072.org/onto#11	0.99	INST
f9	http://data.europa.eu/esco/isco/C11	http://ti-2023-06-2418:12:32.586072.org/onto#11	0.99	INST
f15	http://data.europa.eu/esco/isco/C112	http://ti-2023-06-2418:12:32.586072.org/onto#112	1.0	INST
f16	http://data.europa.eu/esco/isco/C1120	http://ti-2023-06-2418:12:32.586072.org/onto#112	1.0	INST
f1712	http://data.europa.eu/esco/occupation/9f508305-80ce-4111-8722-f2c9b4a44890	http://ti-22023-06-2418:12:32.586072.org/onto#112a	0.87	INST

Tabla 6-1.: Porción o extracción de resultado de las ejecuciones de *LogMap*

Se puede observar como la evaluación ontológica evalúa una clase de la ontología aprendida de TI (OPTI) y muestra la instancia que más se parece en la ontología ESCO. En total las coincidencias entre clases e instancias, según los resultados de LogMap, es de **0.97**.

Para las ocupaciones, la ontología aprendida de TI (OPTI) tiene 188 clases, la ontología ESCO tiene 3.551 instancias. Se encontró que tienen 185 instancias correspondientes. En la Tabla **6-2** se muestra el cálculo de las medidas de calidad correspondientes.

Medida de Calidad	Cálculo	Resultado
Precisión	$(185 / 188) * 100$	98.4 %
Recall	$(185 / 3551) * 100$	5.2 %
F1-score	$(2 (0.984)(0.052) / ((0.984) + (0.052))) * 100$	4.9 %

Tabla 6-2.: Cálculo de medidas de calidad ontología aprendida vs ESCO (ocupaciones)

Para las competencias, la ontología aprendida de TI (OPTI) tiene 1948 clases, la ontología ESCO tiene 13958 instancias. Se encontró que tienen 1556 instancias correspondientes. En la Tabla **6-3** se muestra el cálculo de las medidas de calidad correspondientes.

Medida de Calidad	Cálculo	Resultado
Precisión	$(1556 / 1948) * 100$	79.9 %
Recall	$(1556 / 13958) * 100$	11.1 %
F1-score	$(2 (0.799)(0.111) / ((0.799) + (0.111))) * 100$	9.8 %

Tabla 6-3.: Cálculo de medidas de calidad ontología aprendida vs ESCO (competencias)

Podemos ver que la ontología construida de TI (OPTI) tiene mucha diferencia a la ontología ESCO en cuanto al número de clases e instancias tanto en competencias como en ocupaciones, sin embargo, podemos concluir que la medida de **precisión** muestra que la ontología aprendida de TI es coherente con la ontología ESCO, también que la mayoría de las clases de ocupaciones y competencias de la ontología de TI están en la ontología ESCO. Esto nos permite asegurar que vocabulario del dominio con el que se construyó la ontología de TI es correcto y validado por la ontología ORO.

6.3. Evaluación del criterio 3: Taxonomía

La exactitud de la extracción taxonómica se puede realizar también a través del uso de una *golden standard*, es decir, comparando la ontología construida con una ontología existente. Para ello se seguirá usando la ontología oro ESCO. Se analiza la relación jerárquica y la relación semántica.

6.3.1. Exactitud de la extracción taxonómica

Para el criterio de exactitud de la extracción taxonómica se realiza nuevamente la comparación de la ontología aprendida con la ontología oro ESCO. Se analiza la relación jerárquica y semántica.

De 188 análisis correspondientes a la superclase Ocupación, 178 tienen la misma clasificación tanto en la ontología aprendida como en la ESCO y 10 no corresponden a la misma clasificación. Así pues, se puede decir que el **94.6 %** de los datos de esta superclase corresponden taxonómicamente con la estructura de la ontología oro, mientras que el **5.31 %** no corresponde taxonómicamente con la estructura ESCO.

Para la superclase Competencias no es posible realizar este análisis debido a que ESCO y OPTI tienen una estructura inicial en sus competencias como se muestra en el capítulo 3 en la Figura 3-11. Para poder evaluar esta superclase para la ontología OPTI, se podría proponer acudir a un grupo de expertos que validara dicha clasificación.

6.4. Evaluación del criterio 4: Aplicación

Una ontología de calidad permitirá que las tareas y aplicaciones de software que la utilicen ofrezcan resultados exitosos. Las evaluaciones basadas en tareas y aplicaciones ofrecen un marco útil para medir aspectos prácticos del despliegue de la ontología, como la capacidad humana para formular consultas utilizando el lenguaje de consulta proporcionado por la ontología, la precisión de las respuestas proporcionadas por el sistema, el grado de capacidad de explicación que ofrece el sistema, la cobertura de la ontología en términos del grado de reutilización entre dominios, la escalabilidad de la base de conocimientos y la facilidad de uso del componente de consulta. Estas evaluaciones basadas en tareas tienen en cuenta los requerimientos para los cuales se construye la ontología y los resultados en una aplicación (escenario o caso de uso). El capítulo 7 mostrará de forma más detallada el despliegue de la aplicación para poner a prueba la ontología aprendida.

6.5. Análisis y conclusiones del capítulo

En este capítulo se cumplió el objetivo 5 de la tesis: Evaluar los resultados de la implementación del modelo de aprendizaje ontológico por medio de técnicas o métricas de evaluación de aprendizaje de ontologías, con el fin de verificar los cambios sobre la ontología base.

Se valida que la construcción de la ontología final en cuanto al lenguaje o codificación es correcta a través del OWL en el validador online de Manchester. Contar con esta validación es importante para garantizar que la ontología sea escalable, interoperable y reutilizable.

En la evaluación de las palabras en el lenguaje español se encontró que el **89.9 %** de las palabras de la ontología pertenecen al lenguaje español y el otro **10.1 %** restante no pertenece al idioma español. Sin embargo, la mayoría de las palabras del 10.1 % restante sí pertenecen al dominio, pues describen alguna metodología, técnica, sistema, etc., asociadas a perfiles profesionales de TI.

La ESCO es una ontología muy grande en comparación a la ontología conformada en este trabajo. Es por ello, que cuando se evalúa el vocabulario se obtiene una precisión del 98.4 % para las clases aprendidas de ocupaciones y 79.9 % para las clases aprendidas de competencias, es decir, que casi todos los conceptos de la ontología aprendida existen y coinciden con alguno en a ontología ESCO. Pero debido a que el recall analiza que tanto la ESCO se parece a la ontología aprendida su valor disminuye notablemente a un valor de 5.2 % y 11.1 % para ocupaciones y competencias, respectivamente.

Se encontró una correspondencia taxonómica del 94.6 % de las clases nuevas de ocupaciones con la ontología ESCO. Para la superclase Competencias no es posible realizar este análisis debido a que ESCO y OPTI tienen una estructura diferente en sus competencias.

7. Capítulo 7: Aplicación de la ontología en el dominio de e-recruitment a través de casos de uso

En este capítulo se completará la evaluación de la ontología final aprendida a través del desarrollo de casos de uso para *e-recruitment*. Se pretende poner a prueba la ontología midiendo sus aspectos prácticos en escenarios de contratación en línea (*e-recruitment*).

Teniendo en cuenta que el activo más importante de cualquier empresa u organización es tener a su disposición los empleados y el talento adecuado para desempeñar sus funciones de forma exitosa, pone al área de contratación o recursos humanos en un escenario de responsabilidad interesante. [Sołek-Borowska and Wilczewska, 2018] afirman que el reclutamiento y la selección de candidatos dependen de dos factores estratégicos: la velocidad y la precisión. Con un número cada vez más alto de solicitantes altamente cualificados por puesto de trabajo, los profesionales de recursos humanos se encuentran ante el desafío de clasificar solicitudes rápidamente, asegurándose simultáneamente de que seleccionan a los mejores candidatos.

Lo anterior es un desafío aún mucho más riguroso en perfiles asociados con tecnologías de la información de la comunicación, debido a que en los últimos años ha estado creciendo, modificándose y renovándose en sus cargos y habilidades.

Los sistemas de contratación electrónica o *e-recruitment* han experimentado una expansión exponencial durante las últimas décadas, ya que permiten a las agencias y departamentos de Recursos Humanos dirigirse a un público mucho más amplio a un coste muy reducido [Arroyo Mesa, 2022]. Las empresas han tenido que adaptarse a los nuevos métodos rápidamente emergentes de contratación en línea.

Las características más importantes del *e-recruitment* son:

Del canal de comunicación

- Mayor alcance de candidatos
- Capacidad de proporcionar gran cantidad de información

- Acceso 24/7/365

De la técnica de *recruiting*

- Rápidez
- Economicidad
- confidencialidad

Del canal de comunicación

- Igualdad de oportunidades
- Posibilidad de observar el mercado

Históricamente, el departamento de recursos humanos ha tenido que dedicar mucho tiempo a realizar tareas repetitivas como el filtrado de grandes volúmenes de solicitudes de candidatos. Gracias a la incorporación de herramientas de inteligencia artificial (IA) y la analítica de datos se han podido agilizar muchos procesos como la selección de candidatos mediante algoritmos predictivos, la programación de entrevistas, la comunicación directa con los candidatos mediante correos automatizados o *chatbots* y la facilitación de programas de incorporación, entre otros [Ross, 2021].

Las diferentes investigaciones apuntan a que la IA y la analítica de datos aportan en diferentes procesos de *e-recruitment*. En primer lugar, en la **Identificación de candidatos** a través de herramientas de software que permitan seleccionar currículums a partir de bases de datos de la empresa y seleccionar los mejores perfiles para que los responsables del proceso los examinen. En segundo lugar, en el **Candidate Experience** para agilizar la comunicación con los solicitantes respondiendo a la necesidad de mantener el compromiso del candidato con la empresa. Y en tercer y último lugar, la **Selección de candidatos** a través del análisis de los currículums submitidos por los candidatos e indicar qué candidato es el más apto para cubrir el puesto vacante.

A continuación, se explica detalladamente las herramientas de software desarrolladas a través de Python como casos de uso para probar la ontología en el dominio. Este trabajo apunta a dar respuesta a las características **De la técnica de *recruiting*** y a los procesos de **Identificación de candidatos** y **Selección de candidatos**.

7.1. Caso de uso 1

Esta aplicación consiste en hacer uso de la ontología aprendida para verificar sus aspectos técnicos a través de las respuestas que emite cuando se le asignan tareas por medio de requerimientos para una aplicación específica.

El escenario que se plantea en este primer caso de uso es el de una empresa que desea buscar una persona para ocupar un puesto de trabajo dentro de su organización. Debido a que los perfiles de TI y sus ocupaciones están en un constante cambio, se usará la ontología aprendida para encontrar la ocupación que satisface el puesto requerido.

“Definir correctamente un perfil de puesto es primordial para encontrar al ejecutivo indicado para determinada empresa. Este proceso requiere, entre otras cosas, identificar las competencias necesarias, la personalidad y el tipo de liderazgo que la empresa requiere” [HunterAmericas, 2023].

De acuerdo con los parámetros del *recruiting* una de las cosas más importantes cuando se define un perfil es tener presente el conjunto de habilidades, los rasgos de personalidad y las actitudes que coinciden con la vacante. Cuando las empresas no realizan una buena definición del cargo que necesita, pasa por alto las habilidades que realmente necesita el perfil o simplemente desconoce el cargo que requiere para cumplir cierta función, generalmente comete el error de contratar perfiles que no corresponden con el adecuado. Lo anterior produce dentro de las empresas contrataciones incorrectas, bajas en la productividad, necesidad de reemplazar al ejecutivo dentro de poco tiempo, volver a empezar el proceso de selección y acoplamiento. El resumen de lo anterior hace hincapié en pérdida de tiempo y recursos.

Esto pasa en cargos como los de perfiles de TI, que constantemente se actualizan y se transforman, incluso, emergen nuevos cargos y habilidades importantes.

Para el desarrollo de este caso de uso, se toma como entrada de la aplicación la lista de competencias que necesita una empresa para una vacante dada. Seguido a esto, se consulta en la ontología que ocupaciones se acercan más a las competencias buscadas.

Para el desarrollo del aplicativo se usa nuevamente la librería de *Owlbready2* para *Python*, específicamente las consultas realizadas con SPARQL. Desde la versión 0.30, *Owlbready2* propone 2 métodos para realizar consultas SPARQL: el motor SPARQL nativo y *RDFlib*. El motor SPARQL nativo traduce automáticamente las consultas SPARQL en consultas SQL y luego ejecuta las consultas SQL con *SQLite3*. Es importante definir que para esta aplicación se usa el motor SPARQL nativo que tiene un mejor rendimiento que *RDFlib* (alrededor de 60 veces más rápido) [LAMY, 2019].

Gracias a que la librería permite realizar consultas SPARQL, se implementa un algoritmo que permita recorrer una lista de competencias para poder encontrar las competencias que contengan la mayor cantidad de dichas ocupaciones.

En la Lista 7.1 se muestra un ejemplo de la consulta.

```

/*****/
result = graph.query("""
prefix opti: <http://OPTI.org/onto#>

SELECT
    ?x
    ?y
    ?z
    ?w
WHERE {
    ?x opto:tiene_competencia ?y .
    ?y rdfs:label ?z .
    ?x rdfs:label ?w .
    FILTER regex(?z, "(Manejo de sql server para
    realizacion de consultas)")
}
""")
/*****/

```

Listing 7.1: Consulta SPARQL para informe perfil técnico TI

De la consulta se puede ver que se declara como prefijo la URI de la ontología, esto para seleccionar la ontología OPTI como fuente de consulta.

La variable ?x contiene el URI de todas las ocupaciones encontradas. La variable ?y contiene el URI de las competencias relacionadas. La variable ?z tiene el nombre (label) de la ocupación y la variable ?w tiene el nombre de la competencia. La consulta también tiene un filtro, donde sólo trae los registros donde la ocupación tenga relacionada la competencia "Manejo de sql server para realización de consultas".

Para evaluar diferentes competencias, se realiza la consulta para cada competencia y se unen los resultados, como se muestra en la Lista 7.2. Esto se hace con un algoritmo que permita construir la consulta para todas las competencias.

```
/*
SELECT
    ?x
    ?y
    ?z
    ?w
WHERE {
    {
        ?x opto:tiene_competencia ?y .
        ?y rdfs:label ?z .
        ?x rdfs:label ?w .
        FILTER regex(?z, "(Manejo de sql server para realizacion
de consultas)")
    }
    UNION
    {
        ?x opto:tiene_competencia ?y .
        ?y rdfs:label ?z .
        ?x rdfs:label ?w .
        FILTER regex(?z, "(control de la arquitectura de
las computadoras)")
    }
}
*/
```

Listing 7.2: Consulta SPARQL para informe perfil administrativo TI

El resultado de la consulta se maneja como un *dataframe* de la librería de Pandas. En la Tabla 7-1 se muestra una porción del resultado de la consulta.

Después de tener la consulta, se implementa un *GroupBy* por id de la ocupación, y se realiza el conteo de cuantas competencias tiene relacionada cada ocupación. A su vez, se calcula el valor porcentual de las relaciones entre el número de las competencias buscadas y el número de competencias encontradas para cada ocupación, esto es, $\text{num_ocu_encontradas}/\text{num_ocu_buscadas} * 100$.

Finalmente, se ordena el *dataframe* de mayor a menor número de competencias encontradas y se seleccionan las 10 primeras para generar el reporte. Con esta información se procede a

id_ocu	id_compe	ocupación	competencia
http://OPTI.org/onto#251b	http://OPTI.org/onto#c0011156	analista requerimientos tecnología informacion	Manejo de sql server para realizacion de consultas
http://OPTI.org/onto#251b	http://OPTI.org/onto#c0011159	analista requerimientos tecnología informacion	Metodología scrum
http://OPTI.org/onto#2511	http://OPTI.org/onto#c0011164	Analistas de sistemas	La planificación la proyección la producción y el control de la arquitectura de las computadoras
http://OPTI.org/onto#2512	http://OPTI.org/onto#c0011164	Desarrolladores de software	La planificación la proyección la producción y el control de la arquitectura de las computadoras
http://OPTI.org/onto#2513	http://OPTI.org/onto#c0011164	Desarrolladores Web y multimedia	La planificación la proyección la producción y el control de la arquitectura de las computadoras

Tabla 7-1.: Porción o extracción del resultado de la consulta

construir el informe.

A continuación, se definen las entradas de competencias para las 2 consultas de prueba del caso de uso. Obviamente las 2 consultas se realizan para habilidades de perfiles de TI, uno se hace para un perfil técnico y la otra para un perfil administrativo.

Para la consulta 1 se define como entrada las siguientes competencias: Manejo de *sql server* para realización de consultas, Metodología scrum, Control de la arquitectura de las computadoras, Mantener actualizados los programas lógicos, Preparar gráficos y diagramas sistemáticos para describir en términos lógicos, Desarrollar por escrito y Proporcionar documentación detallada sobre los programas informáticos.

Para la consulta 2 se define como entrada las siguientes competencias: Controlar la selección formación y rendimiento del personal, Desempeñar tareas afines, y Supervisar a otros trabajadores. Después de construir la consulta automáticamente y ejecutar el código, se obtiene el informe que se muestra a continuación.

En la Lista 7.3 y en la Lista 7.4 se muestran los informes arrojado por las consultas SPARQL en Python.

```
ONTOLOGIA DE PERFILES EN TECNOLOGIA DE LA INFORMACION (OPTI)
-----
Informe Generado
-----
Competencias objetivo :
- Manejo de sql server para realizacion de consultas
- Metodologia scrum
- Control de la arquitectura de las computadoras
- Mantener actualizados los programas logicos
- Preparar graficos y diagramas sistematicos para
describir en terminos logicos
- Desarrollar por escrito y proporcionar documentacion
detallada sobre los programas informaticos
-----
Ocupaciones con mayor coincidencias

1.      50.0%   Programadores de aplicaciones
2.      33.3%   analista requerimientos tecnologia informacion
3.      16.7%   Analistas de sistemas
4.      16.7%   Desarrolladores de software
5.      16.7%   Desarrolladores Web y multimedia
6.      16.7%   Desarrolladores y analistas de software
y multimedia y analistas no clasificados bajo otros epigrafes
7.      16.7%   Diseñadores y administradores de bases de datos
8.      16.7%   Administradores de sistemas
9.      16.7%   Profesionales en redes de computadores
10.     16.7%   Especialistas en bases de datos
y en redes de computadores no clasificados bajo
otros epigrafes
-----
URI: http://OPTI.org/onto#
-----
Fecha de generacion de informe: 2023-01-17 03:49:44.755128
```

Listing 7.3: Informe para consulta SPARQL perfil técnico de TI

```
ONTOLOGIA DE PERFILES EN TECNOLOGIA DE LA INFORMACION (OPTI)
-----
Informe Generado
-----
```

```
Competencias objetivo :
- Controlar la seleccion formacion y rendimiento del personal
- Desempenar tareas afines
- Supervisar a otros trabajadores
-----
Ocupaciones con mayor coincidencias

1.      100.0%  Directores financieros
2.      100.0%  Directores de recursos humanos
3.      100.0%  Directores de administracion y
servicios no clasificados bajo otros epigrafes
4.      100.0%  Directores de publicidad y relaciones publicas
5.      100.0%  Directores de investigacion y desarrollo
6.      100.0%  Gerentes de sucursales de bancos
de servicios financieros y de seguros
7.      66.7%   Directores de servicios de cuidados infantiles
8.      66.7%   Supervisores de la construccion
9.      66.7%   Directores de cine de teatro y afines
10.     66.7%   Gerentes de servicios no clasificados
bajo otros epigrafes
-----
URI: http://OPTI.org/onto#
-----
Fecha de generacion de informe: 2023-01-17 02:01:54.410602
```

Listing 7.4: Informe para consulta SPARQL perfil administrativo de TI

En los dos informes se muestran las primeras 10 ocupaciones con mayor coincidencia con las competencias requeridas en la vacante. El porcentaje dado a cada ocupación se determina a través de promedio de coincidencia al número de competencias establecidas en la entrada de la consulta. Se observa que en el informe del perfil técnico la ocupación de mayor coincidencia a las ocupaciones es **Programadores de aplicaciones** con un 50 %, seguido de **Programadores de aplicaciones** con un 33.3 % y el resto son ocupaciones con un porcentaje de coincidencia del 16.7 %, todos cargos técnicos de TI. Por otro lado, se observa que, en el informe del perfil administrativo de TI, hay 6 ocupaciones que satisfacen al 100 % las necesidades de ocupaciones de la vacante y las demás dentro del informe un 66.7 %.

7.2. Caso de uso 2

Esta aplicación también consiste en hacer uso de la ontología aprendida para verificar sus aspectos técnicos a través de las respuestas que emite cuando se le asignan tareas por medio de requerimientos para una aplicación específica.

El escenario que se plantea es el de una empresa que recibe muchas hojas de vida y quiere filtrar toda la base de datos que tiene, dejando sólo las hojas de vida que pertenecen a un perfil de TI. Esto se considera como un primer filtro de clasificación que los ejecutivos de recursos humanos deben hacer para extraer los perfiles de ocupación deseados.

Lo primero, es realizar una extracción de hojas de vida de personas que pertenecen a perfiles relacionados con TI y personas con perfiles relacionados a cocina. La búsqueda de las personas se hizo de nuevo en la red social de LinkedIn, donde el parámetro de búsqueda fue tecnología de la información y Concina, respectivamente. Esta extracción de información se realiza a través de un *web scraping* bajo el método RPA (automatización robótica de procesos) mediante el software UiPath.

nombre	ocupación	ubicación	tipo
Adrián Esteban Giraldo Benítez	Ingeniero de tecnología de la información en la Universidad Distrital Francisco José de Caldas	Bogotá Bogotá DC Colombia	TI
Adrian gallego ospina	Analista tecnología de la información infraestructura en Línea Directa	Colombia	TI
Adriana Patricia Cortes	Consultora tecnología de la información en COLSOF SA	Colombia	TI
Ahoi Zeng	TI developer analyst at Caden SA	Antioquia Colombia	TI
Albein Naranjo Castaneda	auxiliar de cocina en Restaurante chichayo	Colombia	Cocina
Aldo Yate	Consultor tecnología de la información de SOAINT	Colombia	TI
Alejandra Forero Ramirez	Chef Pastelera en INALDE Business School	Colombia	Cocina
Alejandra García Nemocón	Chef at NUGA Cocina Local	Colombia	Cocina

Tabla 7-2.: Porción o extracción de hojas de vida de personas que pertenecen a perfiles relacionados con TI y personas con perfiles relacionados a Cocina

En la Tabla **7-2** se puede observar una parte de la extracción. En la primera columna está el nombre de la persona, en la siguiente columna está la ocupación que tiene actualmente, la ubicación de trabajo y si se obtuvo como resultado de búsqueda de TI o Cocina. Se puede ver como las personas con perfiles de TI tiene ocupaciones como Ingenieros de TI, analistas de TI, etc. Mientras que las personas con perfiles de cocina tienen ocupaciones como auxiliar de cocina, chef pastelero, etc.

La base de datos con la extracción de información de los 2 perfiles tiene en total 601 registros, entre los cuales 490 son del tipo TI y 111 son del tipo cocina.

Después de tener la base de datos con personas de perfiles de TI y cocina, se importan en un ambiente de desarrollo de Python, donde también se importa la Ontología OPTI. Para realizar la búsqueda del perfil en la ontología, se realiza una limpieza de la información quitando caracteres especiales, dejando todo el texto en minúscula y quitando palabras vacías (*stopwords*), todo esto para estandarizar el texto y aumentar la coincidencia en las búsquedas. Cabe resaltar que la búsqueda se hace por medio de una consulta SPARQL, donde OCUPACION_OBJETIVO es la ocupación que se desea encontrar (ver Lista 7.5).

```

/*****/
prefix opti: <http://OPTI.org/onto#>
  SELECT
    ?x
    ?y
  WHERE {
    ?x rdfs:label ?y .
    ?x opti:tiene_competencia ?z .
    FILTER regex( lcase(?y), lcase("(OCUPACION_OBJETIVO)") ) .
  }
/*****/

```

Listing 7.5: Consulta SPARQL para informe de clasificación perfil TI y cocina

Luego, se construye un código iterativo, que recorra cada una de las ocupaciones y busque su pertenencia o parecido con alguna de las instancias de la ontología. En la Figura **7-3** se puede ver una muestra del resultado de la búsqueda de la ontología. Los perfiles con los que se encuentra alguna coincidencia se marcan como *TI*, los demás se marcan como *False.c*

nombre	ocupación	ubicación	tipo	ocu_cleaned	Grupo
Adrián Esteban Giraldo Benítez	Ingeniero de tecnología de la información en la Universidad Distrital Francisco José de Caldas	Bogotá Bogotá DC Colombia	TI	ingeniero tecnologia informacion universidad distrital francisco jose caldas	TI
Adrian gallego ospina	Analista tecnología de la información infraestructura en Línea Directa	Colombia	TI	analista tecnologia informacion infraestructura en linea directa	TI
Adriana Patricia Cortes	Consultora tecnología de la información en COLSOF SA	Colombia	TI	consultora tecnologia informacion colsof sa	TI
Ahoi Zeng	TI developer analyst at Cadena SA	Antioquia Colombia	TI	developer analyst at cadena sa	FALSE
Albein Naranjo Castaneda	auxiliar de cocina en Restaurante chicalayo	Colombia	Cocina	auxiliar cocina restaurante chicalayo	FALSE
Aldo Yate	Consultor tecnología de la información de SOAINT	Colombia	TI	consultor tecnologia informacion soaint	TI
Alejandra Forero Ramirez	Chef Pastelera en INALDE Business School	Colombia	Cocina	chef pastelera in alde business school	FALSE
Alejandra García Nemocón	Chef at NUGA Cocina Local	Colombia	Cocina	chef at nuga cocina local	FALSE
Alejandra Lopez Guerrero	Aprendiz de auxiliar de cocina practicante en DLK SAS restaurante Gamberro	Bogotá Bogotá DC Colombia	Cocina	aprendiz auxiliar cocina practicante dlk sas restaurante gamberro	FALSE

Tabla 7-3.: Resultado de búsqueda de la ontología para personas que pertenecen a perfiles relacionados con TI

Para evaluar el resultado de la búsqueda de perfiles en la ontología, se comparan los valores conocidos, y los valores resultados de la búsqueda en la ontología. En Figura 7-1 se puede ver la matriz de confusión. Una matriz de confusión, también conocida como matriz de error, es una tabla resumida que se utiliza para evaluar el rendimiento de un modelo de clasificación. El número de predicciones correctas e incorrectas se resumen con los valores de conteo y se desglosan por cada clase. Esta gráfica permite analizar el comportamiento de la búsqueda. Se puede observar que de todos los perfiles de cocina (111), 107 no se encontraron en la ontología, mientras que 4 si se encontraron como perfiles de TI. Y de todos los perfiles de TI (490), 31 no se encontraron en la ontología, mientras que 459 si se encontraron como perfiles de TI.

La exactitud encontrada es del 94.2 %, el *recall* como [96.4 %, 93.7 %] y la precisión [77.5 %, 99.1 %].

La exactitud es simplemente igual a la proporción de predicciones que el modelo clasificó correctamente. La precisión también se conoce como valor predictivo positivo y es la propor-

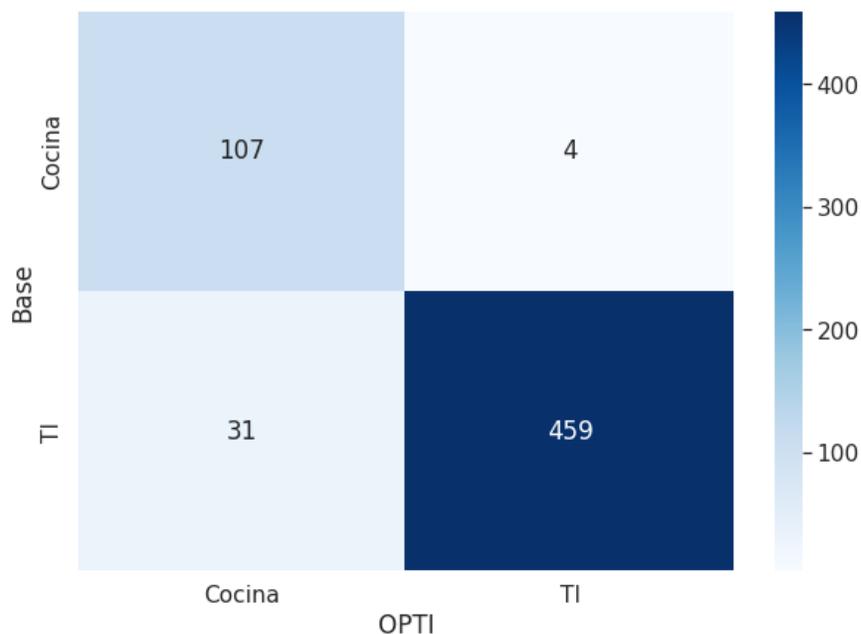


Figura 7-1.: Matriz de confusión de la búsqueda de perfiles TI en la ontología

ción de instancias relevantes entre las instancias recuperadas. En otras palabras, responde a la pregunta ¿Qué proporción de identificaciones positivas fue realmente correcta? y la sensibilidad, corresponde a la tasa de aciertos o tasa positiva real (TPR), siendo la proporción de la cantidad total de instancias pertinentes que se recuperaron realmente. Responde a la pregunta ¿Qué proporción de positivos reales se identificó correctamente?.

A continuación, se analizan algunos resultados de los perfiles que no fueron hallados correctamente. Los perfiles de cocina que sí se encontraron como perfiles de TI fueron los siguientes:

Gerente general en Rapsodia Cocina Fusión, Tecnico Profesional en Cocina en Servicio Nacional de Aprendizaje SENA, Gerente general en Chartoun Cocina Libanesa Utti Party Land, Gerente Propietario en CocinaVino Cocina interactiva Cursos para todo tipo de profesional.

Aunque los perfiles fueron encontrados en la búsqueda de cocina, sus ocupaciones sí corresponden a perfiles de TI, como Gerente general y técnico profesional.

Algunos de los perfiles que no fueron catalogados como TI pero si provienen de la búsqueda de TI son:

TI developer analyst at Cadena SA, Talent Acquisition en Stefanini LATAM, Coordinador de Infraestructuras de tecnología de la información en Aliatec SAS, Analista de Atracción del

Talento TI Sr en GRUPO CINTE, Scrum MasterTeamCoach en ConsultecTI, IT Specialist sénior Architect Ti Davivienda, TI Engineer y Software developer.

Muchos de estos perfiles están en inglés, por lo que no se pudieron encontrar en la ontología, los demás perfiles de TI sí pertenecen a TI pero todavía no están aprendidos en la ontología.

7.3. Caso de uso 3

Después de conocer que la persona postulante tiene un perfil de TI, se hace una búsqueda en la ontología, para conocer, cuáles son las competencias que debería tener esa persona.

Después de importar la base de datos de perfiles de TI y la ontología, se construye una consulta SPARQL para obtener las competencias que tiene la ocupación, que son las competencias que debería tener la persona postulante (Ver Lista 7.6). En el área de recursos humanos es también importante verificar que las competencias buscadas para un cargo correspondan a las del perfil encontrado.

```
/*  
prefix opti: <http://OPTI.org/onto#>  
SELECT  
  ?x  
  ?y  
WHERE {  
  ?x rdfs:label ?y .  
  ?x opti:tiene_competencia ?z .  
  FILTER regex(lower(?y), lower(" " + w + " ")) .  
}
```

Listing 7.6: Consulta SPARQL para obtener las competencias que tiene la ocupación

Así, el resultado muestra la ocupación que tiene la persona, la ocupación con la que hace MATHC y las competencias correspondientes. En la Figura se pueden ver algunos ejemplos de los perfiles y las ocupaciones que debe tener (ver Tabla 7-4).

nombre	ocupación	MATCH	Competencias
Eduard Salazar	Analista de infra-estructura tecnología de la información en Seguros SURA	soporte tecnico informatica	[‘Entre otros utilizando adecuadamente las herramientas informaticas excel powerpoint word realizar cambio de piezas o insumos de equipos cuando se requiera realizar soporte remoto o presencial realizar diagnostico y solucion a problemas de red operar’, ‘Windows y otros sistemas operativos atencion todo tipo de consultas tecnicas atencion call center activacion de servicios mantenimiento datacenter gestion proveedores relacionados directamente con el area de operaciones gestion bilingue español e ingles acceso total bajo vigilancia para servidor perfil requerido titulo tecnico en informatica ti manejo de plataformas y programas ti manejo de herramientas informaticas deseable conocimiento de erp deseable ingles tecnico nivel medio networking basico routers switchs firewalls a nivel conceptual servidores a nivel hardware conceptual marcas dell hp supermicro sistema operativos windows server’]

Sigue en la página siguiente.

nombre	ocupación	MATCH	Competencias
Leonardo Duque Tapias	Coordinador de infraes- tructuras de tecno- logía de la infor- mación en emtelco	profesional tecnología innovacion	['Responsabilidades brindar soporte para fomentar el uso de las herramientas de tecnología que tiene la empresa apoyar en campañas de asimilacion y adopción de nuevas tecnologías brindar soporte a la gestión integral de la gestión de la demanda', 'Apto para realizar prácticas profesionales por meses haber culminado estudios universitarios conocimientos se valorara la experiencia previa como practicante pre profesional manejando temas afines deseable manejo de herramientas office a nivel intermedio avanzado deseable manejo de temas de innovación y transformación tecnológica deseable conocimientos de marketing digital habilidad para realizar talleres para acercar la tecnología a la gente sede rep lima', 'Operación y mantenimiento de redes de transmisión de energía eléctrica en alta tensión rep gestiona a dos importantes empresas del sector consorcio transmantaro e isa peru las tres son empresas de interconexión eléctrica sa isa y cuentan con una red de subestaciones eléctricas y mas de ']

Sigue en la página siguiente.

nombre	ocupación	MATCH	Competencias
Numar Alexis Peña Quevedo	Director Área de proyectos tecnología de la infor- mación	jefe tecno- logia infor- macion	[‘Procurando el correcto funcionamien- to de los sistemas informaticos y la se- guridad de la informacion responsabi- lidades asegurar el exito e implementa- cion de los diferentes proyectos de desa- rrollo de software monitorear el cumpli- miento de los procesos de mantenimien- to a la infraestructura de ti supervisar la correcta administracion de las bases de datos’, ‘O afines conocimientos ges- tion de proyectos’, ‘Mantenimiento de infraestructura’]
Antonio Henao Mejía	Especialista Seguridad Informáti- ca Redes de datos Comuni- caciones Unificadas Dirección tecnología de la infor- mación	ingeniero datos tec- nologia informa- cion	[‘Ingenieria de la informacion o carre- ras afines certificacion en herramien- tas de inteligencia de negocios y admi- nistracion de bases de datos en micro- soft conocimiento y manejo de servicios cloud’, ‘Y virtualizacion de datos’, ‘In- tegracion de datos con tecnologias etl y frameworks de orquestacion’, ‘Herra- mientas de calidad de datos tales como master data servicesmds’, ‘Construc- cion de modelos fisicos tsq’, ‘Y logicos analysis services power bi otros conoci- mientos de controles centrados en datos usando iso y proteccion de datos per- sonales años de experiencia como desa- rrollador en proyectos de inteligencia de negocio o analisis de datos en el ambito empresarial’]

Sigue en la página siguiente.

nombre	ocupación	MATCH	Competencias
Jorge Ivan Escobar	Gerente Transformación tecnología de la información Bancolombia	gerente servicios tecnología informacion	[’Por medio de la implementacion de herramientas tecnologias para optimizar el proceso y funcionamiento del negocio tus responsabilidades clave identificar requerimientos y necesidades de clientes promoviendo la venta de servicios al mercado’, ’Políticas y requerimientos de ey aplicar estilos y estandares de calidad de ey en todas las comunicaciones internas y externas comprender y aplicar la metodología’, ’Herramientas y tecnologia correctas desarrollar un entendimiento de nuestras prioridades estrategicas’, ’Estructura comercial y objetivos de equipo y rol personal contribuir a la cultura de intercambio de conocimiento de ey participando de foros’, ’Diploma o post titulo experiencia minima de años en cargos similares experiencia en labores gerenciales y manejo de equipos ingles avanzado oral y escrito flexibilidad y disponibilidad para viajar dominio y conocimiento de plataforma sap conocimiento de python’, ’Rpa y manejo de bases de datos que estamos buscando personas que se identifiquen con los valores de ey’, ’El aprendizaje y el reconocimiento’]

Tabla 7-4.: Porción o extracción de resultado de búsqueda de competencias de las personas que pertenecen a perfiles relacionados con TI

7.4. Análisis y conclusiones del capítulo

La identificación y selección de candidatos son procesos que requieren rapidez de respuesta dentro de una compañía y en el cual, procesos como el *e-recruitment* pueden aportar de forma sustancial a economizar tiempo y dinero. La ontología aprendida OPTI responde adecuadamente a las preguntas de competencia con la cual fue diseñada, sirve como base de

organización del conocimiento para una ontología de perfiles de TI que pueda ser utilizada en procesos de *e-recruitment* y responde satisfactoriamente a las consultas SPARQL que se realiza en cada caso de uso. Adicionalmente, se puede decir que los aspectos técnicos de la ontología OPTI fueron puestos a prueba con la aplicación de tareas en cada caso de uso y la respuesta de esta fue satisfactoria para cada uno de los requerimientos.

El desarrollo de este capítulo completa el proceso de evaluación de la ontología aprendida, validando y verificando una vez más a partir del criterio de aplicación, que la ontología aprendida es adecuada, está bien construida y responde bien a las necesidades del dominio. Además, da cumplimiento al objetivo 4 de la tesis: Verificar la utilidad del modelo de aprendizaje ontológico implementado por medio de la construcción de un prototipo para el dominio de *e-recruitment*.

8. Capítulo 8: Conclusiones y recomendaciones

8.1. Conclusiones

La propuesta de esta tesis de maestría destaca cómo la combinación de técnicas de aprendizaje de máquinas, aprendizaje profundo y procesamiento de lenguaje natural puede potenciar el desarrollo de sistemas de aprendizaje ontológico en el dominio de *e-recruitment*. Este tipo de enfoques contribuirá a la rapidez de respuesta, automatización de procesos y economicidad en el área de recursos humanos de las compañías.

El *e-recruitment* requiere el uso de nuevas tecnologías para ir remodelando y afinando los procesos para seleccionar y retener el personal talentoso dentro de una organización. Así mismo, cuando de profesionales asociados a las tecnologías de la información se trata, las empresas deben realizar un doble esfuerzo en buscar la ocupación correcta para cumplir ciertas funciones dentro de la empresa, debido a que las habilidades y competencias de estas ocupaciones se están modificando con gran velocidad debido a la incorporación de tecnologías desafiantes en el mercado. Tener una ontología de calidad que modele adecuadamente el dominio y pueda ir modificando las ocupaciones y las competencias de estos puestos de trabajo de forma semiautomática y al ritmo de su evolución, permitirá que las empresas tomen decisiones rápidas, acertadas, que reduzcan la rotación de personal y que disminuya los gastos en los procesos de contratación de personal idóneo.

Trabajar bajo estándares o normas establecidas, local o internacionalmente, garantizan la que el desarrollo de proyectos de ingeniería sea reproducibles y escalables. Los estándares y normas unifican enfoques de un dominio en particular, por ello, se hace necesario que a la hora de gestionar el conocimiento sobre perfiles asociados a las tecnologías de la información y de la comunicación o cualquier otro perfil, se tenga en cuenta la clasificación internacional de ocupaciones (ISCO) u otras clasificaciones, normas o estándares actualizadas y mejoradas que existan acerca de ello.

Se conoce que crear una ontología desde cero es un proceso laborioso y complejo. Es por ello que las cuestiones metodológicas constituyen un aspecto crítico en el proceso de construcción de las ontologías y proporcionan múltiples recursos para la formalización del conocimiento

en un determinado dominio. Aunque no existe una sola forma establecida para el desarrollo de ontologías, las metodologías existentes tienen en algunos casos criterios en común, pero difieren en su objetivo de aplicación. Se hace necesario entonces determinar la conveniencia de la definición de una metodología a la hora de construir una ontología base a la que luego se le aplicarán técnicas de aprendizaje ontológico, ya que esto enriquecerá de forma adecuada el conocimiento, la correcta construcción de la ontología y el buen rendimiento de esta en la aplicación para la cual fue diseñada.

El aprendizaje de ontologías (AO) está orientado en la construcción y el mantenimiento de ontologías. La gran necesidad de una ontología es mantenerse actualizada para cumplir con los criterios de calidad necesarios para el buen desempeño en el uso para la cual se construyó. La recuperación de información, el aprendizaje automático, el *data mining*, el *deep learning*, el procesamiento de lenguaje natural y la representación y razonamiento del conocimiento son técnicas usadas para que el AO pueda generarse en diferentes elementos de las ontologías. La combinación de estas técnicas permite buenos resultados de aprendizaje y la selección adecuada de la técnica específica a usar para cada salida de forma ordenada (clases, jerarquía de clases y restricciones) tendrá como resultado un buen desempeño de la ontología aprendida en aplicaciones de su dominio.

Las redes profesionales contienen actualmente cantidades de información relevante sobre profesionales y vacantes ofrecidas por el mercado laboral. La disponibilidad de todo este conocimiento es de gran importancia para abastecer tecnologías como las ontologías y ser usadas por personal de recursos humanos para optimizar, automatizar, economizar y agilizar los diferentes procesos de la contratación de personal. Redes profesionales como LinkedIn, brindan gran cantidad de conocimiento tanto del mercado laboral como de los diferentes perfiles profesionales que se están gestando en cualquier área de conocimiento. Es importante que la disposición de este conocimiento en la web a veces está etiquetado, otras veces se encuentra medianamente etiquetada como es el caso de LinkedIn y muchas otras veces el texto es completamente no estructurado. De acuerdo con cómo está dispuesto el conocimiento las investigaciones deberán adaptar sus técnicas de *web scraping* para generar información de calidad.

Los *Transformers* han demostrado ser altamente eficientes en la extracción de similitud semántica gracias a su arquitectura de atención, capacidad de contexto, adaptabilidad a diferentes tareas y capacidad para aprender representaciones profundas y ricas del lenguaje. Estas ventajas los han convertido en uno de los enfoques más exitosos y ampliamente utilizados en el campo del procesamiento del lenguaje natural. Los *Transformers* tienen la capacidad de considerar el contexto completo de una oración o un fragmento de texto. A diferencia de modelos como las redes neuronales recurrentes (RNN), que procesan la información secuencialmente, los *Transformers* pueden considerar todas las palabras en una secuencia al

mismo tiempo, lo que les permite comprender mejor el contexto y las relaciones entre las palabras. Por eso es posible ver como en esta tesis, en diferentes partes se hacen análisis de similitud semántica a través de *Sentence-BERT* que es una técnica que utiliza modelos preentrenados de *Transformers*, como BERT (*Bidirectional Encoder Representations from Transformers*), para generar representaciones numéricas (*embeddings*) de oraciones enteras en un espacio semántico de alta dimensión. Esta técnica a través de su codificador siamés permite comparar oraciones con diferentes dimensiones a través de un *padding* que agrega tokens de relleno para igualar la longitud de las dos oraciones al máximo de tokens entre ellas.

En el contexto de aprendizaje de clases y jerarquías de clases en una ontología, se evidencia que los modelos KGE pueden ser usados para capturar la semántica de las relaciones entre clases y aprender representaciones vectoriales de las clases y subclases. Entonces, en esta tesis se usaron los modelos KGE para predecir relaciones jerárquicas entre clases en una ontología, es decir, dado un grafo de conocimiento que representa la jerarquía de clases, un modelo KGE puede aprendió las representaciones vectoriales de las clases y predijo la relación *es_subclase_de* para determinar si una clase es una subclase de otra. Además, permitió aprender representaciones vectoriales de las clases de la ontología y utilizó estas representaciones para clasificar nuevas entidades en las clases adecuadas.

El enriquecimiento de grafos a través de extracción de nuevas tripletas con información de similitud semántica de elementos de la ontología base, permite mejorar las predicciones de un modelo KGE aplicado al aprendizaje de clases y jerarquía de clases de una ontología. Este concepto puede ser ampliado con la aplicación de otras técnicas analíticas que agreguen más información al grafo.

La evaluación de una ontología todavía es un tema de investigación que debe ser estudiado profundamente, debido a que es difícil determinar con un solo criterio si la ontología cumple o no con ciertos parámetros específicos para ser declarada de buena calidad o no. Lo anterior es porque las ontologías tienen diferentes fines y diferentes formas de construirse. Ahora bien, para cumplir con los requisitos mínimos de evaluación de una ontología es necesario que se evalúen 4 criterios importantes, la rigurosidad taxonómica, el lenguaje utilizado para la codificación, el vocabulario utilizado para representar los conceptos y relaciones del dominio modelado y el rendimiento de las aplicaciones o tareas que utilizan las ontologías. A partir de la evaluación de estos criterios, se pueden tomar decisiones de refinamiento de la ontología construida o aprendida.

La aplicación de una ontología en un caso de uso para el dominio hace parte del proceso de evaluación de la ontología. El rendimiento de la aplicación y las repuestas a las consultas de las ontologías corresponden a respuestas lógicas y acertadas en el dominio, se considera que los aspectos técnicos y lógicos de la ontología corresponden a las preguntas de compe-

tencia. En este caso, de este desarrollo de investigación se pudo observar que el diseño de la ontología base y el modelo de aprendizaje ontológico aplicado satisfacen al dominio de *e-recruitment*. Basados en necesidades actuales de la contratación web, se puso a prueba la ontología construida OPTI. Los informes de los diferentes casos de uso corresponden a lo esperado y puede ser implementado en una herramienta de software real para contratación de perfiles profesionales de TI.

8.2. Recomendaciones o trabajo futuro

El aprendizaje ontológico constituye un campo de investigación amplio que aún tiene brechas de conocimiento que abren preguntas de investigación interesantes para trabajos futuros.

En particular, se pueden señalar diferentes aspectos del modelo propuesto que se pueden mejorar y profundizar. A continuación, se enumeran algunas recomendaciones o trabajos futuros que surgen a partir del desarrollo de esta tesis:

- La ontología base puede ampliarse a otros aspectos que no fueron considerados en este trabajo, como lo son las cualificaciones. Así mismo, a esta ontología base pueden asociarse un mayor número de relaciones que permita en el módulo de AO aprender más restricciones por análisis de relaciones u otras técnicas.
- En eslabón de AO puede realizarse una comparación de la respuesta de aplicación de diferentes técnicas o algoritmos para cada una de las capas de aprendizaje ontológico, con el fin de tener una perspectiva más amplia del rendimiento de diferentes algoritmos y seleccionar los mejores para el dominio de la ontología.
- Se pueden explorar técnicas de clusterización u otras técnicas de aprendizaje para extraer nueva información de la ontología y generar nuevas tripletas que puedan enriquecer el grafo y lograr así mejores resultados en las predicciones de los modelos KGE para aprendizaje en las diferentes capas de una ontología.
- Implementar más técnicas de procesamiento de lenguaje natural, con el fin de refinar la información obtenida de la red profesional LinkedIn a través del *web scraping*, específicamente en el trabajo de la superclase competencias que, aunque es una etiqueta de la red social, el empleador puede describir las tareas específicas y competencias en texto libre.
- Se pueden estudiar otros análisis posibles que se harían sobre la ontología base para enriquecer el grafo de conocimiento aún más y verificar que el rendimiento del modelo para aprendizaje de clases y jerarquía de clases mejore.

- Para robustecer el razonamiento lógico de la ontología OPTI, se pueden realizar más análisis de relaciones para extraer más restricciones. Así mismo, estudiar la forma de aplicar otras técnicas para aprender nuevas restricciones e incluso axiomas. El aprendizaje de axiomas es un área de conocimiento que aún tiene mucho por recorrer y podría llevarse a una investigación de doctorado.
- Establecer un proceso de refinamiento de la ontología después de la evaluación y los resultados de la aplicación de caso de uso, puede favorecer la creación de una ontología de más calidad, capaz de modelar y desplegar sistemas de automatización reales e inteligentes en el dominio de *e-recruitment* para profesionales asociados al área de TI.
- La ontología OPTI podría ponerse a punto y usarse para llevarla a aplicaciones académicas e identificar las nuevas carreras esperadas en el área de TI. A través de la creación de algún portal de consultas, los proveedores de educación y capacitación podrían comprender con la ontología OPTI qué habilidades necesitan los mercados laborales y así, adaptar sus planes de estudios para preparar mejor a sus estudiantes.

A. Anexo: Definición de terminología de la ontología base

Ocupación: se refiere al tipo de trabajo realizado en un trabajo. El concepto de ocupación se define como un conjunto de puestos de trabajo cuyas tareas y funciones principales se caracterizan por un alto grado de similitud. Una persona puede estar asociada a una ocupación a través del trabajo principal que tiene actualmente, un segundo trabajo, un trabajo futuro o un trabajo anterior.

Competencia: es la capacidad de llevar a cabo las tareas y deberes de un trabajo.

Directores y gerentes: Los gerentes planifican, dirigen, coordinan y evalúan las actividades generales de empresas, gobiernos y otras organizaciones, o de unidades organizacionales dentro de ellos, y formulan y revisan sus políticas, leyes, normas y reglamentos.

Profesionales científicos e intelectuales: Los profesionales aumentan el stock de conocimiento existente; aplican conceptos y teorías científicas o artísticas; enseñan sobre lo anterior de manera sistemática; o participan en cualquier combinación de estas actividades.

Técnicos y profesionales de nivel medio: Los técnicos y profesionales asociados realizan tareas técnicas y afines relacionadas con la investigación y la aplicación de conceptos científicos o artísticos y métodos operativos, y regulaciones gubernamentales o comerciales.

Personal de apoyo administrativo: Los trabajadores de apoyo administrativo registran, organizan, almacenan, calculan y recuperan información, y realizan una serie de tareas administrativas relacionadas con operaciones de manejo de dinero, arreglos de viaje, solicitudes de información y citas.

Oficiales operarios y artesanos de artes mecánicas y de otros oficios: Los trabajadores artesanales y afines aplican conocimientos y habilidades técnicas y prácticas específicas en los campos para construir y mantener edificios; formar metal; levantar estructuras metálicas; Configurar máquinas herramienta o fabricar, ajustar, mantener y reparar maquinaria, equipo o herramientas; realizar trabajos de impresión; y producir o procesar alimentos, textiles y madera, metal y otros artículos, incluidos artículos artesanales.

Directores ejecutivos personal directivo de la administración pública y miembros del poder ejecutivo y de los cuerpos legislativos: Los directores ejecutivos, altos funcionarios y legisladores formulan y revisan las políticas y planifican, dirigen, coordinan y evalúan las actividades generales de las empresas, los gobiernos y otras organizaciones con el apoyo de otros gerentes.

Directores administradores y comerciales: Los gerentes administrativos y comerciales planifican, organizan, dirigen, controlan y coordinan las actividades financieras, administrativas, de recursos humanos, políticas, planificación, investigación y desarrollo, publicidad, relaciones públicas y ventas y marketing de empresas y organizaciones, o de empresas que proporcionan tales servicios a otras empresas y organizaciones.

Directores y gerentes de producción y operaciones: Los gerentes de producción y servicios especializados planifican, dirigen y coordinan la producción de los bienes y la prestación de los servicios profesionales y técnicos especializados que brinda una empresa u organización, ya sea como gerente de un departamento o como gerente general de una empresa u organización que no tener una jerarquía de gerentes. Son responsables de las operaciones de fabricación, minería, construcción, logística, tecnología de la información y las comunicaciones, de las operaciones agrícolas, forestales y pesqueras a gran escala, y de la prestación de servicios de salud, educación, bienestar social, banca, seguros y otros servicios profesionales y técnicos.

Gerentes de hoteles restaurantes comercios y otros servicios: Los gerentes de hotelería, comercio y servicios relacionados planifican, organizan y dirigen las operaciones de los establecimientos que ofrecen alojamiento, hostelería, venta al por menor y otros servicios.

Profesionales de las ciencias y de la ingeniería: Los profesionales de la ciencia y la ingeniería realizan investigaciones, mejoran o desarrollan conceptos, teorías y métodos operativos, o aplican conocimientos científicos relacionados con campos como la física, la astronomía, la meteorología, la química, la geofísica, la geología, la biología, la ecología, la farmacología, la medicina, las matemáticas, la estadística, la arquitectura, ingeniería, diseño y tecnología.

Profesionales de la salud: Los profesionales de la salud realizan investigaciones; mejorar o desarrollar conceptos, teorías y métodos operativos; y aplicar los conocimientos científicos relacionados con la medicina, enfermería, odontología, medicina veterinaria, farmacia y promoción de la salud.

Especialistas en organización de la administración pública y de empresas: Los

profesionales de administración y negocios realizan labores analíticas, conceptuales y prácticas para brindar servicios en materia financiera, desarrollo de recursos humanos, relaciones públicas, mercadeo y ventas en las áreas técnica, médica, tecnologías de la información y la comunicación; y conducta, revisiones de estructuras, métodos y sistemas organizacionales, así como análisis cuantitativos de la información que afecta los programas de inversión.

Profesionales de tecnología de la información y las comunicaciones: Los profesionales de la tecnología de la información y las comunicaciones realizan investigaciones, planifican, diseñan, escriben, prueban, brindan asesoramiento y mejoran los sistemas de tecnología de la información, el hardware, el software y los conceptos relacionados para aplicaciones específicas; desarrollar documentación asociada que incluya principios, políticas y procedimientos; y diseñar, desarrollar, controlar, mantener y respaldar bases de datos y otros sistemas de información para garantizar un rendimiento óptimo y la integridad y seguridad de los datos.

Profesionales en derecho en ciencias sociales y culturales: Los profesionales del derecho, la sociedad y la cultura realizan investigaciones, mejoran o desarrollan conceptos, teorías y métodos operativos; o aplicar conocimientos relacionados con la ley, almacenamiento y recuperación de información y artefactos, psicología, bienestar social, política, economía, historia, religión, idiomas, sociología, otras ciencias sociales y artes y entretenimiento.

Profesionales de las ciencias y la ingeniería de nivel medio: Los profesionales asociados de ciencia e ingeniería realizan tareas técnicas relacionadas con la investigación y los métodos operativos en ciencia e ingeniería. Supervisan y controlan los aspectos técnicos y operativos de la minería, la fabricación, la construcción y otras operaciones de ingeniería, y operan equipos técnicos, incluidos aviones y barcos.

Profesionales de nivel medio de la salud: Los profesionales asociados de la salud realizan tareas técnicas y prácticas para respaldar el diagnóstico y el tratamiento de enfermedades, lesiones y deficiencias en humanos y animales, y para respaldar la implementación de planes de atención médica, tratamiento y derivación generalmente establecidos por médicos, veterinarios, enfermeras y otros profesionales de la salud.

Profesionales de nivel medio en operaciones financieras y administrativas: Los profesionales asociados de administración y negocios realizan principalmente tareas técnicas relacionadas con la aplicación práctica de conocimientos relacionados con asuntos de transacciones y contabilidad financiera, cálculos matemáticos, desarrollo de recursos humanos, compra y venta de instrumentos financieros, tareas de secretaría especializadas y cumplimiento o aplicación de normas gubernamentales. También se incluyen los trabajadores que brindan servicios comerciales como despacho de aduanas, planificación de conferencias, co-

locación laboral, compra y venta de bienes inmuebles o productos básicos a granel, y que actúan como agentes para artistas, como atletas y artistas.

Profesionales de nivel medio de servicios jurídicos sociales culturales y afines:

Los profesionales adjuntos del ámbito jurídico, social, cultural y afines realizan tareas técnicas relacionadas con la aplicación práctica de los conocimientos relacionados con los servicios jurídicos, el trabajo social, la cultura, la preparación de alimentos, el deporte y la religión.

Técnicos de la tecnología de la información y las comunicaciones: Los técnicos de la información y las comunicaciones prestan apoyo en el día a día de los sistemas informáticos, sistemas y redes de comunicaciones y realizan tareas técnicas relacionadas con las telecomunicaciones, la emisión de imagen y sonido, así como otro tipo de señales de telecomunicaciones en tierra, mar o aeronaves.

Oficinistas: Los empleados generales y de teclado registran, organizan, almacenan y recuperan información y realizan una amplia gama de tareas administrativas y administrativas de acuerdo con los procedimientos establecidos.

Empleados en trato directo con el público: Los empleados de servicios al cliente se ocupan de los clientes en relación con las operaciones de manejo de dinero, arreglos de viaje, solicitudes de información, concertación de citas, operación de centralitas telefónicas y entrevistas para encuestas o para completar solicitudes de elegibilidad para servicios.

Empleados contables y encargados del registro de materiales: Los empleados de registro numérico y material obtienen, compilan y computan datos contables, contables, estadísticos, financieros y otros datos numéricos, y se hacen cargo de las transacciones en efectivo relacionadas con asuntos comerciales. Algunas ocupaciones clasificadas aquí mantienen registros de bienes producidos, comprados, almacenados, despachados y de materiales necesarios en fechas de producción específicas, o mantienen registros de aspectos operativos y coordinan el tiempo de transporte de pasajeros y carga.

Trabajadores especializados en electricidad y la electrotecnología: Los trabajadores de oficios eléctricos y electrónicos instalan, ajustan y mantienen sistemas de cableado eléctrico y maquinaria y otros aparatos eléctricos, líneas y cables de transmisión y suministro eléctrico, y equipos y sistemas electrónicos y de telecomunicaciones. El trabajo se realiza a mano y con herramientas manuales y otras que se utilizan para reducir la cantidad de esfuerzo físico y el tiempo requerido para tareas específicas, así como para mejorar la calidad de los productos. Las tareas exigen una comprensión de la organización del trabajo, los materiales y herramientas utilizados, y la naturaleza y propósito del producto final.

Directores generales y gerentes generales: Los directores gerentes y los ejecutivos en jefe formulan y revisan las políticas y planifican, dirigen, coordinan y evalúan las actividades generales de las empresas u organizaciones (excepto las organizaciones de intereses especiales y los departamentos gubernamentales) con el apoyo de otros gerentes, generalmente dentro de las pautas establecidas por una junta directiva, administradores o un órgano de gobierno ante quien respondan de las operaciones realizadas y de sus resultados.

Directores de administración y servicios: Los gerentes de administración y servicios comerciales planifican, organizan, dirigen, controlan y coordinan las actividades financieras, administrativas, de recursos humanos, de políticas y de planificación de las organizaciones o de las empresas que brindan tales servicios a otras empresas y organizaciones.

Directores de ventas comercialización y desarrollo: Los gerentes de ventas, marketing y desarrollo planifican, organizan, dirigen, controlan y coordinan la publicidad, las relaciones públicas, la investigación y el desarrollo y las actividades de ventas y marketing de empresas y organizaciones, o de empresas que brindan dichos servicios a otras empresas y organizaciones.

Directores de producción agropecuaria silvicultura y pesca: Los gerentes de producción en agricultura, silvicultura y pesca planifican, dirigen y coordinan la producción en operaciones agrícolas, hortícolas, forestales, acuícolas y pesqueras a gran escala, como plantaciones, grandes ranchos, granjas colectivas y cooperativas para cultivar y cosechar cultivos, criar y criar ganado, pescados y mariscos y para capturar y recolectar peces y otras formas de vida acuática.

Directores de industrias manufactureras de minería construcción y distribución: Los gerentes de manufactura, minería, construcción y distribución planifican, organizan y coordinan las operaciones de manufactura, extracción de minerales, construcción, suministro, almacenamiento y transporte, ya sea como gerente de un departamento o como gerente general de una empresa u organización que no tener una jerarquía de gerentes.

Directores de servicios de tecnología de la información y las comunicaciones: Los gerentes de servicios de tecnología de la información y las comunicaciones planifican, dirigen y coordinan la adquisición, el desarrollo, el mantenimiento y el uso de sistemas informáticos y de telecomunicaciones, ya sea como gerente de un departamento o como gerente general de una empresa u organización que no tiene una jerarquía de gerentes.

Directores y gerentes de servicios profesionales: Los gerentes de servicios profesionales planifican, dirigen y coordinan la prestación de servicios de cuidado de niños, salud, bienestar, educación y otros servicios profesionales, y administran las sucursales de instituciones que brindan servicios financieros y de seguros.

Gerentes de hoteles y restaurantes: Los gerentes de hoteles y restaurantes planifican, organizan y dirigen las operaciones de los establecimientos que brindan alojamiento, comidas, bebidas y otros servicios de hospitalidad.

Gerentes de comercios al por mayor y al por menor: Los gerentes de comercio al por menor y al por mayor planifican, organizan, coordinan y controlan las operaciones de los establecimientos que venden mercancías al por menor o al por mayor. Son responsables de los presupuestos, la dotación de personal y la dirección estratégica y operativa de las tiendas o de las unidades organizativas dentro de las tiendas que venden determinados tipos de productos.

Otros gerentes de servicios: Los administradores de otros servicios planifican, organizan y controlan las operaciones de los establecimientos que brindan servicios deportivos, culturales, recreativos, de viajes, contacto con el cliente y otros servicios de esparcimiento.

Físicos químicos y afines: Los profesionales de las ciencias físicas y de la tierra realizan investigaciones; mejorar o desarrollar conceptos, teorías y métodos operativos, o aplicar conocimientos científicos relacionados con la física, la astronomía, la meteorología, la química, la geología y la geofísica.

Matemáticos actuarios y estadísticos: Los matemáticos, actuarios y estadísticos realizan investigaciones, mejoran o desarrollan conceptos, teorías y técnicas y modelos operativos matemáticos, actuariales y estadísticos; y aplicar este conocimiento a una amplia gama de tareas en campos tales como ingeniería, negocios y ciencias sociales y otras ciencias.

Ingenieros excluyendo electrotecnólogos: Los profesionales de la ingeniería (excluida la electrotecnia) diseñan, planifican y organizan las pruebas, la construcción, la instalación y el mantenimiento de estructuras, máquinas y sus componentes, y sistemas y plantas de producción; y planificar programas de producción y procedimientos de trabajo para garantizar que los proyectos de ingeniería se lleven a cabo de manera segura, eficiente y rentable.

Ingenieros en electrotecnología: Los ingenieros en electrotecnología investigan y diseñan, asesoran, planifican y dirigen la construcción y operación de sistemas, componentes, motores y equipos electrónicos, eléctricos y de telecomunicaciones. Organizan y establecen sistemas de control para monitorear el desempeño y la seguridad de los conjuntos y sistemas eléctricos y electrónicos.

Arquitectos urbanistas agrimensores y diseñadores: Los arquitectos, planificadores, topógrafos y diseñadores planifican y diseñan paisajes, exteriores e interiores de edificios, productos para la fabricación y contenido visual y audiovisual para la comunicación de in-

formación. Realizan trabajos topográficos para posicionar con precisión las características geográficas; diseñar, preparar y revisar mapas; y desarrollar e implementar planes y políticas para controlar el uso de la tierra.

Otros profesionales de la salud: Otros profesionales de la salud brindan servicios de salud relacionados con odontología, farmacia, salud e higiene ambiental, salud y seguridad ocupacional, fisioterapia, nutrición, terapias de audición, habla, visión y rehabilitación. Este grupo menor incluye a todos los profesionales de la salud humana, excepto médicos, practicantes de medicina tradicional y complementaria, enfermeras, parteras y profesionales paramédicos.

Especialistas en finanzas: Los profesionales de las finanzas planifican, desarrollan, organizan, administran, invierten, gestionan y realizan análisis cuantitativos de sistemas de contabilidad financiera o de fondos para personas, establecimientos e instituciones públicas o privadas.

Especialistas en organización de administración: Los profesionales de la administración aplican varios conceptos y teorías relacionados con la mejora de la eficacia de las organizaciones y de las personas que la integran.

Profesionales de las ventas la comercialización y las relaciones públicas: Los profesionales de ventas, marketing y relaciones públicas planifican, desarrollan, coordinan e implementan programas de difusión de información para promover organizaciones, bienes y servicios; y representar a empresas en la venta de una gama de bienes y servicios técnicos, industriales, médicos, farmacéuticos y de TIC.

Desarrolladores y analistas de software y multimedia: Los desarrolladores y analistas de software y aplicaciones realizan investigaciones, planifican, diseñan, escriben, prueban, brindan asesoramiento y mejoran los sistemas de tecnología de la información, como hardware, software y otras aplicaciones, para cumplir con requisitos específicos.

Especialistas en bases de datos y en redes de computadores: Los profesionales de bases de datos y redes diseñan, desarrollan, controlan, mantienen y respaldan el rendimiento y la seguridad óptimos de los sistemas e infraestructuras de tecnología de la información, incluidas las bases de datos, el hardware y el software, las redes y los sistemas operativos.

Profesionales en derecho: Los profesionales del derecho realizan investigaciones sobre problemas jurídicos, asesoran a clientes sobre aspectos jurídicos de los problemas, defienden casos o llevan a cabo acciones judiciales en los tribunales de justicia, presiden procedimientos judiciales en los tribunales de justicia y redactan leyes y reglamentos.

Archivistas bibliotecarios curadores y afines: Los bibliotecarios, archiveros y curadores desarrollan y mantienen las colecciones de archivos, bibliotecas, museos, galerías de arte y establecimientos similares.

Artistas creativos e interpretativos: Los artistas creativos e intérpretes comunican ideas, impresiones y hechos en una amplia gama de medios para lograr efectos particulares; interpretar una composición como una partitura musical o un guión para realizar o dirigir la actuación; y organizar la presentación de dicha actuación y otros eventos mediáticos.

Supervisores en ingeniería de minas de industrias manufactureras y de la construcción: Los supervisores de minería, manufactura y construcción coordinan, supervisan, controlan y programan las actividades de los trabajadores en operaciones de manufactura, minería y construcción.

Técnicos en control de procesos: Los técnicos de control de procesos operan y monitorean tableros de distribución, sistemas de control computarizados, maquinaria de control de procesos multifunción y mantienen unidades de procesamiento en generación y distribución de energía eléctrica, aguas residuales, tratamiento de aguas residuales y plantas de eliminación de desechos, refinerías químicas, de petróleo y gas natural, en procesamiento de metales y otras operaciones de procesos múltiples.

Técnicos médicos y farmacéuticos: Los técnicos médicos y farmacéuticos realizan tareas técnicas para ayudar en el diagnóstico y tratamiento de enfermedades, lesiones y deficiencias.

Profesionales de nivel medio en finanzas y matemáticas: Los profesionales financieros y matemáticos asociados asignan un valor a varios artículos y propiedades, mantienen registros de transacciones financieras, analizan la información de la solicitud de préstamo y toman una decisión, compran y venden instrumentos financieros y realizan cálculos matemáticos y relacionados.

Agentes comerciales y corredores: Los agentes y corredores de compras y ventas representan a empresas, gobiernos y otras organizaciones para comprar y vender productos básicos, bienes, seguros, envíos y otros servicios a establecimientos industriales, profesionales, comerciales o de otro tipo, o actúan como agentes independientes para reunir a compradores y vendedores de productos básicos. y servicios.

Agentes de servicios comerciales: Los agentes de servicios comerciales establecen contactos para vender diversos servicios comerciales, como espacios publicitarios en los medios; llevar a cabo actividades de despacho de aduanas asegurándose de que los documentos necesarios estén en orden; relacionar a los solicitantes de empleo con las vacantes; encontrar

trabajadores para los empleadores; concertar contratos para la actuación de atletas, animadores y artistas, así como para la publicación de libros, la producción de obras de teatro o la grabación, interpretación y venta de música; y planificar y organizar conferencias y eventos similares.

Secretarios administrativos y especializados: Los secretarios legales aplican conocimientos especializados de terminología y procedimientos legales para brindar apoyo a los profesionales del derecho con actividades de comunicación, documentación y coordinación gerencial interna, en bufetes de abogados, departamentos legales de grandes firmas y gobiernos.

Profesionales de nivel medio de servicios jurídicos sociales y religiosos: Los profesionales adjuntos jurídicos, sociales y religiosos brindan servicios técnicos y prácticos y funciones de apoyo en procesos e investigaciones legales, programas de asistencia social y comunitaria y actividades religiosas.

Profesionales de nivel medio en actividades culturales artísticas y culinarias: Los profesionales asociados artísticos, culturales y culinarios combinan habilidades creativas y conocimientos técnicos y culturales para tomar y procesar fotografías fijas; diseño y decoración de decorados de teatro, escaparates de comercios e interiores de viviendas; preparación de objetos para exhibición; mantenimiento de colecciones, registros y sistemas de catalogación de bibliotecas y galerías; creación de menús y preparación y presentación de alimentos; apoyo a la producción teatral, cinematográfica y televisiva; y en otras áreas del esfuerzo artístico y cultural.

Técnicos en operaciones de tecnología de la información y las comunicaciones y asistencia al usuario: Los técnicos de operaciones de tecnología de la información y las comunicaciones y de soporte al usuario brindan apoyo para el funcionamiento diario de los sistemas de comunicaciones, los sistemas informáticos y las redes, y brindan asistencia técnica a los usuarios.

Técnicos en telecomunicaciones y radiodifusión: Los técnicos de telecomunicaciones y radiodifusión controlan el funcionamiento técnico de equipos para grabar y editar imágenes y sonido y para transmitir emisiones de radio y televisión de imágenes y sonidos, así como otros tipos de señales de telecomunicación en tierra, mar o aeronaves; realizan tareas técnicas relacionadas con la investigación en ingeniería de telecomunicaciones y con el diseño, la fabricación, el montaje, la construcción, el funcionamiento, el mantenimiento y la reparación de sistemas de telecomunicaciones.

Secretarios general: Las secretarias (generales) usan máquinas de escribir, computadoras

personales u otros equipos de procesamiento de textos para transcribir la correspondencia y otros documentos, verificar y formatear documentos preparados por otro personal, manejar el correo entrante y saliente, filtrar solicitudes de reuniones o citas y realizar una variedad de tareas de apoyo administrativo.

Operadores de máquinas de oficina: Los operadores de teclado ingresan y procesan texto y datos, y preparan, editan y generan documentos para almacenamiento, procesamiento, publicación y transmisión.

Empleados de servicios de información al cliente: Los empleados de información del cliente proporcionan u obtienen información en persona, por teléfono o medios electrónicos, como un correo electrónico en relación con la realización de arreglos de viaje, describiendo los productos o servicios de una organización, registrando y saludando a los invitados y visitantes, haciendo citas, conectando llamadas telefónicas y recopilando información de encuestados o solicitantes de servicios.

Auxiliares contables y financieros: Los empleados numéricos obtienen, compilan y computan datos contables, contables, estadísticos, financieros y otros datos numéricos, y se hacen cargo de las transacciones en efectivo relacionadas con asuntos comerciales.

Instaladores y reparadores de equipos eléctricos: Los instaladores y reparadores de equipos eléctricos instalan, ajustan y mantienen sistemas de cableado eléctrico y equipos relacionados, maquinaria eléctrica y otros aparatos eléctricos y líneas y cables de transmisión y suministro eléctrico.

Oficiales y operarios del tratamiento de la madera ebanistas y afines: Los instaladores y reparadores de electrónica y telecomunicaciones ajustan, mantienen, ajustan y reparan equipos electrónicos tales como máquinas comerciales y de oficina, instrumentos electrónicos y sistemas de control; instalar, reparar y mantener equipos de telecomunicaciones, equipos de transmisión de datos, cables y antenas; y reparar, ajustar y mantener computadoras.

Directores financieros: Los directores financieros planifican, dirigen y coordinan las operaciones financieras de una empresa u organización, en consulta con los altos directivos y con los responsables de otros departamentos o secciones, o de empresas que prestan servicios financieros a otras empresas y organizaciones.

Directores de recursos humanos: Los gestores de recursos humanos, planifican, dirigen y coordinan las políticas relativas a las actividades de personal, relaciones laborales y salud y seguridad en el trabajo de una empresa u organización, o de empresas que prestan

servicios de recursos humanos a otras empresas y organizaciones.

Directores de políticas y planificación: Los gestores de políticas y planificación planifican, organizan, dirigen y coordinan actividades de asesoramiento político y planificación estratégica dentro de la administración pública o para organizaciones no gubernamentales y organismos del sector privado, o gestionan las actividades de empresas que prestan servicios de planificación política y estratégica.

Directores de administración y servicios no clasificados bajo otros epígrafes: Este grupo de unidades abarca a los gerentes de servicios empresariales y de administración no clasificados en otro lugar del grupo menor 121: Gerentes de servicios empresariales y de administración. Por ejemplo, el grupo incluye ocupaciones como gerente de instalaciones, gerente de servicios de limpieza, gerente de servicios administrativos empleado ya sea como gerente de un departamento de una gran empresa y organización, o de una empresa que proporciona tales servicios a otras empresas y organizaciones.

Directores de ventas y comercialización: Los directores de ventas y marketing planifican, dirigen y coordinan las actividades de ventas y marketing de una empresa u organización, o de empresas que prestan servicios de ventas y marketing a otras empresas y organizaciones.

Directores de publicidad y relaciones públicas: Los directores de publicidad y relaciones públicas planifican, dirigen y coordinan las actividades de publicidad, relaciones públicas e información pública de empresas y organizaciones o de empresas que prestan servicios relacionados a otras empresas y organizaciones.

Directores de investigación y desarrollo: Los directores de investigación y desarrollo planifican, dirigen y coordinan las actividades de investigación y desarrollo de una empresa u organización o de empresas que prestan servicios relacionados a otras empresas y organizaciones.

Directores de producción agropecuaria y silvicultura: Los directores de producción agrícola y forestal planifican, dirigen y coordinan la producción en explotaciones agrícolas, hortícolas y forestales a gran escala, como plantaciones, grandes ranchos, granjas colectivas y cooperativas agrícolas para cultivar y cosechar cosechas, y criar y criar ganado.

Directores de producción de piscicultura y pesca: Los gestores de producción acuícola y pesquera planifican, dirigen y coordinan la producción en operaciones acuícolas y pesqueras a gran escala para capturar y recolectar peces y mariscos, y para cultivar peces, mariscos u otras formas de vida acuática, como cultivos comerciales o para su liberación en agua dulce o salada.

Directores de industrias manufactureras: Los directores de fabricación planifican, dirigen y coordinan las actividades relacionadas con la producción de bienes, la producción y distribución de electricidad, gas y agua, y la recogida, tratamiento y eliminación de residuos. Pueden dirigir los departamentos de producción de grandes empresas o ser los gerentes de pequeñas empresas manufactureras.

Directores de explotaciones de minería: Los gerentes de minas planifican, dirigen y coordinan la producción de las actividades de las explotaciones mineras, canteras y extracción de petróleo y gas, ya sea como gerente de un departamento o como gerente general de una empresa u organización que no tiene una jerarquía de gerentes.

Directores de empresas de construcción: Los jefes de obra planifican, dirigen y coordinan la construcción de proyectos de ingeniería civil, edificios y viviendas, ya sea como responsables de un departamento o como directores generales de una empresa u organización que no tiene una jerarquía de directivos.

Directores de empresas de abastecimiento, distribución y afines: Los gestores de suministro, distribución y afines planifican, dirigen y coordinan los sistemas e instalaciones de transporte de pasajeros y el suministro, transporte, almacenamiento y distribución de mercancías, ya sea como gestor de un departamento o como director general de una empresa u organización que no tenga una jerarquía de gestores.

Directores de servicios de tecnología de la información y las comunicaciones: Los gestores de servicios de tecnologías de la información y la comunicación planifican, dirigen y coordinan la adquisición, el desarrollo, el mantenimiento y el uso de sistemas informáticos y de telecomunicaciones, ya sea como responsables de un departamento o como directores generales de una empresa u organización que no tiene una jerarquía de directivos.

Directores de servicios de cuidados infantiles: Los gestores de servicios de cuidado infantil planifican, dirigen, y coordinan y evalúan la prestación de cuidados a niños en centros y servicios preescolares, extraescolares, de vacaciones y de día.

Directores de servicios de salud: Los gestores de servicios sanitarios planifican, dirigen, coordinan y evalúan la prestación de servicios sanitarios clínicos y comunitarios en hospitales, clínicas, organismos de salud pública y organizaciones similares.

Directores de servicios de cuidado de las personas de edad: Los gestores de servicios de atención a la tercera edad planifican, dirigen, coordinan y evalúan la prestación de servicios de atención residencial y personal a personas y familias que los necesitan debido a

los efectos del envejecimiento.

Directores de servicios de bienestar social: Los gestores de bienestar social planifican, dirigen y coordinan la prestación de programas de servicios sociales y comunitarios, como ayudas a la renta, asistencia familiar, servicios infantiles y otros programas y servicios comunitarios.

Directores de servicios de educación: Los gestores educativos planifican, dirigen, coordinan y evalúan los aspectos educativos y administrativos de los servicios educativos, los centros de enseñanza primaria y secundaria, los institutos y las facultades y departamentos de las universidades y otras instituciones educativas.

Gerentes de sucursales de bancos, de servicios financieros y de seguros: Los directores de sucursales de servicios financieros y de seguros planifican, dirigen y coordinan las sucursales de instituciones que prestan servicios financieros y de seguros, como bancos, sociedades de crédito hipotecario, cooperativas de crédito y compañías de seguros. Asesoran y ayudan a los clientes en cuestiones financieras y de seguros.

Directores y gerentes de servicios profesionales no clasificados bajo otros epígrafes: Este grupo de unidades abarca a los gestores que planifican, dirigen, coordinan y evalúan la prestación de servicios profesionales y técnicos especializados y no están clasificados en el grupo menor 121: Gestores de servicios empresariales y administrativos, ni en el grupo menor 134: Directores de servicios profesionales. Por ejemplo, se clasifican aquí los directivos responsables de la prestación de servicios policiales, correccionales, bibliotecarios, jurídicos y de bomberos.

Gerentes de hoteles: Los directores de hotel planifican, organizan y dirigen las operaciones de hoteles, moteles y establecimientos similares para proporcionar alojamiento a los huéspedes y otros servicios.

Gerentes de restaurantes: Los gerentes de restaurantes planifican, organizan y dirigen las operaciones de cafeterías, restaurantes y establecimientos relacionados para proporcionar servicios de restauración y catering.

Gerentes de comercios al por mayor y al por menor: Directores de comercio al por menor y al por mayor, planifican, organizan, coordinan y controlan las operaciones de establecimientos que venden mercancías al por menor o al por mayor. Son responsables de los presupuestos, la dotación de personal y la dirección estratégica y operativa de los comercios, o de las unidades organizativas dentro de los comercios que venden determinados tipos de productos.

Gerentes de servicios no clasificados bajo otros epígrafes: Este grupo de unidades abarca a los directivos que planifican, dirigen y coordinan la prestación de servicios y no están clasificados en el subgrupo principal 13: Directivos de producción y servicios especializados ni en el subgrupo principal 14: Directivos de hostelería, comercio minorista y otros servicios. Por ejemplo, se clasifican aquí los directores de agencias de viajes, centros de conferencias, centros de contacto y centros comerciales.

Físicos y astrónomos: Los físicos y astrónomos investigan y mejoran o desarrollan conceptos, teorías y métodos operativos relativos a la materia, el espacio, el tiempo, la energía, las fuerzas y los campos y la interrelación entre estos fenómenos físicos. Aplican los conocimientos científicos relacionados con la física y la astronomía en campos industriales, médicos, militares u otros.

Meteorólogos: Los meteorólogos elaboran previsiones meteorológicas a corto o largo plazo que se utilizan en la aviación, la navegación, la agricultura y otros ámbitos y para información del público en general. Realizan investigaciones relacionadas con la composición, estructura y dinámica de la atmósfera.

Químicos: Los químicos investigan, mejoran o desarrollan conceptos, teorías y métodos operativos, o aplican conocimientos científicos relacionados con la química, para desarrollar nuevos conocimientos o productos y para el control de la calidad y los procesos.

Geólogos y geofísicos: Los geólogos y geofísicos investigan, mejoran o desarrollan conceptos, teorías y métodos operativos, o aplican conocimientos científicos relacionados con la geología y la geofísica en campos como la exploración y extracción de petróleo, gas y minerales, la conservación del agua, la ingeniería civil, las telecomunicaciones y la navegación, y la evaluación y mitigación de los efectos de los proyectos de desarrollo y eliminación de residuos en el medio ambiente.

Matemáticos, actuarios y estadísticos: Los matemáticos, actuarios y estadísticos investigan y mejoran o desarrollan conceptos, teorías y métodos operativos y técnicos matemáticos, actuariales y estadísticos, y asesoran o se dedican a su aplicación práctica en campos como la ingeniería, la empresa y las ciencias sociales y de otro tipo.

Ingenieros industriales y de producción: Los ingenieros industriales y de producción investigan y diseñan, organizan y supervisan la construcción, el funcionamiento y el mantenimiento de procesos e instalaciones de producción industrial. Establecen programas de coordinación de las actividades de fabricación y evalúan la rentabilidad y la seguridad.

Ingenieros civiles: Los ingenieros civiles investigan, asesoran, diseñan y dirigen la construcción; gestionan la explotación y el mantenimiento de estructuras de ingeniería civil; o estudian y asesoran sobre aspectos tecnológicos de determinados materiales.

Ingenieros medioambientales: Los ingenieros medioambientales investigan, asesoran, diseñan y dirigen la aplicación de soluciones para prevenir, controlar o remediar los efectos negativos de la actividad humana en el medio ambiente utilizando diversas disciplinas de la ingeniería. Realizan evaluaciones medioambientales de proyectos de construcción e ingeniería civil y aplican principios de ingeniería al control de la contaminación, el reciclaje y la eliminación de residuos.

Ingenieros mecánicos: Los ingenieros mecánicos investigan, asesoran, diseñan y dirigen la producción de máquinas, aviones, barcos, maquinaria e instalaciones industriales, equipos y sistemas; asesoran y dirigen su funcionamiento, mantenimiento y reparación; o estudian y asesoran sobre aspectos mecánicos de determinados materiales, productos o procesos.

Ingenieros químicos: Los ingenieros químicos investigan y desarrollan, asesoran y dirigen procesos químicos a escala comercial y la producción de diversas sustancias y artículos como petróleo crudo, derivados del petróleo, productos alimenticios y bebidas, medicamentos o materiales sintéticos. Dirigen el mantenimiento y la reparación de plantas y equipos químicos y estudian y asesoran sobre aspectos químicos de materiales, productos o procesos concretos.

Ingenieros de minas, metalúrgicos y afines: Los ingenieros de minas, metalúrgicos y profesionales afines investigan, diseñan, desarrollan y mantienen métodos a escala comercial para extraer metales de sus menas, o minerales, agua, petróleo o gas de la tierra, y para desarrollar nuevas aleaciones, cerámicas y otros materiales, o estudian y asesoran sobre aspectos mineros o metalúrgicos de materiales, productos o procesos concretos.

Ingenieros no clasificados bajo otros epígrafes: Este grupo de unidades abarca a los profesionales de la ingeniería no clasificados en el grupo menor 214: Profesionales de la ingeniería (excluida la electrotecnología) o en el grupo menor 215: Ingenieros en Electrotecnia. Por ejemplo, el grupo incluye a quienes investigan y asesoran o desarrollan procedimientos y soluciones de ingeniería relacionados con la seguridad en el lugar de trabajo, la ingeniería biomédica, la óptica, los materiales, la generación de energía nuclear y los explosivos.

Ingenieros electricistas: Los ingenieros eléctricos investigan y asesoran, diseñan y dirigen la construcción y el funcionamiento de sistemas, componentes, motores y equipos eléctricos, y asesoran y dirigen su funcionamiento, mantenimiento y reparación, o estudian y asesoran sobre aspectos tecnológicos de materiales, productos y procesos de ingeniería eléctrica.

Ingenieros electrónicos: Los ingenieros electrónicos investigan, diseñan y dirigen el funcionamiento de la construcción, el mantenimiento y la reparación de sistemas electrónicos, y estudian y asesoran sobre aspectos tecnológicos de materiales, productos o procesos de ingeniería electrónica.

Ingenieros en telecomunicaciones: Los ingenieros de telecomunicaciones investigan y asesoran, diseñan y dirigen la construcción, el funcionamiento, el mantenimiento y la reparación de sistemas y equipos de telecomunicaciones. Estudian y asesoran sobre aspectos tecnológicos de materiales, productos o procesos de ingeniería de telecomunicaciones.

Arquitectos: Los arquitectos diseñan edificios comerciales, industriales, institucionales, residenciales y recreativos, y planifican y supervisan su construcción, mantenimiento y rehabilitación.

Arquitectos paisajistas: Los arquitectos paisajistas planifican y diseñan paisajes y espacios abiertos para proyectos como parques, escuelas, instituciones, carreteras, zonas exteriores de centros comerciales, industriales y residenciales, y planifican y supervisan su construcción, mantenimiento y rehabilitación.

Urbanistas e ingenieros de tránsito: Los urbanistas y planificadores de tráfico elaboran y aplican planes y políticas para el uso controlado del suelo urbano y rural y para los sistemas de tráfico. Investigan y asesoran sobre los factores económicos, medioambientales y sociales que afectan al uso del suelo y a los flujos de tráfico.

Cartógrafos y agrimensores: Los cartógrafos y topógrafos determinan la posición exacta o preparan y revisan mapas digitales, gráficos y pictóricos, cartas u otras representaciones visuales de características naturales y construidas y límites de tierras, mares, zonas subterráneas y cuerpos celestes, aplicando principios científicos y matemáticos.

Farmacéuticos: Los farmacéuticos almacenan, conservan, componen y dispensan medicamentos y asesoran sobre el uso adecuado y los efectos adversos de los fármacos y medicamentos siguiendo las recetas emitidas por médicos y otros profesionales de la salud. Contribuyen a investigar, probar, preparar, prescribir y supervisar terapias medicinales para optimizar la salud humana.

Contables: Los contables planifican, organizan y administran sistemas contables para particulares y establecimientos. Algunas ocupaciones clasificadas aquí examinan y analizan los registros contables y financieros de particulares y establecimientos para garantizar su exactitud y el cumplimiento de las normas y procedimientos contables establecidos.

Asesores financieros y en inversiones: Los asesores financieros y de inversión elaboran planes financieros para particulares y organizaciones, e invierten y gestionan fondos en su nombre.

Analistas financieros: Los analistas financieros realizan análisis cuantitativos de la información que afecta a los programas de inversión de instituciones públicas o privadas.

Analistas de gestión y organización: Los analistas de gestión y organización ayudan a las organizaciones a lograr una mayor eficiencia y a resolver problemas organizativos. Estudian estructuras, métodos, sistemas y procedimientos organizativos.

Especialistas en políticas de administración: Los profesionales de la administración política elaboran y analizan políticas que orientan el diseño, la aplicación y la modificación de operaciones y programas gubernamentales y comerciales.

Especialistas en políticas y servicios de personal y afines: Los profesionales de personal y carreras profesionales prestan servicios empresariales profesionales relacionados con las políticas de personal, como la contratación o el desarrollo de empleados, el análisis ocupacional y la orientación profesional.

Especialistas en formación del personal: Los profesionales de la formación y el desarrollo del personal planifican, desarrollan, aplican y evalúan programas de formación y desarrollo para garantizar que la dirección y el personal adquieren las habilidades y desarrollan las competencias que necesitan las organizaciones para cumplir sus objetivos.

Profesionales de la publicidad y la comercialización: Los profesionales de la publicidad y el marketing desarrollan y coordinan estrategias y campañas publicitarias, determinan el mercado para nuevos bienes y servicios e identifican y desarrollan oportunidades de mercado para bienes y servicios nuevos y existentes.

Profesionales de relaciones públicas: Los profesionales de las relaciones públicas planifican, desarrollan, aplican y evalúan estrategias de información y comunicación que crean una comprensión y una opinión favorable de las empresas y otras organizaciones, sus bienes y servicios, y su papel en la comunidad.

Profesionales de ventas técnicas y médicas (excluyendo la TIC): Los profesionales de ventas técnicas y médicas (excluidas las TIC) representan a empresas en la venta de una serie de bienes y servicios industriales, médicos y farmacéuticos a establecimientos industriales, comerciales, profesionales y de otro tipo.

Profesionales de ventas de tecnología de la información y las comunicaciones:

Los profesionales de ventas de tecnologías de la información y la comunicación (TIC) venden, a nivel mayorista, una gama de equipos y programas informáticos y otros bienes y servicios de tecnologías de la información y la comunicación, incluidas instalaciones, y proporcionan información especializada según sea necesario.

Analistas de sistemas: Los analistas de sistemas investigan, analizan y evalúan los requisitos, procedimientos o problemas informáticos de los clientes, y elaboran y aplican propuestas, recomendaciones y planes para mejorar los sistemas de información actuales o futuros.

Desarrolladores de software: Los desarrolladores de software investigan, analizan y evalúan los requisitos de aplicaciones de software y sistemas operativos nuevos o existentes, y diseñan, desarrollan, prueban y mantienen soluciones de software para satisfacer dichos requisitos.

Desarrolladores Web y multimedia: Los profesionales del desarrollo web y multimedia combinan conocimientos técnicos y de diseño para investigar, analizar, evaluar, diseñar, programar y modificar sitios web y aplicaciones que combinan texto, gráficos, animaciones, imágenes, audio y vídeo y otros medios interactivos.

Programadores de aplicaciones: Los programadores de aplicaciones escriben y mantienen código programable descrito en instrucciones y especificaciones técnicas para aplicaciones de software y sistemas operativos.

Desarrolladores y analistas de software y multimedia y analistas no clasificados bajo otros epígrafes: Este grupo de unidades incluye a los desarrolladores y analistas de software y aplicaciones no clasificados en el grupo menor 251: Desarrolladores y analistas de software y aplicaciones. Por ejemplo, el grupo incluye a los profesionales especializados en garantía de calidad, incluidas las pruebas de software.

Diseñadores y administradores de bases de datos: Los diseñadores y administradores de bases de datos diseñan, desarrollan, controlan, mantienen y respaldan el rendimiento y la seguridad óptimos de las bases de datos.

Administradores de sistemas: Los administradores de sistemas desarrollan, controlan, mantienen y respaldan el rendimiento y la seguridad óptimos de los sistemas informáticos.

Profesionales en redes de computadores: Los profesionales de redes informáticas investigan, analizan, diseñan, prueban y recomiendan estrategias para la arquitectura y el desarrollo de redes. Implantan, gestionan, mantienen y configuran hardware y software de red, y supervisan, solucionan problemas y optimizan el rendimiento.

Especialistas en bases de datos y en redes de computadores no clasificados bajo otros epígrafes: Este grupo de unidades incluye a los profesionales de bases de datos y redes no clasificados en otro lugar del grupo menor 252: Profesionales de bases de datos y redes. Por ejemplo, el grupo incluye a los especialistas en seguridad de las tecnologías de la información y las comunicaciones.

Abogados: Los abogados asesoran a sus clientes sobre una gran variedad de temas, redactan documentos jurídicos, representan a sus clientes ante consejos o tribunales administrativos y defienden casos o llevan a cabo procesamientos en tribunales de justicia, o instruyen a abogados para que defiendan casos ante tribunales superiores de justicia. **Jueces:** Los jueces presiden los procedimientos civiles y penales en los tribunales de justicia. **Profesionales en derecho no clasificados bajo otros epígrafes:** Este grupo de unidades abarca a los profesionales del Derecho no clasificados en otro lugar del grupo menor 261: Profesionales del Derecho. Por ejemplo, el grupo incluye a quienes desempeñan funciones jurídicas distintas de la defensa o el enjuiciamiento de causas o la presidencia de procedimientos judiciales.

Archivistas y curadores de museos: Los archiveros y conservadores recopilan, valoran y garantizan la custodia y preservación del contenido de archivos, artefactos y registros de interés histórico, cultural, administrativo y artístico, así como de obras de arte y otros objetos. Planifican, conciben y aplican sistemas de conservación de archivos y documentos de valor histórico.

Bibliotecarios, documentalistas y afines: Los bibliotecarios y profesionales de la información afines recopilan, seleccionan, desarrollan, organizan y mantienen las colecciones de la biblioteca y otros depósitos de información, organizan y controlan otros servicios bibliotecarios y proporcionan información a los usuarios.

Directores de cine, de teatro y afines: Los directores y productores de cine, teatro y afines supervisan y controlan los aspectos técnicos y artísticos de las películas cinematográficas, las producciones televisivas o radiofónicas y los espectáculos teatrales.

Supervisores de la construcción: Supervisores de construcción, coordinan, supervisan y programan las actividades de los trabajadores dedicados a la construcción y reparación de edificios y estructuras.

Técnicos en control de procesos no clasificados bajo otros epígrafes: Este grupo de unidades incluye a los técnicos de control de procesos no clasificados en el grupo menor 313: Técnicos de control de procesos. Por ejemplo, el grupo unitario incluye a quienes manejan equipos de control de procesos múltiples en líneas de montaje de fabricación y en

la producción de papel y pasta de papel.

Técnicos en aparatos de diagnóstico y tratamiento médico: Los técnicos en imagen médica y equipos terapéuticos prueban y manejan equipos radiográficos, de ultrasonidos y otros equipos de imagen médica para producir imágenes de estructuras corporales para el diagnóstico y tratamiento de lesiones, enfermedades y otras deficiencias. Pueden administrar tratamientos de radiación a pacientes bajo la supervisión de un radiólogo u otro profesional sanitario.

Agentes de bolsa, cambio y otros servicios financieros: Los agentes y corredores de valores y finanzas compran y venden valores, acciones, bonos y otros instrumentos financieros, y negocian en los mercados de divisas, al contado o de futuros, en nombre de su propia empresa o para clientes a comisión. Recomiendan operaciones a clientes o altos directivos.

Oficiales de préstamos y créditos: Los responsables de créditos y préstamos analizan y evalúan la información financiera de las solicitudes de créditos y préstamos y determinan la aprobación o desaprobación del cliente para el crédito o préstamo, o recomiendan a la dirección su aprobación o desaprobación.

Tenedores de libros: Los profesionales asociados de la contabilidad mantienen registros completos de las transacciones financieras de una empresa y verifican la exactitud de los documentos y registros relativos a dichas transacciones.

Profesionales de nivel medio de servicios estadísticos, matemáticos y afines: Los profesionales asociados de estadística, matemáticas y afines ayudan a planificar la recogida, el tratamiento y la presentación de datos estadísticos, matemáticos o actuariales y a llevar a cabo estas operaciones, trabajando normalmente bajo la dirección de estadísticos, matemáticos y actuarios.

Tasadores: Los tasadores y evaluadores de siniestros valoran propiedades y bienes diversos y evalúan las pérdidas cubiertas por las pólizas de seguros.

Agentes de seguros: Los representantes de seguros asesoran y venden seguros de vida, accidente, automóvil, responsabilidad civil, dotación, incendio, marítimo y otros tipos de seguros a clientes nuevos y establecidos.

Representantes comerciales: Los representantes de ventas comerciales representan a empresas para vender diversos bienes y servicios a empresas y otras organizaciones y proporcionan información específica sobre productos según sea necesario.

Agentes de compras: Los compradores adquieren bienes y servicios para su uso o reventa por cuenta de establecimientos y organizaciones industriales, comerciales, gubernamentales o de otro tipo. **Agentes de compras y consignatarios:** Los corredores de comercio compran y venden mercancías y servicios de transporte, normalmente a granel, en nombre de su propia empresa o para clientes a comisión.

Declarantes o gestores de aduana: Los agentes de aduanas y transitarios realizan los trámites aduaneros y se aseguran de que el seguro, las licencias de exportación/importación y otras formalidades estén en regla.

Organizadores de conferencias y eventos: Los organizadores de conferencias y actos organizan y coordinan servicios para conferencias, actos, funciones, banquetes y seminarios.

Agentes de empleo y contratistas de mano de obra: Los agentes de empleo y los contratistas ponen en contacto a demandantes de empleo con vacantes, encuentran trabajadores para empresarios y contratan mano de obra para proyectos concretos a petición de empresas y otras organizaciones, incluidas las administraciones públicas y otras instituciones, o encuentran plazas para demandantes de empleo a cambio de una comisión.

Agentes inmobiliarios: Los agentes inmobiliarios y los administradores de fincas organizan la venta, la compra, el alquiler y el arrendamiento de bienes inmuebles, normalmente en nombre de clientes y a comisión.

Agentes de servicios comerciales no clasificados bajo otros epígrafes: Este grupo de unidades abarca a los agentes de servicios empresariales no clasificados en otro lugar del grupo menor 333: Agentes de servicios empresariales. Por ejemplo, el grupo incluye a quienes establecen contactos comerciales, venden servicios empresariales como espacios publicitarios en los medios de comunicación, conciertan contratos para actuaciones de deportistas, artistas del espectáculo y artistas, para la publicación de libros, la producción de obras de teatro o la grabación, interpretación y venta de música, venden propiedades y bienes mediante subasta y quienes diseñan y organizan viajes combinados y en grupo.

Supervisores de secretaría: Los supervisores de oficina supervisan y coordinan las actividades de los trabajadores del grupo principal 4: trabajadores administrativos de apoyo.

Secretarios jurídicos: Los secretarios jurídicos aplican conocimientos especializados de terminología y procedimientos jurídicos para prestar apoyo a los profesionales del Derecho en actividades de comunicación, documentación y coordinación interna de la gestión, en despachos de abogados, departamentos jurídicos de grandes empresas y administraciones públicas.

Secretarios administrativos y ejecutivos: Los secretarios administrativos y ejecutivos realizan tareas de enlace, coordinación y organización en apoyo de directivos y profesionales y/o preparan correspondencia, informes y actas de procedimientos y otra documentación especializada.

Secretarios médicos: Los secretarios médicos, utilizando conocimientos especializados de terminología médica y procedimientos de prestación de asistencia sanitaria, ayudan a los profesionales sanitarios y a otros trabajadores realizando diversas funciones de comunicación, documentación, administración y coordinación interna, para apoyar a los trabajadores sanitarios en centros médicos y otras organizaciones relacionadas con la asistencia sanitaria.

Profesionales de nivel medio del derecho y servicios legales y afines: Los profesionales jurídicos y afines asociados desempeñan funciones de apoyo en tribunales o despachos de abogados, prestan servicios relacionados con asuntos jurídicos como contratos de seguros, transmisión de propiedades y concesión de préstamos y otras transacciones financieras o realizan investigaciones para clientes.

Fotógrafos: Los fotógrafos utilizan cámaras fijas para fotografiar personas, acontecimientos, escenas, materiales, productos y otros objetos.

Técnicos en operaciones de tecnología de la información y las comunicaciones: Los técnicos de operaciones de tecnología de la información y las comunicaciones y de atención al usuario prestan apoyo al funcionamiento cotidiano de los sistemas de comunicaciones, los sistemas informáticos y las redes, y ofrecen asistencia técnica a los usuarios.

Técnicos en asistencia al usuario de tecnología de la información y las comunicaciones: Los técnicos de asistencia al usuario de tecnologías de la información y la comunicación prestan asistencia técnica a los usuarios, ya sea directamente o por teléfono, correo electrónico u otros medios electrónicos, lo que incluye diagnosticar y resolver cuestiones y problemas de software, hardware, equipos periféricos informáticos, redes, bases de datos e Internet, y ofrecer orientación y apoyo en la implantación, instalación y mantenimiento de sistemas.

Técnicos en redes y sistemas de computadores: Los técnicos de redes y sistemas informáticos establecen, operan y mantienen redes y otros sistemas de comunicación de datos.

Técnicos de la Web: Los técnicos web mantienen, supervisan y apoyan el funcionamiento óptimo de los sitios web de Internet e Intranet y del hardware y software de los servidores web.

Técnicos de radiodifusión y grabación audio visual: Los técnicos de radiodifusión y audiovisuales controlan el funcionamiento técnico de los equipos de grabación y edición de imágenes y sonido y de transmisión de emisiones de radio y televisión de imágenes y sonidos, así como de otros tipos de señales de telecomunicación por tierra, mar o aire.

Técnicos de ingeniería de las telecomunicaciones: Los técnicos en ingeniería de telecomunicaciones realizan tareas técnicas relacionadas con la investigación en ingeniería de telecomunicaciones, así como con el diseño, la fabricación, el montaje, la construcción, el funcionamiento, el mantenimiento y la reparación de sistemas de telecomunicaciones.

Operadores de máquinas de procesamiento de texto y mecanógrafos: Los mecanógrafos y operadores de tratamiento de textos escriben, editan e imprimen textos utilizando máquinas de escribir, ordenadores personales u otros procesadores de texto, y taquígrafían asuntos orales o escritos.

Grabadores de datos: Los empleados de entrada de datos introducen datos codificados, estadísticos, financieros y otros datos numéricos en equipos electrónicos, bases de datos informatizadas, hojas de cálculo u otros depósitos de datos utilizando un teclado, ratón, escáner óptico, software de reconocimiento de voz u otras herramientas de entrada de datos. Introducen datos en dispositivos mecánicos y electrónicos para realizar cálculos matemáticos.

Empleados de agencias de viajes: Los asesores y empleados de viajes proporcionan información sobre los destinos de los viajes, organizan itinerarios de viaje, obtienen reservas de viajes y alojamiento y registran a los pasajeros en la facturación y la salida.

Empleados de contabilidad y cálculo de costos: Los empleados de contabilidad y teneduría de libros calculan, clasifican y registran datos numéricos para mantener completos los registros financieros. Realizan cualquier combinación de tareas rutinarias de cálculo, contabilización y verificación para obtener datos financieros primarios que se utilizan en el mantenimiento de los registros contables.

Empleados de servicios estadísticos, financieros y de seguros: Los empleados de contabilidad y teneduría de libros calculan, clasifican y registran datos numéricos para mantener completos los registros financieros. Realizan cualquier combinación de tareas rutinarias de cálculo, contabilización y verificación para obtener datos financieros primarios que se utilizan en el mantenimiento de los registros contables.

Empleados encargados de las nóminas: Los empleados encargados de las nóminas recopilan, verifican y procesan la información de las nóminas y calculan el salario y los derechos a prestaciones de los empleados de un departamento, empresa u otro establecimiento.

Instaladores y reparadores de líneas eléctricas: Los instaladores y reparadores de líneas eléctricas instalan, reparan y unen cables eléctricos de transmisión y suministro y equipos relacionados.

Mecánicos y reparadores en electrónica: Los mecánicos y técnicos electrónicos instalan, mantienen, ajustan y reparan equipos electrónicos como máquinas comerciales y de oficina e instrumentos electrónicos y sistemas de control.

Instaladores y reparadores en tecnología de la información y las comunicaciones: Los instaladores y mantenedores de tecnologías de la información y la comunicación (TIC) instalan, reparan y mantienen equipos de telecomunicaciones, equipos de transmisión de datos, cables, antenas y conductos, y reparan, instalan y mantienen ordenadores.

Competencias: La competencia debe ser entendida como un elemento que integra aspectos que tienen que ver con conocimientos, habilidades y valores, es decir comprende aspectos de tipo cognitivo, procedimental y actitudinal interrelacionados en la búsqueda de desempeños eficientes en entornos de trabajo asociados a un campo laboral concreto; desde esta perspectiva, la competencia es integral e integradora.

Competencias básicas: En el contexto internacional reciben distintos nombres tales como: core skills, key competences, basic skills, entre otros. Permiten el ingreso al trabajo o a la educación superior, por ello deben ser identificadas desde ambos sectores. Para los que ingresan al mundo del trabajo, se consideran como requisitos mínimos necesarios no solo para el desempeño de una ocupación u oficio, sino, y prioritariamente, para desenvolverse adecuadamente en los espacios sociales y ciudadanos en donde se desenvuelve la vida misma. En lo que se refiere a la educación, determinan tanto el perfil de ingreso a la educación superior, como los fundamentos de competencias más complejas que se desarrollaran a lo largo de la formación profesional, en especial los procesos de formación que deben ser introducidos en los programas de los ciclos propedéuticos.

Competencias transversales: Son competencias requeridas en un amplio campo de profesiones y ocupaciones y aportan las herramientas requeridas por un trabajador profesional para analizar los problemas, evaluar las estrategias a utilizar y aportar soluciones pertinentes en situaciones nuevas. Están presentes por lo general en la mayoría de las labores que se le presentan a un sujeto en los distintos campos profesionales.

Competencias instrumentales: Reciben su nombre porque brindan las herramientas claves tanto para el aprendizaje como para el desempeño en el mundo del trabajo. Comprenden una serie de habilidades como las cognoscitivas -que permiten comprender y procesar ideas y

pensamientos; las metodológicas -que dan las herramientas para organizar eficientemente el tiempo, el aprendizaje mismo y tomar decisiones o solucionar problemas; y las tecnológicas -relacionadas con el uso de equipos, incluyendo las TIC y gerenciar la información. Comprende capacidades como: de análisis y síntesis, de organizar y planificar, de hacer uso adecuado de los conocimientos de la profesión, de comunicación oral y escrita, de uso adecuado de una segunda lengua, de manejo de las TIC, de gestión de la información, de resolución de problemas, de toma de decisiones y de crear situaciones seguras en los diversos contextos.

Competencias interpersonales: Son las que permiten mantener una buena relación social y un adecuado comportamiento ciudadano. Se relacionan con la capacidad de expresar los sentimientos, de hacer planteamientos críticos y de autocrítica, de participar en la vida política y de asumir los deberes y derechos ciudadanos en condiciones éticas. Desarrollan en la persona la capacidad de trabajar en equipo, interactuar social y políticamente y cooperar con el desarrollo de su entorno. Comprende capacidades como: crítica y autocrítica, de manejo de conflictos, de trabajo en equipos uni- y multidisciplinares, de aceptación y respeto de la diversidad y la multiculturalidad, de conocimiento y respeto por las costumbres de otras regiones del país y de otros países y de compromiso social y ético.

Competencias sistémicas: Están relacionadas con la visión de conjunto y la capacidad de gestionar integralmente los procesos organizacionales. Se logran mediante una combinación de comprensión, sensibilidad y conocimientos que permiten identificar las partes de un todo y las relaciones entre las partes que generan la estructura de totalidad.

Competencias específicas: Son las requeridas para el desempeño de una ocupación en concreto, están relacionadas más con funciones o puestos de trabajo. Aportan al estudiante o al trabajador los conocimientos, actitudes, habilidades y valores propios de cada profesión y actividad laboral.

Nivel de estudio: El nivel de formación alcanzado por una persona se define como "el máximo nivel que el individuo ha completado" se mide a partir del programa de máximo nivel que haya superado, validado, normalmente, por la correspondiente certificación.

Técnico: Relativo a programas Técnicos Profesionales.

Tecnólogo: Relativo a programas tecnológicos.

Profesional: Relativo a programas profesionales universitarios.

Especialista: relativas a programas de Especialización Técnica Profesional, Especialización Tecnológica y Especializaciones Profesionales, niveles posgrado. Las especializaciones suelen te-

ner una duración de un año y, como su nombre lo indica, tienen el objetivo de que los profesionales se especialicen en alguno de los campos de su disciplina.

Magíster: Una maestría tiene una duración mayor a la de una especialización, dos años aproximadamente. Su objetivo es que los estudiantes obtengan un conocimiento más amplio sobre un área en específico. A su vez, en Colombia hay dos tipos de maestrías: de profundización y de investigación.

Doctor/Doctora: Un doctorado es un nivel de estudio de alto nivel superado solo por el posdoctorado. Su duración es mucho más larga que los demás posgrados, incluso más que un pregrado.

Experiencia: conjunto de aptitudes y conocimientos adquiridos por una persona o grupo en un determinado puesto laboral, o durante un periodo de tiempo específico.

Junior: Experiencia laboral adquirida entre 1 y 3 años.

Máster: Experiencia laboral adquirida entre 3 y 5 años.

Senior: Experiencia laboral adquirida entre 5 años o más.

B. Anexo: Código en Python usando la librería Owlready2 para construcción de la ontología base

```
# Importar la libreria de Ontologias
! pip install Owlready2
from owlready2 import *
# https://owlready2.readthedocs.io/en/v0.37/intro.html

import datetime
now = str(datetime.datetime.now()).replace('-', '')
# Definir el IRI de la Ontologia
onto = []
onto = get_ontology("http://ti-e-recruitment-2"+now+".org/onto")
# Comentario general de la Ontologia
onto.metadata.comment.append("Ontologia-de-perfiles-de-TI
para-aplicacion-de-e-recruitment")

# Diccionario vacio para las clases de la ontologia
class_dict = {}
# lista de los id de las ocupaciones de nivel mas bajo
relation_l = []

# Construir la ontologia
with onto:
    # SuperClase de Ocupaciones
    class_ocu = types.new_class('Ocupaciones', (Thing,))
    # Comentario de la SuperClase de Ocupaciones

    class_ocu.comment = ["""Se refiere al tipo de trabajo realizado
en un trabajo. El concepto de ocupacion se define como un conjunto
de puestos de trabajo cuyas tareas y funciones
```

```

principales se caracterizan por un alto grado
de similitud. Una persona puede estar asociada
a una ocupacion a traves del trabajo principal
que tiene actualmente, un segundo trabajo,
un trabajo futuro o un trabajo anterior."""]
# SuperClase de Competencias
class_com = types.new_class('Competencias', (Thing,))
# Comentario de la SuperClase de Competencias
class_com.comment = ['Es-la-capacidad-de-llevar-a
-cabo-las-tareas-y-deberes-de-un-trabajo. ']

# Se recorren todas las competencias
for id, ocu in ocu_d.items():

    # Caso de Super Ocupacion
    if len(id) == 1:
        # Se crea la clase
        class_dict[id] = types.new_class(ocu, (class_ocu,))
        # Descripcion de la ocupacion como comentario
        class_dict[id].comment = [desc_d[id]]
    # Caso de minima Ocupacion
    elif len(id) == 4:
        # Se crean como instancias y no como clases
        relation_l.append(id)
    # Casos intermedios
    else:
        # print(id, len(id), ocu)
        try:
            # Se crea la clase
            class_dict[id] = types.new_class(ocu, (class_dict[id[:-1]],))
            # Descripcion de la ocupacion como comentario
            class_dict[id].comment = [desc_d[id]]
        except:
            print('Try-Error:-', ocu, 'id:-', id)

# Se definen las ObjectProperty con su dominio y rango
class tiene_competencia(ObjectProperty, AsymmetricProperty,
IrreflexiveProperty):
    domain = [class_ocu]
    range = [class_com]

```

```

# Se definen las DataProperty con su rango
class id(DataProperty, FunctionalProperty):
    range      = [str]
    domain     = [class_ocu, class_com]

    # Characteristic: Functional
class descripcion(DataProperty):
    range      = [str]
    domain     = [class_ocu]

# Se recorren todas las instancias
for id in relation_l:
    # Lista vacia de competencia de cada instancia
    temp_comp_list = []
    # Se recorren las competencias
    cont = 0
    for comp in comp_d[id]:
        if len(comp):
            # Se declara la competencia como una instancia
            instancia_compe = class_com(id+str(cont))
            instancia_compe.label = comp
            temp_comp_list.append(instancia_compe)
            # Label de la clase

            cont += 1
            # if cont == 10:
            # break

# Se toma la clase a la que pertenece
try:
    supClass = class_dict[id[:-1]]
except:
    print('except: -', id)

# Se declara la ocupacion como una instancia
try:
    ocupacion = supClass(ocu_d[id])
except:
    ocupacion = supClass(ocu_d[id]+'_i')
```

```
# Se le asocian todas las competencias
ocupacion.tiene_competencia = temp_comp_list

# Se le asocia el dataProperty ID
ocupacion.id = id
# Se le asocia el dataProperty Descripcion
ocupacion.descripcion = [desc_d[id]]

# Se guarda la ontologia en formato .OWL
onto.save(file = "onto.owl")
```

C. Anexo: Extracción Código en Python para web scraping de LinkedIn

```
# Importar librerias
# https://pypi.org/project/linkedin-jobs-scraper/
!pip install linkedin-jobs-scraper
!pip install urllib3==1.24.2
!pip install attrs==19.1.0
!pip install certifi==2019.3.9
!pip install selenium==3.141.0
!apt-get update # to update ubuntu to correctly run apt install
!apt install chromium-chromedriver
!cp /usr/lib/chromium-browser/chromedriver /usr/bin
import sys
sys.path.insert(0, '/usr/lib/chromium-browser/chromedriver')

# CONFIGURACION del web driver de chrome para realizar el scraping

# Importar librerias
from selenium import webdriver
import time
import pandas as pd
import os
import datetime
import logging
from linkedin_jobs_scraper import LinkedInScraper
from linkedin_jobs_scraper.events import Events, EventData,
EventMetrics
from linkedin_jobs_scraper.query import Query, QueryOptions,
QueryFilters
from linkedin_jobs_scraper.filters import RelevanceFilters,
TimeFilters, TypeFilters, ExperienceLevelFilters,
RemoteFilters
```

```
from google.colab import files

# Definir la ruta del chrome driver
print(os.environ[ 'PATH' ])
PATH = "/usr/lib/chromium-browser/chromedriver"
# PATH = "/usr/bin/chromedriver"

# Configurar el web driver
chrome_options = webdriver.ChromeOptions()
chrome_options.add_argument( '--headless' )
chrome_options.add_argument( '--no-sandbox' )

# Crear listas vacas para llenarlas de info
title = []
company = []
company_link = []
date = []
link = []
insights = []
description = []
```

Bibliografía

- [Python Software Foundation, Ludovico Fabbri, 2022] Python Software Foundation, Ludovico Fabbri (2022). <https://pypi.org/project/linkedin-jobs-scraper/#usage>.
- [Alfonso-Hermelo et al., 2019a] Alfonso-Hermelo, D., Langlais, P., and Bourg, L. (2019a). Automatically learning a human-resource ontology from professional social-network data. In *Canadian Conference on Artificial Intelligence*, pages 132–145. Springer.
- [Alfonso-Hermelo et al., 2019b] Alfonso-Hermelo, D., Langlais, P., and Bourg, L. (2019b). Automatically learning a human-resource ontology from professional social-network data. In *Canadian Conference on Artificial Intelligence*, pages 132–145. Springer.
- [Alfonso-Hermelo et al., 2019c] Alfonso-Hermelo, D., Langlais, P., and Bourg, L. (2019c). *Automatically Learning a Human-Resource Ontology from Professional Social-Network Data*, pages 132–145. Springer, Cham.
- [Apostolico and Galil, 1997] Apostolico, A. and Galil, Z., editors (1997). *Pattern Matching Algorithms*. Oxford University Press.
- [Arendt et al., 2023] Arendt, L., Gałecka-Burdziak, E., Núñez, F., Pater, R., and Usabiaga, C. (2023). Skills requirements across task-content groups in poland: What online job offers tell us. *Technological Forecasting and Social Change*, 187:122245.
- [Arnal et al., 2001] Arnal, E., Ok, W., and Torres, R. (2001). Knowledge, work organisation and economic growth. In *Internet, economic growth and globalization*, pages 327–376. Springer.
- [Arroyo Mesa, 2022] Arroyo Mesa, L. (2022). Tendencias actuales en la búsqueda y selección de personal en el ámbito empresarial.
- [Asim et al., 2018a] Asim, M. N., Wasim, M., Khan, M. U. G., Mahmood, W., and Abbasi, H. M. (2018a). A survey of ontology learning techniques and applications. *Database*, 2018. bay101.
- [Asim et al., 2018b] Asim, M. N., Wasim, M., Khan, M. U. G., Mahmood, W., and Abbasi, H. M. (2018b). A survey of ontology learning techniques and applications. *Database : the journal of biological databases and curation*, 2018.

- [Balachander and Moh, 2018] Balachander, Y. and Moh, T.-S. (2018). Ontology based similarity for information technology skills. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 302–305.
- [Bechhofer et al., 2004] Bechhofer, S., Van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., Stein, L. A., et al. (2004). Owl web ontology language reference. *W3C recommendation*, 10(2):1–53.
- [Bernaras et al., 1996] Bernaras, A., Laresgoiti, I., and Corera, J. M. (1996). Building and reusing ontologies for electrical network applications. In *European Conference on Artificial Intelligence*.
- [BERNERS-LEE et al., 2001] BERNERS-LEE, T., HENDLER, J., and LASSILA, O. (2001). The semantic web. *Scientific American*, 284(5):34–43.
- [Borst, 1999] Borst, W. N. (1999). Construction of engineering ontologies for knowledge sharing and reuse. *CKL*.
- [Brenčič and Norris, 2010] Brenčič, V. and Norris, J. B. (2010). Do employers change job offers in their online job ads to facilitate search? *Economics Letters*, 108(1):46–48.
- [Brewster et al., 2004] Brewster, C., Alani, H., Dasmahapatra, S., and Wilks, Y. (2004). Data driven ontology evaluation. *International Conference on Language Resources and Evaluation, Lisbon, Portugal*.
- [Bui et al., 2021] Bui, T.-C., Le, V.-D., To, H.-T., and Cha, S. K. (2021). Generative pre-training for paraphrase generation by representing and predicting spans in exemplars. In *2021 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 83–90. IEEE.
- [Cañete et al., 2020] Cañete, J., Chaperon, G., Fuentes, R., Ho, J.-H., Kang, H., and Pérez, J. (2020). Spanish pre-trained bert model and evaluation data. In *PML4DC at ICLR 2020*.
- [CB Insights Research, 2020] CB Insights Research (3 de Abril de 2020). <https://www.cbinsights.com/research/hr-tech/f>.
- [Cer et al., 2018] Cer, D., Yang, Y., Kong, S.-y., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., et al. (2018). Universal sentence encoder. *arXiv preprint arXiv:1803.11175*.
- [Chala, 2018] Chala, S. A. (2018). Bidirectional job matching through unsupervised feature learning. In *Master’s thesis. OPUS Siegen Master’s thesis, Universität Siegen, URN 467-12858*.

- [Chen and Gu, 2021] Chen, J. and Gu, J. (2021). Adol: a novel framework for automatic domain ontology learning. *The Journal of Supercomputing*, 77(1):152–169.
- [Chiatti and Daga, 2022] Chiatti, A. and Daga, E. (2022). Neuro-symbolic learning for dealing with sparsity in cultural heritage image archives: an empirical journey.
- [Chow and Chapman, 2013] Chow, S. and Chapman, D. (2013). Gamifying the employee recruitment process. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, pages 91–94.
- [Coggle, 2021] Coggle (2021). Dataset validator relevant for iot ontologies. <https://coggle.it/diagram/WZ8ANciQegABugPa/t/how-to-validate-ontologies>.
- [Cordero Mena and Yunga Tucto, 2022] Cordero Mena, K. I. and Yunga Tucto, C. M. (2022). Diseño y desarrollo de un chatbot usando redes neuronales y procesamiento de lenguaje natural orientado a entidades bancarias. B.S. thesis.
- [Cormen et al., 2022] Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2022). *Introduction to algorithms*. MIT press.
- [Costabello et al., 2023] Costabello, L., Janik, A., Bayram, E., and Pai, S. (2023). Knowledge graph embeddings for nlp: From theory to practice.
- [Costabello et al., 2019] Costabello, L., Pai, S., Van, C. L., McGrath, R., McCarthy, N., and Tabacof, P. (2019). AmpliGraph: a Library for Representation Learning on Knowledge Graphs.
- [Devlin et al., 2018] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- [Dong et al., 2020] Dong, H., Wang, W., Coenen, F., and Huang, K. (2020). Knowledge base enrichment by relation learning from social tagging data. *Information Sciences*, 526:203–220.
- [Ebiele et al., 2022] Ebiele, M., McKenna, L., Bendeche, M., and Brennan, R. (2022). Multi-label classification using bert and knowledge graphs with a limited training dataset.
- [Ehrlinger and Wöß, 2016] Ehrlinger, L. and Wöß, W. (2016). Towards a definition of knowledge graphs. *SEMANTiCS (Posters, Demos, SuCESS)*, 48(1-4):2.
- [ENACHESCU, 2016] ENACHESCU, M.-I. (2016). A prototype for an e-recruitment platform using semantic web technologies. *Informatica Economica*, 20:62–75.
- [Erd et al., 2022] Erd, R., Feddoul, L., Lachenmaier, C., and Mauch, M. J. (2022). Evaluation of data augmentation for named entity recognition in the german legal domain.

- [Europass, 2017] Europass (22 de Mayo de 2017). <https://europa.eu/europass/es/european-qualifications-framework-eqf>.
- [europass Union Europea, 2022] europass Union Europea (2022). europass. <https://europa.eu/europass/es/find-courses>.
- [European Commission, 2022] European Commission (27 de Enero de 2022). <https://ec.europa.eu/esco/lod/static/model.html>.
- [European Commission website, 2022] European Commission website (27 de Enero de 2022). <https://esco.ec.europa.eu/es/about-esco/what-esco>.
- [Fajčíková et al., 2018] Fajčíková, A., Urbancová, H., and Fejfarová, M. (2018). New trends in the recruitment of employees in czech ict organisations. *Scientific Papers of the University of Pardubice, Series D: Faculty of Economics and Administration*, 26:39–49.
- [Faliagka et al., 2012] Faliagka, E., Tsakalidis, A., and Tzimas, G. (2012). An integrated e-recruitment system for automated personality mining and applicant ranking. *Internet research*.
- [Feng et al., 2020] Feng, F., Yang, Y., Cer, D., Arivazhagan, N., and Wang, W. (2020). Language-agnostic bert sentence embedding. *arXiv preprint arXiv:2007.01852*.
- [Fensel et al., 2020] Fensel, D., Şimşek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., Toma, I., Umbrich, J., Wahler, A., Fensel, D., et al. (2020). Introduction: what is a knowledge graph? *Knowledge graphs: Methodology, tools and selected use cases*, pages 1–10.
- [Fensel et al., 2001] Fensel, D., Van Harmelen, F., Horrocks, I., McGuinness, D. L., and Patel-Schneider, P. F. (2001). Oil: An ontology infrastructure for the semantic web. *IEEE intelligent systems*, 16(2):38–45.
- [Fraga, 2016] Fraga, D. C. (2016). Métodos y técnicas para la evaluación del aprendizaje de ontologías de dominio. -.
- [Gao and Chen, 2020] Gao, W. and Chen, Y. (2020). Approximation analysis of ontology learning algorithm in linear combination setting. *Journal of Cloud Computing*, 9(1):1–10.
- [García-Sánchez et al., 2006] García-Sánchez, F., Martínez-Béjar, R., Contreras, L., Fernández-Breis, J. T., and Castellanos-Nieves, D. (2006). An ontology-based intelligent system for recruitment. *Expert Systems with Applications*, 31(2):248–263.
- [Gruber, 1993] Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220.

- [Guarino, 1997] Guarino, N. (1997). Understanding, building and using ontologies. *International journal of human-computer studies*, 46(2-3):293–310.
- [Gusfield, 1997] Gusfield, D. (1997). *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge University Press.
- [Harzallah et al., 2002] Harzallah, M., Leclere, M., and Trichet, F. (2002). Commoncv: modelling the competencies underlying a curriculum vitae. In *Proceedings of the 14th international conference on Software engineering and knowledge engineering*, pages 65–71.
- [Hogan et al., 2021] Hogan, A., Blomqvist, E., Cochez, M., d’Amato, C., Melo, G. D., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., et al. (2021). Knowledge graphs. *ACM Computing Surveys (Csur)*, 54(4):1–37.
- [Howard and Ruder, 2018] Howard, J. and Ruder, S. (2018). Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*.
- [Hsu et al., 2021] Hsu, W.-N., Bolte, B., Tsai, Y.-H. H., Lakhotia, K., Salakhutdinov, R., and Mohamed, A. (2021). Hubert: Self-supervised speech representation learning by masked prediction of hidden units. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3451–3460.
- [Hubert et al., 2022] Hubert, N., Monnin, P., Brun, A., and Monticolo, D. (2022). Knowledge graph embeddings for link prediction: Beware of semantics! In *DL4KG@ ISWC 2022: Workshop on Deep Learning for Knowledge Graphs, held as part of ISWC 2022: the 21st International Semantic Web Conference*.
- [HuntersAmericas, 2023] HuntersAmericas (2023). Las fallas en los perfiles de puesto y el impacto de sus resultados. <http://www.hlghunters.com/blog/96-fallas-perfiles-puestos>.
- [Ilieva, 2013] Ilieva, R. (2013). E-recruitment public services ontological model. *Innovations*, 1(1):22–27.
- [Javed et al., 2019] Javed, Z., Qazi, H., and Khoja, S. A. (2019). An ontology-based knowledge management model for e-recruitment utilizing moocs data. In *2019 8th International Conference on Information and Communication Technologies (ICICT)*, pages 124–128.
- [Jiménez-Ruiz et al., 2012] Jiménez-Ruiz, E., Grau, B. C., Zhou, Y., and Horrocks, I. (2012). Large-scale interactive ontology matching: Algorithms and implementation. In *ECAI*, volume 242, pages 444–449.
- [Jun, 2011] Jun, Z. (2011). Manufacturing process skill systematization based on ontology technology. In *2011 Third International Conference on Measuring Technology and Mechatronics Automation*, volume 2, pages 496–499.

- [Kahlawi, 2020] Kahlawi, A. (2020). An ontology driven esco lod quality enhancement. *International Journal of Advanced Computer Science and Applications*, 11(3).
- [Kasereka, 2020] Kasereka, H. (2020). Importance of web scraping in e-commerce and e-marketing. *SSRN Electronic Journal*.
- [Katyshev and Anikin, 2023] Katyshev, A. and Anikin, A. (2023). Modern approaches for concepts and relations extraction for ontology learning. In *Novel & Intelligent Digital Systems Conferences*, pages 290–296. Springer.
- [Khatiwada et al., 2022] Khatiwada, A., Shirai, S., Srinivas, K., and Hassanzadeh, O. (2022). Knowledge graph embeddings for causal relation prediction. In *Workshop on Deep Learning for Knowledge Graphs (DL4KG@ ISWC2022)*.
- [Khder, 2021] Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, 13(3).
- [Kinsta, Maddy Osman, 2022] Kinsta, Maddy Osman (2022). <https://kinsta.com/es/blog/estadisticas-de-linkedin/>.
- [Kumaran and Sankar, 2013] Kumaran, V. S. and Sankar, A. (2013). Towards an automated system for intelligent screening of candidates for recruitment using ontology mapping (expert). *International Journal of Metadata, Semantics and Ontologies*, 8(1):56–64.
- [Lakzaei and Shamsfard, 2021] Lakzaei, B. and Shamsfard, M. (2021). Ontology learning from relational databases. *Information Sciences*, 577:280–297.
- [LAMY, 2019] LAMY, J.-B. (2019). Consultas sparql owlready2. <://owlready2.readthedocs.io/en/v0.37/sparql.html>.
- [Lawi et al., 2022] Lawi, A., Mesra, H., and Amir, S. (2022). Implementation of long short-term memory and gated recurrent units on grouped time-series data to predict stock prices accurately. *Journal of Big Data*, 9(1):1–19.
- [le Vrang et al., 2014] le Vrang, M., Papantoniou, A., Pauwels, E., Fannes, P., Vandenstein, D., and De Smedt, J. (2014). Esco: Boosting job matching in europe with semantic interoperability. *Computer*, 47(10):57–64.
- [Lievens et al., 2007] Lievens, F., Van Hoye, G., and Anseel, F. (2007). Organizational identity and employer image: Towards a unifying framework. *British Journal of Management*, 18:S45–S59.
- [LinkedIn Corporation, 2022] LinkedIn Corporation (2022). <https://www.linkedin.com/feed/>.

- [Liu et al., 2019] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- [Lopez-Bassols, 2002] Lopez-Bassols, V. (2002). Ict skills and employment. *OECD*.
- [López Bonilla et al., 2012] López Bonilla, M., Guzmán Luna, J. A., and Durley Torres, I. (2012). Metodologías y métodos para la construcción de ontologías. *Scientia Et Technica*.
- [Maedche and Staab, 2001] Maedche, A. and Staab, S. (2001). Ontology learning for the semantic web. *IEEE Intelligent systems*, 16(2):72–79.
- [Manzano-Macho et al., 2008] Manzano-Macho, D., Gómez-Pérez, A., and Borrajo Millán, D. (2008). Unsupervised and domain independent ontology learning: combining heterogeneous sources of evidence. *Sixth International Language Resources and Evaluation (LREC'08)*.
- [Maree et al., 2018] Maree, M., Kmail, A. B., and Belkhatir, M. (2018). Analysis and shortcomings of e-recruitment systems: Towards a semantics-based approach addressing knowledge incompleteness and limited domain coverage. *Journal of Information Science*, 45(6):713–735.
- [McGuinness et al., 2002] McGuinness, D. L., Fikes, R., Hendler, J., and Stein, L. A. (2002). Daml+ oil: an ontology language for the semantic web. *IEEE Intelligent Systems*, 17(5):72–80.
- [McGuinness et al., 2003] McGuinness, D. L., Fikes, R., Stein, L. A., and Hendler, J. A. (2003). Daml-ont: An ontology language for the semantic web. In *Spinning the Semantic Web*, pages 65–93.
- [McGuinness et al., 2004] McGuinness, D. L., Van Harmelen, F., et al. (2004). Owl web ontology language overview. *W3C recommendation*, 10(10):2004.
- [Mhammedi et al., 2021] Mhammedi, S., El Massari, H., and Gherabi, N. (2021). Cb2onto: Owl ontology learning approach from couchbase. In *Intelligent Systems in Big Data, Semantic Web and Machine Learning*, pages 95–110. Springer.
- [Mikolov et al., 2013] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26.
- [Mnih and Kavukcuoglu, 2013] Mnih, A. and Kavukcuoglu, K. (2013). Learning word embeddings efficiently with noise-contrastive estimation. *Advances in neural information processing systems*, 26.

- [Neches et al., 1991] Neches, R., Fikes, R. E., Finin, T., Gruber, T., Patil, R., Senator, T., and Swartout, W. R. (1991). Enabling technology for knowledge sharing. *AI magazine*, 12(3):36–36.
- [NERI, 2005] NERI, F. (2005). Evaluation of ontolearn, a methodology for automatic learning of domain ontologies. *Ontology Learning from Text: Methods, Evaluation and Applications*, 123.
- [Nordsieck et al., 2022] Nordsieck, R., Heider, M., Hummel, A., and Hähner, J. (2022). A closer look at sum-based embeddings for knowledge graphs containing procedural knowledge. In Alam, M., Buscaldi, D., Cochez, M., Osborne, F., and Recupero, D. R., editors, *DL4KG 2022 - Deep Learning for Knowledge Graphs 2022: Proceedings of the Workshop on Deep Learning for Knowledge Graphs (DL4KG 2022) co-located with the 21th International Semantic Web Conference (ISWC 2022), virtual conference, online, October 24, 2022*, page 7.
- [Novalija and Grobelnik, 2018] Novalija, I. and Grobelnik, M. (2018). Towards smart statistics in labour market domain. *Slovenian KDD Conference on Data Mining and Data Warehouses (SiKDD), Ljubljana 2018*.
- [Noy and McGuinness, 2001] Noy, N. and McGuinness, D. (2001). Ontology development 101: A guide to creating your first ontology. *Knowledge Systems Laboratory*, 32.
- [OECD, 2004] OECD (2004). *Information Technology Outlook 2004*. OECD.
- [OECD, 2010] OECD (2010). *OECD Information Technology Outlook 2010*. OECD.
- [OIT, 2022] OIT (2022). Isco international standard classification of occupations. <https://www.ilo.org/public/spanish/bureau/stat/isco/index.htm>.
- [on Information Technology Curricula, 2017] on Information Technology Curricula, T. G. (2017). *Information Technology Curricula 2017: Curriculum Guidelines for Baccalaureate Degree Programs in Information Technology*. Association for Computing Machinery, New York, NY, USA.
- [Organización internacional del trabajo, 2005] Organización internacional del trabajo (30 de Enero de 2005). <https://www.ilo.org/public/spanish/bureau/stat/isco/isco08/index.htm>.
- [Organization, 2022] Organization, I. L. (2022). Isco international standard classification of occupations. <https://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm>.

- [Park et al., 2018] Park, S. H., Kim, B., Kang, C. M., Chung, C. C., and Choi, J. W. (2018). Sequence-to-sequence prediction of vehicle trajectory via lstm encoder-decoder architecture. In *2018 IEEE intelligent vehicles symposium (IV)*, pages 1672–1678. IEEE.
- [Pennington et al., 2014] Pennington, J., Socher, R., and Manning, C. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.
- [Porzel and Malaka, 2004] Porzel, R. and Malaka, R. (2004). A task-based approach for ontology evaluation. In *ECAI Workshop on Ontology Learning and Population, Valencia, Spain*, pages 1–6. Citeseer.
- [Pouliakas, 2021] Pouliakas, K. (2021). *Artificial intelligence and job automation: an EU analysis using online job vacancy data*. Publications Office of the European Union.
- [Prud’hommeaux and Seaborne, 2007] Prud’hommeaux, E. and Seaborne, A. (2007). Sparql query language for rdf. *Live Draft - version*.
- [Puduppully and Lapata, 2021] Puduppully, R. and Lapata, M. (2021). Data-to-text Generation with Macro Planning. *Transactions of the Association for Computational Linguistics*, 9:510–527.
- [Radevski and Trichet, 2006] Radevski, V. and Trichet, F. (2006). Ontology-based systems dedicated to human resources management: An application in e-recruitment. In Meersman, R., Tari, Z., and Herrero, P., editors, *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops*, pages 1068–1077, Berlin, Heidelberg. Springer Berlin Heidelberg.
- [Radziwon et al., 2014] Radziwon, A., Bilberg, A., Bogers, M., and Madsen, E. S. (2014). The smart factory: exploring adaptive and flexible manufacturing solutions. *Procedia engineering*, 69:1184–1190.
- [Ramos et al., 2009] Ramos, E., NÃÃ, H., and CasaÃ, R. (2009). Esquema para evaluar ontologÃas Ãpara un dominio de conocimiento. *Enlace*, 6:57 – 71.
- [Raza et al., 2020] Raza, M. A., Raza, B., Jabeen, T., Raza, S., and Abbas, M. (2020). Using combined list hierarchy and headings of html documents for learning domain-specific ontology. *International Journal of Advanced Computer Science and Applications*, 11(4).
- [Reimers and Gurevych, 2019] Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

- [Rentzsch and Staneva, 2020] Rentzsch, R. and Staneva, M. (2020). Skills-matching and skills intelligence through curated and data-driven ontologies. In *Proceedings of the DELFI Workshops*.
- [Revenko et al., 2022] Revenko, A., Mireles, V., Breit, A., Bourgonje, P., Schneider, J., Khvalchik, M., and Rehm, G. (2022). *Learning Ontology Classes from Text by Clustering Lexical Substitutes Derived from Language Models*.
- [Rohman et al., 2023] Rohman, M. A., Suhartono, S., and Chamidy, T. (2023). Bidirectional gru dengan attention mechanism pada analisis sentimen pln mobile. *Techno. Com*, 22(2):358–372.
- [Ross, 2021] Ross, L. (2021). 25 emerging recruitment trends in 2021. <https://vervoe.com/recruitment-trends-2021/>.
- [Saeeda et al., 2020] Saeeda, L., Med, M., Ledvinka, M., Blaško, M., and Křemen, P. (2020). Entity linking and lexico-semantic patterns for ontology learning. In *European Semantic Web Conference*, pages 138–153. Springer.
- [Salas and Salas, 2020] Salas, A. M. D. L. M. and Salas, L. M. (2020). Brecha de habilidades de los jóvenes en el mercado laboral colombiano. *Vniversitas Económica 018139*, Universidad Javeriana - Bogotá.
- [Salazar et al., 2015] Salazar, O. M., Jaramillo, J. C., Ovalle, D. A., and Guzmán, J. A. (2015). A case-based multi-agent and recommendation environment to improve the e-recruitment process. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 389–397. Springer.
- [Sanh et al., 2019] Sanh, V., Debut, L., Chaumond, J., and Wolf, T. (2019). Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.
- [Seedorf and Schader, 2011] Seedorf, S. and Schader, M. (2011). Towards an enterprise software component ontology. In *A Renaissance of Information Technology for Sustainability and Global Competitiveness.*, volume 5.
- [Shakya and Paudel, 2019] Shakya, A. and Paudel, S. (2019). Job-candidate matching using esco ontology. *Journal of the Institute of Engineering*, 15(1):1–13.
- [Sibarani et al., 2017] Sibarani, E. M., Scerri, S., Morales, C., Auer, S., and Collarana, D. (2017). Ontology-guided job market demand analysis: A cross-sectional study for the data science field. In *Proceedings of the 13th International Conference on Semantic Systems, Semantics2017*, page 25–32, New York, NY, USA. Association for Computing Machinery.
- [Singto and Mingkhwan, 2013] Singto, P. and Mingkhwan, A. (2013). Semantic searching it careers concepts based on ontology. *Journal of Advanced Management Science*, 1:103–106.

- [Soni and Swaminarayan, 2017] Soni, H. H. and Swaminarayan, P. R. (2017). "study of semantic web based e-recruitment system: Review". *International Journal of Advanced Research in Computer Science*, 8(9).
- [Sołek-Borowska and Wilczewska, 2018] Sołek-Borowska, C. and Wilczewska, M. (2018). New technologies in the recruitment process. *Economics and Culture*, 15:25–33.
- [Tebes et al., 2019] Tebes, G., Peppino, D., Becker, P., and Olsina, L. (2019). Especificación del modelo de proceso para una revisión sistemática de literatura . specifying the process model for a systematic literature review.
- [Tong, 2009] Tong, D. Y. K. (2009). A study of e-recruitment technology adoption in malaysia. *Industrial Management & Data Systems*.
- [Turney, 2002] Turney, P. D. (2002). Mining the web for synonyms: Pmi-ir versus lsa on toefl. *Computer information sciences*.
- [ul haq Dar and Dorn, 2018] ul haq Dar, E. and Dorn, J. (2018). Ontology based classification system for online job offers. In *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, pages 1–8. IEEE.
- [Van der Velde, 2020] Van der Velde, L. (2020). Within occupation wage dispersion and the task content of jobs. *Oxford Bulletin of Economics and Statistics*, 82(5):1161–1197.
- [Van Hautte et al., 2020] Van Hautte, J., Schelstraete, V., and Wornoo, M. (2020). Leveraging the inherent hierarchy of vacancy titles for automated job ontology expansion. In *Proceedings of the 6th International Workshop on Computational Terminology*, pages 37–42, Marseille, France. European Language Resources Association.
- [Varen Caballero et al., 2019] Varen Caballero, E., Silega Martínez, N., and Rojas Carbonell, R. (2019). Ontología para gestionar la información científica de los recursos humanos en proyectos de desarrollo de software. *Serie Científica De La Universidad De Las Ciencias Informáticas*, 10(9).
- [Vashishth et al., 2018] Vashishth, S., Jain, P., and Talukdar, P. (2018). Cesi: Canonicalizing open knowledge bases using embeddings and side information. In *Proceedings of the 2018 World Wide Web Conference*, pages 1317–1327.
- [Vaswani et al., 2017] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- [Vrandečić, 2009] Vrandečić, D. (2009). Ontology evaluation. In *Handbook on ontologies*, pages 293–313. Springer.

- [Vrolijk et al., 2022] Vrolijk, J., Mol, S. T., Weber, C., Tavakoli, M., Kismihók, G., and Pelucchi, M. (2022). Ontojob: Automated ontology learning from labor market data. In *2022 IEEE 16th International Conference on Semantic Computing (ICSC)*, pages 195–200.
- [W3C, 2012] W3C (2012). Owl 2 web ontology language profiles (second edition). <https://www.w3.org/TR/owl2-profiles/>.
- [Wong et al., 2017] Wong, I. A., Wan, Y. K. P., and Gao, J. H. (2017). How to attract and retain generation y employees? an exploration of career choice and the meaning of work. *Tourism Management Perspectives*, 23:140–150.
- [World Wide Web Consortium, “Semantic Web - W3C,” W3.org, 2012] World Wide Web Consortium, “Semantic Web - W3C,” W3.org (2012). <https://www.w3.org/standards/semanticweb/>.
- [Xu et al., 2020] Xu, Z., Harzallah, M., Guillet, F., and Ichise, R. (2020). Towards a term clustering framework for modular ontology learning. In Fred, A., Salgado, A., Aveiro, D., Dietz, J., Bernardino, J., and Filipe, J., editors, *Knowledge Discovery, Knowledge Engineering and Knowledge Management*, pages 178–201, Cham. Springer International Publishing.
- [Zhou et al., 2021] Zhou, J., Naseem, T., Fernandez Astudillo, R., Lee, Y.-S., Florian, R., and Roukos, S. (2021). Structure-aware fine-tuning of sequence-to-sequence transformers for transition-based AMR parsing. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6279–6290, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- [Zhou, 2007] Zhou, L. (2007). Ontology learning: State of the art and open issues. *Information Technology and Management*, 8:241–252.