



UNIVERSIDAD DE DEUSTO

**Sistema de recomendación de cursos de  
formación continua para mejorar las habilidades  
profesionales basado en ontología y Machine  
Learning**

Tesis doctoral presentada por María Cora Urdaneta Ponte

dentro del Programa de

Doctorado de Ingeniería para la Sociedad de la Información y Desarrollo Sostenible

Dirigida por Dra. Amaia Méndez Zorrilla y

Dr. Ibon Oleagordia Ruiz





UNIVERSIDAD DE DEUSTO

# Sistema de recomendación de cursos de formación continua para mejorar las habilidades profesionales basado en ontología y Machine Learning

Tesis doctoral presentada por María Cora Urdaneta Ponte

dentro del Programa de

Doctorado de Ingeniería para la Sociedad de la Información y Desarrollo Sostenible

Dirigida por Dra. Amaia Méndez Zorrilla y

Dr. Ibon Oleagordia Ruiz

El doctorando

**MENDEZ  
ZORRILLA,  
AMAIA  
(FIRMA)** Firmado digitalmente por  
MENDEZ  
ZORRILLA, AMAIA  
(FIRMA)  
Fecha: 2022.03.08  
11:38:46 +01'00'

Los directores

**IBON  
OLEAGORDIA  
RUIZ** Firmado digitalmente por  
IBON OLEAGORDIA RUIZ  
Nombre de reconocimiento  
(DN): c=ES, ou=Herritar  
ziurtagiria - Certificado de  
ciudadano, cn=IBON  
OLEAGORDIA RUIZ,  
givenName=IBON,  
sn=OLEAGORDIA RUIZ,  
serialNumber=14262451J  
Fecha: 2022.03.08 14:30:09  
+01'00'

Bilbao, marzo de 2022



UNIVERSIDAD DE DEUSTO

# **Sistema de recomendación de cursos de formación continua para mejorar las habilidades profesionales basado en ontología y Machine Learning**

Tesis doctoral presentada por María Cora Urdaneta Ponte  
dentro del Programa de  
Doctorado de Ingeniería para la Sociedad de la Información y Desarrollo Sostenible

Dirigida por Dra. Amaia Méndez Zorrilla y  
Dr. Ibon Oleagordia Ruiz

El doctorando

Los directores

Bilbao, marzo de 2022

Sistema de recomendación de cursos de formación continua para mejorar las habilidades profesionales basado en ontología y Machine Learning

Autor: María Cora Urdaneta Ponte

Directores: Dra. Amaia Méndez Zorrilla y Dr. Ibon Oleagordia Ruiz

Texto impreso en Bilbao

Primera edición, marzo de 2022

*A Juan y Paola.*



# Agradecimiento

Son muchas las personas a las que debo agradecer por hacer esto posible, no quisiera olvidarme de alguna en este instante en el que estoy escribiendo los agradecimientos, si así fuera, estar seguras de que durante el desarrollo de mi tesis las tuve siempre presente.

Comienzo agradeciendo a mis directores de tesis Amaia Méndez e Ibón Oleagordia, que desde el momento en el que llegué a la universidad estuvieron siempre dispuestos a ayudarme, que confiaron en mí, se preocuparon por todas las situaciones que viví, no solo académicas, si no personales y siempre me animaron a seguir. Al grupo de investigación eVida, de quienes son parte y me hicieron parte de él, ¡Muchas gracias!

A la Cátedra Unesco de la universidad de Deusto, de la que formo parte, gracias al Vicerrectorado de Relaciones Internacionales y Transformación Digital, a nuestros directores Iziar Basterretxea y Sergio Caballero, por su acompañamiento y estar siempre preocupados por nosotros y buscar una solución a cada situación por la que pasamos. A mis compañeros de la cátedra, que nos convertimos en una bonita familia, muchas gracias. No podría darle las gracias a la Cátedra y dejar por fuera a la UCAB, mi universidad en Venezuela, quienes me dieron es espacio y tiempo para poder cumplir con esta meta, son una parte importante de este logro.

Dentro de la Universidad de Deusto, son muchas las personas a quien quiero dar las gracias, a Juanjo Etxeberria, por acompañarme en este proceso, a la Facultad de ingeniería donde estuve colaborando, gracias por hacerme parte y por las oportunidades que me brindaron, gracias a Cristina Giménez, por lo que ha supuesto trabajar juntas en el Premio Ada Byron, una experiencia muy enriquecedora. A la Unidad de Ordenación Académica,

a Asun Barredo, a Manoli, Itzi, María, por hacerme parte de ustedes, ¡por darme la oportunidad!

Tengo que dar las gracias infinitas a mi familia, que me apoyaron en todo momento y de todas las formas posibles, en especial a mi hermana Yluz, que se convirtió en una madre para mi hija mientras duró este proceso. A Jesús que me acompañó y vivió muy de cerca todo este proceso. A mis hijos, Paola que se llevó la peor parte de todo esto, gracias por todo tu sacrificio, sé que fue muy difícil, pero todo tiene su recompensa, y a Juanjo que siempre estuvo pendiente de nosotras y siempre me animaba. Los amo.

Hace poco leí lo siguiente:

“Las personas mágicas existen. Aparecen de la nada. Son aquellas que saben ganarse la confianza rápido y de forma inesperada. De pronto llegan a tu vida, empiezan a hablar de todo, de alegrías, de daños, experiencias, penas y heridas. Cuando te das cuenta, no recuerdas como era tu vida antes que las conocieras. Llegan a tu universo para darte un aire liviano, para brindarte tu amistad, su mano, su energía juntos. Llegan para abrazarte y muchas veces quedarse”

Termino agradeciendo a las personas especiales que Dios me puso en el camino, que me hicieron y hacen las cosas más fáciles, a mis personas mágicas, Marga, Carla, Úrsula y Aitor, muchas gracias por estar.

# Resumen

Los sistemas de recomendación son herramientas que han captado el interés de diversas áreas debido a la posibilidad de sugerir productos o servicios tomando en cuenta los perfiles y el comportamiento de las personas en la red. Estos sistemas aplican algoritmos predictivos sobre un conjunto de datos para sugerir elementos relevantes a las necesidades de los usuarios. En el campo de la formación continua los sistemas de recomendación están adquiriendo mucho peso permitiendo relacionar la oferta formativa y las personas, para ayudarles a mejorar sus habilidades profesionales, frente a la necesidad de contar con personal altamente competitivos en un entorno laboral cada vez más demandante.

Hoy en día en las distintas redes sociales es posible conseguir información actualizada de personas relacionada con su perfil profesional y sus áreas de desempeño, así mismo, es común el uso de este tipo de redes, como LinkedIn, para la oferta de empleos. Por otra parte, en la web hay una gran variedad de oferta de cursos de formación continua donde suele ser muy limitada la información de los participantes referente a la calificación y preferencia de estos. Siendo unos de los mayores retos encontrar relaciones y patrones en esta vasta cantidad de información, que se actualiza constantemente, donde las fuentes de datos son heterogéneas y su naturaleza es que se presentan sin ningún tipo de normalización y de forma no estructurada o en el mejor de los casos de forma semiestructurada. Dado que la calidad de las recomendaciones depende en gran medida de la recuperación y representación de la información, se ha adoptado la representación semántica del conocimiento como una de las teorías para ayudar a resolver este tipo de problemas. Adicionalmente, se han combinado técnicas de web semántica con técnicas

de Machine Learning, lo que permite, a partir de la información disponible, explotar el conocimiento, actualizarlo e inferir nuevas relaciones entre los datos, surgiendo sistemas de recomendación basados en estas técnicas como un enfoque para mejorar la calidad de las recomendaciones.

En este contexto, se desarrolló un sistema de recomendación híbrido, cuyo núcleo aplica un filtrado semántico que utiliza una ontología para modelar sectores de desempeño laboral y áreas de conocimiento, que permite representar habilidades profesionales. La ontología se actualiza vía eventos a partir de datos perfilados obtenidos de registros profesionales de las redes sociales, que, haciendo uso de técnicas de Machine Learning, clusteriza entidades con la finalidad de hacer predicciones para nuevos datos. Etapas posteriores del sistema de recomendación hacen uso de filtrado por contenido y heurísticas con la finalidad de proponer cursos de formación continua para el desarrollo y/o actualización de habilidades profesionales.

En el proceso de desarrollo del sistema, se han explorado diferentes técnicas de filtrado, y desarrollado tres versiones de este, donde la versión final, basada en ontología y Machine Learning, permitió aprovechar la riqueza de la información obtenida de la web y mejorar el rendimiento del sistema de recomendación en cuanto a relevancia, y precisión de las sugerencias.

Con el objetivo de probar el sistema, y validar la hipótesis planteada, se obtuvo un conjunto de datos de perfiles profesionales de LinkedIn y de cursos de formación continua de la web, con los cuales se ejecutaron las diferentes versiones y configuraciones del sistema de recomendación, para poder comparar el rendimiento de estas y compararlas con los resultados de otros sistemas de recomendación encontrados en la revisión del estado del arte.

# Abstract

Recommendation systems are tools that have attracted the interest of several areas due to the possibility of suggesting products or services considering the profiles and behavior of people on the network. These systems apply predictive algorithms on a set of data to suggest items relevant to users' needs. In the field of lifelong learning, recommendation systems are becoming very important to link the training offer and people, to help them improve their professional skills, given the need for highly competitive personnel in an increasingly demanding work environment.

Nowadays, it is possible to get updated information about people related to their professional profile and areas of performance in the different social networks, and it is also common to use this type of networks, such as LinkedIn, for job offers. On the other hand, on the web there is a wide variety of continuing education courses where the information of the participants regarding their qualifications and preferences is usually very limited. One of the most challenging tasks is to find relationships and patterns in this large amount of information, which is constantly updated, where the data sources are heterogeneous and its nature is that it is presented without any standardization and in an unstructured or, at best, semi-structured form. Due to the fact that the quality of recommendations depends to a large extent on information retrieval and representation, semantic knowledge representation has been adopted as one of the theories to help solve this type of problem. Additionally, semantic web techniques have been combined with Machine Learning techniques, which allow, from the available information, to exploit the knowledge, update it and infer new relationships between the data, emerging

recommendation systems based on these techniques as an approach to improve the quality of recommendations.

In this context, a hybrid recommendation system was developed, whose core applies a semantic filtering that uses an ontology to model job performance sectors and areas of knowledge, which allows representing professional skills. The ontology is updated via events from profiled data obtained from professional records of social networks, which, using Machine Learning techniques, cluster entities in order to make predictions for new data. Subsequent stages of the recommender system use filtering by content and heuristics to propose continuing education courses for the development and/or updating of professional skills.

In the process of developing the system, different filtering techniques have been explored, and three versions of the system have been developed, where the final version, based on ontology and ML, allowed taking advantage of the richness of the information obtained from the web and improving the performance of the Recommender System in terms of relevance and accuracy of the suggestions.

In order to test the system, and validate the hypothesis put forward, a data set of LinkedIn professional profiles and web-based lifelong learning courses was obtained, with which the different versions and configurations of the recommender system were run, in order to compare the performance of these and compare them with the results of other recommender systems found in the review of the state of the art.

# Tabla de Contenido

Introducción .....	1
1.1    Introducción .....	1
1.2    Hipótesis y Objetivos .....	6
1.3    Impacto Social.....	8
1.4    Metodología de la Investigación .....	9
1.5    Organización de la tesis .....	11
Estado del Arte.....	13
2.1    Sistemas de Recomendación .....	13
2.1.1    Clasificación de los SR según los algoritmos de filtrado.....	14
2.1.2    Funciones de Similitud en los SR .....	18
2.1.3    Evaluación de los sistemas de recomendación.....	20
2.2    Machine Learning .....	24
2.3    Ontología.....	26
2.4    Sistemas de Recomendación en Educación .....	29
2.5    Resumen del Capítulo .....	49
Fuentes de Información y datos .....	51
3.1    Tamaño de la muestra y método de adquisición de datos .....	52
3.2    Características de los datos .....	58
3.3    Representación de Perfiles de usuarios .....	58
3.4    Representación de Perfiles de Cursos .....	60
3.5    Construcción del Dataset.....	62
3.6    Resumen del Capítulo .....	63
Diseño del Sistema de recomendación .....	65
4.1    Diseño de alto nivel.....	65

4.1.1	Sistema de recomendación híbrido basado en taxonomía.....	68
4.1.2	Sistema de recomendación híbrido basado en ontología y ML.....	82
4.1.3	Sistema de recomendación híbrido basado en ontología y ML con <i>endorsements</i> 89	
4.2	Diseño de bajo nivel.....	92
4.2.1	Sistema de Recomendación híbrido basado en taxonomía .....	92
4.2.2	Sistema de recomendación híbrido basado en ontología y ML.....	105
4.2.3	Sistema de recomendación híbrido basado en ontología y ML con <i>endorsements</i> 121	
4.3	Resumen del Capítulo .....	124
Resultados.....		125
5.1	Resultados del sistema de recomendación híbrido basado en taxonomía. ....	128
5.2	Resultados del sistema de recomendación híbrido basado en ontología y ML.....	135
5.3	Resultados del sistema de recomendación híbrido basado en ontología y ML con <i>endorsements</i> .....	139
5.4	Resumen del capítulo .....	148
Conclusiones y trabajo futuro .....		150
6.1	Conclusiones .....	150
6.2	Trabajo Futuro.....	160
6.3	Difusión Científica.....	161

# Índice de Figuras

Figura 1.1 Metodología de Investigación. Elaboración propia.....	10
Figura 3.1. Estructura de Perfil de Usuario. Elaboración propia .....	59
Figura 3.1. Estructura de Perfil de Usuario. Elaboración propia .....	59
Figura 3.2 Codificaciones utilizadas en los Perfiles de Usuarios. Elaboración propia ..	60
Figura 3.2 Codificaciones utilizadas en los Perfiles de Usuarios. Elaboración propia ..	60
Figura 3.3 Estructura de Perfil de cursos. Elaboración propia .....	61
Figura 3.3 Estructura de Perfil de cursos. Elaboración propia .....	61
Figura 3.4 Codificaciones utilizadas en los Perfiles de Cursos. Elaboración propia.....	62
Figura 3.4 Codificaciones utilizadas en los Perfiles de Cursos. Elaboración propia.....	62
Figura 3.5 Proceso de construcción de perfiles. Elaboración propia.....	63
Figura 3.5 Proceso de construcción de perfiles. Elaboración propia.....	63
Figura 4.1 Propuesta general del Sistema de Recomendación .....	66
Figura 4.1 Propuesta general del Sistema de Recomendación .....	66
Figura 4.2 Procesos del Sistema de Recomendación.....	68
Figura 4.2 Procesos del Sistema de Recomendación.....	68
Figura 4.3 Procesos del Sistema de Recomendación híbrido basado en taxonomía .....	69
Figura 4.3 Procesos del Sistema de Recomendación híbrido basado en taxonomía .....	69
Figura 4.4 Proceso de creación de perfiles del SR híbrido basado en taxonomía .....	70
Figura 4.4 Proceso de creación de perfiles del SR híbrido basado en taxonomía .....	70
Figura 4.5 Taxonomía áreas de conocimiento – Lenguaje de programación .....	71
Figura 4.5 Taxonomía áreas de conocimiento – Lenguaje de programación .....	71
Figura 4.6 Taxonomía de sectores de desempeño laboral en el área de Desarrollo de Software.....	72

Figura 4.6 Taxonomía de sectores de desempeño laboral en el área de Desarrollo de Software .....	72
Figura 4.7 Códigos para las taxonomías áreas de conocimiento y sectores de desempeño laboral .....	73
Figura 4.7 Códigos para las taxonomías áreas de conocimiento y sectores de desempeño laboral .....	73
Figura 4.8 Ejemplos de codificación de habilidades .....	73
Figura 4.8 Ejemplos de codificación de habilidades .....	73
Figura 4.9 Representación con los datos de perfiles de usuarios almacenados.....	74
Figura 4.9 Representación con los datos de perfiles de usuarios almacenados.....	74
Figura 4.10 Proceso de recomendación del SR híbrido basado en taxonomía.....	74
Figura 4.10 Proceso de recomendación del SR híbrido basado en taxonomía.....	74
Figura 4.11 Valoración de una habilidad. Elaboración propia .....	76
Figura 4.11 Valoración de una habilidad. Elaboración propia .....	76
Figura 4.12 Determinación de las habilidades a mejorar y/o desarrollar para un usuario en el SR híbrido basado en taxonomía.....	81
Figura 4.12 Determinación de las habilidades a mejorar y/o desarrollar para un usuario en el SR híbrido basado en taxonomía.....	81
Figura 4.13 Determinación de las habilidades a mejorar y/o desarrollar para un usuario sin cargo propio en el SR híbrido basado en taxonomía.....	82
Figura 4.13 Determinación de las habilidades a mejorar y/o desarrollar para un usuario sin cargo propio en el SR híbrido basado en taxonomía.....	82
Figura 4.14 Procesos del SR híbrido basado en ontología y ML .....	84
Figura 4.14 Procesos del SR híbrido basado en ontología y ML .....	84
Figura 4.15 Procesos de actualización de la ontología del SR híbrido basado en ontología y ML .....	85
Figura 4.15 Procesos de actualización de la ontología del SR híbrido basado en ontología y ML .....	85
Figura 4.16 Representación de los datos basada en ontología.....	86
Figura 4.16 Representación de los datos basada en ontología.....	86
Figura 4.17 Nuevas relaciones soportadas entre entidades de la Ontología.....	86
Figura 4.17 Nuevas relaciones soportadas entre entidades de la Ontología.....	86
Figura 4.18 Determinación de las habilidades en el SR híbrido basado en ontología y ML .....	88

Figura 4.18 Determinación de las habilidades en el SR híbrido basado en ontología y ML .....	88
Figura 4.19 Determinación de las habilidades a partir de ML en la ontología en el SR híbrido basado en ontología y ML .....	89
Figura 4.19 Determinación de las habilidades a partir de ML en la ontología en el SR híbrido basado en ontología y ML .....	89
Figura 4.20 Procesos del SR híbrido basado en ontología y ML con <i>endorsements</i> .....	90
Figura 4.20 Procesos del SR híbrido basado en ontología y ML con <i>endorsements</i> .....	90
Figura 4.21 Determinación de las habilidades a mejorar o desarrollar para un usuario particular en el SR híbrido basado en ontología y ML con <i>endorsements</i> .....	92
Figura 4.21 Determinación de las habilidades a mejorar o desarrollar para un usuario particular en el SR híbrido basado en ontología y ML con <i>endorsements</i> .....	92
Figura 4.22 Arquitectura SR híbrido basado en taxonomía.....	93
Figura 4.22 Arquitectura SR híbrido basado en taxonomía.....	93
Figura 4.23 Algoritmo de recomendación del SR híbrido basado en taxonomía.....	97
Figura 4.23 Algoritmo de recomendación del SR híbrido basado en taxonomía.....	97
Figura 4.24 Algoritmo de recomendación del SR híbrido basado en taxonomía.....	98
Figura 4.24 Algoritmo de recomendación del SR híbrido basado en taxonomía.....	98
Figura 4.25 Desarrollo de Prerrequisitos .....	102
Figura 4.25 Desarrollo de Prerrequisitos .....	102
Figura 4.26 Arquitectura de SR híbrido basado en ontología y ML.....	106
Figura 4.26 Arquitectura de SR híbrido basado en ontología y ML.....	106
Figura 4.27 Clases y relaciones de la ontología.....	111
Figura 4.27 Clases y relaciones de la ontología.....	111
Figura 4.28 Etapa On-line del SR híbrido basado en ontología y ML .....	114
Figura 4.28 Etapa On-line del SR híbrido basado en ontología y ML .....	114
Figura 4.29 Algoritmo de recomendación del SR híbrido basado en ontología y ML. ....	115
Figura 4.29 Algoritmo de recomendación del SR híbrido basado en ontología y ML. ....	115
Figura 4.30 Arquitectura del SR híbrido basado en ontología y ML con <i>endorsements</i> .....	122
Figura 4.30 Arquitectura del SR híbrido basado en ontología y ML con <i>endorsements</i> .....	122
Figura 5.1 MAE para los diferentes valores de $\mu$ y las diferentes configuraciones de similitud y codificación. ....	131

Figura 5.1 MAE para los diferentes valores de $\mu$ y las diferentes configuraciones de similitud y codificación. ....	131
Figura 5.2 $F_1$ para los diferentes valores de $\mu$ y las diferentes configuraciones de similitud y codificación. ....	131
Figura 5.2 $F_1$ para los diferentes valores de $\mu$ y las diferentes configuraciones de similitud y codificación. ....	131
Figura 5.3. Ejemplo de salida del SR híbrido basado en taxonomía con configuración plana.....	132
Figura 5.3. Ejemplo de salida del SR híbrido basado en taxonomía con configuración plana.....	132
Figura 5.4 Ejemplo de salida SR híbrido basado en taxonomía con configuración jerárquica .....	133
Figura 5.4 Ejemplo de salida SR híbrido basado en taxonomía con configuración jerárquica .....	133
Figura 5.5 Matriz de similitud del coseno para una codificación plana .....	134
Figura 5.5 Matriz de similitud del coseno para una codificación plana .....	134
Figura 5.6 Ejemplo de salida SR híbrido basado en ontología y ML.....	137
Figura 5.6 Ejemplo de salida SR híbrido basado en ontología y ML.....	137
Figura 5.7 Ejemplo de salida SR híbrido basado en ontología y ML con <i>endorsements</i> .....	140
Figura 5.7 Ejemplo de salida SR híbrido basado en ontología y ML con <i>endorsements</i> .....	140
Figura 5.8 Media armónica ( $F_1$ ) para las ejecuciones con mejor desempeño del SR ..	146
Figura 5.8 Media armónica ( $F_1$ ) para las ejecuciones con mejor desempeño del SR ..	146
Figura 6.1 Evolución del Sistema de Recomendación .....	151
Figura 6.1 Evolución del Sistema de Recomendación .....	151



# Índice de tablas

Tabla 2.1. Tipo algoritmos de filtrado, características y limitaciones .....	18
Tabla 2.2 Artículos analizados en la revisión de la literatura .....	30
Tabla 2.3 Porcentaje de artículos según el tipo de educación .....	36
Tabla 2.4 Porcentaje de artículos según el tipo de usuario .....	36
Tabla 2.5 Porcentaje de artículos según el tipo de elementos que recomienda .....	37
Tabla 2.6 Porcentaje de artículos según el enfoque de desarrollo .....	39
Tabla 2.7 Uso de taxonomía para la representación de perfiles y recursos .....	43
Tabla 2.8 Uso de Ontología en Sistemas de Recomendación.....	46
Tabla 3.1 Formato del archivo csv LinkedIn con la información personal del usuario	53
Tabla 3.2 Formato del archivo csv de LinkedIn con la información profesional del usuario.....	55
Tabla 3.3 Formato del archivo csv de LinkedIn con la información académica del usuario.....	56
Tabla 3.4 Formato del archivo de cursos .....	57
Tabla 4.1 Clases de la ontología .....	109
Tabla 4.2. Relaciones de la ontología .....	110
Tabla 5.1. Métricas de evaluación para sistemas de recomendación.....	127
Tabla 5.2. Métricas para el SR híbrido basado en taxonomía con codificación plana utilizando similitud de coseno, para diferentes valores de $\mu$ .....	128
Tabla 5.3. Métricas para el SR híbrido basado en taxonomía con codificación plana utilizando similitud de Pearson, para diferentes valores de $\mu$ .....	129

Tabla 5.4. Métricas para el SR híbrido basado en taxonomía con codificación jerárquica utilizando similitud de coseno, para diferentes valores de $\mu$ .....	130
Tabla 5.5 Métricas para el SR híbrido basado en taxonomía con codificación jerárquica utilizando similitud de Pearson, para diferentes valores de $\mu$ .....	130
Tabla 5.6. Métricas para el SR híbrido basado en taxonomía configurado con codificación plana .....	132
Tabla 5.7 Métricas para el SR híbrido basado en taxonomía configurado con codificación jerárquica.....	133
Tabla 5.8 Métricas para las diferentes configuraciones del SR híbrido basado en ontología y ML para el conjunto de datos de entrenamiento.....	137
Tabla 5.9 Métricas para las diferentes configuraciones del SR híbrido basado en ontología y ML para el total de la muestra del conjunto datos.....	138
Tabla 5.10 Métricas para las diferentes configuraciones del SR híbrido basado en ontología y ML para el conjunto datos de pruebas. ....	138
Tabla 5.11 Métricas para las diferentes configuraciones del SR híbrido basado en ontología y ML con <i>endorsements</i> para el conjunto de datos de entrenamiento.....	140
Tabla 5.12 Métricas para las diferentes configuraciones del SR híbrido basado en ontología y ML con <i>endorsements</i> para el total de las muestras del conjunto datos....	141
Tabla 5.13 Métricas para las diferentes configuraciones del SR híbrido basado en ontología y ML con <i>endorsements</i> para el conjunto datos de pruebas. ....	141
Tabla 5.14 Métricas de las diferentes configuraciones, para el conjunto de datos de entrenamiento.....	143
Tabla 5.15 Métricas de las diferentes configuraciones, para el conjunto total de datos .....	144
Tabla 5.16 Métricas de las diferentes configuraciones, para el conjunto datos de pruebas. ....	144
Tabla 5.17 Media armónica ( $F_1$ ) para las ejecuciones con mejor desempeño del SR ..	144
Tabla 5.18 Medidas de evaluación en trabajos similares.....	148



# Capítulo 1

## Introducción

### 1.1 Introducción

Hoy en día las organizaciones están atravesando un cambio significativo, evidenciado en una continua y rápida transformación como respuesta a la globalización y a los avances en las tecnologías de información y comunicación, reflejándose en una constante modificación de los requisitos laborales. Ante esto, los profesionales deben mejorar y/o desarrollar sus competencias y habilidades conforme a las tendencias del mercado laboral para lograr ser competitivos y dar respuesta a la demanda de profesionales altamente especializados e incrementar la competitividad de las empresas.

De acuerdo con (Raitskaya & Tikhonova, 2019) los requisitos de empleabilidad han cambiado para abordar las nuevas realidades y desafíos, apareciendo en algunos ámbitos algunas discrepancias entre las competencias de los empleados y las necesidades del mercado laboral. Así mismo, es posible observar que los conocimientos adquiridos dentro de una asignatura en una institución de educación superior pueden no responder al mercado para cuando los graduados obtienen su titulación.

Los empleadores hoy en día se enfocan en conseguir buenos trabajadores que no sólo tengan habilidades académicas básicas sino también competencias genéricas y/o transversales como, aprender, razonar, pensar creativamente, toma de decisiones y resolución de problemas. Las organizaciones dan importancia a las habilidades de empleabilidad, que son las que se conocen como aquellas necesarias para conseguir, mantener y hacer bien un trabajo (Banu Kenayathulla, Aziah Ahmad, & Rahman Idris, 2019). Estas se pueden definir como una gama de habilidades o competencias que se pueden desarrollar durante la vida a través de la educación, capacitación, experiencia laboral, intereses y actividades extracurriculares.

Según (Kálmán, Molnár, & Szúts, 2018), estos cambios se han visto reflejados en elementos del sistema educativo. El aumento de la competencia económica mundial ha llevado a hacer hincapié en el aprendizaje permanente con un enfoque en la creación de habilidades basadas en el mercado (Stromquist & da Costa, 2017). Hemos pasado de profesionales con competencias adaptadas a un modo industrial de producción, a profesionales que deben desarrollar habilidades más relacionadas con los desarrollos económicos y sociales actuales (van Laar, van Deursen, van Dijk, & de Haan, 2017).

De acuerdo con (Laal, 2011), las personas necesitan actualizar sus conocimientos a lo largo de su vida adulta para hacer frente a la vida moderna, tanto en su trabajo como en su vida privada. Así mismo, define *Lifelong Learning* (en adelante, LL) como el proceso de ampliación de conocimientos y procedimientos, habilidades profesionales y actitudes de una manera deliberada y autodirigida. El objetivo del LL es mejorar eficazmente el rendimiento, y poder adaptarse con éxito a los desafíos emergentes a corto y largo plazo y a los nuevos requisitos a lo largo de la vida útil de la carrera profesional. Para el Instituto de la UNESCO para el Aprendizaje a lo largo de la vida (UIL, 2021), LL promueve la capacidad de las personas para afrontar los cambios que pueden surgir debido a las conocidas amenazas y oportunidades del cambio demográfico, la crisis climática, el rápido avance de la tecnología y, más recientemente, la pandemia COVID-19.

La necesidad de tener profesionales actualizados ha derivado en una gran oferta de cursos formativos de LL, la cual se caracteriza por ser muy variada para poder cubrir las competencias en las diferentes áreas de conocimiento y por su continua actualización para adaptarse a los cambios constantes. De acuerdo con (UIL, 2021), LL existe en cualquier lugar, para cualquier persona, con distintas modalidades y para todos los dominios de conocimiento.

Las competencias están formadas por un conjunto de capacidades y habilidades que se pueden describir como un know-how resultante de la integración, y adaptación de capacidades y habilidades, a situaciones características (Keinänen, Ursin, & Nissinen, 2018), necesarias para desempeñarse eficazmente en el mercado laboral. Se espera que los trabajadores seleccionen eficientemente los conocimientos de la cantidad de información disponible y los apliquen eficazmente (van Laar, van Deursen, van Dijk, & de Haan, 2017).

En la medida en que crece la oferta de cursos de LL en la web y los usuarios necesitan pasar horas para encontrar información que cumpla con sus intereses, surge la necesidad de proporcionar recomendaciones personalizadas y relevantes, al igual que ocurre en otros ámbitos, como por ejemplo el comercio electrónico, las redes sociales, los medios digitales y los libros (Nadine, Cao, & Deng, 2016). Los usuarios necesitan priorizar los datos filtrados para obtener de manera eficiente la información adecuada de la web.

En la era del Big Data el reto es encontrar relaciones y patrones en inmensas cantidades de información, en su mayoría no estructurados y semiestructurados. En este contexto destacan los Sistemas de Recomendación (en adelante, SR), donde el principal objetivo es predecir las preferencias de los elementos que sean valioso para el usuario. Según (Adam, Zulkafli, Soh, & Kamal, 2017), los SR recogen la información de conexión entre los usuarios y los artículos, y la utilizan para hacer sugerencias relevantes y significativas basándose en la predicción del interés del usuario, de acuerdo con las preferencias explícitas o implícitas de estos. Para ello, utilizan diferentes algoritmos de filtrado, donde los enfoques predominantes son el filtrado colaborativo, el filtrado basado en contenido y el filtrado híbrido, que combina los dos enfoques para superar cualquier

problema que pueda surgir mediante el uso individual de las diferentes técnicas, como el problema de arranque en frío o escasez de datos, y mejorar así el rendimiento de las recomendaciones (Suhaim & Berri, 2021).

En el ámbito de la educación se hace cada vez más difícil para los motores de búsqueda tradicionales cumplir con los requisitos relacionados con las búsquedas en línea de información sobre productos y servicios educativos por parte de los estudiantes durante el proceso de aprendizaje (Cui, Guo, & Liang, 2018). En consecuencia, se han propuesto un número significativo de SR orientados a los servicios de enseñanza y asesoramiento académico, siendo los usuarios objetivos: estudiantes, profesores y asesores académicos, y los elementos recomendados: materiales educativos, objetos de aprendizaje, documentos, universidades e información sobre cursos y área de estudio (Shao, 2017).

Por otra parte, las Redes Sociales (en adelante, RRSS) se han convertido en una importante fuente de información tanto para la creación, como la difusión de contenido. El reporte de la Unesco (UIL, 2021) reconoce el papel de las RRSS en la era de la información, e indica que “son los principales proveedores de oportunidades de aprendizaje informal, como tutoriales en video sobre diversos temas”, que representan un desafío para LL en cuanto a la desinformación e información no correcta que se puede encontrar en estas redes.

Al mismo tiempo, cada vez es más común el uso de las RRSS para establecer vínculos profesionales e intercambiar información (Xu, y otros, 2018). El uso de las RRSS se ha incorporado como fuente de información en los procesos de selección de muchas organizaciones, el estudio realizado por (Melão & Reis, 2020) arrojó que las más utilizadas para este fin son Facebook (58%) y LinkedIn (89%). Esto es debido al tipo de información en el dominio laboral que comparten, como, puesto de trabajo, formación y habilidades; características a tomar en cuenta a la hora de elegir un candidato, por tanto, proponer una ruta formativa para mejorar y/o desarrollar habilidades profesionales.

En este contexto, (Patel, Kakuste, & Eirinaki, 2017) indica, que los SR ofrecen una oportunidad a profesionales, al personalizar recomendaciones de ofertas de trabajo e identificar áreas y habilidades a desarrollar. Por otra parte, las RRSS constituyen un

medio potencial para la búsqueda de este tipo de información. De acuerdo (Guo, Yin, Zhang, Li, & He, 2019), la información recuperada de RRSS puede mejorar problemas de escasos de datos, relacionados a la falta de historial de valoraciones.

Los SR manejan una gran cantidad de datos, que, dependiendo de su naturaleza, pueden provenir de diferentes fuentes teniendo como característica la heterogeneidad. La calidad de las recomendaciones depende en gran medida de la recuperación y representación de la información. En este proceso se pueden presentar problemas relacionados con la diversidad de tipos y fuentes de datos utilizadas, o la falta de información ofrecida por los usuarios o elementos recomendados, conocida como arranque en frío (Aggarwal, 2016), convirtiéndose en un desafío para este tipo de sistemas. Ante esto, se ha adoptado la representación semántica del conocimiento como una de las teorías para ayudar a resolver este tipo de problemas (Akmaliah, Krisnadhi, Sensuse, Rahayu, & Wulandari, 2018).

La web semántica utiliza ontologías para representar datos legibles por las máquinas (Shao, 2017). Una ontología es una especificación formal de los tipos, propiedades y las interrelaciones de las entidades para un dominio particular. Dado que las ontologías se construyen generalmente sobre la base del consenso de expertos en dominios, son bases de conocimiento estructuradas altamente confiables, con una amplia gama de aplicaciones (Xu, y otros, 2018). Se han desarrollado SR basados en ontología demostrando que mejoran el rendimiento del sistema en términos de estructura de datos, ampliando el contexto de las recomendaciones basadas en la información disponible en diferentes fuentes de datos (Quang, Luan, Pham, & Chu, 2019). Esto permite extraer la información, y representar el conocimiento sobre los usuarios, artículos y la relación que existe entre ellos en el dominio de aplicación, sin que haya pérdida de información relevante, producto de la representación, en relación con una codificación plana, así como abordar el problema de arranque en frío.

El uso de técnicas de la web semántica en la educación ha traído algunas mejoras tales como, contenido personalizado, recomendación de recursos educativos para los estudiantes, y la recopilación de datos con respecto a la interacción de los estudiantes con

entornos educativos (Pereira, Campos, Ströele, David, & Braga, 2018), permitiendo relacionar a las personas y la oferta formativa para mejorar sus conocimientos y habilidades profesionales, y ayudarlos a la toma de decisión.

Adicionalmente se han combinado técnicas de web semántica con técnicas de Machine Learning (en adelante, ML) permitiendo, a partir de la información disponible, explotar el conocimiento, actualizarlo e inferir nuevas relaciones entre los datos (Obeid, Lahoud, El Khoury, & Champin, 2018). La combinación del sistema de recomendación basado en ontología con técnicas de ML se ha utilizado como un enfoque para mejorar la precisión de las recomendaciones.

Tomando en cuenta la importancia del desarrollo de habilidades profesionales, la amplia oferta de cursos de LL, así como el actual crecimiento de los SR en el ámbito de educación, se propone un SR híbrido, para recomendar cursos de LL a usuarios de RRSS, para desarrollar habilidades profesionales. El núcleo del SR hace uso de un filtrado semántico que utiliza una ontología para modelar sectores de desempeño laboral y áreas de conocimiento, que permite representar habilidades profesionales. Dicha ontología se actualiza vía eventos a partir de datos perfilados obtenidos de registros profesionales de las RRSS, que, haciendo uso de técnicas de ML, clusteriza entidades con la finalidad de hacer predicciones para nuevos datos. Etapas posteriores del sistema de recomendación hacen uso de filtrado por contenido y heurísticas con la finalidad de proponer cursos de formación continua para el desarrollo y/o actualización de habilidades.

## **1.2 Hipótesis y Objetivos**

Con el fin de hacer un sistema de recomendación de cursos de formación continua para desarrollar habilidades profesionales, cuyas recomendaciones sean precisas y relevantes se hace uso de la ontología para la representación del dominio, y por medio de técnicas de ML inferir nuevos datos, que sean de interés para los usuarios. Partiendo de este planteamiento se propone la siguiente hipótesis:

Es posible modelar la información de perfiles de usuario obtenidos en RRSS para mejorar el rendimiento de un SR híbrido basado en ontología y Machine Learning de cursos de formación continua para profesionales, que les permita adaptarse a los cambios que sufren las organizaciones.

A partir de la hipótesis planteada se define el siguiente objetivo general:

Diseñar y desarrollar un Sistema de Recomendación híbrido de cursos de formación continua para la actualización y/o desarrollo de habilidades profesionales a usuarios de redes sociales basado en ontología y Machine Learning.

Para el logro del objetivo general se plantean los siguientes objetivos específicos:

OE1.- Revisar el estado del arte de los SR en el área de LL y necesidades del sector educación

OE2.- Identificar y caracterizar las fuentes de datos en redes sociales y oferta de cursos en internet.

OE3.- Proponer procesos para la construcción de perfiles de usuarios y cursos a partir de registros profesionales en redes sociales y ofertas de cursos en internet.

OE4.- Proponer una estrategia de recomendación para cursos de LL a partir de perfiles de usuarios y de cursos.

OE5.- Definir un modelo basado en ontología y ML para un sistema de recomendación de cursos de LL.

OE6.- Diseñar e impletar las diferentes propuestas de sistemas de recomendación para cursos de LL.

OE7.- Evaluar las diferentes propuestas de sistemas de recomendación para cursos de formación continua a partir de las métricas de rendimiento de los SR, contrastando los resultados con otros SR similares del estado del arte.

### **1.3 Impacto Social**

Como respuesta a la preocupación global y con el objetivo de poner fin a la pobreza, proteger el planeta y mejorar las vidas y las perspectivas de las personas en todo el mundo se plantean los Objetivos de Desarrollo Sostenible (en adelante ODS) (ONU, La Agenda para el Desarrollo Sostenible, 2015) que forman parte de la Agenda 2030, en la cual se establece un plan y una serie de iniciativas para lograrlo, destaca el ODS No. 4: Educación de Calidad, que tiene como objetivo garantizar una educación inclusiva, equitativa y de calidad, y promover oportunidades de aprendizaje durante toda la vida para todos.

Como lo indica (ONU, La Agenda para el Desarrollo Sostenible, 2015) (ONU, Objetivo 4, s.f.), una de las metas del ODS No. 4 es aumentar considerablemente el número de jóvenes y adultos que tienen las competencias necesarias, en particular, técnicas y profesionales, para acceder al empleo, el trabajo decente y el emprendimiento.

Diversas organizaciones han implementado planes de acción para contribuir al logro de las metas correspondientes a los ODS. La (UNESCO, 2016) reconoce el papel significativo que desempeña la educación como motor principal del desarrollo y para la consecución de los ODS, y propone un marco de acción para el ODS No. 4, en donde promueven oportunidades de aprendizaje de calidad a lo largo de la vida para todos, en todos los contextos y en todos los niveles educativos.

Según (UIL, 2021), LL permite a los individuos, dependiendo de sus circunstancias individuales, habilidades existentes y demandas de conocimiento y aprendizaje, elegir trayectorias de aprendizaje. La presente investigación responde a estas iniciativas, y propone un SR que permite sugerir cursos de LL para que los profesionales actualicen sus habilidades, y puedan responder a los cambios producto de los avances tecnológicos y transformaciones de los mercados laborales.

## 1.4 Metodología de la Investigación

De acuerdo con (Sampieri, Collado, & Lucio, 2014) los diseños experimentales se utilizan cuando el investigador pretende establecer el posible efecto de una causa que se manipula, el investigador dirige su actividad con el propósito de crear el fenómeno que se indaga, y así poder observarlo (Rivero, 2008). Según (Bernal, 2010) la investigación experimental se caracteriza porque en ella el investigador actúa conscientemente sobre el objeto de estudio, con el objetivo de conocer los efectos de los actos producidos por el propio investigador como mecanismo o técnica para probar sus hipótesis.

Por tanto, con el propósito de comprobar la hipótesis y cumplir con el objetivo de diseñar y desarrollar un SR híbrido de cursos de formación continua para la actualización y/o desarrollo de habilidades profesionales a usuarios de redes sociales basado en ontología y ML, el presente trabajo de investigación sigue un diseño experimental, con el fin de determinar si los algoritmos propuestos mejoran las recomendaciones en el dominio de la educación.

Se utilizarán técnicas de recolección de datos de RRSS, análisis de documentos utilizando la metodología de revisión sistemática a partir de trabajos de investigación, artículos científicos, entre otros medios impresos de grupos de investigación con lo cual se pueda obtener la información necesaria, para finalmente llevar a cabo el diseño y desarrollo del sistema de recomendación.

La metodología aplicada para el desarrollo de la investigación se muestra en la Figura 1.1, en donde el Fundamento teórico fue una etapa que acompañó todo el desarrollo de la tesis, así mismo se observa un proceso iterativo en la etapa de diseño y desarrollo, cuyo producto llevó a plantear 3 propuestas del SR.

A continuación, se describen cada una de las fases de la metodología:



Figura 1.1 Metodología de Investigación. Elaboración propia

*Fundamento teórico.* En esta primera fase se realizó la revisión sistemática de la literatura lo cual permitió identificar brechas o vacíos en el área de estudio, así como el basamento teórico de la presente investigación. Se recopiló la información relacionada con los SR, naturaleza, tipos, características, algoritmos de filtrado, problemas y evaluación. Así mismo, se revisa información referente a taxonomía y ontología, sus fundamentos, elementos constitutivos y uso para la construcción de conocimiento y organización de elementos de información respectivamente, así como el uso de ML en los SR.

*Planteamiento del problema.* A partir de la revisión del estado del arte, se detectan tendencias y oportunidades de trabajo lo que permitió realizar un planteamiento del problema para cubrir algunos de los vacíos identificados con relación a los SR en el ámbito del desarrollo de habilidades profesionales a través de la formación continua

*Hipótesis y Objetivos.* A partir del resultado de la etapa anterior se define la hipótesis relacionada con la construcción del SR de cursos de formación continua para el desarrollo

y actualización profesional y los objetivos que una vez desarrollados permitirán comprobar esta.

*Diseño y desarrollo del SR.* Esta fase se divide en varias etapas: una primera etapa se realiza el levantamiento, análisis y caracterización de información, se definen estrategias y algoritmos para el sistema de recomendación; en una segunda etapa se realiza los diseños de bajo y alto nivel; en una tercera etapa se construye los diferentes conjuntos de datos (data set), los perfiles de usuarios y cursos, y la ontología y en una cuarta etapa se implementan y verifican los diferentes prototipos del SR.

*Evaluación y pruebas.* En esta fase se define el protocolo de evaluación del SR, se realizan las pruebas y se procede a la evaluación del SR.

*Difusión de los resultados.* Esta última fase permite dar a conocer los resultados obtenidos a la comunidad científica tanto en congresos como en artículos de revista científicas de impacto.

## **1.5 Organización de la tesis**

La presente memoria está organizada en seis capítulos, los cuales se describen a continuación:

Capítulo I, en donde se aborda el planteamiento del problema de estudio, hipótesis, objetivo general y específicos, así como los aportes y el impacto social de la investigación. Siguiendo con la metodología de la investigación y la estructura de ésta.

Capítulo II, en el cual se presenta la revisión de la literatura presentando las bases teóricas necesarias para el desarrollo de la tesis. Se revisan los conceptos de taxonomía y ontología, así como la definición de los sistemas de recomendación, tipos, características, algoritmos y métricas de evaluación.

Capítulo III, se presenta la base de datos utilizada, indicando las principales fuentes de información, tamaño de la muestra, método de adquisición de los datos, selección y representación de estos.

Capítulo IV, describe la arquitectura propuesta para el sistema de recomendación, y se describen los diseños de alto y bajo nivel.

El resultado de la propuesta se presenta en el Capítulo V, en donde se define el protocolo de pruebas, y la evaluación de los resultados.

Finalmente, en el Capítulo VI, se plantean las conclusiones de la investigación y se indican las líneas en las que se pueden orientar el trabajo futuro en el área de sistemas de recomendación en educación.

# Capítulo 2

## Estado del Arte

El desarrollo de esta investigación se enmarca en los SR, en especial, en los orientados a la educación. En este capítulo se presentan los conceptos relacionados a estos ámbitos, encontrados en la revisión del estado del arte, partiendo de los conceptos generales de los SR, técnicas utilizadas y por último se centra en los SR en educación.

### **2.1 Sistemas de Recomendación**

Debido al crecimiento en los últimos años de los servicios en internet, la cantidad de información disponible en la web se ha convertido en un problema para los usuarios en cuanto al tiempo empleado para su búsqueda y a la cantidad de información que esta recupera. Es determinante a la hora de tomar decisiones poder contar con la información pertinente y de manera oportuna.

Los SR han surgido con el fin de hacer frente al problema de sobrecarga de información, con el propósito de ayudar a los usuarios a encontrar aquello que realmente es significativo para sus necesidades, reduciendo eficazmente el volumen y la diversidad de información. Los SR sugieren los artículos más relevantes e importantes basados en predecir el interés del usuario (Adam, Zulkafli, Soh, & Kamal, 2017), ofreciendo a cada usuario una variedad de opciones de acuerdo con sus intereses y preferencias (Venkatraman S. , 2017).

Para (Bokde, Girase, & Mukhopadhyay, 2015), los SR son herramientas de software para ayudar a los usuarios en el proceso de toma de decisiones mediante el filtrado de información, minería de datos y algoritmos de predicción. De acuerdo con (Iqbal, y otros, 2019), un SR se basa en dos entidades, usuarios y artículos, en donde los usuarios proporcionan valoraciones sobre los artículos para que el SR recomiende nuevos artículos, basado en las valoraciones que los usuarios han realizado. Según (Venkatraman S. , 2017), al construir un SR hay que tener presente una serie de cuestiones basadas en el propósito del sistema, como son: “qué, cómo y cuándo se recoge y procesa la información”, lo que lleva a desarrollar diferentes tipos de sistemas de recomendación, que se diferencian en el algoritmo de filtrado.

### **2.1.1 Clasificación de los SR según los algoritmos de filtrado**

Una forma clásica de categorizar los diferentes tipos de SR es la propuesta por (Burke, Brusilovsky, Kobsa, & Nejdl, 2007), quienes distinguen entre seis clases diferentes de enfoques de recomendación: filtrado colaborativo, filtrado basado en contenido, filtrado basado en conocimiento, filtrado basado en contexto, filtrado demográfico y filtrado híbrido, este último surge para superar las limitaciones de los diferentes enfoques.

1. *Filtrado colaborativo*: para (Herlocker, Konstan, Terveen, & Riedl, 2004), los sistemas de filtrado colaborativo son los más simples, calculan la similitud entre los usuarios y predicen las calificaciones de los artículos para el usuario activo de acuerdo con las calificaciones proporcionadas por otros usuarios que tienen preferencias muy similares a las de los usuarios actuales. Singh (2020), distingue dos tipos de filtrado colaborativo: basado en usuario, el

cual sugiere artículos, a partir de usuarios similares al usuario al que se dirige la recomendación, y el filtrado colaborativo basado en artículos, que hace la recomendación basándose en la similitud de los artículos, para lo que toma en cuenta las valoraciones de estos. Entre los algoritmos utilizados en filtrado colaborativo están *K-Nearest Neighbor*, Árboles de decisión y regresión y *Naive Bayes* (Aggarwal, 2016).

2. *Filtrado basado en contenido*: de acuerdo con (Adam, Zulkafli, Soh, & Kamal, 2017) estos sistemas se basan en características proporcionadas previamente por el usuario, que luego se usarán para filtrar todos los elementos del sistema. Los artículos que tengan la puntuación de similitud más alta se recomendarán a los usuarios. En este tipo de filtrado se hace uso de métodos heurísticos o algoritmos de clasificación, tales como clasificadores basados en reglas, modelos basados en la regresión, *Nearest Neighbor*, clasificador de Bayes (Ricci, Rokach, & Shapira, 2015).
3. *Filtrado basado en conocimiento*: según (Anusha & Biradar, 2016), el enfoque basado en conocimiento sugiere elementos al usuario de acuerdo con el conocimiento que tiene sobre ellos y sus relaciones, con el fin de satisfacer las necesidades específicas del usuario.
4. *Filtrado basado en el contexto*: los sistemas de recomendación conscientes del contexto son capaces de reconocer los contextos del usuario y las actividades diarias en tiempo real, como la ubicación, la información del entorno, etc., para sugerir elementos contextualmente relevantes (Otebolaku & Andrade, 2014). Las preferencias del usuario cambian con el tiempo, y esto no es tomado en cuenta por otro tipo de filtrados, (Iqbal, y otros, 2019) señala la importancia de la información contextual para ofrecer mejores recomendaciones. Algunas de las técnicas utilizadas en este tipo de filtrado son la factorización matricial y de regresión lineal (Aggarwal, 2016).
5. *Filtrado demográfico*: las recomendaciones se establecen sobre un perfil demográfico del usuario (Thorat, Goudar, & Barve, 2015), y las recomendaciones se sugieren a partir de similitudes en términos de datos

demográficos en perfiles de usuario como nacionalidad, edad, género, etc. (Ortega, Sánchez, Bobadilla, & Gutiérrez, 2013). De acuerdo con (Aggarwal, 2016), este enfoque está relacionado con la metodología de los sistemas de recomendación sensibles al contexto.

6. *Filtrado Híbrido*: han surgido como un medio para superar cualquier problema que pueda surgir a través del uso individual de las diferentes técnicas de recomendación, combinan dos o más técnicas de entre los enfoques descritos anteriormente para mejorar el rendimiento de las recomendaciones (Burke, Brusilovsky, Kobsa, & Nejdl, 2007). Según (Wita, Bubphachuen, & Chawachat, 2017) el filtrado colaborativo se encuentra con problema de arranque en frío y de escalabilidad, mientras que el enfoque basado en contenido puede enfrentarse a la escasez de datos.

Para Jain, Grover, Thakur & Choudhary (2015), las diferentes técnicas de filtrado utilizadas en los SR tienen muchos desafíos relacionados con los datos: arranque en frío, escalabilidad, escasez, sobre-especialización y con el usuario como privacidad y el problema de la oveja gris (*Grey Sheep*):

*Arranque en frío*:- este problema se refiere a situaciones en las que solo hay unas pocas clasificaciones en las que basar las recomendaciones, lo que suele ocurrir cuando se registran nuevos elementos en el sistema y normalmente no tienen una valoración de los usuarios. También puede presentarse en los nuevos usuarios ya que el sistema no tiene ninguna información relacionada con sus intereses o podría ser posible que no haya valorado ningún artículo todavía, por lo que su gusto es desconocido para el sistema. Este problema se puede presentar en los filtrados colaborativos, basado en contenido y basado en conocimiento (Burke, Brusilovsky, Kobsa, & Nejdl, 2007).

*Escalabilidad*: este problema está relacionado el número de recursos necesarios para el sistema, cuando aumenta el número de usuarios y artículos, el sistema necesita más recursos para dar las recomendaciones más precisas a los usuarios. Es común en los SR colaborativos (Jain, Grover, Thakur, & Choudhary, 2015).

*Escasez*: la escasez de los datos es otro de los problemas de la mayoría de los sistemas de recomendación, esto se presenta cuando los datos de valoración son escasos, se puede decir que la escasez es el problema de la falta de conocimiento (Jain, Grover, Thakur, & Choudhary, 2015).

*“Grey Sheep”*: se produce cuando un usuario puede ser clasificado en más de un grupo de usuarios. La similitud de este usuario con dos o más grupos es igual, lo que hace que las recomendaciones que recibirá sean inexactas (Bouraga, Jureta, Faulkner, & Herssens, 2014).

*Privacidad*: es un problema principalmente de los sistemas demográficos. Para dar la recomendación más precisa el sistema adquiere información demográfica del usuario (edad, sexo, dirección de correo electrónico, aficiones, etc.) y datos sobre la ubicación, lo que puede violar la privacidad del usuario (Jain, Grover, Thakur, & Choudhary, 2015).

*Sobre-especialización*: este es uno de los problemas más comunes a los que se enfrenta el sistema de recomendación basado en contenido. Sucede cuando el conjunto de artículos recomendados es muy homogéneo, los artículos serán muy similares a los que el usuario ya valorado. Impide a los usuarios de descubrir algo nuevo y diferente (Jain, Grover, Thakur, & Choudhary, 2015), (Bouraga, Jureta, Faulkner, & Herssens, 2014).

En la siguiente tabla se pueden apreciar las características de los diferentes filtrados y los problemas a los que se puede enfrentar cada uno de ellos.

Tabla 2.1. Tipo algoritmos de filtrado, características y limitaciones. Elaboración propia

Filtrado	Características	CS	EC	ES	GR	PR	SE
Colaborativo	Utiliza únicamente información sobre los perfiles de valoración de los distintos usuarios y artículos.	X	X	X	X		
Basado en Contenido	Genera recomendaciones a partir de dos fuentes: las características asociadas a los productos y las valoraciones que un usuario les ha dado.	X					X
Basado en conocimiento	Sugiere productos basados en inferencias sobre las necesidades y preferencias de un usuario.	X					
Basado en Contexto	Utiliza la información del contexto del usuario y las actividades diarias en tiempo real.	X					
Demográfico	Proporciona recomendaciones basadas en un perfil demográfico del usuario.	X				X	
CS: arranque en frío	EC: Escalabilidad	ES: Escasez					
GS: Grey Sheep	PR: Privacidad	SE: Sobre especialización					

### 2.1.2 Funciones de Similitud en los SR

De acuerdo con (Ortega, Sánchez, Bobadilla, & Gutiérrez, 2013), sin importar el algoritmo de filtrado que se utilice en los SR, el proceso de recomendación se basa en el cálculo de la similitud. Este valor, generalmente entre  $-1$  y  $1$  indica que tan relacionados están los elementos. Las funciones de similitud más utilizadas se describen a continuación (Hassanieh, Jaoudeh, Abdo, & Demerjian, 2018), (Lin, y otros, 2019), (Jayvardhan, Thomas, & Lal Yadav, 2013):

*Similitud de Correlación de Pearson (CP)*. Se utiliza para medir la similitud entre dos vectores. La similitud entre el usuario  $u$  y el usuario  $v$  se calcula mediante la siguiente ecuación (2.1):

$$CP(u, v) = \frac{\sum_{i \in I_{uv}} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_{uv}} (r_{vi} - \bar{r}_v)^2}} \quad (2.1)$$

Donde  $u$  y  $v$  representan a los usuarios,  $I_{uv}$  es la lista de artículos valorados,  $r_{ui}$  y  $r_{vi}$  es la valoración del usuario  $u$  y  $v$  respectivamente sobre el artículo  $I$ , y  $\bar{r}_{ui}$  y  $\bar{r}_{vi}$  es la media de valoraciones dadas por el usuario  $u$  y  $v$  respectivamente.

*Similitud del vector coseno (SC).* Se calcula a partir de una matriz de puntuación basada en las valoraciones de los usuarios, en donde, establece el vector de filas para representar las valoraciones de diferentes usuarios, y el vector de columnas indica las valoraciones de diferentes usuarios para el mismo artículo. El grado de similitud se mide calculando el coseno del ángulo de los dos vectores como se indica en la siguiente ecuación (2.2):

$$SC = \frac{u \cdot v}{\|u\| \|v\|} \quad (2.2)$$

Donde  $u$  y  $v$  representan los vectores de puntuación de los elementos y,  $\|u\|$  y  $\|v\|$  representan los módulos de los vectores.

La similitud de coseno no toma en cuenta los diferentes criterios de calificación de los usuarios, por lo que se propone una modificación a este método conocida como Similitud de Coseno modificada o ajustada.

*Similitud de Coseno Modificada (SCM).* La similitud de coseno ajustada corrige la deficiencia del método de la similitud del coseno restando el promedio de las calificaciones del usuario correspondiente de cada par clasificado conjuntamente, calculándose como indica (2.3):

$$SCM(u, v) = \frac{\sum_{i \in I_{uv}} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I_u} (r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{u \in I_v} (r_{uj} - \bar{r}_u)^2}} \quad (2.3)$$

Donde  $I_{uv}$  denota un conjunto de elementos que todos los usuarios  $u$  y  $v$  han evaluado conjuntamente,  $I_u, I_v$  respectivamente, representan la colección de elementos de puntuación individuales,  $u$  y  $v$  representan a los usuarios,  $r_{ui}$  y  $r_{vi}$  es la valoración del usuario  $u$  y  $v$  respectivamente sobre el artículo  $I$ , y  $\bar{r}_{ui}$  y  $\bar{r}_{vi}$  es la media de valoraciones dadas por el usuario  $u$  y  $v$  respectivamente.

*Similitud por Distancia Euclidiana (SDE)*. La definición formal de la medición de distancia euclidiana establece que es la raíz cuadrada de la suma de las diferencias al cuadrado entre los elementos correspondientes de dos vectores. Este concepto de medir la distancia se adopta en el cálculo de la similitud entre usuarios al considerar a los usuarios como elementos en el espacio euclídeo cuyo valor de coordenada es su preferencia. Se calcula mediante la siguiente fórmula (2.4):

$$SDE_{uv} = \sqrt{\sum_{i=1}^n (x_{ui} - x_{vi})^2} \quad (2.4)$$

Donde  $i$  es el artículo,  $u$  y  $v$  representan a los usuarios,  $x_{ui}$  y  $x_{vi}$  son las valoraciones dadas por los usuarios  $u$  y  $v$ , al artículo  $i$ .

### 2.1.3 Evaluación de los sistemas de recomendación

Un aspecto importante en los SR es la evaluación. Un diseño incorrecto de la evaluación experimental puede llevar a una subestimación o sobreestimación de la precisión real de un algoritmo o modelo. De acuerdo con (Aggarwal, 2016) y (Ricci, Rokach, & Shapira, 2015), la evaluación de los sistemas de recomendación se puede hacer mediante una experimentación en línea, y fuera de línea.

Las evaluaciones en línea permiten medir directamente los objetivos generales del sistema, pueden utilizarse para entender cómo los objetivos generales del SR se ven afectados por las propiedades del sistema, como la precisión de las recomendaciones y la diversidad de estas. Sin embargo, como es difícil variar estas propiedades y comparar muchos algoritmos mediante pruebas en línea, este tipo de evaluación puede resultar

costosa (Ricci, Rokach, & Shapira, 2015). Adicionalmente, es difícil utilizar este método durante la fase de puesta en marcha puesto que no cuentan con un número suficiente de usuarios, y no se pueden controlar totalmente las acciones de los usuarios en el proceso de evaluación (Aggarwal, 2016).

Por estas razones, es mejor realizar una evaluación en línea después de que un amplio estudio fuera de línea demuestre que los enfoques son razonables, y quizás después de un estudio de usuarios que mida la actitud de éstos hacia el sistema (Ricci, Rokach, & Shapira, 2015).

Por su parte, (Aggarwal, 2016) indica que, en los SR, las pruebas sobre múltiples conjuntos de datos son especialmente importantes para garantizar la generalización de estos, de modo que funcionen en una variedad de escenarios. En estos casos, se utilizan evaluaciones fuera de línea. Así mismo, subraya que son los más habituales para evaluar a los SR desde la perspectiva de la investigación.

En los métodos fuera de línea son importantes una serie de medidas para poder realizar una evaluación completa y precisa del SR. Es importante diseñar el sistema de evaluación cuidadosamente para que las métricas medidas reflejen realmente la eficacia del sistema desde la perspectiva del usuario. Los métodos fuera de línea proporcionan una evaluación aceptada ampliamente debido a robustez de las métricas utilizadas. De acuerdo con (Herlocker, Konstan, Terveen, & Riedl, 2004), la mayoría de las evaluaciones de los sistemas de recomendación se han centrado en la evaluación de la precisión. A continuación, se presentan las métricas más utilizadas para evaluar el rendimiento de los SR:

*Precisión.* La precisión de las recomendaciones es una de las medidas más utilizadas en la evaluación de los SR. Partiendo del hecho de que las valoraciones de los usuarios son precisas y coherentes, la precisión de la predicción suele ser independiente de la interfaz de usuario, por lo que puede medirse en un experimento fuera de línea.

Sea  $S$  el conjunto de entradas especificadas (observadas) y  $E \subset S$ , el conjunto de entradas de prueba utilizado para la evaluación. Cada entrada de  $E$  es un par de índices

usuario-artículo de la forma  $(u, j)$ . Sea  $r_{uj}$  el valor de la calificación de la entrada  $u, j \in E$  que se utiliza en el conjunto de pruebas. Además, sea  $\hat{r}_{uj}$  la calificación predicha de la entrada  $(u, j)$  por el algoritmo de entrenamiento utilizado. El error específico de la entrada viene dado por  $e_{uj} = \hat{r}_{uj} - r_{uj}$ .

Para medir la precisión, se utilizan con frecuencia una serie de métricas comunes, como la raíz del error medio al cuadrado (RMSE) y el error medio absoluto (MAE) (Aggarwal, 2016). Estas se centran en el valor de las puntuaciones que el SR predice y miden qué tan cerca están respecto a la puntuación real que le daría el usuario. Cuanto menor sean estos valores, mayor será la precisión de las predicciones.

El RMSE entre las valoraciones pronosticadas y las reales y el MAE vienen dado por (2.5) y (2.6):

$$RMSE = \sqrt{\frac{\sum_{(u,j) \in E} e_{u,j}^2}{|E|}} \quad (2.5)$$

$$MAE = \frac{\sum_{(u,j) \in E} |e_{u,j}|}{|E|} \quad (2.6)$$

Otra alternativa para el cálculo de la precisión es el error medio cuadrático medio (MSE) dado por la ecuación (2.7):

$$MSE = \frac{\sum_{(u,j) \in E} e_{u,j}^2}{|E|} \quad (2.7)$$

En otros casos, no es de interés evaluar si el sistema predice correctamente las valoraciones de los artículos, suele ser interesante evaluar si el sistema predice correctamente los artículos que el usuario seleccionaría.

Sea  $t$  el tamaño de la lista de objetos recomendados,  $S(t)$  es el conjunto de elementos recomendados y  $G$  representa el conjunto de elementos relevantes que son consumidos

por el usuario, entonces la precisión se define como el porcentaje de elementos recomendados que son relevantes para el usuario y está dada por la ecuación (2.8):

$$Precisión = 100 \frac{|S(t) \cap G|}{|S(t)|} \quad (2.8)$$

*Recuperación.* Otra medida utilizada para la evaluación de los SR es la recuperación, la cual representa la proporción entre la recomendación y la preferencia del usuario, y está dada por (2.9):

$$Recuperación = 100 \frac{|S(t) \cap G|}{|G|} \quad (2.9)$$

*Cobertura.* Aún si el SR tiene un alto porcentaje de precisión puede ser posible que no pueda recomendar nunca una determinada proporción de los artículos, o que no pueda recomendar nunca a una determinada proporción de los usuarios. La relación de usuarios a los que el sistema le ha hecho recomendación, se denomina cobertura y está dada por (2.10):

$$Covertura = \frac{|U_{u=1}^m T_u|}{n} \quad (2.10)$$

Donde,  $T_u$  representa la lista de los  $k$  mejores artículos recomendados al usuario  $u \in \{1..m\}$  y  $n$  representa en número de artículos.

*Media armónica entre precisión y recuperación.* Aunque hay un equilibrio natural entre la precisión y la recuperación, un aumento en la recuperación no siempre conduce a una reducción en la precisión. Una forma de crear una sola medición que resuma tanto la precisión como la recuperación es la media armónica entre la precisión y la recuperación ( $F_1$ ), que se calcula como muestra la ecuación (2.11):

$$F_1 = \frac{2 * Precisión * Recall}{Precisión + Recall} \quad (2.11)$$

*Novedad y Serendipia.* Existen un conjunto de propiedades, además de la recuperación y cobertura, que se definen para medir el éxito de un SR más allá de la precisión, que no siempre capta características importantes de las recomendaciones, como son, la serendipia y la novedad. Estas, son importantes para la experiencia del usuario, ya que tienen un importante impacto a corto y largo plazo en las tasas de conversión de conversión de las recomendaciones (Aggarwal, 2016).

La novedad de un sistema de recomendación evalúa la probabilidad de que un sistema de recomendación ofrezca al usuario recomendaciones que no conoce o que no ha visto antes.

La serendipia es una medida del nivel de sorpresa en las recomendaciones exitosas. En otras palabras, las recomendaciones deben ser inesperadas. En cambio, la novedad sólo requiere que el usuario no haya tenido conocimiento de la recomendación antes. La serendipia es una condición más fuerte que la novedad.

## **2.2 Machine Learning**

En los últimos años, se han introducido técnicas de inteligencia artificial en los SR, en donde se incluyen técnicas de ML (Anusha & Biradar, 2016). Según (Karakaya & Aytakin, 2018), el uso de estas técnicas ha sido una solución prometedora a la hora de diseñar SR en la era del Big Data. Para (Samin & Azim, 2019), indica que el uso de técnicas de ML con filtrado colaborativo ayuda a resolver los problemas de arranque en frío. Así mismo, se utilizan para la selección de la información de forma adecuada y personalizada con el fin de obtener conocimientos del mundo real y así ofrecer recomendaciones convenientes.

De acuerdo con (Akmaliah, Krisnadhi, Sensuse, Rahayu, & Wulandari, 2018), ML se utiliza como una simulación del aprendizaje humano para obtener conocimientos del mundo real, con el objetivo de generalizar un patrón detectable o crear una regla a partir de ejemplos dados. Para (Ruchika, Singh, & Sharma, 2017), ML es un proceso a través

del cual se entrena el sistema y éste es capaz de tomar decisiones como los humanos. El conocimiento se adquiere mediante autoaprendizaje, la experiencia o la enseñanza.

En ML se clasifica en tres categorías: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje de refuerzo.

*Aprendizaje supervisado*, en esta categoría se aprende de datos etiquetados. El sistema analiza el conjunto de datos actual y cada vez que encuentre un nuevo el sistema podrá generar la solución correcta al hacerlo coincidir el conjunto de datos previamente entrenado (Ruchika, Singh, & Sharma, 2017). Los principales segmentos de este tipo son los problemas de clasificación y problemas de regresión (Akmaliah, Krisnadhi, Sensuse, Rahayu, & Wulandari, 2018).

*Aprendizaje no supervisado*, el algoritmo aprende por sí mismo, sin la variable objetivo, para encontrar patrones y relaciones ocultas en los datos proporcionados. Las categorías de este tipo son la reducción de la dimensionalidad y el clustering.

*Aprendizaje por refuerzo*, el agente realiza una serie de acciones decisivas sin supervisión y, al final, se da una recompensa, ya sea +1 o -1. En función de la recompensa final recompensa final, el agente reevalúa sus trayectorias (Akmaliah, Krisnadhi, Sensuse, Rahayu, & Wulandari, 2018).

Algunos SR que utilizan ML son el presentado por (Taheri & Irajian, 2018) que propone un marco de trabajo para la recomendación de películas, así mismo, (Mana & Sasipraba, 2021) proponen dos modelos prototipos que utilizan ML. El primer prototipo es un sistema de recomendación de servicios bancarios y el segundo es un sistema de recomendación de películas.

En los SR, ML se puede combinar con las técnicas de ontología para obtener las recomendaciones más adecuadas según las necesidades, además de abordar la sobrecarga de información, pueden solucionar los problemas de arranque en frío (Akmaliah, Krisnadhi, Sensuse, Rahayu, & Wulandari, 2018). Así mismo, la heterogeneidad y sobrecarga de información hacen necesario el análisis y selección de información de

forma adecuada y de acuerdo con el contexto para el desarrollo de los SR. Con este fin, se apoyan en técnicas de web semántica, como la ontología.

## **2.3 Ontología**

La ontología ha sido ampliamente utilizada como una fuente importante para el enriquecimiento semántico de los datos homogéneos y heterogéneos (Alkhamisi & Saleh, 2020). Adicionalmente, ayuda a la identificación de entidades inequívocas a partir de fuentes de información heterogéneas y de las relaciones entre entidades. Utilizando la ontología como forma de organizar la información, es más fácil extraer la información básica de una entidad (Shu, Lun, Run, & Ye, 2016).

Las ontologías proporcionan una comprensión universalmente acordada de un dominio, que puede ser reutilizado y compartido entre aplicaciones y grupos, facilitando la comprensión computacional y la interoperabilidad entre personas y organizaciones. Se utiliza como solución para resolver los conflictos semánticos entre las fuentes de datos (Alrumaih, Mirza, & Alsalamah, 2020).

La ontología es una conceptualización de la representación del conocimiento que puede traducirse a lenguaje de máquina, con el fin de modelar el del mundo real (Shao, 2017). De acuerdo con (Akmaliah, Krisnadhi, Sensuse, Rahayu, & Wulandari, 2018), una ontología se define como una representación formal de la información, es "conjunto de conceptos dentro de un dominio, utilizando un vocabulario compartido para denotar los tipos, las propiedades y las interrelaciones de esos conceptos". Para esto, la ontología define las clases de cosas que existen en un dominio de aplicación y asocia una definición precisa a cada concepto y tipo de relación que se utiliza (Alrumaih, Mirza, & Alsalamah, 2020).

Los principales componentes de una ontología son, los conceptos, las relaciones, las instancias y los axiomas. Un concepto, representa un conjunto o clase de entidades o "cosas" dentro de un dominio. Las relaciones, describen las interacciones entre los conceptos o las propiedades de un concepto. Las instancias, son las "cosas" representadas

por un concepto. Por último, los axiomas, se utilizan para restringir los valores de las clases o instancias. Una ontología y sus instancias individuales de clases construyen una base de conocimiento (Aissaoui & Oughdir, 2020).

El trabajo de (Alkhamisi & Saleh, 2020) indica que la utilización de ontologías proporciona beneficios potenciales para la integración de datos heterogéneos. Donde algunas de las principales ventajas del uso de ontologías son:

- Proporciona una interfaz conceptual sólida para los esquemas de bases de datos con la información semánticamente rica.
- Soporta la transformación de las múltiples fuentes de datos relevantes basándose en su representación conocimiento.
- Durante la integración de datos, la ontología gestiona de manera eficiente las fuentes de datos heterogéneas y reconoce la información incoherente.
- La integración de datos basada en la ontología no crea un impacto en el coste incluso cuando se incrementa el número de fuentes de datos.
- La integración semántica de datos preserva el contexto de los datos en todas las condiciones, ya que la ontología es flexible.

Las ontologías se han utilizado en distintos ámbitos, con el objetivo principal de poder integrar datos heterogéneos de las distintas fuentes en un dominio específico. El trabajo presentado por (Yang, Cormican, & Yu, 2019) construye una ontología en el ámbito de la ingeniería de sistemas, en donde a partir de diferentes fuentes, se identifican terminología común y las relaciones lógicas entre los conceptos esenciales con el objetivo de demostrar que éstas pueden ser integradas. En este mismo dominio, (Velasquez, Castro, & Gaona, 2018), proponen una ontología para los planes de estudio de la carrera de Ingeniería de Sistemas, con la finalidad de definir al Ingeniero de Sistemas que pueda servir para guiar al ingeniero en su desarrollo y evolución profesional. Por su parte, (ul haq Dar & Dorn, 2018), proponen la creación de ontologías para la clasificación de ofertas

de empleo procedentes de diferentes sitios web de ofertas de empleo, extrayendo la descripción del texto de las ofertas.

Para el diseño y desarrollo de ontologías existen varias metodologías que definen las actividades a seguir. Entre algunas metodologías encontradas en la revisión del estado del arte, destacan la Ontology Development 101 y Methontology.

Ontology Development 101, fue propuesta por la Universidad de Stanford, (Noy & McGuinness, 2001), con la intención de ofrecer una guía inicial, en la cual apoyarse para el desarrollo de ontologías. Sugieren los siguientes pasos: determinar el dominio y ámbito de la ontología, reutilizar ontologías, enumerar los términos importantes del dominio, definir las propiedades de las clases (slots) y crear instancias.

Determinar el dominio y ámbito de la ontología, en este paso se define el dominio que cubrirá la ontología, el uso que se le dará, a qué tipo de preguntas debe dar respuesta la información de la ontología y por último quién la utilizará y mantendrá. Reutilizar ontologías existentes: en este paso se debe considerar si es útil incorporar ontologías ya existentes. Esta reutilización puede ser un requisito si se tiene que interactuar con otras aplicaciones que ya se han comprometido con determinadas ontologías o vocabularios controlados. Enumerar los términos importantes del dominio: esta actividad dará respuesta a las siguientes cuestiones: ¿Cuáles son los términos de los que se quiere hablar? ¿Qué propiedades tienen esos términos? ¿Qué se quiere decir sobre esos términos? Definir las clases y jerarquía de clases, así como definir las propiedades de las clases (slots). Definidas las clases, se debe describir la estructura interna de los conceptos, definir cuáles son sus propiedades y las características asociadas a ellos. Por último, sugieren crear instancias. En esta etapa se debe considerar que la definición de una instancia individual de una clase requiere (1) elegir una clase, (2) crear una instancia individual de la clase y (3) rellenar los valores del slot.

Por su parte, Methontology, fue desarrollada por el Grupo de Ingeniería Ontológica de la Universidad Politécnica de Madrid (Fernández, Gómez-Pérez, & Juristo, 1997). Esta metodología está conformada por las siguientes actividades: especificación, adquisición de conocimiento, integración, implementación, evaluación y documentación.

En especificación, se define el propósito, el nivel de formalidad y alcance de la ontología; adquisición de conocimientos. En la conceptualización de la ontología, se estructurará el conocimiento del dominio en un modelo conceptual. La etapa de integración debe garantizar que los conjuntos de definiciones nuevas y reutilizadas se basen en el mismo conjunto de términos básicos, en caso de utilizar la reutilización de definiciones ya construidas en otras ontologías en lugar de empezar desde cero. Implementación, en esta actividad se codifica la ontología en un lenguaje formal. En la evaluación, se realiza la verificación y la validación, en donde se garantiza la corrección de la ontología, sus entornos de software y la se garantiza que se corresponden con el sistema que se supone que representa. Por último, para la documentación de la ontología no existen directrices consensuadas, la documentación muchas veces está compuesta por el código de la ontología, el texto en lenguaje natural y las definiciones formales de esta.

Las distintas metodologías tienen actividades en común, y su objetivo es representar la realidad de la forma más cercana posible. Para (Alrumaih, Mirza, & Alsalamah, 2020) y (Noy & McGuinness, 2001), el diseño de las ontologías es un proceso creativo y no existe una única metodología correcta para el diseño de estas.

## **2.4 Sistemas de Recomendación en Educación**

Los SR se han convertido en una herramienta ampliamente utilizada en diferentes ámbitos, incluyendo la educación, en donde el número de recursos educativos está en continuo crecimiento. Cada vez se hace más difícil, para los motores de búsqueda tradicionales, cumplir con los requisitos relacionados con las búsquedas en línea de información sobre productos y servicios educativos por parte de los estudiantes durante el proceso de aprendizaje (Cui, Guo, & Liang, 2018). En tal sentido, se ha propuesto un número significativo de SR en el ámbito de la educación, en donde los usuarios objetivo son estudiantes, profesores y asesores académicos, y los elementos recomendados son materiales educativos, objetos de aprendizaje, documentos, universidades e información sobre cursos, rendimiento de los estudiantes y campos de estudio (Deschênes, 2020).

El reto al que se enfrentan los SR educativos es cómo entender mejor el interés y el contexto del alumno, y el propósito del dominio (Tang, Wu, Sun, & Su, 2012). Estos, están condicionados por los diferentes niveles y tipos de educación. Según (Tarus, Niu, & Mustafa, 2018), la implementación de los SR en educación requiere tener en cuenta un amplio conjunto de variables que pueden incluir, entre otras, el nivel de conocimiento y las competencias por parte de los estudiantes. De esta manera, la pertinencia de los recursos educativos recomendados mejorará el proceso de aprendizaje del alumno y, de ahí la importancia de poder garantizar de forma adecuada y fiable información relevante y útil para el usuario.

Dada la rápida evolución de los SR en educación, se realizó una revisión sistemática y de este modo, poder identificar brechas en este ámbito, así como proporcionar una guía marco adecuada para la presente investigación. La búsqueda de artículos se realizó inicialmente en dos de las principales bibliotecas digitales de contenido científico en el área de informática, como lo son IEEE y ACM, y para complementar la búsqueda, se tuvieron en cuenta Web of Science (WoS) y Scopus ya que son dos bases de datos que tienen acceso a un importante y amplio número de aplicaciones en las diferentes áreas. Se tomó en cuenta el periodo comprendido entre los años 2015 y 2020, analizando 98 artículos para conocer las características de los SR en el dominio de la educación. Se analizó el tipo de educación en donde hacen las recomendaciones, usuarios y sus características, elementos de recomendación y las técnicas utilizadas para su desarrollo. La Tabla 2.2 muestra la información obtenida de cada artículo, en donde se observa el crecimiento de los SR en el dominio de educación en los últimos dos años.

Tabla 2.2 Artículos analizados en la revisión de la literatura. Elaboración propia

Referencias	Tipo de educación	Tipo de usuario	Elementos de recomendación	Enfoque de desarrollo	Tipo de plataforma
(Tejeda-Lorente, Moreno, Porcel, Galindo-Moreno, & Herrera-Viedma, 2015)	Formal	Estudiantes	Recursos de aprendizaje	FH	Web
(Bokde, Girase, & Mukhopadhyay, 2015)	Formal	Estudiantes	Universidades	FC	Web
(Borges & Stiubiener, 2015)	Formal	Estudiantes	Recursos de aprendizaje	FKB	N/E

<b>(Ganeshan &amp; Li, 2015)</b>	Formal	Estudiantes	Cursos	FC	Web
<b>(Behar, y otros, 2015)</b>	Formal	Estudiantes	Recursos de aprendizaje	FH	N/E
<b>(Montuschi, Lamberti, Gatteschi, &amp; Demartini, 2015)</b>	Informal	Estudiantes	Cursos	FKB	Web
<b>(Rodríguez, Ovalle, &amp; Duque, 2015)</b>	Formal/ Informal	Estudiantes	Recursos de aprendizaje	FH	N/E
<b>(Sergis &amp; Sampson, 2015)</b>	Formal	Profesores	Recursos para prácticas docente	FKB	N/E
<b>(Zheng, y otros, 2015)</b>	Informal	Estudiantes	Cursos	FC	N/E
<b>(Elbadrawy &amp; Karypis, 2016)</b>	Formal	Estudiantes	Cursos	FC	N/E
<b>(Khacheb &amp; Cheniti Belcadhi, 2016)</b>	Formal/ Informal	Estudiantes	Recursos de aprendizaje	FKB	N/E
<b>(Kurilovas, Kurilova, Kurilova, &amp; Melesko, 2016)</b>	Formal	Estudiantes	Recursos de aprendizaje	FKB	N/E
<b>(Meryem, Douzi, &amp; Chantit, 2016)</b>	Formal	Estudiantes	Secuencia de estudios/programas de estudio	FH	N/E
<b>(Aguilar, Valdiviezo-Díaz, &amp; Riofrio, 2016)</b>	Formal	Estudiantes	Recursos de aprendizaje	FKB	N/E
<b>(Xu, Xing, &amp; van der Schaar, 2016)</b>	Formal	Estudiantes	Secuencia de estudios/programas de estudio	FH	Web
<b>(Hasan, Ahmed, Abdullah, &amp; Rahman, 2016)</b>	Formal	Estudiantes	Cursos de postgrado	FC	N/E
<b>(Bourkougou &amp; El Bachari, 2016)</b>	Formal	Estudiantes	Recursos de aprendizaje	FC	Web
<b>(Fazeli, Rajabi, Lezcano, Drachsler, &amp; Sloep, 2016)</b>	Informal	Estudiantes	Cursos	FH	N/E
<b>(Kuznetsov, Kordík, Tomás Rehorek, &amp; Kroha, 2016)</b>	Formal	Estudiantes	Cursos	FH	Web
<b>(Symeonidis &amp; Malakoudis, 2016)</b>	Informal	Estudiantes	Recursos de aprendizaje	FKB	Web
<b>(Bañeres &amp; Conesa, 2016)</b>	Formal/ Informal	Empleados	Cursos	FKB	N/E

<b>(Ochirbat &amp; Shih, 2017)</b>	Formal	Estudiantes	Profesiones	FC	N/E
<b>(Bakhshinategh, Spanakis, Zaïane, &amp; ElAtia, 2017)</b>	Formal	Estudiantes	Cursos	FC	N/E
<b>(Patel, Kakuste, &amp; Eirinaki, 2017)</b>	Formal	Estudiantes	Secuencia de estudios/ Profesiones	FC	Web
<b>(De Borba, Gasparini, &amp; Lichtnow, 2017)</b>	Formal	Estudiantes	Recursos de aprendizaje	FH	N/E
<b>(Hajri, Bourda, &amp; Popineau, 2017)</b>	Informal	Estudiantes	Recursos de aprendizaje	FKB	Web
<b>(Li, Shi, Zhang, &amp; Yun, 2017)</b>	Formal	Estudiantes	Recursos de aprendizaje	FC	N/E
<b>(Brigui-Chtioui, Caillou, &amp; Negre, 2017)</b>	Informal	Estudiantes	Recursos de aprendizaje	FC	N/E
<b>(Chicaiza, Piedra, Lopez-Varga, &amp; Tovar-Caro, 2017)</b>	Informal	Estudiantes	Recursos de aprendizaje	FKB	Web
<b>(Bhumichitr, Channarukul, Saejiem, Jiamthaphaksin, &amp; Nongpong, 2017)</b>	Formal	Estudiantes	Cursos electivos	FC	N/E
<b>(El Mabrouk, Gaou, &amp; Rtili, 2017)</b>	Informal	Estudiantes	Recursos de aprendizaje	FH	N/E
<b>(Dwivedi &amp; Roshni, 2017)</b>	Formal	Estudiantes	Cursos Electivos	FKB	N/E
<b>(Li &amp; Li, 2017)</b>	Informal	Estudiantes	Cursos	FC	Web
<b>(Yan, Hara, Kazuma, &amp; He, 2017)</b>	Formal	Estudiantes	Cursos	FKB	N/E
<b>(Dang, 2018)</b>	Informal	Profesores/ Estudiantes	Recursos de aprendizaje	FKB	N/E
<b>(Gulzar, 2018)</b>	Informal	Estudiantes	Recursos de aprendizaje	FH	N/E
<b>(Baskota &amp; Ng, 2018)</b>	Formal	Estudiantes	Cursos de postgrado	FH	Web
<b>(Bozyiğit, Bozyiğit, Kılınc, &amp; Nasiboglu, 2018)</b>	Formal	Estudiantes	Cursos Electivos	FC	N/E
<b>(Cheng, Zhang, &amp; Shi, 2018)</b>	Formal	Estudiantes	Recursos de aprendizaje	FKB	N/E
<b>(El Fazazi, Qbadou, Sahi, &amp; Mansouri, 2018)</b>	Informal	Estudiantes	Recursos de aprendizaje	FC	N/E

(Zhu, y otros, 2018)	Formal	Estudiantes	Secuencia de estudios/programas de estudio	FC	N/E
(Li H. , Li, Zhang, Zhong, & Cheng, 2018)	Informal	Estudiantes	Recursos de aprendizaje	FC	Web
(Bauman & Tuzhilin, 2018)	Formal	Estudiantes	Recursos de aprendizaje	FKB	N/E
(Lessa & Brandao, 2018)	Formal/ Informal	Estudiantes	Cursos	FCB	Web
(Nadeem, Dustin, & Shane, 2018)	Formal	Estudiantes	Recursos de aprendizaje	FC	N/E
(Zhao, Wang, Wang, Zhou, & Jiang, 2018)	Informal	Estudiantes	Artículos	FH	N/E
(Campos, dos Santos, & Oliveira, 2018)	Informal	Estudiantes	Recursos de aprendizaje	FKB	N/E
(Hinz & Pimenta, 2018)	Informal	Estudiantes	Recursos de aprendizaje	FC	Web
(Huang, y otros, 2018)	Informal	Estudiantes	Cursos	FH	Web
(Huo, Xiao, & Ni, 2018)	Formal	Estudiantes	Recursos de aprendizaje	FCB	N/E
(Ibrahim, Yang, Ndzi, Yang, & Almaliki, 2018)	Formal	Estudiantes	Cursos	FH	N/E
(Dahdouh, Dakkak, Oughdir, & Ibriz, 2019)	Formal	Estudiantes	Recursos Pedagógicos	ML	N/E
(Gianotti, Cazella, & Behar, 2019)	Formal	Estudiantes	Recursos de aprendizaje	FCT	N/E
(Huang, y otros, 2019)	Formal	Estudiantes	Exámenes multiobjetivo	ML	N/E
(Britto, Prabhu, Gawali, & Jadhav, 2019)	Formal	Estudiantes	Cursos electivos	ML	N/E
(Jiang & Pardos, 2019)	<b>Formal/ Informal</b>	<b>Estudiantes</b>	<b>Cursos</b>	<b>ML</b>	<b>N/E</b>
(Jiang, Pardos, & Wei, 2019)	Formal	Estudiantes	Cursos	ML	N/E
(Shi, Wen, Wang, & Ouyang, 2019)	Formal	Estudiantes	Recursos de aprendizaje	FC	N/E
(Kamila, Subastian, & Rosmasari, 2019)	Formal	Estudiantes	Cursos electivos	ML	N/E

<b>(Morsomme &amp; Alferez, 2019)</b>	Formal	Estudiantes	Cursos	ML	N/E
<b>(Mrhar &amp; Abik, 2019)</b>	Formal	Estudiantes	Cursos	ML	N/E
<b>(Pardos &amp; Jiang, 2019)</b>	Formal	Estudiantes	Cursos	ML	N/E
<b>(Prisco, y otros, 2019)</b>	Formal	Estudiantes	Problemas de programación	N/E	Chatbot
<b>(Ren, Ning, Lan, &amp; Rangwala, 2019)</b>	Formal	Estudiantes	Cursos	ML	N/E
<b>(Rizun, 2019)</b>	Formal	Estudiantes	Recursos de aprendizaje	Grafos	N/E
<b>(Samin &amp; Azim, 2019)</b>	Formal	Académicos	Asesoría académica	ML	Web
<b>(Slim, y otros, 2019)</b>	Formal	Estudiantes	Asesoría académica	ML	N/E
<b>(Vaidhehi &amp; Suchithra, 2019)</b>	Formal	Estudiantes	Programas de estudio	FC	N/E
<b>(Zhang, Liu, &amp; Wang, 2020)</b>	Informal	Estudiantes	Cursos	ML	N/E
<b>(Mondal, Patra, Mishra, &amp; Patra, 2020)</b>	Informal	Estudiantes	Cursos	FC	N/E
<b>(Brik &amp; Touahria, 2020)</b>	Formal	Académicos	Recursos de aprendizaje	FC	N/E
<b>(Na, Xiuyuan, &amp; Na, 2020)</b>	Formal/ Informal	Estudiantes	Recursos de aprendizaje	FH	N/E
<b>(Cheng &amp; Bu, 2020)</b>	Formal	Estudiantes	Recursos de aprendizaje	FH	N/E
<b>(Das &amp; Akour, 2020)</b>	Informal	Estudiantes	Recursos de aprendizaje	Lógica difusa	N/E
<b>(Dhar &amp; Jodder, 2020)</b>	Formal	Estudiantes	Programas de estudio	ML	N/E
<b>(Gong, y otros, 2020)</b>	Informal	Estudiantes	Cursos	Grafos	N/E
<b>(Gordillo, López-Fernández, &amp; Verbert, 2020)</b>	Informal	Estudiantes	Cursos	FH	N/E
<b>(Chen, y otros, 2020)</b>	Informal	Estudiantes	Recursos de aprendizaje	FC	N/E
<b>(Liu, y otros, 2020)</b>	Formal	Estudiantes	Recursos de aprendizaje	ML	N/E
<b>(Hidayat, Suwawi, &amp; Laksitowening, 2020)</b>	Informal	Estudiantes	Aprendizaje en línea	FC	N/E
<b>(Tan, Chang, Liu, &amp; Zhao, 2020)</b>	Informal	Estudiantes	Cursos	ML	N/E
<b>(Wang, Xie, Au, Zou, &amp; Wang, 2020)</b>	Informal	Estudiantes	Cursos	FC	N/E

<b>(Li &amp; Ye, 2020)</b>	Formal	Estudiantes	Recursos de aprendizaje	FC	Web
<b>(Liao, y otros, 2020)</b>	Informal	Estudiantes	Aprendizaje en línea	FC	N/E
<b>(Isma'il, Haruna, Aliyu, Abdulmumin, &amp; Adamu, 2020)</b>	Formal	Estudiantes	Cursos	ML	N/E
<b>(Wan, y otros, 2020)</b>	Formal/ Informal	Estudiantes	Recursos de aprendizaje	FC	Móvil
<b>(Niyigena &amp; Jiang, 2020)</b>	Informal	Estudiantes	Recursos de aprendizaje	FH	N/E
<b>(Park, 2020)</b>	Formal	Estudiantes	Asesoría académica	FC	N/E
<b>(Chen, Yu, Liu, Yuan, &amp; Wang, 2020)</b>	Informal	Estudiantes	Cursos	DL	N/E
<b>(Rismanto, Rachmad, &amp; Agusta, 2020)</b>	Formal	Estudiantes	Asesoría académica	TF-IDF	Web
<b>(Pandey, Lan, Karypis, &amp; Srivastava, 2020)</b>	Informal	Estudiantes	Cursos	ML	N/E
<b>(Sebbaq, el Faddouli, &amp; Bennani, 2020)</b>	Informal	Profesores	Cursos	FH	Web
<b>(Suganya, Mariappan, Dubey, Drolia, &amp; Srihari, 2020)</b>	Formal	Estudiantes	Aprendizaje en línea	ML	N/E
<b>(Morrow, Hurson, &amp; Sedigh Sarvestani, 2020)</b>	Formal	Estudiantes	Secuencia de estudios/programas de estudio	FCT	N/E
<b>(Tavakoli, Mol, &amp; Kismihók, 2020)</b>	Informal	Profesionales	Cursos	FH	N/E
<b>(Wang, y otros, 2020)</b>	Informal	Empleados	Cursos	FC	N/E
<b>(Wu, Wang, Liu, Yin, &amp; Zheng, 2020)</b>	Formal	N/E	Cursos	FC	N/E
<b>(Pan, Li, &amp; Lu, 2020)</b>	Formal	Estudiantes	Cursos	ML	N/E

FC: Filtrado colaborativo  
 ML: Machine Learning  
 FH: Filtrado Híbrido  
 DL: Deep Learning  
 FCT: Filtrado basado en contexto  
 TF-UDF: Term frequency – Inverse document frequency  
 FBC: Filtrado basado en contenido  
 FBK: Filtrado basado en conocimiento  
 N/E: No especifica

Al analizar los artículos, de acuerdo con el tipo de educación que cubren los SR (Tabla 2.3), fue posible identificar que las recomendaciones están principalmente orientadas a la educación formal (59,18%), en particular a los estudios universitarios, el 33,67% se centraban en la educación no formal, y un 7,14% tomaban en cuenta ambas.

En la mayoría de los casos, no se especifica el área de conocimiento a la que están dirigidas las recomendaciones, solo un 5% indican que están relacionados con el campo de la arquitectura y la ingeniería (Gulzar, 2018), (Aguilar, Valdiviezo-Díaz, & Riofrio, 2016), (Yan, Hara, Kazuma, & He, 2017), (Pandey, Lan, Karypis, & Srivastava, 2020) y (Tavakoli, Mol, & Kismihók, 2020), y el trabajo presentado por (Morsomme & Alferez, 2019) está orientado al área de las artes.

Tabla 2.3 Porcentaje de artículos según el tipo de educación

<b>Tipos de Educación</b>	<b>Número de artículos</b>	<b>% de artículos</b>
Formal	58	59,18%
No formal	33	33,67%
Formal/No formal	7	7,14%

Al analizar a qué tipo de usuario están orientados los SR, se determinó que el 91% está dirigido a los estudiantes (Tabla 2.4). La mayoría de los sistemas hacen sugerencias de cursos y recursos educativos, donde solo los trabajos presentados por (Sergis & Sampson, 2015) y (Sebbaq, el Faddouli, & Bennani, 2020) se centra en los profesores, recomendando recursos para la práctica docente, debido a la creciente disponibilidad de dichos recursos en línea y la dificultad resultante para localizarlos. El SR propuesto por (Dang, 2018) se centra tanto en los estudiantes como en los profesores, donde el estudiante recibe sugerencias de material de apoyo para estudiar, mientras que los profesores reciben perfiles de estudiantes para su revisión.

Tabla 2.4 Porcentaje de artículos según el tipo de usuario

<b>Tipo de usuario</b>	<b>Número de artículos</b>	<b>% de artículos</b>
Estudiantes	89	91%

Académicos	2	2%
Profesores	2	2%
Estudiantes/Profesores	1	1%
Profesionales	3	3%
No especifica	1	1%

En cuanto al de tipo de sugerencias que realizan, se pudo observar una variedad de elementos objeto de recomendación (Tabla 2.5). Se observó que los trabajos en su mayoría están orientados a ofrecer recursos de aprendizajes (37,75%), cursos (33,67%), lo que representa un porcentaje de 71,4% le siguen cinco trabajos que recomiendan planes de estudio; 5 trabajos recomiendan cursos electivos de grado; y cuatro trabajos se centran en el asesoramiento académico. Destaca el trabajo propuesto por (Patel, Kakuste, & Eirinaki, 2017) que recomienda trayectorias profesionales para estudiantes y jóvenes profesionales basados en descripciones de puestos de trabajo y el perfil de los usuarios.

Los trabajos presentados por (Bauman & Tuzhilin, 2018), (Brigui-Chtioui, Caillou, & Negre, 2017) y (Dang, 2018) realizan recomendaciones de recursos de aprendizajes para cubrir los vacíos de conocimiento, se basan en el diagnóstico y progreso del estudiante en los diferentes contenidos de los cursos que realiza. El artículo (Zhao, Wang, Wang, Zhou, & Jiang, 2018), propone recomendar cursos en plataformas MOOC (Massive Open Online Courses), teniendo en cuenta la existencia de más de un proveedor en el proceso de recomendación. También utilizará el conocimiento en estos ecosistemas MOOCs para mejorar la recomendación de cursos a través de estas plataformas.

Tabla 2.5 Porcentaje de artículos según el tipo de elementos que recomienda

Elemento para recomendar	Número de artículos	% de artículos
Recursos de aprendizaje	37	37,75%
Cursos	33	33,67%
Cursos electivos	5	5,1%

Secuencia de estudio	5	5,1%
Asesoramiento académico	4	4,1%
Aprendizaje en línea	3	3,06%
Cursos de postgrado	2	2,04%
Programas educativos	2	2,04%
Exámenes Multiobjetivo	1	1,02%
Artículos	1	1,02%
Recursos Pedagógicos	1	1,02%
Recursos Prácticas docentes	1	1,02%
Profesiones	1	1,02%
Problemas de programación	1	1,02%
Universidades	1	1,02%

Los SR presentados por (Bhumichitr, Channarukul, Saejiem, Jiamthapthaksin, & Nongpong, 2017), (Dwivedi & Roshni, 2017) y (Kamila, Subastian, & Rosmasari, 2019), recomiendan cursos electivos basados en los récords académicos de los estudiantes, en donde, (Bhumichitr, Channarukul, Saejiem, Jiamthapthaksin, & Nongpong, 2017) indica que de esta manera se minimiza el tiempo de búsqueda de cursos, a la vez que permite a la universidad planificar las secciones que puede ofrecer de una asignatura.

Por otra parte, resaltan los trabajos presentados por (Bañeres & Conesa, 2016), (Tavakoli, Mol, & Kismihók, 2020) y (Wang, y otros, 2020) que proponen SR para ayudar a los profesionales a identificar aquellas habilidades que deben desarrollar para su actual puesto de trabajo o las que pueden adquirir por ser las más demandadas en mercado laboral, y así, poder planificar su formación a través de cursos de LL.

Cuando se analizan los artículos para indagar sobre el enfoque utilizado en el desarrollo de SR en el campo de la educación, se encuentra que utilizan principalmente filtrado colaborativo; los segundos enfoques más utilizados son el híbrido y ML, seguido de los SR que utilizan filtrado basado en conocimiento; destacándose el uso de las

técnicas de ML en los últimos dos años. La Tabla 2.6 muestra en detalle el número de artículos según el enfoque de desarrollo.

De los artículos que utilizan ML, la mayoría utiliza ML supervisado. Las técnicas más destacadas incluyen redes neuronales: (Jiang, Pardos, & Wei, 2019) (Jiang & Pardos, 2019) (Mrhar & Abik, 2019), (Zhang, Liu, & Wang, 2020), (Gordillo, López-Fernández, & Verbert, 2020), (Tan, Chang, Liu, & Zhao, 2020), (Wang, Xie, Au, Zou, & Wang, 2020) (Pandey, Lan, Karypis, & Srivastava, 2020) y (Pan, Li, & Lu, 2020).

Por otra parte, los artículos presentados por (Mrhar & Abik, 2019) y (Zhang, Liu, & Wang, 2020) utilizan Deep Learning; (Britto, Prabhu, Gawali, & Jadhav, 2019) y (Suganya, Mariappan, Dubey, Drolia, & Srihari, 2020) utilizan *Support Vector Machines*; (Vaidhehi & Suchithra, 2019) y (Mondal, Patra, Mishra, & Patra, 2020) utilizan *k-nearest neighbors* (10,5%), y por último, (Chen, y otros, 2020) y (Slim, y otros, 2019) utilizan K-mean (10,5%).

El trabajo presentado por (Cheng, Zhang, & Shi, 2018), propone un SR basado en conocimiento para sugerir cursos, para generar una ruta de aprendizaje personalizada, de acuerdo con el dominio de los conocimientos de los alumnos, y teniendo en cuenta las preferencias del usuario y el orden de aprendizaje diseñado por expertos como base.

El trabajo presentado por (Bakhshinategh, Spanakis, Zaïane, & ElAtia, 2017) utiliza un filtrado colaborativo para sugerir cursos que fueron tomados por otros estudiantes, y calificados de manera similar con el objetivo de mejorar los atributos de graduación, que definen las habilidades que los estudiantes deben desarrollar durante su estancia en la universidad.

Tabla 2.6 Porcentaje de artículos según el enfoque de desarrollo

Enfoque de desarrollo	Número de artículos	% de artículos
Filtrado colaborativo	32	32,65%
Filtrado Híbrido	20	20,41%
Machine Learning	20	20,41%

Filtrado basado en conocimiento	16	16,33%
Filtrado basado en contenido	2	2,04%
Filtrado basado en contexto	2	2,04%
Grafos	2	2,04%
Lógica Difusa	1	1,02%
Deep Learning	1	1,02%
Term frequency–Inverse document frequency	1	1,02%
No especifica	1	1,02%

Para (Meryem, Douzi, & Chantit, 2016), los SR colaborativos no requieren información del contenido de los artículos, éstos aprovechan el conocimiento sobre el comportamiento y valoración de personas similares. En este contexto y dada la popularidad de las plataformas de cursos masivos, que se han convertido en una importante fuente de recursos de aprendizajes, (Li & Li, 2017) y (Hinz & Pimenta, 2018), proponen SR basados en el comportamiento de los usuarios y el historial de estos en las plataformas MOOCs para recomendar cursos a usuarios similares.

Los algoritmos de filtrado basado en conocimiento son útiles cuando los usuarios quieren especificar explícitamente sus requisitos, y cuando es difícil obtener valoraciones para un tipo específico de artículos, o en los ámbitos en donde las valoraciones pueden ser sensibles al tiempo (Aggarwal, 2016). Teniendo esto en mente, (Bañeres & Conesa, 2016) propone un SR en donde a partir de ofertas de empleo se definen los requerimientos de formación para las habilidades que necesita un usuario para optar al empleo y sugerir los programas académicos y las asignaturas que le permitirán alcanzarlas.

El trabajo de (Meryem, Douzi, & Chantit, 2016), propone combinar técnicas de filtrado colaborativo y basado en contenido, ponderadas, con el fin de proporcionar a los estudiantes los itinerarios académicos más apropiados, en donde se da más ponderación al filtrado basado en contenido porque ofrece resultados objetivos en comparación con el filtrado colaborativo, que se basa en la experiencia de las personas.

El uso de técnicas individuales tiene limitaciones, como puede ser el caso de los filtrados colaborativos en el que los artículos nuevos son poco valorados, conocido como arranque en frío, y por lo tanto raramente se sugieren, o problemas de escasez de datos (Montuschi, Lamberti, Gatteschi, & Demartini, 2015) . Estas dificultades se superan con los SR híbridos, al combinar el uso de diferentes técnicas.

Con el objetivo de mejorar la eficacia del e-learning, (Gulzar, 2018) propone un SR híbrido para facilitar el acceso a la información y proporcionar personalización a los alumnos. El sistema selecciona cursos para el aprendizaje no formal y basado en proyectos, que se ajusten a los usuarios antes de matricularse en un programa de investigación.

Por su parte, el trabajo presentado por (Rodríguez, Ovalle, & Duque, 2015), propone un SR híbrido centrado en el estudiante, en el que combinan tres enfoques: basado en contenido, colaborativo y basado en conocimiento para la recomendación de recursos de aprendizajes.

Algunas propuestas de SR basan su desarrollo en ML para hacer inferencias sobre cómo las acciones pasadas se corresponden con los resultados futuros. (Jiang & Pardos, 2019) aplica esta técnica para predecir el rendimiento futuro con base al registro académico de los estudiantes para recomendar cursos para la preparación de un curso objetivo y preparación de exámenes. (Yanes, Mostafa, Exx, & Almuayqil, 2020) utiliza ML para predecir las acciones adecuadas para proponer acciones para mejorar la calidad de los cursos, basándose en las especificaciones de los cursos, los registros académicos y las evaluaciones de los resultados de aprendizaje de los cursos.

(Samin & Azim, 2019) propone un SR para diferentes usuarios en el ámbito académico. Utiliza ML para extraer información sobre área de investigación de interés de los profesores a partir de publicaciones y currículo, combinándolas con su formación para asignarle cursos y trabajo de investigación a supervisar.

Así mismo, (Tavakoli, Mol, & Kismihók, 2020) utiliza ML para descomponer los puestos de trabajo en componentes de competencias significativas para identificar cuáles

son las más demandadas en el mercado laboral y proponer una ruta de aprendizaje a los usuarios que les permita desarrollarlas.

Los sistemas de recomendación son capaces de estimar el interés del usuario por un recurso o artículo utilizando información sobre el perfil de los usuarios, información de usuarios similares y propiedades de los recursos. Tomando esto en cuenta, en función del objetivo y dominio de los SR se debe definir la estrategia de recomendación para lograr precisión y relevancia en las sugerencias. En este contexto, se deben tomar en cuenta diferentes elementos para la construcción del perfil de los usuarios, que pueden afectar la orientación de las recomendaciones, además del pasado e historia de los estudiantes, como son: género, edad, personalidad, calificaciones, conocimiento previo, habilidades y competencias, factor social y factor socioeconómico (Meryem, Douzi, & Chantit, 2016).

Para orientar a los estudiantes a seleccionar las mejores universidades y trayectorias profesionales, (Meryem, Douzi, & Chantit, 2016), (Bokde, Girase, & Mukhopadhyay, 2015) y (Baskota & Ng, 2018) proponen un SR basado en el perfil de los estudiantes con base a las calificaciones de las universidades realizadas por usuarios similares, en donde (Baskota & Ng, 2018) resalta el hecho de que esta selección permite seleccionar la universidad con miras a conseguir empleos con perspectivas a largo plazo.

(Xu, Xing, & van der Schaar, 2016), (Morrow, Hurson, & Sedigh Sarvestani, 2020) y (Zhu, y otros, 2018) proponen unos SR que tienen en cuenta el perfil del estudiante y los requisitos curriculares de programas de estudio para generar una propuesta de secuencia de cursos, con el objetivo de reducir atributos como el coste y el tiempo de obtención de un título.

Se pudo observar que la mayoría de los SR propuestos tienen en cuenta las preferencias del usuario y basan las recomendaciones en usuarios similares. (Ganeshan & Li, 2015) y (Bhumichitr, Channarukul, Saejiem, Jiamthapthaksin, & Nongpong, 2017) proponen agregar más parámetros al perfil de usuario, como estilos de aprendizaje. En este sentido, se encontraron algunas investigaciones en el dominio de la educación formal que toman en cuenta el estilo de aprendizaje, (Kurilovas, Kurilova, Kurilova, & Melesko, 2016) y (Bourkougou & El Bachari, 2016) proponen el uso del modelo de Estilos de

Aprendizaje de Felder y Silverman, en donde se representa la preferencia del alumno en cuatro dimensiones: procesamiento de la información, percepción de la información, recepción de la información y comprensión de la información. Esta información se identifica a través de un cuestionario que debe completar el alumno, para utilizarla en el perfilado de usuario.

Teniendo en cuenta la posibilidad de representar el perfil de los usuarios en función del estilo de aprendizaje, se realizó una búsqueda de artículos que tomaran en cuenta esto, encontrando que algunos hacen uso de una taxonomía para la representación de diferentes perfiles y recursos de aprendizaje. En la Tabla 2.7 se puede observar el objetivo de estos estudios y la estructura que utilizan.

El uso de la taxonomía por (Bouzeghoub & D., 2006), (Demchenko, A., Bernstein, & Laat, 2014), (Bigotte de Almedia & A, 2015) y (Yildirim & Baur, 2015) tuvo como objetivo definir un perfil de usuario basado en el estilo de aprendizaje para proponer la estructura que debe utilizarse para diseñar herramientas que hagan que la educación sea más efectiva. El estudio de (Wiktorski, Demchenko, & Belloum, 2017) definió una taxonomía para guiar el trabajo en el aula de los asistentes educativos, y los de (Moreira, Ferreira, Collazos, & Cano, 2017) y (Thompson, y otros, 2020) propusieron el diseño de una taxonomía para representar la información contenida en los cursos y en los perfiles de usuario.

Tabla 2.7 Uso de taxonomía para la representación de perfiles y recursos

Artículos	Propuesta	Estructura
(Bouzeghoub & D., 2006)	Taxonomía basada en conocimiento que constituye un perfil de usuario.	El perfil constaba de los siguientes elementos: datos personales, dominio de interés, calidad de los datos, entrega de datos, seguridad e historial de interacciones usuario-sistema.
(Demchenko, A., Bernstein, & Laat, 2014)	Currículos sobre tecnologías de computación en la nube teniendo en cuenta la taxonomía de Bloom	Se propuso la siguiente estructura para los cursos: Título / Duración / Horas (sesiones de laboratorio / teóricas) / Requisitos / Objetivo / Temas / Actividades (Tareas: discusiones / presentaciones) / Evaluación (Exámenes / trabajo de

		autoaprendizaje / Proyecto Final)
(Bigotte de Almedia & A, 2015)	Definición de trayectorias educativas de acuerdo con el perfil de resultados de aprendizaje de cada alumno.	El perfil se definió en base al estilo de aprendizaje del alumno, aplicando la taxonomía de Bloom.
(Yildirim & Baur, 2015)	Taxonomía de aprendizaje para clasificar las metas del curso, los objetivos de aprendizaje de los estudiantes y las herramientas de evaluación.	Taxonomía del aprendizaje/ taxonomía de los objetivos educativos/Descripción del curso / metas del curso /objetivos de aprendizaje de los estudiantes y herramientas de evaluación
(Wiktorski, Demchenko, & Belloum, 2017)	Brindar orientación a universidades y organizaciones de capacitación en la construcción de programas de Ciencia de Datos basados en el perfil del estudiante de acuerdo con el estilo de aprendizaje del estudiante.	Los cursos se estructuraron de acuerdo con los grupos de áreas de conocimiento, y las competencias se definieron por área de conocimiento. Estas competencias se clasificaron por niveles de dominio, y cada nivel de competencia tiene un nivel de evaluación.
(Moreira, Ferreira, Collazos, & Cano, 2017)	Metodología para construir un perfil de estudiante basado en el cuestionario VARK para definir los materiales y herramientas educativas más adecuadas	El perfil del estudiante consistió en habilidades informáticas y estilos de aprendizaje.
(Thompson, y otros, 2020)	Taxonomía para ayudar a categorizar las acciones de los asistentes de aprendizaje.	Esta taxonomía se clasificará de acuerdo con las actividades planificadas y estas actividades se clasifican a su vez en acciones dependientes.

Otra de las características identificadas en el análisis de los artículos es la heterogeneidad de los datos en el dominio de la educación, donde (Ibrahim, Yang, Ndzi, Yang, & Almaliki, 2018) indica que la integración de datos de múltiples fuentes heterogéneas ayuda al sistema a mejorar las recomendaciones. Así mismo, (Patel, Kakuste, & Eirinaki, 2017), (Fazeli, Rajabi, Lezcano, Drachsler, & Sloep, 2016) y (Bozyiğit, Bozyiğit, Kılınç, & Nasiboglu, 2018) sugieren utilizar un mayor volumen de datos para mejorar el rendimiento,

Ante la características de los datos, (Samin & Azim, 2019), (Ibrahim, Yang, Ndzi, Yang, & Almaliki, 2018), (Tavakoli, Mol, & Kismihók, 2020), (Vaidhehi & Suchithra,

2019) y (Sebbaq, el Faddouli, & Bennani, 2020), indican que el uso de la ontología para la representación de la información del usuario ayuda a resolver el problema del arranque en frío.

(Kuznetsov, Kordík, Tomás Rehorek, & Kroha, 2016) utiliza la ontología para representar el conocimiento de un dominio educativo, para generar perfiles de usuarios, y mejorar el rendimiento del SR de proyectos para estudiantes. Combinan técnicas de ML para clasificar las ontologías adaptadas al perfil del usuario. Los datos educativos se transforman en perfiles de estudiantes, lo que ayuda a evitar los problemas de arranque en frío.

En el trabajo de (Aissaoui & Oughdir, 2020) se propone un SR para proporcionar a cada alumno el recurso de aprendizaje adecuado para e-learning basado en ontologías, las cuales se utilizan para representar el conocimiento de los alumnos y los materiales de aprendizaje. El perfil del alumno contiene información demográfica, que se extraen mediante un cuestionario y el estilo de aprendizaje, que se puede captar mediante el análisis del comportamiento del alumno al interactuar con el sistema utilizando técnicas de ML.

Dada la tendencia del uso de ontologías para la representación de la información en los SR, se realiza una revisión de artículos en esta línea, donde se pueden encontrar algunos estudios que sugieren combinar diferentes técnicas de filtrado con el uso de ontologías, con el fin de crear un modelo que pueda describir la relación entre un área de dominio y los usuarios que pertenecen a ella, para recomendaciones más precisas (Tabla 2.8).

En esta revisión destaca la importancia de la representación y el manejo de la información. El estudio de (Valentin, Emrich, . Lahann, Werth, & Loos, 2015) indicó que la cantidad de información manejada es un factor clave para que las recomendaciones tengan éxito.

Asimismo, (Koh & Chew, 2015), (Tarus, Niu, & Yousif, 2017), (Harrathi, Touzani, & & Braham, 2017) y (Gugnani, Kasireddy, & Ponnalagu, 2018), señalaron que las

ontologías permiten inferir información de interés para los usuarios, que puede ser utilizada para enriquecer las recomendaciones. (Cerón-Figueroa, y otros, 2017) concluyeron que aprovechan la función de similitud para encontrar información adicional.

Tabla 2.8 Uso de Ontología en Sistemas de Recomendación

Ref.	Objetivo	Técnicas	Ontología	Evaluación
(Koh & Chew, 2015)	Recomendación de empleos	Ontología Algoritmos de similitud	Puestos de trabajo Tareas	No indica
(Valentin, Emrich, Lahann, Werth, & Loos, 2015)	Recomendación de RRSS para el desarrollo de habilidades en RRSS	Ontología Filtrado por contenido y contexto	RRSS	Calidad de resultados e integración de funcionalidades por parte de usuarios
(Tarus, Niu, & Yousif, 2017)	Recomendación de recursos educativos	Ontología Sequential pattern mining (SPM)	Alumnos Recursos educativos	Comparativa de satisfacción de usuario entre el sistema propuesto, y otros SR, sin ontología y sin SPM
(Harrathi, Touzani, & Braham, 2017)	Recomendación de actividades de aprendizaje masivo	Ontología Filtrado basado en conocimiento	Alumnos Actividades de Aprendizaje	No indica
(Cerón-Figueroa, y otros, 2017)	Recomendación de material de aprendizajes	Ontología Canopy-K means Filtrado Colaborativo	Material Educativo	Comparativa con otros algoritmos
(Balachander & Moh, 2018)	Recomendación de empleo	Ontología	Habilidades	Cálculo de puntuaciones de similitud de la Información Mutua Puntual (PMI)
(Gugnani, Kasireddy, & Ponnalagu, 2018)	Recomendación de trayectoria de empleo	Ontología	Habilidades	Cálculo de parámetros relacionados a

				recuperación de la información
(Obeid, Lahoud, El Khoury, & Champin, 2018)	Recomendar la carrera y la universidad apropiadas para cada uno	Ontología K-means	Institución de educación superior. Alumno Empleo	Propuesta
(Aissaoui & Oughdir, 2020)	Recomendación de recursos de aprendizaje en e-learning	Ontología Filtrado baso en contenido	Alumno Contenido de aprendizaje	Propuesta

Los estudios (Harrathi, Touzani, & & Braham, 2017), (Cerón-Figueroa, y otros, 2017), (Balachander & Moh, 2018) y (Gugnani, Kasireddy, & Ponnalagu, 2018) indicaron que el uso de la ontología mejoró el arranque en frío y los problemas de dispersión de datos, que limitan el rendimiento de los RS. También señalaron que la personalización del perfil del alumno mediante el uso de la ontología significa que las recomendaciones se pueden adaptar mejor a las preferencias de los alumnos destinatarios.

(Aissaoui & Oughdir, 2020) concluyó que la representación del conocimiento es un paso vital en la construcción de una recomendación basada en el conocimiento para el aprendizaje automático; por lo tanto, es esencial representar la información en una estructura legible por máquina. (Obeid, Lahoud, El Khoury, & Champin, 2018) también sugirió que el uso de ontologías facilita el análisis, el razonamiento, el intercambio y la reutilización del conocimiento.

Otro hallazgo en la revisión de la literatura está relacionado con las fuentes de información. La mayoría de los artículos analizados buscan la información en una fuente única, y algunos estudios sugerían que debe explorarse la búsqueda de información en múltiples fuentes para poder asegurar una amplia variedad de elementos y ofrecer un mejor servicio de recomendación. Las RRSS constituyen un medio potencial para buscar información a través de sistemas de recomendación. (Zheng, y otros, 2015) indica que el uso de las RRSS podría mejorar el problema de la dispersión de datos.

Para (Pereira, Campos, Ströele, David, & Braga, 2018) dada la cantidad de los recursos de aprendizaje que se pueden encontrar en diferentes repositorios, la elección

adecuada de éstos se convierte en un reto, en donde es importante seleccionar la información para poder definir el perfil de usuario e identificar sus intereses. El crecimiento y alcance de las RRSS producen un gran volumen de datos que permite identificar diferentes informaciones sobre el perfil, los intereses, preferencias, estilo y comportamiento a partir de la interacción espontánea. Proponen un SR que se apoya en técnicas de extracción de información y tecnologías de la web semántica para la extracción, enriquecimiento y definición del perfil de usuario y el contexto educativo, a partir de la red social Facebook, para recomendar recursos educativos.

El trabajo presentado por (Huang, y otros, 2018), propone un SR híbrido de cursos basado en RRSS académicas, que combina un algoritmo de reglas de asociación y un algoritmo de multisimilaridad de múltiples fuentes de información, para recomendar cursos según las relaciones potenciales entre los cursos y los intereses implícitos de los usuarios. La propuesta promete reducir el problema del arranque en frío y proporcionar una mayor precisión.

El aprendizaje a distancia se enfrenta a muchos desafíos debido a las bajas tasas de abandono y a la rápida evolución del aprendizaje con ayuda de la tecnología. Conocer al usuario y sus necesidades, mejora el desarrollo de cursos en línea. La interacción de las personas en RRSS genera grandes cantidades de datos, y se presentan como oportunidades a los proveedores de MOOCs. El uso de los datos de los usuarios de las RRSS para enriquecer el perfil y conocer mejor las necesidades y preferencias de los alumnos se convierten en un recurso de datos externo para los SR (Assami, Daoudi, & Ajhoun, 2019).

Finalmente, del análisis de los artículos, se puede destacar que, de acuerdo con el tipo de educación, los SR cubren principalmente la educación formal, especialmente orientada a los estudiantes.

Las técnicas de desarrollo más utilizadas son el filtrado colaborativo, seguido de los RS que combinan diferentes técnicas. Revisiones sistemáticas similares, como la presentada por (Rivera, Tapia-Leon, & Luján-Mora, 2018), concuerdan con este resultado, encontrando una brecha en el uso de técnicas inteligentes. Se puede ver en

nuestra revisión cómo se ha cubierto esta área potencial desde 2019, donde se presentan propuestas de SR con ML. Así mismo, se puede observar el uso incipiente de la ontología para la representación de la información y la construcción de perfiles de usuario. Por otra parte, otro elemento que ha resaltado en la revisión es la incorporación de las RRSS, para construir un perfil más completo más allá de las preferencias del usuario y de los usuarios similares.

La revisión del estado del arte permitió identificar, la necesidad de profundizar en el uso de taxonomía y ontología para la representación de la información, como una solución a los diferentes retos de los SR y así mejorar el rendimiento de estos; así mismo, la mayoría de los SR están orientados a la educación formal, siendo el área de LL poco abordada, detectando una oportunidad para explorar en el campo de los SR para la educación no formal con miras a profesionales como usuario objetivo para potenciar el desarrollo de sus. Estos hallazgos permitieron orientar nuestra propuesta para cubrir las lagunas detectadas y profundizar en esta línea de investigación

## **2.5 Resumen del Capítulo**

El crecimiento en los últimos años de los servicios en internet, la sobrecarga de información se ha convertido en un problema para los usuarios en cuanto al tiempo empleado para su búsqueda y a la cantidad de información que esta arroja. Es determinante a la hora de tomar decisiones poder contar con la información pertinente, de utilidad y de manera oportuna. Como consecuencia a este problema han surgido los SR para ayudar a los usuarios a encontrar aquello que realmente es relevante para sus necesidades. A medida que estos evolucionan se han combinado con otras técnicas para mejorar sus resultados. En el presente capítulo se presentó una revisión de los SR, el uso de técnicas de ML y ontología y se mostró una revisión de la literatura relacionada con lo SR en el dominio de la educación, que permitió detectar algunos *gaps* en este campo para abordarlos en esta investigación.



## Capítulo 3

### Fuentes de Información y datos

En el ámbito del aprendizaje, y en los servicios de recomendación personalizada, se deben tener en cuenta diferentes dimensiones del usuario, como datos personales, intereses, niveles de conocimiento, contexto, entre otras, para responder de forma adecuada a sus necesidades (Bedjou, Azouaou, & Berkani, 2014). Con el propósito de conocer algunas características de los usuarios de cursos de LL, en la primera etapa de la investigación se realizó una encuesta entre profesionales para recopilar información sobre sus intereses en la formación, así como los patrones de búsqueda de cursos en la web.

La encuesta fue realizada con Google Forms, poniéndola disponible por Internet. La encuesta fue respondida por 119 profesionales, donde, el 73,1% son de género masculino, la edad predominante esta entre el rango de 20 a 30 años, de los cuales el 88% tienen formación universitaria y las procedencias más destacadas son Venezuela (44%), España (13%), Chile (12) y Colombia (9%). Para el propósito de nuestra investigación, se observa que el 30% de los encuestados toman cursos de LL para mejorar el desempeño en su sector laboral.

En cuanto al objetivo de los participantes a la hora de tomar cursos de LL, 59,4% lo hacen para complementar su formación, seguido por el 30% para obtener una alta especialización.

Con estos resultados, se evaluaron algunas RRSS a fin de determinar cuáles eran las más convenientes a la hora de crear perfiles de usuarios para el propósito de la investigación. Evaluadas LinkedIn, Facebook y Twitter, se concluyó que LinkedIn era la más prometedora ya que contenía datos académicos y profesionales de los usuarios, de los cuales se pueden procesar las habilidades, lo que en las otras RRSS no se identificaba. Esto significaba que se podían crear perfiles basados en su sector laboral y áreas de conocimiento, que resultaban de las habilidades adquiridas por los usuarios a través de la formación o desarrolladas en sus puestos de trabajo.

### **3.1 Tamaño de la muestra y método de adquisición de datos**

Las pruebas, evaluación y el posterior análisis del SR propuesto requieren de un conjunto de datos de profesionales y cursos de formación continua.

El tamaño y diversidad del conjunto de datos de perfiles de usuario fue limitado por la cantidad de registros que se pudieron extraer de LinkedIn. Para la recolección de los datos se utilizó un Bot, recuperándose 609 perfiles profesionales en formato csv. Una vez filtrados por el área de dominio se seleccionaron 128 con los que se construyeron los de perfiles de usuario. El criterio de selección se basó en que los registros recuperados tuvieran la información completa, y para efectos de la evaluación de la propuesta y facilitar la construcción de las taxonomías y ontologías se delimitó al dominio específico del área de desarrollo de software.

La información contenida en el archivo csv está clasificada en tres categorías: personal, trayectoria profesional y trayectoria académica. Para la construcción de los perfiles de usuarios, se tomó en cuenta toda la información relacionada con la trayectoria profesional y académica por contener información relevante para determinar el área de conocimiento, las habilidades y el sector laboral que son tomados en cuenta a la hora de

realizar la recomendación de cursos. De la información personal se descartaron los datos de identificación del usuario, para mantener la privacidad de éstos, así como otros campos que no aportaban a los elementos del perfil definido como fueron: el correo electrónico, el *url* del perfil de LinkedIn, Nombre, dirección, cumpleaños, perfil de Twitter, los campos de información de teléfono, mensajería instantánea, direcciones de sitio web, número de seguidores, contactos y fecha de alta en LinkedIn. En la Tabla 3.1, Tabla 3.2 y Tabla 3.3, se muestra, según el tipo de información, el formato de la descarga del archivo de los datos de los perfiles y cuáles de los datos fueron incluidos o excluidos, para el proceso de construcción de perfiles de usuarios.

Tabla 3.1 Formato del archivo csv LinkedIn con la información personal del usuario.

Elaboración propia

<b>Identificador</b>	<b>Descripción</b>	<b>Tipo de dato</b>	<b>Inclusión/Exclusión</b>
Id	Identificador del usuario	Texto	Incluido
Email	Correo electrónico del usuario	Texto	Excluido
Profile url	url del perfil de LinkedIn	Texto	Excluido
First name	Nombre de usuario	Texto	Excluido
Last name	Apellido de Usuario	Texto	Excluido
Title	Cargo actual	Texto	Incluido
Location	Ubicación actual	Texto	Incluido
Address	Dirección	Texto	Excluido
Birthday	Fecha de Cumpleaños	Fecha	Excluido
Summary	Resumen del perfil de usuario	Lenguaje natural	Incluido
Twitter	Perfil de Twitter	Texto	Excluido
Phone 1			
...	Teléfono	Numérico	Excluido
Phone 3			

Phone 1 type			
...	Tipo de teléfono	Texto	Excluido
Phone 2 type			
Messenger 1			
...	Mensajería instantánea	Texto	Excluido
Messenger 3			
Messenger 1 type			
...	Tipo de mensajería instantánea	Texto	Excluido
Messenger 3 type			
Website 1			
...	Dirección de sitio web del usuario	Texto	Excluido
Website 2			
Followers	Número de contactos en LinkedIn	Numérico	Excluido
Relationship	Grado de contacto	Numérico	Excluido
Connected at	Fecha de alta en LinkedIn	Fecha	Excluido
Mutual Count	Contactos en común	Texto	Excluido
Mutual 1			
...	Contactos	Texto	Excluido
Mutual 2			
<i>Endorsements</i>	Habilidades y el número de validaciones realizadas por otros contactos de esas habilidades	Texto	Incluido
Interests	Fuentes de noticias, Influencers, empresas, universidades y grupos a los que sigue un miembro de LinkedIn.	Texto	Excluido

Tabla 3.2 Formato del archivo csv de LinkedIn con la información profesional del usuario.

Elaboración propia

<b>Identificador</b>	<b>Descripción</b>	<b>Tipo de dato</b>	<b>Inclusión/Exclusión</b>
Industry	Tipo de Sector	Texto	Inclusión
Organization 1			
...	Nombre de la empresa	Texto	Exclusión
Organization 7			
Organization Title 1			
...	Cargo actual en la empresa	Texto	Inclusión
Organization Title 7			
Organization Start 1			
...	Fecha de inicio de la relación laboral en la empresa	Fecha	Inclusión
Organization Start 7			
Organization End 1			
...	Fecha de finalización de relación laboral en la empresa	Fecha	Inclusión
Organization End 7			
Organization Description 1			
...	Descripción del trabajo realizado en la empresa	Lenguaje natural	Inclusión
Organization Description 7			
Organization Location 1			
...	Ubicación de la empresa	Texto	Exclusión
Organization Location 7			

Organization LI URL 1			
...	Perfil de LinkedIn de la empresa	Texto	Exclusión
Organization LI URL 7			
Organization LI ID 1			
...	Identificador de la empresa	Texto	Exclusión
Organization LI ID 7			
Organization www 1			
...	Dirección web de la empresa	Texto	Exclusión
Organization www 7			

Tabla 3.3 Formato del archivo csv de LinkedIn con la información académica del usuario.

Elaboración propia

Identificador	Descripción	Tipo de dato	Inclusión/Exclusión
Education 1			
...	Nombre de Universidad/ Instituto de estudio	Texto	Exclusión
Education 3			
Education Degree 1			
...	Título de grado	Texto	Inclusión
Education Degree 3			
Education FOS 1			
...	Campo de estudio	Texto	Exclusión
Education FOS 3			
Education Grade 1			
...	Nivel de grado	Texto	Inclusión
Education Grade 3			

Education Start 1			
...	Fecha de inicio de los estudios	Fecha	Exclusión
Education Start 3			
Education End 1			
...	Fecha de finalización de los estudios	Fecha	Inclusión
Education End 3			
Education Description 1			
...	Descripción relacionada con el estudio.	Lenguaje natural	Exclusión
Education Description 3			

Para el catálogo de cursos que se utilizó en la etapa de recomendación, se recolectaron datos de diferentes páginas web de universidades e instituciones de cursos de formación continua, encontrándose con una amplia y variada oferta. La búsqueda se limitó al área de dominio seleccionada para la prueba, obteniéndose 100 cursos, los cuales se registraron en un archivo en formato csv, con los campos que se muestran en la siguiente tabla:

Tabla 3.4 Formato del archivo de cursos. Elaboración propia

Identificador	Descripción	Tipo de dato
Curso	Nombre del Curso	Texto
Descripción	Descripción del curso	Lenguaje Natural
Institución	Institución que dicta el curso	Texto
Url	Dirección web del curso	Texto
Localidad	Localidad donde se dicta el curso	Texto
Fecha	Fecha en la que se dicta el curso	Fecha
Modalidad	Tipo de modalidad del curso (Presencial/Virtual/Mixta)	Texto

Idioma	Idioma en la que se dicta el curso	Texto
Habilidades	Habilidades que se desarrollan en el curso	Lenguaje Natural
Requisitos	Requisitos del curso	Lenguaje Natural
Observaciones	Observaciones relacionadas al curso	Lenguaje natural

## 3.2 Características de los datos

Debido a la naturaleza de las diferentes fuentes de información disponible, los datos son no estructurados y en el mejor de los casos semiestructurados, pudiéndose observar las siguientes características:

- En las descripciones de las actividades laborales, los conocimientos y habilidades de las personas hay una gran cantidad de términos que, aunque sean diferentes hacen referencia a una misma cosa o actividad (sinónimos) o siendo iguales referencia a cosas o actividades diferentes (homónimos).
- Dada a la característica variable de los cursos de formación continua no existen una relación explícita entre perfiles profesionales y preferencias de cursos de formación, así mismo, las organizaciones que dictan los cursos suelen tener evaluaciones de los participantes, pero no sus perfiles. Por otra parte, los registros de profesionales, aunque algunas veces tienen información de cursos de formación continua, estos no muestran información detallada de los contenidos, conocimientos y habilidades que desarrollan, ni calificación que la persona hace de los mismos

## 3.3 Representación de Perfiles de usuarios

El sistema de recomendación hace uso de perfiles de usuarios, la información del perfil de usuario consta de los siguientes elementos:

- Información Demográfica. Está formada por su identificador, su URL en LinkedIn, su residencia, idiomas, información académica y años de experiencia.
- Sectores de desempeño laboral. Es un conjunto de códigos que representan el sector de trabajo, el cual viene dando por un código jerárquico definido por una taxonomía para sectores de desempeño laboral, su descripción y fechas.
- Conjuntos de Habilidades, viene dado por un conjunto de códigos de habilidades, donde cada habilidad se codifica con tres argumentos: un código jerárquico que se define por una taxonomía de áreas de conocimiento, un nivel de especialización y grado de actualización.

En la figura 3.1 se muestra la estructura para un perfil de usuario.

```

{
  "ID":25,
  "DemoInfo":{
    "URL_Linkedin":"https://www.linkedin.com/in/.../",
    "Location":"Bilbao-Spain",
    "Language":[
      "Code":"01-3-3-3",
      "Code":"02-3-3-3",
      "Code":"10-3-3-3"
    ],
    "AcadInfo":{
      "Level":4,
      "End":"2015"
    },
    "Experience":5
  }
  "JobSectors":[
    {
      "ID":"001-03-01:02-00",
      "Description":"Desarrollador Senior iOS"
      "Start":"20-may-2019",
      "End":"Actual"
    }
  ],
  "Skills":{
    {"KA":"20-22-00-2:00","SL":4,"UG":3},
    {"KA":"20-10-01-0:01","SL":3,"UG":3},
    {"KA":"20-10-02-0:01","SL":3,"UG":2},
    {"KA":"10-07-01-2:01","SL":4,"UG":3},
    {"KA":"10-07-06-2:00","SL":3,"UG":3}
  }
}

```

Figura 3.1. Estructura de Perfil de Usuario. Elaboración propia

La figura 3.2 muestra la codificación utilizada en la representación de perfiles de usuario y perfiles de curso, indicando el significado de las posiciones dentro de la jerarquía.

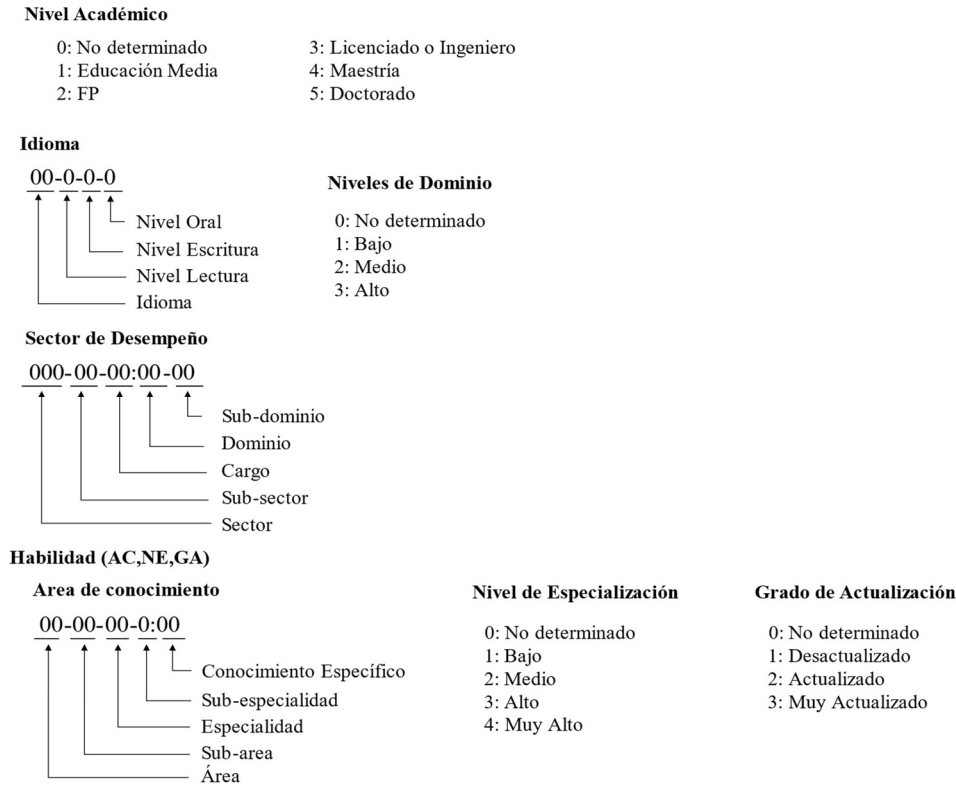


Figura 3.2 Codificaciones utilizadas en los Perfiles de Usuarios. Elaboración propia

### 3.4 Representación de Perfiles de Cursos

El SR mantiene un catálogo de cursos de LL, en donde la información contenida en el perfil es la siguiente:

- Información Demográfica. Está formada el nombre del curso y su URL de la página Web que tiene la información del curso, la localidad donde se imparte si la modalidad es presencial, su duración en horas, el idioma en el cual se imparte, una valoración que se le asigna en función de la valoración dada por personas que lo han tomado, la duración y la valoración.

- **Habilidades a Desarrollar.** Vienen dadas por un conjunto de habilidades, donde cada habilidad se codifica con dos argumentos: un código jerárquico que se define por una taxonomía de áreas de conocimiento y un nivel de especialización, es de notar que, para todas las áreas de conocimiento codificadas en los cursos, se utilizan la misma codificación utilizada para codificar las habilidades de usuario
- **Habilidades Requeridas.** Vienen dadas por un conjunto de códigos de áreas de conocimiento, donde cada una de ellas se codifica con un código jerárquico que se define por una taxonomía de áreas de conocimiento.
- **Habilidades Afines.** Vienen dadas por un conjunto de códigos de áreas de conocimiento donde las habilidades que desarrollan el curso pueden ser de interés o las complementan.

La figura 3.3 muestra la estructura para un perfil de curso

```
{
  "ID":12,
  "Info":{
    "Desc":"Desarrollo Avanzado de App en iOS",
    "URL":"https://.../ ",
    "Location":"Bilbao-Spain",
    "Language":["Code":"01","Code":"02","Code":"03"],
    "Duration":{"Cnt":40, "Unid":"Hrs"},
    "Value":4
  }
  "Skills2Develop":[
    {"KA":"20-10-01-0:01","SL":4}
  ],
  "Prerequisites":[
    "KA":"20-10-01-0:01",
    "KA":"10-06-01-9:02"
  ],
  "Related":[
    "KA":"20-06-01-2:00"
  ],
}
```

Figura 3.3 Estructura de Perfil de cursos. Elaboración propia

En la figura 3.4 se muestran, el significado de las posiciones dentro de la jerarquía, así como los diferentes códigos utilizados para los campos.

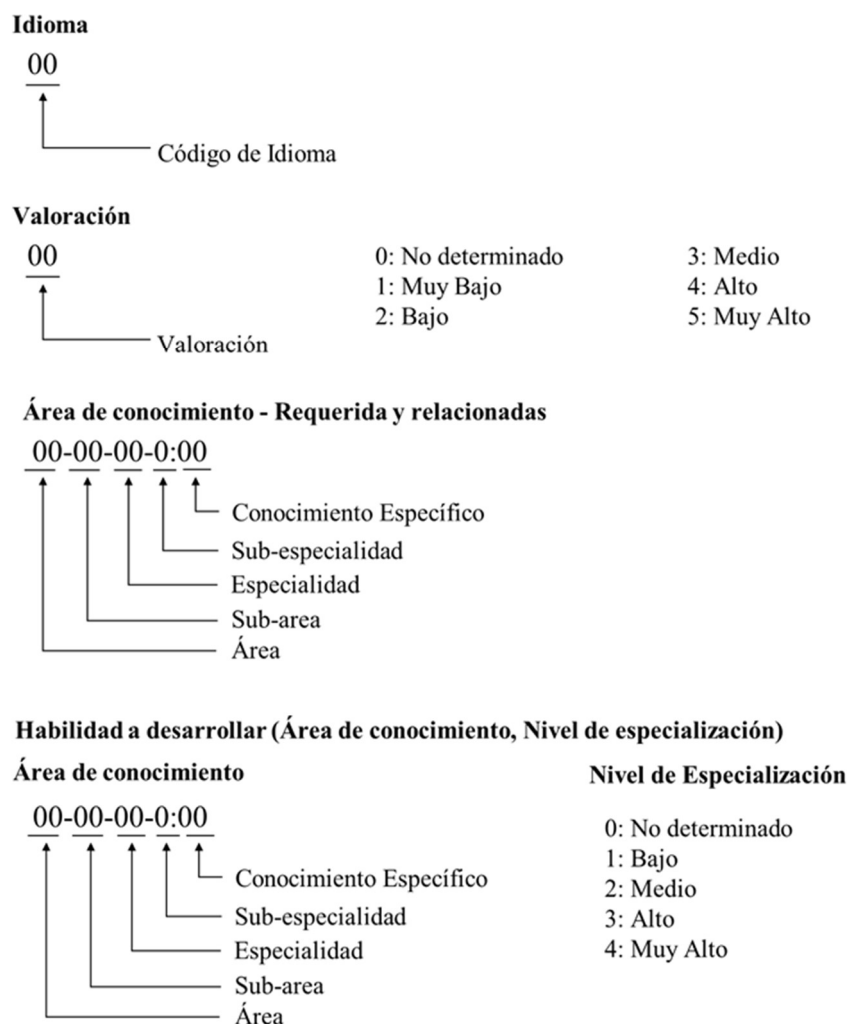


Figura 3.4 Codificaciones utilizadas en los Perfiles de Cursos. Elaboración propia

### 3.5 Construcción del Dataset

Una vez determinado la fuente de los datos y los criterios de selección, se procede a construir los perfiles de usuarios y cursos, para lo cual los procesos automáticos de perfilados hacen uso de transductores semánticos que empleando la ontología puede codificar, a partir de los registros de los usuarios obtenidos por un Bot desde el sitio de

LinkedIn, las habilidades de los mismo y sectores de desempeño laboral. Adicionalmente, de las encuestas realizadas a los usuarios se cargan sus preferencias de cursos con la finalidad de evaluar las recomendaciones dadas por los sistemas de recomendación. En la figura 3.5 se puede ver un diagrama que ilustra el proceso, en el cual resalta los transductores semánticos que, dado los diferentes contextos de los registros de redes sociales, codifica los diferentes campos en lenguaje natural a habilidades y cargos.

### 3.6 Resumen del Capítulo

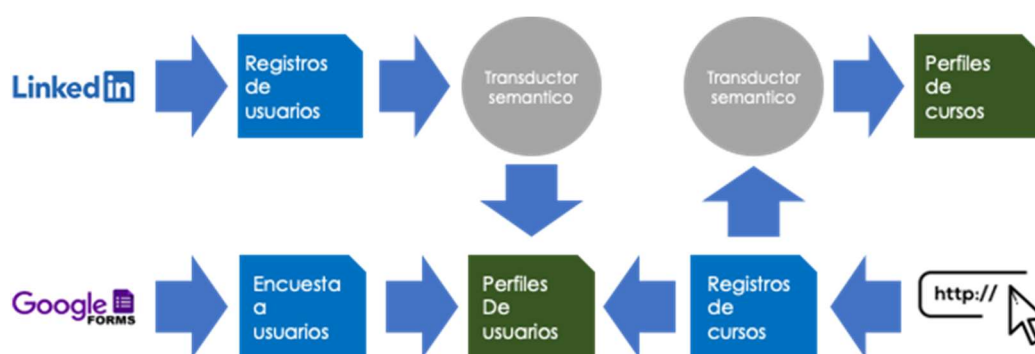


Figura 3.5 Proceso de construcción de perfiles. Elaboración propia

Para el desarrollo del sistema de recomendación, el entrenamiento de los modelos, así como la evaluación de la calidad de la recomendación de las diferentes propuestas, es necesario contar con un conjunto de datos, que no solo permitan el desarrollo de las actividades antes mencionadas, sino que adicionalmente ofrezca la flexibilidad de realizar múltiples ejecuciones de forma consistente. El conjunto de datos debe ser tal que normalicen las diversas naturalezas de las fuentes de datos minimizando la pérdida de información, para efectos de generar la mejor recomendación, y para efecto de evaluación contengan las preferencias de los usuarios. En el presente capítulo se presenta el tamaño de la muestra, método de adquisición de los datos, sus características, representación de perfiles de usuarios y de cursos, así como la construcción del dataset.



## Capítulo 4

# Diseño del Sistema de recomendación

Para demostrar la hipótesis planteada, se ha diseñado y desarrollado un SR híbrido. En el proceso se han explorado diferentes técnicas de filtrado, y desarrollado tres versiones del sistema, donde la versión final, basada en ontología y ML, permitió aprovechar la riqueza de la información obtenida de la web y mejorar el rendimiento del SR en cuanto a relevancia, y precisión de las sugerencias. A partir de la revisión de la literatura y considerando que el objetivo del SR es ofrecer un conjunto de cursos de LL, que ayuden a mejorar y desarrollar las habilidades profesionales de los usuarios en diferentes áreas del conocimiento, se consideraron las RRSS como fuente para la obtención de la información para la creación de usuarios y se plantea el uso de varios algoritmos de filtrado para obtener un buen rendimiento del SR.

### 4.1 Diseño de alto nivel

Teniendo presente el propósito de la investigación, a partir del estudio del estado del arte, y analizadas las diferentes fuentes de información, se plantea utilizar la información profesional de usuarios de las RRSS, particularmente de LinkedIn, luego de un análisis de la información contenida, en donde se puede obtener los datos de los perfiles profesionales y académicos de los usuarios, información en la que se basará la recomendación. La oferta formativa de LL, para la recomendación de cursos que desarrollen las habilidades profesionales de los usuarios, se obtendrá de las páginas web de las distintas organizaciones e instituciones educativas. La propuesta general del SR se muestra en la figura 4.1.

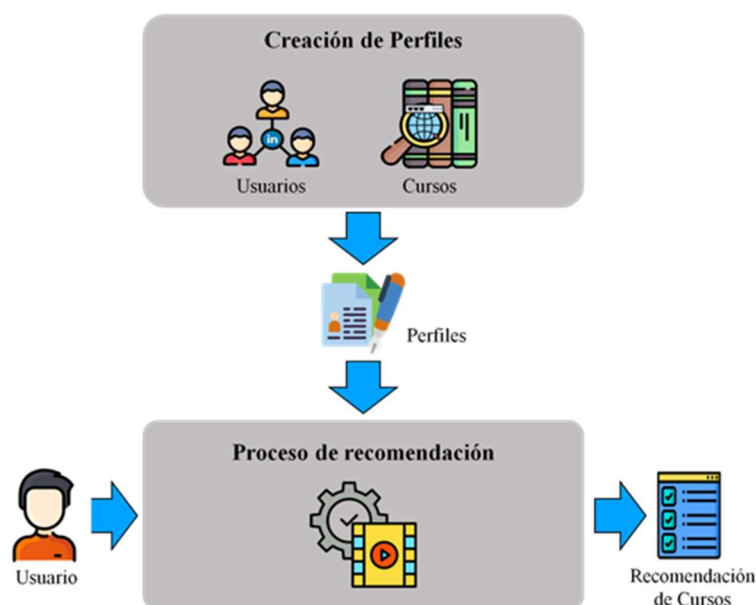


Figura 4.1 Propuesta general del Sistema de Recomendación

Para el diseño del SR se establece como base lo siguiente:

Los datos de entrada del sistema serán obtenidos de las RRSS, para la creación de los perfiles de usuario, y de la web para los cursos, sin requerir carga de información adicional por parte de los usuarios.

La recomendación de cursos se realizará utilizando como criterio las habilidades que estos desarrollan. Por ello, se debe determinar para cada uno los usuarios, las habilidades a desarrollar o mejorar, alineados con las tendencias del mercado laboral, y por tanto las

habilidades en los perfiles de usuarios y de cursos deben estar definidas en los mismos términos.

Diseño modular, lo que permite incorporar nuevos algoritmos al sistema, y ofrecer múltiples opciones de configuración, Así mismo, incluir parametrizaciones para los diferentes algoritmos, habilitar o inhabilitar etapas de filtrado, a fin de validar y evaluar las propuestas.

Tomando en cuenta lo establecido, se plantea un SR híbrido, para solventar las deficiencias del uso individual de técnicas de filtrado, basado en taxonomía para la representación de los perfiles de usuario. A partir de los datos extraídos el uso de transductores semánticos permite representar las descripciones, en lenguaje natural, de los datos recuperados en registros no estructurados o semiestructurados, en codificaciones para la construcción de los perfiles de usuarios y cursos. Los datos, codificados en los perfiles, serán la entrada del SR híbrido, que consta de múltiples técnicas de filtrado, y tiene como salida una lista ordenada de cursos de LL para la mejora y/o desarrollo de las habilidades profesionales del usuario. En la Figura 4.2 se muestran los procesos que se plantean para el SR.

A partir del planteamiento antes descrito, y producto de los resultados obtenidos de la evaluación del sistema, y como mejora de la representación de la información, durante el desarrollo de esta tesis se propusieron tres versiones del SR:

- SR híbrido basado en taxonomía.
- SR híbrido basado en ontología y ML.
- SR híbrido basado en ontología y ML con *endorsements*.

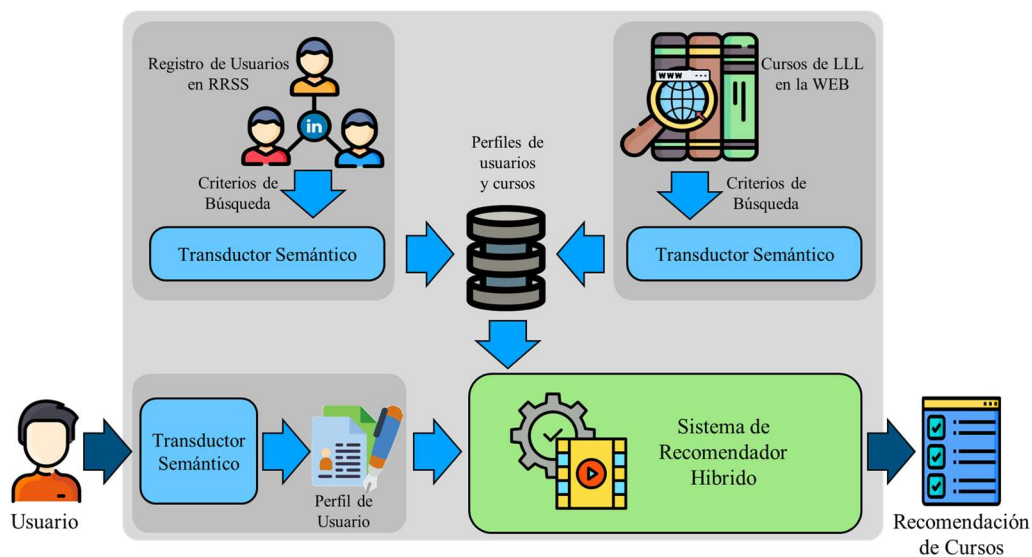


Figura 4.2 Procesos del Sistema de Recomendación

A continuación, se explica el diseño de alto nivel de cada una de estas propuestas.

#### 4.1.1 Sistema de recomendación híbrido basado en taxonomía

La revisión del estado del arte evidencia la importancia de la información, y su representación para el buen rendimiento de los SR, debido a la heterogeneidad y volumen de datos que se puede recuperar de la web. Para la construcción de los perfiles de usuarios, dada las características de los datos, se prestó especial atención a su naturaleza y a las diversas fuentes de información, se definieron dos taxonomías para representar la información recuperada.

Las taxonomías se desarrollaron con el propósito de poder relacionar la información de los usuarios, con la información de los cursos. Para la creación de los perfiles, además de la información demográfica, se toman en cuenta, las habilidades presentes en los registros profesionales de los usuarios y las habilidades que son alcanzadas a través de un curso. Esto permite establecer relaciones entre ambos tipos de fuentes de datos.

Este enfoque sugiere un SR híbrido que hace uso de una taxonomía para la codificación de la información, tanto para los perfiles de usuarios, como para los cursos.

El objetivo, al utilizar códigos jerárquicos, es evitar la pérdida de categorías de la información, asociada al uso de códigos planos, al construir los perfiles a partir de las diferentes fuentes de datos. El sistema se compone de dos procesos principales como se muestra en la Figura 4.3:

(A) Proceso de creación de perfiles.

(B) Procesos de Recomendación.

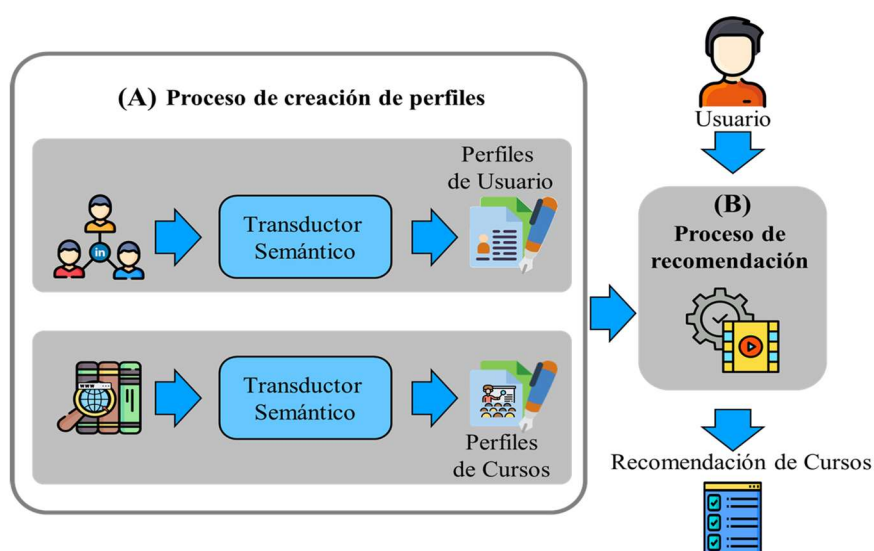


Figura 4.3 Procesos del Sistema de Recomendación híbrido basado en taxonomía

(A) Proceso de Creación de perfiles

El proceso de creación de perfiles es común a todas las propuestas, y se realiza un perfilado semántico basado en ontología, para generar los diferentes perfiles utilizados, haciendo uso de taxonomías. Este proceso está conformado por el proceso de creación de perfiles de usuarios (A.1) y el proceso de creación de perfiles de cursos (A.2), como se puede observar en la Figura 4.4:

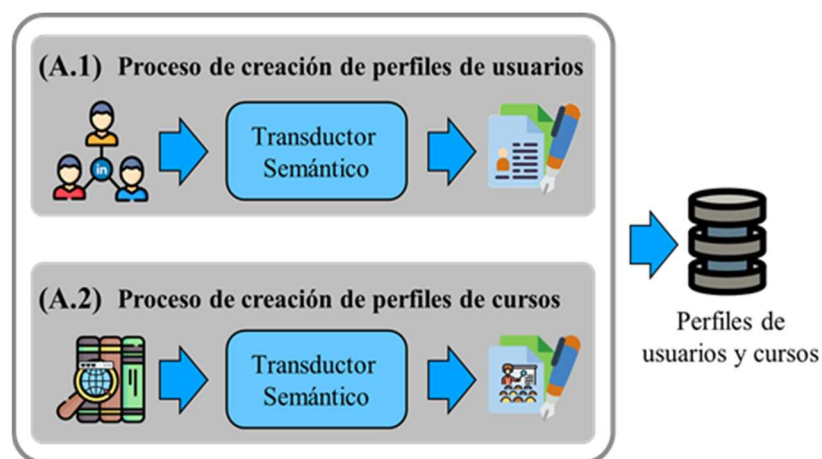


Figura 4.4 Proceso de creación de perfiles del SR híbrido basado en taxonomía

En la creación de perfiles, como se indicó previamente, las habilidades son esencialidades para el proceso de recomendación.

El trabajo de (Iordache, Mariën, & Baelden, 2017), indica que la Comisión Europea, en el marco de trabajo de cualificaciones para apoyar el aprendizaje, trabajo y la movilidad transfronteriza, diferencia entre conocimientos, habilidades y competencias. Los conocimientos los define como "el conjunto de hechos, principios, teorías y prácticas relacionados con un campo de trabajo o estudio". Las habilidades las denomina como "la capacidad de aplicar estos conocimientos", mientras que la competencia se considera "la capacidad demostrada de utilizar estos conjuntos de conocimientos y habilidades para el desarrollo personal". Por su parte (Cornide-Reyes, y otros, 2021), distingue en la literatura dos tipos de habilidades: habilidades blandas, asociadas a las dimensiones de comportamiento profesional, también conocidas como no técnicas, y habilidades técnicas, son las relacionadas con los conocimientos técnicos necesarios para actuar con un determinado nivel de rendimiento en el contexto de una disciplina, teniendo en cuenta que, diferentes disciplinas requieren distintos conocimientos técnicos.

Para efectos de esta investigación, y con base a lo anteriormente expuesto, se define habilidad como un conocimiento y/o destreza (de naturaleza teórica y/o práctica) que permite desempeñarse en un cargo; la misma se circunscribe a un área de conocimiento

específico. Para poder realizar una valoración de la habilidad, se le establece un Nivel de Especialización (en adelante NE) y un Grado de Actualización (en adelante GA).

Luego de analizar los registros de los usuarios de LinkedIn y de los cursos de LL extraídos de la Web, se definieron dos taxonomías para codificar de forma jerárquica las áreas de conocimiento y los sectores de desempeño laboral. En la figura 4.5 se muestra parcialmente, una instancia de la taxonomía de las áreas de conocimiento del dominio de computación, haciendo énfasis en lenguajes de programación.

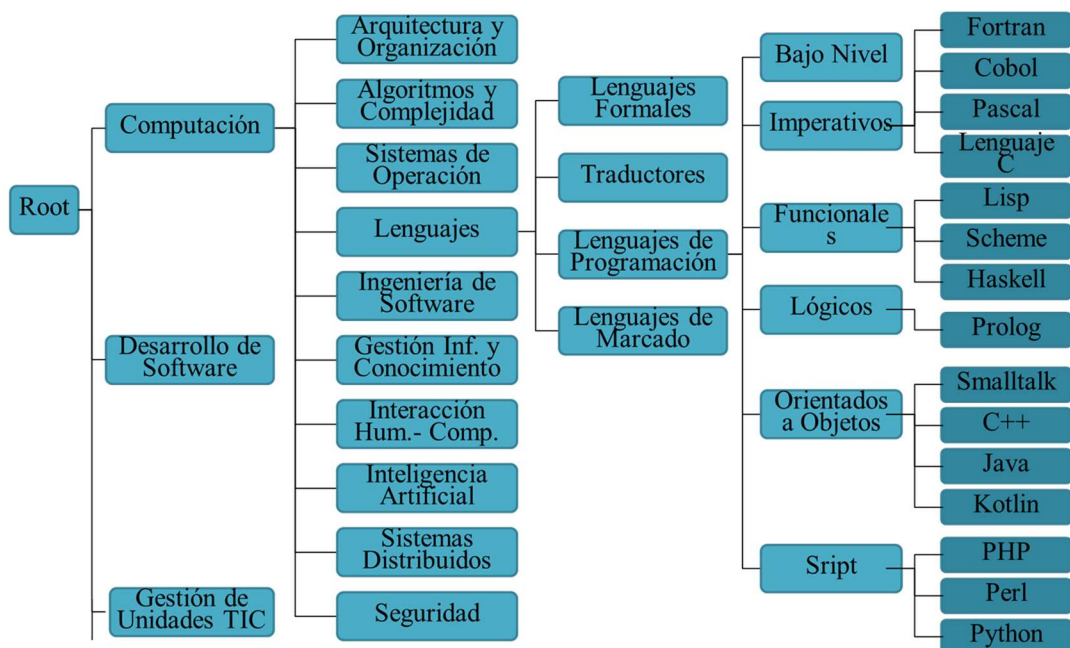


Figura 4.5 Taxonomía áreas de conocimiento – Lenguaje de programación

A sí mismo, la figura 4.6, muestra un extracto de la taxonomía para los sectores de desempeño laboral, en el área de desarrollo de software.

Con la taxonomía de las áreas de conocimiento se codifican, las habilidades de los usuarios con su NE y GA, las habilidades a desarrollar en los cursos de LL con un NE, y las habilidades que son requisito para optar a un curso.

La taxonomía de sector de desempeño laboral se utiliza para codificar jerárquicamente los diferentes cargos ocupados por los usuarios.

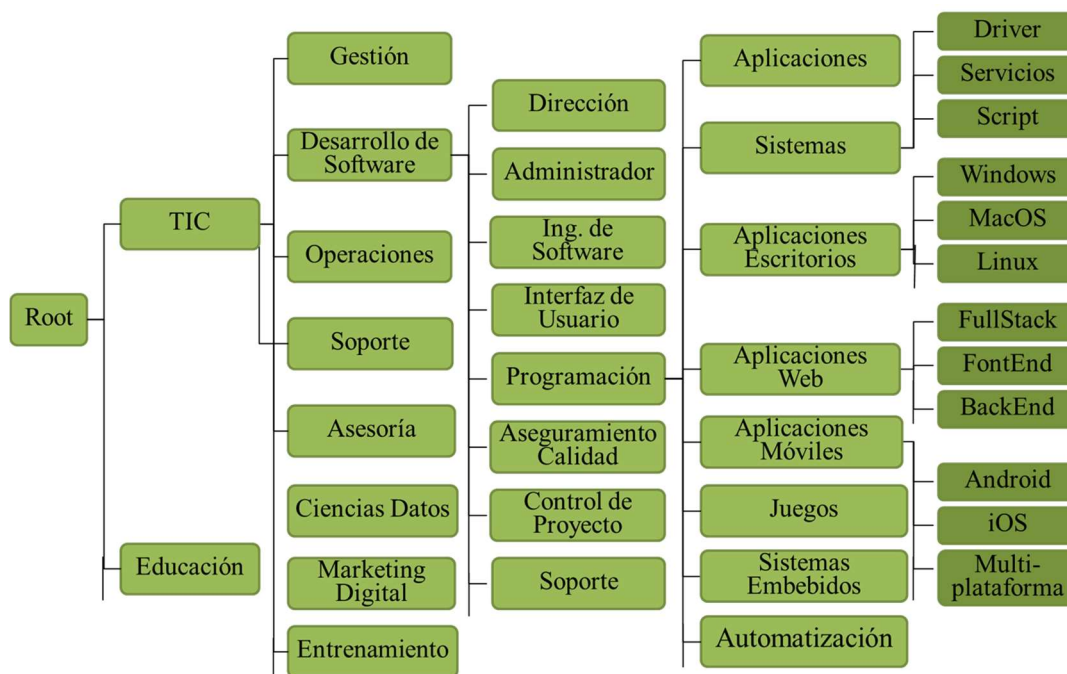


Figura 4.6 Taxonomía de sectores de desempeño laboral en el área de Desarrollo de Software

A cada uno de los nodos, en los diferentes niveles de las taxonomías, se le asocia un nombre, un identificador y una descripción, y se define una codificación basada en estos elementos. En la figura 4.7 se puede observar los nombres de los niveles y la conformación del código para las taxonomías de áreas de conocimiento y sectores de desempeño laboral.

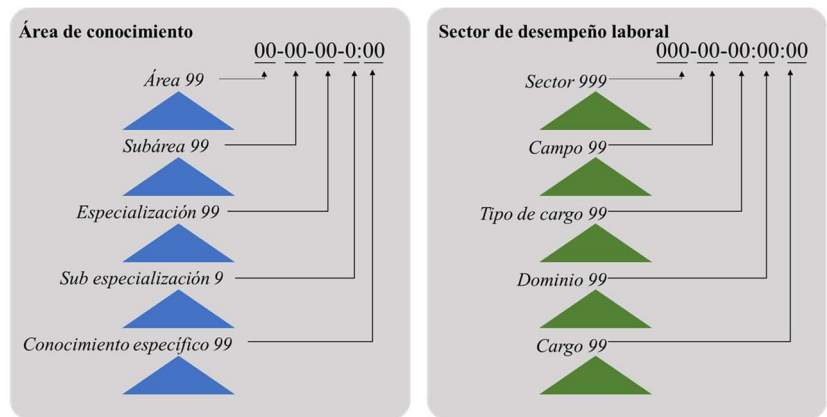


Figura 4.7 Códigos para las taxonomías áreas de conocimiento y sectores de desempeño laboral

En el proceso de perfilado se procesan todos los registros de usuario. Para codificar un término se busca su descripción en los nodos terminales tomando en cuenta su semántica. El código jerárquico se obtiene a partir de la secuencia de identificadores desde la raíz.

A modo de ejemplo, se tienen tres usuarios con habilidades de programación en diferentes lenguajes. El usuario A, con conocimiento de programación en Pascal; el usuario B, con programación en Lenguaje C y, por último, el usuario C, con conocimiento en programación en Kotlin. En la Figura 4.8 se puede observar la codificación de estas habilidades.

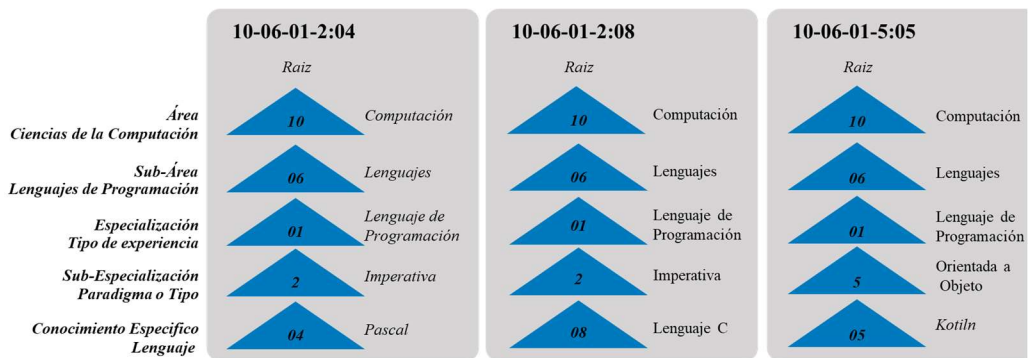


Figura 4.8 Ejemplos de codificación de habilidades

Para cada usuario se almacenan las habilidades relacionadas al sector de desempeño laboral y su cargo, con sus respectivos atributos. En la figura 4.9 se representa el almacenamiento de los perfiles de usuario en relación con las taxonomías definidas. En donde, por cada usuario se tienen sus cargos de desempeño laboral, contenidos en la taxonomía, a los que se les asocia los atributos Fecha de Inicio (FI) y Fecha de Fin (FF), y a cada habilidad se le asocia un área de conocimiento, junto con un NE y GA.

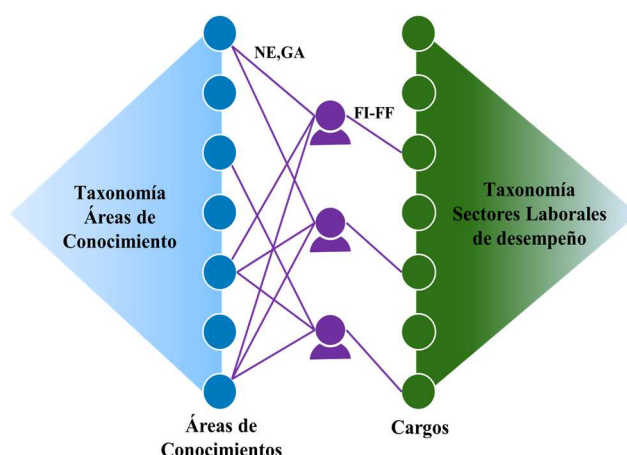


Figura 4.9 Representación con los datos de perfiles de usuarios almacenados

### (B) Proceso de recomendación

Una vez creados los perfiles de usuarios y de cursos, estos serán la entrada del SR, en donde para el proceso de recomendación se plantea combinar distintas técnicas de filtrado a fin de solventar las deficiencias del uso individual de éstos (Figura 4.10).

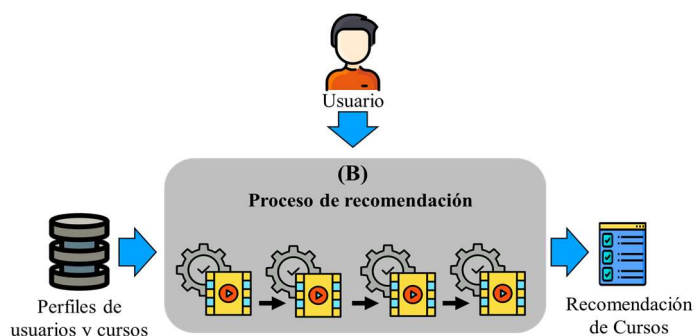


Figura 4.10 Proceso de recomendación del SR híbrido basado en taxonomía

De acuerdo con lo mencionado en el estado del arte, los algoritmos de filtrado basan sus recomendaciones en el cálculo de similitudes. En esta primera propuesta el uso de taxonomía permite la comparación de habilidades a diferentes niveles en la jerarquía, y de esta forma, se pueden comparar habilidades o sectores de desempeño laboral, más allá de una lógica bivaluada.

Retomando el ejemplo (Figura 4.8), si se comparan dos personas que tienen la habilidad del usuario A, en una lógica bivaluada, tuvieran una coincidencia total. Pero si se comparan dos personas, con habilidades diferentes, por ejemplo, el usuario A y el usuario B, ambas tienen conocimiento de programación imperativa, serían iguales solo hasta el cuarto nivel. Finalmente, si se compara al usuario B y al C, hay coincidencia hasta el tercer nivel, es decir, ambas saben de programación. Eso significa que para efectos del cálculo de similitud se pueden considerar diferentes niveles de corte. Un umbral de corte igual a 1, significa que se aplicará la comparación, como si se tratara de una codificación plana, de lo contrario, se pueden calcular distintos grados de similitud, evitando hasta cierto punto, la pérdida de información en la codificación de la fuente.

Para efectos del cálculo de la similitud de las habilidades entre usuarios se deben construir, para cada par de usuarios, vectores de habilidades que permitan hacer la comparación. Dichos vectores deben tener la misma secuencia de habilidades para efectos del cómputo. Como la valoración de una habilidad está en función de un NE y de un GA, y ambos pueden compensarse, son tratados como componentes cartesianos y se define como valor de la habilidad, el módulo del vector, ya que este, es proporcional al valor de sus componentes, siendo siempre positivo o cero en caso del vector nulo (Figura 4.11).

Los algoritmos de recomendación requieren que se calcule la similitud entre los diferentes elementos contenidos en los perfiles de usuarios y de cursos. Si bien, la longitud del trayecto compartido desde la raíz se utiliza con frecuencia para los elementos codificados mediante taxonomías, en este estudio, se dio a cada uno de los niveles, una ponderación según su relevancia. En la propuesta es posible cambiar la ponderación en cada nivel de la taxonomía en el cálculo de la similitud. Por ejemplo, se puede establecer que el conocimiento específico tiene mayor peso que la subespecialidad, al realizar una

comparación. Se debe tener en cuenta que, la suma de los factores de ponderación debe sumar 1 y la función comparación un rango entre 0 y 1, incluyendo los extremos. Del mismo modo, se permite definir un umbral de corte a partir del cual los elementos se consideran similares.

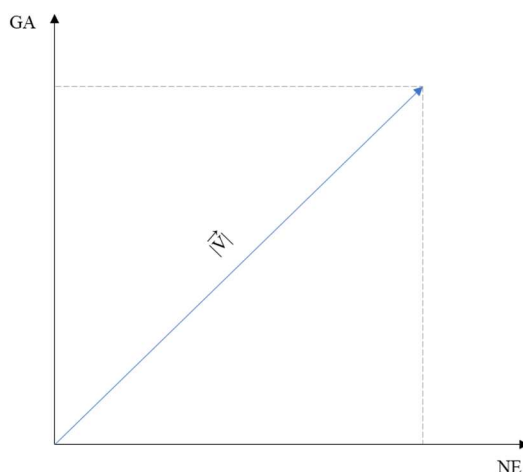


Figura 4.11 Valoración de una habilidad. Elaboración propia

Una vez creados los perfiles y de realizar el cálculo de las similitudes, se debe determinar, para un usuario objetivo, cuáles de sus habilidades debe actualizar y cuáles habilidades nuevas debe desarrollar; estas habilidades son inherentes a su cargo actual o a cargos afines, que podría desempeñar en un futuro.

Para el proceso de recomendación se definieron las siguientes funciones, las cuales se describen a continuación:

- Función de Similitud Taxonómica para sector laboral.
- Función de Similitud Taxonómica para Áreas de Conocimiento.
- Valoración del Conocimiento para la Habilidad
- Función de similitud de coseno para las habilidades de usuario.
- Función de similitud de Pearson para las habilidades de usuarios.

- Construcción de los vectores de valores de conocimiento

### **Función de Similitud Taxonómica para sector laboral ( $JS\_TS$ ).**

Compara dos elementos pertenecientes a sectores de desempeño laboral de acuerdo con la ecuación 4.1

$$JS\_TS(js_i, js_j) = \sum_{l=1}^{D(T)} (K_l * JS\_Match(js_i, js_j, l)) \quad (4.1)$$

Donde:

$js_i, js_j$  : Sectores de desempeño laboral.

$D(T)$  : Profundidad de la Taxonomía, viene dada por el número de elementos en el camino más largo desde la raíz a una hoja.

$K_l$  : Constante de pertinencia, o peso asociado, al nivel  $l$  en la taxonomía.

$JS\_Match(js_i, js_j, l)$ : Función de comparación del nivel  $l$ , con soporte de comodines (wildcard), donde:

$$F: js \times js \times l \rightarrow [0,1]$$

Para los efectos de experimentación, se propone utilizar los siguientes valores iniciales para las Constantes de Pertinencias de Nivel  $l$  ( $K_l$ )

$$K_l = \frac{1}{D(T)} : \text{Constante de pertinencia de nivel } l \text{ en la taxonomía,}$$

$$\text{verificándose que } \sum_{l=1}^{D(T)} K_l = 1.0$$

### **Función de Similitud Taxonómica para Áreas de Conocimiento ( $KA\_TS$ ).**

Compara dos elementos pertenecientes a áreas de conocimiento, entre los usuarios, se define según la ecuación 4.2

$$KA_{TS}(s_i, s_j) = \sum_{l=1}^{D(T)} (K_l * KA_{Match}(KA(s_i), KA(s_j), l)) \quad (4.2)$$

Donde:

$s_i, s_j$  : Habilidades.

$KA(s)$  : Área de Conocimiento para una Habilidad.

$D(T)$  : Profundidad de la taxonomía.

$K_l$  : Constante de pertinencia, o peso asociado, al nivel  $l$  en la taxonomía, se debe verificar que:

$$\sum_{l=1}^{D(T)} K_l = 1.0$$

$KA_{Match}(ka_i, ka_j, l)$  : Función de comparación del nivel  $l$ , con soporte de comodines (wildcard)

$$F: ka \times ka \times l \rightarrow [0,1]$$

Para las funciones de similitud taxonómica,  $k_l$  es definida como una constancia de relevancia para el nivel  $l$  en la taxonomía, donde:

$$k_l = \frac{1}{D(T)}, \text{ verificándose que } \sum_{l=1}^{D(T)} K_l = 1.0$$

### Valoración del Conocimiento para la Habilidad ( $KV$ )

A partir de los atributos de Nivel de Especialización y Grado de actualización de un área de conocimiento, se debe calcular para efectos de comparación una valoración de la habilidad, que viene dada por la ecuación 4.3:

$$KV(s) = \sqrt{SL(s)^2 + UG(s)^2} \quad (4.3)$$

Donde:

$SL(s)$  : Nivel de Especialización para la Habilidad

$UG(s)$  : Grado de Actualización para la Habilidad

### Función de similitud de coseno para las habilidades de usuario.

Para comparar el conjunto de habilidades de los diferentes usuarios, se utilizó la función de similitud del coseno sobre los vectores  $v$  y  $u$ , basada en la valoración del

conocimiento para las habilidades, la cual se calcula utilizando los atributos de nivel de especialización y grado de actualización en un área de conocimiento determinada.

La función de similitud del coseno se define según la ecuación 4.4:

$$Cos\_Sim(v, u) = \frac{\sum_{i=1}^n (v_i * u_i)}{\sqrt{\sum_{i=1}^n v_i^2} + \sqrt{\sum_{i=1}^n u_i^2}} \quad (4.4)$$

Donde:

$v, u$  : Vectores de Valores de Conocimiento para las Habilidades.

$n = |v| = |u|$

### **Función de similitud de Pearson para las habilidades de usuarios.**

Permite comparar el conjunto de habilidades de diferentes usuarios, para lo cual se estima los vectores  $v$  y  $u$  a partir sus habilidades.

La función de similitud de Pearson viene dada por la ecuación 4.5:

$$Pearson\_Sim(v, u) = \frac{\sum_{i=1}^n ((v_i - \bar{v}) * (u_i - \bar{u}))}{\sqrt{\sum_{i=1}^n (v_i - \bar{v})^2} + \sqrt{\sum_{i=1}^n (u_i - \bar{u})^2}} \quad (4.5)$$

Donde:

$v, u$  : Vectores de valores de conocimiento para las habilidades.

$n = |v| = |u|$

### **Construcción de los vectores de valores de conocimiento**

Al ser las áreas de conocimiento del conjunto de habilidades de un usuario, un código jerárquico, el cual viene dado por una taxonomía, es conveniente poder construir los vectores valores del conocimiento  $v$  y  $u$ , dado por las ecuaciones 4.6 y 4.7, tomando en cuenta la función de similitud taxonómica para las Áreas de Conocimiento, la cual permite comparar habilidades afines dado un umbral de corte, permitiendo que todas aquellas habilidades afines se agrupen bajo una misma entrada para cuyos fines se utilizaría la función máximo, como función resumen para la valoración de conocimiento.

Para lo cual se define:

$$\begin{aligned}
S &= S_i \cup S_j \\
[s] &= \{s' \in S \mid s' \sim s\} \\
[s] &= \{s' \in S \mid KA_{TS}(s, s') \geq \theta\} \\
\tilde{S} &= \{[s] \mid s \in S\} \\
v, u &= uvCal(S_i, S_j, \theta)
\end{aligned}$$

Donde:

$S_i$  : Habilidades Usuario i

$S_j$  : Habilidades Usuario j

$\theta$  : Umbral de corte para la función de similitud taxonómica

Se tiene que:

$$v_k = \max(KV(s_l)) \mid s_l \in S_i, s_l \in \tilde{s}_k, \tilde{s}_k \in \tilde{S} \quad (4.6)$$

$$u_k = \max(KV(s_l)) \mid s_l \in S_j, s_l \in \tilde{s}_k, \tilde{s}_k \in \tilde{S} \quad (4.7)$$

Los vectores  $v$  y  $u$  se construyen a partir del conjunto cociente obtenido al definir como relación de equivalencia la función de similitud taxonómica con umbral de corte  $\theta$  de 0,8, sobre el conjunto  $S_i \cup S_j$ , donde cada elemento con índice  $k$  de  $v$  y  $u$  es el valor máximo de las valoraciones de conocimiento de las habilidades de  $S_i$  y  $S_j$ , donde se verifica que las habilidades pertenezcan a  $k$ -ésima clase de equivalencia del conjunto cociente.

Dado un usuario  $U$ , tomando en cuenta sus habilidades propias ( $H_p$ ), se determinan, por similitud de habilidades, los usuarios similares ( $S$ ). Tomando en cuenta los cargos de los usuarios afines ( $Ca$ ), y aquellos usuarios que puedan tener el mismo cargo del usuario  $U$ , se procede a determinar las habilidades afines ( $Ha$ ), estas corresponden a aquellas que aparecen con mayor frecuencia entre los usuarios similares y los usuarios con igual cargo que  $U$ , que son diferentes a las propias del usuario. En la Figura 4.12, se puede apreciar este proceso.

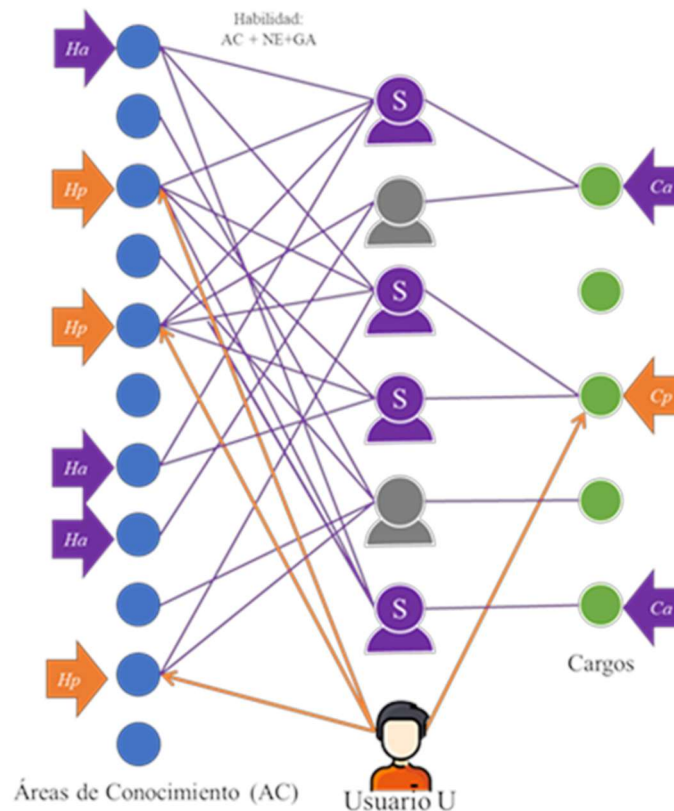


Figura 4.12 Determinación de las habilidades a mejorar y/o desarrollar para un usuario en el SR híbrido basado en taxonomía

Se puede presentar el caso en donde un usuario no tenga un cargo definido, o bien porque es un estudiante recién egresado o porque es un profesional que quiere reorientar su carrera. Se determinan los cargos afines ( $C_a$ ), que serán aquellos que poseen los usuarios similares, basados en habilidades, al usuario  $U$ , y a partir de estos cargos afines se pueden determinar las habilidades afines ( $H_a$ ). Resalta el hecho que el proceso utiliza solamente cargos afines. En ese caso, su conjunto de habilidades no está relacionado a ningún cargo. En la figura 4.13 se observa cómo es el proceso para determinar las habilidades a actualizar y/o a desarrollar.

Tras determinar las habilidades que el usuario debe desarrollar o mejorar, se aplican algoritmos de filtrado colaborativo y basado en contenido, para determinar una predicción inicial de cursos. Se seleccionan los cursos que tienen entre sus habilidades a desarrollar,

las habilidades propias o afines del usuario. La recomendación final de cursos LL se construye a partir de refinar y ordenar la predicción inicial de cursos utilizando heurística.

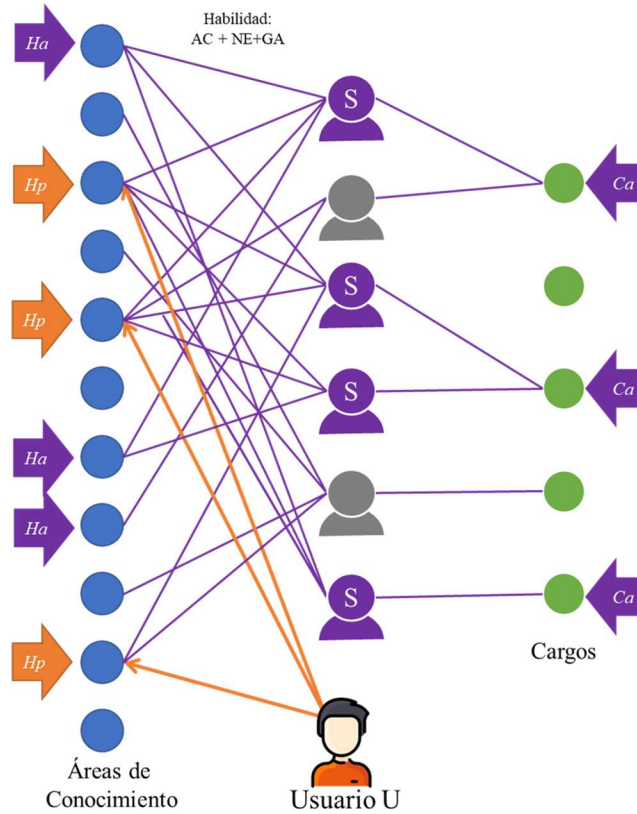


Figura 4.13 Determinación de las habilidades a mejorar y/o desarrollar para un usuario sin cargo propio en el SR híbrido basado en taxonomía

#### 4.1.2 Sistema de recomendación híbrido basado en ontología y ML

La primera propuesta mostró la viabilidad de la estrategia y en la evaluación surgieron oportunidades de mejora relacionadas, especialmente, con el manejo de la información.

En la revisión del estado del arte sobre SR, presentada en el Capítulo 2, se destacó la importancia de la representación y gestión de la información, donde se señala el uso de ontología para modelar y representar el conocimiento de un dominio, así como de inferir información y mejorar los problemas de arranque en frío y los problemas de dispersión de datos, que limitan el rendimiento de los RS.

Basado en lo anteriormente expuesto, para una segunda versión del sistema, se propone pasar de datos almacenados, a un modelo basado en ontología, en donde se incorporan técnicas de ML. Esto permite tener una mejor representación de la realidad y hacer uso de un filtrado semántico para las recomendaciones.

En función de la eficiencia, es recomendable pasar a una representación en la que no se tengan que almacenar todos los perfiles de usuario. Se plantea hacer uso de las taxonomías definidas en la primera versión para utilizar como modelo una ontología. Esto se logra enriqueciendo con nuevas entidades, relaciones y atributos, las taxonomías, que van a formar el sustrato de la ontología. Se ha definido como sustrato de la ontología, aquellas instancias de entidades, relaciones y atributos en la ontología que no cambian, son los elementos sobre los cuales se basan los procesos de actualización vía eventos y ML, desde los conjuntos de datos de entrenamiento.

El sistema se compone de tres procesos principales como se muestra en la Figura 4.14:

- (A) Proceso de creación de perfiles.
- (B) Proceso de actualización de ontología.
- (C) Proceso de recomendación del SR híbrido basado en ontología y ML.

A diferencia del primer sistema, esta versión consta de una fase off-line, en donde se ejecutan los procesos (A) y (B), y una fase on-line, en donde se ejecuta el proceso de recomendación (C), para el usuario objetivo

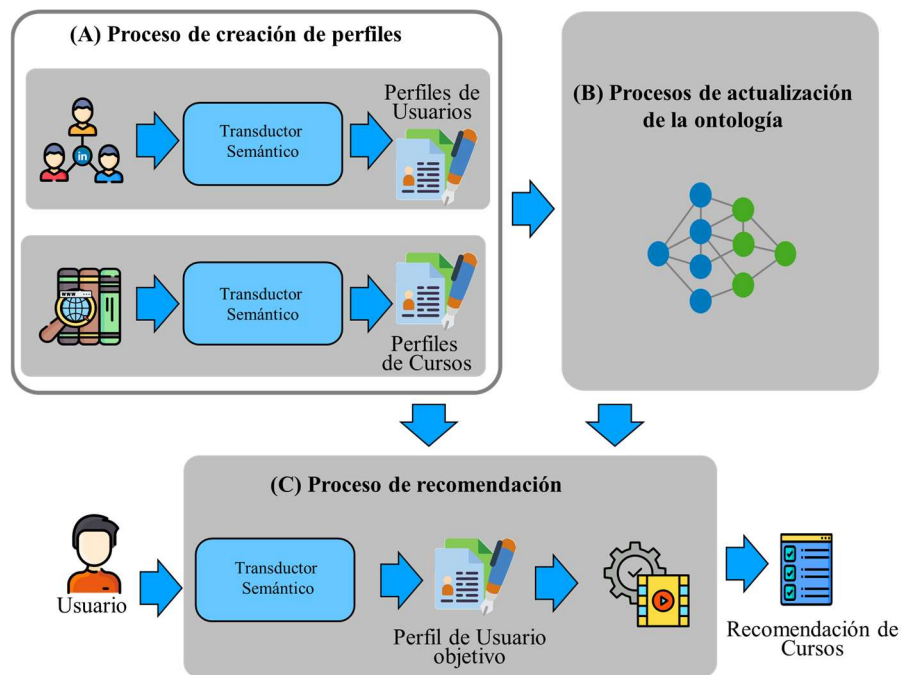


Figura 4.14 Procesos del SR híbrido basado en ontología y ML

Como se indicó en la propuesta anterior, el proceso de creación de perfiles de usuarios (A) es similar en las tres propuestas. A continuación, se describe el proceso de actualización de ontología (B) y el proceso de recomendación (C).

### **(B) Proceso de actualización de ontología**

El proceso de actualización de la ontología se hace en dos etapas que como se puede observar en la Figura 4.15:

(B.1) El proceso de actualización de ontología vía eventos.

(B.2) El proceso de ML para la creación de los clústeres de desempeño laboral

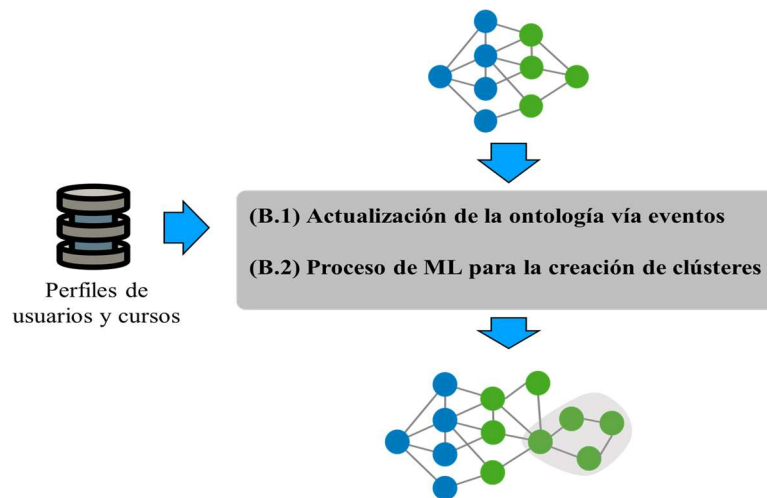


Figura 4.15 Procesos de actualización de la ontología del SR híbrido basado en ontología y ML

### **(B.1) Proceso de actualización de ontología vía eventos**

En la ontología se almacena la relación entre los cargos y las áreas de conocimiento, que viene dada por las habilidades con atributos, así como el número de veces (N) que está presente esa habilidad, el promedio de su NE y el promedio del GA, esto se calcula tomando en cuenta al número de usuarios que poseen la habilidad. Las relaciones son determinadas, y sus atributos calculados, a partir de un conjunto de datos de entrenamiento de perfiles de usuarios vía eventos en la ontología. La Figura 4.16, muestra la representación de los datos basada en ontología.

Para el sustrato de la ontología, se definen nuevas entidades, relaciones y atributos, a las taxonomías y se permite hacer inferencias, que se usarán en un filtrado semántico. Se contempla, entre las nuevas relaciones, sinónimos y usos de términos en otros idiomas. Para efectos de la prueba de esta propuesta, se consideraron el uso de términos en castellano y el inglés de forma indistinta, debido al uso común en el dominio utilizado. En el caso de áreas de conocimiento, se manejan en sus entidades relaciones del tipo “es de interés” para indicar, que a los usuarios que posean una habilidad relacionada con esas entidades, les puede interesar desarrollar otras habilidades con otros conocimientos, para obtener un perfil más integral.

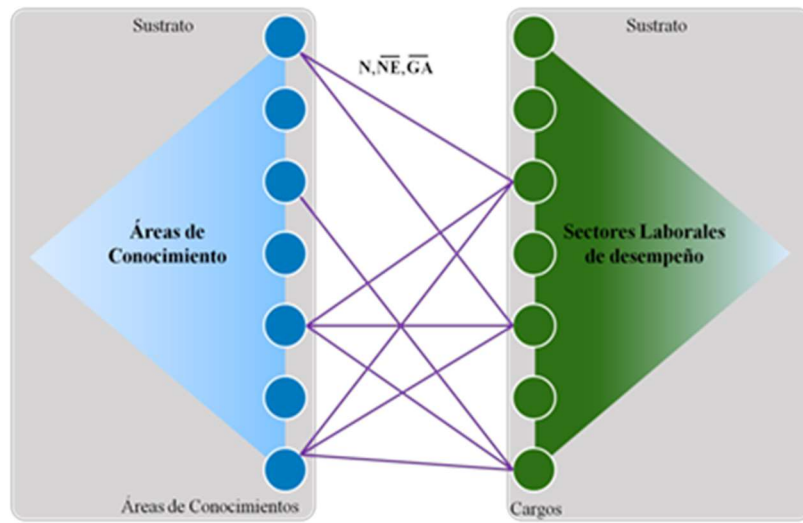


Figura 4.16 Representación de los datos basada en ontología

Adicionalmente, en los sectores de desempeño laboral, las entidades de cargos pueden tener relaciones “optar a” donde desde un cargo se puede optar a cargos similares o mayores, en una relación de orden definida por un escalafón. Esto se puede apreciar en la Figura 4.17.

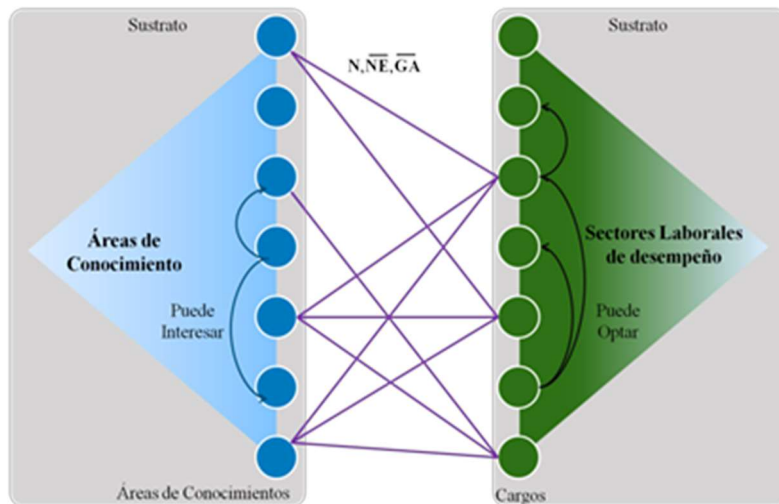


Figura 4.17 Nuevas relaciones soportadas entre entidades de la Ontología

Dada la ontología se propone un filtrado semántico para determinar las habilidades afines y ontológicas. Las primeras se determinan como se indicó anteriormente, y las segundas son el producto de usar relaciones “son de interés” en la ontología.

### **(B.2) Proceso de ML para la creación de los clústeres de desempeño laboral**

Una alternativa para el cálculo de las habilidades afines es utilizando ML. Se propone utilizar algoritmos no supervisados de *clustering*, sobre las instancias de las entidades de cargos, para agrupar cargos similares, estas agrupaciones están basadas en la similitud de las habilidades asociadas a los cargos. La agrupación de cargos de desempeño laboral, basada en la similitud del conjunto de habilidades, puede ser utilizada para determinar el conjunto de cargos afines a un cargo determinado, o un conjunto de habilidades de un usuario en particular. A partir de la agrupación de cargos determinada, se puede establecer como las habilidades afines aquellas que aparecen con mayor frecuencia en los cargos de la agrupación, que el usuario no posee.

### **(C) Proceso de recomendación del SR híbrido basado en ontología y ML**

Para el proceso de recomendación se necesitan determinar las habilidades propias del usuario a mejorar, así como, las habilidades afines y ontológicas que se deben desarrollar. La Figura 4.18 representa el proceso que se lleva a cabo para determinar dichas habilidades.

Para un usuario  $U$ , haciendo uso de sus habilidades propias ( $H_p$ ), se identifican los cargos en que estas habilidades son utilizadas, y se determina vía filtrado de coberturas los posibles cargos afines ( $C_a$ ). Esto se realiza por dos vías, la primera, identificando aquellos cargos donde las habilidades del usuario cubren un porcentaje de las habilidades asociadas al cargo, y la segunda, identificando aquellos cargos, cuyas habilidades asociadas, cubren un porcentaje de las habilidades del usuario. En este proceso, puede pasar, que una habilidad del usuario no está asociada a ningún cargo, en este caso, es posible buscar cargos afines, utilizando habilidades asociadas a áreas de conocimiento vecinas, subiendo un nivel en la taxonomía.

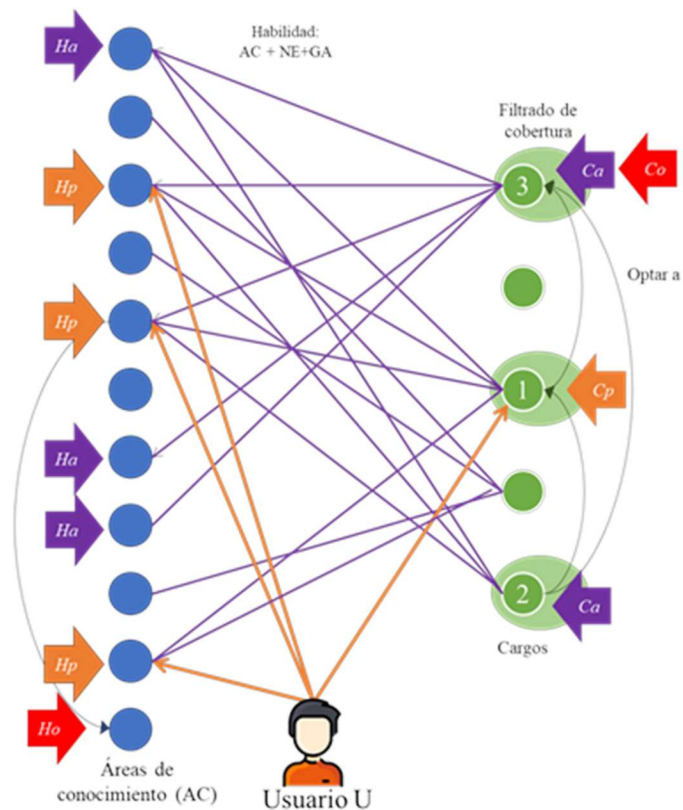


Figura 4.18 Determinación de las habilidades en el SR híbrido basado en ontología

Sobre los posibles cargos afines ( $C_a$ ) se realiza un filtrado utilizando la relación “optar a”, que evita considerar cargos “inferiores al actual”, dado una relación de orden impuesta por un escalafón, obteniéndose los cargos afines filtrados ontológicamente ( $C_o$ ). Con los cargos afines ( $C_a$ ), los cargos ontológicos ( $C_o$ ) y el cargo propio ( $C_p$ ) se proceden a determinar las habilidades afines ( $H_a$ ) “más frecuentes”. Posteriormente, a partir de las habilidades propias del usuario ( $H_p$ ) y las habilidades afines ( $H_a$ ) se obtienen las habilidades ontológicas ( $H_o$ ), las cuales vienen dadas por las relaciones “son de interés”.

La Figura 4.19, representa el procedimiento para el cálculo de las habilidades afines, basadas en los clústeres de áreas de desempeño laboral. Para un usuario ( $U$ ), con un cargo propio ( $C_p$ ) y sus habilidades propias ( $H_p$ ), se predice el clúster al que pertenece; los cargos que lo conforman pasan a ser los cargos afines ( $C_a$ ). Una vez identificados estos cargos, se sigue según el procedimiento similar anterior. Para este fin, es importante

destacar que, si un cargo aparece por primera vez, se debe asignar a uno de los clústeres existentes.

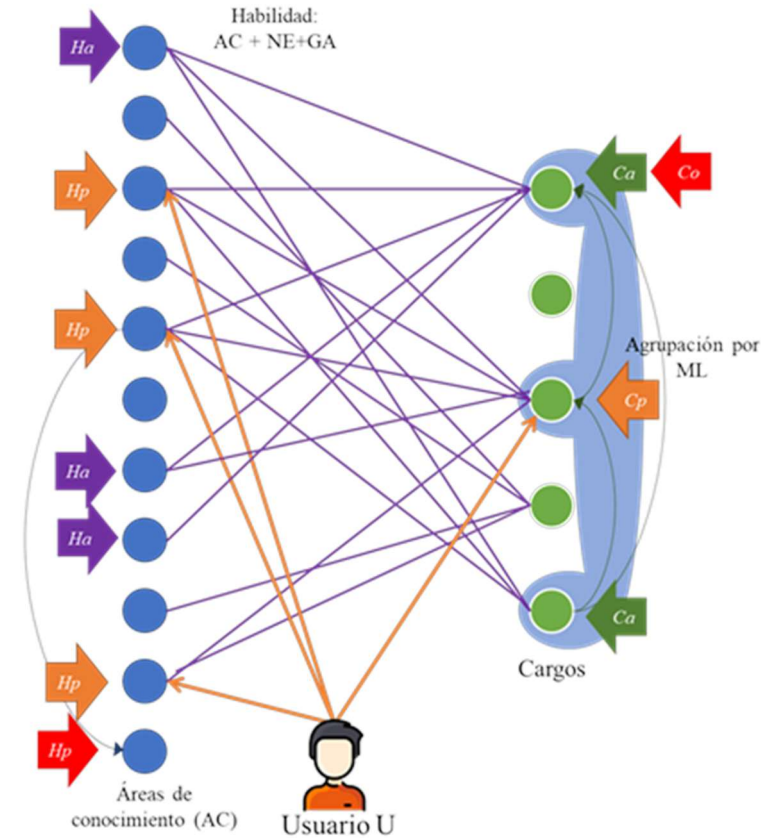


Figura 4.19 Determinación de las habilidades a partir de ML en la ontología en el SR híbrido basado en ontología y ML

Una vez identificadas las habilidades a desarrollar y/o mejorar, prosigue el proceso de recomendación de forma similar al SR basado en taxonomía. Se aplica un filtrado semántico y uno basado en contenido, para determinar una predicción inicial de cursos, que se refina y ordena, aplicando una heurística para la recomendación final.

### 4.1.3 Sistema de recomendación híbrido basado en ontología y ML con *endorsements*

En LinkedIn, los usuarios de la RRSS pueden avalar las habilidades de otro usuario, estas certificaciones se conocen como *endorsements*. Con el objetivo de mejorar la precisión

de las recomendaciones del sistema, se propone para una tercera versión del este, tomar en cuenta, para la construcción del perfil de usuario, los *endorsements* para certificar alguna habilidad que se incorpore en el perfil de un usuario.

El sistema se compone de cuatro procesos principales como se muestra en la Figura 4.20:

- (A) Proceso de creación de perfiles.
- (B) Proceso de actualización de ontología.
- (C) Creación de perfil de usuario objetivo.
- (D) Proceso de recomendación del SR híbrido basado en ontología y ML.

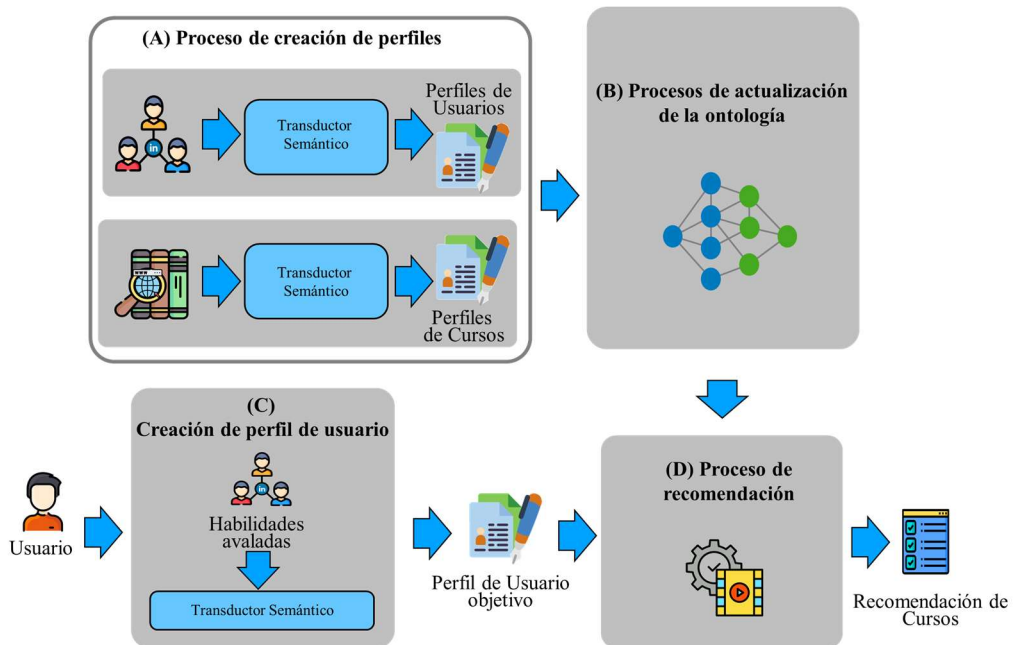


Figura 4.20 Procesos del SR híbrido basado en ontología y ML con *endorsements*

El sistema tiene una fase off-line, conformada por los procesos de construcción de los perfiles (A) y los procesos de actualización de la ontología (B), los cuales se describieron en la segunda propuesta.

La fase on-line, conformada por el perfilado del usuario objetivo (C) y el motor de recomendación (D), donde la principal diferencia con el sistema anterior es que se

considera como entrada adicional del SR, para la creación del perfil del usuario objetivo, los *endorsements*, los cual se toma en cuenta en la construcción de la recomendación.

### **(C) Proceso de creación del perfil del usuario objetivo**

Este proceso difiere con la versión anterior del sistema, en la etapa de transducción semántica, al incorporar la codificación los *endorsements* (E), en el perfil de usuario. En la información obtenida, del registro de LinkedIn del usuario al que se la va a hacer la recomendación, están los *endorsements* y haciendo uso de los algoritmos de codificación de habilidades, se determina el código jerárquico de área de conocimiento, para el NE y GA, se utilizan heurísticas que toma en cuenta la frecuencia relativa de la cantidad de usuarios que avala la habilidad. Los *endorsements* forman parte del perfil del usuario al cual se le va a realizar la recomendación y en el proceso de recomendación se toman en cuenta como habilidades que el usuario ha logrado.

### **(D) Proceso de recomendación**

En el proceso de recomendación, si los *endorsements* no están en el conjunto de las habilidades propias, se debe considerar que el usuario los posee en las etapas de filtrado, de tal forma que no es necesario recomendar cursos que las desarrolle.

En la Figura 4.21 se observa que en la determinación de las habilidades afines ( $H_a$ ) al coincidir con habilidad una avalada (E), la habilidad afín ( $H_a$ ) se considera como habilidad propia ( $H_p$ ), adicionalmente, en el caso de cursos a recomendar, donde aparezca como habilidad requerida una habilidad avalada, esta se considera como lograda.

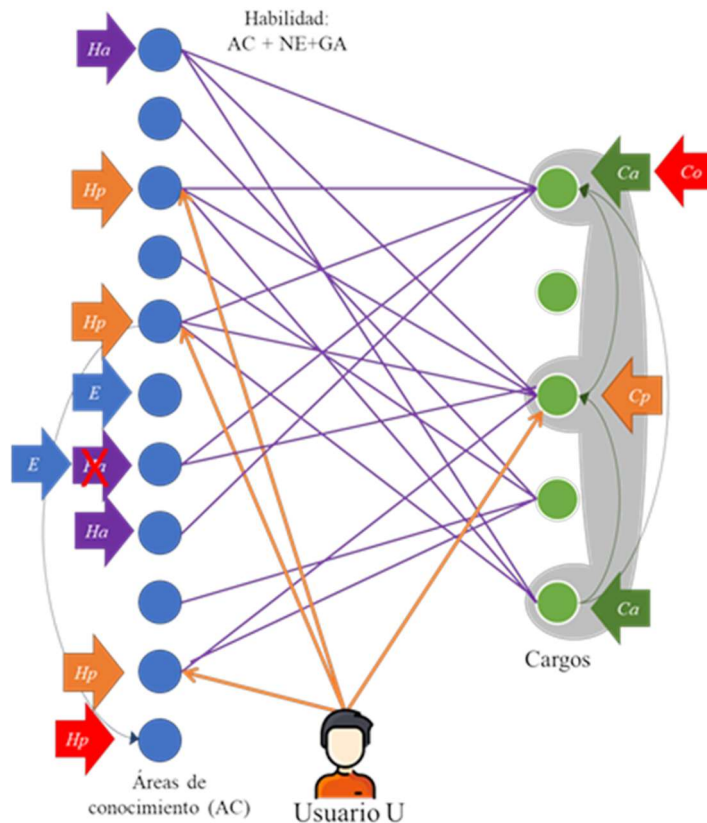


Figura 4.21 Determinación de las habilidades a mejorar o desarrollar para un usuario particular en el SR híbrido basado en ontología y ML con *endorsements*

## 4.2 Diseño de bajo nivel

En las próximas secciones se abordará el diseño de bajo nivel de las diferentes versiones propuestas del sistema.

### 4.2.1 Sistema de Recomendación híbrido basado en taxonomía

En la primera versión del SR se plantea como estrategia mantener almacenados todos los datos de los perfiles de usuarios y cursos, dado que, para dar una recomendación se procesan todos los perfiles. Para el proceso de construcción de los perfiles se hace uso de unas taxonomías.

El motor de recomendación comprende cuatro filtrados, dos etapas de filtrado colaborativo, una de filtrado por contenido y un filtrado heurístico. La arquitectura está representada en la Figura 4.22, en donde se pueden observar la fase de creación de perfiles (A) y la fase de recomendación (B).

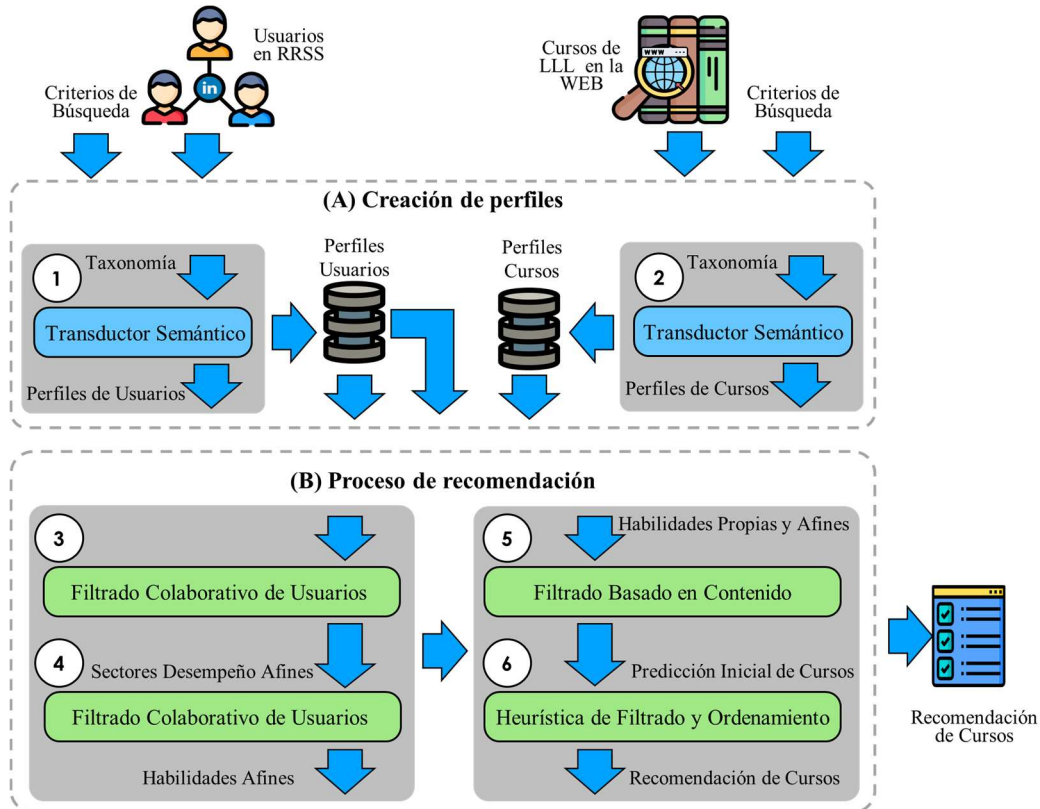


Figura 4.22 Arquitectura SR híbrido basado en taxonomía

La fase (A) consta de dos procesos, el proceso 1, para la construcción de los perfiles de usuarios y el proceso 2 corresponden a la construcción de perfiles de cursos. La fase B está conformada por los procesos del 3 al 6, que corresponden a cada una de las diferentes etapas de filtrado SR. A continuación, se describen cada uno de ellos.

### (A) Creación de los perfiles de usuarios y de cursos

Como se indicó en el capítulo 3, los perfiles de usuario contienen información demográfica, información del sector de desempeño laboral y de las habilidades.

Para la representación del área de conocimiento de las habilidades y de los cargos de desempeño laboral, de los perfiles de cursos y usuarios, se utiliza dos taxonomías. Para su construcción se siguieron los pasos propuestos por (Centelles, 2005), los cuales se detallan a continuación:

*Delimitar la realidad que será representada por la taxonomía*, se representan los sectores de desempeño laboral (cargos) y las áreas de conocimiento, donde se circunscriben las habilidades, para ambos casos se define como dominio profesional el de desarrollo de software.

*Extracción del conjunto de términos o categorías que representan dicha realidad*, para este paso se definen las fuentes de información, para las áreas de conocimiento, se utilizan los registros profesionales de LinkedIn y el currículo de computación (ACM & IEEE, 2020), y para los sectores de desempeño laboral los registros profesionales de LinkedIn y estructuras organizativas de empresas y unidades de desarrollo de software. Como mecanismo de extracción se utilizó el análisis de los registros y documentos.

*Control terminológico de los términos o categorías*, en este paso se realizan dos tareas. En primer lugar, se identifican los términos que designan un mismo concepto, se incluyen abreviaciones y anglicismos, determinando cuales se consideran preferentes y cuales se consideran sinónimos. En segundo lugar, se da una forma correcta y consistente a los términos en ambas taxonomías. Como resultado de este paso es el establecimiento de la relación de equivalencia entre todos los términos de la taxonomía.

*Establecimiento del esquema y la estructura de organización de los términos o categorías*, este paso incluye los criterios utilizados para dividir y agrupar las categorías. El modelo estructural define el tipo de relación que se establece entre las agrupaciones, derivadas del esquema de organización. En el caso de las dos taxonomías definidas, se aplicó modelos arbóreos, basado en la relación 'parte de'. En este paso se colocaron los términos hojas a un mismo nivel y al definir la profundidad, se requería utilizar la mínima profundidad que no comprometiera la capacidad de representación de ambas taxonomías, con la finalidad de poder realizar múltiples evaluaciones.

Luego de definir las taxonomías, para la construcción de los perfiles, transductores semánticos hacen uso de codificadores para los sectores de desempeño laboral y las habilidades de usuarios, dados los diferentes contextos en sus registros de RRSS. La codificación se hace de acuerdo con código jerárquico haciendo uso de las taxonomías. Para determinar el NE y GA de las habilidades, el codificador hace uso de listas de términos y reglas para su estimación.

Un proceso similar es utilizado para perfilar los cursos y codificar sus habilidades, tanto las que se desarrollan, como las que se tienen como requisitos para tomar el curso.

A continuación, se describe el proceso de codificación de las habilidades de usuarios:

Para el proceso se tienen las siguientes entradas:

$Context_0$  : Contexto de la organización  $o$ , agrupa los campos asociados a cada una de las organizaciones en el registro de LinkedIn para cada usuario

$Sieve$  : Lista de términos a filtrar.

$SL\_List_i$  : Lista de términos que indica en nivel de especialidad  $i$ .

Y la salida dado un usuario  $u_i \in U$  se define:

$Skill_i$  : Conjunto de Habilidades del usuario

Para codificar las habilidades del usuario se tiene el siguiente algoritmo:

---



---

**Algoritmo: Codificación de habilidades de usuario**

---



---

**Entrada:**  $Context_0, Sieve, SL\_List$

---

- 1:  $Skill' \leftarrow \emptyset$
- 2:  $Phrases \leftarrow Split(Context_{actual}.Descriptions)$
- 3: **Para**  $Phrase \in Phrases$  **hacer**
- 4:      $SL \leftarrow SL\_Estimator(Phrase, SL\_List, Experience(Context_{actual}))$
- 5:      $UG \leftarrow UG\_Estimator(Phrase, Context_{actual}.FI, Context_{actual}.FF)$
- 6:      $MatchCodes' \leftarrow \emptyset$
- 7:      $Codes' \leftarrow \emptyset$
- 8:     **Para**  $word \in Phrase$  **hacer**

```

9:            $MatchCodes' \leftarrow MatchCodes' \cup Onto\_Get\_KA\_Code(word)$ 
10:        Para  $Code \in MatchCodes'$  hacer
11:           Si  $Code.mandatory$  entonces  $Skill_i \leftarrow Skill_i \cup Code$ 
12:           Si  $Code.leaf$  entonces
13:             Para  $Path \in MatchCodes'$  hacer
14:               Si  $Path \text{ In } Code$  hacer  $Codes' \leftarrow Codes' \cup Code$ 
15:            $SCodes' \leftarrow MS2S(Codes', MostCommon)$ 
16:        Para  $Score \in SCodes'$  hacer
17:            $Skill' \leftarrow Skill' \cup (Score, SL, UG)$ 
18:    $Skill_i \leftarrow MS2S(Skill')$ 

```

---

**Salida:**  $Skill_i$

---

Donde:

$Onto\_Get\_KA\_Code(word)$ , obtiene de la ontología los códigos desde la raíz de todas las entidades perteneciente a taxonomía de áreas de conocimiento, donde el termino word aparece es su descripción.

$SL\_Estimator()$ , Estima el nivel de especialidad.

$UG\_Estimator()$ , Estima el grado de actualización.

$UG\_Split(Text)$ , separa el texto Text en frases según los signos de puntuación.

$MS2S(MS, c)$ , construye un conjunto con los c elementos más frecuentes en el multiconjunto MS, si c es omitido se construye con todos los elementos y en caso de que c sea la constante MostCommon, se construye con los elementos que con la mayor cantidad de frecuencia.

Como requerimiento se tiene que las habilidades deben estar definida en los mismos términos, tanto en los perfiles de usuarios como en los perfiles de cursos, es decir, se tiene que utilizar la misma codificación en ambos casos.

## **(B) Proceso de Recomendación**

Luego de construir los perfiles de los usuarios y de cursos, se ejecutan el proceso de recomendación, el cual consiste en cuatro etapas, como se observa en la Figura 4.23, las etapas 3 y 4 de filtrado colaborativo, donde se determinan los sectores de desempeño

laboral afines y las habilidades afines, respectivamente. La etapa 5 se corresponde a un filtrado basado en contenido, en donde se realiza una predicción inicial de cursos, y finalmente, en la etapa 6 se aplica una heurística de filtrado y ordenación para determinar la predicción final de cursos de LL.

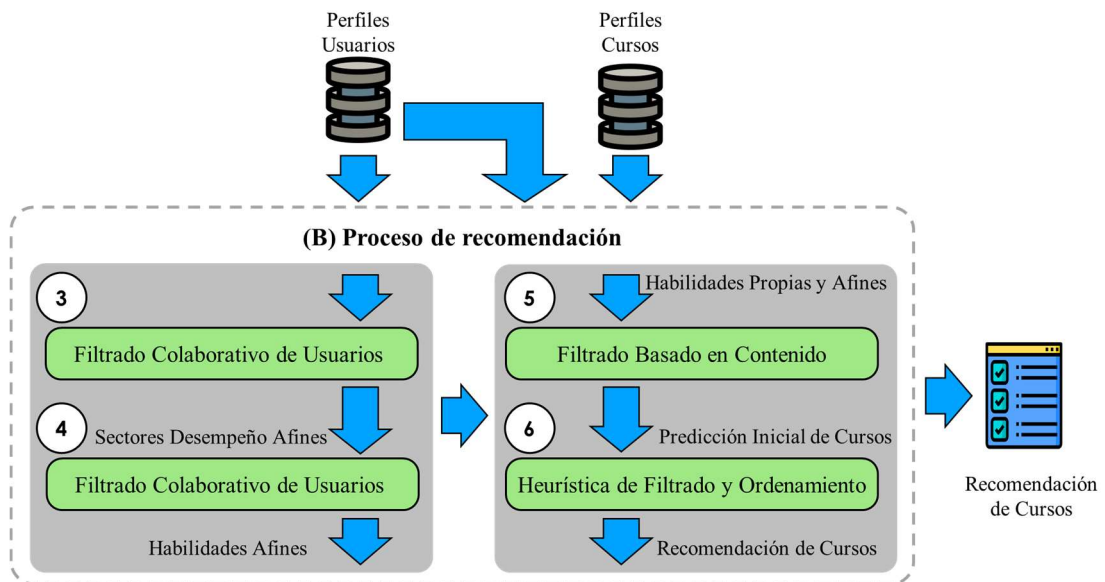


Figura 4.23 Algoritmo de recomendación del SR híbrido basado en taxonomía

El proceso de recomendación sigue el algoritmo general representado en la Figura 4.24, en donde 3, 4, 5 y 6 corresponden a las etapas de filtrado de la Figura 4.23.

A continuación, se explican cada una de las etapas de filtrado.

### (3) Filtrado colaborativo para determinación de los sectores de desempeño laboral afines

Para los perfiles usuarios se determinan sus posibles sectores de desempeño laboral afines. Dado un usuario y haciendo uso de su lista de habilidades propias (S), se procede a buscar en el resto de los usuarios aquellos con mayor similitud, utilizando para ello la función de similitud de coseno con un umbral de corte. Finalmente, se seleccionan, los sectores de desempeño laboral diferentes a los del usuario, más frecuentes.

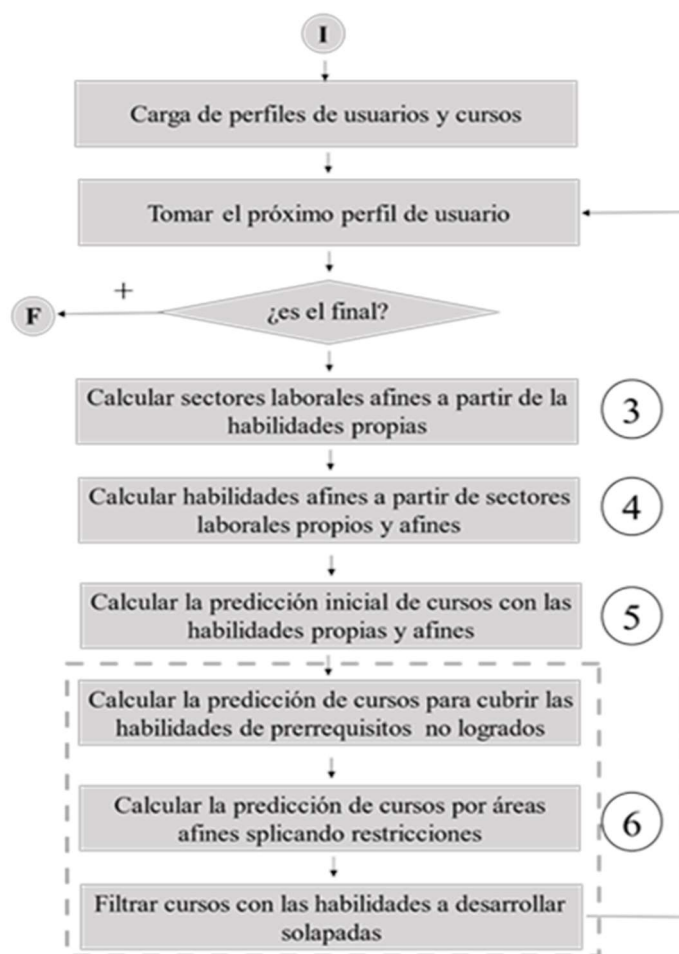


Figura 4.24 Algoritmo de recomendación del SR híbrido basado en taxonomía

Se tienen para el proceso las siguientes entradas:

$U$  : Perfiles de usuarios.

$\eta$  : Máximo sectores de desempeño laboral.

$\theta$  : Umbral de corte para la similitud taxonómica.

$\mu$  : Umbral de corte para la similitud del coseno.

$\tau = \eta - |JS_i|$ , donde  $\tau \geq |AJ S_i|$ , donde representa el máximo número de sectores laborales afines

Y la salida:

Sectores de desempeño laboral afines para los usuarios.

Ahora, dado un usuario  $u_i \in U$  se define:

$JS_i$  : Sectores de desempeño laboral del usuario i.

$AJS_i$  : Sectores de desempeño laboral afines del usuario i.

$S_i$  : Habilidades del usuario i.

Para determinar los sectores de desempeño laboral afines, dadas las habilidades propias de los usuarios, se hace uso de un filtrado colaborativo utilizando la función de similitud del coseno, y se tiene el siguiente algoritmo:

---

---

**Algoritmo: Determinación Sectores de desempeño laboral afines**

---

---

**Entrada:**  $U, \eta, \theta, \mu, \tau$

---

```
1:  Para  $u_i \in U$  hacer
2:       $AJS \leftarrow \emptyset$ 
3:      Para  $u_j \in (\{U\} - \{u_i\})$  hacer
4:           $u, v \leftarrow uv\_Cal(S_i, S_j, \theta)$ 
5:          Si  $Cos\_SIM(u, v) \geq \mu$  entonces
6:               $AJS \leftarrow AJS \cup \{j \mid j \in JS_i, j \notin JS_i\}$ 
7:       $AJS_i \leftarrow MS2S(AJS, \tau)$ 
```

---

**Salida:**  $AJS_i$

---

#### (4) Filtrado colaborativo para la determinación de habilidades afines

Se determinan las habilidades afines de los usuarios dados sus sectores de desempeño laboral, propios y afines. Para este cálculo, se hace uso de una función de similitud taxonómica con la cual se comparan los sectores de desempeño, y, si el resultado es mayor a un umbral dado, se escogen las habilidades diferentes a las del usuario, para finalmente quedarse con las de mayor frecuencia.

Para el proceso se tienen las siguientes entradas:

$U$  : Perfiles de Usuarios.

$\gamma$  : Máximo cantidad de habilidades.

$\varphi$  : Umbral de corte.

$\rho = \gamma - |S_i|$ , donde  $\rho \geq |AS_i|$ , donde representa máxima cantidad de habilidades afines

Y la salida:

Habilidades Afines para los usuarios.

Dado un usuario  $u_i \in U$  se define adicionalmente:

$AS_i$  : Habilidades Afines del usuario  $i$ .

Adicionalmente se define la función:

$SV(s, S)$  : Verifica que una habilidad  $s$  agregue valor a un conjunto de habilidades  $S$ , es decir, que en el conjunto de habilidades no exista la habilidad, o en caso contrario tenga un mayor nivel de especialización o grado de actualización.

Para encontrar habilidades afines dadas las áreas de desempeño propias y afines de un usuario, haciendo uso de un filtrado colaborativo, se tiene el siguiente algoritmo:

---



---

**Algoritmo: Determinación de habilidades afines**

---



---

**Entrada:**  $U, \gamma, \varphi, \rho$

---

- 1: **Para**  $u_i \in U$  **hacer**
  - 2:      $AS \leftarrow \emptyset$
  - 3:     **Para**  $u_j \in (U - \{u_i\})$  **hacer**
  - 4:         Si  $\exists js \in (JS_i \cup AJS_i), \exists js' \in JS_j: JS\_TS(js, js') \geq \varphi$  **entonces**
  - 5:              $AS \leftarrow AS \cup \{s \mid s \in S_k, SV(s, Si)\}$
  - 6:      $AS_i \leftarrow MS2S(AS, \rho)$
- 

**Salida:**  $AS_i$

---

**(5) Filtrado colaborativo para la recomendación inicial de cursos**

Utilizando un filtrado por contenido, se procede a filtrar el catálogo de cursos para cada usuario, de tal forma que los cursos eleven el nivel de la especialización y/o el grado de actualización de las habilidades propias o ayuden a desarrollar las habilidades afines.

Para el proceso se tiene las siguientes entradas:

$U$  : Perfiles de usuarios.

$C$  : Perfiles de catálogo de cursos.

$\lambda$  : Máximo número de cursos para la etapa.

$\delta$  : Umbral de corte para función de similitud.

$\omega$  : Umbral de corte para grado de actualización.

Y la salida:

Predicción de Cursos para los usuarios.

Dado un usuario  $u_i \in U$  adicionalmente se define:

$CP_i$  : Predicción de cursos para el usuario  $i$ .

Así mismo, para la manipulación de cursos se definen las siguientes funciones:

$S2D(c)$  : Habilidades a desarrollar por un curso.

$KA2D(s)$  : Área del Conocimiento de la habilidad a desarrollar.

$SL2D(s)$  : Nivel de Especialización para la habilidad a desarrollar.

Para la construcción de la predicción inicial de cursos para los usuarios, se hace uso de un filtrado por contenido utilizando una función de similitud taxonómica, para lo cual se tiene el siguiente algoritmo:

---

---

**Algoritmo: Recomendación inicial de cursos**

---

---

**Entradas:**  $U, C, \lambda, \delta, \omega$

---

- 1: **Para**  $u_i \in U$  **hacer**
- 2:      $CP \leftarrow \emptyset$
- 3:     **Para**  $c \in C$  **hacer**
- 4:         **Si**  $\exists s \in S_i, \exists s' \in S2D(c): KA_{TS}(KA(s), KA2D(s')) \geq \delta$  **entonces**
- 5:             **Si**  $UG(s) < \omega \vee SL(s) < SL2D(s')$  **entonces**
- 6:                  $CP \leftarrow CP \cup \{c\}$

7: **Continuar**  
8: **Si**  $\exists s \in AS_i, \exists js' \in S2D(c): KA_{TS}(KA(s), K2A2D(s')) \geq \delta$   
**entonces**  
9: **Si**  $SL(s) \leq DL2K(s')$  **entonces**  
10:  $CP \leftarrow CP \cup \{c\}$   
11:  $CP_i \leftarrow MS2S(CP, \lambda)$

---

**Salida:**  $CP_i$

---

### (6) Heurística para el filtrado y el orden de los cursos

Se utiliza una heurística para complementar y filtrar los cursos predichos en la fase anterior. Se deben verificar los siguientes criterios para cada candidato:

1. Para cada curso propuesto, se debe comprobar que el usuario cumpla con los requisitos, de lo contrario se añaden otros cursos, que, cumpliendo las restricciones demográficas, ayuden a desarrollar las habilidades necesarias, en este proceso se limita el número de fases de formación como máximo a  $\sigma$ , este proceso se puede observar en la Figura 4.25, donde:  $f(c_j, c_i)$  es la relación dada cuando  $c_j$  desarrolla las habilidades que son prerrequisitos para  $c_i$  y el p.

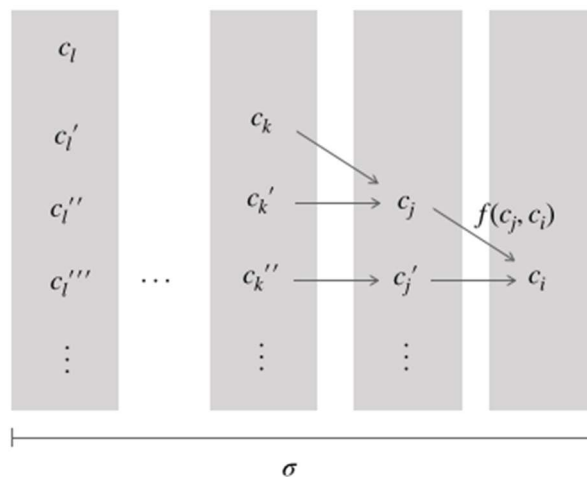


Figura 4.25 Desarrollo de Prerrequisitos

2. Se verifican restricciones demográficas de usuario, tanto a la predicción inicial del curso, como a los cursos añadidos en el paso anterior.

3. Si predicción inicial de cursos es vacía, se recomiendan cursos relacionados con las habilidades del usuario, aplicando restricción demográfica.
4. Para los cursos destinados a desarrollar las mismas habilidades, se seleccionaron los clasificados más altos.
5. Se verifica si hay cursos, cuyas habilidades a desarrollar sean un subconjunto de las habilidades desarrolladas por otro curso, si este es el caso, se eliminan de la predicción inicial.

Para el proceso se tienen las siguientes entradas:

$U$  : Perfiles de Usuarios.

$C$  : Perfiles de Cursos.

$\delta$  : Umbral de corte para función de similitud.

$\phi$  : Máximo número de cursos para los pasos 1 y 3.

$\sigma$  : Máximo número de etapas para lograr los prerrequisitos.

$\psi$ : Mínima cantidad para recomendar cursos relacionados

Y la salida:

Recomendación de Cursos

Dado un usuario  $u_i \in U$  se define:

$CR_i$  : Recomendación de cursos para el usuario  $i$ .

Se definen adicionalmente las siguientes funciones:

$SK2GSK(S)$ : Conjunto de Áreas de Conocimiento de las Habilidades.

$GS2D(c)$  : Conjunto de Áreas de Conocimiento de las Habilidades a desarrollar por un curso.

$PRKA(c)$ : Áreas de Conocimiento prerequisites de un curso.

$RKA(c)$  : Áreas del Conocimientos relacionadas a un curso.

$CV(c)$  : Valoración de un curso.

$ADR(u, c)$  : Aplica restricciones demográficas de usuario a un curso.

$F: u \times c \rightarrow true, false$

Para el filtrado heurístico de la predicción final de cursos para los usuarios, se tiene el siguiente algoritmo:

---

---

**Algoritmo: Recomendación final de cursos**

---

---

**Entradas:**  $U, C, \delta, \phi, \sigma, \psi$

---

```
1:  Para  $u_i \in U$  hacer
2:       $CR \leftarrow \emptyset$ 
3:       $PRCP \leftarrow \emptyset$ 
4:       $CP \leftarrow CP_i$ 
5:       $GS \leftarrow SK2GSK(S_i)$ 
6:      Para  $rango(1.. \sigma)$  hacer                                // paso 1
7:           $CPI \leftarrow \emptyset$ 
8:          Para  $c \in CP$  hacer
9:              Para  $k' \in PRKA(c)$  hacer
10:                 Si  $\neg(\exists k \in GS: KA\_TS(k, k') \geq \delta)$  entonces
11:                     Para  $c' \in C$  hacer
12:                         Si  $\exists k'' \in GS2D(c'): KA\_TS(k', k'') \geq \delta$  entonces
13:                              $CPI \leftarrow CPI \cup \{c'\}$ 
14:                              $GS' \leftarrow GS' \cup GS2D(k')$ 
15:                  $CP \leftarrow CPI$ 
16:                  $PRCP \leftarrow PRCP \cup CPI$ 
17:                  $GS \leftarrow GS \cup GS'$ 
18:          Para  $c \in CP_i$  hacer                                // paso 2
19:              Si  $ADR(U_i, c)$  entonces
20:                   $CR \leftarrow CR \cup \{c\}$ 
21:               $CR \leftarrow CR \cup MS2S(PRCP, \phi)$ 
22:              Si  $|CR| \leq \psi$  entonces                            // paso 3
23:                   $GS \leftarrow SK2GSK(S_i)$ 
24:                   $CP \leftarrow \emptyset$ 
```

```

25:      Para  $c \in C \wedge ADR(u_i, c)$  hacer
26:          Si  $\exists k \in GS, \exists k' \in RKA(c): KA_{TS}(k, k') \geq \delta$  entonces
27:               $CP \leftarrow CP \cup \{c\}$ 
28:           $CR \leftarrow MS2S(CP, \lambda)$ 
29:      Para  $c \in CPR \wedge c' \in (CPR - \{c\})$  hacer           // Paso 4
30:          Si  $S2D(c) \equiv S2D(c') \wedge CV(c) > CV(c')$  entonces
31:               $CPR \leftarrow CPR - \{c\}$ 
32:      Para  $c \in CPR \wedge c' \in (CPR - \{c\})$  hacer           // Paso 5
33:          Si  $S2D(c') \subset S2D(c)$  entonces
34:               $CPR \leftarrow CPR - \{c\}$ 
35:       $CR_i \leftarrow CR$ 

```

---

**Salida:**  $CR_i$

---

#### 4.2.2 Sistema de recomendación híbrido basado en ontología y ML

A partir del SR híbrido basado en taxonomía, se propone utilizar una ontología para mejorar el rendimiento, y optimizar la predicción del SR para los cursos de LL. En tal sentido, se enriquecen las taxonomías existentes con el fin de modelar el conocimiento relacionado con las habilidades de los usuarios y los sectores de desempeño laboral. Utilizando los datos de los perfiles de registros profesionales de LinkedIn, se determinaron las habilidades utilizadas en los diferentes empleos y sectores de desempeño laboral. Estos últimos, se agruparon utilizando técnicas de ML para predecir afinidades dado un nuevo perfil de usuario a partir de los sectores de desempeño laboral o habilidades.

Para el proceso de recomendación, se propuso un sistema híbrido de tres etapas: filtrado semántico, que hace uso de la ontología, para determinar las habilidades a actualizar y/o desarrollar; un filtrado por contenido, para la predicción inicial de los cursos; y una heurística, para obtener la oferta final de cursos a recomendar. La arquitectura del SR propuesto se presenta en la Figura 4.26.

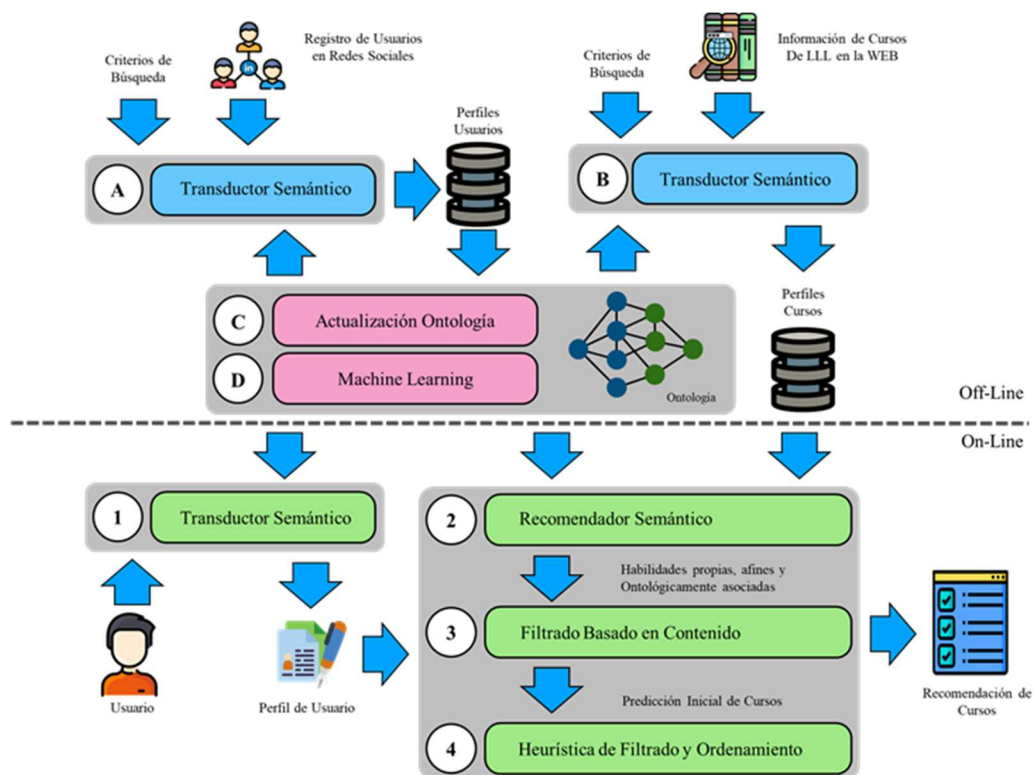


Figura 4.26 Arquitectura de SR híbrido basado en ontología y ML

El sistema se compone de dos fases principales: la fase off-line (I), en la que se construyen los perfiles de usuario (A) y (B), se cargan las definiciones y se construyen los modelos (C) y (D); y una fase on-line (II), en la que se realiza el proceso de recomendación de cursos, para lo cual, debe construir el perfil del usuario al que se le va a dar la recomendación (1), la cual se realiza en tres etapas de filtrado (2,3 y 4).

Antes de explicar el proceso de recomendación, se describe el proceso previo de la definición de la ontología utilizada en esta propuesta y el proceso de ML, que están implicados en la recomendación.

### Definición de la Ontología

Aunque la taxonomía tiene una muy buena relación representación-eficiencia, muestra limitaciones al momento de representar la realidad. Ante la necesidad de tener un modelo, como alternativa al almacenamiento y procesamiento de todos los perfiles de

usuario del SR basado en taxonomía, se hace uso de una ontología por la riqueza que ofrece para representar la realidad y la capacidad de realizar inferencias.

La construcción de la ontología siguió un proceso iterativo, el cual se basó en las metodologías Ontology Development 101 (Noy & McGuinness, 2001) y Methontology (Fernández, Gómez-Pérez, & Juristo, 1997), siguiendo los siguientes pasos:

**Paso 1:** Propósito de uso de la Ontología y las preguntas a las que la información almacenada dará respuesta.

La ontología, en su sentido más amplio, busca representar cargos, habilidades y las relaciones existentes entre ellas, adicionalmente, debe ofrecer la capacidad de representar diferentes agrupaciones de los cargos bajo el criterio de similitud en función de las habilidades que utilizan.

Como ya se había definido antes, una habilidad está referida a conocimiento específico, que puede ser teórico y/o práctico, y este se encuentra circunscrito a diferentes categorías en una relación jerárquica que parte de lo más general del dominio de aplicación. Así mismo, se deben representar las diferentes relaciones que se pueden encontrar entre las habilidades.

Con relación a los cargos, de forma similar, estos se circunscriben a diferentes categorías en una relación jerárquica que parte de lo más general del dominio de aplicación. Así mismo, se deben poder representar sus requisitos y un orden parcial referidos a un escalafón de cargos.

Las principales preguntas a las que dará respuesta la ontología están relacionadas con:

- El conjunto de requisitos de un cargo.
- Categorías en la jerarquía a las que pertenecen un cargo.
- El conjunto de términos asociados a un cargo a lo largo de la jerarquía.

- La pertenencia de un cargo en el grupo formado por los cargos de menor nivel, dado según la relación de orden impuesta por el escalafón.
- El conjunto de habilidades promedio utilizada en el ejercicio de un cargo.
- Las aéreas de conocimientos que son de interés a otra.
- El conjunto de cargos donde es utilizado un conocimiento específico.
- El porcentaje de cobertura de un conjunto habilidades dado, sobre las habilidades utilizadas en un cargo.
- El porcentaje de cobertura de las habilidades de un cargo dado, sobre un conjunto de habilidades.
- Los conocimientos específicos “hermanos” en la jerarquía que comparten un ancestro de conocimiento específico dado.

**Paso 2:** Definición del dominio de la ontología.

Aunque la ontología debe ser lo suficientemente general para representar cualquier dominio de aplicación, el dominio de aplicación sobre el cual se realizó el levantamiento de conocimiento fue desarrollo de aplicaciones de software, basándose en las taxonomías de área de conocimiento y cargos de desempeño laboral de la versión del SR basado en taxonomía.

**Paso 3:** Elicitación de conocimiento.

Para la adquisición de conocimiento se utilizó, el análisis de los registros de usuarios de LinkedIn, la descripción de cursos de LL en la web, estructura organizativa de empresas de desarrollo de software y el currículo de ACM.

**Paso 4:** Modelo conceptual.

**Paso 4.1:** Definición de Clases. La Tabla 4.1 muestra las definiciones de las diferentes clases, su identificación, descripción, así como su naturaleza o tipo: sustrato, si sus instancias son definidas en la elicitación de conocimiento; si son actualizadas vía

eventos, por los registros utilizados para entrenamiento; y, por último, ML, si las instancias son productos del resultado de los diferentes algoritmos de agrupamiento.

Tabla 4.1 Clases de la ontología

<b>Clase</b>	<b>Descripción</b>	<b>Tipo</b>
<b>Área</b>	Categoría general en la que se divide el dominio de conocimiento, por ejemplo: Computación	Sustrato
<b>Sub-Área</b>	Diferentes categorías generales que contiene un área de conocimiento, por ejemplo: Lenguajes	Sustrato
<b>Especialidad</b>	Diferentes ramas de conocimiento específicos en las que contiene una subárea de conocimiento, por ejemplo: Lenguajes de Programación	Sustrato
<b>Sub-Especialidad</b>	Diferentes categorías que contiene una especialidad. Agrupa el conocimiento común de los diferentes conocimientos específicos, por ejemplo: Leguajes de programación Orientados a Objetos	Sustrato
<b>Conocimiento</b>	Conocimiento específico que pertenece a una subespecialidad. Se corresponde con el nivel más particular (hojas) en la jerarquía de área de conocimiento. Pueden detonar el dominio de tecnologías, herramientas, software, etc. Por ejemplo, Lenguaje de Programación Java	Sustrato
<b>Sector</b>	Categoría general en la que se divide el dominio laboral, por ejemplo: Tecnología Información y Comunicación	Sustrato
<b>Campo</b>	Diferentes categorías generales que contiene un sector laboral, por ejemplo: Desarrollo de Software	Sustrato
<b>Tipo</b>	Naturaleza de los diferentes cargos contenidos en campo, por ejemplo: Desarrollador	Sustrato
<b>Dominio</b>	Áreas particulares contenidas en un tipo de cargo, por ejemplo: Desarrollo de Aplicaciones Móviles	Sustrato
<b>Cargo</b>	Cargo específico en un dominio de cargo. Se corresponde con el nivel más particular (hojas) en la jerarquía de sectores de desempeño laboral, por ejemplo: Desarrollo de Aplicaciones móviles iOS Nativo.	Sustrato
<b>Requisito</b>	Requisitos específicos de un cargo. Pueden ser de diferentes tipos (deseables, obligatorios). Están relacionados con la edad, experiencia, formación académica, entre otras. Por ejemplo, más de 10 años de	Sustrato

	experiencia, ingeniero en computación, dominio del idioma inglés.	
<b>Habilidad</b>	Nivel de experticia y grado de actualización asociado a un área de conocimiento	Actualizado vía evento
<b>Clúster</b>	Grupo de cargos afines	ML

**Paso 4.2:** Jerarquía de Clases, ésta viene dada por las diferentes relaciones entre las clases de la ontología, las cuales se describen en la Tabla 4.2.

Tabla 4.2. Relaciones de la ontología

Relación	Descripción	Cardinalidad	Clases
<b>Es un</b>	Pertenece a una categoría	1:1	Entre clase adyacentes perteneciente a una jerarquía
<b>Es igual</b>	Las descripciones son sinónimos	1:1	Entre clase iguales perteneciente a una jerarquía
<b>Es de interés</b>	Un conocimiento es de interés para otros conocimientos	1:N	Entre conocimientos
<b>Involucra</b>	Una habilidad involucra o es sobre un conocimiento específico	1:1	Habilidad → Conocimiento
<b>Posee</b>	Un cargo se hace uso de un conjunto de habilidades Un Clúster posee conjunto de habilidades como Centroide	1:N	Cargo → Habilidad Clúster → Habilidad
<b>Incluye</b>	Un Clúster incluye un conjunto de cargos	1:N	Clúster → Cargo
<b>Tiene</b>	Un cargo tiene un conjunto de requisitos	1:N	Cargo → Requisitos

**Paso 4.3:** Definición de atributos. En la ontología las clases poseen atributos, en la Figura 4.27, se muestran las clases y sus atributos, así como las relaciones entre ellas. Es de notar que un conocimiento se representa jerárquicamente en términos de: área, subárea, especialización, subespecialización y conocimiento específico; y así mismo se utiliza una

jerarquía para representar un sector de desempeño laboral (posición) compuesto por: sector, campo, tipo, dominio y puesto de trabajo específico.

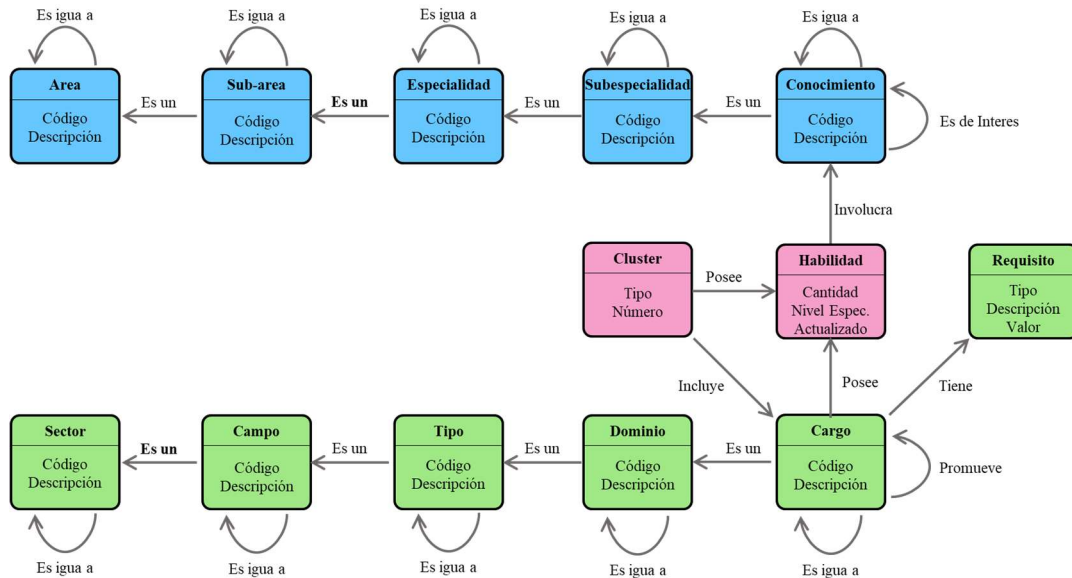


Figura 4.27 Clases y relaciones de la ontología

## Machine Learning

La idea de utilizar ML es que el desempeño del SR mejore con la experiencia, es decir, en la medida en que se tenga una mayor cantidad de usuarios del sistema, este mejore la calidad de las recomendaciones, esto es particularmente útil cuando se tienen grandes cantidades de datos con regularidades implícitas, y donde los programas se deben adaptar de forma dinámica para responder a los cambios en el contexto.

Al utilizar una ontología para representar las relaciones entre áreas de conocimiento, los sectores de desempeño laboral y las habilidades, los dos primeros filtrados del SR basado en taxonomía, son reemplazados por un filtrado semántico y se simplifica el filtrado heurístico. El filtrado semántico se implementa de forma alternativa con dos algoritmos: uno basado en reglas y coberturas de habilidades, y otro que predice la pertenencia de un cargo o un conjunto de habilidades, a una agrupación de cargos, donde los miembros del grupo comparten cierta similitud.

Para la construcción de la agrupación de cargos se hace uso algoritmos de agrupamiento (clustering), que consisten en reunir vectores en grupos según un criterio, que normalmente es similitud o distancia. Dado la naturaleza del problema son interés estos algoritmos, al tratarse de algoritmos de ML no supervisado. En la implementación del sistema se utilizan dos algoritmos de clustering, con la finalidad de ir agrupando a los usuarios, de acuerdo con la similitud de habilidades y puestos de trabajo. Dado que a priori no se tiene una agrupación, para efectos de evaluar el comportamiento del sistema propuesto, se selecciona un algoritmo de clúster globular y uno basado en densidad. Entre los algoritmos de clúster globular se selecciona K-means, por ser de fácil implementación y tener un buen efecto de agrupación tanto para conjuntos de datos grandes o pequeños. De los algoritmos basados en densidad se selecciona DBSCAN, ya que identifican clústeres de formas arbitrarias y son robustos ante la presencia de ruido.

Para aplicar el algoritmo K-means se hace uso de la distancia, determinada a partir de la similitud de las habilidades encontradas en los sectores de desempeño laboral, y encuentra los centroides para los clústeres, cuya cantidad se define a priori utilizando el método del codo; estos algoritmos pueden presentar problemas de ruido y los valores atípicos.

El algoritmo DBSCAN, puede crear agrupaciones de forma arbitraria a partir de identificar regiones densas, estos algoritmos presentan una mejor inmunidad a ruidos y valores atípicos.

#### (I) Fase Off-line

Un elemento clave para el buen rendimiento de los SR es garantizar la riqueza de la minería de datos semánticos, y evitar la pérdida de información obtenida de la recuperación de datos. Esta fase tiene como objetivo construir modelos de datos para representar el dominio de las habilidades y los sectores de desempeño laboral, por áreas de conocimiento.

Se construyen los perfiles de los usuarios y de los cursos mediante transductores (A y B) que hacen uso de la ontología, previamente cargada como grafo de conocimiento. A

partir de los perfiles de usuario, se actualizó la ontología mediante eventos (C), es decir, se actualizó el modelo con las relaciones entre sectores de desempeño laboral y las áreas de conocimiento según las competencias de los perfiles de usuario; y finalmente, mediante ML (D), se agruparon las posiciones de las entidades en la ontología utilizando los algoritmos DB-SCAN y K-means.

## **(II) Fase On-line**

Esta es la etapa en la que se hace la recomendación de cursos. Esto implica dos procesos, un transductor semántico (1) que toma el registro del usuario y genera su perfil, de manera similar a los perfiles de usuario, explicado anteriormente, y el proceso de recomendación híbrido compuesto por tres etapas (2,3 y 4), como se observa en la Figura 4.28.

La primera etapa corresponde a un filtrado semántico (2) para determinar las habilidades propias, afines y las ontológicas; un segundo filtrado basado en contenido (3), para la predicción inicial de cursos, y finalmente, se aplica heurística de filtrado y ordenación (4) para la recomendación final de cursos de LL para la mejora y/o desarrollo de habilidades profesionales con base al registro de un usuario.

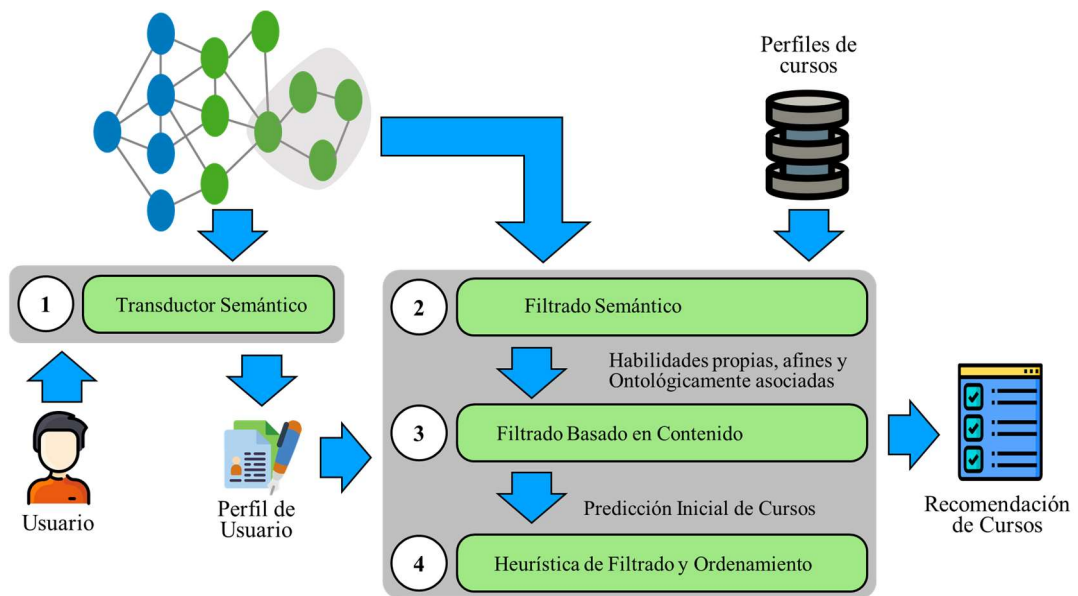


Figura 4.28 Etapa On-line del SR híbrido basado en ontología y ML

Para preparar las recomendaciones de cursos para cada uno de los usuarios, se utiliza el algoritmo general, representado por la Figura 4.29, en donde, 2,3 y 4, corresponde las etapas de recomendación de la Figura 4.28.

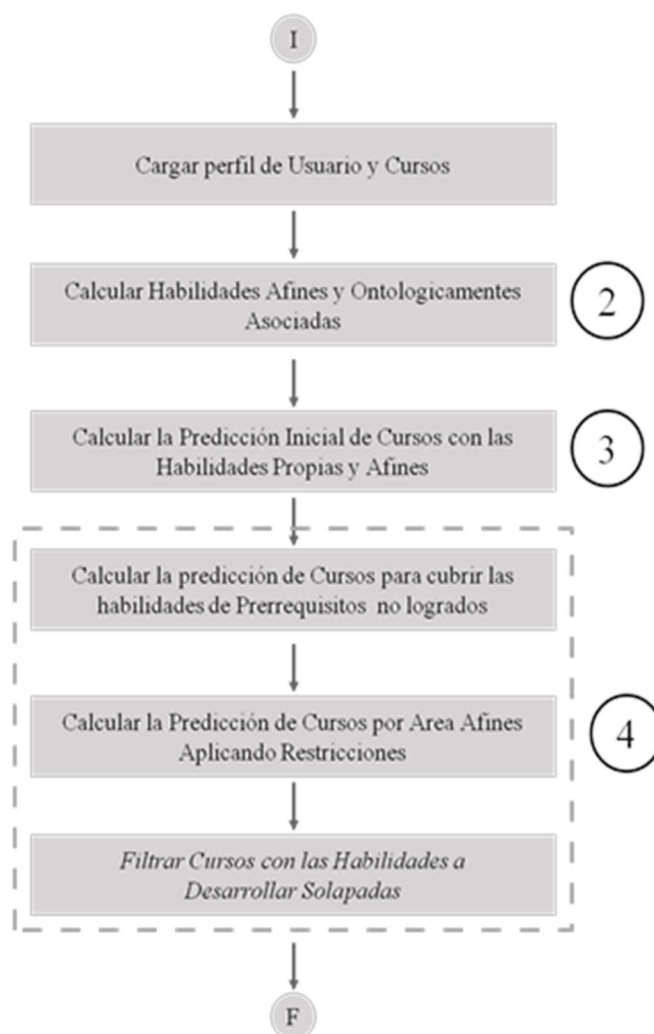


Figura 4.29 Algoritmo de recomendación del SR híbrido basado en ontología y ML

El proceso de recomendación consiste en tres etapas, que se describen a continuación.

## (2) Filtrado semántico para la determinación de las habilidades afines

Con la finalidad de determinar las habilidades afines para un usuario, primero se determinan sus áreas de desempeño laboral afines, a partir de sus habilidades, haciendo uso de la ontología. Para este proceso, se determina el conjunto de sectores de desempeño laboral donde aparecen las habilidades propias del usuario y se realiza un filtrado de estas de la siguiente forma: se seleccionan aquellos sectores de desempeño laboral afines, cuyas habilidades están cubiertas por un porcentaje dado de las habilidades del usuario, y/o se

seleccionan aquellos sectores de desempeño laboral afines cuyas habilidades asociadas cubran un porcentaje de las habilidades propias del usuario. En todo caso, del conjunto de sectores de desempeño laboral afines se excluyen los sectores de desempeño laboral del usuario.

De forma alternativa, a partir de los sectores de desempeño laboral del usuario se puede determinar a qué agrupaciones (clústeres), construidas según los algoritmos de ML, pertenecen. En el caso de un nuevo dato, es decir, un sector laboral de desempeño que no forma parte de ninguna agrupación de las obtenidas durante el entrenamiento, haciendo uso de las habilidades del usuario se puede predecir la agrupación a las que el usuario pertenece. Dada las agrupaciones, el conjunto de los sectores de desempeño laboral que pertenecen a cada una de ellas, conformarán los sectores de desempeño afines del usuario.

La predicción de la agrupación para un nuevo dato, en el caso de DBSCAN, se ubica un nodo núcleo (core) con una distancia menor a un valor épsilon y se asocia al clúster al que este pertenece, para el caso de k-means se ubica el centroide al que se tiene menor distancia y se asocia a ese clúster. En ambos casos se utiliza como distancia la diferencia a uno de la función de similitud, según la ecuación (4.8).

$$d = 1 - Sim(SK_u, SK_c) \quad (4.8)$$

Finalmente, se determina qué sectores de desempeño laboral afines son previos a los sectores de desempeño laboral del usuario, en una relación de orden impuesta por un escalafón y los mismos son removidos de las áreas de desempeño afines. Posteriormente, con sectores de desempeño laboral, propios y afines, se determinan las habilidades afines, que son las habilidades más frecuentes asociadas a sectores de desempeño laboral afines que agregan valor, es decir, que en el conjunto de habilidades no exista la habilidad o en caso contrario, que tenga un mayor nivel de especialización o grado de actualización a las habilidades del usuario, por último, se agregan a las habilidades afines aquellas habilidades que agregan valor y que están bajo la relación “es de interés” en el grafo de conocimiento a las habilidades propias y afines.

Dadas las entradas:

$UJS$ , Conjunto de sectores de desempeño laboral del usuario.

$USK$ , Conjunto de habilidades del usuario.

Para determinar las habilidades afines semánticamente se tiene el siguiente algoritmo:

---



---

**Algoritmo: Determinación semántica de habilidades afines**

---

**Entrada:**  $UJS, USK$

---

- 1:  $RJS \leftarrow \emptyset$
- 2:  $RSK' \leftarrow \emptyset$
- 3: **Para**  $sk \in UKS$  **hacer**
- 4:      $RJS \leftarrow RJS \cup \text{Onto\_GetJobs}(js)$
- 5:  $RJS \leftarrow RJS - UJS$
- 6:  $RJS \leftarrow \text{Onto\_CutOff\_Jobs}(RJS, USK, ujkP, jskP)$
- 7:  $RJS \leftarrow RJS - \text{Onto\_Previous}(RJS, UJS)$
- 8: **Para**  $js \in RJS$  **hacer**
- 9:      $RSK' \leftarrow RSK' \cup \{sk, SV(USK, sk) \wedge sk \in \text{Onto\_Skills}(js)\}$
- 10: **Para**  $sk \in (USK \cup MS2S(RSK, \varphi))$  **hacer**
- 11:      $RSK' \leftarrow RSK' \cup \{sk', SV(USK, sk') \wedge sk' \in \text{Onto\_IntSkills}(sk)\}$
- 12:  $RSK \leftarrow MS2S(RSK', \lambda)$

---

**Salida:**  $RSK$

---

De forma alterna, para determinar las habilidades afines utilizando el clustering de ML se tiene el siguiente algoritmo:

---



---

**Algoritmo: Determinación de habilidades afines usando agrupamiento de ML**

---

**Entrada:**  $UJS, USK$

---

- 1:  $RJS \leftarrow \emptyset$
- 2:  $RSK' \leftarrow \emptyset$
- 3: **Para**  $js \in UJS$  **hacer**
- 4:      $RJS \leftarrow RJS \cup (\text{Onto\_Cluster\_Pred}(js) - \{js\})$
- 5:  $RJS \leftarrow RJS - \text{Onto\_Previous}(RJS, UJS)$
- 6: **Para**  $js \in RJS$  **hacer**
- 7:      $RSK' \leftarrow RSK' \cup \{sk, SV(USK, sk) \wedge sk \in \text{Onto\_Skills}(js)\}$
- 8: **Para**  $sk \in (USK \cup MS2S(RSK, \varphi))$  **hacer**
- 9:      $RSK' \leftarrow RSK' \cup \{sk', SV(USK, sk') \wedge sk' \in \text{Onto\_IntSkills}(sk)\}$
- 10:  $RSK \leftarrow MS2S(RSK', \lambda)$

---

**Salida:**  $RSK$

---

Donde:

$Onto\_Get\_Jobs(sk)$ , realiza sectores de desempeño laboral de desempeño cuyos usuarios tienen la habilidad  $sk$  registrado vía eventos en la ontología, se puede configurar para utilizar habilidades registradas perteneciente a la misma subespecialidad de  $sk$  cuando para la misma no se obtienen sectores de desempeño laboral.

$Onto\_CutOff\_Jobs(jobs, sks, uskp, jskp)$ , realiza el filtrado de los sectores de desempeño laboral,  $js, js \in jobs$ , de acuerdo con un porcentaje  $jskp$  de cobertura del conjunto de habilidades  $sks$  sobre las habilidades asociadas a sector laboral de desempeño  $js, js \in jobs$  o un porcentaje  $skp$  de cobertura de las habilidades del sector  $js, js \in jobs$  sobre el conjunto de habilidades  $sks$

$Onto\_Cluster\_Pred(js)$ , predice el clúster para un sector laboral  $js$  según la agrupación por ML en la ontología.

$Onto\_Previous(RJS, UJS)$ , los sectores previos en  $RJS$  a  $UJS$  según relación de orden dado por el escalafón en la ontología.

$Onto\_Skills(JS_i)$ , son las habilidades de  $JS_i$  representadas en la ontología, producto de actualización vía eventos de los perfiles de usuario.

$Onto\_IntSkills(sk)$ , son las habilidades de interés asociadas a las habilidades  $sk$  representadas en la ontología.

$SV(SK, sk)$ , función que verifica que la habilidad  $sk$  sea de valor para las habilidades  $SK$ .

$MS2S(MS, c)$ , construye un conjunto con los elementos  $c$  más frecuentes en el multiconjunto  $MS$ .

$RSK'$ , Multiconjunto de habilidades afines.

$RSK$ , Conjunto de habilidades afines.

### (3) Filtrado basado en contenido para la predicción Inicial de Cursos

Utilizando un filtrado por contenido, se procede a filtrar del catálogo, los cursos que eleven el nivel de la especialización y/o el grado de actualización de las habilidades propias, o ayuden a desarrollar las habilidades afines del usuario.

Dadas las entradas:

$USK$ , Conjunto de habilidades del usuario.

$RSK$ , Conjunto de habilidades afines del usuario.

$C$ , Perfiles de la oferta de cursos.

Para determinar la predicción inicial de cursos para el usuario se tiene el siguiente algoritmo:

---



---

**Algoritmo: Predicción Inicial de cursos**

---

**Entradas:**  $USK, RSK, C$

---

1:  $CP' \leftarrow \emptyset$

2: **Para**  $c \in C$  **hacer**

3:       **Si**  $\exists s \in USK, \exists s' \in SK2D(c): KA(s) \cong KA(s')$  **entonces**

4:               **Si**  $UG(s) < \omega \vee SL(s) < SL2D(s')$  **entonces**

5:                        $CP' \leftarrow CP' \cup \{c\}$

6:       **Si**  $\exists s \in USK, \exists s' \in SK2D(c): KA(s) \cong KA(s')$  **entonces**

7:               **Si**  $SL(s) \leq SL2D(s')$  **entonces**

8:                        $CP' \leftarrow CP' \cup \{c\}$

9:  $CP \leftarrow MS2S(CP', \delta)$

---

**Salida:**  $CP$

---

Donde:

$SK2D(c)$ , Habilidades a desarrollar por el curso  $c$ .

$KA(s)$ , Área de conocimiento de la habilidad  $s$ .

$SL(s)$ , Nivel de especialización de la habilidad  $s$ .

$UG(s)$ , Grado de actualización de la habilidad  $s$ .

$MS2S(MS, c)$ , construye un conjunto con los  $c$  elementos más frecuentes en el multiconjunto  $MS$ .

$CP'$ , Multiconjunto de predicción de cursos para el usuario.

$CP$  : Predicción de cursos para el usuario.

#### (4) Heurística de Filtrado y Ordenamiento de Cursos para la predicción final

Utilizando heurísticas se complementa y se filtra la predicción inicial de cursos, de la fase anterior. Para el perfil de usuario y catálogo de cursos, se verifica:

1. Qué el usuario tenga las habilidades necesarias para abordar los cursos, en caso contrario, se seleccionan cursos que las desarrollen, aplicando restricciones demográficas.
2. Se aplican las restricciones demográficas de los usuarios a la predicción de cursos inicial y al resultado de la etapa anterior
3. En la predicción de cursos se eliminan los cursos cuyas habilidades a desarrollar sean iguales o un subconjunto de otro, manteniéndose los de mayor valoración.

Dadas las entradas:

$USK$ , Conjunto de habilidades del usuario.

$UDI$ , Información demográfica del usuario.

$CP$ , Predicción inicial de cursos.

$C$ , Perfiles de la oferta de cursos.

Para determinar la recomendación de cursos para el usuario se tiene el siguiente algoritmo:

---

---

**Algoritmo: Recomendación Final de cursos**

---

---

**Entrada:**  $USK, UDI, CP, C$

---

- 1:  $CP' \leftarrow \emptyset$
- 2: **# paso 1**
- 3: **Para**  $c \in CP$  **hacer**
- 4:     **Para**  $sk \in SKP(c)$  **hacer**
- 5:         **Si**  $sk \notin USK \wedge sk \in SK2D(c') \wedge ADR(c', UDI)$  **entonces**
- 6:              $CP' \leftarrow CP' \cup \{c'\}$
- 7: **# paso 2**
- 8:  $CP \leftarrow \{c, c \in CP \wedge ADR(c, UDI)\} \cup MS2S(CP', \phi)$

9: # **paso 3**  
 10: **Para**  $c \in CP \wedge c' \in (CP - \{c\})$  **hacer**  
 11:       **Si**  $S2D(c) \subseteq S2D(c') \wedge CV(c) < CV(c')$  **entonces**  
 12:            $CP \leftarrow CP - \{c\}$   
 13:  $CR \leftarrow (CP)$

---

**Salida:**  $CR$

---

Donde:

$SKP(c)$ , Habilidades prerrequisitos para el curso  $c$ .

$SK2D(c)$ , Habilidades a desarrollar por el curso  $c$ .

$ADR(c, di)$ , Aplica restricciones demográficas  $di$  al curso  $c$ .

$CV(c)$ , Valoración para el curso  $c$ .

$MS2S(MS, c)$ , construye un conjunto con los elementos  $c$  más frecuentes. en el multiconjunto  $MS$ .

$CP'$ , Muticonjunto de predicción de cursos para el usuario.

$CR$ , Recomendación de cursos para el usuario.

### 4.2.3 Sistema de recomendación híbrido basado en ontología y ML con *endorsements*

Para esta última versión del SR, se incorporó como entrada al sistema los *endorsements* de LinkedIn, los cuales se corresponden con las habilidades de los usuarios, que son avaladas por otros miembros de esta red social. Estas son tomadas en cuenta para validar las habilidades determinadas en los perfiles del SR. Para su codificación se utiliza el proceso de codificación de habilidades, modificado, teniendo como entrada el campo *Skills*, el cual se corresponde con los *endorsements* de los registros de LinkedIn.

En la Figura 4.30 se observa la arquitectura del sistema, donde la principal diferencia con la versión anterior se encuentra en la etapa on-line, específicamente en el proceso de

creación del perfil de usuario, que debe incorporar los *endorsements*, y en el proceso filtrado semántico, donde son consideradas, al determinar las habilidades afines y las ontológicamente asociadas.

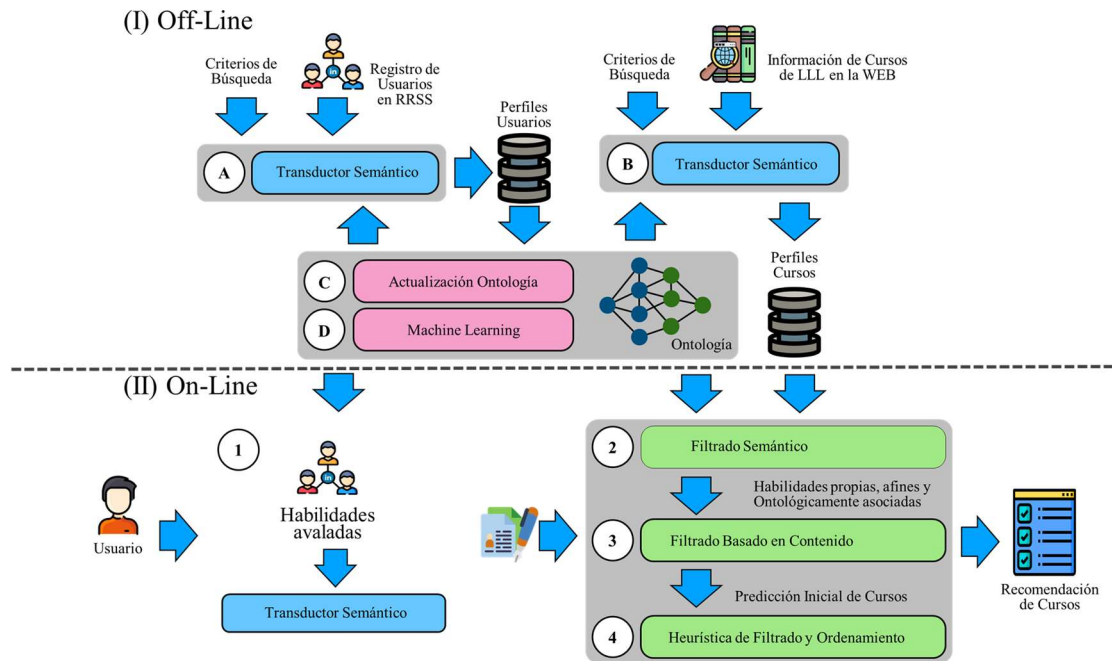


Figura 4.30 Arquitectura del SR híbrido basado en ontología y ML con *endorsements*

El algoritmo general de recomendación es el mismo de la versión anterior, para hacer uso de *endorsements*, se modifican los algoritmos en el filtrado semántico (2):

Dadas las entradas:

$UJS$ , Conjunto de sectores de desempeño laboral del usuario.

$USK$ , Conjunto de habilidades del usuario.

$UESK$ , Conjunto de habilidades validadas del usuario.

Para determinar las habilidades afines semánticamente se tiene el siguiente algoritmo:

---

---

**Algoritmo: Determinación semántica de habilidades afines**

---

**Entrada:**  $UJS, USK, UESK$ 

---

```
1:  $RJS \leftarrow \emptyset$ 
2:  $RSK' \leftarrow \emptyset$ 
3: Para  $sk \in UKS$  hacer
4:    $RJS \leftarrow RJS \cup \text{Onto\_GetJobs}(js)$ 
5:  $RJS \leftarrow RJS - UJS$ 
6:  $RJS \leftarrow \text{Onto\_CutOff\_Jobs}(RJS, USK, ujkP, jskP)$ 
7:  $RJS \leftarrow RJS - \text{Onto\_Previous}(RJS, UJS)$ 
8: Para  $js \in RJS$  hacer
9:    $RSK' \leftarrow RSK' \cup \{sk, SV(USK \cup UESK, sk) \wedge sk \in \text{Onto\_Skills}(js)\}$ 
10: Para  $sk \in (USK \cup MS2S(RSK, \varphi))$  hacer
11:    $RSK' \leftarrow RSK' \cup \{sk', SV(USK \cup UESK, sk') \wedge sk' \in \text{Onto\_IntSkills}(sk)\}$ 
12:  $RSK \leftarrow MS2S(RSK', \lambda)$ 
```

---

**Salida:**  $RSK$ 

---

De forma alterna, al igual que sistema anterior, para determinar las habilidades afines utilizando el clustering de ML se tiene el siguiente algoritmo:

---

---

**Algoritmo: Cálculo de habilidades afines**

---

**Entrada:**  $UJS, USK, UESK$ 

---

```
1:  $RJS \leftarrow \emptyset$ 
2:  $RSK' \leftarrow \emptyset$ 
3: Para  $js \in UJS$  hacer
4:    $RJS \leftarrow RJS \cup (\text{Onto\_Cluster\_Pred}(js) - \{js\})$ 
5:  $RJS \leftarrow RJS - \text{Onto\_Previous}(RJS, UJS)$ 
6: Para  $js \in RJS$  hacer
7:    $RSK' \leftarrow RSK' \cup \{sk, SV(USK \cup UESK, sk) \wedge sk \in \text{Onto\_Skills}(js)\}$ 
8: Para  $sk \in (USK \cup MS2S(RSK, \varphi))$  hacer
9:    $RSK' \leftarrow RSK' \cup \{sk', SV(USK \cup UESK, sk') \wedge sk' \in \text{Onto\_IntSkills}(sk)\}$ 
10:  $RSK \leftarrow MS2S(RSK', \lambda)$ 
```

---

**Salida:**  $RSK$ 

---

Para las etapas posteriores de Predicción Inicial de Cursos (3) y Heurística de Filtrado y Ordenamiento de Cursos (4), se utilizan los mismos algoritmos, agregando a al conjunto de las habilidades del usuario sus habilidades validadas, es decir:

$$USK \leftarrow USK \cup UESK$$

### 4.3 Resumen del Capítulo

Este capítulo presenta las estrategias propuestas para abordar el problema, así como el diseño a alto y bajo nivel del sistema de recomendación para cursos de formación continua. A partir de los registros de usuarios de la red social LinkedIn, e información de cursos de LL en internet, se utilizan transductores semánticos para construir los perfiles de usuarios y cursos. Para codificar los sectores de desempeño laboral y las áreas de conocimiento de las habilidades, que forman parte de los perfiles, se propone un código jerárquico basado en una taxonomía, lo que evita la pérdida de información propia de los códigos planos.

A partir de los perfiles se realizan varias propuestas de sistemas de recomendación híbridos, el primero hace uso de taxonomías y bajo una filosofía de datos almacenados, se definen cuatro fases para el proceso de recomendación: dos fases de filtrado colaborativo, una fase de filtrado por contenido y una heurística para refinar la predicción de cursos. Al evaluar el sistema se tienen buenos resultados, y a partir de las oportunidades de mejora, se propone un sistema basado en modelo, el cual hace uso de una ontología y algoritmos de ML, no supervisado, para agrupar entidades en la ontología. Este sistema tiene tres fases: un filtrado semántico, un filtrado de contenido y una heurística. Frente a las taxonomías, el uso de una ontología permite una mejor representación del dominio abordado, al incorporar nuevas entidades, así como ofrecer mecanismos de inferencias. Siendo esta propuesta más eficiente en términos computacionales, dado el orden de los algoritmos.

Finalmente, al utilizar la red social LinkedIn, hace posible incluir las habilidades avaladas por otros usuarios (*endorsements*), por tanto, se propone en una última versión, utilizando estos datos, para realizar mejoras en los filtrados del sistema de recomendación propuesto.

# Capítulo 5

## Resultados

Para probar la hipótesis planteada en esta tesis, se han propuesto tres versiones del SR basados en las características de los datos de los perfiles de usuarios extraídos de LinkedIn, con el objetivo de recomendar cursos de formación continua que mejoren y/o desarrollen las habilidades profesionales de los usuarios. A continuación, se presentan los resultados de la evaluación de las distintas propuestas.

Para realizar la evaluación de SR, se seleccionó la evaluación off-line, que como se mencionó en el estado del arte, permite evaluar los diferentes algoritmos y enfoques, la cual es utilizada en ambientes de experimentación, dado que, se quiere la mayor consistencia para poder comparar el desempeño de las diferentes propuestas bajo las mismas condiciones. Así mismo, la evaluación off-line, cuenta con métricas diseñadas que pueden reflejar la eficacia del sistema desde la perspectiva del usuario y proporcionan una evaluación aceptada ampliamente debido a robustez de las métricas utilizadas.

Al seleccionar este tipo de pruebas, no fue necesario construir una interfaz de usuario final, y permitió tener consistencia en la preferencia de los usuarios, en las repetidas o múltiples pruebas.

En los métodos fuera de línea, son importantes una serie de medidas para poder tener una evaluación completa y precisa del SR. Para el cálculo de las métricas, para la evaluación de las diferentes propuestas de sistemas de recomendación, es necesario disponer de las calificaciones del usuario en función del universo de cursos, en tres categorías disjuntas: preferencia, novedad y serendipia, que como se indicó en el estado del arte, evalúan si el sistema predice correctamente lo que el usuario seleccionaría. En estas categorías, resalta la novedad, que se refiere a aquellos cursos recomendados, que el usuario no conocía, y la serendipia, que hace referencia a aquellos cursos que se encuentran, como por casualidad y parecieran no ser de interés, pero una vez recomendados, se verifica que son valiosos para el usuario.

Con base a las definiciones del estado del arte, para la evaluación del SR, se define:

**Selección**, conjunto total de curso seleccionados por el usuario.

**Preferencia**, conjunto de cursos que se corresponden a los que el usuario espera recibir en la recomendación.

**Novedad**, conjunto de cursos novedosos al usuario.

**Serendipia**, conjunto de cursos no esperados, pero que le resultan útiles

En donde:

$$Selección \equiv Preferencia \cup Novedad \cup Serendipia$$

$$Preferencia \cap Novedad \cap Serendipia = \emptyset$$

En función de los conjuntos antes mencionados, y de la recomendación dada a cada usuario, se calculan las métricas de error medio absoluto (MAE), la raíz del error medio al cuadrado (RMSE), cobertura, precisión, *recall*, novedad y serendipia, para la evaluación de las diferentes propuestas de los SR.

En la Tabla 5.1 se definen las métricas, tanto para cada usuario, como para el sistema, de acuerdo con las definiciones presentadas en el estado del arte.

Para un usuario  $u_i \in U$ , se tiene:

$CR_i$ : el conjunto de recomendaciones de  $u_i$  realizadas por el sistema.

$CE_i$ : el conjunto de preferencia de  $u_i$ .

Tabla 5.1. Métricas de evaluación para sistemas de recomendación

Métrica	Cálculo para cada usuario	Cálculo para el sistema
MAE	$MAE_i = \frac{ (CE_i \cup CR_i) - (CE_i \cap CR_i) }{ CE_i \cup CR_i }$	$MAE = \frac{\sum_{i=1}^n  (CE_i \cup CR_i) - (CE_i \cap CR_i) }{\sum_{i=1}^n  CE_i \cup CR_i }$
RMSE	$RMSE_i = \sqrt{\frac{ (CE_i \cup CR_i) - (CE_i \cap CR_i) ^2}{ CE_i \cup CR_i }}$	$RMSE = \sqrt{\frac{\sum_{i=1}^n  (CE_i \cup CR_i) - (CE_i \cap CR_i) ^2}{\sum_{i=1}^n  CE_i \cup CR_i }}$
Cobertura	$C(u_i) =  \{u_i \in U, CR_i \neq \emptyset\} $	$C(S) = \frac{ \{u_i \in U, CR_i \neq \emptyset\} }{ U }$
Precisión	$P(u_i) = \frac{ \{c \in CR_i, relevant(c, u_i)\} }{ CR_i }$	$P(S) = \frac{1}{ U } \sum_{u_i \in U} P(u_i)$
Recall	$R(u_i) = \frac{ CE_i - CR_i }{ CE_i }$	$R(S) = \frac{1}{ U } \sum_{u_i \in U} R(u_i)$
Novedad	$N(u_i) = \frac{ CR_i - CE_i }{ CR_i }$	$N(S) = \frac{1}{ U } \sum_{u_i \in U} N(u_i)$
Serendipia	$S(u_i) = \frac{ \{c \in CR_i, serendipity(c, u_i)\} }{ CR_i }$ donde: $\{c \in CR_i, serendipity(c, u_i)\} \subseteq (CR_i - CE_i)$	$S(S) = \frac{1}{ U } \sum_{u_i \in U} S(u_i)$

La evaluación de las propuestas se realiza en lotes (batch), con un conjunto de datos que contiene los perfiles, tanto de usuarios como de cursos, así como las preferencias de usuarios, estas fueron tomadas de una encuesta realizada, con un formulario de Google, en donde, se solicitaba clasificar una lista de cursos en las categorías de: deseable, preferido, novedoso y serendipia, para conocer las preferencias de los usuarios en cuanto a la elección de cursos de LL.

A continuación, se presenta la evaluación y resultados de cada una de las versiones del SR.

## 5.1 Resultados del sistema de recomendación híbrido basado en taxonomía.

El SR híbrido basado en taxonomía puede ser configurado para diferentes funciones de similitud, como la similitud del coseno y la similitud de Pearson, así como configurar diferentes umbrales de corte al aplicar la función de similitud taxonómica. Esto permitió hacer pruebas con diferentes funciones de similitud y parámetros.

Con esta primera versión se pudo demostrar la viabilidad de la propuesta y determinar las configuraciones más eficaces. Fue de particular interés probar diferentes funciones de similitud y observar la ganancia al utilizar una codificación plana versus una jerárquica, con umbral de corte menor a uno, (Urdaneta, Méndez, & Oleagordia, 2020). Así mismo, los resultados obtenidos se convierten en la línea base para la evaluación de las propuestas basadas en ontologías y ML.

Para evaluar la propuesta, se realizó la recomendación, configurando el SR con una codificación plana y con una configuración jerárquica, para el cálculo de la similitud de las habilidades de los usuarios, a fin de determinar los sectores laborales afines.

Con la finalidad de determinar los porcentajes de similitud en donde se obtiene la mejor recomendación, se realizaron múltiples ejecuciones del SR para la codificación plana, que corresponde a un umbral de corte de 1,0, configurando los parámetros  $\mu$  de las funciones de similitud de coseno y Pearson entre 0,5 y 1,0. Estos resultados se pueden observar en la Tabla 5.2 y la Tabla 5.3.

Tabla 5.2. Métricas para el SR híbrido basado en taxonomía con codificación plana utilizando similitud de coseno, para diferentes valores de  $\mu$

$\mu$	MAE	RMSE	Cobertura	Precisión	Recall	Novedad	Serendipia
1,0	0,47	7,57	0,82	0,94	0,50	0,26	0,07
0,9	0,47	7,57	0,82	0,94	0,50	0,26	0,07
0,8	0,47	7,57	0,82	0,94	0,50	0,26	0,07

<b>0,7</b>	0,27	4,38	0,91	0,91	0,74	0,46	0,07
<b>0,6</b>	<b>0,25</b>	<b>4,15</b>	<b>0,91</b>	<b>0,89</b>	<b>0,77</b>	<b>0,47</b>	<b>0,07</b>
<b>0,5</b>	0,27	4,62	0,91	0,87	0,77	0,45	0,05

Tabla 5.3. Métricas para el SR híbrido basado en taxonomía con codificación plana utilizando similitud de Pearson, para diferentes valores de  $\mu$

$\mu$	MAE	RMSE	Cobertura	Precisión	Recall	Novedad	Serendipia
<b>1,0</b>	0,47	7,57	0,82	0,94	0,50	0,26	0,07
<b>0,9</b>	0,47	7,57	0,82	0,94	0,50	0,26	0,07
<b>0,8</b>	0,47	7,57	0,82	0,94	0,50	0,26	0,07
<b>0,7</b>	0,47	7,57	0,82	0,94	0,50	0,26	0,07
<b>0,6</b>	0,47	7,57	0,82	0,94	0,50	0,26	0,07
<b>0,5</b>	0,47	7,57	0,82	0,94	0,50	0,26	0,07

Para explotar las ventajas de trabajar con una codificación jerárquica en el cálculo de similitud de las habilidades de los usuarios, para determinar los sectores laborales afines, y medir el rendimiento del SR, se realizaron pruebas con un umbral de corte de 0,8 para el cálculo de la similitud taxonómica para las áreas de conocimiento, en el cálculo de los vectores  $u$  y  $v$ . Este umbral de 0,8 corresponde al nivel de subespecialidad, ya que, en el caso de no tener concordancia a nivel de conocimiento específico, a nivel de subespecialidad se comparte conocimiento en el área.

Así como se realizó para la codificación plana, se realizaron múltiples ejecuciones del SR, configurando los parámetros  $\mu$  de las funciones de similitud de coseno y Pearson entre 0,5 y 1,0, para la codificación jerárquica. Estos resultados se pueden observar en la Tabla 5.4 y la Tabla 5.5.

Tabla 5.4. Métricas para el SR híbrido basado en taxonomía con codificación jerárquica utilizando similitud de coseno, para diferentes valores de  $\mu$

$\mu$	MAE	RMSE	Cobertura	Precisión	Recall	Novedad	Serendipia
<b>1,0</b>	0,47	7,57	0,82	0,94	0,50	0,26	0,07
<b>0,9</b>	0,47	7,57	0,82	0,94	0,50	0,26	0,07
<b>0,8</b>	0,40	6,46	0,82	0,94	0,55	0,29	0,07
<b>0,7</b>	<b>0,17</b>	<b>2,74</b>	<b>0,91</b>	<b>0,91</b>	<b>0,81</b>	<b>0,51</b>	<b>0,07</b>
<b>0,6</b>	0,18	3,07	0,91	0,89	0,81	0,49	0,07
<b>0,5</b>	0,19	3,18	0,91	0,87	0,83	0,50	0,04

Tabla 5.5 Métricas para el SR híbrido basado en taxonomía con codificación jerárquica utilizando similitud de Pearson, para diferentes valores de  $\mu$

$\mu$	MAE	RMSE	Cobertura	Precisión	Recall	Novedad	Serendipia
<b>1,0</b>	0,47	7,57	0,82	0,94	0,50	0,26	0,07
<b>0,9</b>	0,47	7,57	0,82	0,94	0,50	0,26	0,07
<b>0,8</b>	0,47	7,57	0,82	0,94	0,50	0,26	0,07
<b>0,7</b>	0,47	7,57	0,82	0,94	0,50	0,26	0,07
<b>0,6</b>	0,47	7,57	0,82	0,94	0,50	0,26	0,07
<b>0,5</b>	0,47	7,57	0,82	0,94	0,50	0,26	0,07

A partir de estos resultados, se graficó el error medio absoluto (MAE) (Figura 5.1) y la media armónica entre la precisión y *recall* (Figura 5.2) para los diferentes valores de  $\mu$  para las funciones Pearson y coseno, tanto para la codificación plana y jerárquica.

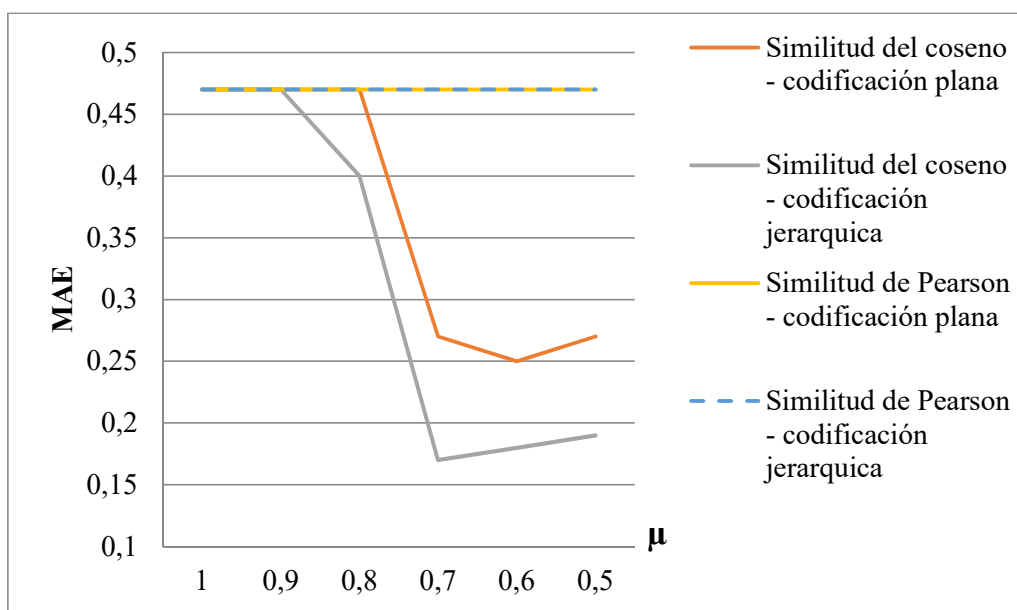


Figura 5.1 MAE para los diferentes valores de  $\mu$  y las diferentes configuraciones de similitud y codificación.

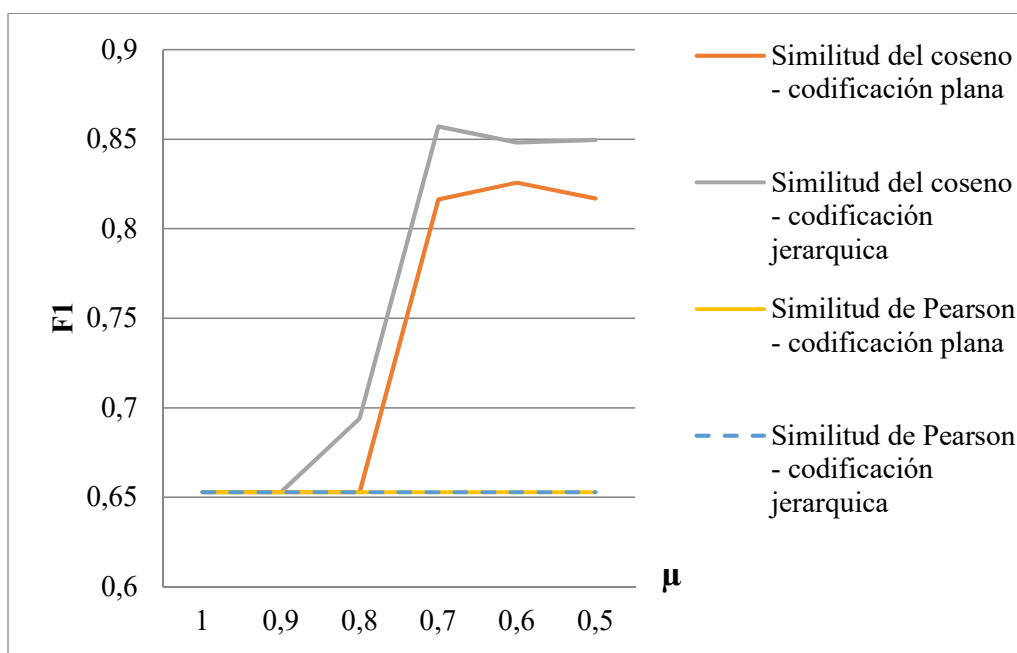


Figura 5.2  $F_1$  para los diferentes valores de  $\mu$  y las diferentes configuraciones de similitud y codificación.

Para la selección del  $\mu$  (similitud entre las habilidades) se busca minimizar el error (el MAE y el RMSE) y maximizar la precisión y el *recall*, para lo cual se utiliza la media armónica ( $F_1$ ) entre ambos, de las gráficas 5.1 y 5.2 se observa que el mejor desempeño se obtiene con  $\mu = 0,7$  para una codificación jerárquica.

Para  $\mu = 0,7$  y un umbral de corte de 1,0 para la construcción de los vectores  $u$  y  $v$ , para el cálculo de la similitud taxonómica para las áreas de conocimiento, en la Figura 5.3 se muestra el ejemplo del resultado de una recomendación para un usuario en particular.

```

Descripción
  Dev iOS
  Sector Laboral
  001-03-07:03-02
  Habilidades
    ('20-10-01-*:01', 3, 2)
    ('10-06-01-9:02', 3, 2)
    ('10-07-09-2:02', 3, 2)
  Sectores Afines
  Habilidades Afines:
    ('20-10-01-*:01', 4, 3)
    ('10-06-01-9:02', 3, 3)
    ('10-07-09-2:03', 3, 2)
    ('10-07-09-2:04', 3, 2)
    ('20-10-01-*:01', 4, 3)
    ('10-06-01-5:03', 4, 2)
  Predicción Inicial de cursos
  [2, 3, 4, 5, 6]
  Recomendación de cursos
  [2, 3, 4]

```

Figura 5.3. Ejemplo de salida del SR híbrido basado en taxonomía con configuración plana

Finalmente, las métricas asociadas al ejecutar el sistema, considerando el umbral de corte  $\mu$  de 0,7 y un umbral de corte de 1,0 para las funciones de similitud de coseno y Pearson, con el conjunto de entrenamiento se muestran en la Tabla 5.6.

Tabla 5.6. Métricas para el SR híbrido basado en taxonomía configurado con codificación plana

Similitud	MAE	RMSE	Cobertura	Precisión	Recall	Novedad	Serendipia
<b>Coseno (S1.1)</b>	<b>0,27</b>	<b>4,38</b>	<b>0,91</b>	<b>0,91</b>	<b>0,74</b>	<b>0,46</b>	<b>0,07</b>
<b>Pearson (S1.2)</b>	0,47	7,57	0,82	0,94	0,50	0,26	0,07

Así mismo, se ejecutó el SR con una configuración jerárquica, para el cálculo de la similitud taxonómica para las áreas de conocimiento, en el cálculo de los vectores u y v. En la figura 5.4 se muestra la salida de una recomendación para un usuario en particular, con configuración jerárquica.

```

Descripción
  Dev iOS
  Sector Laboral
  001-03-07:03-02
  Habilidades
  ('20-10-01-*:01', 3, 2)
  ('10-06-01-9:02', 3, 2)
  ('10-07-09-2:02', 3, 2)
  Sectores Afines
  000-03-07:03-00
  Habilidades Afines:
  ('10-06-01-9:02', 3, 3)
  ('20-10-01-*:01', 4, 3)
  ('10-07-09-2:03', 3, 2)
  ('10-07-09-2:04', 3, 2)
  ('20-10-02-*:01', 3, 2)
  ('10-06-01-5:05', 3, 3)
  ('10-07-09-2:02', 3, 3)
  ('20-10-01-*:01', 4, 2)
  ('10-06-01-5:03', 4, 2)
  Predicción Inicial de cursos
  [2, 3, 4, 5, 6, 7, 8, 9, 10, 11]
  Recomendación de cursos
  [2, 3, 4, 7, 8, 9]
  
```

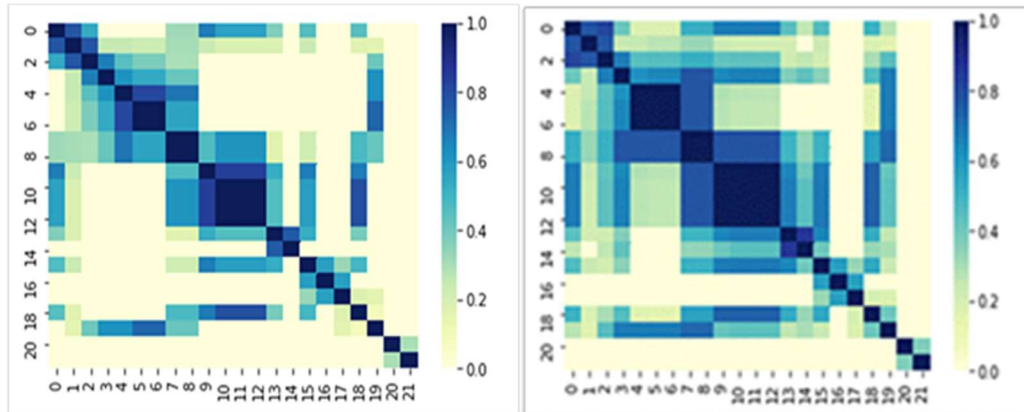
Figura 5.4 Ejemplo de salida SR híbrido basado en taxonomía con configuración jerárquica

El resultado del cálculo de las métricas, utilizando codificación jerárquica, y con parámetro en las funciones de similitud de coseno y Pearson de 0,7, se muestra en la Tabla 5.7.

Tabla 5.7 Métricas para el SR híbrido basado en taxonomía configurado con codificación jerárquica

<b>Similitud</b>	<b>MAE</b>	<b>RMSE</b>	<b>Cobertura</b>	<b>Precisión</b>	<b>Recall</b>	<b>Novedad</b>	<b>Serendipia</b>
<b>Coseno (S1.3)</b>	<b>0,17</b>	<b>2,74</b>	<b>0,91</b>	<b>0,91</b>	<b>0,81</b>	<b>0,51</b>	<b>0,07</b>
<b>Pearson (S1.4)</b>	0,47	7,57	0,82	0,94	0,50	0,26	0,07

Con la finalidad de observar la ganancia al utilizar una codificación jerárquica sobre una plana, haciendo uso de una porción del conjunto de entrenamiento, se ejecutó el sistema y se graficaron los resultados, con umbrales de corte de 1,0 y 0,8 de similitud taxonómica en el cálculo de los vectores  $u$  y  $v$ , obteniéndose como resultado las matrices de similitud que se muestran en las figuras 5.5 (a) y 5.5 (b).



(a) Codificación plana

(b) Codificación jerárquica

Figura 5.5 Matriz de similitud del coseno para una codificación plana

Se observó que, cuando se utiliza un umbral de corte menor a 1,0, que corresponde a una codificación jerárquica, resulta una mayor similitud. Esto es debido a que la similitud se puede establecer a distintos niveles de la taxonomía, permitiendo tomar en cuenta conocimiento que es común en la jerarquía.

El mejor desempeño del sistema en esta primera etapa se obtiene para  $\mu = 0,7$  (70% de coincidencia para establecer similitud entre las habilidades) y un umbral de corte de 0,8 para la construcción de los vectores  $u$  y  $v$  (codificación jerárquica), para el cálculo de la similitud taxonómica para las áreas de conocimiento, resultados que se tomarán como línea base para evaluar las siguientes versiones, que, al estar basadas en los algoritmos propuestos en esta etapa, se utilizan los mismos valores en los parámetros comunes para los efectos de configuración.

## **5.2 Resultados del sistema de recomendación híbrido basado en ontología y ML**

Para la evaluación del SR híbrido basado en ontología y ML, se ejecutó la fase off-line del sistema, en donde se cargaron los datos en la base de datos en Neo4j, previamente poblada con las definiciones del sustrato de la ontología. Esta, se actualizó vía eventos con el conjunto de datos de entrenamiento y se determinaron los clústeres de los sectores de desempeño laboral, utilizando los algoritmos de ML, k-Means y DBSCAN.

Con la finalidad de evaluar el rendimiento de las diferentes características de las implementaciones del SR híbrido basado en ontología y ML, y determinar las configuraciones más eficaces, se calcularon las métricas para cada una de las siguientes configuraciones:

**S2.1:** Filtrado de contenido utilizando solo las habilidades propias, no se determinan habilidades afines, aunque se consideran las habilidades de interés.

**S2.2:** Filtrado colaborativo utilizando solo sectores de desempeño laboral propios, se determinan habilidades afines semánticamente.

**S3.3:** Filtrado semántico utilizando reglas para determinar los sectores de desempeño laboral afines y habilidades afines.

**S2.4:** Filtrado semántico utilizando una cobertura de 75% de las habilidades usuarios para determinar sectores de desempeño laboral afines. Es decir, se seleccionaron como sectores de desempeño laboral afines, aquellos cargos que cubren el 75% de las habilidades del usuario.

**S2.5:** Filtrado semántico utilizando cobertura de 50% de las habilidades del sector de desempeño laboral para determinar sectores de desempeño laboral afines. Es decir, aquellos sectores de desempeño laboral los cuales las habilidades del usuario cubren el 50% de las habilidades asociadas al cargo.

**S2.6:** Filtrado semántico utilizando agrupación de DBSCAN con  $\varepsilon = 0,3$ , para determinar sectores de desempeño laboral afines y reglas semánticas para determinar las habilidades afines.

**S2.7:** Filtrado semántico utilizando agrupación de k-Means para determinar sectores de desempeño laboral afines y reglas semánticas para determinar las habilidades afines.

**S2.8:** Filtrado semántico utilizando agrupación de DBSCAN y k-Means para determinar sectores de desempeño laboral afines y reglas semánticas para determinar las habilidades afines.

Se ejecutaron las diferentes configuraciones del SR de la etapa on-line. En la Figura 5.6 se puede observar un ejemplo de la salida del sistema, haciendo uso de la predicción basada en agrupaciones, obtenidas por el algoritmo k-means, para determinar los sectores de desempeño laboral afines.

Cada una de las configuraciones antes descritas fueron ejecutadas con los diferentes conjuntos de datos: entrenamiento, pruebas y total, obteniéndose los siguientes resultados. Los resultados, utilizando los datos de entrenamiento (70% del conjunto de datos) para validar el desempeño del modelo se muestran en la Tabla 5.8.

Sistema Híbrido de Recomendación							
Utiliza cargos obtenidos de agrupamiento por ML / K-Means							
Salida del Recomendador Semántico							
Sectores Laborales del Usuario:							
001-03-07:03-02: iOS Nativo							
Habilidades del usuario:							
('20-10-01-*:01', 4, 2) : Xcode							
('10-06-01-5:03', 4, 2) : ObjectiveC							
('10-07-09-2:01', 3, 2) : MVC							
('20-09-02-*:00', 3, 3) : Frontend							
Sectores Laborales Previos:							
Sectores Laborales Afines:							
001-03-08:00-04 : Revisor Código							
001-03-07:03-02 : iOS Nativo							
001-03-07:03-00 : Aplicaciones Móviles							
Habilidades Afines:							
('10-06-01-9:02', 3, 3) : Swift							
('10-07-09-2:03', 3, 2) : MVVM							
('10-07-09-2:04', 3, 2) : VIPER							
('10-07-05-*:01', 3, 2) : Mantis							
('10-06-01-5:05', 3, 3) : Kotlin							
('10-07-09-2:02', 3, 3) : MVP							
('20-10-02-*:01', 3, 2) : AndroidStudio							
('10-06-01-5:03', 4, 3) : ObjectiveC							
Habilidades Asociadas							
('10-07-07-*:00', 0, 0) : Gestión de versiones							
('20-09-02-1:00', 0, 0) : Progresivas							
Salida del Filtrado por Contenido (Predicción inicial)							
7 Desarrollo en Android con Kotlin							
2 Desarrollo en iOS con Swift							
3 Fundamentos de Programación con Swift							
4 Introducción a la Programación con Swift - Bilbao							
5 Introducción a la Programación con Swift - Barcelona							
6 Desarrollo Avanzado en Android							
8 Fundamentos de Programación con Kotlin							
9 introducción a la Programación con Kotlin - Bilbao							
10 Introducción a la Programación con Kotlin - Barcelona							
15 Gestión de Versiones con Gif							
Salida de la heurística (recomendación de cursos)							
7 Desarrollo en Android con Kotlin							
2 Desarrollo en iOS con Swift							
3 Fundamentos de Programación con Swift							
6 Desarrollo Avanzado en Android							
8 Fundamentos de Programación con Kotlin							
15 Gestión de Versiones con Gif							

Figura 5.6 Ejemplo de salida SR híbrido basado en ontología y ML

Tabla 5.8 Métricas para las diferentes configuraciones del SR híbrido basado en ontología y ML para el conjunto de datos de entrenamiento.

Configuración	MAE	RMSE	Cobertura	Precisión	Recall	Novedad	Serendipia
S2.1	0,58	9,13	0,64	1,00	0,36	0,14	0,09
S2.2	0,51	8,22	0,73	0,97	0,41	0,21	0,07
S2.3	<b>0,25</b>	<b>4,44</b>	<b>0,91</b>	<b>0,83</b>	<b>0,82</b>	<b>0,46</b>	<b>0,07</b>
S2.4	0,22	3,65	0,82	0,93	0,71	0,45	0,07
S2.5	0,38	6,62	0,91	0,80	0,74	0,39	0,07
S2.6	<b>0,18</b>	<b>2,92</b>	<b>0,91</b>	<b>0,91</b>	<b>0,80</b>	<b>0,52</b>	<b>0,07</b>
S2.7	0,49	9,25	0,91	0,67	0,63	0,31	0,04
S2.8	0,30	5,24	0,95	0,80	0,80	0,48	0,04

Luego de entrenado el modelo con el 70% de los datos, se ejecutan el SR bajo las diferentes configuraciones con el total de las muestras del conjunto de datos. Los resultados se muestran en la Tabla 5.9.

Tabla 5.9 Métricas para las diferentes configuraciones del SR híbrido basado en ontología y ML para el total de la muestra del conjunto datos.

Configuración	MAE	RMSE	Cobertura	Precisión	Recall	Novedad	Serendipia
S2.1	0,59	11,05	0,61	0,99	0,37	0,11	0,06
S2.2	0,50	9,49	0,71	0,97	0,45	0,17	0,05
S2.3	<b>0,25</b>	<b>5,16</b>	<b>0,94</b>	<b>0,82</b>	<b>0,87</b>	<b>0,39</b>	<b>0,06</b>
S2.4	0,23	4,49	0,81	0,94	0,73	0,39	0,05
S2.5	0,34	7,03	0,87	0,84	0,75	0,34	0,06
S2.6	<b>0,27</b>	<b>5,63</b>	<b>0,94</b>	<b>0,84</b>	<b>0,86</b>	<b>0,45</b>	<b>0,05</b>
S2.7	0,51	11,72	0,94	0,62	0,65	0,25	0,04
S2.8	0,35	7,64	0,97	0,75	0,86	0,42	0,03

Por último, entrenado el modelo con el 70% de los datos, se ejecutó el sistema, bajo las diferentes configuraciones, utilizando la porción de prueba (30%) de los datos. Los resultados se muestran en la Tabla 5.10.

Tabla 5.10 Métricas para las diferentes configuraciones del SR híbrido basado en ontología y ML para el conjunto datos de pruebas.

Configuración	MAE	RMSE	Cobertura	Precisión	Recall	Novedad	Serendipia
S2.1	0,62	6,24	0,56	0,96	0,41	0,04	0,00
S2.2	0,47	4,75	0,67	0,96	0,56	0,07	0,00
S2.3	<b>0,23</b>	<b>2,64</b>	<b>1,00</b>	<b>0,80</b>	<b>1,00</b>	<b>0,24</b>	<b>0,04</b>
S2.4	0,26	2,63	0,78	0,96	0,78	0,26	0,00
S2.5	0,25	2,60	0,78	0,94	0,78	0,21	0,02
S2.6	<b>0,43</b>	<b>5,42</b>	<b>1,00</b>	<b>0,67</b>	<b>1,00</b>	<b>0,28</b>	<b>0,00</b>

<b>S2.7</b>	0,56	7,24	1,00	0,49	0,69	0,12	0,06
<b>S2.8</b>	0,45	5,79	1,00	0,62	1,00	0,28	0,00

En todos los casos de pruebas, se realizó un filtrado de los sectores de desempeño laboral afines para descartar los cargos con menor nivel al actual, según la relación de orden impuesta por un escalafón. Así mismo, las habilidades afines fueron enriquecidas con las habilidades de interés descritas en la ontología, se utilizó la similitud del coseno, con umbral de corte  $\mu = 0,7$  y en la construcción de los vectores  $u$  y  $v$  se utilizó una función de similitud taxonómica con un umbral de corte a 0,8, partiendo del hecho de que, en la versión anterior del SR, el mejor desempeño se obtuvo con esa configuración.

De los resultados obtenidos se observa que las configuraciones con mejor desempeño para todos los conjuntos de datos son, el filtrado semántico utilizando reglas para determinar los sectores de desempeño laboral afines y habilidades afines (S2.3), y el filtrado semántico utilizando agrupación de DBSCAN (S2.6) con  $\varepsilon = 0,3$ , medida de disimilitud (distancia) calculada como complemento 0,7 correspondiente al valor de  $\mu$ , para determinar sectores de desempeño laboral afines y reglas semánticas para determinar las habilidades afines.

### **5.3 Resultados del sistema de recomendación híbrido basado en ontología y ML con *endorsements***

Para la evaluación de esta versión, hay que recordar, que el sistema tiene en común con la propuesta anterior, la etapa fuera de línea, con la diferencia de que los perfiles de usuarios contienen los *endorsements*, y estos son tomados en cuenta en las diferentes etapas de recomendación. En la Figura 5.7 se puede observar un ejemplo de la salida al ejecutar el sistema haciendo uso de la predicción basada en agrupaciones obtenidas por el algoritmo DBSCAN para determinar los sectores de desempeño laboral afines.

De forma similar, con la finalidad de evaluar el rendimiento de las características particulares, de las implementaciones del SR híbrido basado en ontología y ML con *endorsements*, las pruebas fueron ejecutadas bajo mismas configuraciones que la propuesta anterior, y con los diferentes conjuntos de datos: entrenamiento, pruebas y total. Los resultados obtenidos se presentan a continuación.

La Tabla 5.11 muestra los resultados, utilizando los datos de entrenamiento (70% del

```

=====
Sistema Híbrido de Recomendación
-----
Utiliza habilidades del mismo nivel de usuarios
Utiliza cargos obtenidos de agrupamiento por ML / DBSCAN
=====
Salida del Recomendador Semántico
-----
Descripción del Usuario: Desarrollador iOS Nativo
Sector Laboral del Usuario:
001-03-07:03-02 : iOS Nativo
Habilidades del usuario:
('20-10-01-*:01', 4, 2) : Xcode
('10-06-01-5:03', 4, 2) : ObjectiveC
('10-07-09-2:01', 3, 2) : MVC
('20-09-02-*:00', 3, 3) : Frontend
Habilidades validadas del usuario:
('10-06-01-5:05', 3, 2) : Kotlin
Sector Laboral Previos:
Sector Laboral Afines:
001-03-08:00-04 : Revisor Código
001-03-07:03-01 : Android Nativo
001-03-07:03-00 : Aplicaciones Móviles
Habilidades Afines:
('20-10-02-*:01', 3, 2) : AndroidStudio
('10-06-01-9:02', 3, 3) : Swift
('10-07-09-2:02', 3, 3) : MVP
('10-06-01-5:04', 3, 3) : Java
('10-07-09-2:03', 3, 2) : MVVM
('10-07-09-2:04', 3, 2) : VIPER
('10-07-05-*:01', 3, 2) : Mantis
('10-07-09-2:02', 3, 3) : MVP
Habilidades Asociadas
('10-07-07-*:00', 0, 0) : Gestión de versiones
('20-09-02-1:00', 0, 0) : Progresivas
=====
Salida del Filtrado por Contenido (Predicción inicial)
-----
2 Desarrollo en iOS con Swift
3 Fundamentos de Programación con Swift
4 Introducción a la Programación con Swift - Bilbao
5 Introducción a la Programación con Swift - Barcelona
6 Desarrollo Avanzado en Android
7 Desarrollo en Android con Kotlin
=====
Salida de la heurística (recomendación de cursos)
-----
2 Desarrollo en iOS con Swift
3 Fundamentos de Programación con Swift
6 Desarrollo Avanzado en Android
7 Desarrollo en Android con Kotlin
=====

```

Figura 5.7 Ejemplo de salida SR híbrido basado en ontología y ML con *endorsements* conjunto de datos).

Tabla 5.11 Métricas para las diferentes configuraciones del SR híbrido basado en ontología y ML con *endorsements* para el conjunto de datos de entrenamiento.

Configuración	MAE	RMSE	Cobertura	Precisión	Recall	Novedad	Serendipia
S3.1	0,58	9,13	0,64	1,00	0,36	0,14	0,09
S3.2	0,51	8,17	0,73	0,97	0,41	0,21	0,07

<b>S3.3</b>	<b>0,24</b>	<b>4,14</b>	<b>0,91</b>	<b>0,85</b>	<b>0,82</b>	<b>0,46</b>	<b>0,07</b>
S3.4	0,21	3,38	0,82	0,95	0,71	0,45	0,07
S3.5	0,36	6,35	0,91	0,83	0,74	0,39	0,07
<b>S3.6</b>	<b>0,16</b>	<b>2,58</b>	<b>0,91</b>	<b>0,94</b>	<b>0,80</b>	<b>0,52</b>	<b>0,07</b>
S3.7	0,48	9,05	0,91	0,69	0,63	0,31	0,04
S3.8	0,28	4,95	0,95	0,82	0,80	0,48	0,04

Los resultados de la ejecución de las diferentes configuraciones con el total de las muestras del conjunto de datos se muestran en la Tabla 5.12.

Tabla 5.12 Métricas para las diferentes configuraciones del SR híbrido basado en ontología y ML con *endorsements* para el total de las muestras del conjunto datos

Configuración	MAE	RMSE	Cobertura	Precisión	Recall	Novedad	Serendipia
S3.1	0,59	11,05	0,61	0,99	0,37	0,11	0,06
S3.2	0,50	9,45	0,71	0,97	0,45	0,17	0,05
<b>S3.3</b>	<b>0,23</b>	<b>4,82</b>	<b>0,94</b>	<b>0,84</b>	<b>0,87</b>	<b>0,39</b>	<b>0,06</b>
S3.4	0,22	4,26	0,81	0,96	0,73	0,39	0,05
S3.5	0,33	6,78	0,87	0,86	0,75	0,34	0,06
<b>S3.6</b>	<b>0,25</b>	<b>5,05</b>	<b>0,94</b>	<b>0,86</b>	<b>0,86</b>	<b>0,45</b>	<b>0,05</b>
S3.7	0,50	11,36	0,94	0,65	0,65	0,25	0,04
S3.8	0,33	7,26	0,97	0,77	0,86	0,42	0,03

Por último, se ejecutó el sistema bajo las diferentes configuraciones utilizando la porción de prueba (30%) de los datos, los resultados se muestran en la Tabla 5.13.

Tabla 5.13 Métricas para las diferentes configuraciones del SR híbrido basado en ontología y ML con *endorsements* para el conjunto datos de pruebas.

Configuración	MAE	RMSE	Cobertura	Precisión	Recall	Novedad	Serendipia
S3.1	0,62	6,24	0,56	0,96	0,41	0,04	0,00

<b>S3.2</b>	0,47	4,75	0,67	0,96	0,56	0,07	0,00
<b>S3.3</b>	<b>0,22</b>	<b>2,48</b>	<b>1,00</b>	<b>0,80</b>	<b>1,00</b>	<b>0,24</b>	<b>0,04</b>
<b>S3.4</b>	0,26	2,63	0,78	0,96	0,78	0,26	0,00
<b>S3.5</b>	0,25	2,60	0,78	0,94	0,78	0,21	0,02
<b>S3.6</b>	<b>0,40</b>	<b>4,92</b>	<b>1,00</b>	<b>0,68</b>	<b>1,00</b>	<b>0,28</b>	<b>0,00</b>
<b>S3.7</b>	0,55	6,90	1,00	0,54	0,69	0,12	0,06
<b>S3.8</b>	0,43	5,54	1,00	0,63	1,00	0,28	0,00

De los resultados obtenidos del SR híbrido basado en ontología y ML haciendo uso de los *endorsements*, se observó que las configuraciones con mejor desempeño para todos los conjuntos de datos son, el filtrado semántico utilizando reglas para determinar los sectores de desempeño laboral afines y habilidades afines (**S3.3**) y el filtrado semántico utilizando agrupación de DBSCAN con  $\varepsilon = 0,3$ , para determinar sectores de desempeño laboral afines y reglas semánticas para determinar las habilidades afines (**S3.6**).

Una vez ejecutadas las pruebas con el objetivo de evaluar el SR, a fin de realizar el análisis de los resultados, se compararon los resultados con los mejores desempeños de cada una de las versiones diferentes propuestas, los cuales se enumeran a continuación:

1. SR híbrido basado en taxonomía con codificación plana y similitud del coseno (**S1.1**).
2. SR híbrido basado en taxonomía con codificación jerárquica con umbral de corte de 0,8 y similitud de coseno (**S1.3**).
3. SR híbrido basado en ontología y ML con filtrado semántico utilizando reglas para determinar los sectores de desempeño laboral afines y habilidades afines (**S2.3**).
4. SR híbrido basado en ontología y ML con filtrado semántico utilizando agrupación de DBSCAN con  $\varepsilon = 0,3$ , para determinar los sectores de desempeño laboral afines y reglas semánticas para determinar las habilidades afines (**S2.6**).

5. SR híbrido basado en ontología y ML haciendo uso de los *endorsements* con filtrado semántico utilizando reglas para determinar los sectores de desempeño laboral afines y habilidades afines (**S3.3**).
6. SR híbrido basado en ontología y ML haciendo uso de los *endorsements* con filtrado semántico utilizando agrupación de DBSCAN con  $\varepsilon = 0,3$ , para determinar sectores de desempeño laboral afines y reglas semánticas para determinar las habilidades afines (**S3.6**).

Para cada una de las configuraciones antes descritas, se procede a comparar los resultados obtenidos con los diferentes conjuntos de datos: entrenamiento (70% del conjunto de datos), pruebas (30% del conjunto de datos) y total (100% del conjunto de datos), las cuales se muestran a continuación.

Los resultados de las configuraciones con los mejores desempeños utilizando los datos de entrenamiento (70% del conjunto de datos) se resumen en la Tabla 5.14.

Tabla 5.14 Métricas de las diferentes configuraciones, para el conjunto de datos de entrenamiento.

Configuración	MAE	RMSE	Cobertura	Precisión	Recall	Novedad	Serendipia
<b>S1.1</b>	0,27	4,38	0,91	0,91	0,74	0,46	0,07
<b>S1.3</b>	0,17	2,74	0,91	0,91	0,81	0,51	0,07
<b>S2.3</b>	0,25	4,44	0,91	0,83	0,82	0,46	0,07
<b>S2.6</b>	0,18	2,92	0,91	0,91	0,80	0,52	0,07
<b>S3.3</b>	0,24	4,14	0,91	0,85	0,82	0,46	0,07
<b>S3.6</b>	<b>0,16</b>	<b>2,58</b>	<b>0,91</b>	<b>0,94</b>	<b>0,80</b>	<b>0,52</b>	<b>0,07</b>

Para los sistemas y configuraciones basados en ontologías y ML una vez entrenados los modelos se obtuvieron los resultados mostrados en la Tabla 5.15 con el total de las muestras del conjunto de datos.

Tabla 5.15 Métricas de las diferentes configuraciones, para el conjunto total de datos

Configuración	MAE	RMSE	Cobertura	Precisión	Recall	Novedad	Serendipia
S2.3	0,25	5,16	0,94	0,82	0,87	0,39	0,06
S2.6	0,27	5,63	0,94	0,84	0,86	0,45	0,05
S3.3	<b>0,24</b>	<b>4,14</b>	<b>0,91</b>	<b>0,85</b>	<b>0,82</b>	<b>0,46</b>	<b>0,07</b>
S3.6	<b>0,16</b>	<b>2,58</b>	<b>0,91</b>	<b>0,94</b>	<b>0,80</b>	<b>0,52</b>	<b>0,07</b>

Para los sistemas y configuraciones basados en ontologías y ML una vez entrenados los modelos los resultados con los datos de prueba (30%), se muestran en la Tabla 5.16.

Tabla 5.16 Métricas de las diferentes configuraciones, para el conjunto datos de pruebas.

Configuración	MAE	RMSE	Cobertura	Precisión	Recall	Novedad	Serendipia
S2.3	0,23	2,64	1,00	0,80	1,00	0,24	0,04
S2.6	0,43	5,42	1,00	0,67	1,00	0,28	0,00
S3.3	<b>0,22</b>	<b>2,48</b>	<b>1,00</b>	<b>0,80</b>	<b>1,00</b>	<b>0,24</b>	<b>0,04</b>
S3.6	<b>0,40</b>	<b>4,92</b>	<b>1,00</b>	<b>0,68</b>	<b>1,00</b>	<b>0,28</b>	<b>0,00</b>

Como se indicó en el estado del arte, una medida que resume, tanto la precisión como la recuperación, es la media armónica entre la precisión y la recuperación ( $F_1$ ). La Tabla 5.17 muestra las medias armónicas para las configuraciones con los mejores desempeños.

Tabla 5.17 Media armónica ( $F_1$ ) para las ejecuciones con mejor desempeño del SR

Configuración	Datos de entrenamiento	Total de datos	Datos de pruebas
S1.1	0,85	-	-
S1.3	0,86	-	-
S2.3	0,82	0,84	0,89
S2.6	0,85	0,85	0,80
S3.3	0,83	0,84	0,89

<b>S3.6</b>	0,86	0,85	0,81
-------------	------	------	------

De la Tabla 5.14 (comportamiento con los datos del entrenamiento) y la Tabla 5.17 (comparativo de la media armónica) se observa, en relación al SR híbrido basado en taxonomía, un mejor desempeño de la codificación jerárquica (**S1.3**) en comparación con la codificación plana, y en relación con el sistema basado en Ontología y ML un desempeño ligeramente superior en el SR que hace uso de los *endorsements*, obteniendo el mejor resultado, la configuración que hace uso del algoritmo de DBSCAN (**S3.6**), este sistema implementa todas las mejoras propuestas, esto se puede observar en la Figura 5.8.

En general de la comparación de los SR taxonómicos y con los SR híbridos que hacen uso de ontologías y ML se observa un desempeño similar de la configuración basada en DBSCAN (**S2.6** y **S3.6**) y un aumento en los errores en la configuración basada en reglas semánticas (**S2.3** y **S3.3**), lo que sugiere un buen modelado de los datos por parte del SR con filtrado semántico y un excelente modelado por parte de SR con la agrupación realizada por DBSCAN.

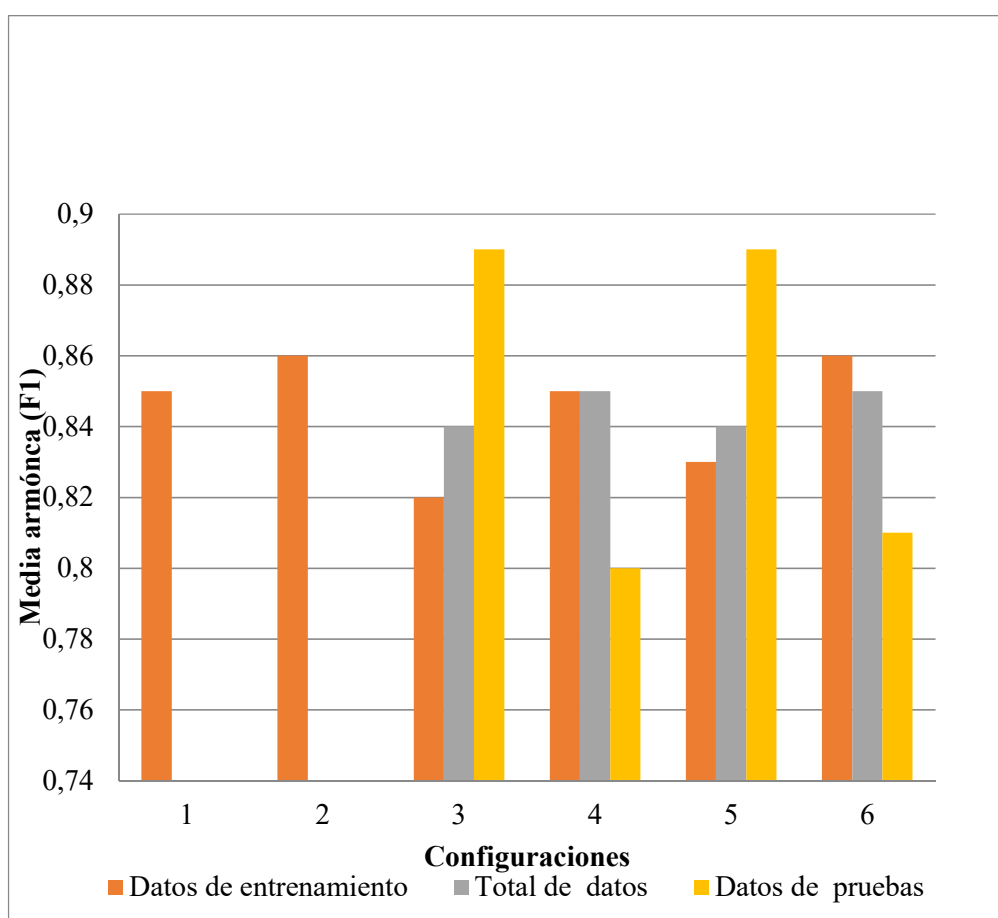


Figura 5.8 Media armónica (F1) para las ejecuciones con mejor desempeño del SR

Del análisis de la Tabla 5.16 (comportamiento con los datos de pruebas), la Tabla 5.14 (comportamiento con los datos del entrenamiento) y la Tabla 5.17 (comparativo de la media armónica) se observa un error similar para los SR basado con filtrado semántico (S2.3 y S3.3), destacándose un aumento del *recall* y una disminución de la precisión, es decir esta aumentado el tamaño de las recomendaciones de tal forma que cubre relevantes pero aumenta sus diferencias, aunque se tiene un aumento de la media armónica. Ahora para los SR basados en DBSCAN (S2.6 y S3.6), a pesar de un aumento en las métricas de errores, se sigue teniendo un muy buen comportamiento en las métricas de *recall* y precisión, destacándose un aumento del *recall* pero una disminución significativa de la precisión, aunque en este caso se observa una disminución de la media armónica. En general se observa un peor comportamiento para las métricas de novedad y Serendipia para los datos de pruebas que para los datos de entrenamiento.

Para el total de los datos, de Tabla 5.16 (comportamiento con el conjunto total de datos) y la Tabla 5.17 (comparativo de la media armónica) se observa que en general se observa un comportamiento similar en relación con el conjunto de datos de pruebas para los SR basados en Ontología y ML, mostrando un mejor desempeño de SR con la agrupación realizada por DBSCAN haciendo uso de los *endorsements* (S3.6).

Las tablas 5.14, 5.15 y 5.16, muestran que incorporar los *endorsements*, representan en algunos casos un muy ligero descenso en los errores y un muy ligero aumento de la precisión, lo que se refleja en la media armónica tal como se puede verificar en la Tabla 5.17.

De las diferentes pruebas, para la agrupación realizada por k-Means se observa un rendimiento inferior, esto podría ser explicado por la naturaleza del dominio y la distribución de los cargos, en el análisis del gráfico se observa en general una mejor agrupación (cargos afines) realizado por DBSCAN que la realizada por k-Means. Mejoras en las métricas de *Recall* y Serendipia se puede asociar al uso en la ontología y la estimación habilidades asociadas o de interés y filtrado de cargos previos al actual en una relación de orden impuesto por un escalafón.

En este punto hay que resaltar que los sistemas basados en modelos tienen la ventaja de tener un mejor rendimiento ya que no necesitan procesar todo el conjunto de datos.

De acuerdo con la Tabla 5.17 el mejor resultado se obtiene con la versión del sistema basado SR híbrido basado en ontología y ML bajo el algoritmo de clusterización DBSCAN y que hace uso de los *endorsement* (S3.6). Al estar basado en un modelo tiene mejor eficiencia computacional y eficacia similar al SR híbrido basado en datos almacenados.

Para poder hacer un análisis completo de los resultados, es conveniente poder comparar el rendimiento de las propuestas con el de otros SR similares. En una revisión de la literatura, encontramos el trabajo de (Kolekar, 2021), el cual se orientado a estudiantes universitarios y coordinadores de empresas, para recomendar empleos de acuerdo con las habilidades de los usuarios. Así mismo, el trabajo de (Ibrahim, Yang,

Ndzi, Yang, & Almaliki, 2018), está pensado para recomendar cursos a estudiantes, utilizando múltiples fuentes de datos. Por su parte, el trabajo de (Patel, Kakuste, & Eirinaki, 2017), recomienda a los usuarios trayectorias profesionales y las habilidades que se requieren para diferentes puestos de trabajo, basado en las habilidades e interés de estos. Finalmente, (Wang C. , y otros, 2020), proponen un trabajo para recomendar a profesionales cursos en línea, según sus competencias de profesionales y preferencias de desarrollo profesional. En la Tabla 5.18, se presenta el resumen de los resultados de cada uno de estos trabajos, junto con el mejor resultado obtenido en esta tesis.

Tabla 5.18 Medidas de evaluación en trabajos similares

SR Similares	MAE	RMSE	Cobertura	Precisión	Recall	Novedad	Serendipia	F <sub>1</sub>
(Kolekar, 2021)	-	-	-	0,83	0,82	-	-	0,82
(Ibrahim, Yang, Ndzi, Yang, & Almaliki, 2018)	-	-	-	0,92	0,7	-	-	-
(Patel, Kakuste, & Eirinaki, 2017)	-	-	-	0,64	-	-	-	-
(Wang C. , y otros, 2020)	-	-	-	-	0,70	-	-	-
<b>S3.6</b>	<b>0,16</b>	<b>2,58</b>	<b>0,91</b>	<b>0,94</b>	<b>0,80</b>	<b>0,52</b>	<b>0,07</b>	<b>0,85</b>

Al comparar estos trabajos con el mejor resultado de las propuestas realizadas en esta investigación, se puede apreciar que el SR híbrido basado en ontología y ML bajo el algoritmo de clusterización DBSCAN con *endorsement* (S3.6) obtiene una mejora de las métricas de todas las propuestas de los SR similares.

## 5.4 Resumen del capítulo

En este capítulo se detalla el protocolo de prueba, describiendo las métricas utilizadas para la evaluación de las diferentes propuestas y configuraciones para las mismas. Se

presentan los resultados para los diferentes escenarios de cada una de las versiones del SR, resaltando en cada uno las configuraciones con mejor desempeño. Obteniendo el mejor resultado con la versión del SR híbrido basado en ontología y ML bajo el algoritmo de clusterización DBSCAN y que hace uso de los *endorsement* (S3.6), este sistema implementa todas las mejoras propuestas.

## Capítulo 6

# Conclusiones y trabajo futuro

Este capítulo presenta las conclusiones del trabajo llevado a cabo durante el desarrollo de la tesis. Así mismo, se señalan líneas de trabajo futuro y las contribuciones científicas producto de esta investigación.

### 6.1 Conclusiones

El desarrollo de la presente tesis parte de la siguiente hipótesis:

*Es posible modelar la información de perfiles de usuario obtenidos en RRSS para mejorar el rendimiento de un SR híbrido basado en ontología y Machine Learning de cursos de formación continua para profesionales, que les permita adaptarse a los cambios que sufren las organizaciones.*

Con el fin de demostrar la hipótesis planteada, el objetivo general de esta tesis fue el desarrollo de un SR híbrido basado en ontología y ML de cursos de LL para la actualización y/o desarrollo de habilidades profesionales a usuarios de redes sociales.

Durante el desarrollo de la investigación se realizaron tres versiones del SR. La Figura 6.1 muestra la evolución del SR, resaltando las diferencias en color.

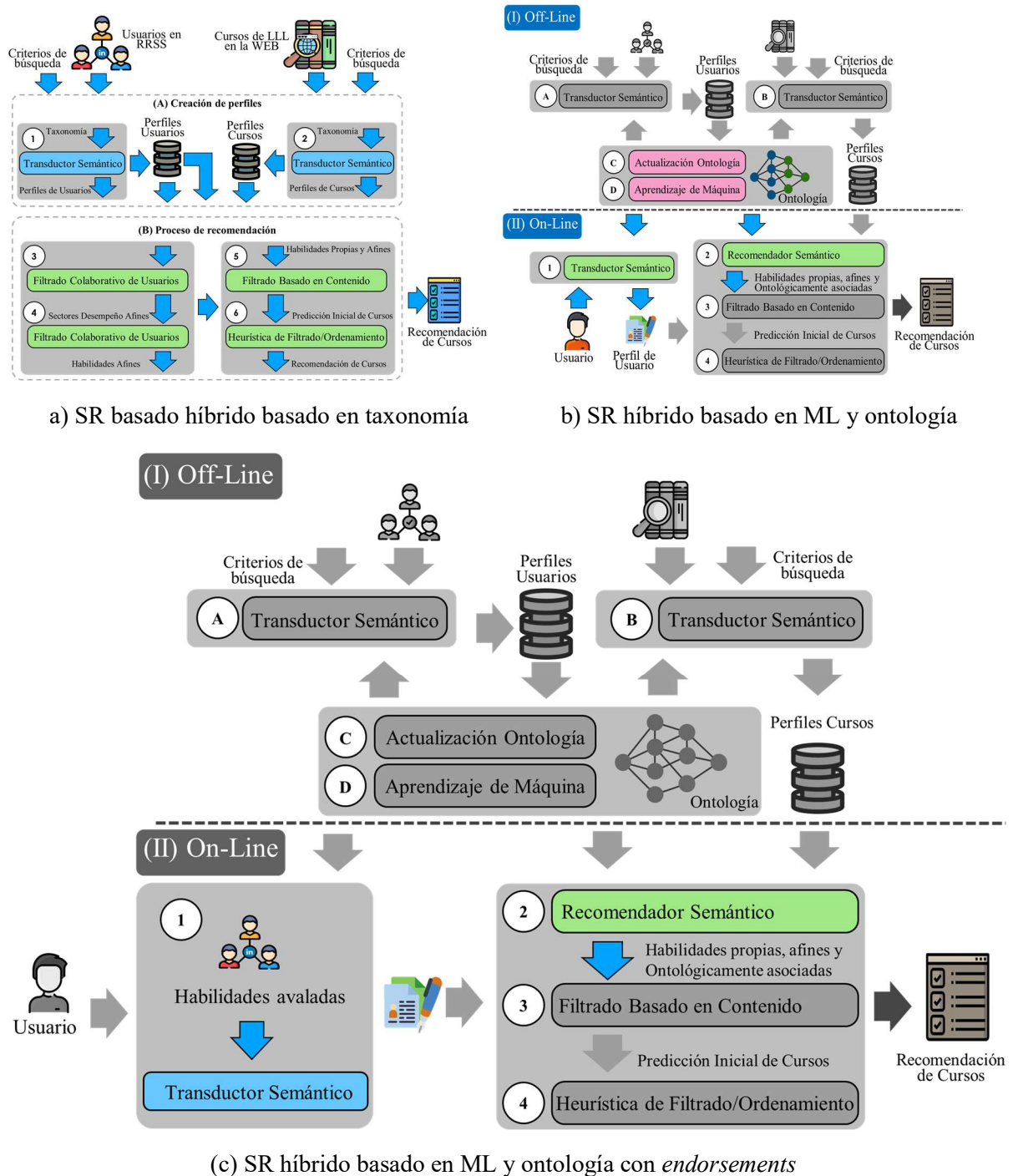


Figura 6.1 Evolución del Sistema de Recomendación

Se partió del desarrollo de un SR híbrido basado en taxonomía (Figura 6.1 a) y, como producto de la evaluación y análisis de los resultados, se planteó una mejora enriqueciendo la taxonomía con nuevas entidades y relaciones, desarrollando un SR basado en ontología, en el que se aplican técnicas de ML (Figura 6.1 b). Por último, se desarrolló la versión final, tomando en cuenta para el perfil de usuario las validaciones de sus habilidades, por parte de otros miembros de LinkedIn (endorsements) (Figura 6.1 c).

A continuación, se presentan las conclusiones obtenidas durante la evolución de la tesis, estas se expondrán relacionadas con cada uno de los objetivos propuestos.

**OE1.-** Revisar el estado del arte de los SR en el área de LL y necesidades del sector educación.

La revisión del estado del arte permitió establecer una base de partida para la presente tesis, señalando un panorama sobre los avances realizados en el área de estudio. Se destaca lo siguiente:

- Los continuos cambios que sufren las organizaciones demandan empleados que se adapten a estos, surgiendo la necesidad de una constante formación por parte de los profesionales para poder responder con éxito ante esta situación. Como respuesta a esta necesidad hay un gran crecimiento de la oferta de cursos de LL en la web. La búsqueda de cursos, que se adapten a las necesidades de formación, requiere de una gran inversión de tiempo.
- Como solución a esta situación y tomando en cuenta la sobrecarga de información, surgen los sistemas de recomendación. En la revisión de la literatura, se pudo identificar cuáles eran los algoritmos de filtrado más utilizados y hacia donde se orientaban los avances en este campo, resaltando en los últimos años, los **SR híbridos** y la incorporación de **técnicas de ML**. Así mismo, resaltaba la importancia de la representación de la información para el buen desempeño de los SR, proponiendo para la construcción de los perfiles de usuarios, el uso de **taxonomías** y **ontologías**.

- En el caso particular de los SR en educación, la mayoría de estos se orientaban a dar soporte en la educación formal, en donde las recomendaciones se basaban principalmente en usuarios similares, estilos de aprendizajes e historial académico. En menos proporción se encontraron SR en el campo de la educación no formal, así como en la sugerencia de planes de formación para profesionales. En estos últimos, se comienzan a explorar el **uso de RRSS**, así como a tomar en cuenta las **habilidades profesionales** de los usuarios para basar la recomendación de cursos.
- La naturaleza de las fuentes de los datos de los SR encontrados en el estado del arte en el ámbito de la educación, en su mayoría, son propiedad de organizaciones y trabajan sobre datos estructurados, en donde se pueden evidenciar relaciones entre los sujetos y los artículos a recomendar.

**OE2.-** Identificar y caracterizar las fuentes de datos en redes sociales y oferta de cursos en internet.

A partir del estado del arte y considerando como una oportunidad para esta investigación el uso de las RRSS y la información extraída de las mismas, era necesario poder identificar aquellas, que para efectos de esta investigación se tomarían en cuenta para el desarrollo de la propuesta. Así mismo, poder caracterizar y representar esta información para el buen desempeño del SR. En este sentido podemos concluir:

- En el estado del arte, los trabajos encontrados, que hacían uso de RRSS, en su mayoría tomaban en cuenta, Facebook, Instagram y **LinkedIn**. Partiendo de que, el objetivo planteado está orientado a la recomendación de cursos de LL para la mejora y/o desarrollo de habilidades profesionales, se seleccionó LinkedIn por ser la que contiene la información más detallada sobre el perfil académico y profesional de los usuarios.
- Una de las características más resaltantes de los datos obtenidos, tanto de los usuarios, como de los cursos de LL que se extrajeron de diferentes sitios web, es que, debido a la heterogeneidad de las fuentes, son datos

semiestructurados, o no estructurados. La siguiente característica, importante a resaltar, está referida a la relación entre usuarios y cursos de LL, en donde, no se encuentra una relación a priori entre ellos, esto debido a la dinámica de los cursos de LL.

- El uso de los transductores permitió pasar estos datos en lenguaje natural a datos estructurados en los perfiles de usuarios y cursos, por lo que pueden ser procesados eficientemente por los SR y por los algoritmos de ML que requieren de datos estructurados.

**OE3.-** Proponer procesos para la construcción de perfiles de usuarios y cursos a partir de registros profesionales en redes sociales y ofertas de cursos en internet.

Partiendo de las características encontradas en las fuentes de datos, para poder crear un perfil de usuario que represente todos los rasgos resaltantes, a fin de poder ofrecer una recomendación que cumpla con las expectativas del usuario, se hizo uso de **ontologías para codificar cada uno de los elementos de los perfiles**, de tal manera de poder tener elementos que relacionaran a los usuarios con los cursos de LL, **y así lograr un buen rendimiento de SR.**

- El uso de taxonomía para codificar jerárquicamente características o propiedades en perfiles no solo es eficiente en su almacenamiento y procesamiento, y evita la pérdida de información al mantener relaciones del tipo “es-un”, sino que, al combinarse con umbrales de corte en las funciones de similitud taxonómica, da una ganancia en el cálculo de similitud entre perfiles de los sujetos, lo que redundará en mejoras en el desempeño del sistema de recomendación.
- El uso de taxonomía, para representar las áreas de conocimiento de las habilidades, permite representar diferentes niveles de conocimiento, evitando la pérdida de información propia de las codificaciones planas, siendo eficiente en su almacenamiento y en su mecanismo de comparación.

- El uso transductores permitió unificar, al codificar los datos provenientes de múltiples fuentes, los perfiles tanto de usuarios como de cursos, así como el tratamiento de entidades iguales o sinónimos

**OE4.-** Proponer una estrategia de recomendación para cursos de LL a partir de perfiles de usuarios y de cursos.

La estrategia planteada para la recomendación de cursos de LL, se basó en establecer una relación entre usuarios basada en sectores de desempeño laboral y habilidades profesionales, para así, identificar aquellas habilidades que debe mejorar o desarrollar para su puesto de trabajo actual, o para acceder a otro puesto de trabajo de mayor nivel, y a partir de estas habilidades identificadas, poder determinar una predicción inicial de cursos que las desarrollen. Donde el uso de heurísticas fue relevante para la construcción de la propuesta. En la primera propuesta se trabajó con el conjunto total de datos de los perfiles de usuario y cursos almacenados. En las versiones siguientes se implementó una mejora al plantearse un modelo de datos que consistió en representar el dominio de los sectores de desempeño laboral y las diferentes áreas de conocimiento de las habilidades. Esta estrategia permitió:

- Establecer un mecanismo para poder relacionar los datos, que como se mencionó previamente, no tenían relaciones a priori con las cuales basar las recomendaciones.
- Se demostró que es posible pasar de una filosofía de datos almacenados a un modelo semántico, que representa el dominio de trabajo, con los consiguientes beneficios computacionales.

**OE5.-** Definir un modelo basado en ontología y ML para un sistema de recomendación de cursos de LL.

Para cumplir con este objetivo, se partió del análisis de las características de los datos y de la estrategia establecida para el SR. **Se definieron dos ontologías, una ontología para la representación de los sectores de desempeño laboral y otra para la**

**representación de las áreas de conocimiento.** En este modelo, las habilidades, son el elemento que permiten relacionar estas dos ontologías. Así mismo, el uso de técnicas de ML sobre los sectores de desempeño laboral permitió enriquecer los datos. En relación con este objetivo, podemos concluir:

- La aplicación de reglas semánticas, haciendo uso de la ontología, mejora las recomendaciones, al hacer uso del filtrado de posiciones de cargos según la jerarquía, y por lo tanto de habilidades, esto permitió considerar habilidades relacionadas, permitiendo así enriquecer las recomendaciones.
- Con relación al uso de ML, dadas las características de las fuentes, no se tenían datos etiquetados, lo que fue determinante para la selección de algoritmos de ML no supervisados. **La implementación de ML permitió, a partir de los clústeres de sectores laborales, inferir más información,** como, por ejemplo, qué habilidades son compartidas por distintos sectores laborales, con lo cual se puede recomendar formación para reorientar la carrera profesional.
- La representación de la ontología con la base de datos NoSQL neo4j mostro ser una alternativa eficiente en el contexto RRSS, caracterizado gran cantidad de datos asociados.

**OE6.-** Diseñar e implementar las diferentes propuestas de sistemas de recomendación para cursos de LL.

Para alcanzar este objetivo se implementaron los algoritmos de las diferentes propuestas definidas en el capítulo 4, permitiendo llegar a las siguientes conclusiones:

- **La arquitectura definida de los diferentes SR hizo posible utilizar diferentes configuraciones** y evaluar cada una de las funcionalidades por separado y, de ser necesario, varias de ellas simultáneamente. **Las configuraciones se pueden parametrizar, lo que permite evaluar una amplia gama de combinaciones.**

- La metodología utilizada hizo posible el refinamiento sucesivo de la propuesta, manteniendo actualizados el conocimiento teórico e incorporando progresivamente elementos, como resultado de una revisión continua del estado del arte y las fuentes de datos.
- Dado el diseño del SR, aunque no era el enfoque de la propuesta, las técnicas de cobertura de habilidades hacen posible que el sistema se utilice para recomendar puestos de trabajo, dado un perfil de usuario, o para buscar candidatos, dado un perfil del sector laboral. El SR puede ser utilizado para ofrecer a un usuario determinado áreas de desempeño profesionales para las cuales está capacitado, y en caso necesario, indicar que habilidades se deben desarrollar o mejorar y/o planes de formación para conseguirlo. Esta funcionalidad puede resultar interesante para encontrar un primer empleo, para la reinserción o para reinventarse en el mundo del ejercicio laboral. Otro uso interesante del SR es poder identificar, de entre un grupo de usuarios, los “mejores” candidatos, para una oferta de empleo de acuerdo con las habilidades requeridas para éste. Así como determinar oportunidades de formación, lo que puede resultar beneficioso para la oferta de cursos LL.

**OE7.-** Evaluar las diferentes propuestas de sistemas de recomendación para cursos de formación continua a partir de las métricas de rendimiento de los SR, contrastando los resultados con otros SR similares del estado del arte.

Para la consecución de este objetivo se realizó una evaluación off-line del SR, en la que **se determinaron un conjunto de métricas que permitieron determinar el rendimiento de cada una de las propuestas.** Esta evaluación hizo posible hacer mayor cantidad de pruebas, aumentar la cantidad de muestras y tener consistencia en la preferencia de los usuarios en repetidas o múltiples pruebas, dado que se quiere la mayor consistencia para poder comparar el desempeño de las diferentes propuestas bajo las mismas condiciones. Con las métricas obtenidas **se pudo comparar esta propuesta con otros trabajos similares.** Las conclusiones obtenidas se discuten a continuación:

- **Las propuestas finales combinan la eficiencia de las taxonomías**, en cuanto a su almacenamiento, procesamiento y su capacidad para evitar la pérdida de la información al codificar, **con el poder de las ontologías**, en la representación de la realidad y poder hacer inferencia sobre ella. Adicionalmente, **el uso de ML** y el refinamiento de las recomendaciones, implementado con heurísticas, **mejoran el desempeño de la recomendación**, proporcionando mayor eficiencia computacional en relación con la propuesta inicial.
- Al evaluar la primera propuesta, el SR híbrido basado en taxonomía, se pudo evidenciar, en las diferentes ejecuciones del sistema, que el mejor desempeño se obtiene con al configurar el SR con la similitud del coseno con un umbral de corte  $\theta$  de 0,8 para el cálculo de la similitud taxonómica, en el cálculo de los vectores  $u$  y  $v$ , y un umbral de corte de  $\mu$  de 0,7 para la similitud del coseno. De esta manera se pueden comparar las habilidades,  $y$ , aunque no sean iguales, en todos sus elementos, permite obtener un grado de similitud, que resulta útil, a la hora de dar las recomendaciones. **El uso del umbral de corte en la similitud taxonómica permite establecer similitudes más allá de los conocimientos específicos, como, por ejemplo, a nivel de subespecialización.** Comparar las habilidades, hasta niveles diferentes de las hojas, redundante en una ganancia en el valor de la similitud, con la respectiva mejora en la calidad de las recomendaciones del sistema.
- De la evaluación de la segunda propuesta, el SR híbrido basado en ontología y ML, se concluye que **el uso de ontología y filtrado semántico no solo es útil para solventar los problemas relacionados con el arranque en frío, sino que permiten representar a mayor detalle la realidad y poder hacer inferencias.** En cuanto a la evaluación de las diferentes configuraciones ejecutadas del SR, aunque se obtuvo un buen desempeño con la que utilizan el filtrado semántico con reglas para determinar los sectores de desempeño y habilidades afines, se obtuvo el mejor desempeño con la propuestas que hacen uso de filtrado semántico utilizando algoritmo de ML, en particular, la que

utiliza DBSCAN, el cual tiene mejor desempeño, dada las características de los datos, que las basadas en k-mean, ya que estas calculan los clústeres por densidades. Este sistema presenta una mejora del desempeño del SR en general con relación a la primera propuesta, evidenciándose en particular en métricas como novedad y serendipia.

- La evaluación de la propuesta final, en la que se incorporan los *endorsements* en la creación de los perfiles de usuario, permitió observar una mejora en el rendimiento del SR, en la configuración que hace uso del algoritmo DBSCAN, mejorando los valores del error cuadrático medio (RMSE) y error absoluto medio (MAE), así como en las métricas de precisión, novedad y cobertura. Con la incorporación de los *endorsements* se pudo obtener más información del perfil de usuario, ya que permitió incorporar habilidades que no se evidenciaban en los datos relacionados con empleo actual de este, que fueron útil a la hora de refinar la recomendación de cursos.

Sobre el desempeño del sistema, en relación con la capacidad y el rendimiento, respecto a otros sistemas similares, en donde sus objetivos estaban orientados a recomendar cursos, empleos o trayectorias profesionales, destaca lo siguiente:

- Una de las principales diferencias está centrada en la naturaleza de los datos, se pudo observar que, en su mayoría, trabajan con datos estructurados. El trabajo presentado por (Patel, Kakuste, & Eirinaki, 2017) obtuvo los datos directamente de la base de datos de una universidad. (Kolekar, 2021), los obtuvo de un repositorio de datos para experimentación con ML, mientras que (Wan, y otros, 2020) los extrajo de los perfiles de empleados y del plan de formación directamente de una empresa. Por su parte, el trabajo de (Ibrahim, Yang, Ndzi, Yang, & Almaliki, 2018), propuso utilizar información de LinkedIn para complementar los datos introducidos directamente de los usuarios a través de la aplicación u obtenidos desde la base de datos de la universidad. A partir de estos datos el propone una taxonomía para la representación de la información de los cursos, empleos y estudiantes

- Al analizar la evaluación de estos trabajos, para comparar sus resultados con esta propuesta, las métricas utilizadas por la mayoría fueron *recall* y precisión. Con relación a los valores obtenidos, podemos concluir que, si comparamos con los resultados obtenidos, con las pruebas realizadas al conjunto total de los datos, esta propuesta mejora dos puntos porcentuales en precisión y 10 puntos porcentuales en *recall*, las mejores métricas obtenidas por estos trabajos.

## 6.2 Trabajo Futuro

Concluida la evaluación de las propuestas y analizados los resultados, se presentan algunas líneas orientativas para continuar los avances en esta área relacionados con la presente tesis:

Para el desarrollo de esta tesis se tomó como dominio el área de desarrollo de software, investigaciones futuras podría implicar trabajar con varios dominios relacionados, como, por ejemplo, desarrollo de software, soporte técnico, consultoría, capacitación y academia, con el fin de evaluar el comportamiento del sistema para múltiples dominios, actualizando la ontología con nuevas instancias asociadas de nuevos dominios. Así como evaluar el uso de ontologías ya construidas, nuevas formas de representación de áreas de conocimiento y sectores laborales de desempeño.

Los transductores semánticos son un elemento clave en la construcción de los perfiles, se debe trabajar en mejorar su eficacia, para lo cual se debe enriquecer las ontologías utilizadas y refinar las estrategias de codificación.

En relación con el uso de algoritmos de ML, se utilizaron algoritmos de *hard clustering* en la implementación de la propuesta, trabajos futuros pueden hacer uso técnicas de *soft clustering* y evaluar el rendimiento.

Con relación a las fuentes de información, dado los continuos cambios evaluar tanto el uso de otras redes sociales, así como distintas plataformas, como, por ejemplo, rocketreach y data-lead, en donde se pueden conseguir perfiles profesionales.

Ofrecer aplicaciones en línea en Internet para recomendación de cursos nos solo para una evaluación en línea del sistema, sino para determinar oportunidades de mejora a partir de las sugerencias de los usuarios.

Explotar las relaciones del grafo de conocimiento. Se puede evaluar el uso de técnicas algorítmicas en el grafo de conocimiento para mejorar las recomendaciones basadas en las posiciones y habilidades más demandadas, y construir caminos que describan secuencias para lograr habilidades, o una carrera profesional.

### 6.3 Difusión Científica

Durante el proceso de elaboración de esta tesis se realizaron varias aportaciones a la comunidad científica.

Se han publicado los siguientes artículos de revista:

- **Urdaneta Ponte María Cora**, Méndez Zorrilla Amaia, Oleagordia Ruiz Ibon. "Lifelong Learning Courses Recommendation System to Improve Professional Skills Using Ontology and Machine Learning" *Applied Sciences*. 2021, 11, 9: 3839. <https://doi.org/10.3390/app11093839>.

*Applied Sciences* cuenta con un factor de impacto de 2.679, y los siguientes índices JCR:

Q2: Engineering, Multidisciplinary

Q2: Physics, Applied

Q3: Chemistry, Multidisciplinary

Q3: Materials Science, Multidisciplinary

- **Urdaneta Ponte María Cora**, Mendez Zorrilla Amaia., Oleagordia Ruiz Ibon. "Recommendation Systems for Education: Systematic Review". *Electronics*. 2021, 10,1611. <https://doi.org/10.3390/electronics10141611>

*Electronics* cuenta con un factor de impacto de 2.397, y los siguientes índices JCR:

Q3: Engineering, Electrical & Electronic

Q3: Computer Science, Information Systems

Q3: Physics, Applied

- **Urdaneta Ponte María Cora**, Mendez Zorrilla Amaia., Oleagordia Ruiz Ibon. “Using LinkedIn endorsements to reinforce an ontology and machine learning-based recommender system to improve professional skills.”. *Electronics* **2022**, *11*, 1190. <https://doi.org/10.3390/electronics11081190>  
Electronics cuenta con un factor de impacto de 2.397, y los siguientes índices JCR:

Q3: Engineering, Electrical & Electronic

Q3: Computer Science, Information Systems

Q3: Physics, Applied

En congresos se han realizado los siguientes aportes:

- **Urdaneta Ponte María Cora**, Méndez Zorrilla Amaia y Oleagordia Ruiz Ibon "Taxonomy-Based Hybrid Recommendation System for Lifelong Learning to Improve Professional Skills," *2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, 2020, pp. 595-600, <https://doi.org/10.1109/TALE48869.2020.9368398>
- **Urdaneta Ponte María Cora**, Méndez Zorrilla Amaia, Oleagordia Ruiz Ibon. “Sistema de Recomendación híbrido basado en una nueva taxonomía para el aprendizaje permanente con el objetivo de mejorar las competencias profesionales”. Congreso Universitario Internacional sobre Comunicación, Investigación y Docencia 2020 (CUICIID 2020) - 7 y 8 de octubre de 2020 <https://cuiciid.net/wp-content/uploads/2021/04/Libro-de-actas-CUICIID-2020-con-comite-1.pdf>
- **Urdaneta Ponte María Cora**, Méndez Zorrilla Amaia, Oleagordia Ruiz Ibon. “Use of LinkedIn Endorsements in Recommender Systems”. The International Conference on Education and New Developments (END Conference 2022). Aceptado.

Se ha publicado un capítulo de libro:

- **Urdaneta Ponte Maria Cora**, Méndez Zorrilla Amaia, Oleagordia Ruiz Ibon. “Sistema de Recomendación híbrido basado en una nueva taxonomía para el aprendizaje permanente con el objetivo de mejorar las competencias profesionales”. Investigación e innovación académicas para una sociedad interconectada. Coordinadores: Juan Enrique González Vallés, Jelena Bobkina y Cristina Manchado Nieto. Editorial Tirant Humanidades. 2021

Se ha realizado la solicitud del registro de la propiedad intelectual del software:

Nº de expediente: 765-713570.

## Bibliografía

- ACM, & IEEE. (2020). *Computing Curricula 2020*. Obtenido de <https://www.acm.org/binaries/content/assets/education/curricula-recommendations/cc2020.pdf>
- Adam, N. L., Zulkafli, M. A., Soh, S. C., & Kamal, N. A. (2017). Preliminary study on educational recommender system. *IEEE Conference on e-Learning, e-Management and e-Services (IC3e)*. Malaysia.
- Aggarwal, C. (2016). *Recommender Systems. The Textbook*. New York, USA: Springer.
- Aguilar, J., Valdiviezo-Díaz, P., & Riofrio, G. (2016). A Fuzzy Cognitive Map like Recommender System of Learning Resources. *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. Vancouver.
- Aissaoui, O. E., & Oughdir, L. (2020). A learning style-based Ontology Matching to enhance learning resources recommendation. *1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*. Meknes, Morocco.
- Akmaliah, F., Krisnadhi, A., Sensuse, D., Rahayu, P., & Wulandari, I. (2018). Role of Ontology and Machine Learning in Recommender Systems. *2018 Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)*. Batu, Indonesia.
- Alkhamisi, A. O., & Saleh, M. (2020). Ontology Opportunities and Challenges: Discussions from Semantic Data Integration Perspectives. *6th Conference on Data Science and Machine Learning Applications (CDMA)*, (págs. 134-140). Riyadh, Saudi Arabia.
- Alrumaih, H., Mirza, A., & Alsalamah, H. (2020). Domain Ontology for Requirements Classification in Requirements Engineering Context. *IEEE Access*, 8, 89899-89908.
- Anusha, U., & Biradar, S. (2016). Recommender Systems: A Survey. *International Journal of Latest Technology in Engineering*, (V), 2278–2540.

- Assami, S., Daoudi, N., & Ajhoun, R. (2019). Exploring Social Media Data for MOOC Recommendation. *International Conference on Systems of Collaboration Big Data, Internet of Things Security (SysCoBioTS)*. Casablanca, Morocco.
- Bakhshinategh, B., Spanakis, G., Zaiane, O., & ElAtia, S. (2017). A Course Recommender System based on Graduating Attributes. *9th International Conference on Computer Supported Education*. Porto, Portugal.
- Balachander, Y., & Moh, T. (2018). Ontology Based Similarity for Information Technology Skills. *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, (págs. 302-305). Barcelona.
- Banu Kenayathulla, H., Aziah Ahmad, N., & Rahman Idris, A. (2019). Gaps between competence and importance of employability skills: evidence from Malaysia. *Higher Education Evaluation and Development*, 13(2), 97-112.
- Bañeres, D., & Conesa, J. (2016). eOrient@—A Recommender System to Address Life-Long Learning and Promote Employability. *International Conference on Intelligent Networking and Collaborative Systems (INCoS)*. Ostrava, República Checa.
- Baskota, A., & Ng, Y. (2018). A Graduate School Recommendation System Using the Multi-Class Support Vector Machine and KNN Approaches. *IEEE International Conference on Information Reuse and Integration (IRI)*. Salt Lake City.
- Bauman, K., & Tuzhilin, A. (2018). Recommending Remedial Learning Materials to the Students by Filling their Knowledge Gaps. *International Workshop on Educational Recommender Systems (EdRecSys)*. Omaha.
- Bedjou, K., Azouaou, F., & Berkani, L. (2014). Semantic recommendation of web services in the context of on-line training. *th International Symposium ISKO-Maghreb: Concepts and Tools for knowledge Management (ISKO-Maghreb)*. Algiers, Algeria.
- Behar, P., da Silva, K., Schneider, D., Cazella, S., Torrezan, C., & Heis, E. (2015). Development and System Assessment of Learning Object Recommendation based on Competency—RecOAComp. *7th International Joint Conference on Knowledge Discovery; Knowledge Engineering and Knowledge Management*. Lisboa.
- Bernal, C. A. (2010). *Metodología de la investigación. Administración, economía, humanidades y Ciencias Sociales*. Colombia: Prentice Hall - Pearson.
- Bhumichitr, K., Channarukul, S., Saejiem, N., Jiamthapthaksin, R., & Nongpong, K. (2017). Recommender Systems for university elective course recommendation. *14th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. Nakhon Si Thammarat, Thailand.
- Bigotte de Almedia, M., & A, G. (2015). The CeAMatE-on project: An online mathematical Support Centre in engineering. *International Symposium on Computers in Education (SIIE)*, (págs. 81-85). Setubal.
- Bokde, D., Girase, S., & Mukhopadhyay, D. (2015). An Approach to a University Recommendation by Multi-criteria Collaborative Filtering and Dimensionality Reduction

- Techniques. *2015 IEEE International Symposium on Nanoelectronic and Information Systems*. Indore, India.
- Borges, G., & Stiubiener, I. (2015). Recommending learning objects based on utility and learning style. *IEEE Frontiers in Education Conference (FIE)*. España.
- Bouraga, S., Jureta, I., Faulkner, S., & Herssens, C. (2014). Knowledge-Based Recommendation Systems: A Survey. *International Journal of Intelligent Information Technologies (IJIT)*, 1-19.
- Bourkougou, O., & El Bachari, E. (2016). E-learning personalization based on collaborative filtering and learner's preference. *Journal of Engineering Science and Technology*, 11, 1565 – 1581.
- Bouzeghoub, M., & D., K. (2006). Data Personalization: a Taxonomy of User Profiles Knowledge and a Profile Management Tool.
- Bouzeghoub, M., & Kostadinov, D. (2006). Data Personalization: a Taxonomy of User Profiles Knowledge and a Profile Management Tool.
- Bozyiğit, A., Bozyiğit, F., Kılınç, D., & Nasiboglu, E. (2018). Collaborative Filtering based Course Recommender using OWA operators. *International Symposium on Computers in Education (SIIE)*. Jerez de la Frontera, España.
- Brigui-Chtioui, I., Caillou, P., & Negre, E. (2017). Intelligent Digital Learning: Agent-Based Recommender System. *9th International Conference on Machine Learning and Computing (ICMLC)*. Singapore.
- Brik, M., & Touahria, M. (2020). Contextual Information Retrieval within Recommender System: Case Study, E-learning System. *TEM Journal*, 9, 1150-1162.
- Britto, J., Prabhu, S., Gawali, A., & Jadhav, Y. (2019). A Machine Learning Based Approach for Recommending Courses at Graduate Level. *International Conference on Smart Systems and Inventive Technology (ICSSIT)*, (págs. 127-131). Tirunelveli, India.
- Burke, R., Brusilovsky, P., Kobsa, A., & Nejdl, W. (2007). Hybrid Web Recommender Systems. En *The Adaptive Web: Methods and Strategies of Web Personalization* (Vol. 4321, págs. 377–408). Berlin: Springer.
- Campos, R., dos Santos, P., & Oliveira, J. (2018). Web-Based Recommendation System Architecture for Knowledge Reuse in MOOCs Ecosystems. In *Proceedings of the 2018 IEEE R. International Conference on Information Reuse and Integration (IRI)*, , (págs. 193-200). Salt Lake City, USA.
- Centelles, M. (2005). Taxonomías para la categorización y la organización de la información en sitios web. *Hipertext.net: Revista Académica sobre Documentación Digital y Comunicación Interactiva*(3).
- Cerón-Figueroa, S., López-Yáñez, I., Alhalabi, W., Camacho-Nieto, O., Villuendas-Rey, Y., Aldape-Pérez, M., & Yáñez-Márquez, C. (2017). Instance-based ontology matching for e-learning material using an associative pattern classifier. *Computers in Human Behavior*, 69, 218-225.

- Chen, H., Yin, C., Li, R., Rong, W., Xiong, Z., & David, B. (2020). Enhanced learning resource recommendation based on online learning style model. *Tsinghua Sci. Technol.*, 348-356.
- Chen, Q., Yu, X., Liu, N., Yuan, X., & Wang, Z. (2020). Personalized Course Recommendation Based on Eye-Tracking Technology and Deep Learning. *IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, (págs. 692-698). Sydney, NSW, Australia.
- Cheng, B., Zhang, Y., & Shi, D. (2018). Ontology-based personalized learning path recommendation for course learning. *2018 9th International Conference on Information Technology in Medicine and Education*. Hangzhou, China.
- Cheng, Y., & Bu, X. (2020). Research on Key Technologies of Personalized Education Resource Recommendation System Based on Big Data Environment. *Journal of Physics: Conference Series*, 1437.
- Chicaiza, J., Piedra, N., Lopez-Varga, J., & Tovar-Caro, S. (2017). Recommendation of open educational resources. An approach based on linked open data. *IEEE Global Engineering, E. Education Conference (EDUCON)*. Athens, Greece.
- Cornide-Reyes, H., Riquelme, F., Noel, R., Villarroel, R., Cechinel, C., Letelier, P., & Munoz, R. (2021). Key Skills to Work With Agile Frameworks in. *IEEE Access*, 9, 84724-84738.
- Cui, L.-Z., Guo, F.-L., & Liang, Y.-J. (2018). Research Overview of Educational Recommender Systems. *2nd International Conference on Computer Science and Application Engineering (CSAE)*. China.
- Dahdouh, K., Dakkak, A., Oughdir, L., & Ibriz, A. (2019). Large-scale e-learning recommender system based on Spark and Hadoop. *Journal of Big Data*, 6.
- Dang, Q. (2018). Implementing an individualized recommendation system using latent semantic analysis. *International Conference on Information and Education Technology (ICIET 2018)*, (págs. 239–243). Osaka, Japan.
- Das, A., & Akour, M. (2020). Intelligent Recommendation System for E-Learning using Membership Optimized Fuzzy Logic Classifier. *IEEE Pune Section International Conference (PuneCon)*, (págs. 1-10). Pune, India.
- De Borba, E., Gasparini, I., & Lichtnow, D. (2017). The Use of Time Dimension in Recommender Systems for Learning. *19th International Conference on Enterprise Information Systems (ICEIS 2017)*. Porto, Portugal.
- Demchenko, Y., A., B., Bernstein, D., & Laat, D. (2014). Experience of Profiling Curricula on Cloud Computing Technologies and Engineering for Different Target Groups. *IEEE 6th International Conference on Cloud Computing Technology and Science*, (págs. 949-955). Conference on Cloud Computing Technology and Science, Singapore.
- Dhar, J., & Jodder, A. (2020). An Effective Recommendation System to Forecast the Best Educational Program Using Machine Learning Classification Algorithms. *Ingénierie Systèmes Inf.*, 25, 559–568.

- Dwivedi, S., & Roshni, V. (2017). Recommender system for big data in education. *2017 5th National Conference on E-Learning and E-Learning Technologies (ELELTECH)*. Hyderabad, India.
- El Fazazi, H., Qbadou, M., Sahi, I., & Mansouri, K. (2018). Personalized recommender system for e-Learning environment based on student's preferences. *Int. J. Comput. Sci. Netw. Secur.*, *18*, 173–178.
- El Mabrouk, M., Gaou, S., & Rtili, M. (2017). Towards an Intelligent Hybrid Recommendation System for E-Learning Platforms Using Data Mining. *Int. J. Emerg. Technol. Learn.*, *52-76*.
- Elbadrawy, A., & Karypis, G. (2016). Domain-Aware Grade Prediction and Top-n Course Recommendation. *ACM Conference on Recommender Systems (RecSys '16)*. Boston.
- Fazeli, S., Rajabi, E., Lezcano, L., Drachsler, H., & Sloep, P. (2016). Supporting Users of Open Online Courses with Recommendations: An Algorithmic Study. *IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*. Austin, Texas.
- Fernández, M., Gómez-Pérez, A., & Juristo, N. (1997). METHONTOLOGY: From Ontological Art Towards Ontological Engineering. *AAAI 1997*, 33-40.
- Ganeshan, K., & Li, X. (2015). An intelligent student advising system using collaborative filtering. *IEEE Frontiers in Education Conference (FIE)*. El Paso, Texas.
- Gianotti, R., Cazella, S., & Behar, P. (2019). A Model for Integrating Personality Traits into an Educational Recommender System. *IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, (págs. 383-385). Maceio, Brazil.
- Gong, J., Wang, S., Wang, J., Feng, W., Peng, H., Tang, J., & Yu, P. (2020). Attentional Graph Convolutional Networks for Knowledge Concept Recommendation in MOOCs in a Heterogeneous View. *43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20)*, (págs. 78-88). Virtual Event, Xi'an, China.
- Gordillo, A., López-Fernández, D., & Verbert, K. (2020). Examining the Usefulness of Quality Scores for Generating Learning Object Recommendations in Repositories of Open Educational Resources. *Appl. Sci.* .
- Gugnani, A., Kasireddy, V., & Ponnalagu, K. (2018). Generating Unified Candidate Skill Graph for Career Path Recommendation. *IEEE International Conference on Data Mining Workshops (ICDMW)*, (págs. 328-333). Singapore.
- Gulzar, Z. &. (2018). PCRS: Personalized Course Recommender System Based on Hybrid Approach. *Computer Science*.
- Guo, X., Yin, S.-C., Zhang, Y.-W., Li, W., & He, Q. (2019). Cold Start Recommendation Based on Attribute-Fused Singular Value Decomposition. *IEEE Access*, *7*, 11349 - 11359.
- Hajri, H., Bourda, Y., & Popineau, F. (2017). MORS: A System for Recommending OERs in a MOOC . *IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*. Timisoara, Romania.

- Harrathi, M., Touzani, N., & Braham, R. (2017). A Hybrid Knowledge-Based Approach for Recommending Massive Learning Activities. *14th International Conference on Computer Systems and Applications (AICCSA)*.
- Hasan, M., Ahmed, S., Abdullah, D., & Rahman, M. (2016). Graduate school recommender system: Assisting admission seekers to apply for graduate studies in appropriate graduate schools. *5th International Conference on Informatics; Electronics and Vision (ICIEV)*. Dhaka, Bangladesh.
- Hawashin, B., Aqel, D., AlZu'bi, S., & Jararweh, Y. (2019). Novel Weighted Interest Similarity Measurement for Recommender Systems Using Rating Timestamp. *Sixth International Conference on Software Defined Systems (SDS)*. Rome, Italy.
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22(1), 5-53.
- Hidayat, A., Suwawi, D., & Laksitowening, K. (2020). Learning Content Recommendations on Personalized Learning Environment Using Collaborative Filtering Method. *8th International Conference on Information and Communication Technology (ICoICT)*, (págs. 1-6). Yogyakarta, Indonesia.
- Hinz, V. T., & Pimenta, M. S. (2018). Integrating Reputation to Recommendation Techniques in an e-learning Environment. *17th International Conference on Information Technology Based Higher Education and Training (ITHET)*. Olhao, Portugal.
- Huang, X., Tang, Y., Qu, R., Li, C., Yuan, C., Sun, S., & Xu, B. (2018). Course Recommendation Model in Academic Social Networks Based on Association Rules and Multi-similarity. *IEEE 22nd International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. Nanjing, China.
- Huang, Z., Liu, Q., Yin, Y., Chen, E., Gao, W., & Hu, G. (2019). Exploring Multi-Objective Exercise Recommendations in Online Education. *28th ACM International Conference on Information and Knowledge Management (CIKM '19)*, (págs. 1261–1270). Beijing, China.
- Huo, Y., Xiao, J., & Ni, L. (2018). Towards Personalized Learning Through Class Contextual Factors-Based Exercise Recommendation. *IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)*, (págs. 85-92). Singapore.
- Ibrahim, M., Yang, Y., Ndzi, D., Yang, G., & Almaliki, M. (2018). Ontology-based personalised course recommendation framework. *IEEE Access*, 7, 5180–5199.
- Iordache, C., Mariën, I., & Baelden, D. (2017). Developing Digital Skills and Competences: A Quick-Scan Analysis of 13 Digital Literacy Models. *Italian Journal of Sociology of Education*, 9, 6-30.
- Iqbal, M., Ghazanfar, M. A., Sattar, A., Maqsood, M., Khan, S., Mehmood, I., & Baik, S. W. (2019). Kernel Context Recommender System (KCR): A Scalable Context-Aware Recommender System Algorithm. *IEEE Access*, 24719-24737.
- Isma'il, M., Haruna, U., Aliyu, G., Abdulmumin, I., & Adamu, S. (2020). An Autonomous Courses Recommender System for Undergraduate Using Machine Learning Techniques.

- International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS)*, (págs. 1-6). Ayobo, Nigeria.
- Jain, S., Grover, A., Thakur, P. S., & Choudhary, S. K. (2015). Trends, problems and solutions of recommender system. *International Conference on Computing, Communication Automation*, (págs. 955-958). Greater Noida, India.
- Jiang, W., & Pardos, Z. (2019). Time slice imputation for personalized goal-based recommendation in higher education. *13th ACM Conference on Recommender Systems (RecSys '19)*. New York.
- Jiang, W., Pardos, Z., & Wei, Q. (2019). Goal-based Course Recommendation. *Learning Analytics and Knowledge*, (págs. 36-45). Tempe, USA.
- Kálmán, A., Molnár, G., & Szűts, Z. (2018). Issues of Lifelong Learning - Behavioral ends of teaching and learning through ICT. *9th IEEE International Conference on Cognitive Informatics (Cog InfoCom)*. Hungary.
- Kamila, V. Z., Subastian, E., & Rosmasari. (2019). KNN and Naive Bayes for Optional Advanced Courses Recommendation. *International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*. Denpasar, Indonesia.
- Keinänen, M., Ursin, J., & Nissinen, K. (2018). How to measure students' innovation competences in higher education : Evaluation of an assessment tool in authentic learning environments. *Studies in Educational Evaluation*, 58, 30-36.
- Khacheb, A., & Cheniti Belcadhi, L. (2016). Open assessment resources recommendation for competence assessment. *Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM '16)*. Salamanca.
- Koh, M., & Chew, Y. (2015). Intelligent Job Matching with Self-learning Recommendation Engine. *Procedia Manu-facturing*, 1959-1965.
- Kurilovas, E., Kurilova, J., Kurilova, I., & Melesko, J. (2016). Personalised learning system based on students' Learning styles and application of intelligent technologies. *International Technology, Education and Development Conference*.
- Kuznetsov, S., Kordík, P., Tomás Rehorek, J. D., & Kroha, P. (2016). Reducing cold start problems in educational recommender systems. *International Joint Conference on Neural Networks (IJCNN)*. Vancouver, BC, Canada.
- Laal, M. (2011). Lifelong Learning: What Does It Mean? *Procedia-Social and Behavioral Sciences*, 28, 470-474.
- Lessa, L., & Brandao, W. (2018). Filtering Graduate Courses based on LinkedIn Profiles. *WebMedia 2018*. Salvador, Brazil.
- Li, H., Li, H., Zhang, S., Zhong, Z., & Cheng, J. (2018). Intelligent learning system based on personalized recommendation technology. *Neural Comput. Z Appl.*, 31, 1-8.
- Li, H., Shi, J., Zhang, S., & Yun, H. (2017). Implementation of intelligent recommendation system for learning resources. *12th International Conference on Computer Science and Education (ICCSE)*, . Houston, USA.

- Li, J., & Ye, Z. (2020). Course Recommendations in Online Education Based on Collaborative Filtering Recommendation Algorithm. *Complexity*.
- Li, Y., & Li, H. (2017). MOOC-FRS: A new fusion recommender system for MOOCs. *IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*. Chongqing, China.
- Liao, T., Feng, X., Sun, Y., Wang, H., Liao, C., & Li, J. (2020). Online Teaching Platform Based on Big Data Recommendation System. *5th International Conference on Information and Education Innovations (ICIEI 2020)*, (págs. 35-39). London, UK.
- Liu, H., Ding, J., Yang, L., Guo, Y., Wang, X., & Deng, A. (2020). Multi-Dimensional Correlative Recommendation and Adaptive Clustering via Incremental Tensor Decomposition for Sustainable Smart Education. *IEEE Trans. Sustain. Comput.* 2020, 5, 389–402.
- Mana, S. C., & Sasipraba, T. (2021). A Machine Learning Based Implementation of Product and Service Recommendation Models. *7th International Conference on Electrical Energy Systems (ICEES)*, (págs. 543-547). Chennai, India.
- Melão, N., & Reis, J. (2020). Using Social Networks in Personnel Selection: A Survey of Human Resource Professionals. *15th Iberian Conference on Information Systems and Technologies (CISTI)*. Sevilla.
- Meryem, G., Douzi, K., & Chantit, S. (2016). Toward an E-orientation platform: Using hybrid recommendation systems. *11th International Conference on Intelligent Systems: Theories and Applications (SITA)*. Morocco.
- Mondal, B., Patra, O., Mishra, S., & Patra, P. (2020). A course recommendation system based on grades. *International Conference on Computer Science; Engineering and Applications (ICCSEA)*, (págs. 1-5). Gunupur, India.
- Montuschi, P., Lamberti, F., Gatteschi, V., & Demartini, C. (2015). A Semantic Recommender System for Adaptive Learning in IT Pro- 503. *IT Professional*, 17(5), 50-58.
- Moreira, F., Ferreira, M., Collazos, C., & Cano, S. (2017). Profile-oriented programming teaching to non technical students: A case study. *Iberian Conference on Information Systems and Technologies (CISTI)*, (págs. 1-6). Lisbon.
- Morrow, T., Hurson, A. R., & Sedigh Sarvestani, S. (2020). Algorithmic Support for Personalized Course Selection and Scheduling. *IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*. Madrid.
- Morsomme, R., & Alferez, S. C.-b. (2019). Content-based Course Recommender System for Liberal Arts Education. *International Conference on Educational Data Mining (EDM)*. Montreal, Canada.
- Mrhar, K., & Abik, M. (2019). Toward a Deep Recommender System for MOOCs Platforms. *3rd International Conference on Advances in Artificial Intelligence (ICA AI 2019)*, (págs. 173-177). Istanbul, Turkey.
- Na, C., Xiuyuan, Z., & Na, L. (2020). Research on Network Personalized Learning Recommendation Algorithm Based on Knowledge Representation. *International*

- Conference on Information Science, Parallel and Distributed Systems (ISPDS)*, (págs. 348-351). Xi'an, China.
- Nadeem, M., Dustin, S., & Shane, M. (2018). Neural Educational Recommendation Engine (NERE). *IEEE International Conference on Data Mining Workshops (ICDMW)*. Singapore.
- Nadine, U., Cao, H., & Deng, J. (2016). Competitive recommendation algorithm for E-commerce. *12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. Changsha, China.
- Niyigena, J., & Jiang, Q. (2020). A Hybrid Model for E-Learning Resources Recommendations in the Developing Countries . *4th International Conference on Deep Learning Technologies (ICDLT 2020)*, (págs. 21-25). Beijing, China.
- Noy, N., & McGuinness, D. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford University, Knowledge Systems Laboratory. 32, Stanford, CA.
- Obeid, C., Lahoud, I., El Khoury, H., & Champin, P.-A. (2018). Ontology-Based Recommender System in Higher Education. *International World Wide Web Conferences Steering Committee*. Republic and Canton of Geneva.
- Ochirbat, A., & Shih, T. (2016). Occupation Recommendation with Major Programs for Adolescents. *International Symposium on Grids and Clouds*. Tapei.
- ONU. (2015). *La Agenda para el Desarrollo Sostenible*. Recuperado el 25 de 07 de 2021, de <https://www.un.org/sustainabledevelopment/es/development-agenda/>
- ONU. (s.f.). *Objetivo 4*. Recuperado el 25 de 07 de 2021, de <https://www.un.org/sustainabledevelopment/es/education/>
- Ortega, F., Sánchez, J.-L., Bobadilla, J., & Gutiérrez, A. (2013). Improving Collaborative Filtering Based Recommender Systems Using Pareto Dominance. *Information Sciences*, 239, 50-61.
- Otebolaku, A., & Andrade, M. (2014). Context-Aware Media Recommendations. *28th International*, (págs. 191-196). Canadá.
- Pan, X., Li, X., & Lu, M. (2020). A MultiView Courses Recommendation System Based on Deep Learning. *International Conference on Big Data and Informatization Education (ICBDIE)*, (págs. 502-506). Zhangjiajie, China.
- Pandey, S., Lan, A., Karypis, G., & Srivastava, J. (2020). Learning Student Interest Trajectory for MOOC Thread Recommendation. *International Conference on Data Mining Workshops (ICDMW)*, (págs. 400-407). Sorrento, Italy.
- Pardos, Z., & Jiang, W. (2019). Designing for serendipity in a university course recommendation system. *Tenth International Conference on Learning Analytics and Knowledge (LAK '20)*, (págs. 350-359). Frankfurt, Germany.
- Park, Y. (2020). Supporting Student Success in Computing Courses through Personalized Peer Advice. *51st ACM Technical Symposium on Computer Science Education (SIGCSE '20)*. Portland, OR, USA.

- Patel, B., Kakuste, V., & Eirinaki, M. (2017). CaPaR: A Career Path Recommendation Framework. *2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService)*. San Francisco.
- Pereira, C., Campos, F., Ströele, V., David, J., & Braga, R. (2018). BROAD-RSI—Educational recommender system using social networks interactions and linked data. *Journal of Internet Services and Applications*, 9(7).
- Prisco, A., dos Santos, R., Bez, J., Tonin, N., Neves, M., Teixeira, D., & Botelho, S. A. (2019). A Facebook chat bot as recommendation system for programming problems. *IEEE Frontiers in Education Conference (FIE)*, (págs. 1-5). Covington, KY, USA.
- Quang, N., Luan, H., Pham, L., & Chu, T. (2019). Ontology-Based Recommender System for Sport Events. *13th International Conference on Ubiquitous Information Management and Communication (IMCOM)*.
- Raitskaya, L., & Tikhonova, E. (2019). Skills and Competencies in Higher Education and Beyond. *Journal of Language and Education*, 5(4), 4-8.
- Ren, Z., Ning, X., Lan, A., & Rangwala, H. (2019). Grade Prediction with Neural Collaborative Filtering. *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, (págs. 1-10). Washington, DC, USA.
- Ricci, F., Rokach, L., & Shapira, B. (2015). *Recommender Systems Handbook*. New York 2015: Springer Science+Business Media.
- Rismanto, R., Rachmad, S., & Agusta, B. (2020). Research Supervisor Recommendation System Based on Topic Conformity. *Int. Journal Modern Education and Computer Science*, 26-34.
- Rivera, A., Tapiá-Leon, M., & Luján-Mora, S. (2018). Recommendation Systems in Education: A Systematic Mapping Study. *International Conference on Information Technology and Systems (ICITS 2018)*, (págs. 10-12). Ecuador.
- Rivero, D. S. (2008). *Metodología de la Investigación*. Editorial Shalom.
- Rizun, M. (2019). Concept of Recommender System for Building an Individual Educational Profile. *18th International Conference on Perspectives in Business Informatics Research (BIR 2019)*. Katowice, Poland.
- Rodríguez, P. A., Ovalle, D. A., & Duque, N. D. (2015). A Student-Centered Hybrid Recommender System to Provide Relevant Learning Objects from Repositories. *Learning and Collaboration Technologies*, 291 - 300.
- Ruchika, Singh, A. V., & Sharma, M. (2017). Building an effective recommender system using machine learning based framework. *International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS)*, (págs. 215-219). Dubai, United Arab Emirates.
- Samin, H., & Azim, T. (2019). Knowledge Based Recommender System for Academia Using Machine Learning: A Case Study on Higher Education Landscape of Pakistan. *IEEE Access*, 7, 67081-67093.

- Sampieri, R. H., Collado, C. F., & Lucio, P. B. (2014). *Metodología de la Investigación* (Sexta ed.). México: McGraw-Hill.
- Schwarz, M., Lobur, M., & Stekh, Y. (2017). Analysis of the effectiveness of similarity measures for recommender systems. *14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*. Lviv, Ukraine.
- Sebbaq, H., el Faddouli, N., & Bennani, S. (2020). Recommender System to Support MOOCs Teachers: Framework based on Ontology and Linked Data. *13th International Conference on Intelligent Systems: Theories and Applications*. New York.
- Sergis, S., & Sampson, D. (2015). Enhancing Learning Object Recommendations for Teachers Using Adaptive Neighbor Selection. *IEEE 15th International Conference on Advanced Learning Technologies*. Hualien, Taiwan.
- Shao, J. (2017). Research on Fuzzy Ontology E-learning Based on User Profile. *International Conference on E-Education, E-Business and E-Technology (ICEBT 2017)*. Toronto, Canada.
- Shi, K., Wen, B., Wang, J., & Ouyang, Z. (2019). Research for Personalized Learning Resource Recommendation Model based on Academic . *IEEE International Conference on Computer Science and Educational Informatization (CSEI)*, (págs. 55-58). Kunming, China.
- Singh, J. (2020). Collaborative Filtering based Hybrid Music Recommendation System. *Third International Conference on Intelligent Sustainable Systems [ICISS 2020]*, (págs. 186-190). Thoothukudi, India.
- Slim, A., Hush, D., Ojha, T., Abdallah, C., Heileman, G., & El-Howayek, G. (2019). An Automated Framework to Recommend A Suitable Academic Program; Course and Instructor. *IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*, (págs. 145-150). Newark, USA.
- Stromquist, N., & da Costa, R. (2017). Popular universities: An alternative vision for lifelong learning in Europe. *International Review of Education*, 63(5), 725-774.
- Suganya, G., Mariappan, P., Dubey, P., Drolia, A., & Srihari, S. (2020). Subjective Areas of Improvement: A Personalized Recommendation. *Procedia Comput. Sci.*, 235–239.
- Suhaim, A. B., & Berri, J. (2021). Context-Aware Recommender Systems for Social. (B. P., K. A., & N. W., Edits.) *IEEE Access*.
- Symeonidis, P., & Malakoudis, D. (2016). MoocRec.com: Massive Open Online Courses Recommender System . *10th ACM Conference on Recommender System*. Boston.
- Taheri, S. M., & Irajian, I. (2018). DeepMovRS: A unified framework for deep learning-based movie recommender systems. *6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*, (págs. 200-204). Kerman, Iran.
- Tan, J., Chang, L., Liu, T., & Zhao, X. (2020). Attentional Autoencoder for Course Recommendation in MOOC with Course Relevance. *International Conference on Cyber-*

- Enabled Distributed Computing and Knowledge Discovery (CyberC)*, (págs. 190-196). Chongqing, China.
- Tang, J., Wu, S., Sun, J., & Su, H. (2012). Cross-domain collaboration recommendation. *18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*, (págs. 1285-1293). Beijing.
- Tarus, J., Niu, Z., & Mustafa, G. (2018). Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. *Artificial Intelligence Review*, 21-48.
- Tarus, J., Niu, Z., & Yousif, A. (2017). A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining. *Future Generation Computer Systems*, 37-48.
- Tavakoli, M., Mol, S., & Kismihók, G. (2020). Labour Market Information Driven; Personalized; OER Recommendation System for Lifelong Learners. *12th International Conference on Computer Supported Education*. On line.
- Tejeda-Lorente, Á., Moreno, J., Porcel, C., Galindo-Moreno, P., & Herrera-Viedma, E. (2015). A Dynamic Recommender System as Reinforcement for Personalized Education by a Fuzzy Linguistic Web System. *3rd International Conference*. Rio De Janeiro, Brazil.
- Thompson, A., Talbot, R., Doughty, L., Huvar, H., Le, P., & Hartley, L. (2020). Development and application of the Action Taxonomy for Learning Assistants (ATLAs). *International Journal of STEM Education*.
- Thorat, P., Goudar, R., & Barve, S. (2015). Survey on Collaborative Filtering and Content-Based Recommending. *Int. J. Comput. Appl.*(110), 31-36.
- UIL. (2021). *Adoptar una cultura de aprendizaje a lo largo de la vida*. UNESCO. Hamburgo: Unesco.
- ul haq Dar, E., & ) Dorn, J. (2018). Ontology based classification system for online job offers. *International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, (págs. 1-8). Sukkur, Pakistan.
- UNESCO. (2016). *Educación 2030: Declaración de Incheon - Hacia una educación inclusiva y equitativa de calidad y un aprendizaje a lo largo de la vida para todos*. Unesco.
- Vaidhehi, V., & Suchithra, R. (2019). An Enhanced Approach Using Collaborative Filtering For Generating Under Graduate Program Recommendations. *Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, (págs. 1-6). Gangtok, India.
- Valentin, C., Emrich, A., . Lahann, J., Werth, D., & Loos, P. (2015). Adaptive Social Media Skills Trainer for Vocational Education and Training: Concept and Implementation of a Recommender System," 2015 ., *48th Hawaii International Conference on System Sciences*. Kauai, HI.

- van Laar, E., van Deursen, A., van Dijk, J., & de Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in human behavior*, 72, 577-588.
- Velasquez, T., Castro, H., & Gaona, D. (2018). Ontological Representation Applied to the Systems Engineering. *Congreso Argentino de Ciencias de la Informática y Desarrollos de Investigación (CACIDI)*, (págs. 1-5). Buenos Aires, Argentina.
- Venkatraman, S. (2017). A proposed Business Intelligent Framework for Recommender System. *Informatics*, 4(40).
- Venkatraman, S. (2017). A Proposed Business Intelligent Framework for Recommender Systems. *Informatics*.
- Wan, N., Wu, X., Guo, S., Yang, L., Han, Q., & Yin, R. (2020). Research on Personalized Recommendation of learning Resource Based on Big Data of Education. *IEEE 2nd International Conference on Computer Science and Educational Informatization (CSEI)*, (págs. 306-311). Xinxiang, China.
- Wang, C., Zhu, H., Zhu, C., Zhang, X., Chen, E., & Xiong, H. (2020). Personalized Employee Training Course Recommendation with Career Development Awareness. *The Web Conference 2020*. Taipei, Taiwan.
- Wang, J., Xie, H., Au, O., Zou, D., & Wang, F. (2020). Attention-Based CNN for Personalized Course Recommendations for MOOC Learners. *International Symposium on Educational Technology (ISET)*, (págs. 180-184). Bangkok, Thailand.
- Wiktorski, T., Demchenko, Y., & Belloum, A. (2017). Model Curricula for Data Science EDISON Data Science Framework. *IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, (págs. 69-74). Hong Kong.
- Wita, R., Bubphachuen, K., & Chawachat, J. (2017). Content-based Filtering Recommendation in Abstract Search using Neo4j. *21st International Computer Science and Engineering Conference (ICSEC)*, (págs. 136-140). Bangkok, Thailand.
- Wu, W., Wang, B., Liu, Y., Yin, L., & Zheng, W. (2020). Higher Education Online Courses Personalized Recommendation Algorithm Based on Score and Attributes. *Journal of Physics Conference Series*.
- Xu, J., Xing, T., & van der Schaar, M. (2016). Personalized Course Sequence Recommendations. *IEEE Transactions on Signal Processing*, 64(20), 5340-5352.
- Xu, Z., Tifrea-Marcuska, O., Lukasiewicz, T., Martinez, M., Simari, G., & Chen, C. (2018). Lightweight Tag-Aware Personalized Recommendation on the Social Web Using Ontological Similarity. *IEEE Access*, 6, 35590–35610.
- Yan, Y., Hara, K., Kazuma, T., & He, A. (2017). A Method for Personalized C Programming Learning Contents Recommendation to Enhance Traditional Instruction. *IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*, (págs. 320–327). Taipei, Taiwan.

- Yanes, N., Mostafa, A. M., Exx, M., & Almuayqil, S. N. (2020). A Machine Learning-Based Recommender System for Improving Students Learning Experiences. *IEEE access*, 8, 201218 - 201235.
- Yang, L., Cormican, K., & Yu, M. (2019). Learning Systems Engineering Domain Ontologies from Text Documents. *International Symposium on Systems Engineering (ISSE)*, (págs. 1-7). Edinburgh, UK.
- Yildirim, S., & Baur, S. (2015). Development of Learning Taxonomy for an Undergraduate Course Based-On Active Learning in Architectural Engineering Program. *American Society for Engineering Education*.
- Zhang, M., Liu, S., & Wang, Y. (2020). STR-SA: Session-based Thread Recommendation for Online Course Forum with Self-Attention. *IEEE Global Engineering Education Conference (EDUCON)*, (págs. 374-381). , Porto, Portugal.
- Zhao, Q., Wang, C., Wang, P., Zhou, M., & Jiang, C. (2018). A Novel Method on Information Recommendation via Hybrid Similarity. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 448-459.
- Zheng, X., Chen, C., Hung, J., He, W., Hong, F., & Lin, Z. (2015). A Hybrid Trust-Based Recommender System for Online Communities of Practice. *2015 IEEE 15th International Conference on Advanced Learning Technologies*. Hualien, Taiwan.
- Zhu, H., Liu, Y., Tian, F., Ni, Y., Wu, K., Chen, Y., & Zheng, Q. (2018). A Cross-Curriculum Video Recommendation Algorithm Based on a Video-Associated Knowledge Map. *IEEE Access*, 6, 57562-57571.