

Centro Nacional de Investigación y Desarrollo Tecnológico

Subdirección Académica

Departamento de Ciencias Computacionales

TESIS DE MAESTRÍA EN CIENCIAS

**Servicios de Recomendación para la Explotación de Repositorios de Ítems
Heterogéneos**

presentada por
L. I. Jorge Lara Ontiveros

como requisito para la obtención del grado de
Maestro en Ciencias de la Computación

Director de tesis
Dr. Juan Gabriel González Serna

Codirector de tesis
Dr. Máximo López Sánchez

Cuernavaca, Morelos, México. Junio de 2016.

Cuernavaca, Morelos a 23 de mayo del 2016

OFICIO No. DCC/075/2016

Asunto: Aceptación de documento de tesis

C. DR. GERARDO V. GUERRERO RAMÍREZ
SUBDIRECTOR ACADÉMICO
PRESENTE

Por este conducto, los integrantes de Comité Tutorial del **C. Jorge Lara Ontiveros**, con número de control M14CE013, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis profesional titulado **“Servicios de Recomendación para la Explotación de Repositorios de Ítems Heterogéneos”** y hemos encontrado que se han realizado todas las correcciones y observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.

DIRECTOR DE TESIS



Dr. Juan Gabriel González Serna
Doctor en Ciencias de la
Computación
7820329

CO-DIRECTOR DE TESIS



Dr. Máximo López Sánchez
Doctor en Ciencias de la
Computación
7498547

REVISOR 1



Dr. Noé Alejandro Castro Sánchez
Doctor en Ciencias de la
Computación
08701806

REVISOR 2



Dra. Alicia Martínez Rebollar
Doctora en Informática
7399055

C.p. Lic. Guadalupe Garrido Rivera - Jefa del Departamento de Servicios Escolares.
Estudiante
Expediente

AMR/lmz

Cuernavaca, Mor., 25 de mayo de 2016
OFICIO No. SAC/177/2016

Asunto: Autorización de impresión de tesis

**C. JORGE LARA ONTIVEROS
CANDIDATO AL GRADO DE MAESTRO EN CIENCIAS
DE LA COMPUTACIÓN
PRESENTE**

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado "**Servicios de Recomendación para la Exploración de Repositorios de Ítems Heterogéneos**", ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

ATENTAMENTE
"CONOCIMIENTO Y TECNOLOGÍA AL SERVICIO DE MÉXICO"



DR. GERARDO VICENTE GUERRERO RAMÍREZ
SUBDIRECTOR ACADÉMICO



SEP TecNM
CENTRO NACIONAL
DE INVESTIGACIÓN
Y DESARROLLO
TECNOLÓGICO
SUBDIRECCIÓN
ACADÉMICA

C.p. Lic. Guadalupe Garrido Rivera.- Jefa del Departamento de Servicios Escolares.
Expediente

GVGR/mcr

Agradecimientos

En primer lugar quiero dar gracias a Dios por permitirme cumplir esta meta.

A mi esposa Trini que estuvo conmigo durante todo este tiempo; sus palabras de aliento y apoyo incondicional siempre estuvieron presentes.

A mis padres que siempre me dijeron que estudiar sería la mejor herencia que ellos me podrían dejar, siempre me han apoyado y guiado de la mejor manera para lograr mis metas y me inculcaron el amor a Dios.

Deseo agradecer también a mi Director de Tesis el Dr. Juan Gabriel González Serna por su tiempo, brindarme su apoyo y por ser una persona profesional que influyo mucho en mí para llevar acabo el desarrollo de esta tesis. A mis revisores por el tiempo dedicado a las correcciones y consejos acerca del trabajo desarrollado.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico brindado para el desarrollo de esta tesis.

Al Centro Nacional de Investigación y Desarrollo Tecnológico que me permitió realizar la Maestría en Ciencias y crecer en el aspecto profesional.

A todos mis compañeros que me brindaron su amistad en este tiempo.

Resumen

Las instituciones de educación superior cuentan con capital intelectual. Entre los que podemos encontrar está el capital humano, el capital estructural y el capital relacional. Los cuales forman un conjunto de ítems heterogéneos que se encuentran contenidos en repositorios de información. Sin embargo, con el crecimiento de las instituciones la localización de información relevante se ha convertido en un problema.

Como solución a esto han surgido los sistemas de recomendación, los cuales se han planteado como una posibilidad de apoyo a los usuarios ayudándolos en la localización automática de ítems interesantes. Los Sistemas de Recomendación son herramientas que brindan una solución para el manejo de la gran cantidad de información que día a día se incrementa en la WWW (Pinter, Marušić, & Radosav, 2012), ayudando al usuario en el filtrado de la información que puede ser de utilidad según sus necesidades específicas.

Este trabajo tiene dos objetivos: el primero es implementar un servicio de recomendación que utilice el algoritmo basada en contenido para explotar un repositorio de ítems heterogéneos, el cual contiene los siguientes tipos de ítems: académicos, objetos de conocimiento, instalaciones especializadas, infraestructura especializada, certificaciones, cursos, servicios tecnológicos y proyectos; el segundo es implementar la metodología de explicación que permita crear explicaciones de las recomendaciones inferidas para los usuarios, enfocándose en crear una para cada tipo de ítem.

Para llevar a cabo la evaluación del sistema de recomendación se utilizaron las curvas ROC. Estas intentan medir el grado en que un sistema de filtrado de información puede distinguir con éxito entre ruido y señal (relevancia). Las curvas ROC tienen un filtro o límite de predicción, a partir del cual todos los ítems que el sistema sitúa por encima del límite son vistos por el usuario. Se decidió emplear la curva ROC para evaluar el rendimiento del algoritmo de recomendación, ya que se busca encontrar los parámetros que mejores resultados arrojen, que recuperen la mayor cantidad de ítems relevantes y descarten aquellos que sean irrelevantes. Estas pruebas permitirán fijar los mejores parámetros para el sistema de recomendación.

Dentro de estas pruebas se ha determinado la necesidad de realizar las evaluaciones dividiéndolas según los tipos de ítems existentes dentro de la plataforma, es decir, académico, objeto de conocimiento, instalación especializada, infraestructura tecnológica, curso, servicio tecnológico, certificación y proyecto de investigación.

Los umbrales que mejores resultados ofrecieron tomando en consideración que extraiga una mayor cantidad de ítems relevantes y una baja cantidad de ítems irrelevantes fueron:

- Académico es $> .40$
- Objeto de conocimiento es $> .25$
- Instalación especializada es $> .31$
- Infraestructura tecnológica es $> .26$
- Curso es $> .26$
- Servicio tecnológico es $> .30$

Abstract

Universities and colleges education have human capital, structural capital and relational capital. Which form a heterogeneous set of items stored in repositories of information. However, with the growth of universities retrieve relevant information has become a problem.

A solution to the problem of retrieving relevant information are recommender systems, which help the user automatically retrieve relevant information. Recommendation Systems (RS) are tools that provide solution to handle large amounts of information, that growing every day in the WWW (Pinter, Marušić, & Radosav, 2012), the RS helping the user to filter information that can be useful according to their needs.

Therefore, in this thesis two objectives were proposed: the first was to develop a recommender service using a content-based algorithm to retrieve relevant information from a repository of heterogeneous items, which contains the following types of items: teachers, knowledge objects, specialized facilities, specialized infrastructure, certifications, courses, technology services and projects; the second objective was to implement a methodology to generate explanations of recommendations provided to users.

To carry out the evaluation of recommendation system, ROC curves were used. This attempt to measure the extent to which an information filtering system can successfully distinguish between noise and signal (relevance). The ROC curves have a filter or prediction limit, from which all the items that the system positioned above the limit are user views. It was decided to use the ROC curve to evaluate the performance of the algorithm recommendation, as it seeks to find the parameters that best results shed, to recover the largest amount of relevant items and discard those that are irrelevant. These tests will set the best parameters for the recommendation system.

Within these tests, it has determined the need for the dividing assessments according to the types of items within the platform, i.e., teachers, knowledge object, specialized facility, specialized infrastructure, course, technological service, certification and research project.

The thresholds that gave better results because they get more items that are relevant and few irrelevant items were:

- Teachers is > .40
- Knowledge Object is > .25
- Specialized facility is > .31
- Specialized infrastructure is > .26
- Course is > .26
- Technology service is > .30

Contenido

CAPÍTULO 1	INTRODUCCIÓN	1
1.1	Introducción	2
1.2	Antecedentes	3
1.3	Descripción del problema	8
1.4	Objetivos	9
1.5	Justificación	9
1.6	Estructura del documento	10
CAPÍTULO 2		11
2.1	Sistemas de recomendación	12
2.2	Clasificaciones de sistemas de recomendación	13
2.3	Técnicas de filtrado colaborativo	14
2.3.1.	Técnicas de filtrado basado en contenidos	14
2.3.2.	Técnicas de filtrado híbrido	14
2.4	Sistemas de Recomendación Conscientes del Contexto (SRCC)	15
2.5	Sistemas de Recomendación en ambientes organizacionales	15
2.6	Explicación y Argumentación	17
2.6.1.	Explicaciones en los sistemas de recomendación	17
2.6.2.	Mapas conceptuales	18
2.6.3.	Mapas Mentales	19
2.7	Evaluación de SR centrada en el usuario	20
CAPÍTULO 3		21
3.1	Recomendación de artículos académicos a través de propósitos de lectura de los usuarios	22
3.2	Sistema de recomendación basado en Argumentación de confianza habilitada	24
3.3	Asesoramiento automatizado de lectores para hacer recomendaciones de libros para lectores K-12	26
3.4	Modelo basado en características contextuales más influyentes para sistema de recomendación sensible al contexto	28
3.5	Combinación de filtrado colaborativo y similitud de texto para recomendaciones de perfiles de expertos en sitios Web Sociales	30
3.6	Un marco de trabajo para la creación rápida de prototipos de sistemas de recomendación basados en conocimiento en el dominio de aprendizaje	33

3.7 SAPRS: sistema de recomendación proactivo consciente de la situación con explicaciones.....	34
3.8 Recomendaciones para la empresa: eventos, contacto y grupo	35
3.9 Repensando el ecosistema de investigación en recomendación: reproducibilidad, la apertura y lenskit.....	36
3.10 Estudio comparativo	37
CAPÍTULO 4	39
4.1 Introducción	40
4.2 Fase de análisis	40
4.2.1. Análisis del funcionamiento del SR y las técnicas utilizadas	40
4.2.2. Fase de Diseño	50
4.2.3. Análisis de los estilos de explicación a implementar	50
4.2.4. Identificación de los objetivos de las explicaciones en el SR	50
4.2.5. Definición de la estructura sintáctica de las explicaciones	51
4.2.6. Definición de la presentación de las explicaciones	59
CAPÍTULO 5	62
5.1 Arquitectura del servicio de recomendación para la PCCT del TecNM.	63
5.1.1. Módulo de poblado y actualización de estatus de ítem	65
5.1.2. Módulo de poblado y actualización de estatus de usuario.....	66
5.1.3. Módulo motor de recomendación	67
5.1.4. Módulo de explicaciones.....	68
5.1.5. Interfaz de usuario	69
5.1.6. Módulo de valoración	69
5.2 Implementación	70
5.2.1. Aplicación backend.....	70
5.2.2. Implementación del motor de recomendación	73
5.2.3. Módulo de explicación	76
5.2.4. Interfaz de usuario	77
5.2.5. Módulo de valoración.	78
CAPÍTULO 6	80
6.1 Introducción	81
6.2 I. Tareas del usuario para el sistema de recomendación.	81
6.2.1. Anotación en contexto	81

6.2.2. Encontrar Buenos ítems	81
6.3 II. Dataset para evaluación.....	81
6.3.1. Sintetizado vs. Conjuntos de Datos Naturales	82
6.4 III. métricas.	82
6.4.1. Métricas de la exactitud de la predicción.	82
6.4.2. Mediciones de precisión de clasificación	82
6.4.3. Métricas de exactitud en ordenamiento.....	82
6.5 Pruebas.....	84
6.5.1. I. Tareas del usuario para el sistema de recomendación.....	84
6.5.2. II. Dataset para evaluación.....	84
6.5.3. Métricas.....	85
6.6 Resultados	86
6.6.1. Ítem académico	86
6.6.2. Ítem objeto de conocimiento.....	87
6.6.3. Ítem instalación	88
6.6.4. Ítem infraestructura	89
6.6.5. Ítem curso.....	90
6.6.6. Ítem servicio	91
6.6.7. Ítem certificación e ítem proyecto	92
6.7 Pruebas realizadas a la aplicación web.	93
6.7.1. Personalización de las plantillas con el nombre, género y grado académico de los usuarios	93
6.7.2. Información utilizada por el servicio de recomendación	93
6.7.3. Recuperación de información de los campos de texto variables para cada tipo de ítem.	94
6.8 Pruebas realizadas al funcionamiento de las valoraciones de los ítems.	100
CAPÍTULO 7	103
7.1 Conclusiones	104
7.2 Aportaciones	104
7.3 Trabajos futuros	105
REFERENCIAS.....	106

Índice de Figuras

Figura 1.1 Guiado del sistema T-Guía, primera versión	3
Figura 1.2 Modelo multidimensional de la base de hechos.....	5
Figura 1.3 Arquitectura del SRSSC T-Guía, segunda versión.....	6
Figura 1.4 Arquitectura del SRSSC T-Guía, tercera versión.....	7
Figura 3.1 Grafo ponderado.....	31
Figura 4.1 Cuerpo de texto del usuario profesor.....	42
Figura 4.2 Cuerpo de texto del ítem Objeto de conocimiento.....	43
Figura 4.3 Proceso de ponderación del algoritmo de filtrado basado en contenido.....	45
Figura 4.4 Ejemplo de etiquetado con <i>freeling</i>	45
Figura 4.5 Plantilla de explicación.....	60
Figura 4.6 Explicación mapa mental.....	61
Figura 5.1 Arquitectura del servicio de recomendación.....	63
Figura 5.2 Modelo entidad relación del estatus de los ítems.....	66
Figura 5.3 Modelo entidad relación del estatus de los usuarios.....	67
Figura 5.4 Modelo entidad relación de estatus de las recomendaciones.....	68
Figura 5.5 Ejemplo de explicación de texto.....	68
Figura 5.6 Ejemplo de recomendación y explicación.....	69
Figura 5.7 Modelo entidad relación de valoraciones de usuario.....	69
Figura 5.8 Vista de los módulos del servicio de recomendación.....	70
Figura 5.9 Vista funcional de los elementos del Servicio de Recomendación.....	72
Figura 5.10 Trigger para proyecto de investigación.....	73
Figura 5.11 Módulo de poblado.....	74
Figura 5.12 Módulos del motor de recomendación.....	75
Figura 5.13 Módulo de recomendación.....	76
Figura 5.14 Módulo de explicaciones.....	76
Figura 5.15 Recomendación ítem académico.....	77
Figura 5.16 Módulo de interfaz de usuario.....	77
Figura 5.17 Recomendaciones agrupadas por tipo de ítem.....	78
Figura 5.18 Módulo de valoración.....	78
Figura 5.19 Valoración de recomendación.....	79
Figura 5.20 Explicación en mapa mental.....	79
Figura 6.1 Ejemplo de curva ROC, los valores p que se muestran en la curva representan diferentes puntos de corte de predicción.....	84
Figura 6.2 Sensibilidad.....	85
Figura 6.3 Especificidad.....	86
Figura 6.4 Curva ROC ítem académico.....	87
Figura 6.5 Curva ROC ítem objeto de conocimiento.....	88
Figura 6.6 Curva ROC ítem instalación.....	89
Figura 6.7 Curva ROC ítem infraestructura.....	90
Figura 6.8 Curva ROC ítem curso.....	91

Figura 6.9 Curva ROC ítem servicio.....	92
Figura 6.10 Explicación utilizando grado y género.....	93
Figura 6.11 Explicación donde usuario no tiene grado académico.....	93
Figura 6.12 Vector de palabras claves.....	94
Figura 6.13 Ítem académico.....	95
Figura 6.14 Ítem objeto de conocimiento.....	95
Figura 6.15 Visualización de archivo.....	96
Figura 6.16 Ítem instalación.....	97
Figura 6.17 Ítem infraestructura.....	97
Figura 6.18 Certificación.....	98
Figura 6.19 Ítem curso.....	99
Figura 6.20 Ítem servicio.....	99
Figura 6.21 Ítem proyecto.....	100
Figura 6.22 Sistema solicita a usuario valoración de ítem.....	101
Figura 6.23 Ítem valorado por el usuario.....	102
Figura 6.24 Actualización del histograma de ratings.....	102

Lista de tablas

Tabla 3.1 Factores de apelación.....	26
Tabla 3.2 Tabla comparativa.....	37
Tabla 3.3 Tabla comparativa algoritmos de recomendación.....	38
Tabla 4.1 Descriptores del perfil del usuario académico.....	41
Tabla 4.2 Descriptores del perfil del usuario Estudiante.....	42
Tabla 4.3 Descriptores del perfil del usuario Egresado.....	42
Tabla 4.4 Descriptores del perfil del usuario Empresa.....	42
Tabla 4.5 Descriptores del perfil del ítem Objeto de conocimiento.....	42
Tabla 4.6 Descriptores del perfil del ítem profesor.....	43
Tabla 4.7 Descriptores del perfil del ítem Infraestructura.....	43
Tabla 4.8 Descriptores del perfil del ítem Instalación.....	43
Tabla 4.9 Descriptores del perfil del ítem Servicio.....	44
Tabla 4.10 Descriptores del perfil del ítem Proyecto.....	44
Tabla 4.11 Descriptores del perfil del ítem Certificación.....	44
Tabla 4.12 Descriptores del perfil del Curso.....	44
Tabla 4.13 Matriz de valoraciones usuario-ítem.....	47
Tabla 4.14 Objetivos de las explicaciones.....	51
Tabla 6.1 Resultados de ítem académico.....	86
Tabla 6.2 Resultados de ítem objeto de conocimiento.....	87
Tabla 6.3 Resultados de ítem instalación.....	88
Tabla 6.4 Resultados de ítem infraestructura.....	89
Tabla 6.5 Resultados de ítem curso.....	90
Tabla 6.6 Resultados de ítem servicio.....	91

CAPÍTULO 1 INTRODUCCIÓN

1.1 Introducción

Diariamente estamos expuestos a una gran cantidad de información que aumenta mucho más rápido que nuestra capacidad para procesarla (Ruiz Iniesta A. , 2014). Hoy en día existe una gran diversidad de información, la cual se encuentra disponible a través de internet y en los medios de comunicación masiva. Dentro de la WWW (World Wide Web) se ha producido un aumento en la información disponible para los usuarios que está compuesta desde archivos de audio, video, email, páginas web, documentos electrónicos, servicios en línea, entre otros. Esto constituye una gran cantidad de datos que están a disposición de los usuarios en repositorios heterogéneos.

Tener la información correcta en el momento adecuado representa una ventaja, esto ocurre con los usuarios que saben lo que buscan y también conocen el lugar donde encontrarla, caso contrario ocurre con los usuarios que no saben qué ni dónde buscar. La cantidad de resultados que arrojan los buscadores no siempre es la esperada o está distante de ser la información que se requiere. En cualquier ámbito, ya sea empresarial, educativo, comercial, etc., se puede tener disponible un gran volumen de información de todo tipo lo que dificulta el acceso de una manera rápida y sencilla a la información que realmente sea de nuestro interés o necesidad.

Esto representa una desventaja para los usuarios que tienen que invertir tiempo en clasificar y seleccionar la información idónea de lo que están buscando. En ocasiones los usuarios no cuentan con suficiente tiempo para dedicarse a clasificarla y desisten en continuar la búsqueda.

Las desventajas anteriormente mencionadas se pueden mitigar brindando ayuda al usuario mediante la implementación de un Sistema de Recomendación (SR). Los Sistemas de Recomendación son herramientas que brindan una solución para el manejo de la gran cantidad de información que día a día se incrementa en la WWW (Pinter, Marušić, & Radosav, 2012), ayudando al usuario en el filtrado de la información que puede ser de utilidad según sus necesidades específicas.

Desde hace ya algunos años ha aumentado el uso de los sistemas de recomendación para realizar filtrados de información de una gran cantidad de esta que el usuario no puede procesar, debido al tiempo y al esfuerzo necesario en la búsqueda de información, lo que disminuye la habilidad del usuario para identificar qué información le es útil de acuerdo a las necesidades. El área en la que mayor demanda ha tenido ha sido sobre sistemas de comercio electrónico, donde se almacenan grandes bases de datos tanto de productos como de servicios, ofreciendo soporte a los usuarios con un perfil determinado.

Algunos Sistemas de Recomendación toman como punto de partida el perfil del usuario para determinar sus gustos o preferencias, para luego aplicar el algoritmo de recomendación y generar una lista de los ítems más relevantes de acuerdo a su perfil.

Este tipo de sistemas aplican algoritmos especializados para el filtrado de información invisible para el usuario y determinando si un ítem reúne las características que cumpla con los intereses del usuario y así recomendar el ítem al usuario.

1.2 Antecedentes

En el CENIDET, dentro del grupo de Sistemas de Recomendación Conscientes del Contexto (por sus siglas en inglés *Context-Aware Recommender Systems: CARS*) se ha realizado una serie de trabajos de investigación dentro del marco del proyecto T-Guía, que consiste en el desarrollo de un Sistema de Recomendación Semánticos Conscientes del Contexto (SRSCC) diseñado para realizar recomendaciones de personas, objetos de conocimiento, lugares, eventos, actividades, recursos tecnológicos y servicios a los miembros o visitantes de una organización, con especial énfasis en Instituciones de Educación Superior (IES).

En (Arjona, 2009), se define la primera versión funcional del proyecto T-Guía. Originalmente, T-Guía consistía de un sistema de guiado automático para la localización de personas mediante tecnologías de posicionamiento heterogéneas; esto es, que con base al posicionamiento inicial del usuario y una tarea determinada (reunión, clase, préstamo de libro, etc.) se muestra mediante un mapa tipo croquis la ubicación del usuario en tiempo real y la ubicación final de donde se realiza la tarea.

Este sistema fue desarrollado con ciertas características tales como: desarrollo para teléfonos móviles con sistema operativo Android y la utilización de tecnologías de localización heterogéneas (RFID, QR Codes y GSM). En la Figura 1.1 se muestra el proceso de guiado utilizando T-Guía en su primera versión.



Figura 1.1 Guiado del sistema T-Guía, primera versión

Este trabajo de investigación consta de 3 aplicaciones: el cliente sobre un dispositivo móvil Android, el servidor que atiende las peticiones del cliente y la aplicación Web que gestiona la información de tareas, ubicaciones, usuarios y recursos.

Posteriormente, después de las pruebas de concepto logradas con la primera versión del T-Guía, se inicia una nueva etapa con el desarrollo de la segunda versión. Este nuevo sistema T-Guía se enfoca en los distintos elementos susceptibles de recomendación dentro de los distintos tipos de organizaciones. Se busca obtener un sistema de recomendación semántico sensible al contexto (SRSSC) que incorpore un repositorio de información (base de hechos) implementado en una red de ontologías capaz de soportar inferencias, un algoritmo de localización en interiores aprovechando los distintos sensores de los dispositivos móviles, mapas interactivos basados en los datos contenidos en el repositorio de información y una nueva interfaz interactiva mediante el uso de técnicas de realidad aumentada.

En (González, 2012) se presenta una extensión del modelo de la base de hechos, añadiendo dimensiones adicionales; es decir, se añaden: modelos que describen dimensiones contextuales como localización y tiempo, modelos que describen factores genéricos de la organización como publicidad, actividad económica, infraestructura, objetos de conocimiento y comercio, modelos que complementan la descripción de las competencias organizacionales como la ocupación, educación y habilidad de una persona, y modelos que describen las características de los dispositivos e interfaces utilizadas para la explotación del sistema. Esto, a grandes rasgos, definido como modelo multidimensional, tal y como se muestra en la Figura 1.2.

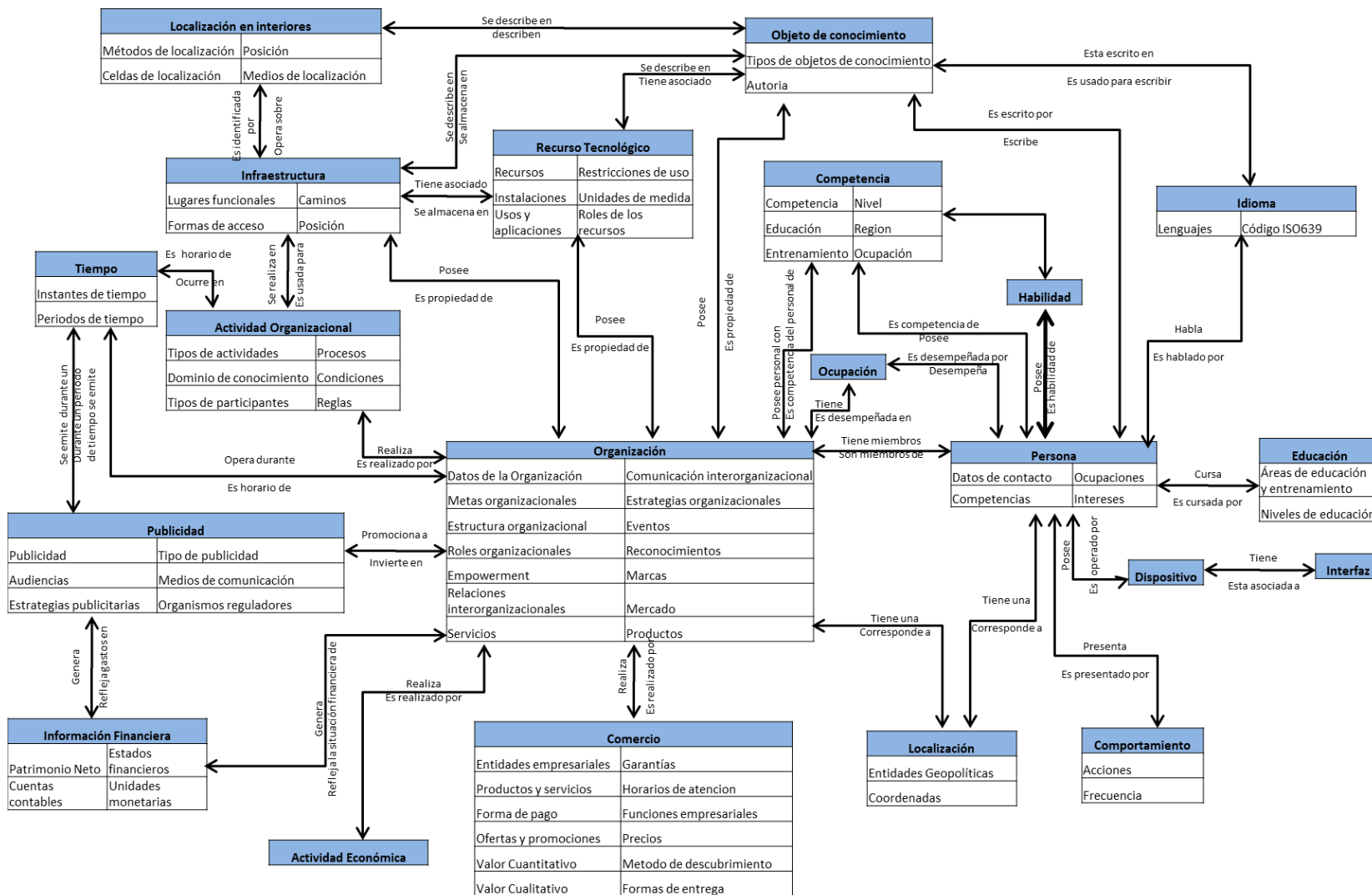


Figura 1.2 Modelo multidimensional de la base de hechos.

En (González, 2012) se presenta una nueva arquitectura modular para el sistema T-Guía, y adquiere el nombre de Sistema de Recomendación Semántico Sensible al Contexto (SRSSC) Organizacional T-Guía (Figura 1.3).

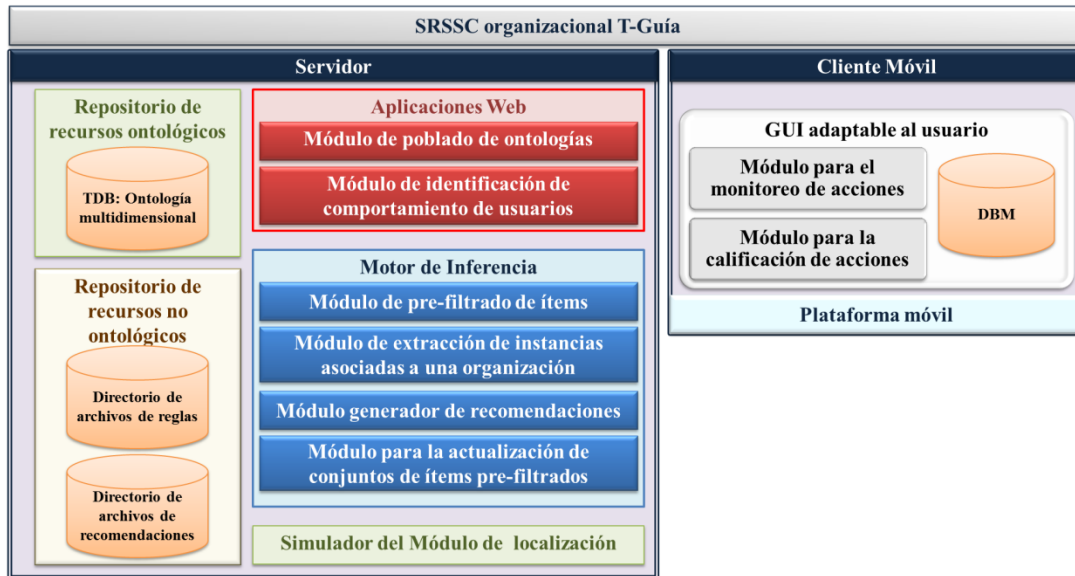


Figura 1.3 Arquitectura del SRSSC T-Guía, segunda versión.

Posteriormente, se desarrollaron varios módulos que se integraron al sistema T-Guía, tales como el desarrollo de un algoritmo de localización en interiores utilizando RFID y redes inalámbricas Wifi y Bluetooth; la creación de mapas interactivos contextuales SVG o actualmente el desarrollo de una interfaz interactiva utilizando realidad aumentada.

Todos estos módulos forman parte del sistema T-Guía, pero no tienen un impacto directo en el desarrollo de este proyecto de tesis. En el año 2012, se define una nueva arquitectura del SRSSC T-Guía (Figura 1.4).

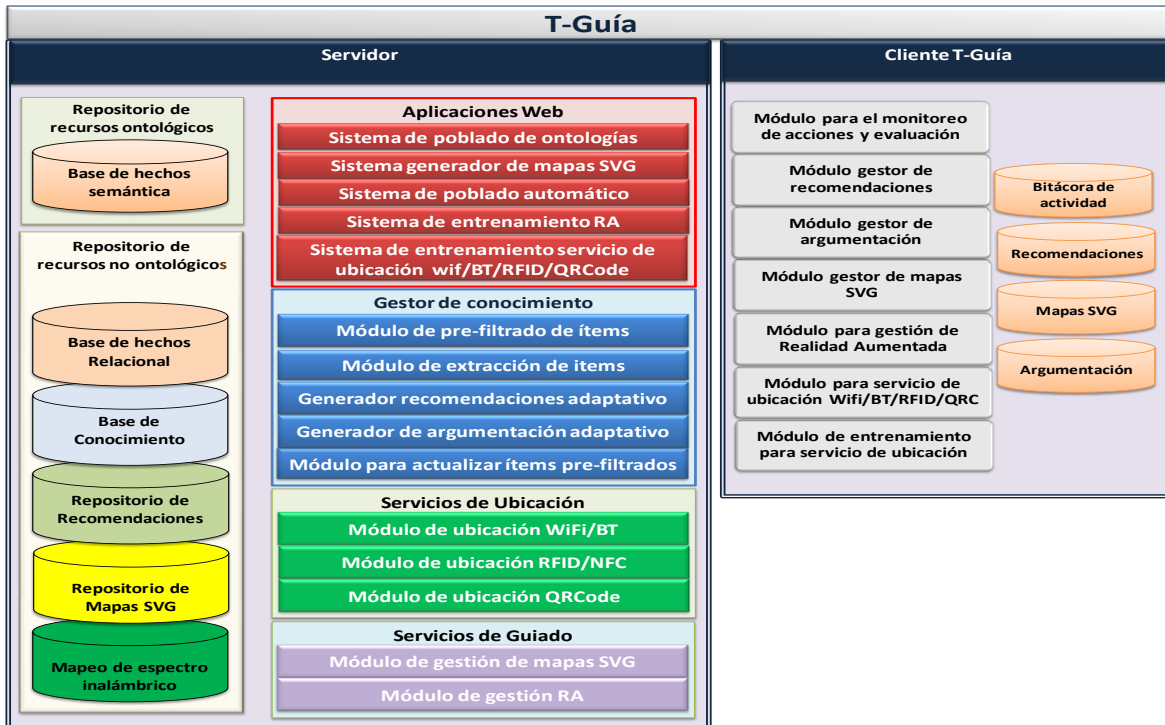


Figura 1.4 Arquitectura del SRSSC T-Guía, tercera versión.

En (Alejandres H. , 2012) se desarrolló un modelo de evaluación de sistemas de recomendación semánticos sensibles al contexto mediante la utilización de métricas que permitan obtener una valoración objetiva de los aspectos subjetivos de la experiencia del usuario, con base en las características de usabilidad de eficiencia, confianza y satisfacción, analizando escenarios con esquemas de argumentación e interfaces multimodales interactivas.

En el trabajo de investigación (Rodríguez, 2014) se desarrolló una metodología para el diseño de interfaces multimodales para SRSC sobre dispositivos móviles multisensoriales, para presentar la información de recomendaciones y explicaciones de ítems heterogéneos de acuerdo a las características del usuario. Tomando en cuenta patrones de diseño, métricas de usabilidad y experiencia del usuario.

En el trabajo de investigación de (Arana, 2014) se desarrolló una metodología de evaluación basada en las mejores técnicas existentes, adaptándolas para el caso de estudio “find it!”, estas técnicas serán analizadas y adaptadas según las necesidades propias del estudio, determinando las mayormente funcionales y con resultados exitosos, cabe mencionar que existe la posibilidad del desarrollo de nuevas técnicas que permitan la obtención de resultados más precisos para fines de la investigación.

En el trabajo de investigación (Gálan, 2014) se desarrolló una metodología adaptable y extensible para la generación de explicaciones textuales mediante plantillas, en los sistemas de recomendación, aplicada a sistemas de recomendación sensibles al contexto”.

1.3 Descripción del problema

Los sistemas de recomendación (SR) clásicos presentan algunos problemas en relación a las técnicas de recomendación que implementan, ya que se basan únicamente en usuarios e ítems. Entre estos problemas se encuentra que un SR se enfoca a un determinado tipo de ítem (música, película, noticias, etc.), además cuentan con perfiles de usuarios basados en el mismo dominio al que pertenece el ítem, y estos SR sólo usan un algoritmo de recomendación para inferir la lista de ítems que van a recomendar al usuario.

Aplicar SR en dominios organizacionales presenta cambios con respecto a los SR clásicos: cambia la premisa de “compartir recomendaciones” a “compartir conocimiento”. Un SR aplicado en una organización es capaz de explotar de manera eficiente su capital intelectual y otros ítems que no se consideran en los modelos clásicos. Las Instituciones de Educación Superior (IES) y los Centros de Investigación (CI) cuentan con ítems heterogéneos y varios tipos de perfil de usuario, que permiten poblar un repositorio organizacional de ítems y usuarios heterogéneos. La caracterización de estas entidades se centran básicamente en el conjunto de competencias propias del capital intelectual (Podgorelec, Pavlič, & Heričko, 2006) (Gronau & Fröming, 2007) (Eiselta & Marianov, 2006), sin considerar otros elementos asociados al capital intelectual de IES y de CI, como lo son: las tesis, las publicaciones, los manuales técnicos, los reportes, el material de cursos académicos, su portal Web, su infraestructura tecnológica, sus capacidades tecnológicas y sus servicios.

La problemática que se abordó en este proyecto de tesis fue el diseño de un modelo sobre el cual se aplicaron algoritmos de recomendación altamente especializados para cada clase de ítem, en combinación con el tipo de usuario. El escenario que se planteó modelar comprende aspectos organizacionales: capital humano, capital estructural y capital relacional (De Castro, Navas-López, López-Sáenz, & Delgado-Verde, 2010). El capital humano lo conforma el conocimiento incorporado a las personas incluyendo experiencia, capacidades tecnológicas y habilidades. El capital estructural comprende todo el conocimiento generado en una organización. Por último, el capital relacional se constituye de las relaciones de la organización con los distintos actores del entorno. La problemática asociada a estos escenarios organizacionales implica una alta heterogeneidad en los tipos de ítems, lo cual implica el diseño de SR adaptativos que descubran y recomienden ítems de diferente clase.

Los SR clásicos presentan problemas en sus tasas de fracaso que siguen siendo altas para ciertos dominios o individuos (Herlocker, 1999). Mientras que un usuario puede estar dispuesto a arriesgarse a comprar un CD de música basado en la recomendación de un SR, es probable que no se arriesgue a la elección de un lugar para las vacaciones de luna de miel, basado en una recomendación de este tipo. La mayoría de sistemas de recomendación clásicos no cuentan con una explicación que le ayude al usuario a decidir y comprender si el ítem que el sistema le recomienda es de utilidad para sus intereses, lo cual puede generar desconfianza por parte del usuario. Los SR clásicos cometen errores de vez en cuando, no importa lo bien implementado que estén, aunado a ello, los SR son cajas negras, oráculos computarizados que dan consejos, pero no pueden ser cuestionados. Un usuario no intuye cuándo confiar en una recomendación y cuándo dudar de una recomendación. Estos problemas pueden causar una baja aceptación de los SR como ayuda de apoyo a la decisión.

1.4 Objetivos

Objetivo general.

Diseñar e implementar un servicio de recomendación para la Plataforma de las Capacidades Científicas y Tecnológicas (PCCT), que utilice técnicas basadas en contenido, para recomendar ítems heterogéneos a usuarios de la PCCT mediante la aplicación de plantillas que incluyan explicaciones que aumenten la efectividad del sistema de recomendación.

Objetivos específicos:

- Evaluar los algoritmos de recomendación sobre el repositorio de ítems heterogéneos.
- Desarrollar un módulo que implemente los algoritmos de recomendación para inferir las recomendaciones.
- Definir la estructura de las explicaciones para cada tipo de ítem.
- Personalizar las explicaciones para cada tipo de usuario.
- Implementar un módulo que permitan la extracción de información de explicación para cada ítem contenida en la lista de recomendación.
- Desarrollar un servicio web de algoritmos de recomendación que explote un repositorio de ítems heterogéneos.
- Implementar un módulo que permitan la valoración de las recomendaciones.

1.5 Justificación

Es necesario que un SR organizacional cuente con un conjunto de algoritmos de recomendación que permitan facilitar la toma de decisiones y además, cuente con un mecanismo que le permita seleccionar la técnica de recomendación más adecuada, tomando en cuenta cada una de las clases de ítems, y además implemente las explicaciones en los ítems recomendados.

Actualmente, el proceso de recomendación en la plataforma de las capacidades Científicas y Tecnológicas sólo se emplea un tipo de algoritmo de recomendación. Esto ocasiona que no se aprovechen de la mejor forma todos los ítems. Por ello se optó por implementar una serie de algoritmos de recomendación sobre un repositorio de ítems heterogéneos; es decir, para cada una de las clases de ítems del repositorio, tomando el tipo de usuario hacia el que van dirigidas las recomendaciones, se determinó qué algoritmos hacen el proceso de selección de ítems más exhaustivo para ofrecer al usuario un mayor nivel de confianza en las recomendaciones.

Para lograr este objetivo, es necesario definir algún tipo de arquitectura que permita implementar los algoritmos de recomendación y crear la interacción entre ellos. Para este fin se desarrollaron los servicios de recomendación para la plataforma de las capacidades Científicas y Tecnológicas del TecNM.

El considerar implementar varios algoritmos de recomendación en un SR trae los siguientes beneficios:

- Definir un mayor grado de confianza hacia el usuario con base a una mejor selección de ítems a ser recomendados.

- Poder ofrecer recomendaciones con base a cualquier tipo de usuario con información del mismo.
- Hacer un mejor filtrado de ítems heterogéneos para cada una de las clases tomando como referencia el algoritmo de recomendación que mejor se adecue, tomándolo del repositorio de algoritmos de recomendación.

1.6 Estructura del documento

El presente documento se encuentra organizado en siete capítulos, los cuales describen el trabajo de investigación en sus diversas etapas como se indica a continuación:

En el capítulo II: Fundamento Teórico, se abordan los temas relevantes para la presente investigación.

En el capítulo III: Estado del Arte, se presenta la información obtenida de la investigación sobre los enfoques actuales en el desarrollo de Sistemas de Recomendación.

En el capítulo IV: La Propuesta de solución, se define detalladamente la metodología para el diseño del servicio de recomendación y la creación de las plantillas de explicación.

En el capítulo V: Arquitectura para el servicio de recomendación, en este capítulo se describe la arquitectura que permite al servicio de recomendación realizar las tareas involucradas con su funcionamiento.

En el capítulo VI: Pruebas y Resultados, En este capítulo se detalla la etapa de pruebas al servicio de recomendación, mostrando los resultados obtenidos. Además se describe el comportamiento del servicio para el conjunto de ítems heterogéneos.

Finalmente en el capítulo VII, se presentan las conclusiones obtenidas con la realización de este trabajo, además de sugerir posibles trabajos futuros para ampliar esta investigación.

CAPÍTULO 2

MARCO TEÓRICO

Esta tesis se enfoca en los Sistemas de Recomendación (SR). Por este motivo en esta sección, se detallan los temas y conceptos específicos relacionados con áreas relevantes para este trabajo.

Se abordan los conceptos de SR, las técnicas utilizadas en el proceso de recomendación; las explicaciones en los sistemas de recomendación y el análisis de los mapas mentales y conceptuales.

2.1 Sistemas de recomendación

Los sistemas de recomendación (SR) son herramientas de software que proveen un mecanismo útil para sugerir *ítems* que sean de interés para el usuario (Resnik & Varian, 1997) (Burke, Hybrid web recommender systems, 2007) (Mahmood & Ricci, 2009). Estos sistemas han tenido gran auge en el comercio electrónico, su principal objetivo es ofrecer sugerencias acerca de otros productos de interés para el usuario. Uno de los SR para comercio electrónico más conocidos es el utilizado por Amazon.com (Linden, Smith, & York, 2003), el cual muestra al usuario recomendaciones de productos de interés con base en valoraciones de la comunidad de usuarios y los productos adquiridos con mayor frecuencia.

Estas recomendaciones se relacionan con diversos procesos para la toma de decisiones Y dan respuesta a las siguientes preguntas: ¿qué ítems comprar?, ¿qué música escuchar? o ¿qué noticias leer? Generalmente, un SR se enfoca en un determinado tipo de *ítem*¹ (música, películas, noticias) y, de acuerdo con el diseño del mismo; la interfaz de usuario y la técnica de recomendación utilizada para la emisión de recomendaciones se personalizan para ofrecer recomendaciones útiles y efectivas para cada tipo de ítem.

Las recomendaciones se ofrecen normalmente como una lista ordenada de ítems, también conocida como *ranking*. Cada *ítem* dentro de la lista obtiene una valoración o *rating* de acuerdo con la técnica utilizada por el recomendador. La asignación del rating es un proceso en el cual el SR predice cuáles son los elementos más adecuados con base en las preferencias y restricciones del usuario. Con este fin, el SR recolecta las preferencias de los usuarios de manera *explícita* o *implícita*. La información *explícita* es aquella que el usuario proporciona de forma directa como votar o contestar una pregunta; la información *implícita* es la que se obtiene del usuario de forma indirecta a través de la interacción con los ítems (Ferran, y otros, 2005). Por ejemplo, cuando el usuario visita una página, el SR obtiene información sobre qué ítems o enlaces ha visitado y cuáles no, durante cuánto tiempo, además de comparar esa información con el resto de la obtenida de diferentes usuarios. Por ejemplo, un SR considera el tiempo que un usuario permanece en la descripción de un artículo como una señal implícita de preferencia o rechazo hacia dicho artículo.

Con el desarrollo de los sitios Web para el comercio electrónico, surgió la necesidad de proveer recomendaciones derivadas del filtrado de diversas alternativas disponibles. Antes los usuarios enfrentaban dificultades para encontrar opciones adecuadas entre la diversidad de ítems que los sitios ofrecían. El crecimiento y variedad de información disponible en la

¹ *Ítem* es el término general utilizado para nombrar los elementos que el sistema recomienda al usuario.

Web y la rápida introducción de nuevos servicios rebasaron a los usuarios, llevándolos a tomar decisiones incorrectas (Ricci, Rokach, & Shapira, 2011). Fue cuando los SR se presentaron como alternativa a esta problemática, orientando a los usuarios hacia ítems nuevos aún no experimentados y que fueran relevantes para el usuario en un contexto específico. Una vez que el SR recibe la petición del usuario, éste genera recomendaciones usando distintos tipos de información acerca de los usuarios. Los ítems disponibles y las transacciones previas usuario-sistema almacenadas. Posteriormente, el usuario puede navegar entre las recomendaciones, aceptarlas o no y proporcionar una retroalimentación explícita o implícita. Todas estas acciones del usuario junto a las retroalimentaciones, son almacenadas por el recomendador para usarlas posteriormente en la emisión de nuevas recomendaciones.

El gran auge que han tenido los SR se debe a su utilidad en los sistemas de comercio electrónico, porque su funcionalidad se refleja en el incremento del beneficio económico. Hay una gran diversidad de dominios de aplicación para los SR en el comercio electrónico: farmacéuticos en Drugstore.com (Schafer, Konstan, & Ryedl, 2001), programas de televisión con TiVo (Ali & Van Stam, 2004), películas en Netflix (Bennet & Lanning, 2007) y canciones mediante Genius (Apple, 2016) para iTunes.

Un SR se compone de cinco elementos

- **Base de conocimientos** (*Knowledge base*): conocimiento modelado extraído del diálogo con un experto.
- **Base de hechos**: información explícita sobre una entidad o problema que se ha descubierto durante el análisis.
- **Motor de inferencia**: modela el proceso de razonamiento humano.
- **Módulos de justificación** (argumentación): explica el razonamiento utilizado por el sistema para llegar a una determinada conclusión.
- **Interfaz de usuario**: interacción entre el SR y el usuario.

El desarrollo de los SR comenzó a partir de una simple observación: los usuarios casi siempre confían en las recomendaciones proporcionadas por otros usuarios (Mahmood & Ricci, 2009). Por ejemplo, es común confiar en la recomendación de un amigo cuando se selecciona un libro o se elige una película en el cine o los reclutadores confían en las cartas de recomendación en el proceso de contratación. Tratando de imitar este comportamiento, los primeros sistemas de recomendación aplicaron algoritmos para soportar recomendaciones producidas por una comunidad de usuarios y entregar al usuario activo.

2.2 Clasificaciones de sistemas de recomendación

Existen dos clasificaciones de sistemas de recomendación la primera de ellas define las siguientes categorías (González, 2012):

- **Basados en contenido**: al usuario se le recomiendan ítems basados en lo que le ha gustado previamente.
- **Colaborativos**: al usuario se le recomiendan ítems que han sido del agrado de usuarios con preferencias similares.

- **Basados en conocimiento:** el conocimiento que proporciona el usuario sobre sus necesidades y el conocimiento que tiene el sistema sobre las entidades para realizar recomendaciones de mejores productos para cubrir las necesidades de los usuarios.
- **Híbridos:** estos sistemas combinan el enfoque basado en contenidos con el enfoque colaborativo.

La segunda clasificación agrupa a los sistemas de recomendación basados en reglas de filtrado colaborativo; y de personalización basada en contenidos (González, 2012).

2.3 Técnicas de filtrado colaborativo

Los algoritmos de filtrado colaborativo proporcionan recomendaciones o predicciones de ítems basándose en la opinión de usuarios con características similares, siendo que esta opinión puede ser obtenida de manera explícita o implícita. En otras palabras, la meta de los algoritmos de filtrado colaborativo es sugerir nuevos ítems o predecir la utilidad de cierto ítem para un usuario particular basándose en los gustos y opiniones definidos previamente por usuarios similares (Sarwar, 2001).

2.3.1. Técnicas de filtrado basado en contenidos

Los SR basados en contenidos utilizan algoritmos que analizan la descripción de los *ítems* con el objetivo de identificar aquellos que resulten de interés para el usuario (Balabanović, 1997). En este tipo de algoritmos se cuenta con un perfil de usuario que describe sus preferencias con base en las características de los objetos que ha calificado, lo que en la literatura se conoce como correlación ítem-a-ítem (Burke, 2002).

Existen diferentes formas de representar los ítems. Una técnica para crear estructuras que representen el contenido de los ítems es el *stemming*, cuya meta es determinar un término que refleje el significado común existente entre diferentes palabras (Balabanović, 1997). El valor de una palabra (variable) asociada a un término es un número real que representa su relevancia.

2.3.2. Técnicas de filtrado híbrido

Los sistemas de recomendación híbridos combinan dos o más técnicas de recomendación para tener un mejor desempeño. A continuación se describen algunas de las combinaciones más comunes (Burke, 2002).

- **Ponderación o *Weighted*:** en este enfoque se calcula un puntaje para un ítem a partir de todas las técnicas de recomendación disponibles en el sistema.
- **Conmutación o *Switching*:** en este enfoque el sistema utiliza algún criterio para cambiar entre distintas técnicas de recomendación dependiendo de la situación presente al momento de efectuar la recomendación.
- **Mixta:** en este enfoque se utiliza más de una técnica de recomendación que son ejecutadas simultáneamente, lo cual es útil cuando se necesita generar múltiples recomendaciones al mismo tiempo.
- **Función de combinación:** en este enfoque se realiza una mezcla entre algoritmos colaborativos y basados en contenido. Al tratar la información colaborativa simplemente como datos adicionales y usar técnicas basadas en contenidos sobre el

conjunto de datos extendido. De este modo, se pueden considerar datos colaborativos sin tener que depender exclusivamente de ellos, lo que reduce la sensibilidad del sistema a la cantidad de usuarios que han calificado un ítem. Otra particularidad es que permite una apreciación de la similitud entre los ítems que no sería visible usando únicamente técnicas colaborativas.

- **Cascada:** la hibridación cascada implica un proceso por etapas, primero se emplea una técnica para producir un conjunto de ítems candidatos, a partir de éste se aplica una segunda técnica que refina la recomendación de ítems. Su eficiencia radica en que la segunda técnica se aplica sólo a los elementos que requieren un segundo filtrado, mientras que en otros enfoques cada técnica se aplica a todos los elementos.
- **Función de incremento:** en este enfoque, se emplea una técnica para producir la clasificación de un ítem incorporando esta información en el proceso de la siguiente técnica de recomendación.
- **Meta-nivel:** en este enfoque se usa el modelo generado por una técnica como entrada de otra. El meta-nivel híbrido se enfoca exclusivamente en la recomendación conocida como "colaboración vía contenido" y tiene la ventaja de que el modelo es una versión comprimida de los intereses del usuario y un mecanismo colaborativo que opera fácilmente sobre su representación de densidad de información.

2.4 Sistemas de Recomendación Conscientes del Contexto (SRCC)

Los SR clásicos presentan algunos problemas en relación a las técnicas que implementan ya que se basan únicamente en usuarios e ítems. Entre estos problemas se encuentran las diferencias significativas en la distribución usuario-ítem, comportamiento inusual de los usuarios y la disponibilidad de sus preferencias.

Una manera de subsanar estos problemas consiste en hacer uso de información del contexto que rodea al usuario durante el proceso de emisión de recomendaciones. En un SR, el contexto del usuario puede ser la localización del usuario, identidad de las personas, objetos a su alrededor, fecha, estación del año, temperatura, información que puede caracterizar la interacción entre un usuario y una aplicación o bien su nivel de experiencia, dispositivos utilizados y sus características.

La inclusión del contexto permite mejorar la calidad de las recomendaciones entregadas al usuario porque se pondera el contexto específico en el cual se encuentra y es posible inferir información implícita no contenida en un perfil de usuario.

2.5 Sistemas de Recomendación en ambientes organizacionales

En el ámbito organizacional, la implementación de los SR no ha tenido el mismo auge que en el comercio electrónico porque las organizaciones han centrado su atención en la recuperación de información. En este sentido, los SR representan una alternativa proactiva a los sistemas de recuperación de información y son utilizados para la explotación de la información corporativa. Esto incluye la obtención de información implícita como

habilidades, competencias y redes de personas así como información de infraestructura, eventos, etc.

Las recomendaciones sobre este tipo de información pueden ser de tres tipos: 1) sugerencias, presentación al usuario de ítems que pueden serle de utilidad; 2) predicciones, suposiciones de lo que el usuario puede hacer con el ítem; 3) evaluaciones, presentación de las opiniones de los usuarios sobre un ítem. Estas recomendaciones se realizan sobre el capital intelectual de la organización, el cual comprende tres vertientes principales: capital humano, capital estructural y capital relacional (De Castro, Navas-López, López-Sáenz, & Delgado-Verde, 2010). El capital humano comprende el conocimiento incorporado a las personas incluyendo experiencia, capacidades y habilidades. El capital estructural comprende el conocimiento de la organización incluyendo rutinas, estrategias, manuales, procedimientos, entre otros. Por último, el capital relacional se constituye de las relaciones de la empresa con distintos actores del entorno (clientes, proveedores y competidores).

Se aprecia que el enfoque de los SR en este dominio presenta cambios con respecto a los SR clásicos. Se parte de la idea de compartir recomendaciones a compartir conocimiento y de construir comunidades a soportar comunidades (Glance, Arregui, & Dardenne, 1999). Haciendo uso de los SR, una organización es capaz de explotar de manera eficiente su capital intelectual y, de esta forma, obtener un beneficio.

Entre los elementos que pueden ser de interés para un usuario y por ende, susceptibles de ser recomendados en un entorno organizacional se encuentran personas expertas en determinada área, documentos sobre temas específicos, eventos dentro y fuera de la organización, servicios propios y organizaciones externas. Bajo la premisa de la división del capital intelectual de la organización, se agrupan estos elementos bajo tres rubros:

- **Recomendación de personas expertas:** la experiencia de una persona en determinada área es un recurso muy valioso en un entorno organizacional. Esta experiencia es difícil de representar de forma explícita y para su obtención es necesario realizar análisis de los repositorios que contiene este tipo de información. En esta área, los SR se utilizan para recomendar personas expertas con conocimiento en resolución de problemas específicos dentro de la organización.
- **Recomendación de documentos:** dentro de las organizaciones es necesario emplear esquemas de colaboración con base en grupos de trabajo. Estos grupos producen una gran cantidad de documentos que contienen conocimiento implícito que resulta valioso como parte del capital intelectual. Existen trabajos enfocados en administración del conocimiento (*knowledge management*); es decir, al almacenamiento y extracción de información de este tipo de documentos. Sin embargo carecen de la proactividad de los SR. Debido a esto, se han presentado trabajos que se enfocan en la recuperación de este conocimiento mediante la implementación de SR.
- **Recomendación de recursos:** se centran en recomendaciones de eventos, servicios y organizaciones externas que fungen como socios potenciales. Una característica de este tipo de trabajos es la utilización de información contextual de los usuarios para mejorar el proceso de emisión de recomendaciones.

2.6 Explicación y Argumentación

De acuerdo a la Real Academia de la Lengua Española una explicación se define como “Una declaración o exposición de cualquier materia, doctrina o texto con palabras claras o ejemplos, para que se haga más perceptible.”

Desde el punto de vista de la ciencia moderna, la formulación de leyes y teorías científicas es lo más adecuado a la hora de prever eventos naturales y controlarlos. De esta manera, la ciencia dispone de métodos sistemáticos, muy sofisticados, para describir los diversos aspectos de los fenómenos naturales para poder explicarlos (Chirinos Bossio, 2007).

Una manera de comprender estos métodos utilizados por la ciencia, es concebirlos como modos de dar respuestas a cierta clase de preguntas sobre los fenómenos. En la incesante búsqueda de respuestas, se trata de responder a la pregunta del ¿por qué? esos fenómenos son lo que son y ¿por qué? se dan o pueden darse ciertos sucesos.

El término explicación desde la perspectiva científica clásica está restringido, a la explicación del “porqué” de los hechos; considerándose sólo como explicación científica aquella que busca su marco y conceptos de referencia únicamente en las relaciones causales, es decir, “¿por qué?, dada ciertas condiciones antecedentes, se podría haber esperado que ocurriese el acontecimiento a explicar” Lambert y Britan 1975.

Por otra parte argumentación proviene del latín “*argumentum*” que significa prueba o razón para justificar algo como verdad o como acción razonable. La teoría de la argumentación parte de los trabajos publicados por Oswald Ducrot y Jean-Claude Anscombe en los años setenta. En la formulación de Ducrot, argumentar tiene el valor de “*hacer admitir*” un razonamiento (sea lógico o no) para poder llegar a una conclusión que sea aceptable por el destinatario de la enunciación.

(Pons, 2003) Afirma que *un aspecto fundamental en la teoría de la argumentación es la defensa de la idea de que es el propio discurso el que propicia la interpretación argumentativa. No son los hechos extralingüísticos los que imponen una lectura argumentativa en una dirección determinada, sino la propia lengua. Es la orientación argumentativa interna de los enunciados (y no el contenido informativo) la que favorece que se espere la continuación de un discurso en un sentido determinado en virtud de una serie de inferencias.*

En resumen, una argumentación y una explicación pueden tener formas lingüísticas similares, así como términos comunes de razones, motivos o causas, pero su distinción se puede apreciar en que la argumentación posee la existencia de una cuestión debatida y de unos compromisos con posiciones o proposiciones enfrentadas al respecto, pero en la explicación existe la ausencia o la irrelevancia de estos aspectos dialécticos e intencionales.

2.6.1. Explicaciones en los sistemas de recomendación

Las explicaciones tienen su origen en el área de sistemas expertos. Estos sistemas utilizan un módulo de explicación mediante el cual proporcionan una explicación al usuario de porqué está haciendo una pregunta y cómo ha llegado a una conclusión. Este módulo

proporciona beneficios tanto al diseñador del sistema como al usuario. El diseñador puede usarlo para detectar errores y el usuario se beneficia de la transparencia del sistema.

Actualmente existen varios tipos de explicaciones e implementaciones en los sistemas que modelan el mundo real. Las explicaciones en los sistemas de recomendación son información acerca de las recomendaciones, apoyan los objetivos definidos por el diseñador del sistema de recomendación (Friedrich & Zanker, 2011). Cuando un usuario recibe una explicación, puede aceptar una recomendación fácilmente porque el sistema proporciona la transparencia de su recomendación (Papadimitriou, Symeonidis, & Manolopoulos, 2011); además de proporcionar diversos objetivos como se plantean en (Tintarev & Masthoff, 2007).

Los estilos explicativos proporcionan un nivel de transparencia a las técnicas de recomendación (Gálan, 2014). Asimismo, exponen los detalles del proceso de razonamiento y la información utilizada para la emisión de una recomendación. Las explicaciones tienen distintos objetivos como: mejorar la confianza del usuario; incrementar la satisfacción de los resultados obtenidos; hacer más fácil y rápida la búsqueda de ítems relevantes e influir en la decisión del usuario para seleccionar determinado ítem. Frente a las recomendaciones emitidas por el SR, si éste tiene dudas sobre el origen de una de ellas se preguntará ¿por qué recibí esta recomendación?, para responder esta pregunta el SR debe ser capaz de explicar *porqué* ha realizado tal sugerencia. De esta forma, el usuario puede analizar la lógica detrás de la recomendación y decidir si existe suficiente evidencia para tomarla como válida. Para generar explicaciones convincentes, se considerarán varios atributos que contribuyen para la aceptación de una recomendación y su utilización en la toma de decisiones. Entre estos atributos se destacan: transparencia, confianza, efectividad, persuasión, eficiencia, satisfacción, etc.

2.6.2. Mapas conceptuales

Los mapas conceptuales, fueron desarrollados por Joseph D. Novak en los años 70. Se usan como un medio para la descripción y comunicación de conceptos dentro de la teoría de asimilación, una teoría del aprendizaje con una enorme influencia en la educación. La teoría está basada en un modelo centrado en los procesos cognitivos humanos. Es decir describe cómo el estudiante construye conceptos y cómo se organizan en su estructura cognitiva (Cañas, y otros, 2000). Los mapas conceptuales presentan de manera explícita las relaciones lógicas y psicológicas relevantes para un individuo dentro de un conjunto de conceptos para construir un aprendizaje significativo sobre un tema. Esta relación se describe por medio de palabras de enlace formando proposiciones. La forma más simple de un mapa conceptual constaría de dos conceptos unidos por una palabra de enlace para formar una proposición.

Una característica importante de los mapas conceptuales, es que tienen una jerarquía gráfica, es decir los conceptos más generales se explicitan en la parte superior y descienden por el mapa, encontrando los conceptos de jerarquía intermedia y luego los más específicos (Ciliberti, y otros, 1999).

En un mapa conceptual, también es posible expresar la jerarquía de los conceptos mediante nodos, haciendo gráficos de los diferentes niveles de inclusión, desde los más generales hasta los más específicos. Además se pueden relacionar distintas ramas jerárquicas entre sí mediante enlaces cruzados, estableciendo conexiones o nexos (Costamanga, 2001). En general para construir un mapa conceptual es necesario identificar los conceptos, ordenarlos del más general al más específico, enlazar los conceptos y opcionalmente añadir ejemplos. Las relaciones subordinadas entre conceptos pueden cambiar en diferentes segmentos, es decir, que cualquier concepto puede elevarse a una posición superior y seguir manteniendo una relación proposicional significativa con otros conceptos del mapa. Por otra parte, las redes semánticas son grafos que proporcionan una representación del conocimiento mediante objetos, propiedades y relaciones sin estructura jerárquica vertical, por lo tanto las conexiones entre nodos se representan por medio de flechas para guiar el sentido de la lectura. Los nodos de una red semántica están unidos por arcos, los cuales indican la relación que existe entre ellos, los arcos representan las relaciones conceptuales. Estas redes fueron propuestas por Quillian y Collins en 1968. Su concepción se basa en la asociación de conocimientos que realiza la mente humana y no requieren jerarquía gráfica vertical.

2.6.3. Mapas Mentales

El término mapa mental se le atribuye al psicólogo británico Tony Buzan, el cual lo define como *“Un método de análisis que permite organizar con facilidad los pensamientos y utilizar al máximo las capacidades mentales”* (Buzan, 2002), los mapas mentales poseen una estructura común compuesta por ramas que irradian de una imagen central, además del uso de colores, símbolos, dibujos y palabras que se enlazan de manera sencilla y lógica.

Buzan argumenta que los diagramas tradicionales requieren que el lector asimile la información de izquierda a derecha y de arriba abajo; mientras que lo natural para el cerebro es explorar toda la página sin un patrón lineal definido. Se basa en los hemisferios cerebrales para proponer los mapas mentales como la herramienta válida a la hora de expresar ideas de forma visual de forma que cada dato se representa como una esfera de la cual pueden irradiar miles de ramificaciones y cada ramificación representa una asociación que a su vez cuenta con una red infinita de conexiones. Los mapas mentales son una forma de representar dicho concepto llamado pensamiento radial.

Como se mencionan en (Rodríguez Piña, y otros, 2008) según el sociólogo mexicano Antonio Emmanuel Berthier, un mapa mental es una herramienta que permite la organización y representación de la información con el propósito de facilitar los procesos de aprendizaje, administración, planeación organizacional, así como el proceso de toma de decisiones.

Al igual que en el cerebro, el mapa mental procede mediante la asociación de ideas. Una vez ubicada la idea central se desprenden por asociación “ramas” hacia todas las ideas relacionadas con ella, mostrando las diferentes dimensiones o aspectos de un mismo tema. Tomando en consideración que es más fácil asimilar la información cuando se representa por medio de imágenes, como se demuestra en diferentes estudios (Perales, y otros, 2005), los mapas mentales permiten conectar, relacionar, expandir y familiarizarnos con la

información de una manera sencilla y eficaz, partiendo de una idea central de la cual se derivan distintas ramificaciones.

2.7 Evaluación de SR centrada en el usuario

Los SR se han convertido en un componente ampliamente aceptado que provee beneficios al usuario y a las organizaciones que los implementan. Para los usuarios, los beneficios incluyen mayor eficiencia para encontrar ítems necesarios, mayor confianza para la toma de decisiones y una oportunidad potencial para descubrir ítems nuevos. Para una organización, esta tecnología mejora significativamente la intención del usuario al comprar productos, regresar al sitio y recomendarlo a otros usuarios.

De esta manera, la evaluación de la percepción del usuario sobre un SR ayuda a los desarrolladores y organizaciones a entender de forma precisa si aquellos realmente experimentan y aprecian los beneficios esperados. En este sentido, la evaluación centrada en el usuario (por sus siglas en inglés, UCE) comprende tres metas principales: verificar la calidad del SR, detectar problemas de funcionamiento y ayudar al usuario en la toma de decisiones.

CAPÍTULO 3

ESTADO DEL ARTE

El contenido de este capítulo abarca el análisis de diferentes estilos y técnicas en los sistemas de recomendación con características afines o útiles para el presente trabajo de investigación.

Los trabajos son presentados basados en las técnicas de recomendación de colaboración, contenido, basados en conocimiento e híbridos.

3.1 Recomendación de artículos académicos a través de propósitos de lectura de los usuarios

En el trabajo realizado en la universidad de Pekín en China por (Jiang, 2012), consiste en un sistema de recomendación de artículos (*papers*) basado en contenido. Este sistema de recomendación toma en cuenta los propósitos de lectura del usuario. Hace recomendaciones de los documentos relacionados que resuelven el mismo problema y documentos relacionados que utilizan la misma solución.

Este sistema está compuesto por 4 módulos:

División del resumen

Este módulo divide el resumen en dos partes: descripción del problema y descripción de la solución. Por cuestiones de derecho de autor no pueden acceder a la información de texto completo de un artículo, en este caso sólo utilizan los resúmenes.

Modelo de similitud de artículos

Este módulo extrae las características del artículo. Se asigna el contenido del problema y el contenido de solución al respectivo corpus de entrenamiento. Emplea 3 modelos de similitud: TF*IDF, modelo tópico y concepto basado en modelo de tópicos.

- El modelo TF*IDF: es una medida estadística que refleja la importancia de una palabra en un documento.
- El modelo tópico está basado sobre la idea de que un documento se puede ver como una mezcla de varios temas, dentro del cual, un tema es una distribución multinomial sobre las palabras. El modelo latente de asignación de *Dirichlet* (LDA) estima la distribución de palabra-tema $P(t/z)$ y la distribución de documento-tema $P(z/d)$ a partir de un corpus sin etiqueta. Se obtiene $P(z/d)$ para cada documento y además se construye el vector de característica, Fórmula 3.1.

$$F = (P(z1|d), P(z2|d), \dots, P(zk|d)) \text{ (Fórmula 3.1)}$$

El concepto basado en modelo de tópicos: consiste en que un grupo de términos técnicos tales como *latent Dirichlet Allocation*, si se trata a las tres palabras individualmente, nunca sabrá que son en realidad es el nombre de un modelo. Une estos términos académicos en una sola palabra utilizando "_", para formar *latent_Dirichlet_Allocation*. Para definir los términos académicos, extrajeron la información de etiquetado del *dataset* de *citeulike*.

Selección de candidatos

Los artículos candidatos son seleccionados mediante el uso de las citas que contiene en él, con el objetivo de no calcular la similitud entre todos los artículos contenidos en la base de

datos. Las citas de las publicaciones académicas ofrecen relación significativa entre ellos. Los investigadores citan otros artículos similares a su trabajo.

Para un artículo i , se representa el conjunto de referencia como Ref_i , su conjunto de citas como Cit_i , y definimos el conjunto candidato de i , Fórmula 3.2.

$$CandidateSet = Ref_i \cup Cit_i \cup \left(\bigcup_{j \in Cit_i} Ref_j \right) \cup \left(\bigcup_{k \in Ref_i} Cit_k \right) \quad (\text{Fórmula 3.2})$$

Clasificación de artículos

Clasifica los artículos candidatos dado el artículo objetivo. Después de dividir los resúmenes y construir los modelos de similitud en ambas partes, problema y solución, se obtienen los vectores ponderados para cada uno. A continuación, calcula la similitud coseno entre el artículo destino y cada uno en la lista de candidatos, Fórmula 3.3.

$$Sim = \cos \theta = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (\text{Fórmula 3.3})$$

Donde A y B son vectores de características de dos artículos. Clasificándolos por similitud coseno para ambos candidatos: problema y solución. Como resultado, se obtiene una lista relacionada con un problema y la lista de solución relacionada para cada artículo.

3.2 Sistema de recomendación basado en Argumentación de confianza habilitada

TABRS (Vashisth, 2012) es un sistema de recomendación basado en agentes. TABRS combinan un sistema híbrido con la argumentación entre los agentes. La argumentación constituye una alternativa para mejorar las tecnologías de recomendación al proporcionar mecanismos inferenciales adecuados para el razonamiento cualitativo. Proporciona recomendaciones sobre una base de argumentos. Este enfoque se basa en el modelado de criterios de preferencia del usuario por medio de hechos y reglas codificadas en el formalismo de argumentación. Estos criterios de preferencia se combinan con información adicional de antecedentes (perfil de usuario), utilizados por el marco de argumentación para priorizar sugerencias potenciales, aumentando así los resultados finales previstos para el usuario.

El enfoque híbrido de TABRS es conocido como “Feature Augmentation Hybrid Technique” (técnica híbrida de aumento de características) que proporciona la salida de una técnica de filtrado basado en contenido en una entrada a una técnica de filtrado colaborativo. TABRS permite la recomendación multi-agentes autónomos con capacidades de inferencia, de modo que puede explorar y razonar acerca de los motivos profundos subyacentes detrás de las preferencias personales del usuario. De esta manera, es capaz de resolver los conflictos entre las preferencias y las recomendaciones y dar sugerencias acompañadas de argumentos.

TABRS propone un modelo de confianza con base en todos los argumentos aceptados y no aceptados generados por los agentes en un sistema de recomendación basado en la argumentación. Este sistema utiliza los agentes Intención-Deseo-Creencia, con razonamiento y capacidad de argumentación. En las interacciones, los agentes pueden discutir acerca de las demás creencias, metas y planes, utilizando la argumentación. En estas interacciones, los agentes tendrán que razonar acerca de qué tanto deben confiar en esas otras entidades, si confía en las entidades para llevar a cabo alguna tarea, o si ellos confían en las entidades para no hacer mal uso de información crucial.

Utiliza el razonamiento basado en argumento para proporcionar una perspectiva cualitativa en la toma de decisiones en la forma de un sistema de recomendación basado en argumentos. Esto se logra mediante la integración de las capacidades de inferencia para ofrecer sugerencias razonadas, modeladas en términos de argumentos a favor y en contra de una decisión particular. Utiliza el enfoque de argumentación para mejorar el desempeño de los mecanismos de confianza. Así, la confianza será calculada por un agente en forma aislada o cuando los agentes intercambian y comparten información sobre la confiabilidad de posibles objetivos.

El sistema de recomendación basado en agentes en el concepto propuesto proporciona y también mejora las recomendaciones de forma autónoma por el razonamiento. Cada agente realiza una tarea dada y diferentes agentes trabajan simultáneamente para alcanzar la tarea general. La información de confianza se utiliza en el proceso de recomendación convencional, así como la fase de la argumentación para generar recomendaciones

mejoradas. El marco propuesto se compone de tres fases: la fase de modelado, la fase recomendación y la fase de la argumentación.

En la fase de modelado

El agente de usuario se activa cuando este inicia sesión en el sistema. Este agente recolecta los requerimientos, intereses y preferencias para incluirlos en el modelo de usuario para construir su perfil. Este agente además tiene la responsabilidad de analizar y extraer los atributos de los ítems que son la entrada al sistema de recomendación basado en contenido.

En la fase de la recomendación

En esta fase el agente recomendador es el responsable de generar las recomendaciones. Los ítems recomendados se filtran utilizando el algoritmo basado en contenido, donde las recomendaciones se generan con base en los atributos de los ítems.

Este agente además evalúa el desempeño del recomendador, determina el número de recomendaciones correctas y el promedio de ítems aceptados por los usuarios. Con este resultado el agente puede actualizar los gustos y disgustos de los usuarios.

Fase de argumentación

En este artículo el autor afirma que la argumentación es: "un medio social y verbal que tratar de resolver o, al menos, lidiar con un conflicto o diferencia que haya surgido o que existe entre dos (o más) partes". La argumentación es un enfoque para razonar con conocimiento inconsistente y ayuda en la toma de decisiones. En esta fase la argumentación es utilizada para encontrar información inconsistente en el perfil del usuario que repercute en los ítems recomendados.

3.3 Asesoramiento automatizado de lectores para hacer recomendaciones de libros para lectores K-12

Este artículo realizado en la universidad de Brigham Young en Utah, EE. UU. (Pera, 2014) Es un sistema de recomendación de libros (llamado *Rabbit*) basados en las habilidades de lectura del usuario. *Rabbit* simula servicio de asesoramiento que se ofrece a los lectores en las bibliotecas escolares.

Este sistema analiza la capacidad de lectura del usuario, examina los factores atractivos (*appeal-term*) de un libro, para determinar las razones por las cuales un libro es interesante para un lector. Además Explora los elementos literarios de los libros que identifican la velocidad a la que las historias se desarrollan (ritmo), su estructura general (historia), los sentimientos que evocan estas historias en un lector (tono), temas que algunos lectores podrían encontrar como desagradables u ofensivos (temas especiales), así como cualidades de los personajes (caracterización), el lenguaje y el nivel de detalles (estilo de escritura) de las historias.

Los factores más prominentes, como se establece en la literatura incluyen: (i) la caracterización, (ii) marco, (iii) la estimulación, (iv) historia, (v) el lenguaje y el estilo de escritura, (vi) el tono, y (vii) temas especiales. Cada factor atractivo se asocia con un vocabulario que es un conjunto de palabras clave, llamados términos atractivos (*Appeal Terms*), empleados para describir el factor con el que se asocia. Los factores atractivos considerados por *rabbit* y sus respectivos términos se muestran en la Tabla 3.1.

Factores atractivos (<i>Appeal Factors</i>)	Términos atractivos (<i>Appeal Terms</i>)
Caracterización:	Crefible, distante, dramático.
Marco:	Agridulce, contemporáneo, descriptivo.
El lenguaje y el estilo de escritura:	Franca, complejo, conversacional, extravagante, poético, prosaico.
Ritmo:	Fácil, rápido, lento.
Temas especiales:	La adicción, la intimidación, la violencia.
Historia:	Orientado a la acción, carácter centrado.
Tono:	Oscuro, feliz, surrealista.

Tabla 3.1 Factores de apelación.

Descripción de términos atractivos

Los descriptores de términos atractivos solo están disponibles en bases de datos en línea que son de paga, o son definidas manualmente por expertos en el tema. Para extraer las descripciones (*appeal term*) de los libros utiliza ABET una herramienta que extrae automáticamente de los comentarios, las descripciones de los libros disponibles en los sitios web relacionados con el libro, tales como *Amazon*, *Bertram*, *Bookfinder4u*, *Bookmooch*, *Dogobooks* y *Fishpond*.

Las opiniones de los lectores describen los factores que ayudan a identificar porqué a los lectores les interesa un libro, basado en sus elementos literarios.

Es imprescindible asociar correctamente los términos de apelación y los factores de apelación, para lograr esta tarea. Han definido una serie de reglas de extracción para ABET basado en relaciones de tipo de dependencia entre pares de palabras en frases extraídas de los comentarios. Las reglas utilizadas por ABET se basan en patrones escritos identificados en reseñas de libros y capturan el vínculo semántico entre los factores de apelación y sus términos correspondientes.

Los libros son seleccionados de repositorios web de confianza como OpenLibrary.org o WorldCat.org y de sitios de marcadores de libros como BiblioNasium.com. Los libros candidatos son seleccionados tomando en cuenta el nivel de lectura del lector, el cual lo determina con la herramienta Troll, una herramienta para el análisis de regresión de los niveles de alfabetización. La selección de libros es menos de medio nivel de grado del nivel de lectura promedio del lector, utiliza el modelo espacio vectorial para determinar el grado de distribución de los temas entre el perfil del usuario y los libros candidatos. Además de determinar los temas de interés, también identifica los libros de interés para un lector mediante el análisis de la descripción del contenido. El cálculo lo determina empleando la medida de similitud coseno.

3.4 Modelo basado en características contextuales más influyentes para sistema de recomendación sensible al contexto

En este trabajo realizado en la Universidad de Manipal, en Karnataka India por (Rana, 2013) propone una metodología para incorporar la información contextual en el sistema de recomendación. Esta metodología propone un algoritmo MICF para identificar las variables contextuales más relevantes (*Most Influential Contextual-Features* por sus siglas en inglés) y generar el conjunto de características contextuales que muestran similar patrón de calificación para cada clase de entidades. La información contextual de los colaboradores (usuarios vecinos), es decir, grupo de edades, ocupación, género, etc. que han valorado los ítems debe ser analizada cuidadosamente con el fin de predecir mejor la calificación probable que un usuario podría dar a un ítem, por lo que es esencial incorporar la información contextual en el sistema de recomendación. Por ejemplo, en un sistema de canciones para una niña de 12 años de edad pueden ser muy diferentes que para un hombre de 47 años de edad.

Esta metodología extiende el enfoque de filtrado colaborativo para incorporar la información contextual de los usuarios, a fin de diseñar un nuevo y mejorado algoritmo. El algoritmo propuesto utiliza el clasificador *Bayes* para seleccionar el conjunto de variables contextuales relevantes e ignorar la información irrelevante (que puede actuar como el ruido) pertenecientes a cada clase de entidad. El conjunto relevante de variables contextuales obtenido de este modo se utilizan en la formación de grupos de características contextuales que tienen patrón de calificación similar perteneciente a cada clase. La información obtenida se integra en el sistema de recomendación para determinar el conjunto de vecinos del usuario y predecir la calificación que el usuario final va a dar a un ítem que no ha puntuado antes.

La información de contexto del usuario y los colaboradores se han denominado como variables contextuales, que son factores como la edad, sexo, ocupación, estado civil, entre otros para un usuario.

Características contextuales: Las variables contextuales consideradas son categorías en la naturaleza. Por ejemplo, la categoría edad las variables contextuales puedan ser: por debajo de 18-24, 25-34, 35-49, 50-64, 65 o más. Los valores de las variables contextuales dadas por un usuario conforman su característica contextual, tal como [18-24, Mujer, Estudiante, Soltero]

El término MICF se refiere al conjunto de características contextuales que muestran similar patrón de calificación al del usuario en todas las clases de entidades. Estas MICF son luego utilizadas en la determinación del conjunto de vecinos del usuario que desempeñan en la estimación de la calificación que un usuario final le daría a un elemento que no ha puntuado antes.

Para la extracción de variables contextuales relevantes, crean subconjuntos de variables y luego utilizan el clasificador bayesiano usando la técnica de aprendizaje supervisado. El subconjunto de las variables contextuales con mínima tasa de error se selecciona como el conjunto de variables contextuales relevantes. Las variables contextuales de cada clase, la

información contextual del usuario final y de los colaboradores se utiliza para predecir la calificación media que el usuario final le dará a una clase.

El conjunto de características contextuales se incorpora entonces en el sistema de recomendación para predecir la calificación que un usuario final daría a un elemento sin valorar. El conjunto de los colaboradores en la base de datos cuya información contextual está contenida en el conjunto MICF de un usuario se denomina: conjunto de vecino del usuario.

3.5 Combinación de filtrado colaborativo y similitud de texto para recomendaciones de perfiles de expertos en sitios Web Sociales

En el trabajo desarrollado por (Spaeth, 2013) en la Escuela Politécnica de Montreal en Quebec, Canadá, propone un sistema de recomendación que infiere gente experta en un tema (enfocado en compartir experiencia). El objetivo es recomendar expertos para otros expertos. Propone la integración de las técnicas de minería de texto para mejorar el algoritmo de filtrado colaborativo.

Utiliza varias técnicas para determinar la similitud entre los usuarios y los ítems, a continuación se describen cada una:

Similitud texto

Para determinar la similitud texto se sigue un proceso: primero, lematiza el perfil que describe la experiencia del experto. Toman en consideración que algunas palabras son más significativas que otras. Para dar valor a las palabras más significativas utilizan la técnica de TF-IDF, Fórmula 3.4.

$$idf_i = \log \frac{|D|}{|\{d_j: t_i \in td_j\}|} \quad (\text{Fórmula 3.4})$$

Donde $|D|$ es el número total de descripciones de experiencia y $|\{d_j: t_i \in d_j\}|$ el número de descripciones de experiencia con el lema t_i . Para calcular la similitud entre los ítems se utiliza la similitud coseno, Fórmula 3.5.

$$S_d = \frac{T_d^T x T_d}{\|T_d^T x T_d\|} \quad (\text{Fórmula 3.5})$$

Similitud etiqueta

Esta similitud se utiliza para aliviar las debilidades de la similitud texto. El vocabulario puede ser grande, que resulta en una matriz dispersa, y aliviar el problema de polisemia (la gente puede usar diferentes palabras para decir la misma cosa).

Para solucionar estas debilidades hacen uso de etiquetas en los perfiles que ayudan a reducir el espacio de vocabulario y los usuarios pueden tener una tendencia natural a evitar términos ambiguos en la elección de las etiquetas.

Utilizan el mismo algoritmo que en similitud texto, pero con etiquetas en lugar de descripciones lematizadas para calcular las recomendaciones basadas únicamente en la similitud etiqueta.

Combinando similitud texto y similitud etiquetas con el análisis semántico latente

En este punto se combinan las medidas de similitud de texto y similitud etiquetas en un solo espacio y se emplea el análisis semántico latente (LSA) para aumentar la relevancia de la medida de similitud.

En primer lugar se construye una nueva matriz M_c de etiquetas y palabras combinadas en el espacio de la columna y los perfiles de experiencia en el espacio de la fila. A continuación se utiliza la técnica de TF-IDF para construir la nueva matriz ponderada T_c y calcular la similitud coseno entre las descripciones de experiencia S_c .

El análisis semántico latente puede combinar términos y etiquetas. Y utilizando la descomposición de valor singular (SVD) se reduce el número de dimensiones. Para ello, la matriz de T_c se descompone mediante SVD: $T_c = U\Sigma V$. Al reducir el número de dimensiones se anulan los valores más pequeños y se construye una nueva matriz $TSVD$ y una nueva similitud coseno de descripciones de experiencia $SSVD$ las cuales son útiles para determinar los valores de similitud. Dentro de las pruebas realizadas los resultados mostraron que las dimensiones con 50 factores latentes dan los mejores resultados.

Filtrado colaborativo

El enfoque de filtrado colaborativo utiliza datos de navegación junto con dos niveles de interacciones: mensajes (envío de mensajes a otros expertos); y reuniones (contactar con el experto para tener un encuentro). Crean una matriz MCF de valoraciones inicializado a 0 y se agregan los siguientes valores a la entrada correspondiente en la matriz:

Perfil Visto: 1 - el usuario (columna) ha visto el perfil del objetivo potencial (fila)

Mensaje: 2 - el usuario (columna) ha enviado un mensaje al objetivo potencial (fila)

Reunión: 4 - el usuario (columna) ha reunido con el objetivo (fila)

La elección de los valores de 1/2/4 se basa en intuición.

Los valores son acumulativos. Por ejemplo, si un usuario ha visto un perfil y se envía un mensaje, el valor resultante será 3. El matriz MCF de valoraciones puede ser considerada como un grafo dirigido, donde los pesos son asignados a las aristas. La Figura 3.1 muestra un ejemplo de dicha estructura, con S y R representan a los usuarios.

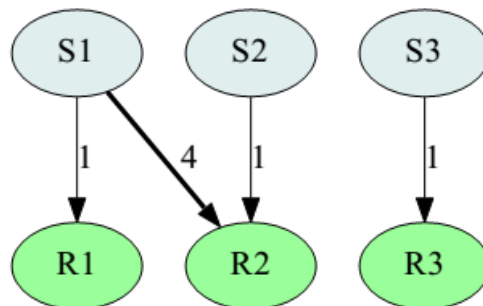


Figura 3.1 Grafo ponderado.

Para evaluar la proximidad entre dos usuarios se utiliza la similitud coseno Fórmula 3.6.

$$S_{CF} = \frac{M_{CF}^T M_{CF}}{\|M_{CF}^T M_{CF}\|} \quad (\text{Fórmula 3.6})$$

Combinando similitud texto y similitud etiquetas con filtrado colaborativo

Primero se calcula la técnica combinada de similitud texto y la similitud etiqueta. Ésta arroja los primeros valores representados como $R_{desc+tags}$. Los valores de la técnica de filtrado colaborativo se representan en R_{CF} .

Se combinan estas puntuaciones mediante el cálculo de una media geométrica ponderada entre los dos valores. Los mejores resultados se lograron con la Fórmula 3.7.

$$R_{CF+tags+sem} = (R_{CF} + 1)^{1/3} * (R_{desc+tags} + 1)^{2/3} \quad (\text{Fórmula 3.7})$$

3.6 Un marco de trabajo para la creación rápida de prototipos de sistemas de recomendación basados en conocimiento en el dominio de aprendizaje

En la investigación desarrollada por (Ruiz Iniesta, 2012) en la Universidad Complutense de Madrid, propone un *framework* que permite la rápida creación de prototipos de sistema de recomendación utilizando el algoritmo basado en conocimiento. Este trabajo se enmarca dentro de esta línea de investigación y más concretamente en la aplicación de estos sistemas al ámbito educativo orientado al dominio de aprendizaje como soporte al acceso personalizado de recursos existentes en repositorios electrónicos.

Esta investigación se enfoca principalmente a tres puntos:

- la introducción de diversidad en las listas de recomendación,
- la incorporación de información contextual en el proceso de recomendación
- y la exploración de modelos de interacción que alivien la carga de trabajo del usuario.

La introducción de diversidad se propone como un método para otorgar al usuario variedad dentro de los productos recomendados.

La incorporación de la información contextual en los sistemas de recomendación brinda nuevas oportunidades de personalización en éstos, ya que el sistema tiene más conocimiento del usuario pudiendo adaptar mejor el conjunto de productos recomendados. Proponen hacer uso de la actividad e información contextual en forma de un modelo de usuario, que puede ser un buen método para mejorar la calidad de las recomendaciones. La incorporación del contexto se modela a través de ontologías.

Finalmente, **los modelos de interacción alternativa** a la formulación de una consulta directa brindan nuevas facilidades de interacción con el recomendador, haciendo que usuarios poco expertos en el dominio o con una actitud pasiva hacia el mismo obtengan recomendaciones sin suponer un esfuerzo para ellos. Estos modelos de interacción también facilitan la obtención de las preferencias del usuario por parte del sistema de una manera no intrusiva para él. Por este motivo, implementaron un modelo de navegación por propuesta que sugiera de manera proactiva al estudiante un primer conjunto de objetos de aprendizaje adaptados a sus necesidades, y que permita de manera sencilla la obtención de realimentación del usuario.

3.7 SAPRS: sistema de recomendación proactivo consciente de la situación con explicaciones

Un sistema de recomendación multi agente (Bedi, 2014) basado en el algoritmo filtrado colaborativo, recomienda restaurantes para dispositivos móviles. Este sistema de recomendación proactivo consciente de la situación (SAPRS) propone recomendaciones relevantes y justificables al usuario en el contexto adecuado. También ofrece una explicación al usuario acerca de porqué se le recomienda un ítem.

SAPRS funciona en dos fases; (i) fase de evaluación de la situación, el sistema analiza la situación actual es decir, si en el contexto actual el usuario necesita una recomendación, esta información se calcula empleando la lógica difusa; (ii) la fase de evaluación de los ítems, SAPRS genera recomendaciones para el usuario destino mediante el algoritmo de filtrado colaborativo tomando en cuenta la ubicación del usuario (el contexto, el cual determina por la posición GPS).

3.8 Recomendaciones para la empresa: eventos, contacto y grupo

En esta investigación, realizada en el corporativo MITRE en Bedford Massachusetts (Gertner, 2012), implementa 3 prototipos de sistemas de recomendación: recomendador de eventos, recomendador de contactos y recomendador de grupos. Implementa el *framework* de recomendación de código abierto “Apache Mahout” y utiliza el algoritmo de filtrado colaborativo.

Recomendador de Eventos

MITRE es una empresa de diversidad tecnológica que emplea los eventos para difundir y compartir información tecnológica. Cuenta con un programa de innovación que alberga eventos que reúne personal con interés en los mismos temas de investigación. Además organiza reuniones formales de intercambio tecnológico para apoyar la creación de redes y apoyar en la solución de desafíos tecnológicos.

Este recomendador se alimenta de los eventos de la organización. Para hacer las recomendaciones de los ítems toma en cuenta en los que está interesado el usuario y que pertenecen a la organización. Uno de los desafíos fue la recopilación de los datos de interés del usuario, debido a que muy pocos eventos en MITRE requieren registro, por lo que no hay datos de registros de asistencia a eventos pasados. Para determinar cuáles son de interés para el usuario analizan las acciones de “añadir al calendario” y “eliminarlos de la lista.

Recomendador de contactos

El recomendador de contactos llamado *Handshake* está diseñado para recomendar personas con intereses o actividades comunes. El recomendador utiliza la información de múltiples fuentes: marcadores sociales, wikis, listas de correo electrónico, así como la pertenencia a grupos y conexiones con otros contactos.

Recomendador de grupos

Handshake tiene cientos de grupos creados por los usuarios. Y la única manera de conocer los grupos son de boca en boca o navegando manualmente por la lista de grupos. Solucionaron este problema implementando un recomendador de grupo, utilizan el algoritmo de filtrado colaborativo de *Apache Mahout*. Este recomendador sufre del problema de arranque en frío en que los usuarios que más necesitan recomendaciones son los nuevos miembros de *Handshake* que pertenecen a pocos o ningún grupos.

3.9 Repensando el ecosistema de investigación en recomendación: reproducibilidad, la apertura y lenskit

En este artículo realizado en la Universidad de Minnesota en Minneapolis Estados Unidos (Ekstrand, 2011) proponen un *framework* de código abierto llamado *lenskit*. Este *framework* proporciona una implementación cuidadosamente afinada de los principales algoritmos de filtrado colaborativo, un API (por sus siglas en inglés *Application Programming Interface*) para los casos de uso común de los sistemas de recomendación y un *framework* para realizar las evaluaciones fuera de línea de estos algoritmos.

El código fuente de *LensKit* está disponible al público bajo la licencia GNU GPL (versión 2 o posterior), lo que permite ser utilizado en una variedad de contextos. Está programado en Java, por lo que se puede utilizar y ser extendió en muchos ambientes.

LensKit permite la implementación de los tres algoritmos de filtrado colaborativo más comúnmente utilizados: basado en usuario, basado en ítem y descomposición en valores singulares (SVD por sus siglas en inglés *Singular Valúé Decompositions*). Estos algoritmos son divididos cada uno en componentes que se pueden recombinar y sustituir de forma independiente.

Además este *framework* permite realizar evaluaciones fuera de línea, entre las estrategias de evaluación que proporciona *LensKit* se encuentran: *train-test* y validación cruzada *k-fold*. *LensKit* incluye las métricas de precisión MAE y RMSE (a nivel promedio global y promedio por usuario), así como los indicadores de cobertura, las métricas basadas en *rank*: *Normalized discounted cumulative gain* (NDCG) y de la utilidad *half-life*.

NDCG se emplea como un evaluador de la predicción, clasificando los ítems en orden de predicción, calcula el NDCG utilizando los *ratings* reales del usuario para cada ítem. El resultado es una medida de la capacidad del predictor para clasificar los ítems de manera consistente con el orden impuesto por los *ratings* del usuario.

Error medio absoluto (MAE por sus siglas en inglés *Mean Absolute Error*). Mide la desviación de las recomendaciones predichas y los valores reales. A menor MAE mejor predice el sistema las evaluaciones de los usuarios.

Los experimentos realizados en este trabajo demuestran que las implementaciones de *LensKit* igualan o superan a las implementaciones de los algoritmos más conocidos.

Es importante resaltar que uno de los principales objetivos de *LensKit* es invitar a la comunidad, a probar este *framework*, extender la funcionalidad de los algoritmos de recomendación y contribuir para formar un ecosistema de una variedad más amplia de algoritmos de recomendación.

Este *framework* de recomendación abierto provee la base de los algoritmos basados en filtrado colaborativo listo para su ejecución, con el fin de proporcionar una implementación completa y que la comunidad pueda probar con otro conjunto de datos o comparar los resultados que predice LenKit contra las mejoras propuestas de la comunidad.

3.10 Estudio comparativo

A continuación se presentan los factores tomados en cuenta para la comparación de los trabajos encontrados en el estado de arte y el presente trabajo de investigación.

Los trabajos analizados abordan diferentes puntos de interés para este trabajo como: la forma como se utiliza ampliamente la técnica de filtrado basado en contenido para recomendar ítems entre los cuales están: libros, documentos de investigación, entre otros; la técnica de filtrado colaborativo en la cual se encontraron dos *frameworks* para implementar esta técnica en un dominio específico siempre que se cumpla con los requerimientos.

Tipo de ítem: se especifica si el trabajo recomienda este tipo de ítem, “✓” en caso de si y “X” en caso de no ser un ítem recomendado. El resultado de la comparación se muestra en la Tabla 3.2.

Nomenclatura utilizada en la tabla 3.2.

PR	Profesores	CU	Cursos
OC	Objetos de conocimiento	PI	Proyectos de investigación
IN	Instalación especializada	NO	Noticias, eventos
IT	Infraestructura tecnológica	OA	Objetos de aprendizaje
SE	Servicios tecnológicos	TR	Técnica de recomendación
CE	Certificación		

Tipo de ítem	PR	IN	IT	OC	IT	SE	CE	CU	PI	NO	OA
Gertner,2012	✓	X	X	X	X	X	X	X	X	✓	X
Rana, 2013	X	X	X	X	X	X	X	X	X	X	X
Jiang,2012	X	X	X	✓	X	X	X	X	X	X	X
Pera,2014	X	X	X	✓	X	X	X	X	X	X	X
Bedi,2014	X	X	X	X	X	✓	X	X	X	X	X
Vashisth, 2012	X	X	X	✓	X	X	X	X	X	X	X
Ekstrand, 2011	X	X	X	X	X	X	X	X	X	X	X
Spaeth, 2013	X	X	X	X	X	X	X	X	X	X	X
Ruiz Iniesta, 2012	X	X	X	X	X	X	X	X	X	X	✓
Propuesta	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	X

Tabla 3.2 Tabla comparativa.

En la tabla 3.2 se aprecia que una de las principales características que diferencian este trabajo de investigación con los trabajos encontrados en el estado del arte, es la cantidad y el tipo de ítems que se recomiendan. En este trabajo de investigación se cuenta con un repositorio que contiene un conjunto de ítems heterogéneos, y los demás trabajos únicamente recomienda uno o dos diferentes tipos de ítems.

Algoritmo de recomendación	Filtrado Colaborativo	Basado en Contenido	Basado en Conocimiento	Técnica Híbrida	Explicación de la recomendación
Gertner,2012	✓	X	X	X	X
Rana, 2013	✓	X	X	X	X
Jiang,2012	X	✓	X	X	X
Pera,2014	X	✓	X	X	X
Bedi,2014	✓	X	X	X	✓
Vashisth, 2012	✓	✓	✓	✓	X
Ekstrand, 2011	✓	X	X	X	X
Spaeth, 2013	✓	✓	✓	✓	X
Ruiz Iniesta, 2012	X	X	✓	X	X
Propuesta	✓	✓	X	X	✓

Tabla 3.3 Tabla comparativa algoritmos de recomendación.

La tabla 3.3 muestra las diferentes técnicas de recomendación utilizadas en el estado del arte. Se aprecia que son ampliamente utilizadas la técnica basada en contenido y filtrado colaborativo, la técnica basada en conocimiento es la que se utiliza en menor frecuencia. Esta tabla también muestra que son pocos los trabajos que hacen uso de las explicaciones en las recomendaciones.

Para este trabajo de investigación son de interés las técnicas basadas en contenido y filtrado colaborativo para inferir las recomendaciones. Así como el uso de las explicaciones para presentar la recomendación acompañada de una explicación que exponga los motivos que llevaron a recomendar un ítem.

CAPÍTULO 4

PROPUESTA DE SOLUCIÓN

El contenido descrito en este capítulo presenta la propuesta de solución que se empleó para la realización del servicio de recomendación para la PCCT del TecNM.

4.1 Introducción

Las tendencias en el desarrollo de Sistemas de Recomendación (SR) mantienen una evolución constante, la cual está motivada en gran medida por la búsqueda de una mayor eficiencia en las recomendaciones que se le ofrecen al usuario (González, 2012).

El servicio de recomendación realizado en este trabajo de tesis se desarrolló sobre el dominio de Instituciones de Educación Superior (TecNM); realizando recomendaciones sobre el conjunto de ítems heterogéneos: Personas, objeto de conocimientos, instalaciones especializadas, servicios tecnológicos, infraestructura tecnológica especializada, certificaciones, cursos y proyectos de investigación; mediante diferentes técnicas de recomendación: filtrado colaborativo y filtrado basado en contenido; sobre cuatro tipos de usuario: estudiantes, profesores/investigadores, egresado y empresas.

Las explicaciones (definidas en la sección 2.6.1) generadas mediante las plantillas, están motivadas para aumentar la confianza del usuario hacia el sistema, las cuales son llenadas con la información correspondiente a campos de texto variable, exponiendo la información utilizada por el algoritmo de recomendación del por qué se originaron dichas recomendaciones.

El algoritmo de recomendación y las plantillas utilizadas para generar las explicaciones, son dos actividades que se pueden hacer juntas o una independiente de la otra. Se pueden presentar dos escenarios: el primero es cuando el SR se encuentre terminado y se requiere implementar un módulo de explicaciones, el segundo escenario es cuando el SR se desarrolla a la par con el módulo de explicaciones, siendo este último el que se empleó en esta tesis.

Para integrar el motor de recomendación y las plantillas de explicación se realizaron las actividades propuestas por (Gálan, 2014). Las actividades se dividen en tres fases: análisis, diseño e implementación; cada fase contempla diferentes actividades.

4.2 Fase de análisis

4.2.1. Análisis del funcionamiento del SR y las técnicas utilizadas

Esta actividad consta de dos tareas las cuales son el análisis del sistema de recomendación utilizado y el análisis de las técnicas utilizadas por el sistema de recomendación, en la primera tarea se analizó el funcionamiento básico del SR, tipo de recomendaciones, dominio, usuarios, técnica/s de recomendación utiliza/s, etc.

4.2.1.1. Análisis del sistema de recomendación utilizado

Como se mencionó anteriormente, el servicio de recomendación se desarrolló para la Plataforma de las Capacidades Científicas y Tecnológicas (PCCT), el cual se implementó en un dominio de Instituciones de Educación Superior; realizando recomendaciones sobre el conjunto de ítems heterogéneos: Personas, objeto de conocimientos, instalaciones especializadas, servicios tecnológicos, infraestructura especializada, certificaciones, cursos y proyectos de investigación; mediante diferentes técnicas de recomendación: *filtrado*

colaborativo y filtrado basado en contenido; sobre cuatro tipos de usuario: estudiantes, profesores/investigadores, egresado y empresas.

4.2.1.2. Análisis de las técnicas utilizadas por el sistema de recomendación

4.2.1.3. Algoritmo de recomendación basada en contenido

Los Sistemas de Recomendación (SR) basados en contenidos utilizan algoritmos que analizan los campos descripción de los ítems, con el objetivo de identificar aquellos que resulten de interés para el usuario (Balabanović, 1997). Los recomendadores basados en contenido son aquellos que realizan la recomendación en base a la descripción y/o contenido de los ítems que va a recomendar (Ruiz Iniesta A. , 2014).

Este tipo de algoritmo utiliza contenido en forma de texto, para comparar un usuario con un ítem. Tanto el perfil del usuario como los atributos que definen a un ítem es texto. Esta información se utiliza para determinar el nivel de similitud entre ellos. Esto nos condujo a analizar los atributos que mejor describen a cada tipo de usuario e ítem. A continuación se muestran los descriptores que se seleccionaron.

4.2.1.4. Usuarios de la plataforma

Usuario: Profesor

La Tabla 4.1 muestra los descriptores tomados para hacer el perfil del usuario profesor.

Usuario Profesor
Reseña
Línea de Investigación Título
Línea de Investigación Descripción
Línea de Investigación Tópicos de Interés

Tabla 4.1 Descriptores del perfil del usuario académico.

Los valores de estos atributos se concatenaron en un cuerpo de texto, la Figura 4.1 muestra un ejemplo del perfil del usuario profesor.

```
@_id_9. FRDMS@TITULO: . es egresado del centro de investigación en computación
del instituto politécnico nacional (cic-ipn), obteniendo los grados de
maestría y doctorado en ciencias de la computación, este último con
mención honorífica, en el año 2012. actualmente es miembro del sistema
nacional de investigadores (sni) y se desempeña como vocal de la sociedad
mexicana de inteligencia artificial (smia).su área de interés es la lingüística
computacional, donde se encuentra desarrollando trabajos de investigación y
dirigiendo tesis de maestría y doctorado en las líneas principalmente de
análisis de sentimientos y emociones, análisis automático de noticias,
lexicografía computacional y recursos léxicos grandes (diccionarios y
corpus).su producción científica incluye la dirección y participación en
el desarrollo de proyectos de investigación nacionales e internacionales,
publicaciones en revistas indizadas y arbitradas, y en revistas científicas
populares. ha elaborado capítulos de libros, impartido ponencias tanto en méxico
como el extranjero y ha participado como revisor técnico de congresos nacionales e
internacionales.universidad autónoma de guerrerofacultad de ingenieríaingeniería
en computacióncédula: 4535178? procesamiento de lenguaje natural (nlp)? entendimiento
de lenguaje natural (nlu)? recursos léxicos grandes: diccionarios, gramáticas,
córpora de textos? morfología, sintaxis y semántica minería de opiniones .
procesamiento de lenguaje natural cenidet . técnicas y herramientas
computacionales para el procesamiento automático de los textos escritos en el
lenguaje natural,en español. realiza investigación en el área de procesamiento de
lenguaje natural y computo lingüístico y su objetivo es desarrollar nuevas técnicas
y herramientas computacionales para el procesamiento automático de los textos
escritos en el lenguaje natural.areas:manejo inteligente de documentos traducción
automática entre los lenguajes humanos extracción de información de internet y
cantidades grandes de textos comprensión del lenguaje natural por computadora .
```

Figura 4.1 Cuerpo de texto del usuario profesor.

Usuario: Estudiante

La Tabla 4.2 muestra los descriptores tomados para el perfil del usuario Estudiante.

Usuario Estudiante
Reseña
VinculacionAcademico Descripción (Estancia)
VinculacionAcademico Actividades(Estancia)

Tabla 4.2 Descriptores del perfil del usuario Estudiante.

Usuario Egresado

La Tabla 4.3 muestra los descriptores tomados para hacer el perfil del tipo de usuario Egresado.

Usuario Egresado
Reseña
Certificación Título
Certificación Descripción
Experiencia Laboral Descripción
Experiencia Laboral Actividades

Tabla 4.3 Descriptores del perfil del usuario Egresado

Usuario Empresa

La Tabla 4.4 muestra los descriptores tomados para hacer el perfil del usuario Empresa.

Usuario Empresa
Título
Descripción
Responsabilidades
Habilidades

Tabla 4.4 Descriptores del perfil del usuario Empresa.

4.2.1.5. Tipos de ítems recomendables

Tipo de Ítem: Objeto de conocimiento

La Tabla 4.5 muestra los descriptores tomados para hacer el perfil del ítem Objeto de conocimiento.

Ítem objeto de conocimiento
Nombre
Tipo
Descripción
Palabras claves

Tabla 4.5 Descriptores del perfil del ítem Objeto de conocimiento.

Los valores de estos atributos se unieron en un cuerpo de texto. La Figura 4.2 muestra un ejemplo del cuerpo de texto del ítem Objeto de conocimiento.

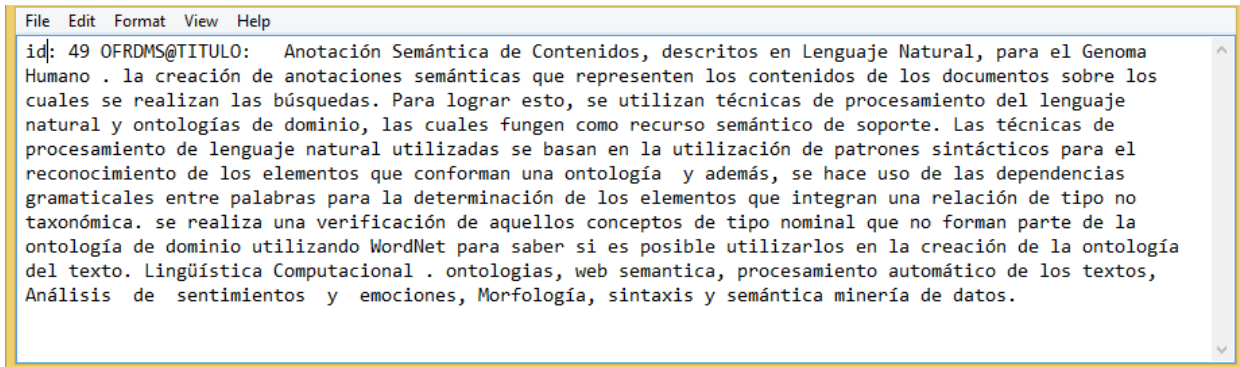


Figura 4.2 Cuerpo de texto del ítem Objeto de conocimiento.

Tipo de Ítem: Profesor

La Tabla 4.6 muestra los descriptores tomados para hacer el perfil del ítem Profesor.

Ítem Profesor (Persona)
Reseña
Línea de Investigación Titulo
Línea de Investigación Descripción
Línea de Investigación Tópicos de Interés

Tabla 4.6 Descriptores del perfil del ítem profesor.

Tipo de Ítem: Infraestructura

La Tabla 4.7 muestra los descriptores tomados para hacer el perfil del ítem Infraestructura.

Ítem Infraestructura
Nombre
Tipo
Descripción

Tabla 4.7 Descriptores del perfil del ítem Infraestructura.

Tipo de Ítem: Instalación

La Tabla 4.8 muestra los descriptores tomados para hacer el perfil del ítem Instalación.

Ítem Instalación
Nombre
Tipo
Descripción

Tabla 4.8 Descriptores del perfil del ítem Instalación.

Tipo de Ítem: Servicio

La Tabla 4.9 muestra los descriptores tomados para hacer el perfil del ítem Servicio.

Ítem Servicio
Nombre
Tipo
Descripción

Tabla 4.9 Descriptores del perfil del ítem Servicio.

Tipo de Ítem: Proyecto

La Tabla 4.10 muestra los descriptores tomados para hacer el perfil del ítem Proyecto.

Ítem Proyecto
Título
Descripción
Objetivo
Keywords

Tabla 4.10 Descriptores del perfil del ítem Proyecto.

Tipo de Ítem: Certificación

La Tabla 4.11 muestra los descriptores tomados para hacer el perfil del ítem Certificación.

Ítem Certificación
Título
Descripción
Departamento

Tabla 4.11 Descriptores del perfil del ítem Certificación.

Tipo de Ítem: Curso

La Tabla 4.12 muestra los descriptores tomados para hacer el perfil del ítem Certificación.

Ítem Certificación
Título
Descripción
Departamento

Tabla 4.12 Descriptores del perfil del Curso.

4.2.1.6. Sistema de recomendación basado en contenido

Proceso de ponderación del algoritmo filtrado basado en contenido

Una vez definidos los perfiles de los tipos de usuarios y de los tipos de ítems, se procede a ponderar éstos para determinar las recomendaciones que recibirá cada tipo de usuario. Este proceso se realiza en 4 pasos, los cuales se muestran en la Figura 4.3.

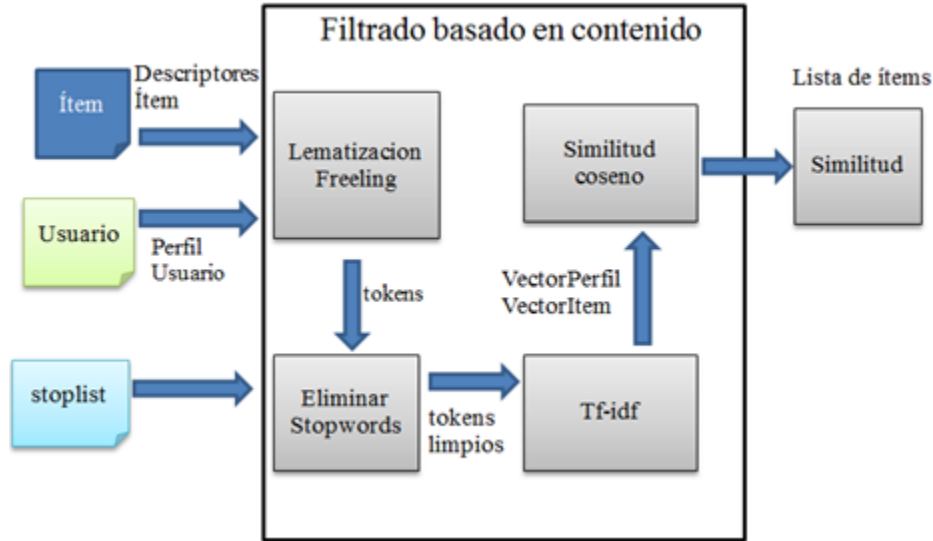


Figura 4.3 Proceso de ponderación del algoritmo de filtrado basado en contenido.

A continuación se describe cada uno de los pasos del proceso llevado a cabo por el algoritmo de filtrado basado en contenido.

Etiquetado de palabras con la herramienta *freeling*: el programa *freeling*² es el encargado de etiquetar el contenido del perfil tanto de los tipos de usuarios como de los tipos de ítems. En el trabajo realizado por (Spaeth, 2013) utilizan una técnica para convertir las palabras a su lema y utiliza el lema³ en lugar de la palabra como la escribe el usuario en su perfil. La Figura 4.4 muestra un ejemplo del proceso de etiquetado con *freeling*.

El	gato	come	pescado	y	bebe	agua	.
<i>el</i>	<i>gato</i>	<i>comer</i>	<i>pescado</i>	<i>y</i>	<i>beber</i>	<i>agua</i>	<i>.</i>
DA0MS0	NCMS000	VMIP3S0	NCMS000	CC	VMIP3S0	NCCS000	Fp
1	1	0.994868	0.608233	0.999962	0.994868	0.99177	1

Figura 4.4 Ejemplo de etiquetado con *freeling*.

Eliminar *stopwords*: estas son una lista de palabras de uso frecuente que, tanto en la indexación como en la búsqueda, no se tienen en consideración, se omiten. Este paso nos permite eliminar palabras que no aportan significado en la descripción de los perfiles.

TF-idf: es una medida estadística que refleja la importancia de una palabra en un documento (Jiang, 2012), esta medida se utiliza como un factor de ponderación, la cual aumenta dependiendo del número de veces que una palabra aparece en el documento, pero es compensada por la frecuencia de la palabra en la colección de documentos, lo que

² <http://nlp.lsi.upc.edu/freeling/>

³ Lema: base o raíz de una palabra.

permite tomar en cuenta de que algunas palabras son más significativas que otras (Spaeth, 2013). La Fórmula 4.1 se utilizó en este proceso.

$$idf_i = \log \frac{|D|}{|\{d_j: t_i \in d_j\}|} \quad (\text{Fórmula 4.1})$$

Donde $|D|$ es el número total de documentos y $|\{d_j: t_i \in d_j\}|$ el número de documentos donde aparece el lema t_i .

Para calcular el valor de tf-idf se utiliza la Fórmula 4.2. Donde w_j es el valor tf*idf del lema j . con el valor de tf-idf llenamos la matriz del perfil del usuario y de los ítems. La cual utilizamos para medir la similitud entre estos.

$$w_j = tf * idf (j, d, D) = tf (j, d) x idf (j, D) \quad (\text{Fórmula 4.2})$$

Similitud coseno: es la medida de similitud utilizada por el algoritmo de filtrado basado en contenido. Entre los trabajos que utilizan esta técnica como medida de similitud para el algoritmo basado en contenido se encuentran: (Jiang, 2012), (Spaeth, 2013), (Pera, 2014). La Fórmula 4.3 se utilizó para calcular la similitud coseno. Donde A es el vector del perfil del usuario y B es el vector del perfil del ítem.

$$\text{Sim} = \cos \theta = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (\text{Fórmula 4.3})$$

4.2.1.7. Sistemas de recomendación basados en filtrado colaborativo

Los algoritmos de filtrado colaborativo proporcionan recomendaciones o predicciones de ítems basándose en la opinión de usuarios con características similares. La meta de los algoritmos de filtrado colaborativo es sugerir nuevos ítems o predecir la utilidad de cierto ítem para un usuario particular, basándose en los gustos y opiniones definidos previamente por usuarios similares (Sarwar, 2001).

Los algoritmos de filtrado colaborativo añaden valoraciones de objetos, reconocen características comunes entre usuarios a partir de las valoraciones y generan nuevas recomendaciones mediante la comparación entre usuarios. Un perfil de usuario típico en un SR basado en filtrado colaborativo se compone de un vector de ítems y sus valoraciones, modificadas dinámicamente conforme el usuario utiliza el sistema (Alejandres H. , 2012). En algunos casos, las valoraciones pueden ser de tipo binario o valores numéricos que indican cierto grado de preferencia. Estas calificaciones son usadas para identificar usuarios que son similares en calificaciones (Tintarev N. &. , 2011.).

La información con la que trabaja el algoritmo de recomendación consta de una serie de ítems, usuarios y valoraciones proporcionada por los usuarios sobre esos ítems. El algoritmo de filtrado colaborativo puede ser representado como el problema de predecir los valores faltantes en una matriz de usuario-ítem. La Tabla 4.13 muestra un ejemplo de una matriz de valoraciones de usuario-ítem donde las celdas que tienen un valor representan la

valoración que el usuario asigno a ese ítem. El algoritmo está intentando proveer a Nathan una predicción para la película *Titanic*.

	Star Wars	Hoop Dreams	Contact	Titanic
Joe	5	2	5	4
John	2	5		3
Al	2	2	4	2
Nathan	5	1	5	?

Tabla 4.13 Matriz de valoraciones usuario-ítem.

Las valoraciones pueden generarse de dos formas: implícitas y explícitas. Las implícitas son aquellas que se pueden generar a partir de la interacción usuario-ítem, y las explícitas son aquellas en que el usuario califica directamente un ítem (Herlocker J. L., 2004).

Método basado en vecindario

Este método puede ser separado en 3 pasos: ponderar todos los usuarios con respecto a la similitud con el usuario activo; seleccionar un subconjunto de usuarios (selección de vecinos) a utilizar como un conjunto de predictores (posiblemente para un ítem específico), normalizar las valoraciones y calcular la predicción con la combinación de las valoraciones de los vecinos seleccionados.

Ponderar todos los usuarios con respecto a la similitud con el usuario activo

Para seleccionar el conjunto de vecinos que comparten gustos similares al usuario activo, se utiliza la técnica de correlación de Pearson. La cual calcula la correlación estadística entre las valoraciones que tienen en común dos usuarios para determinar su similitud entre ellos (Ekstrand M. D., 2011). La Fórmula 4.4 se utiliza para determinar la similitud entre dos usuarios.

$$s(u, v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_u \cap I_v} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_u \cap I_v} (r_{v,i} - \bar{r}_v)^2}} \quad (\text{Fórmula 4.4})$$

La desventaja de la correlación de Pearson es que calcula alta similitud entre usuarios con pocas valoraciones en común. Este problema puede ser solucionado estableciendo un umbral del número de ítems co-valorados, devaluando la similitud cuando el número de ítems co-valorados cae por debajo de este umbral. Un valor del umbral igual a 50 es útil para mejorar la precisión de la predicción (Ekstrand M. D., 2011). Si dos usuarios tienen menos de 50 ítems co-valorados, se aplica la fórmula de $n/50$, donde n es el número de ítems co-valorados. Si tienen más de 50 ítems co-valorados se asigna el valor de 1. De esta manera correlaciones con pequeñas cantidades de ítems co-valorados son apropiadamente devaluados, y las correlaciones con 50 o más ítems co-valorados no dependen del número de ítems co-valorados.

Selección de los vecinos

Después de asignar los valores de similitud a los usuarios, se debe determinar qué datos de los usuarios serán utilizados en el cálculo de la predicción para el usuario activo. Es útil, tanto para la precisión como para el rendimiento, seleccionar un subconjunto de usuarios (el vecindario) para utilizarlos en el cálculo de la predicción, en lugar de todos los usuarios. En el trabajo de (Ekstrand M. D., 2011) dice que el error medio absoluto de la predicción aumenta a medida que aumenta el tamaño del vecindario. Las correlaciones altas (tales como aquellos con correlaciones superiores a 0,5) puede ser más valioso para calcular las predicciones que aquellos con correlaciones bajas. Las técnicas de umbral de correlación y mejor n-vecinos son utilizadas para determinar la cantidad de vecinos.

La primera técnica establece un umbral de correlación, donde se seleccionan todos los vecinos con correlaciones mayores al umbral especificado. El establecer un umbral alto limita el vecindario que contiene buenas correlaciones, pero muchos usuarios no cuentan con correlaciones altas, lo que resulta en un vecindario pequeño que no puede proporcionar una predicción para mucho ítems. El establecimiento de un umbral de correlación más bajo resulta en un gran número de correlaciones bajas, anulando el propósito del umbral, donde un umbral de peso 0,1 proporciona aproximadamente la misma precisión que el uso de todos los usuarios disponibles en el conjunto de datos.

La segunda técnica consiste en seleccionar las mejores n correlaciones. Esta técnica funciona razonablemente bien, ya que no limita la cobertura de predicción. Sin embargo, la selección de un número mayor de n resultará en demasiado ruido para aquellos que tienen altas correlaciones. Elegir un n más pequeño puede causar predicciones pobres para aquellos usuarios que no tienen correlaciones altas.

La cuestión de cuántos vecinos seleccionar. En algunos sistemas, tales como el originalmente utilizado por *GroupLens* $N = T \setminus \{u\}$ (todos los usuarios se consideran como vecinos); en otros sistemas, los vecindarios son seleccionados para cada ítem basado en un umbral de similitud o tamaño del vecindario de tal manera que N_i son los k usuarios más similares a u que han evaluado el ítem i . Limitar el tamaño del vecindario puede dar lugar a predicciones más precisas, ya que los vecinos con baja correlación introducen más ruido que señal en el proceso. El umbral particular a usar es específico de cada dominio y de cada sistema, por lo que es necesario un análisis del conjunto de datos para elegir el tamaño de la vecindad para una implementación particular.

En análisis fuera de línea de datos disponibles en el dominio de películas, (Herlocker J. L., (1999, August).) Encontró que $k = 20$ es un buen umbral; valores en el rango de 20-50 son un punto de partida razonable en muchos dominios.

Calculando predicciones

Una vez que el vecindario ha sido seleccionado, las valoraciones de esos vecinos se combinan para calcular la predicción, después de escalar las valoraciones a una distribución común. La forma básica de combinar todas las valoraciones de los vecinos en una predicción, es calcular un promedio de las valoraciones, usando las correlaciones como los pesos.

Las dos técnicas más utilizadas para calcular las predicciones son:

Media promedio

Normalmente esto se hace mediante el cálculo de la media ponderada de las valoraciones de los usuarios vecinos, utilizando la similitud como los pesos. Restar la valoración media del usuario \bar{r}_u compensa las diferencias en el uso de la escala de valoraciones de los usuarios. La justificación para este enfoque es que los usuarios pueden distribuir las valoraciones en torno a diferentes puntos. Un usuario tiende asignar valoraciones más altas a los ítems, donde los buenos ítems obtienen 5s (cinco estrellas) e ítems pobres consiguen 3s (tres estrellas), mientras que otros usuarios pueden dar principalmente 1s, 2s, 3s. Intuitivamente, si un usuario da con poca frecuencia valoración de 5s, entonces este usuario no debería recibir muchas predicciones de 5s a menos que sean muy significativas. Este enfoque se muestra en la Fórmula 4.5.

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) * w_{a,u}}{\sum_{u=1}^n w_{a,u}} \quad (\text{Fórmula 4.5})$$

$P_{a,i}$ representa la predicción para el usuario activo a para el ítem i , n es el número de vecinos y $w_{a,i}$ es la similitud entre el usuario y el vecino u tal como se define por el coeficiente de correlación de Pearson.

Promedio de puntuación Z

La Fórmula 4.6, se puede extender para normalizar las valoraciones de los usuarios a z-score, dividiendo el desplazamiento de la media de las valoraciones por la desviación estándar de las valoraciones de cada usuario.

$$p_{a,i} = \bar{r}_a + \sigma_a \frac{\sum_{u=1}^n \frac{(r_{u,i} - \bar{r}_u)}{\sigma_u} * w_{a,u}}{\sum_{u=1}^n w_{a,u}} \quad (\text{Fórmula 4.6})$$

Z-score se emplea para usuarios que distribuyen sus calificaciones más o menos a través de la escala de evaluación.

Configuración de inicio

¿Cuál sería una buena configuración base para el algoritmo de filtrado colaborativo basado en usuario?

Los detalles exactos pueden variar de una aplicación a otra, pero si no tiene ninguna configuración. Un buen punto de partida es usar los vecinos más similares. Si sólo intenta calcular la predicción 25 o 30 vecinos es suficiente. Si se necesita obtener mayor cobertura, se tendrían que utilizar más vecinos para hacer recomendaciones.

Utilizar mean average o z-score para calcular la predicción. Z-score da un pequeño beneficio sobre mean average y toma un poco más de tiempo al realizar los cálculos. Se tendrían que realizar pruebas para el dominio en el cual el sistema de recomendación va hacer implementado, para determinar si vale la pena utilizar el algoritmo de z-score que tiene un error medio absoluto un poco menor que mean average, aunque el tiempo de procesamiento lleve un poco más de tiempo.

4.2.2. Fase de Diseño

4.2.3. Análisis de los estilos de explicación a implementar

Para generar las explicaciones se consideró implementar estilos de explicación según las técnicas de recomendación utilizadas por el sistema las cuales corresponden a: *Estilo basado con colaboración y estilo basado en contenido*, con el fin de explotar las características de cada técnica utilizada para la generación de las recomendaciones del sistema de recomendación.

El estudio de (Gálan, 2014) nos dice que los estilos de explicación pueden ser implementados en un sistema de recomendación de acuerdo a la técnica utilizada por el motor de recomendación sin importar su dominio de aplicación, siempre y cuando se exploten adecuadamente los recursos de la técnica de recomendación que originaron la recomendación de un ítem.

Los recursos utilizados por el motor de recomendación para crear las explicaciones se describen a continuación:

Contenido: Se emplean técnicas de similitud basadas en las características definidas para cada tipo de ítem.

Colaborativo: Se definen comunidades de usuarios similares al usuario activo.

En el proceso de selección de los recursos extraídos de acuerdo al tipo de ítem recomendado. Se recibe como entrada el archivo que contiene la lista de las recomendaciones, se identifica cada tipo de ítem recomendado (Personas, Objeto de Conocimientos, Servicios, etc.) para extraer los recursos, llenar las plantillas y generar las explicaciones para cada ítem. Por extracción de recursos se refiere a la recuperación de las características del ítem y lista de vecinos cercanos al usuario.

4.2.4. Identificación de los objetivos de las explicaciones en el SR

En el trabajo realizado por (Gálan, 2014), define que es necesario realizar un análisis detallado de cada uno de los objetivos de las explicaciones para determinar cuáles son los pros y contras de cada uno, ya que existe un problema de tradeOff (es decir, una situación en la cual se puede perder cierta cualidad a cambio de otra).

(Tintarev, 2011.) Presenta siete objetivos de las explicaciones, ver Tabla 4.14, estos objetivos son aplicables a las recomendaciones de un solo ítem, es decir, cuando el sistema ofrece una recomendación única.

Objetivo	Definición
Transparencia	Explican cómo trabaja el SR.
Escrutabilidad	Permite a los usuarios manejar posibles errores del SR.
Confianza	Incrementa la confianza del usuario en el SR.
Eficacia	Ayuda al usuario a tomar decisiones correctas.
Persuasión	Convince al usuario de probar o comprar determinado ítem.
Eficiencia	Ayuda a los usuarios a tomar decisiones rápidamente.
Satisfacción	Incrementa la facilidad de uso del SR, al ser más fácil y rápida la búsqueda de ítems relevantes.

Tabla 4.14 Objetivos de las explicaciones.

Cuando un Sistema de Recomendación proporciona explicaciones de las recomendaciones que proporciona al usuario, éstas le permite entender porqué recibió dichas recomendaciones y promueven objetivos como: la satisfacción y la confianza con relación al Sistema de Recomendación (Gálan, 2014). Cuando un usuario recibe una explicación, aumenta la probabilidad de que éste acepte una recomendación, debido a que el sistema proporciona transparencia en su proceso (Papadimitriou, 2011). Los sistemas que ofrecen transparencia exponen el razonamiento detrás de una recomendación y los usuarios son más propensos a confiar en una recomendación cuando conocen las razones detrás de esa recomendación (Herlocker, 2000).

El servicio que se desarrolló en esta investigación, recomienda un conjunto de ítems heterogéneos, los cuales se presentan al usuario mediante una lista, en el cual se busca generar explicaciones individuales para cada ítem y no una explicación para la lista completa de ítems recomendados. Los atributos de las explicaciones pueden contribuir al cumplimiento de múltiples objetivos. Una explicación comprensible ayuda a aumentar la confianza y la satisfacción del usuario.

El objetivo de implementar explicaciones para el servicio de recomendación es aumentar la confianza del usuario en el uso del servicio. Brindar explicaciones claras y comprensibles que le ayuden en proceso de recomendación, exponiendo la información utilizada para la emisión de la misma. De esta manera el usuario puede decidir si existe suficiente evidencia para tomarla como válida.

El motivo de inclinarnos por el incremento en la confianza del usuario, es que el sistema está orientado a instituciones de educación superior y se intenta que el usuario confíe en las recomendaciones, más que intentar convencerlo de aceptar una recomendación.

4.2.5. Definición de la estructura sintáctica de las explicaciones

Es necesario estructurar textualmente una explicación de acuerdo a la técnica de recomendación utilizada, para crear las plantillas de explicación llenadas con los campos variables que utiliza cada algoritmo de recomendación. El proceso para definir la estructura sintáctica del texto explicativo, requiere el análisis de los textos explicativos de cada uno de los ítems recomendados. Analizando los recursos que el servicio utilizó para crear las

recomendaciones, ya sean vectores de palabras clave, relaciones de los ítems o cualquier otro recurso, de tal forma que los textos explicativos presentados al usuario sean claros y coherentes.

Con base al estudio realizado por (Gálan, 2014) en el cual determino que las relaciones causales y los marcadores lingüísticos causales son los más adecuados para definir una estructura básica y simple para las plantillas de explicación textuales.

Se han tomado las relaciones causales presentadas por (Gálan, 2014), las cuales podemos adecuar a las explicaciones creadas por el servicio, estas son:

- **Causa - Efecto:** **a causa de** que en tu perfil tienes..., se te recomendó el ítem x
- **Efecto - Causa:** se recomendó el ítem x..., **debido a que** en tu perfil tienes...
- **Inferencial:** si en tu perfil tienes..., **entonces**, se te recomendó el ítem x

Esto nos proporciona una estructura y marcadores causales que utilizamos en la definición de las plantillas, para expresar la causa que originó una recomendación.

4.2.5.1. Selección de la estructura para las plantillas

El objetivo de las explicaciones es aumentar la confianza de los usuarios hacia el servicio, por lo que el tipo de relación causal **Efecto-Causa** definido en (Gálan, 2014), es el que mejor se adecua, es el que brinda una mayor claridad del texto explicativo. Aunado con el conectivo *porque* como relación causal, ya que en el lenguaje oral y escrito de los mexicanos es común emplearlo para explicar una causa o razón de alguna cosa.

Ejemplo

Efecto-Causa: Se recomendó el ítem x (**marcador**) **porque** lista de argumentos.

La lista de marcadores de relación causal más comúnmente usados son:

Al fin y al cabo/a la postre, después de todo, porque, es que y pues, debido a.

Personalización y definición de formas de redacción

Ya que se utilizan plantillas con campos de texto variables para crear las explicaciones, las cuales se llenan con las características de los ítems, que existen diferentes tipos de usuarios y que la orientación del sistema es hacia Instituciones de Educación Superior (IES); tomando en cuenta estos aspectos se empleó una forma de redacción personalizada hacia el usuario, donde se incluye su nombre y grado académico.

4.2.5.2. Personalización de los textos explicativos basados en la segmentación de mercado

Se consideraron diversos aspectos de la mercadotecnia como la personalización de los productos hacia el usuario y segmentación del mercado, con el fin de definir los elementos en el ámbito de mercadotecnia, los cuales se pueden considerar en las explicaciones, para una mayor personalización e impacto en el usuario.

La personalización se refiere al proceso de entregar contenido o servicios a un usuario, con base a sus preferencias, intereses y necesidades, con el propósito de adaptarlos a las

características específicas del usuario con el fin de lograr un óptimo desempeño (Tsetsos, 2009).

La segmentación de mercado es el proceso de dividir o segmentar el mercado total (heterogéneo) en grupos uniformes más pequeños que tengan características y necesidades semejantes, la segmentación de un mercado se puede dividir de acuerdo a sus características o variables que puedan influir en su comportamiento. Estos segmentos son grupos homogéneos. Debido a esta similitud dentro de cada grupo, es probable que respondan de modo similar a determinadas estrategias de mercadotecnia. Es decir, probablemente tendrán las mismas reacciones sobre un determinado producto. La segmentación sirve para determinar los rasgos básicos y generales que tendrá el consumidor del producto, teniendo en cuenta que el mismo no va dirigido para todo público, sino para el público objetivo identificado.

Es importante señalar que no existe una clasificación única y depende en gran medida del producto o servicio ofrecido. La clasificación de Sahui (Sahui Maldonado, 2008) es la siguiente:

- La **geográfica**, misma que utiliza la división del mercado en diferentes unidades geográficas (como naciones, regiones, estados, municipios, ciudades). En este tipo de variable, la empresa decide operar en una o en unas cuantas áreas geográficas, atendiendo a las necesidades y preferencias geográficas.
- La **demográfica**, que consiste en dividir el mercado en grupos tales como la edad, el sexo, el tamaño de la familia, el ciclo de vida familiar, la ocupación, la religión, la raza y la nacionalidad. Esta variable es la más utilizada por la mercadotecnia.
- La **psicográfica**, ésta divide a los consumidores en diferentes grupos con base en la clase social, el estilo de vida o las características de la personalidad.
- La **conductual**, trata de dividir a los compradores de acuerdo a sus conocimientos, actitudes y uso o respuesta a un producto.

Para la personalización de las explicaciones se tomaron como características las variables demográficas que afectan al consumidor, tales como: nivel de estudios, realizando una segmentación demográfica de acuerdo al tipo de usuario del sistema.

Ya que no hay una forma única que permita segmentar el mercado, la mercadotecnia tiene que utilizar diferentes variables de segmentación, para hacer llegar sus productos a los consumidores. La segmentación demográfica es la más utilizada ya que divide el mercado en grupos tales como: la edad, el sexo, el tamaño de la familia, el ciclo de vida familiar, la ocupación, la religión, la raza y la nacionalidad.

Debido a los diferentes tipos de usuario del sistema podemos realizar la siguiente segmentación por la variable ocupación:

Profesor
Estudiante
Egresado
Empresa

Para estos segmentos se definieron reglas las cuales pueden extenderse tanto como lo permita la información definida en el perfil del usuario:

Reglas para personalizar las plantillas con el nombre y grado académico de los usuarios Si existe un usuario (X) y tiene-nombre (X, Y) entonces utilizar el nombre del usuario en la plantilla.

Si existe un usuario (X) y tiene-nombre (X, Y) y tiene-ocupación ($X, \text{"Profesor"}$) y tiene-grado de estudio (X, Z) entonces obtener el grado de estudios " Z " y nombre del usuario " Y " y emplear reglas de género para personalizar la plantilla.

Si existe un usuario (X) y tiene-nombre (X, Y) y tiene-ocupación ($X, \text{"Empresa"}$) entonces obtener nombre del usuario " Y " y utilizar una forma de redacción formal tratar al usuario de usted para personalizar la plantilla.

Si existe un usuario (X) y tiene-nombre (X, Y) y tiene-ocupación ($X, \text{"Egresado"}$) y tiene grado de estudio (X, Z) entonces obtener el grado de estudios " Z " y nombre del usuario " Y " y emplear reglas de género para personalizar la plantilla.

Si existe un usuario (X) y tiene-nombre (X, Y) y tiene-ocupación ($X, \text{"Egresado"}$) y tiene grado de estudio (X, falso) entonces obtener nombre del usuario " Y " y complementar con reglas de edad, para personalizar la plantilla.

Si existe un usuario (X) y tiene-nombre (X, Y) y tiene-ocupación ($X, \text{"Estudiante"}$) y tiene-grado de estudio (X, Z) entonces obtener nombre del usuario " Y " emplear reglas de edad, si edad es ≥ 21 y ≤ 30 solo emplear el nombre del usuario sin apellido para personalizar la plantilla.

Si existe un usuario (X) y tiene-nombre (X, Y) y tiene-ocupación ($X, \text{"Estudiante"}$) y tiene-grado de estudio (X, Z) entonces obtener nombre del usuario " Y " y emplear reglas de edad, si edad es > 30 entonces obtener el grado de estudios " Z " y emplear reglas de género y utilizar el nombre y grado para personalizar la plantilla.

Por Edad

18, 18-30, Mayor de 30

Si existe un usuario (X) y tiene-edad ($X, \geq 21$) y tiene-edad ($X, \leq 30$) entonces utilizar una forma de redacción directa en segunda persona (Tu).

Si existe un usuario (X) y tiene-edad ($X, > 30$) entonces utilizar una forma de redacción formal tratar al usuario de Usted.

Por Género

Masculino, Femenino

Si existe un usuario (X) y tiene-género ($X, \text{Femenino}$) entonces utilizar los títulos adecuados a una mujer; Lic., Dra., etc.

Si existe un usuario (X) y tiene-género (X , Masculino) entonces utilizar los títulos adecuados a una mujer; Lic, Dr., etc.

4.2.5.3. Recuperación de información e Identificación de campos variables

Para mostrar las explicaciones al usuario se utilizan plantillas con campos de texto variables, las cuales son llenadas a partir de los recursos utilizados por el motor de recomendación, la información que utiliza el servicio de recomendación para realizar una recomendación específica.

Para la creación de las plantillas explicativas con campos variables es necesario considerar aspectos como:

- La creación de plantillas es de acuerdo al dominio y a la técnica utilizada por el sistema de recomendación.
- Las plantillas se crean por cada tipo de ítem recomendado.
- Es necesario contar con la definición del perfil del usuario y la definición de las propiedades o características de cada ítem.
- Identificar la información a considerar en las explicaciones, de acuerdo a los recursos utilizados por el servicio de recomendación para realizar cada recomendación.
- Analizar la cantidad y tipo de información necesaria para cada plantilla, descartando la que no resulte relevante para el usuario.

Los tipos de ítems recomendados por el servicio son: académicos, objeto de conocimientos, instalaciones especializadas, infraestructura especializada, certificación, cursos, servicios y proyectos de investigación. Las plantillas son definidas de acuerdo al dominio de IES y para cada tipo de ítem recomendado.

La creación de plantillas se realizó para cada ítem y para cada técnica utilizada: basada en contenido y filtrado colaborativo. A continuación se presenta los tipos de plantilla para los diferentes tipos de ítems.

Definiciones:

- [Campo variable]
- (Selección de texto fijo)

Cada plantilla inicia de la siguiente forma:

- [grado_academico]/ [título] [nombre_usuario] se (le/te) ha recomendado (este/esta)...

La información que se utiliza para la personalización de la explicación, se extrae del perfil del usuario, características como:

[grado_academico]: licenciatura, maestría, doctorado, etc. Obtenido de grado académico.

[título]: Ciudadano abreviado como C .

[nombre_usuario]: nombre del usuario.

4.2.5.3.1. Explicaciones basadas en colaboración

Para las explicaciones basadas en filtrado colaborativo se emplean las siguientes plantillas, las cuales son para cualquier tipo de ítem recomendado por el sistema. Se presentan dos tipos.

Basada en ítems:

[grado_academico]/ [título] [nombre_usuario] se (le/te) ha recomendado [este/esta] [tipo_item], [marcador_linguistico] **usuarios** quienes [acción] [el/la] [ítem_visitado], también [acción] [el/la] [ítem_recomendado]

Basada en Usuarios:

[grado_academico]/ [título] [nombre_usuario] se (le/te) ha recomendado [este/esta] [tipo_item], [marcador_linguistico] **usuarios similares** a [usted/ti] en [lista de características similares], le han asignado una calificación alta.

4.2.5.3.2. Explicaciones Basadas en contenido

En el caso de las explicaciones basadas en contenido se requiere una mayor personalización, considerando las características individuales de cada ítem recomendado.

Para explicación basada en contenido obtener:

Propiedades: ítem Persona

- **[nombre_persona]**: nombre de la persona recomendada.
- **[tipo_persona]**: tipo de persona recomendada (Profesor-Investigador, Estudiante, Egresado).
- **[vec_dominio_conocimiento]**: vector de palabras contenidas en la descripción de las líneas de investigación o tópicos de interés de la persona recomendada; que coinciden con el perfil del usuario.
- **[certificación]**: Certificaciones con las que cuenta la persona recomendada.
- **[experiencia_laboral]**: Experiencia laboral de la persona recomendada.
- **[estancia]**: Estancias con las que cuenta un alumno (Residencia Profesional, Estancia Científica, Estancia Tecnológica).
- **[reseña]**: reseña de la persona.
- **[línea de Investigación]**: línea de investigación de la persona.

Personas: [grado_academico] [nombre_persona] se (le/te) ha recomendado a (este/esta) [tipo_persona], [marcador_lingüístico]

- posee conocimientos sobre: **[vec_dominio_conocimiento]**; lo cual podría ser de interés para (usted/ti).
- posee conocimientos sobre **[vec_dominio_conocimiento]**, cuenta con **[certificación]**, | ha realizado estancia en: **[estancia]**.

Propiedades: objeto de conocimiento

Para explicación basada en contenido obtener:

- **[nombre_objeto_conocimiento]**: nombre del objeto de conocimiento recomendado.
- **[tipo_oc]**: tipo de objeto de conocimiento (artículo, revista, tesis, entre otros).
- **[vec_dominio_conocimiento]**: vector de palabras contenidas en las áreas de conocimiento o líneas de investigación al cual pertenece el objeto de conocimiento; que coinciden con el perfil del usuario.
- **[vec_palabra_clave]**: palabras clave contenidas en el objeto de conocimiento.
- **[descripcion_objeto_conocimiento]**: resumen del objeto de conocimiento.
- **[archivo del objeto de conocimiento]**: es el archivo adjunto al objeto de conocimiento guardado en formato pdf.

Objetos de conocimiento: [grado_Academico] [nombre_usuario] se (le/te) ha recomendado (este/esta) [tipo_oc], [marcador_linguistico]

- Corresponde con tópicos de: **[vec_dominio_conocimiento]** | y contiene los siguientes términos: **[vec_palabra_clave]**; los cuales podrían ser de (tu/su) interés.

Propiedades: Instalación especializada

Para explicación basada en contenido obtener:

- **[nombre_instalación]**: nombre de la instalación recomendado.
- **[tipo_instalación]**: laboratorio, biblioteca, taller, entre otros.
- **[descripción_instalación]**: resumen de la instalación.
- **[vec_dominio_conocimiento]**: vector de palabras contenidas en la instalación; que coinciden con el perfil del usuario.

Instalación: [grado_academico] [nombre_usuario] se (le/te) ha recomendado (el/la) **[tipo_Instalación]**, [marcador_linguistico]

- tiene asociado los siguientes términos (la siguiente característica/las siguientes características): **[vec_dominio_conocimiento]** (la cual/las cuales) podrían ser de (tu/su) interés.

Propiedades de la infraestructura tecnológica.

Para explicación basada en contenido obtener:

- **[nombre_infraestructura]**: nombre de la infraestructura tecnológica recomendado.
- **[tipo_infraestructura]**: tipo de infraestructura (equipo, maquinaria, software, entre otros).
- **[descripción_infraestructura]**: información descriptiva de la infraestructura.
- **[vec_dominio_conocimiento]**: vector de palabras contenidas en la descripción de la infraestructura; que coinciden con el perfil del usuario.

Infraestructura: [grado_academico] [nombre_usuario] se (le/te) ha recomendado [este/esta] [**tipo_infraestructura**], [marcador_linguistico]

- posee características como: [**vec_dominio_conocimiento**] las cuales son de (tu/su) interés;

Propiedades de servicios tecnológicos

Para explicación basada en contenido obtener:

- [**nombre_servicio**]: nombre del servicio recomendado.
- [**vec_dominio_conocimiento**]: vector de palabras contenidas en la descripción del servicio; que coinciden con el perfil del usuario.
- [**descripción_servicio**]: resumen del servicio.
- [**tipo_servicio**]: tipo de servicio recomendado (conferencia, diseño, consultoría, análisis, entre otros).
- [**archivo del servicio**]: es el archivo adjunto al servicio guardado en formato pdf.

Servicios: [grado_Academico] [nombre_usuario] se (le/te) ha recomendado el [tipo_servicio], [marcador_linguistico]

- posee (los siguientes términos/ el siguiente término) en su descripción: [**vec_descripcion_servicio**] (el cual/los cuales) podrían ser de (tu/su) interés.

Propiedades de certificación

Para explicación basada en contenido obtener:

- [**nombre_certificación**]: nombre de certificación recomendado.
- [**vec_dominio_conocimiento**]: vector de palabras contenidas en la descripción de certificación; que coinciden con el perfil del usuario.
- [**descripción_certificación**]: resumen de la certificación.
- [**tipo_certificación**]: tipo de certificación recomendado.
- [**archivo del servicio**]: es el archivo adjunto al servicio guardado en formato pdf.

Certificación: [grado_Academico] [nombre_usuario] se (le/te) ha recomendado [tipo_certificación], [marcador_linguistico]

- posee (los siguientes términos/ el siguiente término) en su descripción: [**vec_dominio_conocimiento**] (el cual/los cuales) son de (tu/su) interés.
- tiene asociado los siguientes términos (la siguiente característica/las siguientes características): [**vec_dominio_conocimiento**] (la cual/las cuales) podrían ser de (tu/su) interés.

Propiedades de curso

Para explicación basada en contenido obtener:

- **[nombre_curso]:** nombre de curso recomendado.
- **[vec_dominio_conocimiento]:** vector de palabras contenidas en la descripción de curso; que coinciden con el perfil del usuario.
- **[descripción_curso]:** resumen del curso.
- **[tipo_curso]:** tipo de curso recomendado (curso, taller, entre otros).
- **[archivo del servicio]:** es el archivo adjunto al servicio guardado en formato pdf.

Curso: [grado_academico] [nombre_usuario] se (le/te) ha recomendado el [tipo_curso]: [nombre_curso], [marcador_linguistico]

- posee (los siguientes términos/ el siguiente término) en su descripción: **[vec_dominio_conocimiento]** (el cual/los cuales) son de (tu/su) interés.
- tiene asociado los siguientes términos (la siguiente característica/las siguientes características): **[vec_dominio_conocimiento]** (la cual/las cuales) podrían ser de (tu/su) interés.

Propiedades de los proyectos de investigación

Para explicación basada en contenido obtener:

- **[nombre_proyecto]:** nombre del proyecto recomendado.
- **[tipo_proyecto]:** tipo del proyecto recomendado (investigación, desarrollo).
- **[vec_dominio_conocimiento]:** vector de palabras contenidas en la descripción del proyecto; que coinciden con el perfil del usuario.
- **[vec_palabra_clave]:** vector de palabras clave contenidas en el objeto de conocimiento.
- **[descripción_proyecto]:** resumen del proyecto.
- **[archivo del servicio]:** es el archivo adjunto al servicio guardado en formato pdf.

Proyectos: [grado_Academico] [nombre_usuario] se (le/te) ha recomendado este proyecto, [marcador_linguistico]

- Corresponde con tópicos de: **[vec_dominio_conocimiento]** | y contiene los siguientes términos: **[vec_palabra_clave]**; los cuales podrían ser de (tu/su) interés.

4.2.6. Definición de la presentación de las explicaciones

En (Tintarev, 2011.) se menciona que la presentación de las recomendaciones y el modo de interacción con el usuario, son factores que pueden afectar los tipos de explicación, además la interfaz de un sistema de recomendación puede afectar la credibilidad del sistema (Gálan, 2014); las formas más comunes de presentar las recomendaciones son el top ítem, top *N-ítem*, predicción de valoraciones para todos los ítems y resumen estructurado.

Hay varias formas de presentar las explicaciones, las cuales en (Gálan, 2014) se clasifican en: explicaciones comprendidas por texto e imagen, solamente texto, tablas e histogramas,

y propone la presentación mediante mapas mentales para explicar recomendaciones más complejas.

Para mostrar las explicaciones se utilizó la forma de texto e imagen. El sistema sobre el cual se desarrolló esta investigación, recomienda un conjunto de ítems heterogéneos y genera explicaciones individuales para cada ítem, la explicación se personaliza de acuerdo al tipo de usuario como en el trabajo realizado en (Tintarev N. , 2009). Los resultados de este trabajo muestran que los usuarios que recibieron una explicación personalizada están más satisfechos que aquellos que no la recibieron, es decir, las explicaciones personalizadas aumentan la satisfacción del usuario. Además se realizaron experimentaciones para determinar la influencia que tiene el uso de imágenes en las explicaciones, lo que concluyo que las explicaciones sin imágenes podría dañar considerablemente la satisfacción del usuario, eliminar la imagen del ítem recomendado en una explicación, la cual al no contar con suficiente información, los usuarios se pierden y pueden optar por no calificar los ítems o abandonar la interfaz.

También se utilizó la forma de histograma para presentar explicaciones como en el trabajo realizado por (Herlocker J. L., 2000), con la variación de que en vez de poner solo los ratings de los vecinos, se muestra al usuario información global y resumida de las valoraciones que ha recibido el ítem; se muestra la valoración promedio así como el total de valoraciones agrupadas por el valor de la valoración (de 1 a 5 estrellas en este trabajo).

En (Gálan, 2014) se concluye que son necesarios tres elementos en las interfaces de explicación: una imagen o ícono que represente al ítem recomendado, el nombre del ítem y un texto explicativo, la Figura 4.5 muestra un ejemplo.

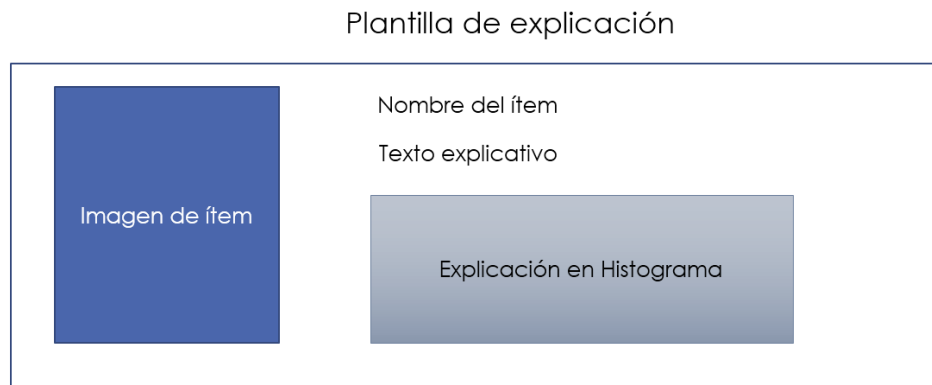


Figura 4.5 Plantilla de explicación.

4.2.6.1. Explicación mediante mapa mental

(Gálan, 2014) Propone la presentación mediante mapas mentales para explicar recomendaciones más complejas, la Figura 4.6 muestra un ejemplo.

Los mapas mentales son herramientas con un gran poder de utilidad, al utilizar un mapa mental el cerebro se concentra en una idea principal y realiza las asociaciones de la información de forma natural.

Se puede establecer como centro del mapa mental la “idea principal de recomendación” o incluso al “usuario” y en las ramificaciones mostrar los ítems recomendados por el sistema especificando información básica en las ramificaciones.

Una imagen se recuerda mejor y puede ser más sugestiva, precisa y poderosa que las palabras cuando se requiere realizar un gran número de asociaciones de información, lo que se conoce como “Efecto de superioridad de las imágenes” (Gálan, 2014).

Las explicaciones mediante mapas mentales pueden implementarse cuando no se desea sobrecargar al usuario con demasiada información textual, de esta forma la información y distribución de los ítems resultará más comprensible y amigable gráficamente;

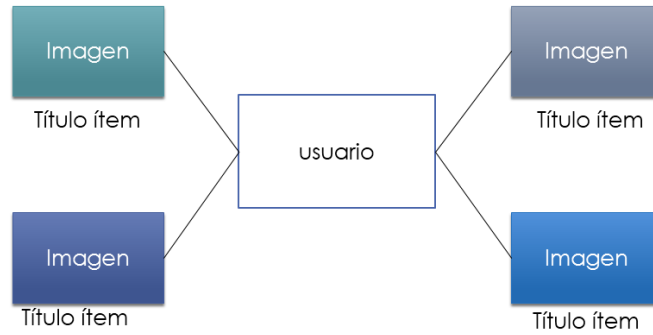


Figura 4.6 Explicación mapa mental.

CAPÍTULO 5
ARQUITECTURA PARA
EL SERVICIO DE
RECOMENDACIÓN

En este capítulo se presenta la arquitectura del servicio de recomendación para la plataforma de las capacidades científicas y tecnológicas del TecNM, tomando como base las características definidas en la metodología de solución. Así también se describe el diseño e implementación del servicio de recomendación.

5.1 Arquitectura del servicio de recomendación para la PCCT del TecNM.

A partir de la metodología se diseñó una arquitectura para el proceso de recomendación, dicha arquitectura se compone de varios módulos que realizan las tareas que integran el proceso de recomendación, creación de las explicaciones y de valoraciones.

La arquitectura del servicio de recomendación se muestra en la Figura 5.1, el cual se conforma de tres módulos principales y como salida se obtiene la recomendación en una página web (la interfaz de usuario).

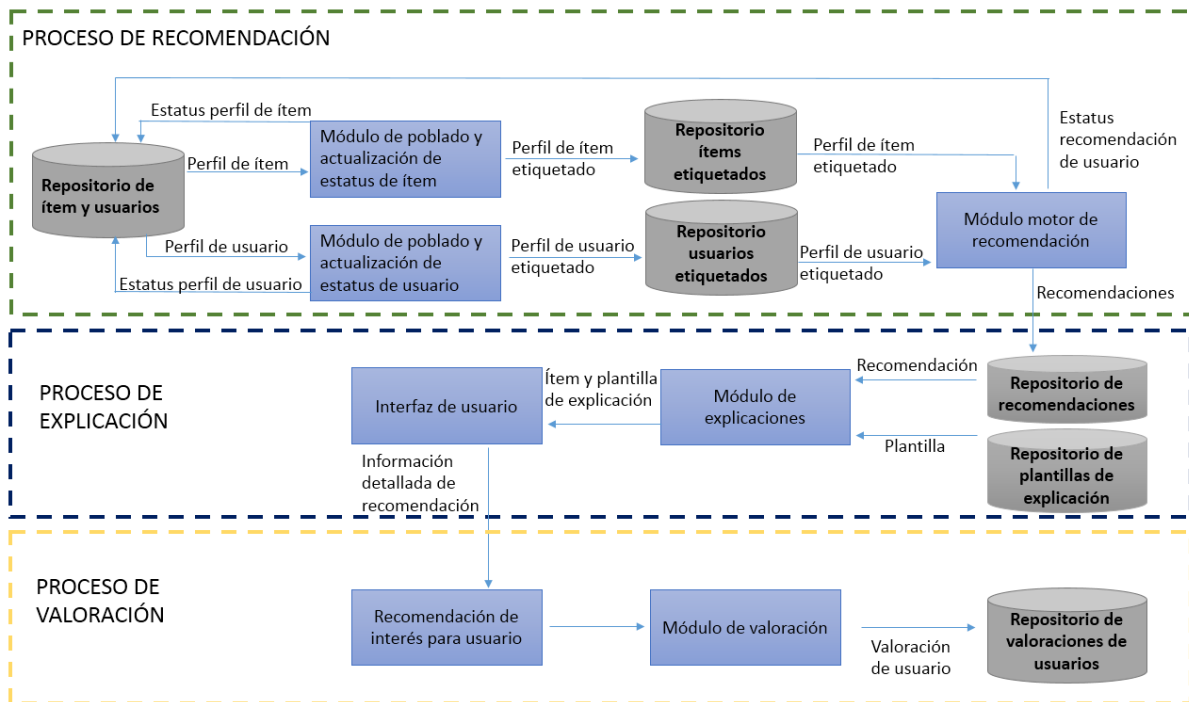


Figura 5.1 Arquitectura del servicio de recomendación.

A continuación se describen cada uno de los módulos que incluye la arquitectura del servicio de recomendación.

- **Módulo de poblado y actualización de perfil de usuario:** este módulo es el encargado de poblar y mantener actualizado el perfil de los usuarios, almacenados en el *repositorio de perfil de usuario*. La información del perfil de los usuarios se guarda en el disco del servidor en archivos de texto (extensión txt).
- **Módulo de poblado y actualización de perfil de ítem:** este módulo es el encargado de poblar y mantener actualizado el perfil de los ítems, almacenados en

el *repositorio de perfil de ítem*. La información del perfil de los ítems es guardada en el disco del servidor en archivos de texto.

- **Módulo de motor de recomendación:** este módulo es el encargado de inferir los ítems que se recomiendan a los usuarios, además de clasificar y ordenar las recomendaciones. Estas son agrupadas por tipo de ítem y ordenadas en forma de top-ítem tomando el peso como valor para ordenarlas. Para el proceso de inferencia este módulo extrae el perfil de usuario y el perfil de los ítems de sus respectivos repositorios; después de que el motor de recomendación pondera los ítems, extrae aquellos ítems que están por encima de un umbral preestablecido y finalmente almacenando los ítems seleccionados en el *repositorio de recomendaciones*.
- **Módulo de explicación:** este módulo crea una explicación para cada tipo de usuario y cada tipo de ítem recomendado. Para la creación de la explicación primero extrae del *repositorio de recomendaciones* los ítems inferidos por el *motor de recomendación* y también extrae del *repositorio de plantillas de explicación*, los cuales se utilizan juntos para formar la explicación que se presentará al usuario. El sistema cuenta con diferentes tipos de usuario, con el objetivo de tratar a cada usuario de manera única se utilizan técnicas discursivas, donde se toman en cuenta el tipo de usuario, el grado académico y el género; en base a estas características este módulo crea la explicación.
- **Interfaz de usuario:** este módulo permite adaptar la forma en que las recomendaciones son mostradas al usuario, dependiendo del tipo de usuario y el tipo de ítem. Muestra la explicación, el histograma de valoraciones, la información detallada del tipo de ítem y la interfaz para valorar el ítem.
- **Módulo de valoración:** este módulo permite implementar el mecanismo para identificar el grado de satisfacción del usuario con respecto a la recomendación recibida. El sistema solicita al usuario que evalúe la recomendación recibida, este módulo implementa un mecanismo de evaluación en escala Likert de 1 a 5, implementando una interfaz de imágenes en forma de estrellas, donde una estrella es el mínimo valor de satisfacción y cinco estrellas el máximo grado de satisfacción. Además este módulo almacena las valoraciones que los usuarios asignan a los ítems en el *repositorio de valoraciones de usuarios*, para hacer persistente las valoraciones de los usuarios, esta información servirá de entrada al algoritmo de recomendación de filtrado colaborativo.
- **Repositorio de ítem y usuarios:** la arquitectura propuesta contempla una base de datos que almacena los ítems e usuarios usados por el servicio de recomendación. La información contenida en este repositorio se utiliza para crear los perfiles de los distintos tipos de usuarios, los perfiles de los ítems que se crean con las características de cada tipo, los cuales utiliza el servicio para inferir las recomendaciones. Además actualiza el estatus tanto de los ítems como de los usuarios, ya que si un perfil de un usuario o la información de un ítem son modificadas o eliminadas, este repositorio guarda esta información, la cual le indica al servicio el estatus de estos.
- **Repositorio de usuarios etiquetados:** La información contenida en este repositorio es creada por el *módulo de poblado y actualización de estatus de usuario*, el cual es el encargado de extraer el perfil del usuario del *repositorio de ítem y usuario*, etiquetar el perfil y almacenar las etiquetas en un archivo de texto.

- **Repositorio de ítems etiquetados:** este repositorio es similar al repositorio anterior. La información contenida en este repositorio es creada por el *módulo de poblado y actualización de perfil de ítem*, el cual es el encargado de extraer los atributos del ítem del *repositorio de ítem y usuario*, etiquetar el perfil y almacenar las etiquetas en un archivo de texto.
- **Repositorio de recomendaciones:** este repositorio es una estructura de directorios de archivos. La información contenida en este repositorio es creada por el *motor de recomendación*, contiene las recomendaciones que se presentan a los diferentes usuarios, la información de las recomendaciones se almacena en archivos en formato JSON⁴.
- **Repositorio de plantillas de explicación:** este repositorio es una base de datos en formato de plantillas. La información contenida en este repositorio es adaptada de la metodología creada en la tesis de (Gálan, 2014), este repositorio contiene las plantillas de explicación que se presentan a los diferentes usuarios, la información de las plantillas de recomendaciones se almacenan en archivos en formato vm⁵.
- **Repositorio de valoraciones de usuarios:** este repositorio almacena las valoraciones que los usuarios asignan a los ítems.

Descripción de los principales módulos.

En este apartado se describe el análisis realizado para definir los módulos de la arquitectura implementada en el servicio de recomendación. Las funciones desarrolladas por estos módulos son la extracción del perfil del usuario y la extracción de las características de los ítems; el etiquetado del perfil del usuario y de los ítems; el proceso de inferencia, y la explicación e interacción del usuario con las recomendaciones.

5.1.1. Módulo de poblado y actualización de estatus de ítem

Este módulo se encarga de poblar el *repositorio de ítems etiquetados* y de modificar el estatus del perfil del ítem una vez que ha finalizado el procesamiento del ítem. Este módulo extrae el perfil del ítem del *repositorio* y guarda el perfil del ítem en una carpeta temporal. Este módulo extrae atributos del perfil de los ítems del repositorio, posteriormente ejecuta el programa *freeling*⁶, este programa es el encargado de etiquetar los archivos temporales que contienen el perfil de los ítems, además de almacenarlos en el *repositorio de ítems etiquetados*. Cuando el programa *freeling* termina de procesar todos los ítems, se eliminan los archivos de texto temporales y se actualiza el estatus del ítem a estado *procesado* en el *repositorio de ítem y usuarios*.

Los ítems cuyo estatus es *procesado* o *eliminado* ya no son procesados por este módulo, esto con el objetivo de procesar únicamente los ítems cuyo estatus sea modificado o sea un ítem nuevo. Esto reduce la carga de procesamiento y acelera el proceso que realiza el *servicio de recomendación*. En el caso de que no se encuentren nuevos ítems o si no se han realizado modificaciones el *servicio de recomendación* no procesa nuevas recomendaciones, y evita que cada vez que este servicio se ejecute procese nuevamente

⁴ <http://www.json.org/>

⁵ <http://velocity.apache.org/>

⁶ <http://nlp.lsi.upc.edu/freeling/>

todos los ítems, dado que obtendría las mismas recomendaciones que en anteriores ejecuciones. La Figura 5.2 muestra el diagrama entidad relación del estatus de los ítems.

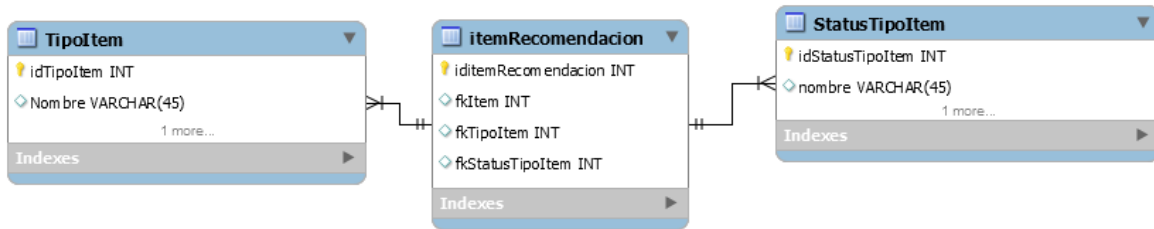


Figura 5.2 Modelo entidad relación del estatus de los ítems.

5.1.2. Módulo de poblado y actualización de estatus de usuario

Este módulo es el encargado de poblar el *repositorio de usuarios etiquetados* y de modificar el estatus del perfil del usuario una vez finalizado el procesamiento del repositorio. Este módulo extrae el perfil del usuario del *repositorio de ítem y usuarios*, guarda el perfil del usuario en una carpeta temporal. Cuando este módulo termina de extraer el perfil de todos los usuarios, ejecuta el programa *freeling*, este programa es el encargado de etiquetar los archivos temporales que contienen el perfil de los usuarios además de almacenarlos en el *repositorio de usuarios etiquetados*. Cuando el programa *freeling* termina de procesar todos los archivos, este módulo elimina los archivos de texto temporales y actualiza el estatus del usuario a *procesado* en el *repositorio de ítem y usuarios*.

De igual forma que ocurre con los ítems únicamente la primera vez que se ejecuta el servicio de recomendación procesando cada uno de los perfiles de usuario. Las ejecuciones siguientes dependerán del estatus del usuario, a continuación se describen cada uno de los estatus:

- Los usuarios tienen más estados que los ítems, cuando el estatus de un perfil de usuario es *modificado* el *módulo de poblado y actualización de estatus de usuario* nuevamente procesa el *nuevo* perfil de usuario.
- Si el estatus del perfil del usuario es *eliminado* este módulo borra el perfil del usuario del *repositorio de usuario etiquetado* y el estatus del perfil del usuario no se modifica por este módulo quedando el estatus en *eliminado*.
- Si el usuario ya ha sido procesado y recibe alguna recomendación, el estatus es *procesado con resultados*, con el objetivo de no sobrecargar el *servicio de recomendación* y periódicamente procesar el perfil del usuario, se define una variable que nos indica los días en que se volverá a procesar el perfil del usuario (por defecto esta variable se le asignó un valor de 15 días).
- Si el usuario ya ha sido procesado y no recibió recomendaciones el estatus es *procesado sin resultados*, en este caso el valor asignada a la variable para volver a procesar este usuario es de 7 días.

La Figura 5.3 muestra el diagrama entidad relación del estatus de los usuarios se.

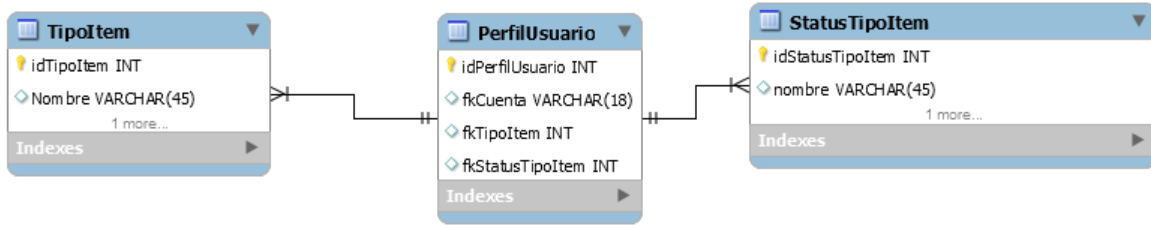


Figura 5.3 Modelo entidad relación del estatus de los usuarios.

5.1.3. Módulo motor de recomendación

Los procesos descritos en los módulos anteriores dan soporte al *motor de recomendación*: en primer lugar, mantiene los repositorios de ítem y el de usuario actualizados, busca nuevas recomendaciones cuando se agrega un nuevo usuario o el perfil del mismo es modificado, se deduce que el usuario tiene nuevos intereses y por lo tanto o elimina atributos que ya no le interesan o agrega nuevos atributos a su perfil que son los nuevos temas de interés para él.

Además cuando se agrega un nuevo ítem o cuando un ítem existente es actualizado, el *motor de recomendación* debe tomarlo en cuenta, para volver a ponderar estos ítems con los perfiles de los usuarios y si infiere nueva(s) recomendación(es) presentársela(s) a los usuarios. Si el ítem es eliminado debe ser removido de aquellas recomendaciones donde fue seleccionado.

El *motor de recomendación* toma como entrada el *perfil de ítem etiquetado* y *perfil del usuario etiquetado*, pondera el perfil del usuario con el de todos los ítems y extrae aquellos ítems que están por encima del umbral. Estos ítems son almacenados en el *repositorio de recomendaciones*.

Además este módulo se encarga de guardar el *estatus de la recomendación de usuario*, esta información indica si el usuario ya fue procesado por el *motor de recomendación*. Los estatus de las recomendaciones son:

- **Procesado con resultados:** este estatus indica que el *motor de recomendación* encontró algún ítem que recomendar al usuario.
- **Procesado sin resultados:** este estatus indica que el *motor de recomendación* no encontró ítems que recomendar al usuario.

Es importante indicar que el servicio de recomendación guarda un registro con la combinación de usuario e ítem, esto con el objetivo de tener información de las recomendaciones que recibe un usuario de los diferentes tipos de ítems; junto con el estatus se guarda la fecha en que esta información fue procesada, la fecha se almacena para conocer cuándo fue la última vez que este usuario recibió una recomendación de este tipo de ítem, para después de cierto lapso de tiempo volver a someter a este usuario al proceso de inferencia contra los tipos de ítems.

La información de estatus de las recomendaciones se modela en la Figura 5.4.

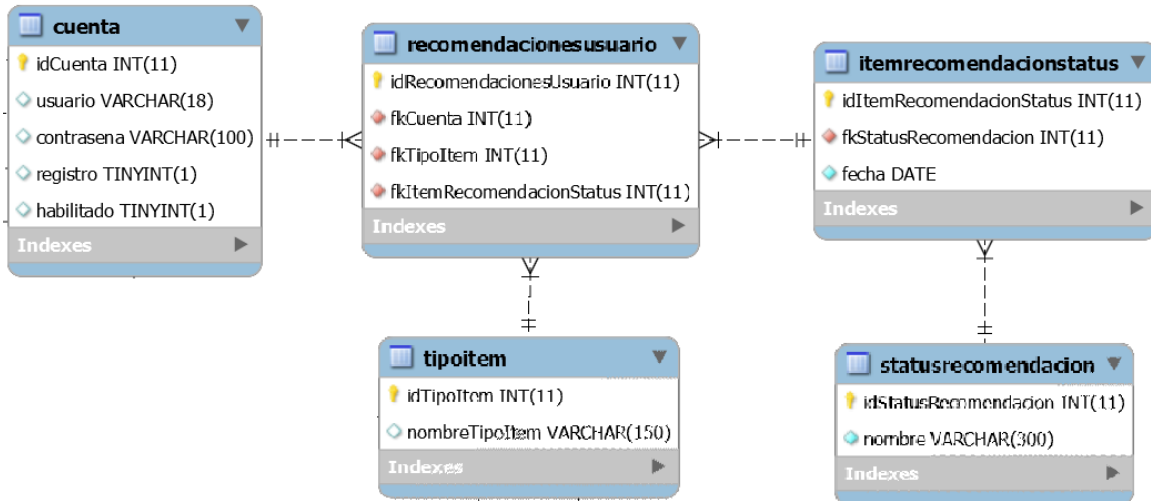


Figura 5.4 Modelo entidad relación de estatus de las recomendaciones.

5.1.4. Módulo de explicaciones

Este módulo es el encargado de crear la explicación que acompaña a cada recomendación. Las explicaciones se conforman de acuerdo al algoritmo descrito en la sección 4.2.2. Para generar las explicaciones, se considera implementar estilos de explicación según las técnicas de recomendación utilizadas por el sistema. Como el ítem fue recomendado a partir de la técnica basada en contenido, se recuperan las características presentes en el ítem que son comunes a las características presentes en el perfil del usuario.

Este módulo crea explicaciones individuales para cada tipo de ítem, toma las recomendaciones inferidas por el motor de recomendación, la plantilla de explicación adecuada, busca la información del tipo de ítem para agregar a la recomendación, recupera la información estadística que llena el histograma y crea la plantilla de explicación.

La Figura 5.5 muestra un ejemplo de la explicación en formato de texto y los campos de texto que se utilizaron para crearla.

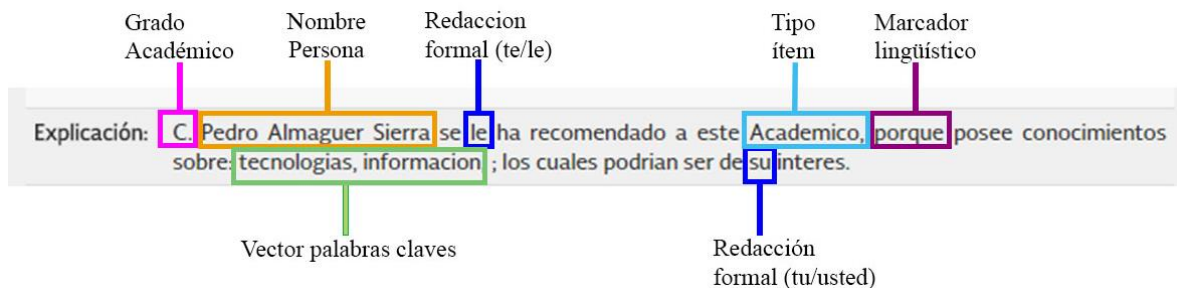


Figura 5.5 Ejemplo de explicación de texto.

La Figura 5.6 muestra un ejemplo de una recomendación y explicación como el usuario la visualiza en el navegador web.



Figura 5.6 Ejemplo de recomendación y explicación.

5.1.5. Interfaz de usuario

La interfaz de usuario es la encargada de la interacción con el usuario. Cuando el usuario ingresa en la aplicación web, el sistema solicita al servidor las recomendaciones creadas por el *módulo de explicación*. La interfaz muestra las recomendaciones, la información estadística del histograma y un enlace para mostrar información detallada del ítem recomendado. Cuando el usuario da click sobre este enlace la interfaz le muestra más información de la recomendación, un enlace para ver el archivo pdf (si está disponible), adicionalmente el usuario puede realizar una valoración de la recomendación recibida.

Un componente adicional en esta interfaz muestra en un mapa mental las recomendaciones que el usuario recibió en forma resumida.

5.1.6. Módulo de valoración

Este módulo es el encargado de guardar y actualizar los ratings que los usuarios asignan a los ítems. Estas valoraciones permiten identificar el grado de satisfacción de un usuario con respecto a una recomendación. Estos ratings son almacenados en el *repositorio de valoraciones de usuario*. Este módulo poblará el repositorio que servirá de entrada para el algoritmo de filtrado colaborativo. La Figura 5.7 muestra el diagrama entidad relación de las valoraciones de los usuarios.

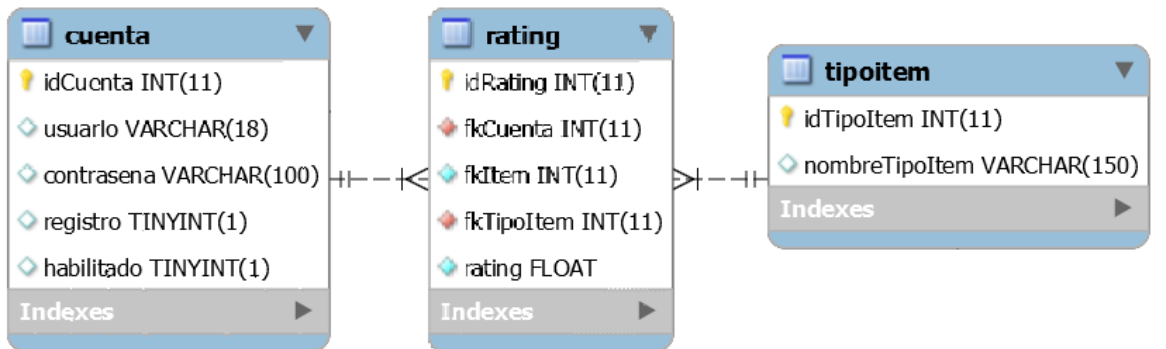


Figura 5.7 Modelo entidad relación de valoraciones de usuario.

5.2 Implementación

El prototipo de *servicio de recomendación* fue desarrollado usando los módulos de la arquitectura propuesta, con el fin de demostrar la efectividad de los conceptos, métodos y técnicas descritos en los apartados anteriores. Estos módulos se implementaron mediante el desarrollo de diversas aplicaciones que se distribuyen en dos partes principales, una aplicación servidor (backend) y otra aplicación web (frontend), como se muestra en la Figura 5.8.

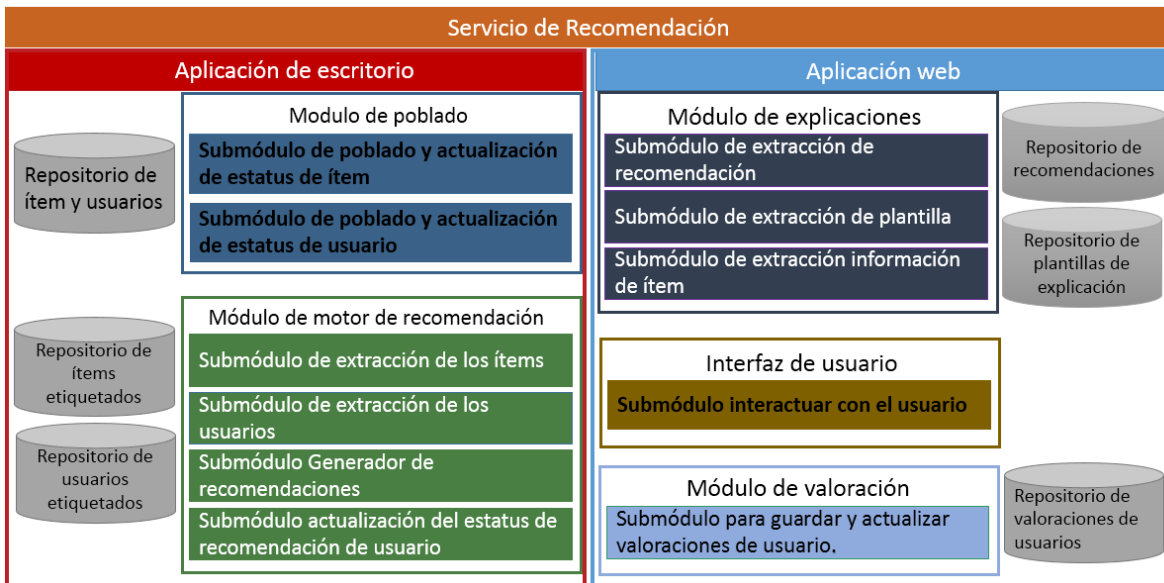


Figura 5.8 Vista de los módulos del servicio de recomendación.

5.2.1. Aplicación backend

Los módulos de la aplicación backend se desarrollaron utilizando el IDE NetBeans 8.0.2, soportado sobre Linux distribución Debian 7 y se describen a continuación:

5.2.1.1. Módulo de poblado

Es un conjunto de aplicaciones que se implementaron con el objetivo de poblar el *repositorio de ítems etiquetados* y *repositorio de usuarios etiquetados* para tener un conjunto de información sobre la cual aplicar el proceso de recomendación.

5.2.1.2. Módulo de motor de recomendación

Los métodos de este módulo están codificados en un proyecto de NetBeans llamado *motor de recomendación*. Estos métodos extraen el *perfil del usuario* y el *perfil de ítems* para generar las recomendaciones, las cuales se almacenan en archivos en formato Json en directorios locales. También guardan el estatus de las recomendaciones, esta información permite reducir el proceso para genera una recomendación, ya que únicamente se procesa periódicamente o cuando el perfil de usuario es modificado.

5.2.1.3. Repositorio de ítem y usuarios

Este repositorio está en la base de datos principal del sistema. La información almacenada en este repositorio contiene los *perfiles de los ítems* y los *perfiles de los diferentes usuarios* del sistema. Esta información es la que utiliza el *motor de recomendación*.

5.2.1.4. Repositorio de ítems etiquetados

Mediante una estructura de directorios en el servidor, los archivos que contienen el perfil de los ítems lematizados son almacenadas en archivos de texto.

5.2.1.5. Repositorio de usuarios etiquetados

Este repositorio es semejante al anterior, una estructura de directorios en el servidor. Se separan los ítems de los usuarios, para que el motor de recomendación procese todos los ítems del mismo tipo, por ejemplo todos los académicos.

5.2.1.6. Repositorio de recomendaciones

Este repositorio contiene las recomendaciones que el motor de recomendación ponderó para los usuarios. La información está organizada en directorio de archivos, a la carpeta raíz se le asigna el nombre del tipo de usuario (ejemplo profesor), y las subcarpetas son nombradas con los ids de la persona. La carpeta de cada usuario tiene un archivo Json por cada tipo de ítem del cual ha recibido una recomendación.

5.2.1.7. Repositorio de plantillas de explicación

Este repositorio es una base de datos en formato de plantillas. La cual contiene las plantillas de explicación que se le presentan a los diferentes usuarios, la información de las plantillas de recomendaciones son almacenadas en archivos en formato vm.

5.2.1.8. Repositorio de valoraciones de usuario

Este repositorio es el encargado de guardar las valoraciones que los usuarios asignan a las recomendaciones. Para almacenar esta información se crearon tres tablas en la base de datos. La Figura 5.7 muestra el diagrama entidad relación de las valoraciones de los usuarios.

Una vista dinámica de la arquitectura, considerando las aplicaciones desarrolladas durante la implementación del servicio de recomendación y cómo interactúan entre sí, se presenta en la Figura 5.9.

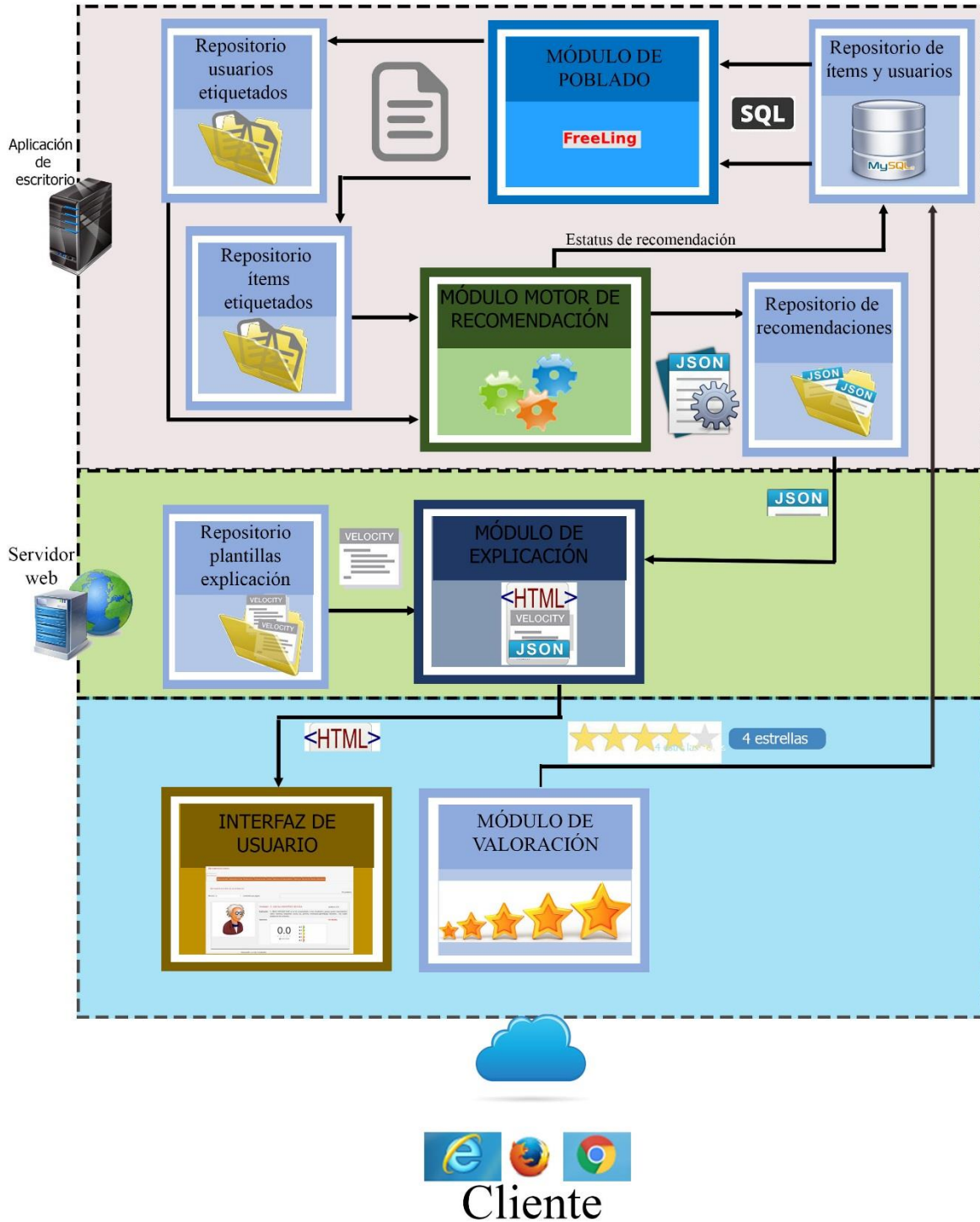


Figura 5.9 Vista funcional de los elementos del Servicio de Recomendación.

5.2.2. Implementación del motor de recomendación

5.2.2.1. Poblado del Repositorio de ítems y usuarios

Para implementar el *servicio de recomendación* se utilizó la base de datos de la Plataforma de las Capacidades Científicas y Tecnológicas (PCCT), se seleccionó y adaptó toda la información considerada como relevante para el proceso de recomendación. Las modificaciones hechas a la base de datos incluyeron la agregación de nuevas tablas y la creación de disparadores (triggers). Se agregaron un total de 9 tablas y 27 disparadores. Estas adaptaciones permiten que el *servicio de recomendación* procese las recomendaciones e identifique el estatus de las mismas.

La primera etapa del poblado de este repositorio se realizó con la colaboración de los 30 Institutos Tecnológicos, este poblado se realizó con el apoyo de una aplicación de escritorio para cargar la información de manera masiva. En esta etapa se dieron de alta un total de 4788 académicos, 525 instalaciones especializadas, 550 infraestructura especializada, 388 objetos de conocimiento, 7 certificaciones, 172 cursos, 195 servicios tecnológicos, 6 proyectos, 58025 alumnos y 7955 egresados.

5.2.2.2. Actualización de los ítems y usuarios

Dentro de las actividades realizadas por el *módulo de poblado* está el monitorear el estatus de los ítems y de los usuarios. Para llevar a cabo esta tarea se implementó en la base de datos un conjunto de *disparadores*, que se activan cada vez que se modifica el perfil de un ítem, de un usuario o se elimina cualquiera de estos. Los disparadores modifican la tabla que contiene la información de estatus del ítem y es la que consulta el *módulo de poblado* para conocer el estatus de los mismos. La Figura 5.10 muestra el trigger que se dispara cuando se modifica el ítem proyecto de investigación.

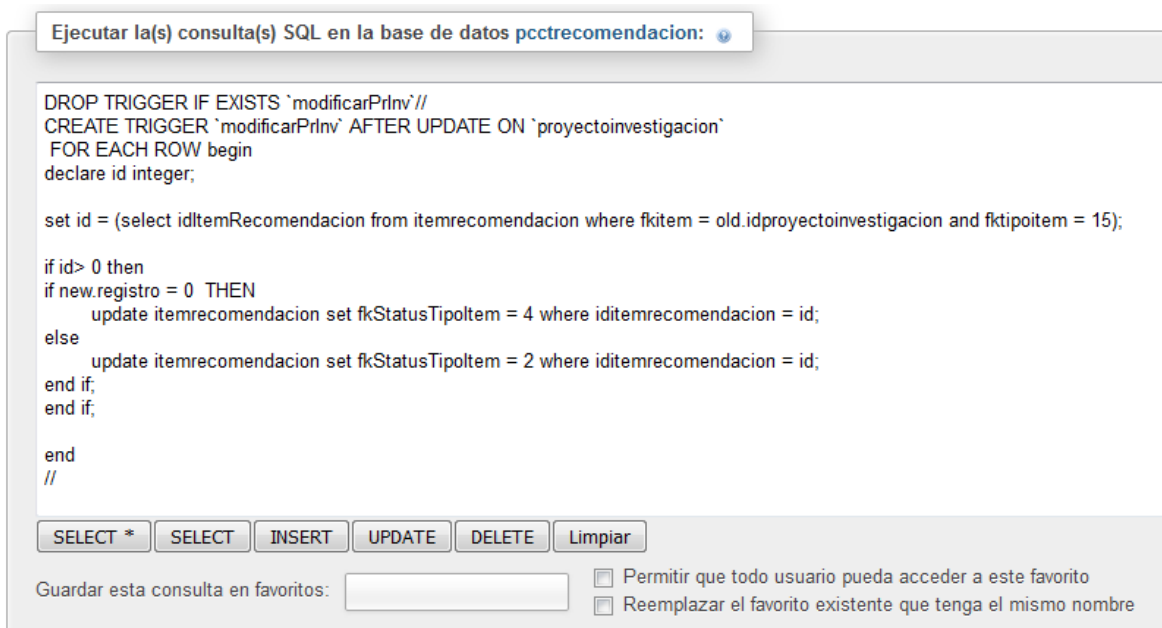


Figura 5.10 Trigger para proyecto de investigación.

5.2.2.3. Poblado de los repositorios de *ítems etiquetados* y *usuarios etiquetados*

Para poblar el *repositorio de ítems etiquetados*, en la primera etapa del proceso se toma la información que describe cada ítem, esta información es única, por lo que cada tipo de ítem es almacenado en una carpeta, agrupando los ítems por tipo. Para poblar el *repositorio de usuarios etiquetados*, primero se toma la información que describe a cada usuario su información es única, por lo que cada tipo de usuario es almacenado en una carpeta diferente, agrupando los usuarios por tipo. La información es extraída de la base de datos de la PCCT llamada *repositorio de ítems y usuarios*, esta información se extrae utilizando la aplicación de escritorio *motor de recomendación*, se utiliza el *módulo de poblado* el cual emplea el framework hibernate para extraer la información y después se guarda en archivos de texto en una carpeta temporal que sirve de entrada para la herramienta *freeling*.

En la segunda etapa, el *módulo de poblado* ejecuta la herramienta *freeling*, que es la encargada de etiquetar todos los archivos de texto que contienen los perfiles de los usuarios y los perfiles de los ítems, el resultado de este proceso se almacena en el *repositorio de ítems etiquetados* y *repositorio usuarios etiquetados* según corresponda. Después de que el módulo de poblado termina de ejecutar *freeling*, si el proceso fue exitoso, se elimina el archivo de texto contenido en la carpeta temporal y se actualiza el estatus del ítem o usuario en la base de datos, con el objetivo de procesar sólo cuando la bandera de estatus es nuevo o modificado. La Figura 5.11 muestra este procedimiento.

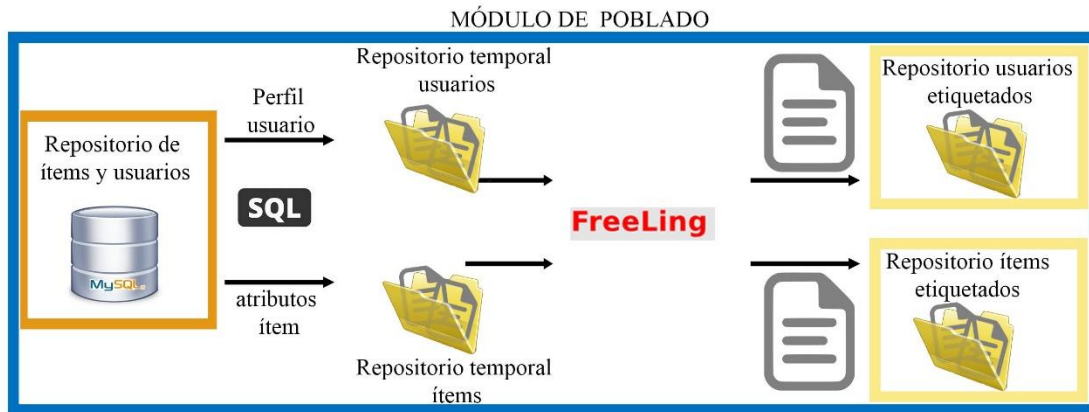


Figura 5.11 Módulo de poblado.

5.2.2.4. Implementación del Motor de recomendación

El proceso de recomendación implementado está dividido en sub-procesos, ejecutados por diversos submódulo del *motor de recomendación* de acuerdo a las etapas del proceso de inferencia presentadas en la Figura 5.12.

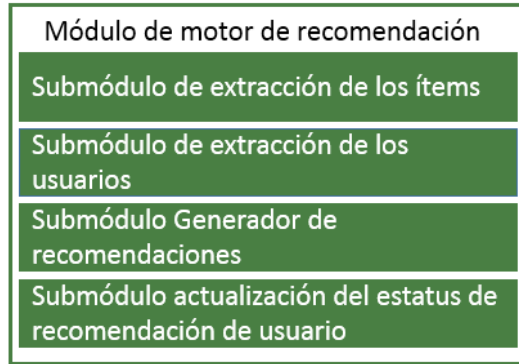


Figura 5.12 Módulos del motor de recomendación.

La aplicación backend es la encargada de realizar todo el proceso inmerso en la recomendación. Esta aplicación está programada para ejecutarse cada cierto tiempo, por defecto una vez por día, esto se realiza con la ayuda del administrador de procesos del sistema operativo llamado *cron*.

Como primer paso se ejecuta el *submódulo de extracción de ítems*, el cual es el encargado de extraer todos los ítems del *repositorio de ítems etiquetados*, esta información es almacenada en la memoria RAM y se envía al *módulo generador de recomendaciones*. En el segundo paso se ejecuta el *submódulo de extracción de usuarios*, este es responsable de extraer todos los usuarios del *repositorio de usuario etiquetado*, esta información también se almacena en la memoria RAM y se pasa al *submódulo generador de recomendaciones*.

En el tercer paso, una vez que el *submódulo generador de recomendaciones* recibió los ítems y los usuarios, inicia el proceso de ponderación. Este se describe a continuación:

1. Se aplica el algoritmo TF-IDF descrito en la sección 4.2.1.6
2. Una vez obtenidos todos los valores IDF, se asignan estos valores en los perfiles de los ítems y el perfil de usuario, en este paso el *motor de recomendación* convierte los perfiles en vectores de datos numéricos, que son los datos de entrada del proceso de similitud coseno.
3. En el proceso de similitud se utiliza la información calculada en el paso 2, se pondera cada perfil de usuario contra todos los ítems. Para calcular la similitud que hay entre un usuario y un ítem se utiliza la medida coseno descrita en la sección 4.2.1.6.
4. Cuando finaliza el proceso de similitud, el *motor de recomendación* extrae del archivo *pcct.properties* el umbral mínimo para filtrar los ítems que están por encima de este valor. Estos son los ítems recomendables que el algoritmo filtro para ser recomendados al usuario.

Una vez que se tiene la lista de ítems recomendados, el *motor de recomendación* recupera la información del ítem de la base de datos. Esta información y el valor de similitud es almacenada en el archivo Json en el *repositorio de recomendaciones*. Este submódulo trabaja en paralelo con el *submódulo actualización del estatus de recomendación*. Este submódulo determina si el usuario recibió o no una recomendación. En caso de ser afirmativo se guarda el estatus de las recomendaciones con el valor de *procesado con*

resultados, y de ser negativo se guarda con el valor de *procesado sin resultados*. La Figura 5.13 muestra el funcionamiento del módulo de recomendación.

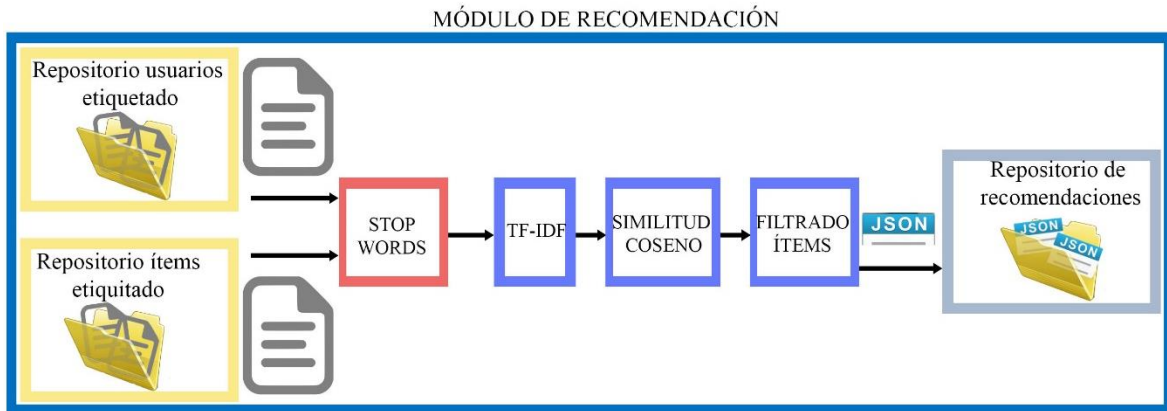


Figura 5.13 Módulo de recomendación

Implementación de la aplicación web

El servidor web corre sobre el sistema operativo Linux (Debian 7), la aplicación fue desarrollada utilizando el IDE Spring Tool Suite. El usuario utiliza el navegador web para conectarse al servidor, después de que el usuario inicia sesión, la aplicación cliente solicita al servidor las recomendaciones, las cuales se encuentran en un archivo Json. El servidor envía las recomendaciones y las plantillas de explicación al proceso cliente, que es el encargado de mostrar las recomendaciones al usuario. Además desde la aplicación cliente se envían al servidor las valoraciones que el usuario asigna a las recomendaciones recibidas.

La aplicación cliente permite administrar la información que será considerada durante el proceso de recomendación. Esta información es almacenada en la base de datos en el servidor.

5.2.3. Módulo de explicación

El proceso implementado en el servicio de recomendación para crear las explicaciones está dividido en sub-procesos, ejecutados por diversos submódulo presentadas en la Figura 5.14.

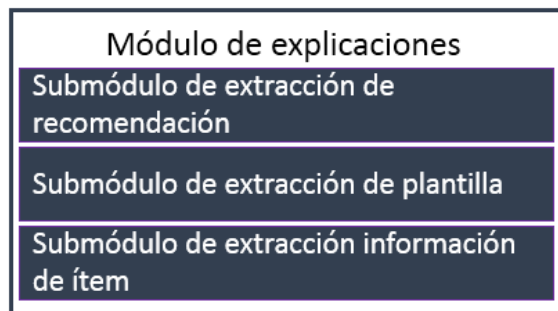


Figura 5.14 Módulo de explicaciones.

El *submódulo de extracción de recomendación* es el encargado de buscar el archivo Json del *repositorio de recomendaciones*, si no existe el archivo termina el proceso ya que no

hay recomendaciones para el usuario y se envía un mensaje que indica que no encontraron recomendaciones. Si encuentra el archivo se inicia el *submódulo de extracción de plantilla* el cual extrae la plantilla de recomendación específica para el tipo de usuario y el tipo de ítem; una vez seleccionada la plantilla se inicia el *submódulo de extracción de información* del ítem que es responsable de buscar la información estadística del ítem para llenar el histograma. Una vez que se recopila toda la información extraída por los tres módulos anteriores se llena la plantilla de explicación y se envía a la aplicación web en formato html.

La Figura 5.15 muestra la interfaz de usuario y la explicación del ítem profesor.

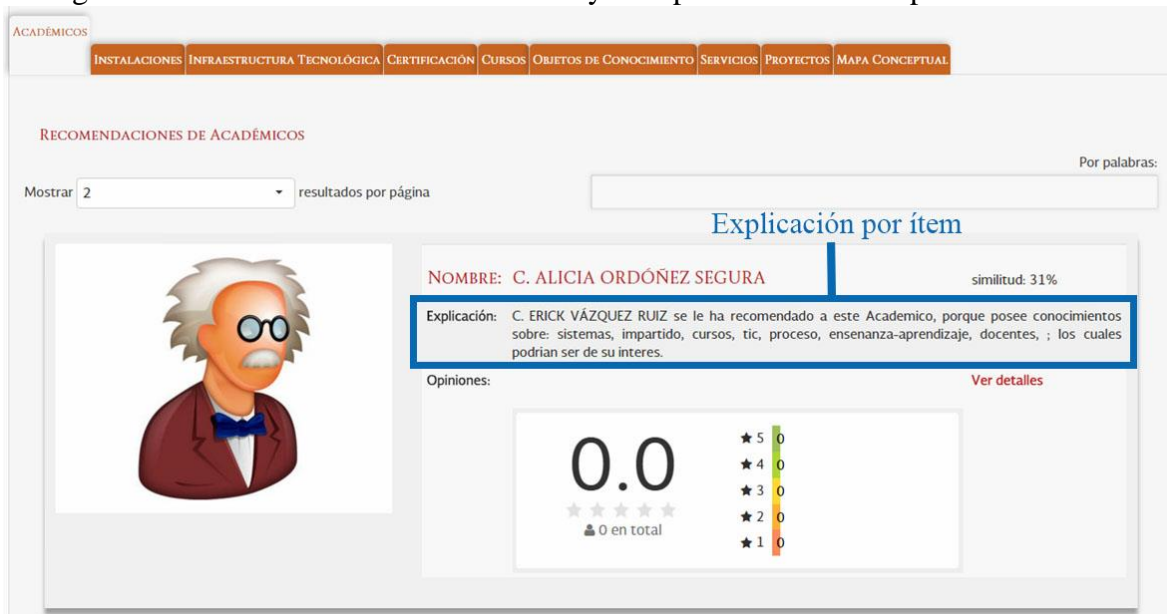


Figura 5.15 Recomendación ítem académico.

5.2.4. Interfaz de usuario

Es utilizada para visualizar las recomendaciones inferidas por el motor de recomendación. El usuario inicia el proceso al iniciar la sesión en la aplicación y éste solicita las recomendaciones al servidor que implementa el módulo para interactuar con el usuario, el cual envía las recomendaciones solicitadas por el usuario, la interfaz web muestra las recomendaciones por tipo de ítem, la Figura 5.17 muestra un ejemplo. Esta interfaz permite al usuario interactuar con las recomendaciones ya que el usuario solicita la recomendación por tipo de ítem. La Figura 5.16 muestra el submódulo utilizado por la interfaz web.

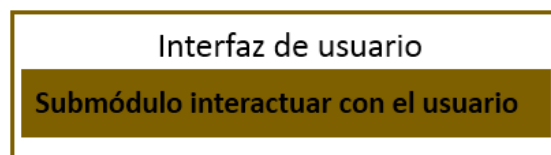


Figura 5.16 Módulo de interfaz de usuario.



Figura 5.17 Recomendaciones agrupadas por tipo de ítem.

5.2.5. Módulo de valoración.

Este módulo es el encargado de administrar las valoraciones que el usuario asigna las recomendaciones. Guarda y actualiza las recomendaciones que el usuario valora. La Figura 5.18 muestra el módulo de valoración. La Figura 5.19 muestra como la interfaz solicita al usuario que valore la recomendación.

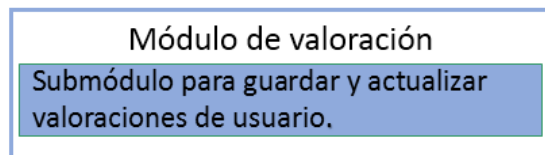


Figura 5.18 Módulo de valoración



Figura 5.19 Valoración de recomendación.

Las explicaciones en mapa mental muestra todas las recomendaciones que recibe un usuario se muestra en la Figura 5.20.

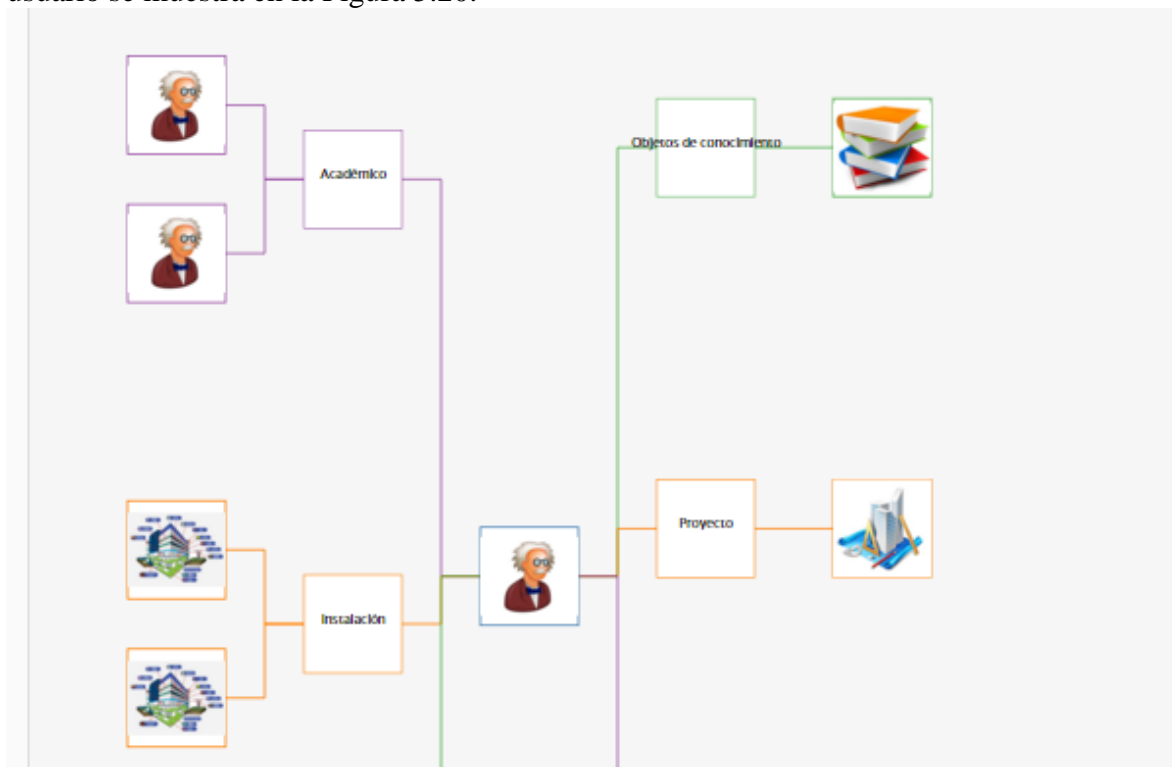


Figura 5.20 Explicación en mapa mental.

CAPÍTULO 6
PRUEBAS Y RESULTADOS

6.1 Introducción

Este capítulo expone las pruebas para verificar la funcionalidad del servicio de recomendación, centrándose en el proceso ejecutado para generar las recomendaciones y la forma en que éstas se le presentan al usuario. Las pruebas se realizan en el motor de recomendación del lado del servidor y en la aplicación web del lado del cliente.

Los sistemas de recomendación utilizan las opiniones de una comunidad de usuarios para ayudar a las personas en esa comunidad a identificar con mayor eficacia, el contenido de interés de un conjunto potencialmente abrumador de opciones (Herlocker, 2004).

Las pruebas realizadas al servicio ayudaran a detectar los puntos fuertes y débiles del funcionamiento del sistema, diagnosticar las causas de las fallas que permitan optimizar los resultados del sistema.

Para llevar a cabo la evaluación del sistema de recomendación se siguieron una serie de pasos propuestos por (Herlocker, 2004), los cuales se detallan en las secciones siguientes.

6.2 I. Tareas del usuario para el sistema de recomendación.

Para evaluar un sistema de recomendación es importante entender los objetivos y tareas para los que se está utilizando. Las tareas del usuario pueden ser Anotación en contexto *Annotation in Context* o Encontrar Buenos ítems *Find Good Items*.

6.2.1. Anotación en contexto

Esta tarea requiere retener el orden y el contexto de las recomendaciones, y utiliza predicciones para anotar las recomendaciones en contexto. Los usuarios utilizan las predicciones mostradas para decidir qué recomendación leer (o que link seguir), y por lo tanto el factor más importante a evaluar es qué tan exitosamente ayudan las predicciones a los usuarios a distinguir entre el contenido deseado y no deseado.

6.2.2. Encontrar Buenos ítems

Sugieren ítems específicos a los usuarios, proporcionando una lista clasificada de los ítems recomendados, junto con las predicciones un valor que indica que tanto les agrada a los usuarios. El usuario puede seleccionar uno o más de los ítems sugeridos. En este caso sólo se puede presentar una parte del conjunto de ítems inferidos. Por lo tanto, es probable que muchos ítems preferidos no se encuentren en la lista. En esta tarea, es más importante que no se presente ningún ítem que no sea de su agrado, que encontrar todos los buenos ítems.

6.3 II. Dataset para evaluación.

El dataset es un aspecto importante en la evaluación de un sistema de recomendación. (Herlocker, 2004) plantea las siguientes preguntas ¿Puede la evaluación llevarse a cabo fuera de línea en un conjunto de datos existente o requiere pruebas de usuario en vivo?, ¿Si un conjunto de datos no está disponible actualmente, puede llevarse a cabo la evaluación con datos simulados?, ¿Qué propiedades debe tener el conjunto de datos para modelar mejor las tareas para las que se está evaluando el algoritmo de recomendación? , las cuales

ayudan a determinar que dataset es el más adecuado de acuerdo a la métrica que se utilizara para evaluar el algoritmo de recomendación y los objetivos del sistema de recomendación.

Tomando en cuenta que el dataset con el que se cuenta para hacer la evaluación del algoritmo de recomendación es de un dominio en el cual no existe un dataset público, el cual contenga el conjunto de ítems heterogéneos y que el idioma de los ítems es español. Para hacer uso de este dataset tomamos como base lo definido en (Herlocker, 2004), menciona que cuando se evalúa un algoritmo de recomendación en un nuevo dominio donde hay una investigación significativa sobre las preferencias de los usuarios, pero no hay conjunto de datos, puede ser apropiado primero evaluar el algoritmo contra conjunto de datos sintéticos para identificar los mejores resultados.

6.3.1. Sintetizado vs. Conjuntos de Datos Naturales

En el trabajo de [Konstan et al. 1997; Miller et al. 1997], experimenta con una variedad de conjuntos de datos sintetizados. Modelan artículos de noticias que tienen un número fijo de propiedades y los usuarios que tienen preferencias por esas propiedades.

Los conjuntos de datos simulados ayudan a probar los algoritmos para encontrar defectos obvios, estos en ninguna manera modelan con precisión la naturaleza de los usuarios y datos reales.

6.4 III. métricas.

En el campo de sistemas de recomendación se han utilizado varios tipos de medidas para la evaluación de la calidad de un sistema de recomendación. En (Herlocker, 2004) se presentan varias métricas que miden la calidad en los Sistemas de Recomendación las cuales son clasificadas en: métricas de exactitud de la predicción, métricas de precisión en clasificación y métricas de exactitud en ordenamiento.

6.4.1. Métricas de la exactitud de la predicción.

Miden qué tan cerca las predicciones previstas por el sistema de recomendación son las verdaderas valoraciones de los usuarios. Métricas de precisión predictiva son particularmente importantes para evaluar tareas en las que la capacidad de predicción se mostrará al usuario como anotación en contexto. Entre estas métricas se encuentran el error medio absoluto y otras métricas relacionadas como el error medio absoluto normalizado y el error cuadrático medio.

6.4.2. Mediciones de precisión de clasificación

Miden la frecuencia con la que un sistema de recomendación toma las decisiones correctas o incorrectas acerca de si un ítem es bueno. Métricas de clasificación son por lo tanto apropiadas para tareas tales como encontrar buenos ítems. Entre estas métricas están precisión and recall y las curvas ROC.

6.4.3. Métricas de exactitud en ordenamiento

En la clase de métricas de precisión de ordenamiento miden la habilidad de un sistema de recomendación para ordenar los ítems recomendados que coincida con el orden que un

usuario haría de los mismos ítems. Entre estas métricas están la correlación predicción calificación, métrica half-life utility y la medida NDPM.

6.4.3.1. Curvas ROC

El modelo de la curva ROC trata de medir el grado con el cual un sistema de filtrado de información puede distinguir de manera exitosa entre relevancia y señal (Herlocker J. L., 2004). El modelo ROC asume que el sistema de información asignará un nivel previsto de relevancia a cada potencial ítem.

La curva ROC asume que hay un filtro z_c , límite de predicción, a partir del cual todos los ítems que el sistema sitúa por encima del límite son vistos por el usuario, y aquellos por debajo no son vistos por el usuario. Como se muestra en la Figura 6.1, en cada valor de z_c , habrá un valor diferente de recall (porcentaje de buenos artículos devueltos) y fallout (porcentaje de malos ítems recuperados). La curva ROC representa una gráfica de recall frente a fallout, donde los puntos de la curva corresponden a cada valor de z_c . Un ejemplo de una curva de ROC se muestra en la Figura 6.1.

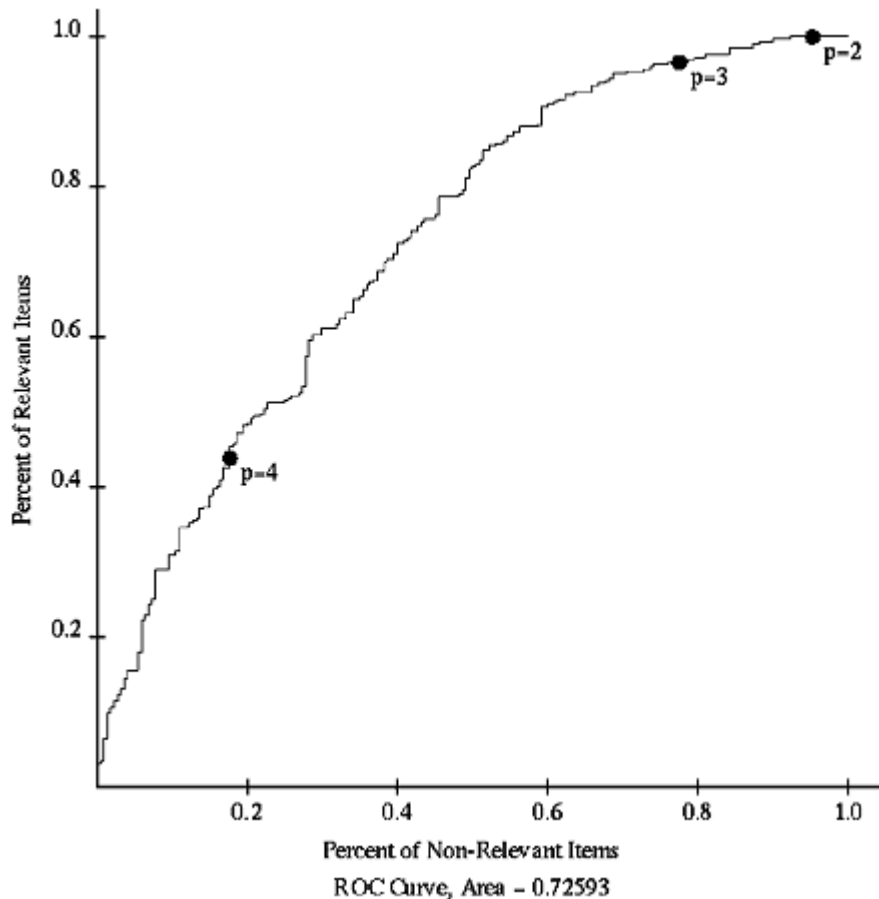


Figura 6.1 Ejemplo de curva ROC, los valores p que se muestran en la curva representan diferentes puntos de corte de predicción.

Las curvas de ROC son útiles para ajustar el equilibrio de la relación señal/ruido en sistemas de información. Es más apropiado cuando existe una relación clara, relevante, y la tarea es similar a encontrar buenos ítems, donde el usuario quiere ver tantas respuestas relevantes como sea posible.

6.5 Pruebas.

Siguiendo las actividades descritas en las secciones anteriores, las pruebas que se realizaron al servicio de recomendación se describen en los siguientes puntos.

6.5.1. I. Tareas del usuario para el sistema de recomendación.

Dentro de los objetivos del servicio de recomendación están el que el usuario use el sistema de recomendación y éste le sea de utilidad al recomendarles ítems de interés o de similitud de acuerdo a su perfil. La tarea de *encontrar buenos ítems* recomienda una parte del conjunto de ítems que son más probables que sean del agrado del usuario, presentando los más similares en la parte más alta de la lista de recomendación.

6.5.2. II. Dataset para evaluación.

El dataset con el que se cuenta para realizar las pruebas contiene:

- 6186 académicos,
- 574 objetos de conocimiento,
- 596 instalaciones,
- 652 infraestructuras,
- 3 certificaciones,
- 187 cursos,
- 198 servicios y
- 7 proyectos de investigación.

De este conjunto de datos los ítems que pasaron el primero filtro, el cual consistió en tener más de 4 palabras, fueron los siguientes:

- 306 académicos,
- 571 objetos de conocimiento,
- 596 instalaciones,
- 537 infraestructuras,
- 2 certificaciones,
- 184 cursos,
- 198 servicios y
- 7 proyectos de investigación.

Como se puede apreciar la mayoría de los académicos no cuenta con la información suficiente para crear un perfil que nos ayude a inferir ítems que pueden ser de interés, a su vez se aprecia que la mayoría de los usuarios no reciben una recomendación.

6.5.3. Métricas

Las métricas que nos ayudaran a medir el servicio de recomendación pueden ser dos de las mencionadas en las secciones anteriores. La primera es la precisión y el recall, ya que esta métrica mide la frecuencia con la que un sistema de recomendación toma las decisiones correctas o incorrectas acerca de si un ítem es bueno.

La segunda es la curva ROC, la cual nos ayudará a decidir los mejores resultados para el algoritmo de recomendación. A medida que la longitud de la lista de recomendación aumenta, más elementos se incluyen, hasta que todo el conjunto de ítems (incluidos todos los artículos relevantes e irrelevantes) ha sido devuelto.

Se decidió emplear la curva ROC para evaluar el rendimiento del algoritmo de recomendación, ya que el servicio de recomendación está en etapa de implementación en su primera fase y se busca encontrar los parámetros que mejores resultados arrojen

La sensibilidad la relacionaremos con el True Positive Rate.

La especificidad la relacionaremos con el False Positive Rate.

Sensibilidad es equivalente al recall (True Positive Rate) mientras que la especificidad mide la fracción de ítems irrelevantes desechados correctamente (False Positive Rate).

La sensibilidad (Ekstrand, 2011) mide la eficiencia en la clasificación de todos los elementos relevantes. La Figura 6.2 muestra como calcular la sensibilidad (Ekstrand, 2011).

	Relevant	Irrelevant
Retrieved	TP	FP
Not retrieved	FN	TN

$$(a) \text{ Sensitivity: } \frac{TP}{TP+FN}$$

Figura 6.2 Sensibilidad.

Se calcula cuantitativamente como el cociente entre los reales positivos (True Positive) entre la suma del total de instancias reales. La primera fila de la matriz de confusión.

La especificidad (Ekstrand, 2011) mide la eficiencia en la clasificación de todos los elementos irrelevantes. La Figura 6.3 muestra como calcular la especificidad (Ekstrand, 2011).

	Relevant	Irrelevant
Retrieved	TP	FP
Not retrieved	FN	TN

$$(b) \text{ Specificity: } \frac{TN}{TN+FP}$$

Figura 6.3 Especificidad.

Se calcula como el cociente entre los reales negativos (True Negative) entre la suma del total de instancias irrelevantes. La segunda fila de la matriz de confusión.

6.6 Resultados

6.6.1. Ítem académico

Definidas las fórmulas se pueden realizar los cálculos de la sensibilidad y la especificidad, se procede a realizar las pruebas. Se calculan los valores de la sensibilidad y la especificidad variando el parámetro. El parámetro que se utilizó es la similitud que existe entre el perfil del usuario y el perfil del ítem, la cual se calcula con la medida similitud coseno. Estos valores se calcularon para cada tipo de ítem.

La Tabla 6.1 muestra el umbral, la sensibilidad y la especificidad obtenidos de las pruebas realizadas, estos datos se utilizan para graficar la curva ROC.

Umbral	Sensibilidad	1 – especificidad
> 91	0.249	0
> 51	0.763	0.112
> 41	0.886	0.306
> 36	0.938	0.516
> 31	0.989	0.84

Tabla 6.1 Resultados de ítem académico.

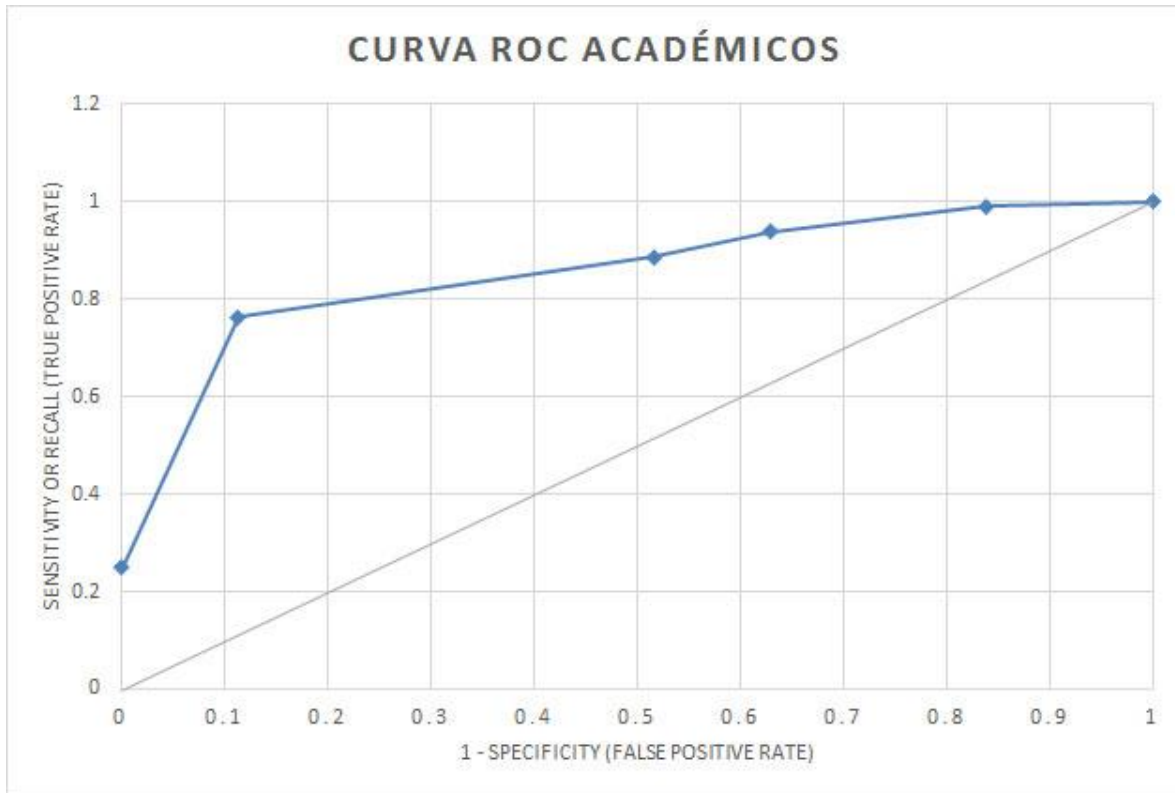


Figura 6.4 Curva ROC ítem académico

Se puede apreciar en la figura 6.4 que el umbral definido con el valor > 51 , el algoritmo está extrayendo el 76% de los ítems correctos (verdaderos positivos), además está extrayendo un 11% de ítems incorrectos (falsos positivos); cambiando el umbral a >41 , el algoritmo está extrayendo el 88% de los ítems correctos (verdaderos positivos), además está extrayendo un 30% de ítems incorrectos (falsos positivos); cambiando el umbral a >36 , el algoritmo está extrayendo el 93% de los ítems correctos (verdaderos positivos), además está extrayendo un 63% de ítems incorrectos (falsos positivos). Podemos concluir que el umbral que mejor resultados está arrojando es el que tiene un valor mayor a 40, tomando en consideración que extrae 88% de ítems relevantes, sacrificando un poco el rendimiento del algoritmo al extraer el 30% de irrelevantes.

6.6.2. Ítem objeto de conocimiento

La Tabla 6.2 muestra los resultados que se obtuvieron del ítem objeto de conocimiento.

Umbral	Sensibilidad	1 – Especificidad
> 51	0.037	0
> 31	0.26	0.04
> 26	0.666	0.16
> 21	0.9	0.9
> 11	1	1

Tabla 6.2 Resultados de ítem objeto de conocimiento.

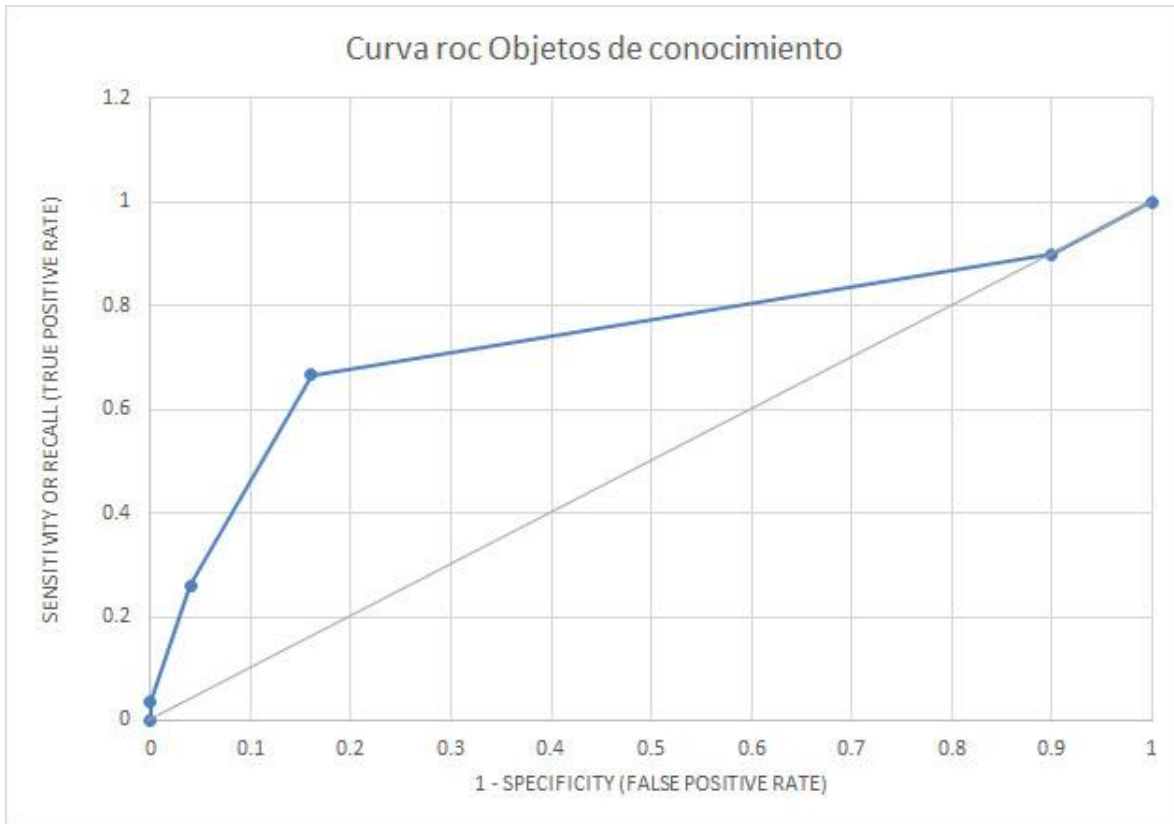


Figura 6.5 Curva ROC ítem objeto de conocimiento

La Figura 6.5 muestra que el umbral definido con el valor > 31 está extrayendo el 26% de los ítems correctos y el 4% de ítems incorrectos; cambiando el valor del umbral >26 , extrae el 66% de los ítems correctos y el 16% de ítems incorrectos; el valor del umbral > 21 , extrae el 90% de los ítems correctos y el 99% de ítems incorrectos arrojando malos resultados. Concluimos que el valor de umbral $>26\%$ es el adecuado para este tipo de ítem.

6.6.3. Ítem instalación

La Tabla 6.3 muestra los resultados que se obtuvieron del ítem instalación.

Umbral	Sensibilidad	1 - Especificidad
> 51	0.064	0
> 41	0.227	0.016
> 31	0.513	0.081
> 26	0.708	0.278
> 21	0.9	0.9

Tabla 6.3 Resultados de ítem instalación

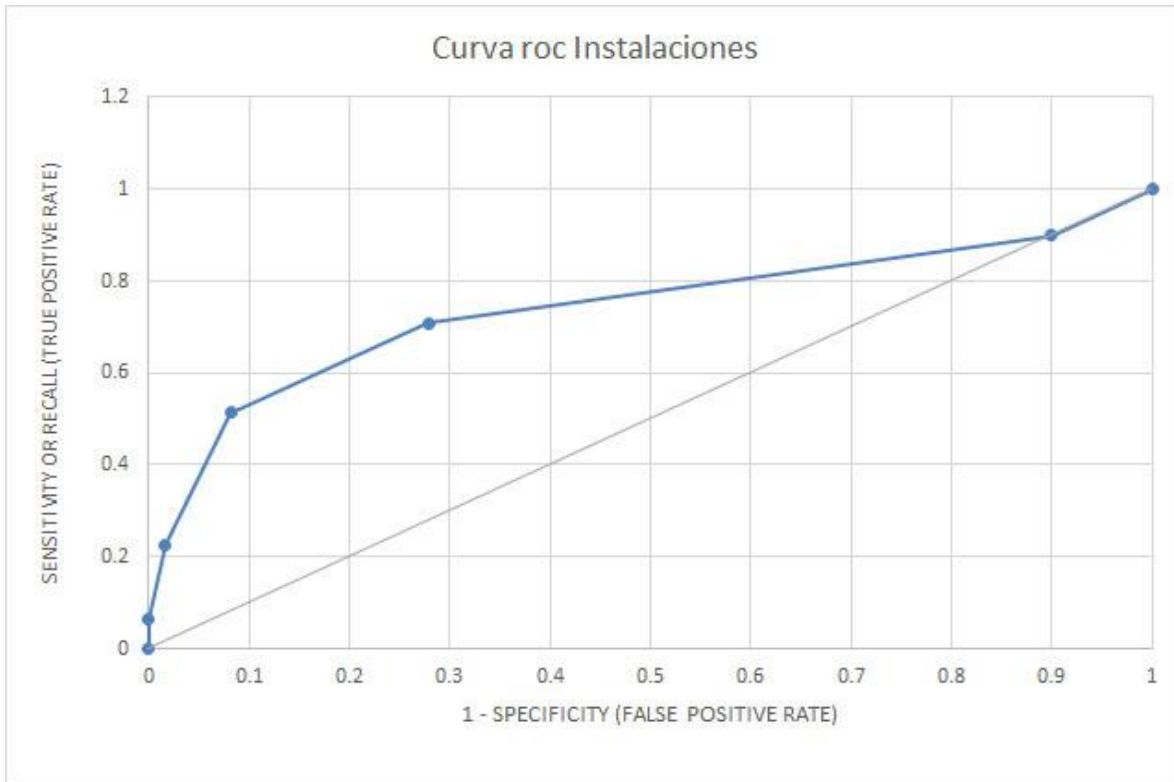


Figura 6.6 Curva ROC ítem instalación

La Figura 6.6 muestra que el umbral definido con el valor > 41 está extrayendo el 22% de los ítems correctos, y el 1% de ítems incorrectos; cambiando el valor del umbral >31 , extrae el 51% de los ítems correctos y el 8% de ítems incorrectos; el valor del umbral > 26 , extrae el 70% de los ítems correctos y el 27% de ítems incorrectos; el valor del umbral > 21 , extrae el 90% de los ítems correctos y el 90% de ítems incorrectos arrojando los peores resultados. Se concluyó que el valor de umbral $>26\%$ es el adecuado.

6.6.4. Ítem infraestructura

La Tabla 6.4 muestra los resultados que se obtuvieron del ítem infraestructura.

Umbral	Sensibilidad	1 - Especificidad
> 41	0.130	0.009
> 31	0.463	0.109
> 26	0.579	0.200
> 21	0.942	0.309

Tabla 6.4 Resultados de ítem infraestructura.

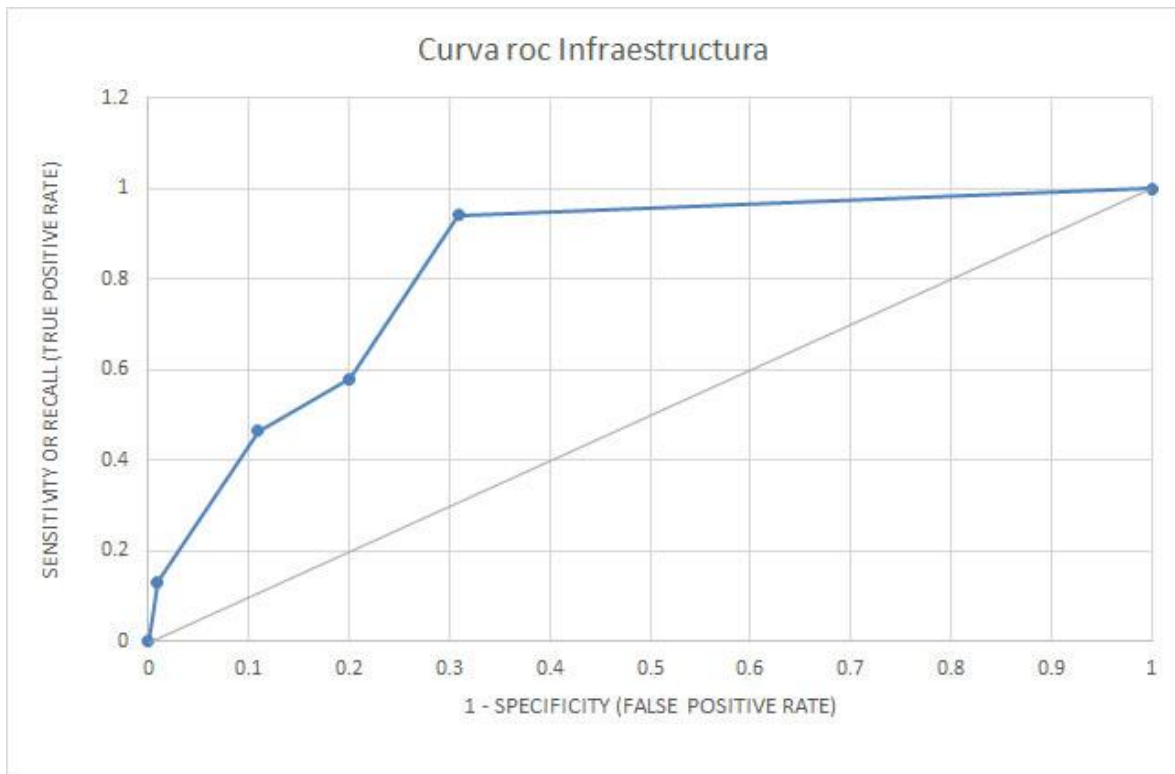


Figura 6.7 Curva ROC ítem infraestructura.

La Figura 6.7 muestra que el umbral definido con el valor del umbral >31 , extrae el 46% de los ítems correctos y el 10% de ítems incorrectos; el valor del umbral > 26 , extrae el 57% de los ítems correctos y el 20% de ítems incorrectos; el valor del umbral > 21 , extrae el 90% de los ítems correctos y el 30% de ítems incorrectos. Concluimos que el valor de umbral $>26\%$ es el adecuado para este tipo de ítem.

6.6.5. Ítem curso

La Tabla 6.5 muestra los resultados que se obtuvieron del ítem curso.

Umbral	Sensibilidad	1 - Especificidad
> 41	0.091	0
> 31	0.326	0.040
> 26	0.530	0.162
> 21	0.938	0.391

Tabla 6.5 Resultados de ítem curso

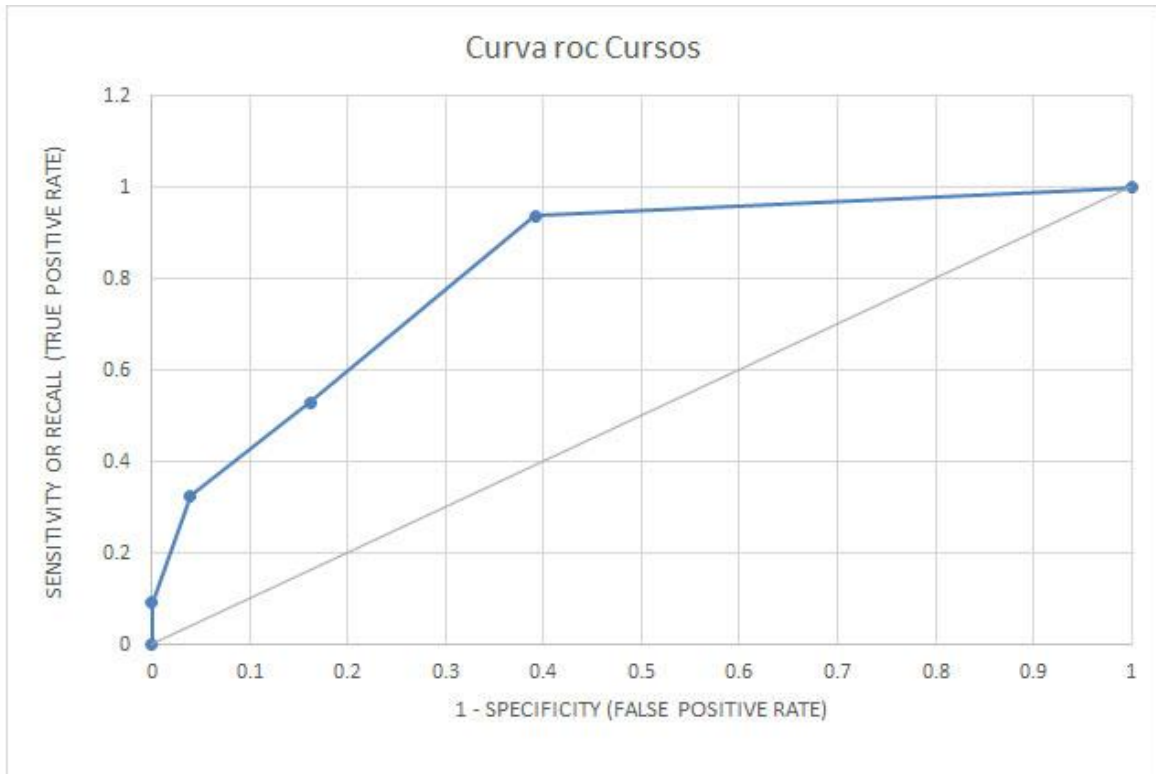


Figura 6.8 Curva ROC ítem curso.

La Figura 6.8 muestra que el umbral definido con el valor del umbral >31 , extrae el 32% de los ítems correctos y el 4% de ítems incorrectos; el valor del umbral > 26 , extrae el 53% de los ítems correctos y el 16% de ítems incorrectos; el valor del umbral > 21 , extrae el 93% de los ítems correctos y el 39% de ítems incorrectos. Concluimos que el valor de umbral $>26\%$ es el adecuado para este tipo de ítem.

6.6.6. Ítem servicio

La Tabla 6.6 muestra los resultados que se obtuvieron del ítem servicio.

Umbral	Sensibilidad	1 - Especificidad
> 41	0.129	0
> 31	0.403	0
> 26	0.58	0.5
> 20	0.95	0.9

Tabla 6.6 Resultados de ítem servicio

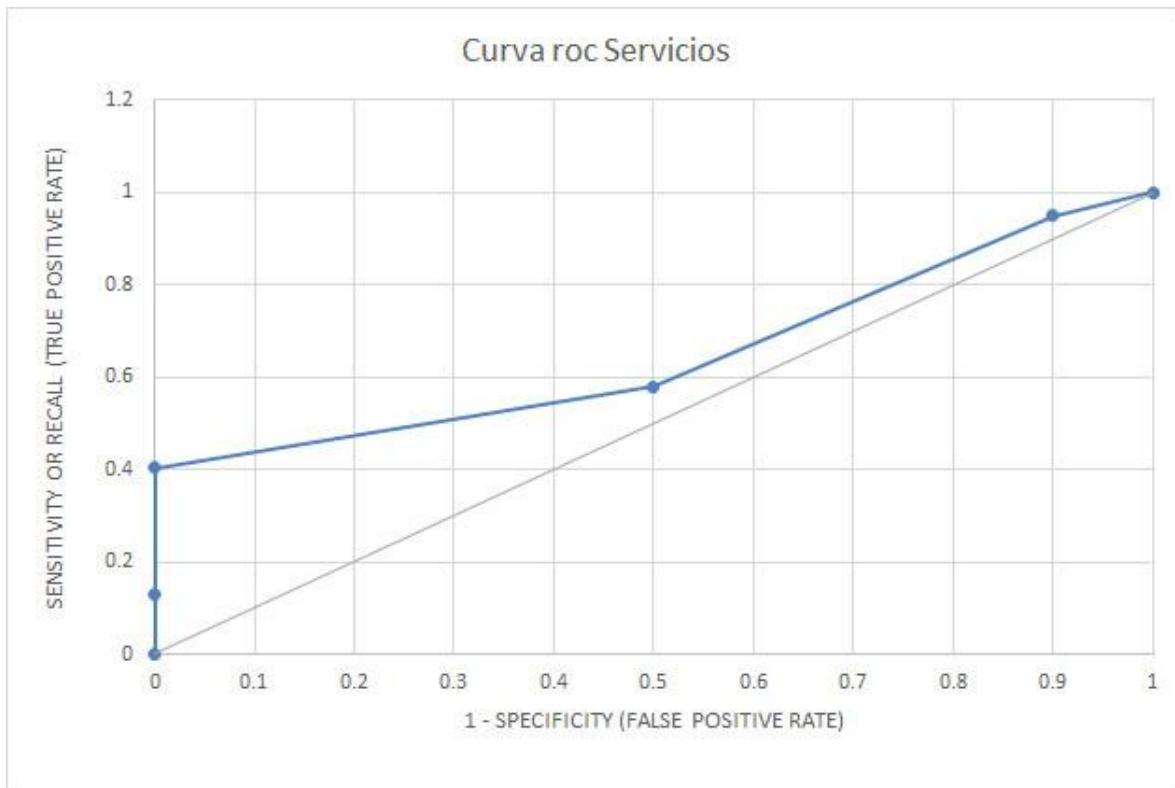


Figura 6.9 Curva ROC ítem servicio.

La Figura 6.9 muestra que el umbral definido con el valor del umbral >31 , extrae el 40% de los ítems correctos y el 0% de ítems incorrectos; el valor del umbral >26 , extrae el 58% de los ítems correctos y el 50% de ítems incorrectos; el valor del umbral >21 , extrae el 95% de los ítems correctos y el 90% de ítems incorrectos. El valor de umbral $>30\%$ es el adecuado para este tipo de ítem.

6.6.7. Ítem certificación e ítem proyecto

Respecto a los tipos de ítem certificación y proyecto, no fue posible realizar un cálculo de la sensibilidad y la especificidad, esto debido a la cantidad de ítems con los que se cuenta actualmente en la base de datos; en el caso del ítem certificación únicamente se cuenta con dos ítems, y en el caso de proyectos se cuenta con 7 ítems.

En ambos casos se hicieron los cálculos correspondientes y el algoritmo arrojaba valores muy variables de similitud tanto baja como alta. En el caso del ítem certificación el algoritmo arrojaba valores entre 20% a 50% de similitudes, siendo estos totalmente irrelevantes o falsos positivos. En el caso del ítem proyecto el algoritmo no encontró ningún ítem con un valor de similitud arriba del 20% de similitud, que fuera un buen ítem a recomendar y los ítems debajo de este umbral son ítems irrelevantes.

6.7 Pruebas realizadas a la aplicación web.

El servicio sobre el cual se desarrolló este trabajo recomienda un conjunto de ítems heterogéneos las cuales se presentan al usuario mediante una lista, se busca generar explicaciones individuales para cada tipo de ítem y no una explicación para la lista completa de los ítems recomendados.

Ya que se utilizan plantillas con campos de texto variable para la generación de las explicaciones, llenadas a partir de los descriptores de características de los ítems; y debido a la orientación del sistema hacia Instituciones de Educación Superior (IES) y la diversidad de tipos de usuarios; se plantea una forma de redacción más personalizada hacia el usuario, incluyendo su nombre así como el grado académico.

Con el fin de personalizar las explicaciones se tomaron como características las variables demográficas que afectan al consumidor, tales como: nivel de estudios, realizando una segmentación demográfica de acuerdo al tipo de usuario del sistema.

Con el fin verificar que se esté cumpliendo con los objetivos para las cuales fueron definidas las plantillas de explicación, y que la información que las plantillas necesitan para la creación de la explicación por cada ítem, se hicieron las siguientes pruebas a las plantillas y a la interfaz web.

6.7.1. Personalización de las plantillas con el nombre, género y grado académico de los usuarios

En este paso se verificó que el servicio de recomendación este llenando de manera correcta y coherente la información del usuario que nos dice cuál es el grado académico y genero del usuario, para definir el título que acompañara el nombre del usuario. En la Figura 6.10 se muestra en ejemplo.

Explicación: Licenciado Luis Fernando González Aldrete se le ha recomendado a este académico, porque posee conocimientos sobre: presente, establecidos, nacional, meg, sistema, alumnos, educativo, basado, proporcionar, servicio, ambiental, modelo, genero, iso, equidad, gestion, integral, calidad, norma, internacional, equivalente, nmx-cc-9001-imnc-2008, orientacion, nmx-ssa-14001-imnc-2004, cumpliendo, requisitos, objetivos, tec, ; los cuales podrian ser de su interes.

Figura 6.10 Explicación utilizando grado y género.

Cuando el servicio no encuentra información del grado académico del usuario, llena la plantilla utilizando el título C. para completar la información, como se muestra en la Figura 6.11.

Explicación: C. Ramon De Jesus Villar Hernandez se le ha recomendado a este académico, porque posee conocimientos sobre: mecanica, ; los cuales podrian ser de su interes.

Figura 6.11 Explicación donde usuario no tiene grado académico.

6.7.2. Información utilizada por el servicio de recomendación

Además del grado académico del usuario y el género también se extrae la información utilizada por el algoritmo de recomendación para completar la explicación, el algoritmo que

se utilizó para inferir las recomendaciones es basado en contenido, la información que se utiliza para completar la explicación de la recomendación, es el vector de palabras claves. En la Figura 6.12 se muestra el vector de palabras claves que completa la plantilla de recomendación.

Explicación: Licenciado Luis Fernando González Aldrete se le ha recomendado a este académico, porque posee conocimientos sobre: presente, establecidos, nacional, meg, sistema, alumnos, educativo, basado, proporcionar, servicio, ambiental, modelo, genero, iso, equidad, gestion, integral, calidad, norma, internacional, equivalente, nmx-cc-9001-imnc-2008, orientacion, nmx-ssa-14001-imnc-2004, cumpliendo, requisitos, objetivos, tec, ; los cuales podrian ser de su interes.

Figura 6.12 Vector de palabras claves.

6.7.3. Recuperación de información de los campos de texto variables para cada tipo de ítem.

En este paso se verificó que el servicio recupere los campos de texto variable para cada tipo de ítem, información que es importante para completar la plantilla de explicación de cada ítem. Los tipos de ítems recomendados por el sistema son: académicos, objeto de conocimientos, instalaciones especializadas, infraestructura especializada, certificación, cursos, servicios y proyectos de investigación. Estas plantillas son definidas de acuerdo al dominio de IES y para cada tipo de ítem recomendado.

6.7.3.1. Ítem académico

La información que se necesita del tipo de ítem académico, es la que define mejor los conocimientos y experiencia que el académico ha adquirido. Los campos de texto que se recuperan son la reseña, la línea de investigación y la certificación. La Figura 6.16 muestra la plantilla donde se puede apreciar la explicación y la información recuperada por el servicio de recomendación para el ítem académico. Es importante mencionar que las recomendaciones deben ir acompañadas de una imagen, y como no todos los usuarios cuentan con una imagen de perfil, se está mostrando una imagen predefinida que hace referencia al usuario de tipo académico misma que se muestra en la Figura 6.13.



Figura 6.13 Ítem académico.

6.7.3.2. Ítem objeto de conocimiento

La información que se necesita del tipo de ítem objeto de conocimiento, la que define mejor este tipo de ítem son el título, descripción y palabras claves. La Figura 6.14 muestra la plantilla de explicación y la información recuperada por el servicio de recomendación para el ítem objeto de conocimiento. La imagen utilizada por el servicio de recomendación para mostrar junto con la explicación de este ítem es la primera página del documento que se adjunta con el ítem, cuando este ítem no cuente con un documento de donde extraer la imagen, el servicio muestra una imagen predefinida.



Figura 6.14 Ítem objeto de conocimiento.

Cuando este tipo de ítem cuenta con un archivo, el usuario puede visualizar el archivo o descargarlo en formato pdf. La Figura 6.15 muestra la visualización del archivo.



Figura 6.15 Visualización de archivo.

6.7.3.3. Ítem instalación

La información que se necesita del tipo de ítem instalación es el título, la descripción y la persona responsable de la instalación. La Figura 6.16 muestra la plantilla de explicación y la información recuperada por el servicio de recomendación para este ítem. La imagen utilizada por el servicio de recomendación para este ítem es la imagen del perfil del ítem.



Figura 6.16 Ítem instalación

6.7.3.4. Ítem infraestructura

La información que se necesita del tipo de ítem infraestructura son el título, la descripción y la persona responsable del ítem. La Figura 6.17 muestra la plantilla de explicación y la información recuperada por el servicio de recomendación. La imagen utilizada por el servicio de recomendación para este ítem es la imagen del perfil del ítem.



Figura 6.17 Ítem infraestructura.

6.7.3.5. Ítem certificación, curso y servicio.

La información que se necesita para los tipos de ítem certificación, curso y servicio tecnológico son el título, la descripción y la persona responsable del ítem. La Figura 6.18, 6.19 y 6.20 muestra la plantilla de explicación y la información recuperada por el servicio de recomendación para los ítems certificación, curso y servicio respectivamente.



Figura 6.18 Certificación



Figura 6.19 Ítem curso.



Figura 6.20 Ítem servicio.

La imagen utilizada por el servicio de recomendación para mostrar junto con la explicación de estos tipos de ítem es la primera página del documento que se adjunta con el ítem,

cuando este ítem no cuente con un documento de donde extraer la imagen, el servicio muestra una imagen predefinida. Cuando estos tipos de ítems cuentan con un archivo, el usuario puede visualizar el archivo o descargarlo en formato pdf.

6.7.3.6. Ítem proyecto

La información utilizada para el tipo de ítem proyecto son el título, la descripción, las palabras claves y la persona responsable del ítem. La Figura 6.21 muestra la plantilla de explicación y la información recuperada por el servicio de recomendación. La imagen utilizada por el servicio de recomendación es la primera página del documento que se adjunta con el ítem, cuando este ítem no cuente con un documento de donde extraer la imagen, el servicio muestra una imagen predefinida. Cuando estos tipos de ítems cuentan con un archivo, el usuario puede visualizar el archivo o descargarlo en formato pdf.

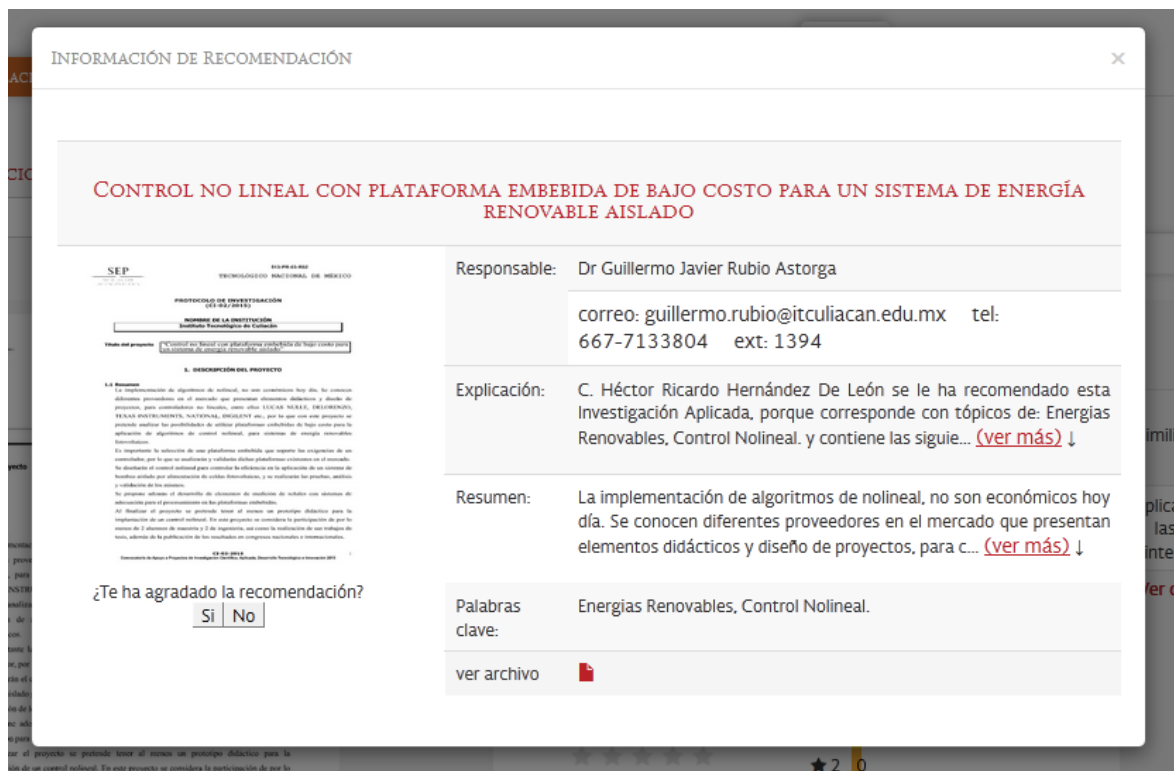


Figura 6.21 Ítem proyecto.

6.8 Pruebas realizadas al funcionamiento de las valoraciones de los ítems.

Las valoraciones son la forma en que los usuarios retroalimentan el sistema. Es importante que este módulo funcione de manera correcta. Las pruebas que se hicieron a este módulo incluyo checar cada tipo de ítem y que cada valor del mismo se almacenara en la base de datos. Debido a la naturaleza de este servicio de recomendación web y la diversidad de navegadores que existen en el mercado. Las pruebas se realizaron considerando probar con varios tipos de navegadores los cuales únicamente fueron seleccionados Firefox y google chrome.

Se realizaron un total de 1500 pruebas individuales. Las cuales consistieron en iniciar sesión con la cuenta del usuario que había recibido una recomendación probar cada tipo de ítem y verificar que el navegador desplegaba de manera correcta la recomendación. Así como verificar que la información de las valoraciones de los usuarios es almacenada en la base de datos. Y checar que la información guardada en la base de datos se viera reflejada en el histograma de valoraciones en la interfaz web en tiempo real.

En la figura 6.22 muestra la manera como el sistema le solicita al usuario que le asigne una valoración al ítem que recibe como recomendación.

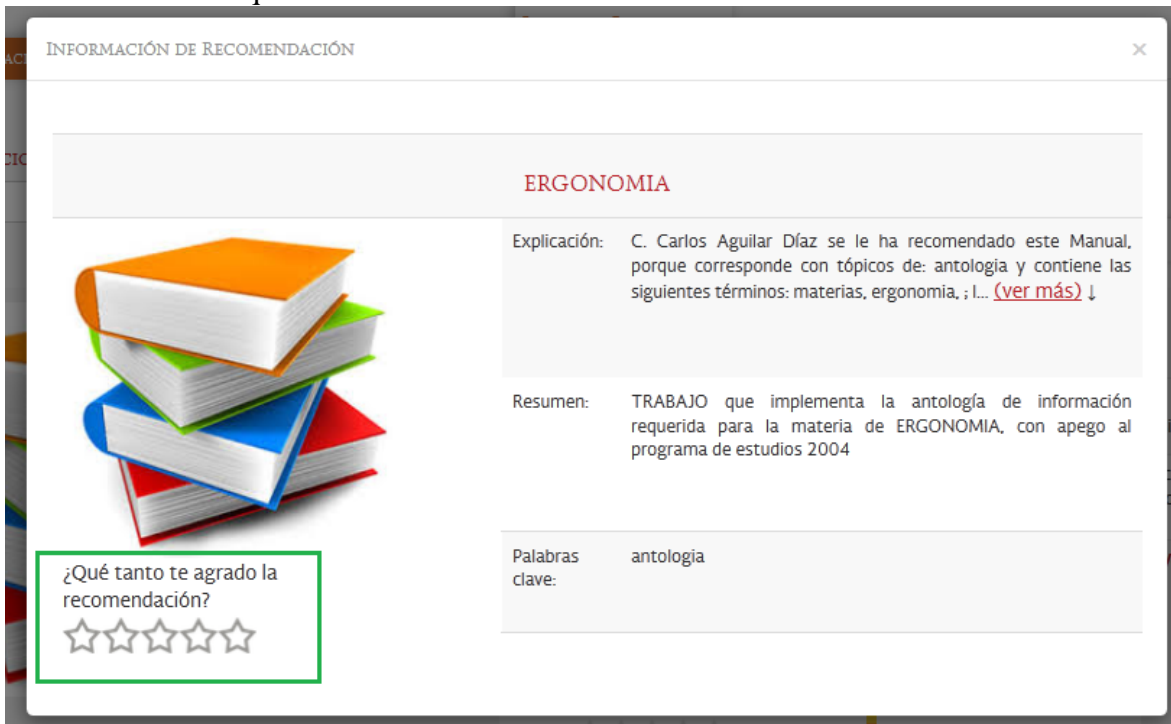



Figura 6.22 Sistema solicita a usuario valoración de ítem.

La Figura 6.23 muestra la valoración que el usuario asigno al ítem valorado.

INFORMACIÓN DE RECOMENDACIÓN
✕

ERGONOMIA



Gracias por tu voto



Explicación: C. Carlos Aguilar Díaz se le ha recomendado este Manual, porque corresponde con tópicos de: antología y contiene las siguientes términos: materias, ergonomía, ; l... [\(ver más\)](#) ↓

Resumen: TRABAJO que implementa la antología de información requerida para la materia de ERGONOMIA, con apego al programa de estudios 2004

Palabras clave: antología

Figura 6.23 Ítem valorado por el usuario.

La Figura 6.24 muestra la actualización del histograma del ítem recomendado.



TIPO: MANUAL:

TÍTULO: ERGONOMIA similitud: 25%

Explicación: C. Carlos Aguilar Díaz se le ha recomendado este Manual, porque corresponde con tópicos de: antología y contiene las siguientes términos: materias, ergonomía, ; los cuales podrian ser de su interes.

Opiniones: Ver detalles

3.5

★★★★☆
1 en total

★ 5	0
★ 4	1
★ 3	0
★ 2	0
★ 1	0



Figura 6.24 Actualización del histograma de ratings.

CAPÍTULO 7

CONCLUSIONES

7.1 Conclusiones

En este trabajo de tesis se realizó la investigación para el desarrollo de un servicio de recomendación, se implementó la metodología desarrollada por (Gálan, 2014) para dotar el servicio de explicaciones. El servicio se diseñó para un entorno web, para el dominio de institutos de educación superior, el cual contiene un conjunto de ítems heterogéneos. Como parte de este trabajo se desarrolló el servicio de recomendación para la PCCT, la cual está orientada a IES, específicamente el TecNM.

Dentro de las conclusiones obtenidas se encontraron las siguientes. En un principio se planteó definir el mismo umbral para todos los tipos de ítems, con los resultados obtenidos se puede apreciar que definir el mismo umbral para todos los tipos de ítems no es la mejor opción. Cada tipo de ítem arroja diferentes resultados según se varía el umbral de similitud entre el perfil del usuario y el tipo de ítem.

Estos resultados nos permiten ajustar los valores que cada tipo de ítem debe tener, el cual nos ayudará a extraer las recomendaciones más relevantes para cada usuario. Se espera que al tener un umbral que extraiga una buena cantidad de ítems positivos (o relevantes) y descarte la mayor parte de ítems negativos (o irrelevantes), las recomendaciones sean de utilidad e interés de los usuarios.

No todos los usuarios reciben una recomendación, se decidió no presentar recomendaciones con un bajo grado de similitud, para que los usuarios que reciben una recomendación utilicen el sistema y ofrezcan una opinión respecto a la recomendación ofrecida por servicio de recomendación.

Los umbrales que mejores resultados ofrecieron tomando en consideración que extraiga una mayor cantidad de ítems relevantes y una baja cantidad de ítems irrelevantes fueron:

- académico es $> .40$
- objeto de conocimiento es $> .25$
- instalación es $> .31$
- infraestructura es $> .26$
- curso es $> .26$
- servicio es $> .30$

Respecto al algoritmo basado en texto se encontró que la cantidad de ítems es importante, como en el caso de certificaciones y proyectos. La cantidad de texto de cada ítem es relevante, en el caso de certificaciones donde solo hay 2 ítems y la media de texto es de 11 palabras, lo que nos da recomendaciones con similitud muy alta que no son relevantes. Caso contrario con proyectos donde la media de palabras es de 55, en este caso la similitud máxima que se obtuvo es de $.10$.

7.2 Aportaciones

Como resultado de este trabajo se obtuvo el conocimiento para modelar el algoritmo de recomendación, así como emplear este tipo de algoritmo para implementar el servicio de recomendación, quedando reflejado en el servicio para la PCCT.

Se adaptó e implemento el algoritmo de filtrado basado en contenido para instituciones de educación superior, el cual se enfocó en un repositorio de usuarios e ítems heterogéneos. Se utilizó la metodología para crear plantillas de explicación, esta metodología implementa plantillas de manera manual y en este trabajo las plantillas son creadas automáticamente por el servicio, el cual utiliza plantillas de explicación para cada tipo de ítem.

Se diseñó un repositorio de valoraciones que serán la entrada para el algoritmo de filtrado colaborativo. Este repositorio también puede ser útil para crear perfiles de los usuarios, pero basado en las características de los ítems que son del agrado del usuario y utilizar estos perfiles empleando otra técnica de recomendación diferente a filtrado colaborativo.

7.3 Trabajos futuros

Como un complemento al trabajo realizado, es posible realizar diferentes trabajos futuros, sugiriéndose los siguientes:

- La implementación del algoritmo de filtrado colaborativo empleando el framework LensKit, el cual obtiene buenos resultados en el dominio de películas. Además ofrece varios algoritmos de filtrado colaborativo y un módulo de pruebas que indica cuál algoritmo se desempeña mejor. Este puede ser útil para adaptarlo al repositorio de la PCCT.
- La Implementación del monitoreo de las acciones efectuadas por los usuarios para aplicar minería de navegación y detectar intenciones de los usuarios que puedan mejorar las recomendaciones inferidas.
- Implementar técnica de valoración unaria que permitan al usuario valorar el ítem si le agrada o no lo agrada como segunda opción a la técnica de valoración tipo liker. Que permita obtener información de los gustos de los usuarios.
- Implementar la técnica que permite al usuario asignarle una etiqueta semántica a los ítems, que permitan mejorar las recomendaciones.
- Desarrollar una aplicación para sistemas operativos móviles.
- Implementar técnicas que permitan recomendar ítems con baja similitud con el perfil del usuario pero que sean de acuerdo a sus intereses, la cual permitirá que el usuario reciba recomendaciones variadas y no únicamente muy similares a su perfil.

REFERENCIAS

- Alejandres, H. (2012). *Desarrollo de un modelo de interacción para smartphones utilizando realidad aumentada y PLN para la explotación de información contextual de ontologías multidimensionales de organizaciones, Propuesta Tesis Doctoral*. Cuernavaca, Morelos.
- Alejandres, H. (2012). *Reporte de Tesis Doctoral: Evaluación centrada en el usuario de sistemas de recomendación semánticos sensibles al contexto: efecto de interfaces multimodales interactivas y esquemas de argumentación en la experiencia del usuario*.
- Ali, K., & Van Stam, W. (2004). TiVo: making show recommendations using a distributed collaborative filtering architecture. *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Apple. (2016). *Genius*. Recuperado el 11 de 05 de 2016, de <http://www.apple.com/legal/internet-services/itunes/es/genius.html>
- Arjona, R. (2009). *Servicios de Localización del Contexto Aplicando Perfiles de Movilidad y Tecnologías de Localización Heterogéneas, Tesis de Maestría*. Cuernavaca, Morelos: CENIDET.
- Balabanović, M. &. (1997). Fab: content-based, collaborative recommendation. *Communications of the ACM*. 66-72.
- Bedi, P. A. (2014). SAPRS: Situation-Aware Proactive Recommender system with explanations. . En *International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (págs. 277-283).
- Bennet, J., & Lanning, S. (2007). The Netflix Prize. *KDD Cup and Workshop*.
- Burke, R. (2002). *Hybrid Recommender Systems: Survey and Experiments: User Modeling and User-Adapted Interaction*, 331-370.
- Burke, R. (2007). Hybrid web recommender systems. *The Adaptive Web*, 377-408.
- Buzan, T. (2002). Como crear mapas mentales. *El instrumento clave para desarrollar tus*. Ediciones Urano, S.A.
- Cañas, A., Ford, K., Coffey, J., Reichherzer, T., Carff, R., Shamma, D., y otros. (2000). Herramientas para construir y compartir modelos de conocimiento basados en mapas conceptuales. *13*(2).
- Chirinos Bossio, R. (2007). El problema de la explicación en la ciencia: Las explicaciones causales en Bas Van Fraassen. *Opción*, 23(53). 140-155.
- Ciliberti, N., & Galagovsky, L. (1999). LAs redes conceptuales como instrumento para evaluar el nivel de aprendizaje conceptual de los alumnos. Buenos Aires, Argentina.

-
- Costamanga, A. (2001). Mapas conceptuales como la expresion de procesos de interrelacion para evaluar la evaluacion del conocimiento de alumnos universitarios. *19*(2).
- De Castro, G. M., Navas-López, J. E., López-Sáenz, P., & Delgado-Verde, M. (2010). El capital intelectual de la empresa. Evolución y desarrollo futuro. *Economía Industrial*(378), 37-44.
- Eiselta, H., & Marianov, V. (2006). Employee positioning and workload allocation. *Computers & Operations Research*.
- Ekstrand, M. D. (2011). Collaborative filtering recommender systems. *Foundations and Trends in Human-Computer Interaction Vol. 4, No. 2*, 81-173.
- Ekstrand, M. D. (2011). Rethinking the recommender research ecosystem: reproducibility, openness, and lenskit. In Proceedings of the fifth ACM conference on Recommender systems ACM.
- Ferran, N., & Minguillon, J. (2005). Información cualitativa sobre el uso de los objetos de aprendizaje. Barcelona España.
- Friedrich , G., & Zanker, M. (2011). *Taxonomy for Generating Explanations in Recommender* .
- Gálan, J. I. (2014). *Metodología para la Generación de Explicaciones para Sistemas de Recomendación Sensibles al Contexto*. Cuernavaca.
- Gertner, A. L. (2012). Recommenders for the enterprise: event, contact, and group. En *Proceedings of the sixth ACM conference on Recommender systems* (págs. 299-300).
- Glance, N. S., Arregui, D., & Dardenne, M. (1999). Making Recommender Systems Work for Organizations. *Practical Application of Intelligent Agents and Multi-Agent Technology*.
- González, N. (2012). *Sistema de Recomendación Contextual Basado en Ontologías para Ambientes Organizacionales y de Usuario en Entornos de Cómputo Móvil, Tesis de Maestría*. Cuernavaca, Morelos: CENIDET.
- Gronau, N., & Fröming, J. (2007). Approach for requirement oriented team building in industrial processes. *Computers in industry*.
- Herlocker, J. (1999). Position Statement | Explanations in Recommender Systems. *Workshop, Interacting with Recommender Systems*.
- Herlocker, J. L. ((1999, August).). An algorithmic framework for performing collaborative filtering. *In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval ACM.*, 230-237.
- Herlocker, J. L. (2000). Explaining collaborative filtering recommendations. En *Proceedings of the 2000 ACM conference on Computer supported cooperative work* (págs. 241-250).
- Herlocker, J. L. (2004). Evaluating collaborative filtering recommender systems. . *ACM Transactions on Information Systems (TOIS)*, *22*(1), 5-53.
- Jiang, Y. J. (2012). Recommending academic papers via users' reading purposes. En *Proceedings of the sixth ACM conference on Recommender systems ACM*. (págs. 241-244).
-

-
- Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*(7), 76-80.
- Mahmood, T., & Ricci, F. (2009). Improving recommender systems with adaptive conversational strategies. *Hypertext 2009* (págs. 73-82). ACM.
- Martin, F. (2009). Top 10 lessons learned developing, deploying and operating real-world recommender systems.
- Papadimitriou, A. S. (2011). A generalized taxonomy of explanations styles for traditional and social recommender systems. En *Data Mining and Knowledge Discovery* (págs. 555-583).
- Papadimitriou, A., Symeonidis, P., & Manolopoulos, Y. (2011). *A generalized taxonomy of explanations styles for traditional and social recommender systems*.
- Pera, M. S. (2014). Automating readers' advisory to make book recommendations for K-12 readers. En *In Proceedings of the 8th ACM Conference on Recommender systems* (págs. 9-16).
- Perales, J., & Romero, J. (2005). Procesamiento conjunto de lenguaje e imágenes en contextos didácticos: Una aproximación cognitiva. *ISSN edición impresa: 0212-9728*. España: Servicio de Publicaciones de la Universidad de Murcia. Murcia (España).
- Pinter, R., Marušić, T., & Radosav, D. (2012). Recommender System in E-student web-based adaptive educational hypermedia. 1270-1073.
- Podgorelec, V., Pavlič, L., & Heričko, M. (2006). Using semantic web technologies for project team building. *Proc. of the First international conference New trends in knowledge management. Knowledge management in organizations*.
- Pons, M. D. (2003). Aportaciones de la Teoría de la Argumentación a la Lingüística. *Histórica. - [s.l.] : Res diachronicae,*.
- Rana, S. J. (2013). Most Influential Contextual-Features [MICF] Based Model for Context-Aware Recommender System. . *In Emerging Trends in Communication, Control, Signal Processing & Computing Applications (C2SPCA)*, 1-6.
- Resnik, P., & Varian, H. R. (1997). Recommender Systems. *Communications of the ACM*, 56-58.
- Ricci, F., Rokach, L., & Shapira, B. (2011). *Recommender Systems Handbook*. Springer.
- Rodriguez Piña, R. A., Mas, B. A., Ochoa Avila, M., & Quevedo Alballe, Y. (2008). Mapas mentales y servicios de inteligencia empresarial. *17(6)*. Vancouver: Acimed.
- (2014). Metodología para la Implementación de Interfaces sobre Dispositivos Multisensoriales Aplicada a Sistemas de Recomendación Sensibles al Contexto. En J. Rodríguez.
- Ruiz Iniesta, A. (2014). *Estrategias de recomendación basadas en conocimiento para la localización personalizada de recursos en repositorios educativos [Tesis Doctoral]*.
- Ruiz Iniesta, A. J. (2012). A framework for the rapid prototyping of knowledgebased recommender systems in the learning domain. En *Journal of Research and Practice in Information Technology* (pág. 16).
-

-
- Sahui Maldonado, J. A. (2008). Factores que influyen en la conducta del consumidor. Una aproximación desde las ciencias sociales. *Trabajo presentado para el Premio Nacional de Investigación del COLPARMEX 2008 en el área de Mercadotecnia.*
- Sarwar, B. K. (2001). Item-based collaborative filtering recommendation algorithms. . *In Proceedings of the 10th international conference on World Wide Web . ACM., 285-295.*
- Schafer, J., Konstan, J., & Ryedl, J. (2001). E-Commerce Recommendation Applications. *Journal of Data Mining and Knowledge Discovery, 5.*
- Spaeth, A. &. (2013). Combining Collaborative Filtering and Text Similarity for Expert Profile Recommendations in Social Websites. . *En User Modeling, Adaptation, and Personalization* (págs. 178-189). Berlin Heidelberg.: Springer.
- Tintarev, N. &. (2011.). *Recommender Systems Handbook: Designing and evaluating explanations for recommender systems.* Springer US.
- Tintarev, N. (2009). *Explaining Recommendations.* Aberdeen: University of Aberdeen.
- Tintarev, N., & Masthoff, J. (2007). A Survey of Explanations in Recommender Systems. *En IEEE 23rd International Conference on Data Engineering Workshop.*
- (2009). Personalization based on Semantic Web Technologies. . *En V. P. Tsetsos, Semantic Web Engineering in the Knowledge Society* (págs. 52-75).
- Vashisth, P. B. (2012). Trust enabled Argumentation Based Recommender System. *En 1. I. ISDA.*