



Evaluación de Sistemas Recomendadores de Contenidos Educativos a través de Estudios de Usuarios

Tesista: Lic. Enrique Antonio Cingolani

Director: Dr. Alejandro Fernández

**Tesis presentada para obtener el grado de
Magister en Tecnología Informática**

**Facultad de Tecnología Informática
Universidad Abierta Interamericana**

Agosto de 2014

Agradecimientos

Deseo expresar mi más sincero agradecimiento a todos aquellos que hicieron posible la realización de este trabajo.

- ? *A mi esposa y mis hijas por su apoyo constante e incondicional.*
- ? *Al Dr. Alejandro Fernández, director de esta tesis, por su permanente ayuda, sus acertados consejos, su amabilidad y disposición a lo largo de todo este recorrido.*
- ? *A Iván Dackiewicz y al Dr. Carlos Neil por sus importantes colaboraciones.*
- ? *A los docentes y ayudantes de las materias Orientación a Objetos I y II y Diseño y Arquitectura de Software, de la Universidad Nacional de La Plata y de la Universidad Abierta Interamericana, así como a los programadores de la plataforma Moodle de la UNLP, sin cuya ayuda no hubieran podido aplicarse las encuestas.*
- ? *A los estudiantes y expertos que participaron de las respectivas encuestas, aportando información invaluable.*
- ? *En general agradezco a todos quienes, de una u otra manera, cooperaron en la realización de este trabajo, sin cuyas ayudas jamás podría haberlo concretado.*

A todos, muchas gracias.

Resumen

En este trabajo se realiza una presentación introductoria de los sistemas de recomendación, se caracterizan los diversos tipos de sistemas recomendadores existentes y en particular se describen los sistemas de recomendación de contenidos educativos y objetos de aprendizaje.

Se efectúa una reseña de los diferentes métodos utilizados para la evaluación de los sistemas recomendadores y se profundiza especialmente en las características y estrategias de evaluación propias para los sistemas de recomendación de contenidos educativos, continuándose con un panorama de las investigaciones en esta área.

Como núcleo central del trabajo, se presenta una propuesta metodológica para la evaluación de sistemas recomendadores de contenidos educativos en entornos de aprendizaje formal, a través de estudios de usuarios. Para esto se brinda una herramienta de evaluación que comprende una encuesta a usuarios y un método para la adquisición de los datos y su posterior análisis. Se completa con un estudio de caso correspondiente a la aplicación de esta metodología para la evaluación de un sistema recomendador en concreto y se analizan los resultados obtenidos.

Por último se elaboran las conclusiones finales y se propone una línea de investigación a seguir para continuar con los estudios en este campo.

Palabras clave

Sistemas recomendadores de contenido educativo, sistemas recomendadores de objetos de aprendizaje, evaluación de sistemas recomendadores, estudios de usuarios, metodología para la evaluación de sistemas recomendadores.

Índice General

| | |
|---|-----------|
| AGRADECIMIENTOS..... | 2 |
| RESUMEN..... | 3 |
| PALABRAS CLAVE | 3 |
| ÍNDICE GENERAL..... | 4 |
| ÍNDICE DE TABLAS..... | 6 |
| ÍNDICE DE ILUSTRACIONES | 7 |
| 1. INTRODUCCIÓN..... | 8 |
| 1.1. DESCRIPCIÓN DEL TEMA..... | 8 |
| 1.2. OBJETIVOS DEL TRABAJO PRESENTADO | 8 |
| 1.3. ESTRUCTURA DEL TRABAJO..... | 9 |
| 2. ESTADO DEL ARTE..... | 11 |
| 2.1. INTRODUCCIÓN A LOS SISTEMAS RECOMENDADORES..... | 11 |
| 2.2. TAXONOMÍA DE LOS SISTEMAS RECOMENDADORES | 13 |
| 2.2.1. <i>Basado en el Contenido</i> | 13 |
| 2.2.2. <i>Filtrado Colaborativo</i> | 13 |
| 2.2.3. <i>Demográfico</i> | 13 |
| 2.2.4. <i>Basado en el Conocimiento</i> | 13 |
| 2.2.5. <i>Basado en la Comunidad</i> | 14 |
| 2.2.6. <i>Híbridos</i> | 14 |
| 2.3. SISTEMAS RECOMENDADORES DE CONTENIDOS EDUCATIVOS | 15 |
| 2.3.1. <i>Introducción</i> | 15 |
| 2.3.2. <i>Redes de Aprendizaje y Capacitación Continua</i> | 15 |
| 2.3.3. <i>Sistema Recomendador Multiagente</i> | 16 |
| 2.4. EVALUACIÓN DE SISTEMAS RECOMENDADORES | 18 |
| 2.4.1. <i>Estrategias y Consideraciones</i> | 18 |
| 2.4.2. <i>Evaluación a través de Experimentos Fuera de Línea</i> | 18 |
| 2.4.3. <i>Evaluación a través de Estudios de Usuarios</i> | 19 |
| 2.4.4. <i>Evaluación a través de Experimentos en Línea</i> | 21 |
| 2.4.5. <i>Propiedades a Evaluar en un Sistema Recomendador</i> | 22 |
| 2.5. EVALUACIÓN DE SISTEMAS RECOMENDADORES DE CONTENIDOS EDUCATIVOS..... | 26 |
| 2.5.1. <i>Características de estos Sistemas Recomendadores</i> | 26 |
| 2.5.2. <i>Estrategias para la evaluación de Sistemas Recomendadores de Contenido Educativo</i> | 27 |
| 3. EXPLICITACIÓN DEL PROBLEMA..... | 31 |
| 3.1. DELIMITACIÓN DEL PROBLEMA | 31 |
| 3.2. JUSTIFICACIÓN DE LA RELEVANCIA DEL TEMA..... | 32 |
| 4. SOLUCIÓN PROPUESTA..... | 34 |
| 4.1. HACIA UNA METODOLOGÍA PARA LA EVALUACIÓN DE SISTEMAS RECOMENDADORES DE CONTENIDOS EDUCATIVOS..... | 34 |
| 4.1.1. <i>Consideraciones para la Evaluación Fuera de Línea</i> | 34 |
| 4.1.2. <i>Consideraciones para la Evaluación a través de Estudio de Usuarios</i> | 35 |
| 4.2. ESTRATEGIA PARA LA EVALUACIÓN DE SISTEMAS RECOMENDADORES DE CONTENIDOS EDUCATIVOS EN ENTORNOS DE EDUCACIÓN FORMAL..... | 35 |
| 4.2.1. <i>Generalidades</i> | 35 |
| 4.2.2. <i>Requerimientos a cumplir</i> | 36 |

| | |
|---|-----------|
| 4.3. METODOLOGÍA PARA LA EVALUACIÓN DE SISTEMAS RECOMENDADORES DE CONTENIDOS EDUCATIVOS A TRAVÉS DE ESTUDIOS DE USUARIOS | 36 |
| 4.3.1. <i>Propuesta de Variables a Medir</i> | 37 |
| 4.3.2. <i>Dimensiones a Evaluar por Grupos y Divisiones</i> | 39 |
| 4.3.3. <i>Propuesta de la Evaluación Específica a Realizar</i> | 39 |
| 4.4. ENCUESTA A USUARIOS | 40 |
| 4.4.1. <i>Formulario de la Encuesta a Usuarios del Sistema Recomendador</i> | 41 |
| 4.5. CATEGORÍAS DEFINIDAS PARA LA EVALUACIÓN DEL SISTEMA RECOMENDADOR..... | 44 |
| 4.6. IMPLEMENTACIÓN DE LA ENCUESTA | 45 |
| 4.6.1. <i>Adquisición y almacenamiento de los datos</i> | 45 |
| 4.6.2. <i>Análisis de los Datos Adquiridos</i> | 45 |
| 4.6.3. <i>Propuestas para Mejorar el Sistema Recomendador Evaluado</i> | 46 |
| 4.7. VALIDACIÓN A TRAVÉS DE UNA ENCUESTA A EXPERTOS..... | 46 |
| 4.7.1. <i>Formulario de la Encuesta a Expertos</i> | 47 |
| 4.7.2. <i>Categorías Cubiertas en la Encuesta a Expertos</i> | 48 |
| 5. CASO DE APLICACIÓN | 49 |
| 5.1. DESCRIPCIÓN DEL SISTEMA RECOMENDADOR EVALUADO..... | 49 |
| 5.2. APLICACIÓN DE LA ENCUESTA A USUARIOS | 49 |
| 5.3. ANÁLISIS DE LOS DATOS OBTENIDOS EN LA ENCUESTA A USUARIOS..... | 51 |
| 5.4. APLICACIÓN DE LA ENCUESTA A EXPERTOS | 54 |
| 5.5. ANÁLISIS DE LOS DATOS OBTENIDOS EN LA ENCUESTA A EXPERTOS..... | 55 |
| 5.6. COMPARACIÓN DE LOS RESULTADOS OBTENIDOS..... | 57 |
| 6. CONCLUSIONES Y LÍNEAS DE INVESTIGACIÓN FUTURAS..... | 59 |
| 7. REFERENCIAS BIBLIOGRÁFICAS..... | 61 |
| ANEXO A | 63 |
| DATOS PROVISTOS AL EXPERTO PARA COMPLETAR LA ENCUESTA A EXPERTOS..... | 63 |
| ANEXO B | 69 |
| CAPTURAS DE PANTALLAS DE LA PLATAFORMA MOODLE CON EL SISTEMA RECOMENDADOR EVALUADO | 69 |
| ANEXO C | 73 |
| OBSERVACIONES Y COMENTARIOS DE USUARIOS CORRESPONDIENTES A LA ENCUESTA A USUARIOS..... | 73 |
| ANEXO D | 75 |
| OBSERVACIONES Y COMENTARIOS DE EXPERTOS CORRESPONDIENTES A LA ENCUESTA A EXPERTOS..... | 75 |

Índice de Tablas

| | |
|--|----|
| Tabla 1: <i>Comparación de las predicciones con el gusto del usuario</i> | 22 |
| Tabla 2: <i>Comparación de características entre SR de comercio electrónico y de contenido educativo</i> | 26 |
| Tabla 3: <i>Modelo clásico de Kirkpatrick</i> | 29 |
| Tabla 4: <i>Dimensiones a evaluar por grupos y divisiones</i> | 39 |
| Tabla 5: <i>Items de la encuesta a usuarios bajo cada categoría</i> | 44 |
| Tabla 6: <i>Items de la encuesta a expertos bajo cada categoría</i> | 48 |
| Tabla 7: <i>Frecuencia de los ítems en la encuesta a usuarios</i> | 50 |
| Tabla 8: <i>Frecuencia por categoría en la encuesta a usuarios</i> | 50 |
| Tabla 9: <i>Resultados obtenidos en la evaluación por usuarios</i> | 53 |
| Tabla 10: <i>Frecuencia de los ítems en la encuesta a expertos</i> | 54 |
| Tabla 11: <i>Frecuencias por categoría en la encuesta a expertos</i> | 55 |
| Tabla 12: <i>Resultados obtenidos en la evaluación por expertos</i> | 57 |
| Tabla 13: <i>Comparación de los resultados obtenidos en ambas encuestas</i> | 58 |

Índice de Ilustraciones

| | |
|--|----|
| Ilustración 1: <i>Histograma de valores para la categoría Interacción e Interfaz</i> | 51 |
| Ilustración 2: <i>Histograma de valores para la categoría Calidad / Novedad</i> | 51 |
| Ilustración 3: <i>Histograma de valores para la categoría Confianza</i> | 51 |
| Ilustración 4: <i>Histograma de valores para la categoría Utilidad / Relevancia</i> | 52 |
| Ilustración 5: <i>Histograma de valores para la categoría Satisfacción</i> | 52 |
| Ilustración 6: <i>Histograma de valores para la evaluación General del SR</i> | 52 |
| Ilustración 7: <i>Histograma de valores para la categoría Calidad / Novedad</i> | 55 |
| Ilustración 8: <i>Histograma de valores para la categoría Confianza</i> | 55 |
| Ilustración 9: <i>Histograma de valores para la categoría Utilidad / Relevancia</i> | 56 |
| Ilustración 10: <i>Histograma de valores para la evaluación General del SR</i> | 56 |
| Ilustración 11: <i>Estructura de la Red al Tiempo 1</i> | 64 |
| Ilustración 12: <i>Estructura de la Red al Tiempo 2</i> | 66 |
| Ilustración 13: <i>Pantalla principal Moodle 25/06/2013</i> | 69 |
| Ilustración 14: <i>Pantalla repositorio externo Wikilor 25/06/2013</i> | 70 |
| Ilustración 15: <i>Pantalla principal Moodle 01/07/2013</i> | 71 |
| Ilustración 16: <i>Pantalla repositorio externo Wikilor 01/07/2013</i> | 72 |

1. Introducción

1.1. Descripción del tema

Los sistemas recomendadores (recommender systems), también llamados sistemas de recomendación, comprenden herramientas de software y técnicas utilizadas en determinados entornos informáticos, para proveer a los usuarios de sugerencias acerca de contenidos (por ejemplo, páginas web) disponibles que podrían resultarles de interés. Estas sugerencias tienen por objetivo ayudar al usuario en los procesos de toma de decisiones relacionadas con situaciones tales como qué ítems comprar, qué música escuchar, qué noticias leer, entre otras (Ricci, 2011).

El desarrollo de sistemas recomendadores comprende un campo multidisciplinario que involucra áreas muy diversas, como Inteligencia artificial, Interacción humano-computadora, Tecnología de la información, Minería de datos, Estadística, Interfaces adaptables al usuario, Sistemas de soporte para toma de decisión, Mercadotecnia y Análisis de comportamiento del consumidor, entre las más relevantes.

Los sistemas recomendadores resultan de gran utilidad para asistir a los usuarios ante la gran sobrecarga de información, y se han transformado en una de las herramientas más potentes y populares para el comercio electrónico. Durante la última década estos sistemas han tenido un enorme crecimiento, habiéndose propuesto diversas técnicas para su construcción, muchas de las cuales fueron implementadas con gran éxito en los más variados entornos comerciales.

En los últimos tiempos, ante la creciente oferta educativa, tanto formal como no formal, disponible en el ámbito del aprendizaje mejorado con la tecnología (technology enhanced learning), los sistemas recomendadores se han expandido a esta área con el objeto de asistir a las personas en el proceso de personalizar sus aprendizajes. Sin embargo, se observa que esta expansión no fue seguida de la correspondiente evaluación sistemática de los sistemas recomendadores que se utilizan (Manouselis, 2011), lo cual puede vincularse con la falta de una metodología apropiada para tal fin.

Con el fin de contribuir en cierta medida a cubrir esta falencia, en el presente trabajo se expone una propuesta metodológica para realizar la evaluación de sistemas recomendadores de contenidos educativos, a partir de estudios de usuarios, presentándose la herramienta de evaluación utilizada y su aplicación a un caso de estudio.

1.2. Objetivos del trabajo presentado

El objetivo general perseguido en este trabajo, es encontrar una metodología adecuada para realizar la evaluación de sistemas recomendadores de contenidos educativos.

Para la concreción de este objetivo general, se cumplen además con los siguientes objetivos específicos:

- Se presenta, el estado del arte de los sistemas recomendadores, y específicamente de los relacionados con contenidos educativos.

- Se identifican las distintas estrategias de evaluación que actualmente se están utilizando para evaluar sistemas recomendadores en general y, en particular, para los sistemas recomendadores de contenido educativo.
- Se exponen las limitaciones, beneficios y dificultades, de las diferentes estrategias de evaluación.
- Se determinan cuáles de estas formas de evaluación se adaptan más a los sistemas recomendadores de contenido educativo.
- Se proponen métricas específicas, relacionadas con las variables que se deben medir y se presenta una herramienta de medición, correspondiente a una encuesta a usuarios, que cumple con los requisitos de validez y confiabilidad necesarios.
- Se propone una estrategia de evaluación que tiene en cuenta los diversos aspectos estudiados, y se la aplica para evaluar un sistema recomendador de contenidos educativos en particular.
- Por último, se extraen las conclusiones obtenidas y se sugieren nuevas líneas de investigación en esta área.

1.3. Estructura del trabajo

El desarrollo de la propuesta metodológica presentada para la evaluación de sistemas recomendadores de contenidos educativos, a partir de estudios de usuarios en experimentos controlados, comprendió la ejecución de varias etapas, las cuales se describen a continuación.

En la primera etapa se indagó sobre el estado del arte en el área de sistemas recomendadores de contenidos educativos, en particular en la temática relacionada con la evaluación de los mismos, a partir de información obtenida en libros, revistas especializadas, publicaciones de congresos y centros de investigación y documentación disponible en Internet, resultando en un mapa detallado del estado de las investigaciones actuales en este aspecto.

Para poner a prueba la metodología de evaluación propuesta, debió aplicarse la misma en la evaluación de un sistema de recomendación de contenidos educativos existente. Por lo tanto, en la segunda etapa se determinó el escenario concreto sobre el que se conduciría dicho estudio (esto es, se identificó el sistema de recomendación, reconociendo sus características y objetivos de sus usuarios, etc.). Este caso concreto sirvió también como punto de referencia en el diseño de la metodología de evaluación (la cual claramente se elaboró de modo que fuera aplicable a otros sistemas).

La tercera etapa, el corazón de esta investigación, consistió en el diseño de la metodología de evaluación. Se desarrollaron procesos de evaluación, herramientas (cuestionarios, encuestas), y estrategias de interpretación de los resultados, a partir de bibliografía específica para el diseño de herramientas para la toma y análisis de datos provenientes de encuestas a usuarios, y se estudiaron diversos métodos estadísticos a aplicar sobre los datos recolectados.

Posteriormente se condujo el estudio de usuarios, aplicando la metodología desarrollada. El foco de esta etapa consistió en determinar la aplicabilidad de la metodología propuesta, enfocando aspectos de la misma como esfuerzo requerido para su aplicación, reproducibilidad, claridad de los resultados obtenidos, capacidad para

establecer comparaciones, entre otros. Como resultado de la evaluación de efectuaron los ajustes necesarios a la metodología.

Para finalizar el trabajo, se elaboraron las conclusiones a partir de los resultados obtenidos y se formularon propuestas para trabajos futuros en esta área.

2. Estado del Arte

2.1. Introducción a los Sistemas Recomendadores

Los sistemas recomendadores consisten en herramientas de software y técnicas utilizadas para brindar sugerencias en forma automatizada, acerca de tópicos (ítems) que pueden resultar de utilidad, a los usuarios de determinados entornos informáticos.

Están dirigidos, en principio, a personas que carecen de la experiencia, la capacidad o el tiempo necesarios para evaluar la inmensa cantidad de tópicos que generalmente están a su disposición en un sitio Web y se basan en la premisa de que, habitualmente, la gente tiene en cuenta las recomendaciones provistas por otros para tomar decisiones rutinarias (Ricci, 2011).

Los tópicos o ítems recomendados resultan de lo más variados, dependiendo del entorno en el que se encuentre el usuario. Por ejemplo, un sitio Web de contenidos musicales, puede disponer de un sistema de recomendación que sugiera temas musicales que resultarían del agrado del usuario, en tanto que un sitio de venta de libros puede disponer de un sistema de recomendación de la literatura que estime que será del agrado del usuario.

Los sistemas recomendadores se encuentran actualmente en innumerables aplicaciones que exponen al usuario a enormes colecciones de elementos o artículos entre los cuales puede elegir y su función principal es la de proveer al visitante o usuario de una lista de artículos recomendados que podrían resultar de su interés. Para lograr este objetivo, dada la gran diversidad en cuanto a los gustos de los usuarios, existen diferentes métodos. En algunos casos, el sistema recomendador cuantifica la preferencia de cada artículo, por ejemplo a partir de calificaciones (rankings) que les han asignado otras personas. Ante una elección del usuario por un determinado artículo, ciertos sistemas recomendadores inclusive presentan una lista de tópicos adicionales que resultaron de interés para otros usuarios que eligieron el mismo artículo.

Un diseñador que desee incorporar un sistema recomendador a su aplicación, dispone de una gran variedad de algoritmos entre los cuales seleccionar el que mejor se adapte a la concreción de sus objetivos dentro del área de dominio de que se trate (contenidos, comercio electrónico, servicios, entretenimiento, etc.). La decisión de cuál algoritmo utilizar, en la mayoría de los casos, se basa en la realización de experimentos que comparen el rendimiento de los diferentes algoritmos, principalmente su capacidad de predecir, con las restricciones impuestas en cada caso, como por ejemplo la veracidad de los datos que se poseen, la memoria disponible, el uso de CPU, entre otras.

Sin embargo, las predicciones exactas son cruciales pero no suficientes para un buen sistema recomendador. En muchas aplicaciones un sistema recomendador puede ser usado no solamente para predecir los gustos de un usuario, sino también para descubrir artículos más rápidamente, explorar diversos artículos, mantener la privacidad del usuario durante la interacción, obtener respuestas rápidas y sobre todo mejorar la experiencia del usuario.

Los sistemas recomendadores se han convertido en la actualidad en una excelente herramienta para tratar el problema de la sobrecarga de información, guiando al usuario

durante el proceso de selección, ofreciéndole tópicos nuevos o que aún no haya explorado y que pueden resultarle relevantes. A partir del requerimiento del usuario, el sistema de recomendación utiliza distintos enfoques y algoritmos (como se verá en la siguiente sección), para efectuar sus propuestas a partir de los datos de que dispone sobre el usuario y el entorno y genera una lista de recomendaciones, que el usuario puede inspeccionar, aceptar o rechazar.

En base a esta interacción, el usuario provee al sistema de una retroalimentación (feedback), que puede darse en forma implícita (basada en el comportamiento del usuario) o explícita (basada en respuestas dadas por el usuario, por ejemplo puntuaciones que asigne a los diferentes tópicos recomendados). Todas las acciones que realiza el usuario y el feedback recibido al interactuar con el sistema, se almacenan en una gran base de datos del recomendador, lo cual hace posible utilizar este conocimiento para generar nuevas recomendaciones, ya sea para el mismo usuario, o para otros usuarios, de acuerdo con el perfil de cada persona o grupo de pertenencia.

El estudio de los sistemas de recomendación es un área muy nueva, dentro de la ingeniería informática, comparada con otras como las de investigación en bases de datos, motores de búsqueda o robótica. Surgida a mediados de la década de 1990, ha despegado, con un crecimiento sostenido en los últimos diez años, a partir de la explosión mundial en el uso de Internet y la web. Es prueba de esto que desde 2007 se realizan anualmente las conferencias y workshops ACM Recommender Systems (RecSys), y cada vez se encuentran más secciones sobre sistemas recomendadores en distintas conferencias de tecnología informática.

Como indica Ricci (Ricci, 2011), los sistemas de recomendación juegan un papel importantísimo en sitios como Amazon.com, YouTube, Netflix, Yahoo, Tripadvisor, Last.fm e IMDb. En este sentido basta mencionar que Netflix, empresa de alquiler de películas en línea, estableció un premio de un millón de dólares para el equipo que pudiera producir una mejora sustancial en el rendimiento de su sistema de recomendación. Este premio fue otorgado en septiembre de 2009 a un grupo de científicos de AT&T Labs que logró una mejora en el rendimiento de alrededor de un diez por ciento sobre el sistema recomendador utilizado en ese momento por Netflix.

2.2. Taxonomía de los Sistemas Recomendadores

Los sistemas recomendadores pueden clasificarse de acuerdo con diferentes aspectos relacionados con su abordaje y los algoritmos de recomendación que utilizan. La clasificación indicada por (Ricci, 2011) es la provista por (Burke, 2007), la cual se ha convertido en una manera clásica de referirse a estos sistemas, y comprende las seis diferentes clases de tratamiento que se presentan a continuación.

2.2.1. Basado en el Contenido

El sistema aprende a recomendar ítems similares a los que le han gustado al usuario en el pasado. La similitud de los ítems se calcula en base a características asociadas con los ítems con los que se compara. Por ejemplo, si un usuario ha calificado en forma positiva a una película del género comedia, entonces el sistema aprende a recomendarle otras películas del mismo género.

2.2.2. Filtrado Colaborativo

La implementación original más simple de esta forma de recomendación (Schafer, 2007), recomienda al usuario los tópicos que han elegido otros usuarios que comparten gustos similares. Aquí la similitud en el gusto de dos usuarios se calcula en base a las coincidencias encontradas en las calificaciones que dieron ambos a distintos tópicos. Es por esto que Schafer (Schafer, 2001) se refiere al filtrado colaborativo como “correlación persona-persona”. El filtrado colaborativo está considerado en la actualidad como la técnica de recomendación más popular y ampliamente implementada.

2.2.3. Demográfico

Estos sistemas recomiendan los tópicos en base al perfil demográfico del usuario. Asumen que se deben generar diferentes recomendaciones de acuerdo al nicho demográfico del usuario. Muchos sitios web se personalizan de manera simple y efectiva de acuerdo con el criterio demográfico, por ejemplo redirigiendo a los usuarios a sitios web de su propio país o en su propio lenguaje. Algunos sistemas dan recomendaciones de acuerdo con la edad del usuario. Sin embargo este enfoque no se ha extendido mucho entre los sistemas recomendadores (Mahmood, 2007).

2.2.4. Basado en el Conocimiento

Estos sistemas recomiendan los tópicos basados en conocimiento específico de dominio. Interpretan cómo ciertas características de un tópico cumplen con los requerimientos y preferencias del usuario y en qué medida este tópico puede serle útil. Muchos de estos sistemas están basados en casos (Bridge, 2006), (Ricci, 2006). Utilizan una *función de similitud* que estima cuánto de lo que el usuario necesita (descripción del problema) coincide con las recomendaciones (solución del problema). El valor que toma esta función se interpreta directamente como la utilidad que tiene la recomendación para el usuario.

Un tipo especial de sistemas recomendadores basados en conocimiento, son los sistemas basados en restricciones. Ambos tipos son similares desde el punto de vista del conocimiento utilizado. La diferencia más grande entre sistemas recomendadores

basados en casos y basados en restricciones es que los primeros usan métricas de similitud, en tanto que los segundos utilizan por lo general bases de conocimiento que contienen reglas explícitas que indican cómo relacionar los requerimientos del usuario con las características del tópico.

Los sistemas recomendadores basados en el conocimiento tienden a trabajar mejor que otros al principio, pero si no están bien equipados con componentes de aprendizaje son pronto superados por otros que utilizan métodos que explotan las interacciones de las personas con las computadoras (como los de filtrado colaborativo, en comunidades con número creciente de usuarios).

2.2.5. Basado en la Comunidad

Este tipo de sistema recomienda los tópicos en base a las preferencias de los amigos del usuario. La evidencia sugiere que las personas tienden a confiar más en las recomendaciones de sus amigos que en las de desconocidos (Sinha, 2001). Esto unido a la creciente popularidad de las redes sociales, está incrementando el interés en este tipo de sistemas, también conocidos como *sistemas de recomendación social* (Golbeck, 2006). Estos sistemas recomendadores adquieren información acerca de los contactos o relaciones sociales del usuario y de las preferencias de sus amigos. Las recomendaciones se basan en calificaciones dadas por los amigos del usuario. El crecimiento continuo de las redes sociales posibilitan la adquisición de estos datos.

2.2.6. Híbridos

Estos sistemas recomendadores se basan en la combinación de dos o más de las técnicas anteriores. Se pueden combinar dos técnicas para aprovechar las ventajas de una y solucionar las desventajas de la otra, en un modelo en el cual ambas técnicas se apoyan mutuamente.

Por ejemplo los sistemas de filtrado colaborativo tienen el problema asociado con la aparición de un nuevo tópico, del cual no existe calificación. Por otro lado los sistemas basados en contenido no tienen esta limitación, dado que fácilmente pueden conocer las características de un nuevo tópico y verificar si éstas son del gusto del usuario.

Actualmente existen diferentes propuestas para combinar dos o más técnicas de recomendación y crear un nuevo sistema híbrido (Burke, 2007).

2.3. Sistemas Recomendadores de Contenidos Educativos

2.3.1. Introducción

Los últimos años han sido testigos de la expansión de los sistemas recomendadores en las más diversas áreas (Jannach 2011). Además de las aplicaciones más tradicionales de estos sistemas ligadas a dominios de contenidos, comercio electrónico, servicios y entretenimiento, día a día se suman otros relacionados con nuevas aplicaciones, como recomendación de *amigos*, *tweets* a seguir y sistemas de recomendación embebidos en entornos de aprendizaje mejorado con tecnología (TEL technology enhanced learning).

Estos últimos, en particular, pueden beneficiarse ampliamente al integrar sistemas recomendadores en el proceso de personalizar el aprendizaje del usuario, adaptándolo a sus conocimientos previos, habilidades y preferencias. Al existir usuarios con distintos niveles de conocimiento, desde novatos hasta expertos, cada uno de estos grupos necesitará diferentes tipos de recomendación: los novatos recomendaciones mas generales y los expertos mas específicas.

De acuerdo con Manouselis (Manouselis, 2011 en Ricci 2011), el aprendizaje mejorado con tecnología (TEL) tiene por objetivo diseñar, desarrollar y probar innovaciones tecnológicas que mejoren las prácticas de enseñanza y aprendizaje tanto de los individuos como de las organizaciones, constituyendo un dominio de aplicación que cubre las tecnologías que sirven de soporte a las actividades de enseñanza y aprendizaje. Dado que la recuperación de información correspondiente a recursos educativos para docentes y estudiantes, constituye una actividad central de TEL, la incorporación de sistemas recomendadores que ayuden a tal fin ha atraído un interés creciente.

En las siguientes subsecciones, para dar una idea de algunas de las propuestas que se están manejando en el campo de los sistemas recomendadores de contenidos educativos, se presentan dos enfoques: el primero debido a Drachsler (Drachsler, 2008), acerca de un *sistema de recomendación personal para el aprendizaje continuo* en las redes de aprendizaje, y el segundo debido a Casali (Casali, 2011), acerca de un sistema recomendador de objetos de aprendizaje basado en conocimiento, utilizando un *sistema multiagente*.

2.3.2. Redes de Aprendizaje y Capacitación Continua

En el ámbito de la educación, el desarrollo y utilización de sistemas recomendadores ha tomado un gran impulso en los últimos tiempos, dentro de los más diversos entornos educativos. Hoy en día se hace evidente la necesidad de una educación continua, y las tecnologías informáticas ponen al alcance de una inmensa cantidad de personas las ventajas del *e-learning*, en donde cada estudiante puede manejar sus propios tiempos en el camino de su capacitación continua.

Drachsler (Drachsler, 2008) habla de los aprendices de por vida (*lifelong learners*) en las redes de aprendizaje (*learning networks*), e indica la necesidad de un sistema recomendador personal que provea a los aprendices de consejos acerca de las actividades de aprendizaje que más les conviene seguir. Estas redes de aprendizaje están dirigidas a aprendices de por vida, en cualquier situación de aprendizaje, para todos los niveles educativos y en los más diferentes contextos. Son redes a las que llama “manejadas por la comunidad” porque en ellas cada miembro también puede contribuir con material de aprendizaje.

Drachsler indica que un sistema recomendador personal en una red de aprendizaje, no puede desarrollarse tomando y ajustando de manera simple un sistema recomendador de productos de consumo. Por el contrario, este autor indica un número de requerimientos específicos creados por el contexto de aprendizaje, que deben tenerse en cuenta, entre ellos: la importancia del contexto de aprendizaje, la novedad inherente de la mayoría de las actividades de aprendizaje, la necesidad de seguir una estrategia de aprendizaje, la necesidad de tener en cuenta cambios en el propio proceso de aprendizaje, el objetivo del aprendizaje, los saberes previos del aprendiz, la calificación que dieron otros aprendices de un grupo similar a las diferentes tareas de aprendizaje, el comportamiento en cuanto a los logros de estudiantes más avanzados de la misma red de aprendizaje, el conocimiento de los caminos de aprendizaje de mayor éxito, la aplicación de estrategias derivadas de la psicología educativa.

Drachsler se enfoca en el uso de técnicas de recomendación basadas en la memoria, las cuales calculan las recomendaciones en base a un conjunto de datos de usuarios y tópicos que son monitoreados continuamente. Estas técnicas comprenden principalmente los siguientes grupos: técnicas de filtrado colaborativo (CF), técnicas basadas en el contenido (CB) y técnicas híbridas. Las técnicas CF recomiendan tópicos que fueron elegidos antes por usuarios (aprendices) similares, en tanto que las técnicas CB recomiendan tópicos similares a los que el propio usuario eligió en el pasado, sin tener en cuenta la opinión de otros usuarios. Las técnicas híbridas combinan las dos primeras para producir recomendaciones más precisas. El autor provee un análisis de las características principales de cada una de estas técnicas de recomendación al ser aplicadas a una red de aprendizaje, e indica las ventajas y desventajas en cada caso.

A partir de estas consideraciones, presenta un “modelo inicial” para un sistema recomendador personal en una red de aprendizaje, descrito a través de un diagrama UML con dos capas (interfaz y objetos) y un motor de recomendación.

Concluye que la mayoría de los sistemas recomendadores de productos de consumo basan sus recomendaciones en un conocimiento limitado de los usuarios e ítems, sin combinar los perfiles de los usuarios y los atributos de los ítems para proveer recomendaciones. Por otro lado los sistemas recomendadores para el aprendizaje continuo, deberían utilizar en forma eficiente los recursos disponibles en una red de aprendizaje, para mejorar la propuesta educativa, teniendo en cuenta las características de los aprendices. Los sistemas recomendadores personales en una red de aprendizaje deberían ser conducidos por reglas pedagógicas que formen parte de la estrategia de recomendación. Esta estrategia debería buscar los datos disponibles para decidir qué técnicas seleccionar para cada situación. Por ejemplo, en la fase inicial el sistema recomendador personal se vería limitado por el problema del “arranque en frío” (cold start), y la estrategia de recomendación debería ser capaz de elegir las técnicas que provean la recomendación que mejor se adapte a este caso.

2.3.3. Sistema Recomendador Multiagente

Casali et al. (Casali, 2011), han desarrollado un sistema recomendador de objetos de aprendizaje, basado en conocimiento, utilizando un sistema multiagente. Este sistema permite a un aprendiz encontrar los objetos de aprendizaje de acuerdo con sus preferencias, buscándolos en un repositorio o en una federación de repositorios, a través del análisis de los metadatos correspondientes a cada objeto de aprendizaje y la

comparación de éstos con el tema de la consulta, el perfil del usuario y los criterios y preferencias que haya determinado.

Para esta tarea los autores diseñaron una arquitectura implementada a través de varios agentes, cada uno de ellos encargado de una parte del trabajo. Cada agente realiza una tarea concreta con la menor intervención humana posible. Los agentes que forman parte de este diseño son:

- El agente de Interfaz, captura los datos ingresados por el usuario y le muestra los resultados hallados, en una interfaz gráfica.
- El agente Refinador Semántico, produce la estrategia de búsqueda asociada al interés del usuario.
- El agente Perfil de Usuario, construye el perfil del usuario.
- Los agentes Buscadores, cuyo objetivo es encontrar en los diferentes repositorios los objetos de aprendizaje que satisfacen el tema y las restricciones impuestas por el usuario
- El agente Mediador, soluciona posibles conflictos entre los objetos de aprendizaje que encontraron los agentes Buscadores
- El agente Recomendador, cuya misión es seleccionar los mejores objetos de aprendizaje de entre los hallados. Este consiste en un agente BDI (Belief, Desire, Intention) graduado, que se encarga de recuperar los recursos y presentarlos en un formato de lista ordenada, comenzando con la prioridad más alta de acuerdo con el perfil del aprendiz.

En esta arquitectura, el agente Recomendador resulta la pieza clave del sistema. Está desarrollado alrededor de contextos y reglas. En el contexto *Belief* (creencia) el agente aplica reglas para determinar su grado de creencia de que un dado objeto de aprendizaje cumpla con las preferencias del aprendiz analizando los metadatos del objeto de aprendizaje. En el contexto *Desire* (deseo) el agente tiene como “deseo” encontrar el objeto de aprendizaje que mejor satisfaga al aprendiz, teniendo en cuenta el tema, las preferencia y las restricciones concretas (por ejemplo de idioma o duración del desarrollo del objeto de aprendizaje). Por último en el contexto *Intention* (intención) el agente combina distintas variables para determinar cómo influye el grado de intención de cada objeto de aprendizaje en los objetivos educativos que se intentan alcanzar.

Los autores indican, como conclusión del estudio de un caso de uso del sistema recomendador implementado, que “la propuesta del sistema recomendador, del agente Recomendador y de las reglas definidas para computar el ranking de objetos, es promisoria”, pero establecen la complejidad de evaluar si efectivamente (con el uso de este sistema recomendador) se han alcanzado mejores resultados en el proceso de enseñanza-aprendizaje. Por otro lado, señalan como dificultad adicional la falta de información en muchos de los metadatos de los objetos de aprendizaje en los repositorios que utilizaron.

2.4. Evaluación de Sistemas Recomendadores

2.4.1. Estrategias y Consideraciones

Resulta de suma importancia y surge como una necesidad, determinar una metodología adecuada para realizar la evaluación de los sistemas recomendadores empleados. Esta evaluación se requiere en diferentes etapas del ciclo de vida del sistema (Ricci 2011).

En el momento del diseño, la evaluación tiene por objeto determinar si se ha escogido el tipo de sistema recomendador apropiado. En esta etapa, se ejecutan evaluaciones “fuera de línea” que consisten en hacer correr distintos algoritmos sobre el mismo conjunto de datos de interacciones con los usuarios, para verificar cuál tiene el mejor rendimiento. Para cumplir con este objetivo se utilizan generalmente bases de datos históricos.

La evaluación también debe efectuarse una vez que el sistema recomendador ha sido puesto en funcionamiento, realizándose una “evaluación en línea” con usuarios reales. Aquí entran en juego diferentes aspectos como la exactitud del sistema al brindar recomendaciones, su interacción con los usuarios, su velocidad de respuesta, su aceptación por parte de los usuarios.

Cuando la evaluación en línea no es posible o resulta demasiado arriesgada, puede realizarse otro tipo de evaluación, que consiste en un experimento controlado o *estudio de usuarios* (Shani 2011). En estos experimentos, un pequeño grupo de usuarios efectúa distintas tareas sobre el sistema, a veces con diferentes versiones del mismo, y completan cuestionarios donde informan sus experiencias. La utilización de estos cuestionarios unidos a mediciones del rendimiento del sistema en cada caso, permite obtener resultados tanto cualitativos como cuantitativos sobre el desempeño del sistema recomendador evaluado.

De acuerdo con Shani y Gunawardana (Shani 2011), los estudios de usuarios “pueden responder el conjunto más amplio de cuestiones” de entre los tres tipos de evaluaciones citadas. Estos autores proponen una serie de propiedades que deben ser tenidas en cuenta al evaluar un sistema recomendador, a la vez que indican que “el efecto de muchas de estas propiedades en la experiencia del usuario -con el sistema- no es claro”, y agregan que “mientras nosotros podemos especular que a los usuarios les gustarían diversas recomendaciones... es esencial demostrar que esto es realmente importante en la práctica”.

Las evaluaciones realizadas durante las diversas etapas del diseño, desarrollo, puesta en funcionamiento y uso del sistema recomendador, así como los experimentos controlados a través de estudios de usuarios, permiten conocer su rendimiento, la exactitud de las predicciones que devuelve, lo adaptable que resulta su interfaz, la rapidez con que entrega resultados, la confidencialidad y seguridad de uso y el grado de satisfacción del usuario, entre otros aspectos. Estas evaluaciones son las herramientas de que se valdrán los desarrolladores para efectuar los ajustes necesarios tendientes a optimizar el funcionamiento del sistema recomendador.

2.4.2. Evaluación a través de Experimentos Fuera de Línea

Los experimentos fuera de línea (*offline*) se realizan utilizando conjuntos de datos de tópicos elegidos por usuarios, que fueron recolectados previamente. Utilizando estos datos se trata de simular el comportamiento de los usuarios que interactúan con un

sistema recomendador. Se asume que el comportamiento de los usuarios que produjeron estos datos va a ser similar con los de los usuarios que van a interactuar con el nuevo sistema recomendador que se está evaluando.

La ventaja de estos experimentos fuera de línea es que son fáciles de efectuar porque no requieren interacción con usuarios reales y permiten correr estos estudios repetidas veces con diferentes sistemas o diversos “retoques” al mismo sistema, a bajo costo. La desventaja es que las preguntas que pueden responder estos estudios están bastante acotadas, en general solamente al poder de predicción del algoritmo, y no a tener un verdadero conocimiento acerca de la forma en que el nuevo sistema recomendador puede influenciar efectivamente sobre los usuarios. Las evaluaciones fuera de línea son, por lo tanto, preferiblemente útiles para descartar algoritmos de recomendación inapropiados y para ajustar parámetros de un sistema en desarrollo.

Un punto importante a tener en cuenta al realizar este tipo de estudios es que los datos históricos que se utilicen sean lo más parecidos a los datos que se esperan encontrar cuando se ponga en funcionamiento al sistema recomendador. Es por eso que debe evitarse el uso de datos que presenten cualquier tendencia (*bias*), ya sea porque provengan de usuarios con alguna particularidad especial que no se va a encontrar entre los usuarios del sistema a prueba, o porque los datos a utilizar fueron previamente filtrados con algún criterio (por ejemplo si se descartaron los tópicos menos votados para achicar la base de datos) que pudiera introducirles una tendencia o desviación.

Por último, las evaluaciones fuera de línea requieren la elaboración de un “modelo de usuario”, para cuya formulación deben hacerse suposiciones acerca del comportamiento de los usuarios. Esto se efectúa habitualmente presentándole al sistema recomendador parte de los datos históricos y ocultándole otros (de un dado usuario y/o de todos los usuarios), para poder comparar las recomendaciones que realice el sistema con las elecciones de los usuarios. Para ello se puede optar por “esconder” datos posteriores al momento en que el sistema debe indicar una recomendación a un usuario. Esto puede efectuarse si los datos históricos tienen un registro de tiempo (fecha y hora), en caso de que no se disponga de esta información, a veces se utilizan, entre otras, estrategias de “mostrar n” o “esconder n” datos al sistema recomendador. En general el modelo de usuario puede hacerse muy complejo, lo importante es que este propio modelo no introduzca tendencias y pueda ser verificado.

2.4.3. Evaluación a través de Estudios de Usuarios

La dificultad de crear verdaderas interacciones con los usuarios hace que las evaluaciones fuera de línea sean difíciles, desde el punto de vista de poder recrear simulaciones que se acerquen lo más posible a la realidad de la interacción. Para subsanar este inconveniente pueden conducirse estudios con usuarios reales.

Los estudios de usuarios se llevan a cabo convocando a un conjunto limitado de personas para interactuar con el sistema recomendador, realizando determinadas tareas que se les encomiendan. Durante el desarrollo de estas tareas se debe observar el comportamiento de las personas y registrar la mayor cantidad de mediciones cuantitativas, tales como el porcentaje de la tarea que fue completada, la exactitud de los resultados, el tiempo que llevó realizar la tarea, si las personas utilizaron las recomendaciones dadas por el sistema, cuántas veces se hizo “clic” sobre cada recomendación, etc.

Por otro lado, en un estudio de usuarios también deben reunirse datos cualitativos, a través de preguntas a los usuarios que pueden realizarse antes, durante o después de que la tarea se haya completado. Estas preguntas permiten reunir información que no es fácilmente observable a través de mediciones directas, como por ejemplo si les gustó la interfaz, si la tarea les resultó fácil de completar, si piensan que las recomendaciones del sistema fueron relevantes, entre otros aspectos.

Según Shani (Shani, 2011) los estudios de usuarios tienen la gran ventaja que “pueden responder el conjunto más amplio de cuestiones” de entre los tres tipos de evaluaciones citadas (*fuera de línea, de usuarios y en línea*). Estos estudios permiten comprobar cuánto influyen las recomendaciones del sistema en el comportamiento del usuario. Es además el único tipo de evaluación que permite reunir datos cualitativos, los cuales son a menudo de gran importancia para interpretar los resultados cuantitativos. Los estudios de usuarios también permiten reunir mucha cantidad de datos cuantitativos, dado que los usuarios pueden ser monitoreados de cerca mientras realizan las tareas interactuando con el sistema recomendador.

Por otra parte, los estudios de usuarios presentan algunas desventajas, la principal es su alto costo, ya sea en tiempo de las personas que se ofrezcan como voluntarios, o económico, si se les debe abonar a las personas participantes. Por este motivo, usualmente este tipo de evaluaciones se realiza sobre un pequeño conjunto de usuarios y un pequeño grupo de tareas, debiéndose tomar la mayor cantidad posible de datos de las interacciones de los usuarios con el sistema durante el tiempo que dure el estudio.

Es conveniente ejecutar alguna *prueba piloto* antes de efectuar el estudio de usuarios. Estas pruebas consisten en experimentos a pequeña escala para probar que el sistema no presente mal funcionamiento o tenga errores que arruinen el desarrollo del estudio con usuarios. En las pruebas piloto se pueden realizar ajustes para mejorar el sistema y dejarlo preparado para el estudio posterior.

Un punto importante al efectuar estudios de usuarios es asegurarse que los usuarios representen lo más fielmente posible a la población de personas que van a utilizar el sistema recomendador. Asimismo es conveniente que los participantes no sepan de antemano cuál es el objetivo al que apunta la prueba, ya que esto podría crear ciertas tendencias (positivas o negativas) en los resultados obtenidos.

Cuando se realizan estudios de usuarios para probar varios sistemas, existen dos alternativas: que todos los usuarios prueben todos los sistemas, o que los usuarios sean divididos en grupos y cada grupo evalúe un único sistema. Cada una de estas formas de distribuir a los usuarios presenta ventajas y desventajas que deben tenerse en cuenta a la hora de interpretar los resultados obtenidos.

Otro punto importante cuando un sistema presenta recomendaciones es la forma como lo hace, si las presenta secuencialmente o todas a la vez y en qué zonas de la pantalla. La presentación de recomendaciones en forma secuencial, produce que la recomendación previa influya a la actual. Si la recomendación previa no fuera del agrado del usuario, la actual le parecerá mejor de lo que realmente pudiera ser, y viceversa. La presentación de recomendaciones en forma simultánea, también puede influenciar al usuario, dependiendo del lugar de la pantalla en que se ubique y de la herencia cultural del sujeto (por ejemplo si la persona lee de derecha a izquierda o de izquierda a derecha), haciéndole pensar que una recomendación es “mejor” que otra.

Los estudios de usuarios permiten reunir datos cualitativos, para lo cual generalmente se recurre a cuestionarios con preguntas que se realizan antes, durante y una vez finalizada la interacción con el sistema recomendador. Estos cuestionarios deben confeccionarse en forma muy cuidadosa, a través de preguntas neutrales, que no sugieran cuál es la respuesta “correcta”, y evitando preguntas que puedan ser percibidas como demasiado privadas o que la persona pueda interpretar que ponen en evidencia algún aspecto personal que no desea exponer.

2.4.4. Evaluación a través de Experimentos en Línea

En los sistemas recomendadores los diseñadores desean influenciar en el comportamiento de los usuarios a través de las recomendaciones presentadas. Para evaluar efectivamente un sistema recomendador es necesario medir el cambio en el comportamiento del usuario cuando interactúa con diferentes sistemas. El sistema que el usuario encuentre más útil y en el que siga más a menudo sus recomendaciones, puede ser considerado superior a otro en el cual los demás aspectos sean iguales.

El efecto que produce una recomendación en un usuario varía de acuerdo a muchos factores, tales como las necesidades del usuario, el contexto y la interfaz a través de la cual se presentan las recomendaciones. La evidencia más fuerte para evaluar un sistema recomendador será corriendo un estudio en línea (*online*) con usuarios reales. Aquí se hace bastante difícil otorgar puntajes absolutos a cada sistema en cuanto a su desempeño y, en la mayoría de los casos, este tipo de estudios consiste en realizar comparaciones entre varios sistemas diferentes y determinar cuál obtuvo un mejor ranking relativo a los demás.

Para llevar adelante estas evaluaciones muchos sistemas recomendadores reales emplean un sistema de *prueba en línea*, redirigiendo una parte del tráfico de usuarios a diferentes motores de recomendación e interfaces. Al realizar estas pruebas en línea es importante que el tráfico se redirija de manera aleatoria, para evitar cualquier tendencia en los resultados. Por otro lado, es fundamental conservar la independencia de los aspectos puestos a prueba, de modo que cuando se deseen probar diferentes algoritmos de recomendación debe mantenerse la misma interfaz, así como para probar distintas interfaces debe mantenerse el mismo algoritmo de recomendación.

Este tipo de pruebas en línea pueden resultar riesgosas en sistemas comerciales, ya que un algoritmo o interfaz a prueba que resulten poco satisfactorios para un usuario, pueden provocar que este usuario no desee utilizar más el sistema. Por este motivo es que las evaluaciones en línea se deben realizar solamente después de haberse efectuado extensos estudios fuera de línea que garanticen que el desempeño del sistema recomendador sea al menos aceptable.

Las evaluaciones en línea tienen una ventaja que no posee ninguno de los otros métodos, en cuanto permiten realizar pruebas reales de funcionamiento del sistema interactuando con usuarios reales, muchas veces durante períodos prolongados. Las variables que deben monitorearse son muchas y diversas y un desafío para la prueba es lograr independizarlas entre sí y medir la influencia que tiene cada una de ellas en particular para el rendimiento global del sistema.

2.4.5. Propiedades a Evaluar en un Sistema Recomendador

Existen diversas propiedades que se espera que posea un buen sistema recomendador. La evaluación correcta de cada una de estas propiedades depende de la estrategia que se haya implementado y de las métricas utilizadas. Tanto en (Shani, 2011) como en (Jannach, 2011), se describen diferentes métricas para cuantificarlas. A continuación se presentan las principales propiedades que deberían ser tenidas en cuenta al evaluar un sistema recomendador.

Preferencia del usuario

Para determinar la preferencia del usuario por un determinado sistema, la opción más simple es realizando directamente una pregunta en un estudio de usuarios, en el cual los usuarios estén probando dos o más sistemas recomendadores. En este caso debe prestarse mucha atención a que no existan desviaciones debidas a características personales de las personas convocadas para el estudio, que no reflejen correctamente las características de los usuarios verdaderos que el sistema tendrá cuando sea puesto en funcionamiento.

Exactitud de las predicciones

Todo sistema recomendador posee un motor de predicciones, y la exactitud de sus predicciones dependerá principalmente del algoritmo utilizado y no de la interfaz de usuario. Esto conduce a que las primeras pruebas del motor de predicciones se realicen habitualmente con experimentos fuera de línea. El siguiente ejemplo muestra algunas de las dificultades y complejidad presentes al efectuar estas evaluaciones.

Al comparar las predicciones del sistema con los gustos del usuario, pueden darse las siguientes posibilidades ante una recomendación propuesta:

| ITEM | Propuesto por el Sistema | Omitido por el Sistema |
|---------------------------|--------------------------|------------------------|
| Aceptado por el Usuario | Predicción Correcta | Falso Negativo |
| Descartado por el Usuario | Falso Positivo | Omisión Correcta |

Tabla 1: Comparación de las predicciones con el gusto del usuario

En experimentos fuera de línea, efectuados a partir de datos históricos, sólo se dispone de información sobre ítems que han sido evaluados por los usuarios, sin hacerse suposiciones acerca de los ítems que no fueron evaluados porque los usuarios no sabían de su existencia. De este modo una recomendación propuesta por el sistema sobre uno de esos ítems no evaluados conduce a un “falso positivo” que penaliza al sistema recomendador, mientras que en una interacción real podría haber sido del agrado del usuario. Por otro lado, al interactuar con usuarios reales, no se sabe si un tópico omitido por el sistema hubiera sido del agrado del usuario, es decir que quedan sin determinar los “falsos negativos”.

Cobertura de las recomendaciones ofrecidas

En muchos casos, sobre todo en sistemas de filtrado colaborativo, las predicciones de un sistema se vuelven más exactas a medida que aumenta la cantidad de datos de que se dispone. Algunos algoritmos disponen de recomendaciones muy precisas sobre un pequeño conjunto de tópicos de los que hay muchos datos, y simplemente no poseen *cobertura* sobre el resto de los tópicos existentes. Este aspecto de la cobertura se define como *cobertura del espacio de ítems*.

Otro aspecto de la cobertura, denominado *cobertura del espacio de usuarios*, se refiere a cuál es la proporción de usuarios o interacciones de usuarios para los cuales el sistema puede recomendar tópicos. Está relacionado con la riqueza que tenga el perfil del usuario en el sistema.

Capacidad de “arranque en frío”

Se refiere al rendimiento del sistema recomendador cuando debe trabajar con nuevos ítems o con nuevos usuarios. Se lo considera a veces como un problema de cobertura, ya que está relacionado con la cobertura que presenta el sistema sobre un conjunto nuevo de ítems o de usuarios.

Confianza

Puede definirse como la confianza que el propio sistema le asigna a la predicción o recomendación que brinda. Habitualmente esta confianza crece con la cantidad de datos disponibles. Hay sistemas que le indican al usuario el puntaje que asigna a una recomendación dada, lo que lo ayuda a tomar ciertas medidas o realizar alguna investigación posterior, en caso que la confianza dada por el sistema a una recomendación sea baja. Desde el punto de vista estadístico, cuando se trata de una distribución normal, el intervalo de confianza en una predicción puede calcularse como la desviación estándar multiplicada por un factor de cobertura.

Credibilidad

Se refiere a la credibilidad o confianza del usuario hacia el sistema recomendador. Por ejemplo, para ganar credibilidad el sistema puede recomendar algunos tópicos que conozca que son del gusto del usuario. Si bien esto no le aporta valor al usuario, el sistema logra ganar en credibilidad, para posteriormente intentar recomendaciones más novedosas, diversas, o de mayor utilidad, sobre tópicos que el usuario desconoce. Otra forma de ganar credibilidad es explicando las recomendaciones presentadas.

La credibilidad puede evaluarse a partir de una pregunta en un estudio de usuarios, en cierta medida y con menor certeza a partir del retorno de los usuarios a un sistema en línea, pero no se han probado métodos confiables de evaluarla en experimentos fuera de línea.

Novedad u Originalidad

Comprende la recomendación de ítems que el usuario desconocía hasta ese momento. Es fácil de evaluar en estudios de usuarios a través de una simple pregunta acerca de un tópico recomendado. También puede evaluarse en experimentos fuera de línea, partiendo el conjunto de datos en un instante dado y analizando qué recomendaciones novedosas propone el sistema a partir de ese instante, las cuales se comparan con las elecciones posteriores del usuario.

Diversidad

Se define como lo opuesto a similitud en las recomendaciones. Muchas veces se prefiere brindar recomendaciones que no sean todas tan similares, intentando presentar alternativas más diversas. La similitud entre ítems se mide a partir de algoritmos basados en el contenido, que utilizan funciones como la suma, el promedio, la mínima o la máxima distancia entre pares de ítems. Al evaluar esta característica, se buscará el sistema recomendador que presente las recomendaciones más diversas pero que, a su vez, obtenga la mayor precisión.

Utilidad

Esta propiedad corresponde a la utilidad ganada por el sistema o por el usuario a partir de una recomendación. En los entornos de comercio electrónico, se ve directamente como el beneficio económico que aporta al sitio la utilización del sistema recomendador. En este caso existen diversas propiedades que pueden ser vistas todas como diferentes *funciones de utilidad*, inclusive la exactitud, la novedad o la diversidad son formas de optimizar la utilidad. En este aspecto sería la evaluación más importante de realizar.

Desde el punto de vista del usuario esto es diferente, y la utilidad no solamente tiene en cuenta minimizar sus costos económicos, sino que para un usuario puede ser de mayor utilidad, por ejemplo recibir una buena recomendación acerca de cómo invertir su tiempo viendo una película o leyendo un libro que sean efectivamente gratificante para su gusto.

Riesgo

Hay casos en los cuales una recomendación puede conllevar un cierto riesgo. Por ejemplo en sistemas recomendadores de inversiones, el sistema puede ajustarse de acuerdo al riesgo, determinando si se prefieren inversiones conservadoras con bajo riesgo y bajo beneficio o si se desea optar por inversiones con mayor riesgo y alto beneficio.

Robustez

Esta propiedad tiene en cuenta la estabilidad del sistema recomendador ante diversos aspectos que involucren situaciones externas. Uno de ellos sería la alta demanda provocada por un gran número de consultas y tiene que ver con la administración del sistema.

Otro aspecto importante es cómo responde el sistema ante el ingreso de información falsa o ataques deliberados tendientes a que el sistema realice cierto tipo de recomendaciones en desmedro de otras. Para evaluar este último aspecto, debe analizarse teóricamente el costo de modificar la puntuación de algún ítem en el sistema. Como esto puede ser muy difícil de realizar por diversos motivos ligados a la cantidad de variables que deben ser tenidas en cuenta, otra alternativa sería simular ataques introduciendo información falsa en el conjunto de datos y medir empíricamente el costo de un ataque exitoso.

Privacidad

Se refiere a la confidencialidad con que el sistema maneja la información personal de sus usuarios. Este es un punto muy importante a ser tenido en cuenta, debido a que los usuarios deben revelar sus gustos personales para obtener las mejores recomendaciones del sistema. El sistema recomendador debe garantizar la privacidad de esos datos y no exponerlos, aún a costa de pérdida en la exactitud de sus predicciones.

Adaptabilidad

Los sistemas recomendadores deben funcionar en entornos en los cuales los ítems a recomendar cambian frecuentemente. Un ejemplo extremo sería un servicio de noticias en donde los tópicos se renuevan permanentemente. Por otro lado las preferencias generales de los usuarios pueden volcarse en instantes, por ejemplo cuando ocurre algún hecho que hace que las personas comiencen a buscar elementos relacionados con el

mismo, ocurre que tópicos que nunca tuvieron altos puntajes, de pronto se vuelven los principales. Un sistema recomendador debe en esos momentos ser capaz de sacrificar algo de exactitud en las predicciones por rapidez en las recomendaciones.

Otro tipo de adaptación ocurre cuando un usuario en particular modifica o agrega nueva información. El espera que las recomendaciones que le propone el sistema cambien en consecuencia. Aquí debe evaluarse al sistema recomendador a partir del estudio de las recomendaciones propuesta antes y después de estas modificaciones.

Escalabilidad y Rendimiento

Dado que los sistemas recomendadores están primariamente diseñados para ayudar a los usuarios a moverse entre enormes colecciones de datos, se espera que escalen bien cuando sean expuestos a datos reales que comprendan conjuntos de millones de ítems a recomendar y en continuo crecimiento. En estas circunstancias se debe estudiar la complejidad de los algoritmos en cuanto a requerimientos de tiempo, espacio de memoria y consumo de recursos al trabajar sobre grandes conjuntos de datos y cantidades crecientes de usuarios.

Las mediciones del rendimiento relativo de distintos sistemas, aún sobre conjuntos de datos más reducidos, puede dar idea del comportamiento que tendrán cuando crezca la demanda. Se espera que los sistemas puedan proveer recomendaciones rápidas en línea, por lo cual una evaluación importante para caracterizar este aspecto es medir cuántas recomendaciones por segundo (throughput) puede proveer el sistema y el tiempo de respuesta (latencia) que requiere para efectuar una recomendación.

Es importante destacar que la mayoría de los sistemas recomendadores no trabajan en tiempo real, sino que efectúan el recalcado cada cierto periodo que depende de las modificaciones que se produzcan en el sistema, con lo cual los parámetros de throughput y latencia pueden verse afectados de acuerdo a la actividad que se esté desarrollando.

2.5. Evaluación de Sistemas Recomendadores de Contenidos Educativos

2.5.1. Características de estos Sistemas Recomendadores

Los sistemas recomendadores de contenidos educativos poseen características que los diferencian de los sistemas utilizados para aplicaciones comerciales. Mientras que el objetivo principal de un sistema recomendador para comercio electrónico es proveer a los consumidores de información para ayudarlos a decidir qué productos comprar (*amazon.com, netflix.com, ebay.com*), el objetivo principal de un sistema recomendador de contenidos educativos en un entorno de aprendizaje mejorado con la tecnología, es proveer a los aprendices de actividades de aprendizaje adecuadas para ayudarlos a desarrollar sus competencias.

Para alcanzar este objetivo el sistema recomendador debe poseer conocimientos de reglas pedagógicas relevantes (Drachler, 2009), por ejemplo, los contenidos o actividades educativos a recomendar deben estar levemente por encima del nivel de conocimientos del aprendiz, dentro de su *zona de desarrollo próximo* (Vygotsky, 1978). De modo que un sistema recomendador de este tipo debe tener en cuenta los niveles de competencia para poder sugerir un tópico adecuado.

Por otra parte un sistema recomendador debe tener siempre presente el objetivo del usuario, y en este aspecto la característica principal de un recomendador de actividades de aprendizaje es que sus usuarios desean alcanzar una competencia específica en un cierto tiempo, a diferencia de un usuario de un sistema de comercio electrónico cuyo objetivo es comprar un producto de una determinada calidad dentro de un cierto rango de precios.

En la tabla siguiente se comparan las características principales de acuerdo a los objetivos de la recomendación que debe ofrecer un sistema recomendador para comercio electrónico y uno para contenidos educativos.

| SR para Comercio Electrónico | SR para Contenidos Educativos |
|--|---|
| Ayudar a los consumidores a decidir qué productos comprar. | Ayudar a los aprendices a desarrollar competencias para alcanzar un objetivo específico de aprendizaje. |
| Convertir visitantes en compradores. | No pueden convertirse visitantes en aprendices. El aprendizaje es una actividad altamente individual. |
| El estado final es la compra del producto. | No se alcanza un estado final en un tiempo fijo. Se alcanza un <i>nivel de competencia</i> que tiene niveles por debajo y niveles por encima. |
| Incrementar las ventas cruzadas, sugiriendo productos adicionales a partir de los productos en el "carrito de compra". | Sugerir actividades de aprendizaje adicionales a partir de las cumplidas. |
| No puede recomendarse el mismo producto que ya lleva (ejemplo: el mismo libro con otra encuadernación). | A veces, para reforzar, deberá recomendarse una actividad igual o muy similar en el mismo nivel de competencia. |
| Garantizar la Lealtad del Comprador, generando una relación de confianza. | La satisfacción del aprendiz dependerá del apoyo que el sistema le provea a lo largo de todo el proceso de aprendizaje. |

Tabla 2: Comparación de características entre SR de comercio electrónico y de contenido educativo

Los sistemas recomendadores para comercio electrónico poseen además la característica de trabajar con miles de artículos provenientes de catálogos, con metadata adecuada y relaciones semánticas entre ellos, desarrollados por expertos en mantener este tipo de información. Por otro lado, los sistemas recomendadores de contenido educativo trabajan con objetos de aprendizaje que comprenden contenidos, actividades, etc., la mayoría de las veces sin catalogar apropiadamente.

Dentro de los entornos de aprendizaje mejorado con la tecnología (TEL), deben distinguirse dos categorías: aquellos de aprendizaje formal y los de aprendizaje informal. Los primeros son entornos de aprendizaje ofrecidos principalmente por escuelas y universidades, altamente estructurados, que otorgan acreditaciones y tienen expertos en su dominio, lo que garantiza su calidad (Drachsler, 2009). Estos entornos están contruidos “de arriba hacia abajo”, con los contenidos y actividades de aprendizaje provistos por la institución educativa, en la cima de la pirámide. En estos casos los objetos de aprendizaje comprenden una cantidad limitada y bastante bien definida de recursos.

Por otro lado, los entornos de aprendizaje informal, se desarrollan de la mano de la web 2.0 y están ligados a los “aprendices de por vida” (Drachsler, 2008). En este tipo de aprendizaje los sujetos se autogestionan en cuanto al camino y ritmo que toman sus aprendizajes. Los materiales provienen de las más diversas fuentes, y los propios aprendices no tienen un rol muy fijo, sino que pueden pasar de aprendices a maestros y a proveedores de materiales para otras actividades. Son entornos contruidos “de abajo hacia arriba”. Ejemplos de estos son sitios como *wikipedia.org*, *flickr.com* y *youtube.com*, en donde se encuentra enorme cantidad de objetos de aprendizaje, sin una estructura fija, en los que contribuyen los mismos usuarios.

En estos entornos también existe el problema de la navegación para los usuarios, que pueden encontrar muy difícil acceder a los materiales disponibles e identificar los que les resultarán más apropiados. En este contexto se espera que un sistema recomendador sea capaz de:

- Estructurar las actividades de aprendizaje de manera pedagógica para mejorar el desarrollo de las competencias del aprendiz.
- Sugerir caminos de aprendizaje a partir de datos emergentes en la red de aprendizaje y producir *recomendaciones de secuencias de actividades de aprendizaje* (Drachsler, 2009), de la misma manera como algunos sistema recomendadores (*youtube.com*, *grooveshark.com*, *dilandau.eu*) proponen secuencias de temas musicales, para alcanzar el objetivo de aprendizaje requerido.

2.5.2. Estrategias para la evaluación de Sistemas Recomendadores de Contenido Educativo

En la evaluación de sistemas recomendadores para comercio electrónico, se utilizan diversas estrategias presentadas anteriormente. Una de las preferidas consiste en realizar experimentos fuera de línea en los cuales se utilizan conjuntos de datos históricos (como los provistos por *MovieLens* o *EachMovie*). Entre las propiedades que se buscan medir para evaluar estos sistemas, aparecen principalmente la exactitud, que mide cuán cerca está la predicción del sistema recomendador con respecto a la elección del usuario, la cobertura, que mide para qué proporción de tópicos el sistema es capaz de realizar recomendaciones, y el rendimiento, correspondiente al tiempo de respuesta del sistema.

Para evaluar sistemas recomendadores en entornos de aprendizaje mejorado con la tecnología no existen conjuntos de datos históricos a los que se pueda recurrir, ni procedimientos de evaluación estandarizados. En este caso deberán evaluarse otros aspectos más allá de los puramente técnicos (Drachsler 2009), en los cuales se tengan en cuenta mediciones del rendimiento educativo y del desempeño e interacciones del estudiante dentro de la red de aprendizaje (Wasserman, 1999).

En este sentido Drachsler realiza una propuesta de una estrategia para evaluar sistemas recomendadores de contenido educativo, y establece la necesidad de investigación para determinar si los estudiantes se benefician de la utilización de un sistema recomendador de actividades de aprendizaje y en qué medida. Para esto propone un marco para la evaluación de tales sistemas recomendadores, en el cual indica qué propiedades deben tenerse en cuenta en dicha evaluación. Estas propiedades pueden asociarse, de acuerdo con sus características, en distintos grupos de mediciones (técnicas, educativas y en la red social), según se indica a continuación.

Mediciones técnicas

Exactitud, Cobertura y Rendimiento

Estos aspectos ya fueron analizados anteriormente y comprende el grupo de mediciones técnicas propuestas.

Mediciones Educativas

Estas mediciones evalúan si los aprendices se beneficiaron del uso del sistema recomendador en sus actividades educativas y comprenden:

Efectividad

Cantidad de actividades de aprendizaje vistas, estudiadas o completadas durante la fase de aprendizaje.

Eficiencia

Tiempo que los aprendices necesitaron para cumplir con el objetivo de aprendizaje.

Satisfacción

Grado de satisfacción, acuerdo y agrado del aprendiz con las recomendaciones recibidas del sistema.

Abandono

Proporción de estudiantes que abandonaron la actividad durante la fase de aprendizaje.

Mediciones en la Red Social

Estas mediciones estiman el beneficio proveniente de las contribuciones de los aprendices dentro de la red de aprendizaje y comprenden:

Variedad

Mide el *nivel de emergencia* en una red de aprendizaje a través de la combinación de caminos de aprendizaje individuales hacia las rutas de aprendizaje más exitosas.

Centralidad

Es un indicador de la conectividad del aprendiz en la red de aprendizaje, se determina contando la cantidad de enlaces que tiene con otros aprendices.

Cercanía

Mide el grado de cercanía de un aprendiz con los demás, y representa la capacidad que tiene para acceder a la información, directa o indirectamente, a través de la conexión con otros miembros de la red de aprendizaje.

Cohesión

Indica cuán fuertemente están conectados entre sí los aprendices por *uniones cohesivas*. Grupos unidos de pares pueden identificarse si cada aprendiz está enlazado directamente con todos los demás aprendices de la red de aprendizaje.

La incorporación de todos estos componentes en un marco evaluativo, seguramente presentará muchas dificultades metodológicas y prácticas. Una idea (Manouselis, 2011) es la de adaptar marcos de evaluación clásicos al contexto de la evaluación de sistemas recomendadores. Manouselis propone, como ejemplo, tomar el modelo de Kirkpatrick (Kirkpatrick, 1959) diseñado originalmente para medir el resultado de cursos de entrenamiento, y adaptarlo para evaluar un sistema recomendador en el contexto del aprendizaje mejorado con la tecnología (TEL), de la siguiente manera:

| Modelo clásico de Kirkpatrick | Pregunta correspondiente para evaluar el Sistema Recomendador |
|---|--|
| 1. Reacción del usuario | |
| Lo que pensó y sintió | ¿Disfrutó las recomendaciones que recibió? |
| 2. Aprendizaje | |
| El aumento resultante de nuevos conocimientos y competencias | ¿Aprendió lo que necesitaba y obtuvo nuevas ideas con la ayuda de las recomendaciones recibidas? |
| 3. Comportamiento | |
| Medida en la cual los conocimientos y competencias adquiridas pueden implementarse / aplicarse en el mundo real | ¿Utilizará la nueva información e ideas de las recomendaciones recibidas? |
| 4. Resultados | |
| Efectos del rendimiento del usuario en el entorno de aprendizaje o trabajo | ¿Las ideas e información recibidas del recomendador mejorarán su eficiencia y resultados? |

Tabla 3: Modelo clásico de Kirkpatrick

A partir de un recorrido realizado sobre una veintena de sistemas recomendadores de contenido educativo en entornos de aprendizaje mejorado con la tecnología, Manouselis encuentra que actualmente existe una falencia con respecto a estudios de evaluación sistemáticos. Se informa que han sido evaluados a través de estudios que involucren interacción o pruebas con usuarios, solamente diez de los veinte sistemas examinados.

De acuerdo con Manouselis, en la definición de un marco de evaluación para sistemas recomendadores de contenidos educativos deberán incluirse:

- Un análisis detallado de los métodos y herramientas de evaluación que pueden ser empleados para evaluar cada uno de los componentes seleccionados (modelo de usuario, modelo de dominio, estrategia de recomendación, algoritmo, etc.). Por ejemplo, en el modelo de Kirkpatrick, debería incluirse el método asociado para medir el efecto del sistema recomendador, sobre cada una de las cuatro dimensiones presentadas.
- La especificación de métricas / indicadores de evaluación para medir el éxito de cada componente del sistema recomendador (por ejemplo exactitud, cobertura, etc.).
- La elaboración de métodos e instrumentos capaces de medir cada una de estas métricas especificadas, con el objeto de adquirir datos de evaluación suministrados por los participantes, en forma explícita o implícita, por ejemplo midiendo, entre otros, el grado de satisfacción y el impacto producido por la incorporación del sistema recomendador, en las tareas cumplidas.

3. Explicitación del problema

3.1. Delimitación del problema

Retomando lo expuesto en el capítulo anterior, actualmente los sistemas recomendadores se encuentran en muchas aplicaciones que exponen al usuario a enormes colecciones de elementos o artículos entre los cuales puede elegir. La función principal de estos sistemas es la de proveer al visitante o usuario de una lista de artículos recomendados (ya sean libros, CDs de música, etc., según el caso) que podrían resultar de su interés.

Para realizar esta tarea, los sistemas recomendadores utilizan diversas estrategias, por ejemplo, cuantifican la preferencia de cada artículo para el usuario, a partir de calificaciones (rankings) que les han asignado otras personas. Cuando el usuario elige un determinado artículo, ciertos sistemas recomendadores pueden buscar y presentar una lista de ítems adicionales que resultaron de interés para otros usuarios que eligieron el mismo artículo.

En este sentido, el sistema recomendador debe ser capaz de predecir los gustos del usuario. Si bien las predicciones exactas son cruciales, no resultan suficientes para un buen sistema recomendador. Se espera que estos sistemas puedan también ser usados para descubrir artículos más rápidamente, explorar diversos artículos, mantener la privacidad del usuario durante la interacción, obtener respuestas rápidas y sobre todo mejorar la experiencia del usuario.

Paralelamente a las aplicaciones más tradicionales de estos sistemas ligadas a dominios de contenidos, comercio electrónico, servicios y entretenimiento, día a día se suman otros relacionados con nuevas aplicaciones, como recomendación de *amigos*, *tweets* a seguir y sistemas de recomendación embebidos en entornos de aprendizaje mejorado con tecnología (TEL technology enhanced learning). Estos últimos, en particular, pueden beneficiarse ampliamente al integrar sistemas recomendadores en el proceso de personalizar el aprendizaje del usuario, adaptándolo a sus conocimientos previos, habilidades y preferencias. Por ejemplo, un sistema recomendador de este tipo puede, teniendo en cuenta la actividad que está desarrollando un alumno, recomendarle contenidos que leyeron y resultaron útiles para otros alumnos mientras realizaban la misma actividad.

Tratándose de un área tan sensible, un sistema recomendador de contenidos educativos debe, necesariamente, ser evaluado extensivamente en cuanto a sus aspectos técnicos y educativos, antes de incorporarse a un entorno de aprendizaje mejorado con tecnología.

Con respecto a las estrategias de evaluación, son diversas las que se han planteado y aplicado en el campo de los sistemas recomendadores para comercio electrónico. Dado que en esta área se dispone de gran cantidad de datos históricos de las interacciones de los usuarios con los sistemas recomendadores, uno de los métodos de evaluación preferidos para un nuevo sistema corresponde a los llamados estudios *fuera de línea*, en los cuales se “alimenta” al sistema recomendador con datos históricos y se estudia su comportamiento, determinándose la exactitud de sus predicciones, la cobertura de tópicos y su rendimiento, entre otros.

Como contrapartida, en el campo del aprendizaje mejorado con la tecnología no se dispone de gran cantidad de datos históricos, por lo cual debe recurrirse a otras estrategias, como por ejemplo estudios sobre grupos de usuarios, y desarrollarse métodos específicos de evaluación para los sistemas recomendadores de este tipo.

El problema planteado aquí, que constituye el objetivo general del presente trabajo, es encontrar una metodología adecuada para realizar la evaluación de sistemas recomendadores de contenidos educativos.

Para la concreción de este objetivo general, se cumplen además con los siguientes objetivos específicos:

- Se presenta, el estado del arte de los sistemas recomendadores, y específicamente de contenidos educativos.
- Se identifican las distintas estrategias de evaluación que actualmente se están utilizando para evaluar sistemas recomendadores en general y, en particular, para los sistemas recomendadores de contenido educativo.
- Se exponen las limitaciones, beneficios y dificultades, de las diferentes estrategias de evaluación.
- Se determinan cuáles de estas formas de evaluación se adaptan más a los sistemas recomendadores de contenido educativo.
- Se proponen métricas específicas, relacionadas con las variables que se deben medir y se presenta una herramienta de medición, correspondiente a encuesta a usuarios, que cumple con los requisitos de validez y confiabilidad necesarios.
- Se propone una estrategia de evaluación que tiene en cuenta los diversos aspectos estudiados, y se la aplica para evaluar un sistema recomendador de contenidos educativos en particular.
- Por último, se extraen las conclusiones obtenidas y se sugieren nuevas líneas de investigación en esta área.

3.2. Justificación de la relevancia del tema

Los sistemas recomendadores de contenidos educativos poseen características que los diferencian de los sistemas utilizados para aplicaciones comerciales. Mientras que el objetivo principal de un sistema recomendador para comercio electrónico es proveer a los consumidores de información para ayudarlos a decidir qué productos comprar (*amazon.com, netflix.com, ebay.com*), el objetivo principal de un sistema recomendador de contenidos educativos en un entorno de aprendizaje mejorado con la tecnología, es proveer a los usuarios de actividades de aprendizaje, contenidos educativos, herramientas y relaciones con otras personas, adecuados para ayudarlos a desarrollar sus competencias.

Es importante destacar en este punto el enorme impacto que un sistema recomendador puede tener al ser aplicado al área educativa, dada la sensibilidad de este dominio. Una recomendación apropiada puede producir resultados positivos, en tanto que una inadecuada puede originar el efecto contrario. Desde el punto de vista del estudiante, la recomendación de un trayecto educativo (cursos a tomar) y de contenidos adecuados, pueden mejorar notablemente su desempeño y comprensión de un tema, pero la recomendación de trayectos o materiales para los cuales el estudiante no esté preparado para abordar, podrían conducir a su frustración y potencial fracaso. Resulta entonces de

suma importancia y surge como una necesidad la evaluación de los sistemas recomendadores empleados.

La presente propuesta comprende la evaluación de los sistemas recomendadores de contenidos educativos a través de estudios de usuarios, dado que existen diversas características de estos sistemas en particular, que los hacen más aptos para ser evaluados a través de este tipo de estudios, que a partir de otras técnicas.

Por un lado, de acuerdo con (Shani, 2011), los estudios de usuarios, llevados a cabo sobre pequeños grupos, tienen la gran ventaja de que “pueden responder el conjunto más amplio de cuestiones”, ya que permiten comprobar directamente cuánto influyen las recomendaciones del sistema en el comportamiento del usuario. Estos estudios son, además, el único tipo de evaluación que permite reunir a la vez datos cualitativos y cuantitativos, a partir del empleo de encuestas y del monitoreo de los usuarios mientras realizan tareas interactuando con el sistema recomendador.

Por otra parte, los sistemas recomendadores de contenidos educativos en el marco del aprendizaje formal, son aplicados a cursos con cantidades limitadas de estudiantes, alrededor de una treintena, resultando el ámbito ideal para efectuar estudios personalizados y exhaustivos sobre cada uno de los usuarios.

A las características descritas hasta aquí, se suma la actual falta de los elementos mínimos necesarios, correspondientes a un conjunto apropiado y validado de datos históricos, con un volumen de información suficiente como para encarar una evaluación fuera de línea de los sistemas recomendadores para aprendizaje mejorado con la tecnología (Drachsler, 2010).

Por último, puede hacerse un comentario adicional sobre la falta de propuestas con respecto a evaluaciones *en línea* (en funcionamiento) de los sistemas recomendadores para aprendizaje mejorado con la tecnología en entornos de educación formal. En este ámbito las cantidades de usuarios son relativamente pequeñas en comparación con la masividad de los entornos de comercio electrónico, en los cuales se aplican métodos de evaluaciones en línea, muchas veces derivando pequeños porcentajes del tráfico para interactuar con distintas interfaces y algoritmos, y medir los resultados obtenidos. Estas técnicas son costosas, de gran envergadura en cuanto a la cantidad de usuarios y recursos involucrados, y por el momento no resultan aplicables a la evaluación de los sistemas recomendadores referenciados en la presente propuesta.

Todos los factores reseñados anteriormente refuerzan la elección propuesta, consistente en el desarrollo de una metodología para la evaluación de sistemas recomendadores de contenido educativo en entornos de aprendizaje formal, a través de la realización de estudios de usuarios, que no requieren de datos históricos y pueden ser aplicados a diferentes contextos particulares de aprendizaje, cada uno con su propia problemática.

Las evaluaciones realizadas sobre el sistema recomendador permitirán conocer su rendimiento, el valor de las predicciones que devuelve, lo adaptable que resulta su interfaz, la rapidez con que entrega resultados, la confidencialidad y seguridad de uso y el grado de satisfacción del estudiante, entre otros aspectos. Estas evaluaciones son las herramientas de que se valdrán los desarrolladores para efectuar los ajustes necesarios tendientes a optimizar el funcionamiento del sistema recomendador, lo que justifica la relevancia del tema dentro de este dominio.

4. Solución propuesta

En el presente capítulo se presenta una propuesta metodológica para la evaluación de sistemas recomendadores de contenidos educativos, a partir de estudios de usuarios en experimentos controlados.

Con el propósito de brindar una metodología para tal evaluación, se presenta un sistema que cubre los siguientes aspectos:

- a) La definición de una herramienta, consistente en una encuesta aplicable, válida y confiable.
- b) La facilitación de una metodología que permita su aplicación así como la adquisición y almacenamiento de los datos obtenidos de manera simple y organizada.
- c) Una herramienta que permita el posterior análisis sistematizado de los datos de la encuesta, para determinar las fortalezas y debilidades del sistema recomendador evaluado.

4.1. Hacia una Metodología para la Evaluación de Sistemas Recomendadores de Contenidos Educativos

En la evaluación de sistemas recomendadores orientados al consumo, como ya se vio antes, es una práctica común utilizar conjuntos de datos históricos de dominio público, como los suministrados por *MovieLens*, *BookCrossing* y *EachMovie*, para probar nuevos algoritmos de recomendación, en evaluaciones *fuera de línea*.

Por otra parte, dentro de los entornos de aprendizaje mejorado con la tecnología (TEL), ya sea formales o no formales, cada uno tiene sus propias características, modelos, tareas para recomendar y algoritmos de recomendación, pero no existen conjuntos adecuados de datos históricos de disposición pública a los que se pueda recurrir, para efectuar evaluaciones fuera de línea.

4.1.1. Consideraciones para la Evaluación Fuera de Línea

De acuerdo con Drachsler, (Drachsler, 2010) los resultados acerca del rendimiento de diferentes sistemas recomendadores en los entornos TEL son difíciles de comparar, y no existe un conocimiento válido acerca de qué tipo de algoritmo y sistema de recomendación puede ser aplicado con éxito en cada contexto de aprendizaje. Dado que la validez y generalización de un experimento científico depende de la posibilidad de verificación, repetición y comparación de los resultados, este autor propone la necesidad de comenzar por elaborar una colección pública de conjuntos de datos que puedan ser utilizados para evaluar sistemas recomendadores en los entornos de aprendizaje mejorado con la tecnología.

Esta propuesta debe, necesariamente, contar con el apoyo de la comunidad de investigadores para desarrollar un conjunto de datos adecuados provenientes de diferentes campos de investigación en TEL, con una estructura de metadata apropiada, procedimientos de evaluación normalizados aplicables a los sistemas de recomendación que sean probados utilizando estos datos, y políticas de protección legal de derechos

sobre los datos que permitan que los mismos sean compartidos por todos los investigadores.

El camino planteado por esta propuesta está aún en sus inicios pero, evidentemente, llevará mucho tiempo llegar a un acuerdo en cuanto a todos los puntos que deberán ser tenidos en cuenta para lograr la preparación de una base de datos históricos, que surja de un proyecto de cooperación científica internacional, y pueda ser utilizada ampliamente en estudios de evaluaciones fuera de línea.

En esta línea se ha realizado recientemente uno de los primeros aportes, Verbert (2011), utilizando datasets provistos por Mendeley, APOSDLE, ReMashed, Organic Edunet, Mace y Travel well, los cuales capturan las interacciones de los usuarios con herramientas y recursos, para evaluar y comparar el rendimiento de distintos algoritmos de recomendación para aprendizaje mejorado con la tecnología (TEL). Estos autores presentaron una comparación experimental de la exactitud de diversos algoritmos de filtrado colaborativo (basados en usuarios y basados en ítems) elaborada a partir de datos implícitos tales como tags y downloads, aplicados a estos datasets.

4.1.2. Consideraciones para la Evaluación a través de Estudio de Usuarios

De acuerdo con lo expuesto hasta aquí, la falta de datos históricos, las cantidades de usuarios involucrados y dadas las ventajas reseñadas en el capítulo anterior (3.2), nuestra propuesta comprende la evaluación de los sistemas recomendadores de contenidos educativos a través de estudios de usuarios, debido a las características propias de estos sistemas en particular, que los hacen más aptos para ser evaluados a través de este tipo de estudios, que a partir de otras técnicas.

4.2. Estrategia para la Evaluación de Sistemas Recomendadores de Contenidos Educativos en Entornos de Educación Formal

4.2.1. Generalidades

La estrategia propuesta en el presente trabajo corresponde a la evaluación de sistemas recomendadores de contenidos educativos utilizados en entornos de educación formal, principalmente de nivel terciario y universitario. En estas situaciones, la elección de la evaluación a través de estudios de usuarios parece ser la más adecuada tanto para las etapas de desarrollo, prueba y refinamiento de los sistemas recomendadores, como así también, una vez implementado el sistema, para su posterior seguimiento y mejora.

De acuerdo a lo expuesto anteriormente, esto es así debido, entre otras consideraciones, a que la cantidad de usuarios es más limitada que en entornos de aprendizaje informal, de acceso más masivo, y los métodos de evaluación propuestos pueden aplicarse a la gran mayoría de los usuarios del sistema, obteniéndose resultados más completos que en aquellos casos en que deban tomarse muestras, las cuales podrían resultar sesgadas de acuerdo al criterio de selección utilizado.

4.2.2. Requerimientos a cumplir

La metodología para la evaluación de estos sistemas recomendadores debe tener características y cumplir con ciertos requerimientos que permitan su generalización para ser aplicada en distintos entornos de aprendizaje mejorado con la tecnología. Entre los aspectos más importantes, la metodología debe ser:

Clara (sin ambigüedades)
Aplicable
Reproducible (que sea repetitiva)
Verificable
Comparable

Además esta metodología debe reflejar las particularidades de los sistemas recomendadores de contenido educativo, identificando qué es lo que se está buscando al hacer recomendaciones de objetos de aprendizaje.

4.3. Metodología para la Evaluación de Sistemas Recomendadores de Contenidos Educativos a través de Estudios de Usuarios

La metodología para la evaluación que se presenta a continuación, está diseñada para ser aplicable a pequeños grupos de usuarios dentro de entornos controlados de aprendizaje mejorado con la tecnología, en sistemas de educación formal.

Pretende cumplir con los requerimientos que permitan su generalización para ser utilizada en distintos entornos particulares de aprendizaje mejorado con la tecnología, en cuanto a claridad, aplicabilidad, reproducibilidad, verificabilidad y comparabilidad.

En su forma más ambiciosa, esta propuesta metodológica contempla la realización de mediciones cuantitativas y cualitativas sobre distintos atributos del sistema recomendador.

En principio esta metodología propone incorporar la evaluación de aspectos técnicos y educativos (Drachler, 2009), como así también aspectos centrados en el usuario, a partir de la aplicación de un cuestionario, que posibilite determinar la calidad percibida por el usuario acerca de las recomendaciones que recibió (Pu, 2010), (Chen, 2010). Dentro de esta estrategia se evalúan también características derivadas de marcos evaluativos tradicionales (Manouselis, 2011) en particular aquellas tenidas en cuenta en el modelo clásico de Kirkpatrick (Kirkpatrick, 1959).

Así, la metodología propuesta para la evaluación, en su aspecto más amplio, comprende dos divisiones fundamentales: por un lado se efectúan mediciones que surgen de las interacciones de los usuarios con el sistema recomendador que pueden ser cuantificadas de manera objetiva (mediciones sobre el sistema) y, por otra parte, se aplica una encuesta para medir aquellos aspectos más subjetivos, que tienen que ver con las opiniones del usuario referentes a diversas cuestiones relativas al trabajo con el sistema recomendador (mediciones sobre la encuesta).

4.3.1. Propuesta de Variables a Medir

En la elección de las variables a medir de acuerdo con la presente propuesta, se tiene en cuenta sintetizar los aspectos recomendados para su evaluación por los autores mencionados en el punto anterior (4.3), así como incorporar características que consideramos relevantes para lograr una correcta evaluación del sistema recomendador.

Las variables a medir abarcan tres grupos: mediciones técnicas, mediciones educativas y mediciones centradas en el usuario. Los grupos de mediciones técnicas y de mediciones educativas tienen en cuenta la propuesta del framework de evaluación de SR en TEL efectuada por Drachsler (Drachsler, 2009).

Dentro de las mediciones centradas en el usuario, el grupo de mediciones correspondiente al nivel general, tiene en cuenta las recomendaciones de Manouselis (Manouselis, 2011), en lo relativo a la incorporación de marcos evaluativos tradicionales como el de Kirkpatrick (Kirkpatrick, 1959).

Dentro del mismo grupo anterior, las mediciones correspondientes al nivel específico, incluyen los aspectos recomendados por Chen (Chen, 2010) y por Pu (Pu, 2010) en la propuesta de su modelo ResQue (Recommender systems' Quality of user experience).

A continuación se describen los grupos de mediciones a efectuar y las variables a medir correspondientes a cada uno de ellos.

Mediciones técnicas

-Exactitud

Cantidad de actividades propuestas por el sistema y aceptadas por el usuario en relación con la cantidad de actividades propuestas por el sistema y no aceptadas por el usuario.

-Cobertura

Cantidad de actividades realizadas por el usuario y no propuestas por el sistema.

-Rendimiento

Rapidez con que el sistema entrega las recomendaciones.

Mediciones Educativas

-Efectividad

Cantidad de actividades de aprendizaje vistas, estudiadas o completadas durante la fase de aprendizaje.

-Eficiencia

Tiempo que los usuarios necesitaron para cumplir con el objetivo de aprendizaje.

-Satisfacción

Grado de satisfacción, acuerdo y agrado del usuario con las recomendaciones recibidas del sistema.

-Abandono

Proporción de estudiantes que abandonaron la actividad durante la fase de aprendizaje.

Mediciones Centradas en el Usuario

Nivel General:

-Reacción

Lo que el usuario piensa y siente (nivel de agrado/malestar) en relación con las recomendaciones recibidas.

-Aprendizaje

En qué medida el usuario aprendió gracias a las recomendaciones recibidas.

-Comportamiento

En qué medida el usuario piensa que podrá aplicar en su actividad las nuevas ideas e informaciones recomendadas.

-Resultados

La mejora percibida por el usuario en sus aprendizajes, a partir de las recomendaciones recibidas del sistema.

Nivel Específico:

-Calidad de las recomendaciones

Cómo percibe el usuario la calidad de las recomendaciones recibidas en cuanto a exactitud, novedad, atractivo, diversidad y contextualización.

-Interacción adecuada

Grado de comunicación del sistema con el usuario y explicación acerca de las recomendaciones propuestas.

-Interfaz adecuada y clara

Agrado del usuario con respecto a la interfaz, visibilidad y claridad.

-Facilidad de uso

Grado de comodidad del usuario para interactuar en forma rápida y correcta con el sistema recomendador. Facilidad de aprendizaje de uso.

-Utilidad de las recomendaciones

Medida en que el usuario percibe que el uso del sistema recomendador va a contribuir a mejorar su rendimiento.

-Control

Medida en que el usuario puede revisar su historial de recomendaciones, aceptar, rechazar o pedir una nueva recomendación.

-Transparencia

Medida en que el sistema le permite al usuario entender su lógica interna y le explica el porqué de sus recomendaciones.

-Satisfacción

Nivel de satisfacción general del usuario al utilizar el sistema recomendador.

-Confiabilidad

Nivel de confianza y veracidad que el usuario siente por las recomendaciones que entrega el sistema.

-Actitud respecto al Sistema

Indica si el usuario volvería a utilizar el sistema recomendador, o si lo recomendaría.

4.3.2. Dimensiones a Evaluar por Grupos y Divisiones

En la siguiente tabla, se describen las variables que se propone medir dentro de sus grupos de pertenencia. Se indican además cuáles de estas variables deben medirse directamente a partir de los datos suministrados por el sistema de aprendizaje mejorado con la tecnología, y cuáles de ellas se recaban a partir de una encuesta a usuarios.

| MEDICIONES | Técnicas | Educativas | Centradas en el Usuario |
|-------------------|---------------------------------------|---------------------------------------|---|
| Sobre el Sistema | Exactitud Cobertura Rendimiento | Efectividad Eficiencia Abandono | |
| Sobre la Encuesta | | Satisfacción | Reacción Aprendizaje Comportamiento Resultados Calidad Interacción Interfaz Facilidad de uso Utilidad Control Transparencia Satisfacción Confiabilidad Actitud |

Tabla 4: Dimensiones a evaluar por grupos y divisiones

4.3.3. Propuesta de la Evaluación Específica a Realizar

En el presente trabajo, nuestra acción está principalmente orientada a la evaluación del sistema recomendador en los temas relacionados con las variables registradas a través de los datos recabados a partir de la encuesta a los usuarios, con lo cual nos interesa indagar directamente sobre las mediciones centradas en el usuario.

Estas comprenden aspectos sutiles y subjetivos, que dependen de cada persona y tienen que ver en gran medida con la confianza (Chen, 2010) que el usuario sienta hacia el sistema recomendador y las recomendaciones recibidas, y el grado de satisfacción alcanzado al haberlo utilizado.

Estos aspectos son los que determinarán en última instancia la aceptación o no de un determinado sistema recomendador y el consecuente éxito o fracaso en su utilización.

Es importante destacar que la evaluación sobre la encuesta centrada en los usuarios resulta perfectamente aplicable aun cuando el evaluador no tenga acceso al funcionamiento interno del sistema de recomendación, y cubre un amplio espectro de sistemas de recomendación educativos que pueden ser evaluados con este método.

No obstante se propone para trabajos posteriores, incorporar los grupos de mediciones técnicas y educativas que se realizan sobre el sistema, las cuales no fueron abordadas aquí. Para cumplir esta acción será necesario trabajar en colaboración estrecha con los desarrolladores del sistema recomendador a evaluar, para implementar herramientas internas del sistema que permitan efectuar este tipo de medidas.

4.4. Encuesta a Usuarios

A continuación se presentan las preguntas / frases correspondientes a la encuesta a aplicar sobre los usuarios del sistema recomendador. Cada una de estas preguntas debe ser contestada utilizando una escala de Likert con valores de 1 a 5 (Marradi, 2007), en la cual el valor 1 indica “completamente en desacuerdo” y el valor 5 indica “completamente de acuerdo”. Aquellas preguntas que no sean evaluadas se consideran como respondidas con “No sabe / No contesta / No aplica”.

Nivel General

Reacción

- 1) Las recomendaciones recibidas resultaron de mi agrado.

Aprendizaje

- 2) Las recomendaciones recibidas me ayudaron a obtener nuevas ideas y alcanzar los aprendizajes que necesitaba.

Comportamiento

- 3) Estimo que podré aplicar en mi actividad la información e ideas logradas a partir de las recomendaciones que recibí.

Resultados

- 4) Las recomendaciones recibidas me ayudaron a mejorar mi efectividad y los resultados de mi aprendizaje.

Nivel Específico

Calidad de las recomendaciones

- 5) Las recomendaciones recibidas se ajustaron exactamente a mis necesidades.
- 6) Las recomendaciones de mis compañeros resultaron mejores que las del sistema.
- 7) Las recomendaciones me resultaron novedosas e interesantes.
- 8) Las recomendaciones fueron demasiado generales y poco específicas.
- 9) Las recomendaciones estaban fuera de contexto.

Interacción adecuada

- 10) El sistema recomendador me dio explicaciones claras acerca de las recomendaciones y porqué las realizaba.

Interfaz adecuada

- 11) La interfaz del sistema es atractiva, clara y adecuada.

Facilidad de uso

- 12) El sistema recomendador me resultó fácil de utilizar.
- 13) Me resultó fácil comunicar mis preferencias al sistema recomendador.

Utilidad de las recomendaciones

- 14) Las recomendaciones propuestas me resultaron útiles para lograr la comprensión de los temas estudiados.
- 15) Las recomendaciones recibidas me ayudaron a mejorar mi rendimiento.

Control

- 16) El sistema recomendador presentaba el historial de recomendaciones.
- 17) El sistema permitía aceptar o rechazar fácilmente una recomendación.

Transparencia

- 18) Sentí que el sistema recomendador reflejaba correctamente los cambios en mis preferencias.

Satisfacción

- 19) En general estuve de acuerdo con las recomendaciones recibidas del sistema.

Confianza

- 20) Sentí confianza acerca de las recomendaciones que recibí del sistema.
- 21) Las recomendaciones recibidas me parecieron acertadas.
- 22) Las recomendaciones recibidas del sistema a veces me confundieron.

Actitud respecto al sistema

- 23) Si se diera la oportunidad, volvería a utilizar nuevamente este sistema recomendador.
- 24) Recomendaría este sistema recomendador a mis amigos.

Observaciones

Notas o comentarios libres que el usuario del sistema recomendador desee realizar.

Dentro de las frases formuladas en la encuesta, existen varias que corresponden a una escala de Likert inversa, es decir una alta puntuación corresponde a desacuerdo con el sistema recomendador, es el caso, por ejemplo, de la frase 9: “Las recomendaciones estaban fuera de contexto”. Esto apunta a intentar minimizar ciertos desvíos técnicos asociados a este tipo de encuestas y marcados por Archenti (Marradi, 2007).

4.4.1. Formulario de la Encuesta a Usuarios del Sistema Recomendador

A continuación se presenta la encuesta diseñada, con el formato propuesto, para ser distribuida a los usuarios del sistema recomendador a evaluar.

Encuesta a Usuarios del Sistema Recomendador

A partir de su interacción con este sistema recomendador, le agradeceremos completar la siguiente encuesta, que nos permitirá conocer sus opiniones las cuales nos serán de mucha utilidad para evaluar el sistema.

Por favor para cada uno de los tópicos indicados a continuación, marque con una cruz en la casilla correspondiente al valor que usted considere más adecuado, teniendo en cuenta que:

El valor 1 indica “completamente en desacuerdo” y el valor 5 indica “completamente de acuerdo” con la descripción dada.

¡Muchas gracias!

| IT | TOPICO | Evaluación | | | | |
|----|--|-----------------|---|---|---|---------|
| | | Desacuerdo..... | | | | Acuerdo |
| N° | Descripción | 1 | 2 | 3 | 4 | 5 |
| 1 | Las recomendaciones recibidas resultaron de mi agrado. | | | | | |
| 2 | Las recomendaciones recibidas me ayudaron a obtener nuevas ideas y alcanzar los aprendizajes que necesitaba. | | | | | |
| 3 | Estimo que podré aplicar en mi actividad la información e ideas logradas a partir de las recomendaciones que recibí. | | | | | |
| 4 | Las recomendaciones recibidas me ayudaron a mejorar mi efectividad y los resultados de mi aprendizaje. | | | | | |
| 5 | Las recomendaciones recibidas se ajustaron exactamente a mis necesidades. | | | | | |
| 6 | Las recomendaciones de mis compañeros resultaron mejores que las del sistema. | | | | | |
| 7 | Las recomendaciones me resultaron novedosas e interesantes. | | | | | |
| 8 | Las recomendaciones fueron demasiado generales y poco específicas. | | | | | |
| 9 | Las recomendaciones estaban fuera de contexto. | | | | | |
| 10 | El sistema recomendador me dio explicaciones claras acerca de las recomendaciones y porqué las realizaba. | | | | | |
| 11 | La interfaz del sistema es atractiva, clara y adecuada. | | | | | |
| 12 | El sistema recomendador me resultó fácil de utilizar. | | | | | |
| 13 | Me resultó fácil comunicar mis preferencias al sistema recomendador. | | | | | |

| | | | | | | |
|----|---|--|--|--|--|--|
| 14 | Las recomendaciones propuestas me resultaron útiles para lograr la comprensión de los temas estudiados. | | | | | |
| 15 | Las recomendaciones recibidas me ayudaron a mejorar mi rendimiento. | | | | | |
| 16 | El sistema recomendador presentaba el historial de recomendaciones. | | | | | |
| 17 | El sistema permitía aceptar o rechazar fácilmente una recomendación. | | | | | |
| 18 | Sentí que el sistema recomendador reflejaba correctamente los cambios en mis preferencias. | | | | | |
| 19 | En general estuve de acuerdo con las recomendaciones recibidas del sistema. | | | | | |
| 20 | Sentí confianza acerca de las recomendaciones que recibí del sistema. | | | | | |
| 21 | Las recomendaciones recibidas me parecieron acertadas. | | | | | |
| 22 | Las recomendaciones recibidas del sistema a veces me confundieron. | | | | | |
| 23 | Si se diera la oportunidad, volvería a utilizar nuevamente este sistema recomendador. | | | | | |
| 24 | Recomendaría este sistema recomendador a mis amigos. | | | | | |

Observaciones

Notas o comentarios que desee realizar:

4.5. Categorías definidas para la evaluación del Sistema Recomendador

Para poder brindar información que resulte de utilidad, tanto para evaluar al sistema recomendador, como para hallar sus fortalezas y debilidades, las preguntas formuladas en la encuesta fueron agrupadas en una cantidad más limitada de categorías, cinco en este caso, que cubren los aspectos principales a ser tenidos en cuenta, comprendiendo:

- Interacción e interfaz
- Calidad y Novedad
- Confianza
- Utilidad y Relevancia
- Satisfacción

En la siguiente tabla se detallan los ítems de la encuesta relacionados con cada una de las categorías definidas y a continuación se presenta una breve reseña de las mismas.

| Categoría | Ítems de la encuesta bajo esta categoría |
|---------------------------------|--|
| Interacción e Interfaz | 10, 11, 12, 13, 16, 17 |
| Calidad / Novedad | 05, 06, 07, 08, 09 |
| Confianza | 18, 20, 21, 22 |
| Utilidad percibida / Relevancia | 02, 03, 04, 14, 15 |
| Satisfacción | 01, 19, 23, 24 |

Tabla 5: *Ítems de la encuesta a usuarios bajo cada categoría*

Interacción e Interfaz

En esta categoría se tienen en cuenta los aspectos relativos a la forma como interactúa el usuario con el sistema recomendador, la claridad y adecuación de la interfaz, facilidad de uso, comodidad con que el usuario puede indicar sus preferencias al sistema, aceptar o rechazar las recomendaciones ofrecidas, claridad con que el sistema presenta sus recomendaciones, el historial y explica el porqué de las mismas.

Calidad / Novedad

En esta categoría se incluyen los aspectos percibidos por el usuario en cuanto a la exactitud, novedad, atractivo, diversidad y contextualización de las recomendaciones.

Confianza

Esta categoría tiene en cuenta el grado de confianza y veracidad que el usuario siente en cuanto a las recomendaciones recibidas, así como la transparencia con que el sistema recomendador refleja correctamente los cambios en las preferencias del usuario.

Utilidad / Relevancia

Esta categoría corresponde a la medida en que el usuario percibe que las recomendaciones del sistema contribuyen a mejorar su rendimiento, obtener ideas para alcanzar y mejorar la efectividad durante el aprendizaje, mejorar los resultados del aprendizaje y la comprensión de los temas estudiados y poder aplicar las ideas obtenidas en su actividad.

Satisfacción

En esta categoría se mide el nivel general de satisfacción que siente el usuario al utilizar el sistema recomendador, su nivel de agrado y acuerdo hacia las recomendaciones recibidas y su actitud futura en cuanto al uso, o recomendación por su parte, del sistema recomendador utilizado.

4.6. Implementación de la Encuesta

4.6.1. Adquisición y almacenamiento de los datos

La encuesta a usuarios, desarrollada en el presente trabajo, se implementa a través de la utilización de un formulario en línea elaborado con el sistema de documentos en la nube de Google. La encuesta fue creada con “Google Forms”.

Esta aplicación de Google brinda un formulario en el cual presentar la encuesta y a la vez se convierte en un sistema poderoso para la aplicación de la misma. La encuesta se ofrece al usuario del sistema recomendador que se quiere evaluar a través de un enlace dentro del mismo sistema. Cada usuario tiene la posibilidad de acceder a la misma con un simple click del mouse en este enlace.

Una vez que el usuario completa y envía la encuesta, los resultados se almacenan automáticamente en una hoja de cálculo de Google en la que aparecen la fecha y hora de envío y los valores de cada respuesta, en un formato de columnas (una por cada ítem) y con una fila para cada encuesta completada por un usuario.

Esta hoja de cálculo reside en el sistema de almacenamiento en la nube “Google Drive” y se actualiza de manera automática, agregándose una nueva fila cada vez que ingresa una nueva encuesta completada.

Una vez cumplido un plazo determinado u obtenidas una determinada cantidad de respuestas a la encuesta, la hoja de cálculo puede descargarse en un formato estándar, por ejemplo Microsoft Excel, para efectuar los correspondientes análisis sobre los datos adquiridos. Esto no inhibe la adquisición de datos sobre la hoja de cálculo de Google, que seguirá almacenando las nuevas respuestas de usuarios, las cuales son fácilmente identificables utilizando los rótulos de fecha y hora de envío con la que comienza cada fila.

4.6.2. Análisis de los Datos Adquiridos

El análisis de los datos se automatiza a través de un sistema de cálculo basado en Microsoft Excel (Bott, 1999), que toma los datos de las encuestas respondidas a partir de la hoja de cálculo de Google.

Teniendo en cuenta los valores asignados por los usuarios a cada ítem de la encuesta, la planilla de cálculo ajusta aquellos ítems que tengan escalas Likert inversas y luego realiza un histograma de frecuencias. De esta manera se determina para cada una de las preguntas formuladas, la cantidad de usuarios que asignó cada valor, por ejemplo, para el ítem 16 de la encuesta se determina cuántos usuarios le asignaron “1”, cuántos le asignaron “2”, y así siguiendo hasta “5”.

Utilizando estos valores, en la planilla de cálculo Excel se agrupan los ítems por categorías y para cada una de las categorías se suman las cantidades de respuestas correspondientes a cada valor, provenientes de los ítems que conforman esa categoría, para realizar un histograma de frecuencias por categoría. Por ejemplo, la categoría “Confianza”, está conformada por los ítems de la encuesta números 18, 20, 21 y 22, por lo que se suman para dichos ítems las cantidades de valores “1” obtenidos, las cantidades de “2” obtenidos, y así siguiendo hasta “5”, para realizar un histograma de frecuencias para esa categoría.

Finalmente, el puntaje que se asigna a cada categoría corresponde a la moda, o valor con mayor frecuencia en la distribución de los datos, dentro de la categoría considerada. A partir de estos valores, la planilla Excel presenta los gráficos correspondientes a los histogramas obtenidos, para cada una de las categorías estudiadas.

Se destaca la facilidad de uso de este sistema de análisis de datos, ya que basta copiar y pegar en la planilla de cálculo Excel el bloque de datos recogidos en la planilla generada por Google Forms, para obtener resultados instantáneos.

El sistema aquí presentado permite de esta manera la automatización, tanto de la adquisición de los datos, como del análisis de los mismos, ofreciendo rápidamente resultados que posibilitan una primera evaluación del sistema de recomendación a partir de los gráficos generados.

4.6.3. Propuestas para Mejorar el Sistema Recomendador Evaluado

Los resultados obtenidos una vez analizados los datos de la encuesta, más allá de proporcionar una evaluación del sistema recomendador, pueden ser utilizados además para encontrar debilidades y fortalezas del mismo y, eventualmente, brindar recomendaciones para mejorar el sistema evaluado en aquellas categorías con bajos puntajes. Incluso se pueden detectar aspectos puntuales a mejorar dentro de cada categoría, según lo revelado por determinadas preguntas aplicadas para esa categoría dentro de la encuesta.

Si bien este tópico no está cubierto en el presente trabajo, se presenta como propuesta para investigar y desarrollar en trabajos futuros.

4.7. Validación a través de una Encuesta a Expertos

Como elemento de validación de la herramienta de evaluación presentada, se diseñó una herramienta adicional consistente en una encuesta a expertos. Esta encuesta está orientada para ser respondida por expertos docentes y consiste en una serie más limitada de preguntas.

El experto no debe necesariamente interactuar con el sistema recomendador, sino que responde la encuesta a partir del análisis que realiza acerca de las recomendaciones efectuadas por el sistema a los distintos estudiantes, y los datos proporcionados para diferentes momentos y usuarios (alumnos) utilizando las “fotos” del sistema recomendador, *ego network de radio 1* (ver Anexo A).

4.7.1. Formulario de la Encuesta a Expertos

A continuación se presenta la encuesta diseñada, con el formato propuesto, para ser distribuida a los expertos.

Encuesta a Expertos para el Sistema Recomendador a Evaluar

A partir de su conocimiento experto, le agradeceremos completar la siguiente encuesta para evaluar este sistema recomendador. Sus opiniones nos serán de mucha utilidad.

Por favor para cada uno de los tópicos indicados a continuación, marque la casilla correspondiente al valor que usted considere más adecuado, teniendo en cuenta que: El valor 1 indica “completamente en desacuerdo” y el valor 5 indica “completamente de acuerdo” con la descripción dada.

¡Muchas gracias!

| IT | TOPICO | Evaluación | | | | |
|----|---|------------------------|---|---|---|---|
| | | Desacuerdo.....Acuerdo | | | | |
| N° | Descripción | 1 | 2 | 3 | 4 | 5 |
| 1 | Las recomendaciones ofrecidas son relevantes para alcanzar los aprendizajes. | | | | | |
| 2 | Las recomendaciones ofrecidas resultan aplicables a las tareas del estudiante. | | | | | |
| 3 | Las recomendaciones ofrecidas se ajustan a las necesidades del estudiante, teniendo en cuenta los contenidos a los que accedió. | | | | | |
| 4 | Las recomendaciones ofrecidas brindan diversidad en sus contenidos. | | | | | |
| 5 | Las recomendaciones ofrecidas resultan nuevas para el estudiante. | | | | | |
| 6 | Las recomendaciones ofrecidas no se encuentran personalizadas. | | | | | |
| 7 | Las recomendaciones ofrecidas están fuera del contexto de los temas en estudio. | | | | | |
| 8 | Las recomendaciones ofrecidas resultan útiles para lograr la comprensión de los temas en estudio. | | | | | |
| 9 | Las recomendaciones ofrecidas reflejan los cambios en las preferencias del estudiante. | | | | | |
| 10 | Las recomendaciones ofrecidas resultan acertadas. | | | | | |
| 11 | No queda claro porqué el sistema efectúa ciertas recomendaciones. | | | | | |

Observaciones

Notas o comentarios que desee realizar:

4.7.2. Categorías Cubiertas en la Encuesta a Expertos

En la siguiente tabla se detallan los ítems de la encuesta a expertos relacionados con cada una de las categorías definidas anteriormente.

| Categoría | Ítems de la encuesta bajo esta categoría |
|---------------------------------|--|
| Interacción e Interfaz | -- |
| Calidad / Novedad | 03, 04, 05, 06, 07 |
| Confianza | 09, 10, 11 |
| Utilidad percibida / Relevancia | 01, 02, 08 |
| Satisfacción | -- |

Tabla 6: *Ítems de la encuesta a expertos bajo cada categoría*

Como se puede apreciar, la encuesta a expertos no brinda validación en las categorías de Interacción e Interfaz del sistema recomendador ni tampoco acerca de la Satisfacción del usuario. En el primer caso dado que, como se indicó anteriormente, el experto no necesariamente interactúa con el sistema, y en el segundo caso debido a que la Satisfacción del usuario resulta una apreciación completamente personal del estudiante que utilizó el sistema recomendador.

5. Caso de aplicación

Con el objeto de realizar un sondeo del instrumento de evaluación propuesto, se efectuó una prueba, aplicando la encuesta a un grupo de usuarios de un sistema recomendador de contenidos educativos.

5.1. Descripción del Sistema Recomendador Evaluado

La encuesta se aplicó sobre un grupo de usuarios del sistema de recomendación de contenidos educativos, que funciona sobre una plataforma Moodle utilizada en la Facultad de Ingeniería de la Universidad Nacional de La Plata.

El sistema recomendador evaluado toma información de fuentes heterogéneas, construye un gráfico y aplica una variación del algoritmo PageRank (Brin, 1998) para computar las recomendaciones ofrecidas. El mismo se encuentra inmerso en un sistema de aprendizaje mejorado con la tecnología (TEL) consistente en una plataforma Moodle LMS y un repositorio de objetos de aprendizaje implementado sobre una wiki semántica.

Esta plataforma Moodle maneja tres cursos de grado diferentes: “Orientación a Objetos 1”, “Orientación a Objetos 2” y “Diseño y Arquitectura de Software”. En todos ellos se desarrollan distintos tópicos dentro del área de programación orientada a objetos.

Durante el cuatrimestre en que se realizó la evaluación del sistema recomendador, en la plataforma Moodle se encontraban registrados 332 usuarios, de los cuales 304 eran estudiantes y 28 profesores. Sus publicaciones incluían 75 recursos propios (archivos pdf, presentaciones, etc.) con más de 13000 vistas, y 108 hilos de discusión con más de 8000 visitas.

El repositorio de objetos de aprendizaje contaba en el mismo período con 178 recursos relacionados con los cursos en Moodle, sumando alrededor de 280 vistas, 580 ediciones y 15 calificaciones.

En el Anexo B se presentan, a modo de ejemplo, algunas capturas de pantalla correspondientes a esta plataforma.

5.2. Aplicación de la Encuesta a Usuarios

La prueba de evaluación se llevó a cabo sobre un grupo de usuarios del sistema de recomendación de contenidos educativos, que funciona sobre una plataforma Moodle utilizada en la asignatura “Orientación a Objetos 2”, de Ingeniería Informática de la Universidad Nacional de La Plata, el cual fue descrito anteriormente.

Una vez procesados los datos de la encuesta, los puntajes asignados por los usuarios (p) a las preguntas con escalas Likert inversas (números 6, 8, 9 y 22), fueron ajustados de la siguiente manera:

$$p' = 6 - p$$

de modo que un valor mínimo ($p = 1$) en esas respuestas correspondieran a un valor máximo ($p' = 5$) y consecuentemente con el resto de los puntajes.

Posteriormente, para cada uno de los ítems de la encuesta, se determinó la frecuencia correspondiente a cada uno de los valores asignados por los usuarios. Por ejemplo, para el ítem número 1 de la encuesta, la cantidad de usuarios que lo evaluaron con “1”, la cantidad de usuarios que lo evaluaron con “2”, etc.

Así se obtuvieron los valores indicados en la tabla siguiente, los que permiten efectuar un histograma para cada ítem de la encuesta.

| Frecuencia de los Ítems 1 a 12 | | | | | | | | | | | | |
|--------------------------------|---|---|---|---|---|---|---|---|---|----|----|----|
| Valor | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | 0 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 6 | 0 | 0 |
| 2 | 2 | 2 | 1 | 3 | 4 | 4 | 2 | 3 | 2 | 1 | 1 | 1 |
| 3 | 3 | 2 | 2 | 4 | 3 | 3 | 5 | 4 | 2 | 3 | 2 | 1 |
| 4 | 2 | 5 | 4 | 3 | 3 | 3 | 1 | 3 | 1 | 2 | 5 | 2 |
| 5 | 5 | 2 | 4 | 1 | 1 | 0 | 3 | 1 | 5 | 0 | 4 | 8 |

| Frecuencia de los Ítems 13 a 24 | | | | | | | | | | | | |
|---------------------------------|----|----|----|----|----|----|----|----|----|----|----|----|
| Valor | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 1 | 2 | 1 | 1 | 5 | 3 | 4 | 1 | 0 | 2 | 1 | 2 | 3 |
| 2 | 2 | 2 | 3 | 2 | 1 | 2 | 3 | 3 | 2 | 2 | 2 | 2 |
| 3 | 6 | 1 | 4 | 3 | 5 | 3 | 3 | 2 | 3 | 3 | 1 | 2 |
| 4 | 2 | 3 | 1 | 1 | 1 | 2 | 3 | 3 | 2 | 4 | 1 | 1 |
| 5 | 0 | 5 | 3 | 1 | 2 | 1 | 2 | 4 | 3 | 2 | 6 | 4 |

Tabla 7: Frecuencia de los ítems en la encuesta a usuarios

A partir de la categorización definida en (4.5), se determinó para cada categoría la frecuencia de cada puntaje asignado por los usuarios. Para ello se sumaron las ocurrencias de cada valor dentro de los ítems correspondientes a cada categoría. Por ejemplo, para la categoría *Interacción e Interfaz*, se sumaron todas las ocurrencias del valor 1 dentro de los ítems 10, 11, 12, 13, 16 y 17 de la encuesta, procediendo de manera análoga con los demás puntajes del 2 al 5.

De esta forma se obtuvieron los valores indicados en la tabla siguiente, los que permiten efectuar un histograma para cada categoría definida a partir de la encuesta.

| Frecuencias por Categoría y General | | | | | | |
|-------------------------------------|------------------------|-------------------|-----------|-----------------------|--------------|---------------|
| Valor | Interacción e Interfaz | Calidad / Novedad | Confianza | Utilidad / Relevancia | Satisfacción | Total General |
| 1 | 16 | 7 | 7 | 5 | 6 | 41 |
| 2 | 8 | 15 | 9 | 11 | 9 | 52 |
| 3 | 20 | 17 | 11 | 13 | 9 | 70 |
| 4 | 13 | 11 | 11 | 16 | 7 | 58 |
| 5 | 15 | 10 | 10 | 15 | 17 | 67 |

Tabla 8: Frecuencia por categoría en la encuesta a usuarios

5.3. Análisis de los Datos Obtenidos en la Encuesta a Usuarios

Para el análisis de los datos se realizaron, a partir de los valores adquiridos, los gráficos correspondientes a los histogramas obtenidos para cada una de las categorías estudiadas, los cuales se presentan a continuación.

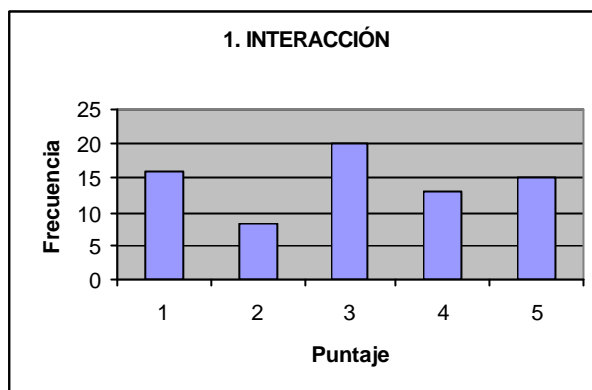


Ilustración 1: Histograma de valores para la categoría Interacción e Interfaz

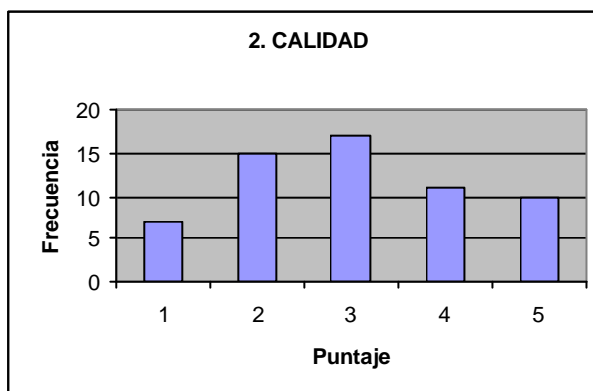


Ilustración 2: Histograma de valores para la categoría Calidad/Novedad

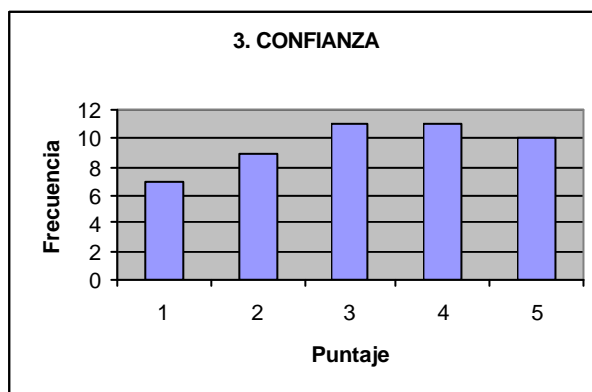


Ilustración 3: Histograma de valores para la categoría Confianza

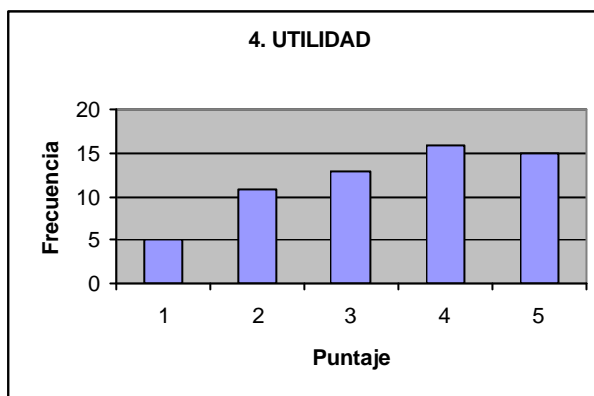


Ilustración 4: Histograma de valores para la categoría Utilidad / Relevancia

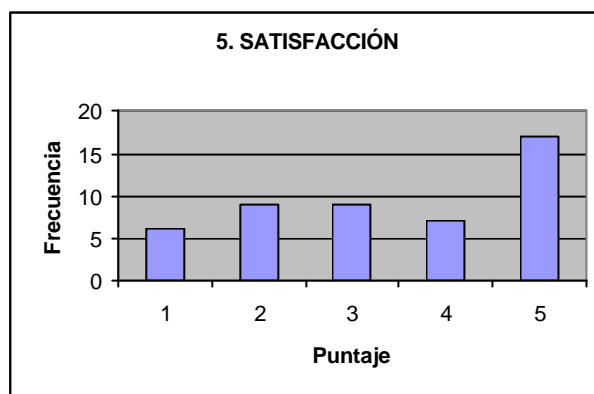


Ilustración 5: Histograma de valores para la categoría Satisfacción

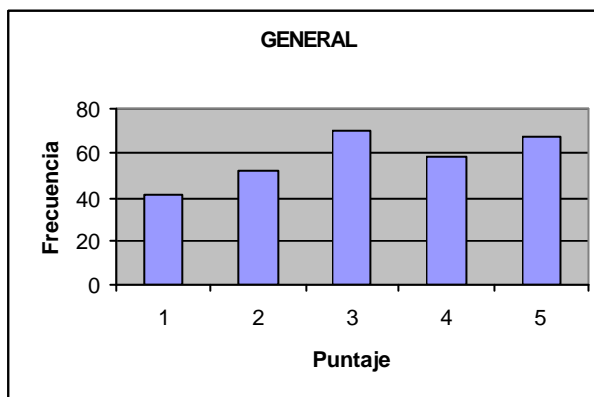


Ilustración 6: Histograma de valores para la evaluación General del SR

Es importante destacar que tanto la adquisición como el posterior análisis de los datos fueron realizados de manera automática, utilizando los formularios Google Forms para la adquisición y la planilla de cálculo diseñada con Microsoft Excel para el análisis, de acuerdo con lo explicado anteriormente en las secciones 4.6.1 y 4.6.2.

Dado que la cantidad de ítems de la encuesta que corresponde a cada categoría no es la misma, es importante tener esto en cuenta si se pretende trabajar con valores absolutos.

En el presente trabajo, se utilizó la moda, que es el valor con mayor frecuencia en la distribución de los datos, para determinar los “puntajes” correspondientes a cada categoría, según se presentan a continuación.

| Resultados Obtenidos en la Evaluación por Usuarios | |
|---|-----------------------------|
| Categoría | Puntaje (según moda) |
| Interacción e Interfaz | 3,0 |
| Calidad / Novedad | 3,0 |
| Confianza | 3,5 |
| Utilidad / Relevancia | 4,0 |
| Satisfacción | 5,0 |
| | |
| General | 3,0 |

Tabla 9: Resultados obtenidos en la evaluación por usuarios

En la categoría de Confianza, se obtuvo igual cantidad de calificaciones con los valores 3 y 4, por lo que se le asignó un puntaje de 3,5 correspondiente al promedio aritmético de ambos valores, al simple efecto de dar cuenta de esta situación.

Dado que los puntajes asignados por los usuarios que completaron la encuesta cubren un rango de 1 a 5 puntos para manifestar total desacuerdo (1) a total acuerdo (5), respectivamente, se pueden tomar los valores 1 y 2 como calificaciones negativas hacia el sistema recomendador, los valores 4 y 5 como calificaciones positivas hacia el sistema y el valor 3 como una calificación neutra, que no indica aprobación ni desaprobación.

Teniendo en cuenta esto, un rápido análisis de los resultados obtenidos en la encuesta correspondiente al sistema recomendador evaluado, permite señalar que los usuarios obtuvieron un alto grado de *Satisfacción* (5 puntos) al utilizar este sistema, en tanto que apreciaron la *Utilidad y Relevancia* de sus recomendaciones (4 puntos).

En cuanto a la *Confianza* que los usuarios manifestaron sobre las recomendaciones recibidas, la balanza se inclina hacia un grado de acuerdo algo mayor que el promedio (3 a 4 puntos).

Con respecto a los aspectos relativos a la *Interfaz e Interacción* con el sistema así como la *Calidad / Novedad* de sus recomendaciones, los usuarios se mostraron solamente satisfechos (3 puntos), lo que indicaría que son las cuestiones sobre las que más habría que trabajar para mejorar el sistema recomendador.

Por último, un análisis combinado de los valores asignados por los usuarios para todos los tópicos de todas las categorías, ubica a este sistema recomendador con un puntaje promedio, neutro (3 puntos), de acuerdo con la moda del histograma correspondiente a la evaluación *General*.

Para terminar, a modo informativo, en el Anexo C se presentan algunas observaciones y comentarios realizados por estudiantes que participaron en la encuesta a usuarios.

5.4. Aplicación de la Encuesta a Expertos

La prueba de evaluación por expertos se llevó a cabo para el mismo sistema de recomendación de contenidos educativos sobre el que se efectuó la encuesta a usuarios, el cual funciona sobre una plataforma Moodle utilizada en la asignatura “Orientación a Objetos 2”, de Ingeniería Informática de la Universidad Nacional de La Plata, descrito anteriormente.

Para efectuar esta evaluación, se utilizaron los datos proporcionados para diferentes momentos y usuarios (alumnos) utilizando las “fotos” del sistema recomendador, *ego network de radio 1* (ver Anexo A), de manera que el experto respondiera la encuesta a expertos (ver 4.7.1) a partir del análisis que realice acerca de las recomendaciones efectuadas por el sistema a los distintos estudiantes, en función de dichos datos.

Una vez procesados los datos de la encuesta, los puntajes asignados por los expertos (p) a las preguntas con escalas Likert inversas (números 6, 7 y 11), fueron ajustados de la siguiente manera:

$$p' = 6 - p$$

de modo que un valor mínimo ($p = 1$) en esas respuestas correspondieran a un valor máximo ($p' = 5$) y consecuentemente con el resto de los puntajes.

Posteriormente, para cada uno de los ítems de la encuesta, se determinó la frecuencia correspondiente a cada uno de los valores asignados por los expertos. Por ejemplo, para el ítem número 1 de la encuesta, la cantidad de expertos que lo evaluaron con “1”, la cantidad de expertos que lo evaluaron con “2”, etc.

Así se obtuvieron los valores indicados en la tabla siguiente, los que permiten efectuar un histograma para cada ítem de la encuesta.

| | Frecuencia de los Ítems 1 a 11 | | | | | | | | | | |
|-------|--------------------------------|---|---|---|---|---|---|---|---|----|----|
| Valor | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 2 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 3 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 4 | 1 | 3 | 1 | 2 | 1 | 2 | 0 | 1 | 2 | 3 | 1 |
| 5 | 2 | 1 | 2 | 1 | 3 | 1 | 3 | 2 | 1 | 1 | 0 |

Tabla 10: Frecuencia de los ítems en la encuesta a expertos

A partir de la categorización definida en (4.5), se determinó para cada categoría la frecuencia de cada puntaje asignado por los expertos. Para ello se sumaron las ocurrencias de cada valor dentro de los ítems correspondientes a cada categoría. Resulta importante destacar que, como se indicó antes, la encuesta a expertos no cubre las categorías de *Interacción e Interfaz* y *Satisfacción* (del usuario), por lo que no se registraron valores para las mismas.

De esta forma se obtuvieron los valores indicados en la tabla siguiente, los que permiten efectuar un histograma para cada categoría definida a partir de la encuesta a expertos.

| Frecuencias por Categoría y General | | | | | | |
|-------------------------------------|------------------------|-------------------|-----------|-----------------------|--------------|---------------|
| Valor | Interacción e Interfaz | Calidad / Novedad | Confianza | Utilidad / Relevancia | Satisfacción | Total General |
| 1 | -- | 1 | 3 | 0 | -- | 4 |
| 2 | -- | 1 | 0 | 0 | -- | 1 |
| 3 | -- | 2 | 1 | 2 | -- | 5 |
| 4 | -- | 6 | 6 | 5 | -- | 17 |
| 5 | -- | 10 | 2 | 5 | -- | 17 |

Tabla 11: Frecuencias por categoría en la encuesta a expertos

5.5. Análisis de los Datos Obtenidos en la Encuesta a Expertos

Para el análisis de los datos se realizaron, a partir de los valores adquiridos, los gráficos correspondientes a los histogramas obtenidos para cada una de las categorías estudiadas, los cuales se presentan a continuación.

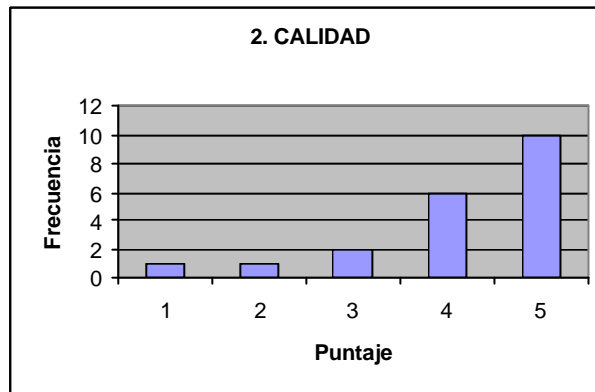


Ilustración 7: Histograma de valores para la categoría Calidad / Novedad

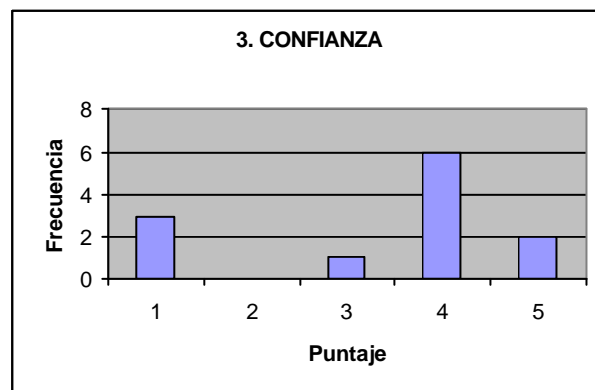


Ilustración 8: Histograma de valores para la categoría Confianza

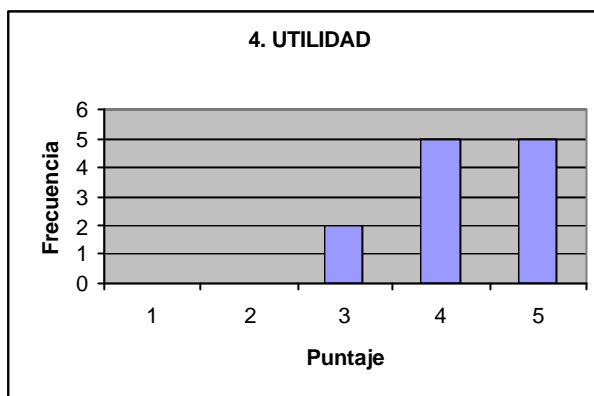


Ilustración 9: *Histograma de valores para la categoría Utilidad / Relevancia*

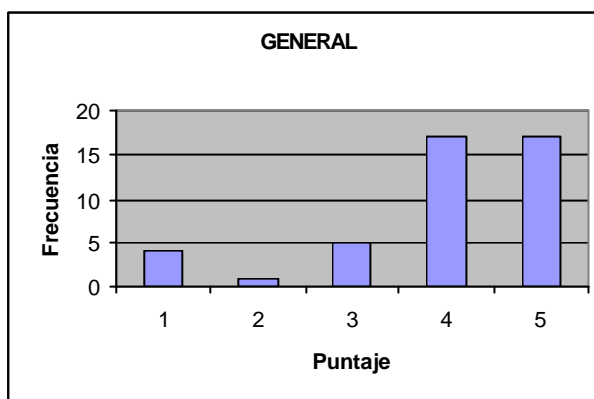


Ilustración 10: *Histograma de valores para la evaluación General del SR*

Es importante destacar que tanto la adquisición como el posterior análisis de los datos fueron realizados de manera automática, utilizando los formularios Google Forms para la adquisición y la planilla de cálculo diseñada con Microsoft Excel para el análisis, de acuerdo con lo explicado anteriormente en las secciones 4.6.1 y 4.6.2.

Dado que la cantidad de ítems de la encuesta que corresponde a cada categoría no es la misma, es importante tener esto en cuenta si se pretende trabajar con valores absolutos. En el presente trabajo, se utilizó la moda, que es el valor con mayor frecuencia en la distribución de los datos, para determinar los “puntajes” correspondientes a cada categoría, según se presentan a continuación.

| Resultados Obtenidos en la Evaluación por Expertos | |
|---|-----------------------------|
| Categoría | Puntaje (según moda) |
| Interacción e Interfaz | -- |
| Calidad / Novedad | 5,0 |
| Confianza | 4,0 |
| Utilidad / Relevancia | 4,5 |
| Satisfacción | -- |
| | |
| General | 4,5 |

Tabla 12: Resultados obtenidos en la evaluación por expertos

En la categoría de Utilidad / Relevancia, se obtuvo igual cantidad de calificaciones con los valores 4 y 5, por lo que se le asignó un puntaje de 4,5 correspondiente al promedio aritmético de ambos valores, al simple efecto de dar cuenta de esta situación. Una situación análoga se observó en la evaluación general.

Dado que los puntajes asignados por los expertos que completaron la encuesta cubren un rango de 1 a 5 puntos para manifestar total desacuerdo (1) a total acuerdo (5), respectivamente, se pueden tomar los valores 1 y 2 como calificaciones negativas hacia el sistema recomendador, los valores 4 y 5 como calificaciones positivas hacia el sistema y el valor 3 como una calificación neutra, que no indica aprobación ni desaprobación.

Teniendo en cuenta esto, un rápido análisis de los resultados obtenidos en la encuesta a expertos correspondiente al sistema recomendador evaluado, permite señalar que los expertos otorgaron un muy alto grado a la *Calidad* (5 puntos) de las recomendaciones, y asimismo destacaron tanto la *Confianza* en las mismas (4 puntos) como su *Utilidad* y *Relevancia* (4,5 puntos).

Por último, un análisis combinado de los valores asignados por los expertos para todos los tópicos de las tres categorías evaluadas, ubica a este sistema recomendador con un puntaje promedio positivo alto (4 a 5 puntos), de acuerdo con la moda del histograma correspondiente a la evaluación *General*.

Para terminar, a modo informativo, en el Anexo D se presentan algunas observaciones y comentarios realizados por docentes que participaron en la encuesta a expertos.

5.6. Comparación de los Resultados Obtenidos

La comparación entre los resultados obtenidos en la encuesta a los usuarios del sistema recomendador (estudiantes) y la encuesta a los expertos (docentes), se efectúa a los efectos de validar la herramienta de evaluación aquí presentada.

Como se indicó anteriormente, dentro de las cinco categorías evaluadas, Interacción e Interfaz, Calidad/Novedad, Confianza, Utilidad/Relevancia y Satisfacción, la encuesta a expertos se limitó a realizar mediciones únicamente sobre las categorías de Calidad/Novedad, Confianza y Utilidad/Relevancia.

En el siguiente cuadro se muestra la comparación de los resultados provenientes de la aplicación de ambas encuestas al sistema recomendador evaluado, indicándose el valor modal obtenido.

| Comparación de los Resultados Obtenidos | | |
|--|----------------------------|----------------------------|
| Categoría | Evaluación Usuarios | Evaluación Expertos |
| Calidad / Novedad | 3 | 5 |
| Confianza | 3 a 4 | 4 |
| Utilidad / Relevancia | 4 | 4 a 5 |

Tabla 13: *Comparación de los resultados obtenidos en ambas encuestas*

Puede apreciarse que tanto en la categoría Confianza como en Utilidad/Relevancia, ambas evaluaciones coinciden significativamente. En cuanto a la categoría de Calidad/ Novedad, los usuarios la evaluaron en forma neutra, en tanto que los expertos lo hicieron en forma muy positiva, resultando, de todas formas, en una diferencia no contradictoria, dado que ambas evaluaciones se ubican en valores de 3 a 5 de la escala de Likert (entre neutro y acuerdo).

En este último aspecto, es importante tener en cuenta que en un sistema recomendador de contenidos educativos, las recomendaciones brindadas no siempre van a reflejar completamente el gusto de los usuarios porque, al mismo tiempo, debe tenerse presente cuáles son los temas del curso que el docente desea reforzar. Así que las estrategias del recomendador pueden no resultar tan evidentes para los estudiantes, quienes podrían interpretar ciertas recomendaciones como fuera de contexto (como se indicó en discusiones anteriores, recomendar un objeto de aprendizaje no es lo mismo que vender un artículo de consumo masivo).

Por otro lado, es posible que el “arranque en frío” sea de mayor duración para un sistema recomendador con las características presentes en este caso, en el cual la cantidad de usuarios es mucho más limitada que en sistemas de uso masivo. Esto haría que se requiriera mayor tiempo e interacción con cada usuario para que el sistema recomendador vaya adaptándose a sus gustos y necesidades, sin descuidar el logro educativo a alcanzar.

También debe notarse que el cambio del comportamiento del sistema recomendador y de los estudiantes en tiempos de cursada o de exámenes, agregaría otra variable a este análisis, pudiendo darse el caso que la encuesta arroje resultados diferentes de acuerdo al momento en que se la aplique y las necesidades de los usuarios en cada ocasión.

6. Conclusiones y Líneas de Investigación Futuras

En el presente trabajo se formula una propuesta metodológica para la evaluación de sistemas recomendadores de contenidos educativos en entornos de aprendizaje formal. Para alcanzar este objetivo, se desarrolla una herramienta consistente en una encuesta a usuarios, la cual permite evaluar el sistema recomendador en cinco categorías diferentes, correspondientes a Interacción e Interfaz, Calidad/Novedad, Confianza, Utilidad/Relevancia y Satisfacción del usuario.

Esta encuesta a usuarios consiste de un conjunto de veinticuatro tópicos a evaluar individualmente, utilizando una calificación determinada por una escala de Likert, entre 1 (completo desacuerdo) y 5 (completo acuerdo). Los tópicos a evaluar se agrupan dentro de las cinco categorías descritas previamente.

Se determinó la aplicabilidad de esta herramienta de evaluación, a través de un estudio de caso, llevado a cabo sobre un grupo de usuarios del sistema de recomendación de contenidos educativos, que funciona sobre una plataforma Moodle utilizada en la Facultad de Ingeniería de la Universidad Nacional de La Plata, en las asignaturas Orientación a Objetos 1, Orientación a Objetos 2 y Diseño y Arquitectura de Software.

Con el fin de validar la metodología de evaluación propuesta, se desarrolló una herramienta de validación del método, consistente en una encuesta a expertos que cubre tres categorías centrales de las evaluadas a través de la encuesta a usuarios, que comprenden Calidad/Novedad, Confianza y Utilidad/Relevancia.

Esta encuesta a expertos consiste de un conjunto de once tópicos a evaluar individualmente, utilizando una calificación determinada por una escala de Likert, entre 1 (completo desacuerdo) y 5 (completo acuerdo). Los tópicos a evaluar se agrupan dentro de las tres categorías descritas anteriormente. Para completar esta encuesta los expertos no interactúan con el sistema recomendador, sino que utilizan información extraída del mismo (ver Anexo A).

Una vez aplicadas las encuestas a usuarios y expertos, se efectuó la comparación de los resultados obtenidos en ambas. Se observó que en dos de las categorías evaluadas (Confianza y Utilidad/Relevancia), no se encontraron diferencias significativas, en tanto que en la tercera categoría (Calidad/Novedad), se obtuvieron diferencias no contradictorias (neutro, en encuesta a usuarios, a positivo, en encuesta a expertos) entre ambas encuestas realizadas.

A partir de los resultados expuestos, se concluye que la metodología propuesta cumple con los objetivos perseguidos, en tanto resulta ser clara, aplicable, reproducible, verificable, comparable y brinda, en principio, un grado de confiabilidad razonable para realizar la evaluación de sistemas de recomendación de contenidos educativos en entornos de aprendizaje formal, como el evaluado en el estudio de caso presentado en este trabajo.

El trabajo aquí presentado, abre además una cantidad de posibilidades en cuanto a futuros desarrollos e investigaciones a seguir. En lo inmediato, los resultados obtenidos una vez analizados los datos de la encuesta a usuarios, pueden ser utilizados para determinar los aspectos a mejorar en el sistema recomendador evaluado, encontrando sus debilidades y fortalezas.

De esta manera, podrían ofrecerse sugerencias para mejorar el sistema evaluado en aquellas categorías con bajos puntajes, e incluso detectar aspectos puntuales a mejorar dentro de cada categoría, según lo revelado por las respuestas a determinados ítems correspondientes a esa categoría dentro de la encuesta.

La automatización de estos procesos resultaría de gran ayuda para el diseñador de sistemas de recomendación en esta área, ya que brindaría una forma rápida y eficiente de realizar pruebas ante cualquier cambio introducido en el sistema, que se evaluarían en forma continua a partir de la aplicación de las encuestas a usuarios y el análisis automático de los resultados obtenidos.

La clave para la mejora permanente durante el ciclo de vida del sistema recomendador, se encontrará seguramente en su evaluación continua, a partir del diseño de encuestas cada vez más sencillas de implementar, simples de completar para los usuarios y, utilizando técnicas para su análisis automatizado, fáciles de interpretar para los desarrolladores.

7. Referencias bibliográficas

Brin, S., Page, L. (1998). *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. Computer Science Department, Stanford University, Stanford, CA, USA.

Bott, E., Leonhard, W. (1999). *Edición Especial Microsoft Office 2000*. Prentice Hall. Madrid. España.

Bridge, D., Göker, M., McGinty, L., Smyth, B. (2006). *Case-based recommender systems*. The Knowledge Engineering review 20(3), 315–320.

Burke, R. (2007). *Hybrid web recommender systems*. The AdaptiveWeb, pp. 377–408. Springer. Berlin / Heidelberg, Alemania.

Casali, A., Gerling, V., & Deco, C. (2011). Sistema inteligente para la recomendación de objetos de aprendizaje. *Ariadne, Revista Generación Digital*, 9(1), 88-95.

Chen, L., Pu, P. (2010). User Evaluation Framework of Recommender Systems. *Workshop SRS'10* (ACM 978-1-60558-995-4/10/01). Hong Kong, China. Feb. 7, 2010.

Drachsler, H., Bogers, T., Vuorikari, R., Verbert, K., Duval, E., Manouselis, N., Beham, G., et al. (2010). Issues and considerations regarding sharable data sets for recommender systems in technology enhanced learning. *Procedia Computer Science*, 1(2), 2849-2858. Elsevier. doi:10.1016/j.procs.2010.08.010

Drachsler, H., Hummel, H. G. K., & Koper, R. (2008). Personal recommender systems for learners in lifelong learning networks: the requirements, techniques and model. *International Journal of Learning Technology*, 3(4), 404-423. Inderscience. doi:10.1504/IJLT.2008.019376.

Drachsler, H., Hummel, H. G. K., & Koper, R. (2009). Identifying the Goal, User model and Conditions of Recommender Systems for Formal and Informal Learning. *Journal Of Digital Information*, 10(2), 4-24.

Golbeck, J. (2006). *Generating predictive movie recommendations from trust in social networks*. Trust Management, 4th International Conference, iTrust 2006, Pisa, Italy, May 16-19, 2006, Proceedings, pp. 93–104.

Jannach D., Zanker M., Felfernig A., Friedrich G. (2011). *Recommender System An Introduction*. Cambridge University Press. New York. USA .

Kirkpatrick, D.L. (1959). *Evaluating Training Programs* (2nd ed.). Berrett Koehler, San Francisco, USA.

Mahmood, T., Ricci, F. (2007). *Towards learning user-adaptive state models in a conversational recommender system*. A. Hinneburg (ed.) LWA 2007: Lernen - Wissen - Adaption, Halle, September 2007, Workshop Proceedings, pp. 373–378. Martin-Luther-University Halle-Wittenberg.

- Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., & Koper, R. (2011). Recommender systems in technology enhanced learning. *Recommender Systems Handbook*, 387–415. Springer.
- Marradi, A., Archenti, N., Piovani, J. (2007). *Metodología de las Ciencias Sociales*. Emece Editores S.A., Buenos Aires, Argentina.
- Pu, P., Chen, L. (2010). A User-Centric Evaluation Framework of Recommender Systems. *Proceedings of the ACM RecSys 2010 Workshop on User-Centric Evaluation of Recommender Systems and their Interfaces* (Vol. i, pp. 14-21). Barcelona, España. Sep.30, 2010.
- Ricci, F., Cavada, D., Mirzadeh, N., Venturini, A. (2006). *Case-based travel recommendations*. D.R. Fesenmaier, K.Woeber, H.Werthner (eds.) *Destination Recommendation Systems: Behavioural Foundations and Applications*, pp. 67–93. CABI 2006.
- Ricci F., Rokach L, Shapira B., Kantor P. (2011). *Recommender System Handbook*. Springer. New York. USA.
- Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S. (2007). *Collaborative filtering recommender systems*. *The Adaptive Web*, pp. 291–324. Springer Berlin / Heidelberg, Alemania.
- Schafer, J.B., Konstan, J.A., Riedl, J. (2001). *E-commerce recommendation applications*. *Data Mining and Knowledge Discovery* 5(1/2), pp. 115–153.
- Shani G., Gunawardana A. (2011). *Evaluating Recommendation Systems*. In Ricci (2011) *Recommender Systems Handbook*, Chapter 8. Springer. New York. USA.
- Sinha, R.R., Swearingen, K. (2001). *Comparing recommendations made by online systems and friends*. *DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries*.
- Verbert, K., Drachsler, H., Manouselis, N., Wolpers, M., Vuorikari, R., & Duval, E. (2011). Dataset-driven Research for Improving Recommender Systems for Learning. *Technology*, 44-53. ACM Press.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*: Harvard University Press.
- Wasserman, S., & Faust, K. (1999). *Social Network Analysis: methods and applications*: Cambridge University Press.

ANEXO A

Datos Provistos al Experto Para Completar la Encuesta a Expertos

El sistema bajo estudio es una integración de Moodle con un Repositorio de Objetos de Aprendizaje externo. La red de recursos resultante se compone de:

- Curso/s (en Moodle)
 - Sección/es
 - Recurso/s
 - Foro/s
 - Discusión/es
- Usuario/s (en Moodle)
- Objeto/s de aprendizaje (en el repositorio)

Se tomaron dos instantáneas de los datos del sistema, a dos tiempos con dos meses de diferencia entre ellas (13/5/13 y 7/7/2013). En ambos momentos el sistema estaba siendo utilizado simultáneamente para atender a tres cursos en temáticas de orientación a objetos.

Se identificaron dos usuarios, pertenecientes a dos cursos diferentes. Se generaron recomendaciones para los dos momentos capturados en las instantáneas.

A continuación se muestra, para cada instantánea:

1. Estructura de la red (simplemente para tener una idea de la complejidad de la información manejada).
2. Conexiones de cada usuario con elementos contenidos y otros usuarios al momento de la instantánea (Ego-Network de radio 1).
3. Recomendaciones que ofrece el sistema al usuario, al momento de la instantánea.

Se espera que esta información sirva al experto para hacerse una idea del contexto del usuario al momento de la instantánea. Este es también el contexto en el que se calculan las recomendaciones. Por consiguiente, se busca que el experto pueda evaluar la pertinencia de las recomendaciones.

Tiempo 1

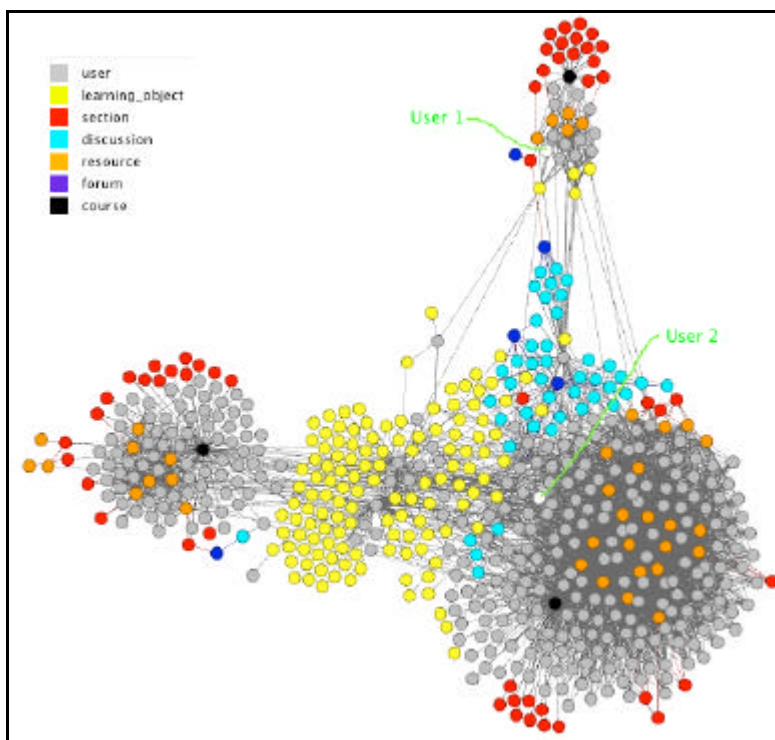


Ilustración 11: Estructura de la Red al Tiempo 1

Usuario 1

Ego-network de radio 1

A continuación se presenta la lista de recursos a los que se encuentra conectado el usuario al momento de la instantánea.

| Label | type |
|---|-----------------|
| Diseño y arquitectura - 2013 | course |
| Repaso Objetos - Arbol familiar | resource |
| MaquinaDeCafeRolePlaying | resource |
| Proyecto BlueJ del Oraculo para el Trabajo Práctico 2 | resource |
| Repaso Objetos - Figuras | resource |
| Solución de referencia al diagnóstico | resource |
| Add Parameter | learning_object |
| Consolidate Conditional Expression | learning_object |
| Introduce Assertion | learning_object |

Recomendaciones para el usuario 1

A continuación se indican los contenidos que el sistema recomienda al usuario.

| | |
|----------|------------------------------------|
| resource | TP2 |
| resource | Práctica 1 |
| resource | Programa - Orientación a Objetos 2 |

| | |
|-----------------|--|
| learning_object | Abstract Classes in Smalltalk |
| learning_object | Bad Smells |
| learning_object | Hot-Spot-Driven Framework Development by W Pree |
| discussion | Ejemplo de composite en Visualworks |
| discussion | Ejemplo de template method en Cuentas-Bancarias con error? |
| discussion | Explicacion práctica |

Usuario 2

Ego network de radio 1

A continuación se presenta la lista de recursos a los que se encuentra conectado el usuario al momento de la instantánea.

| Label | type |
|---|-----------------|
| Orientación a Objetos 2 - 2013 | course |
| Programa - Orientación a Objetos 2 | resource |
| Codigo fuente utilizado en clase - versiones usando dependencias, AspectAdaptor y DependencyTransformer | resource |
| Ejemplo de composite en Visualworks | resource |
| Ejemplo de template method en Cuentas-Bancarias | resource |
| Práctica 1 | resource |
| Transparencias de la teoría 4 | resource |
| Trabajo Práctico Nro. 6 | resource |
| Transparencias de la teoría 3 | resource |
| Tp03 | resource |
| Material adicional trabajo práctico Nro. 5 | resource |
| TP2 | resource |
| Trabajo Practico Nro 5 | resource |
| Trabajo Práctico Nro 4 | resource |
| Consulta colecciones Practica 5 | discussion |
| Consulta parcel | discussion |
| Consulta practica 5 (menu) | discussion |
| Composite Design Pattern (Video) | learning_object |
| Composite Design Pattern | learning_object |
| Dispatch | learning_object |
| Chain of Responsibility | learning_object |
| Abstract Classes in Smalltalk | learning_object |
| Building a Composite with the Builder Pattern | learning_object |
| Eliminate Duplicate Code with the Template Method | learning_object |

Recomendaciones para el usuario 2

A continuación se indican los contenidos que el sistema recomienda al usuario.

| | |
|----------|------------------------------------|
| resource | TP2 |
| resource | Práctica 1 |
| resource | Programa - Orientación a Objetos 2 |

| | |
|-----------------|--|
| learning_object | Abstract Classes in Smalltalk |
| learning_object | Bad Smells |
| learning_object | Hot-Spot-Driven Framework Development by W Pree |
| discussion | Ejemplo de composite en Visualworks |
| discussion | Ejemplo de template method en Cuentas-Bancarias con error? |
| discussion | Explicación practica |

Tiempo 2

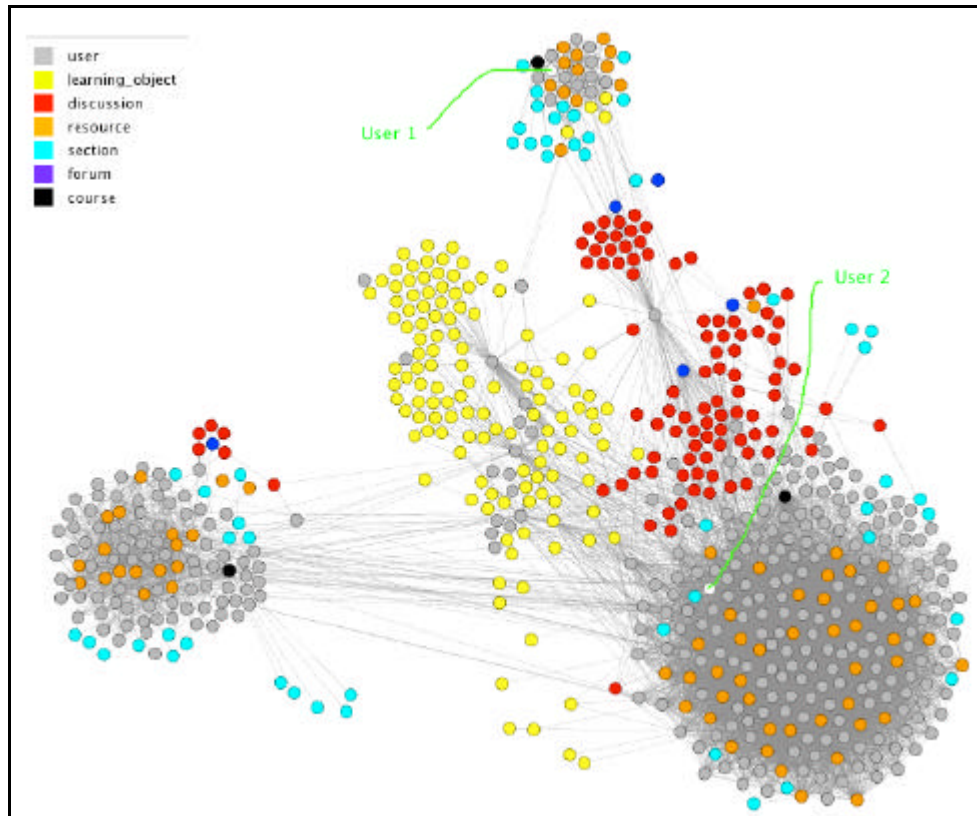


Ilustración 12: Estructura de la Red al Tiempo 2

Usuario 1

Ego network de radio 1

A continuación se presenta la lista de recursos a los que se encuentra conectado el usuario al momento de la instantánea.

| Label | type |
|---|-----------------|
| Diseño y arquitectura - 2013 | course |
| Add Parameter | learning_object |
| Consolidate Conditional Expression | learning_object |
| Introduce Assertion | learning_object |
| Repaso Objetos - Arbol familiar | resource |
| MaquinaDeCafeRolePlaying | resource |
| Ejemplo utilizado para diseño conducido por responsabilidades | resource |
| Proyecto BlueJ del Oraculo para el Trabajo Práctico 2 | resource |

| | |
|---------------------------------------|----------|
| Repaso Objetos - Figuras | resource |
| Solución de referencia al diagnóstico | resource |
| Ejercicio de revisión de colecciones | resource |

Recomendaciones para el usuario 1

A continuación se indican los contenidos que el sistema recomienda al usuario.

| | |
|-----------------|--|
| resource | Programa - Orientación a Objetos 2 |
| resource | Transparencias de la teoría 3 |
| resource | Práctica 1 |
| learning_object | Bad Smells |
| learning_object | Hot-Spot-Driven Framework Development by W Pree |
| learning_object | Abstract Classes in Smalltalk |
| discussion | Ejemplo de composite en Visualworks |
| discussion | Ejemplo de template method en Cuentas-Bancarias con error? |
| discussion | Muestra Parcialito Refactoring |

Usuario 2

Ego network de radio 1

A continuación se presenta la lista de recursos a los que se encuentra conectado el usuario al momento de la instantánea.

| Label | type |
|---|-----------------|
| Orientación a Objetos 2 - 2013 | course |
| Consulta colecciones Practica 5 | discussion |
| Consulta parcel | discussion |
| Consulta practica 5 (menu) | discussion |
| Composite Design Pattern (Video) | learning_object |
| Composite Design Pattern | learning_object |
| Dispatch | learning_object |
| Chain of Responsibility | learning_object |
| Abstract Classes in Smalltalk | learning_object |
| Building a Composite with the Builder Pattern | learning_object |
| Eliminate Duplicate Code with the Template Method | learning_object |
| Programa - Orientación a Objetos 2 | resource |
| Trabajo práctico 8 | resource |
| SUnit Explained | resource |
| Codigo fuente utilizado en clase - versiones usando dependencias, AspectAdaptor y DependencyTransformer | resource |
| Transparencias de la teoría 5 | resource |
| Ejemplo de composite en Visualworks | resource |
| Ejemplo de template method en Cuentas-Bancarias | resource |
| Transparencias de la teoría 12 - Refactoring to patterns | resource |
| Práctica 1 | resource |
| Transparencias de la teoría 4 | resource |
| Transparencias de la teoría 3 | resource |
| Tp03 | resource |
| Transparencias de la teoría 11 - Refactoring según Martin Fowler | resource |
| Johnson, R. E. (1997). Frameworks = (components + patterns). Communications of the ACM, 40(10), 3942. | resource |
| Material adicional trabajo práctico Nro. 5 | resource |
| TP2 | resource |
| Transparencias de la teoría 2 | resource |

| | |
|--|----------|
| Transparencias de la teoría 7 (Patrón observer) | resource |
| Transparencias de la teoría 9 - Frameworks | resource |
| Fayad, M., & Schmidt, D. C. (1997). Object-oriented application frameworks. Communications of the ACM, 40(10), 3238. | resource |
| Trabajo Práctico Nro 4 | resource |
| TP 11 | resource |
| Trabajo Práctico 10 | resource |
| Resultados parcialito de patrones | resource |
| Listado integral de notas de los 3 parcialitos | resource |
| Trabajo Práctico Nro. 6 | resource |
| Resultados del parcial - listado actualizado y completo | resource |
| Notas Parcialito Refactoring | resource |
| Transparencias de la teoría 8 (Model-View-Controller y como hacer custom views) | resource |
| Trabajo Practico Nro 5 | resource |
| Trabajo Práctico 9 | resource |
| Trabajo práctico 7 | resource |

Recomendaciones para el usuario 2

A continuación se indican los contenidos que el sistema recomienda al usuario.

| | |
|-----------------|--|
| resource | Programa - Orientación a Objetos 2 |
| resource | Transparencias de la teoría 3 |
| resource | Práctica 1 |
| learning_object | Bad Smells |
| learning_object | Hot-Spot-Driven Framework Development by W Pree |
| learning_object | Abstract Classes in Smalltalk |
| discussion | Ejemplo de composite en Visualworks |
| discussion | Ejemplo de template method en Cuentas-Bancarias con error? |
| discussion | Muestra Parcialito Refactoring |

ANEXO B

Capturas de Pantallas de la Plataforma Moodle con el Sistema Recomendador Evaluado

A continuación se presentan, a modo de ejemplo, algunas capturas de pantallas correspondientes al sistema Moodle y a repositorios externos de objetos de aprendizaje, pertenecientes a la asignatura “Orientación a Objetos 2”. En la sección de la derecha de cada pantalla pueden observarse las recomendaciones del sistema recomendador sobre el cual se realizó la evaluación.

Pantalla principal del Moodle al ingresar el 25/06/2013

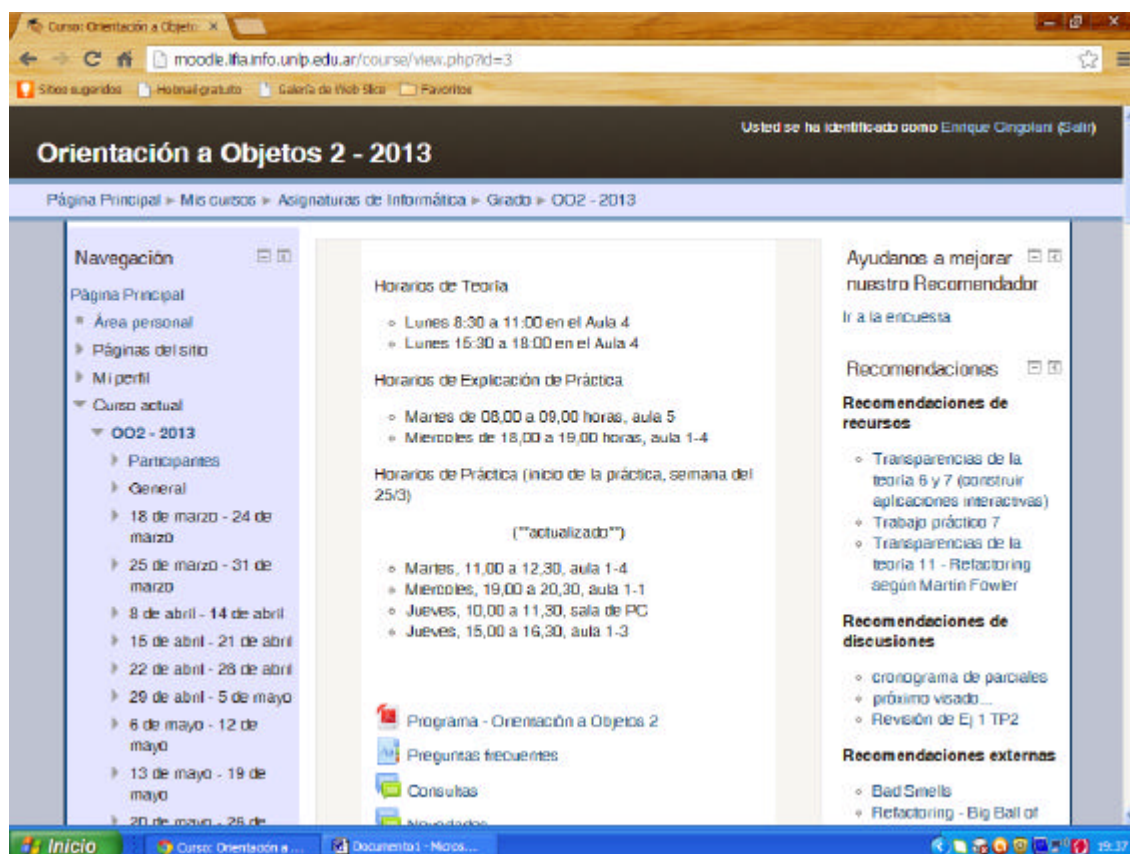


Ilustración 13: Pantalla principal Moodle 25/06/2013

Pantalla repositorio externo Wikilor 25/06/2013
(Recomendaciones Externas desde pantalla principal)



Ilustración 14: Pantalla repositorio externo Wikilor 25/06/2013

Pantalla principal del Moodle al ingresar el 01/07/2013

The screenshot shows the Moodle course interface for 'Orientación a Objetos 2 - 2013'. The browser address bar indicates the URL 'moodle.iaa.info.unip.edu.ar/course/view.php?id=3'. The user is logged in as 'Enrique Gonzalez (Gonz)'. The page layout includes a navigation sidebar on the left, a main content area, and a right sidebar with recommendations.

Curso: Orientación a Objetos 2 - 2013

Usted se ha identificado como Enrique Gonzalez (Gonz)

Página Principal » Mis cursos » Asignaturas de Informática » Grado » OO2 - 2013

Navegación

- Página Principal
- Área personal
- Páginas del sitio
- Mi perfil
- Curso actual
 - OO2 - 2013**
 - Participantes
 - General
 - 18 de marzo - 24 de marzo
 - 25 de marzo - 31 de marzo
 - 8 de abril - 14 de abril
 - 15 de abril - 21 de abril
 - 22 de abril - 28 de abril
 - 29 de abril - 5 de mayo
 - 6 de mayo - 12 de mayo
 - 13 de mayo - 19 de mayo
 - 20 de mayo - 26 de mayo
 - 27 de mayo - 2 de junio
 - 3 de junio - 9 de junio
 - 10 de junio - 16 de junio
- Mis cursos

Ajustes

- Administración del curso
- Calificaciones

Horarios de Teoría

- Lunes 8:30 a 11:00 en el Aula 4
- Lunes 15:30 a 18:00 en el Aula 4

Horarios de Explicación de Práctica

- Martes de 08:00 a 09:00 horas, aula 5
- Miércoles de 18:00 a 19:00 horas, aula 1-4

Horarios de Práctica (Inicio de la práctica, semana del 25/3) (**actualizado**)

- Martes, 11:00 a 12:30, aula 1-4
- Miércoles, 19:00 a 20:30, aula 1-1
- Jueves, 10:00 a 11:30, sala de PC
- Jueves, 15:00 a 16:30, aula 1-3

Programa - Orientación a Objetos 2

- Preguntas frecuentes
- Consultas
- Novedades
- Turno Martes 11,00 horas
- Turno Miércoles 19,00 horas
- Turno Jueves 10,00 horas
- Turno Jueves 15,00 horas
- Listado integral de notas de los 3 parcialitos

18 de marzo - 24 de marzo

- Transparencias de la teoría 1

Ayudanos a mejorar nuestro Recomendador

Ir a la encuesta

Recomendaciones

Recomendaciones de recursos

- Transparencias de la teoría 6 y 7 (construir aplicaciones interactivas)
- Trabajo práctico 7
- Transparencias de la teoría 11 - Refactoring según Martin Fowler

Recomendaciones de discusiones

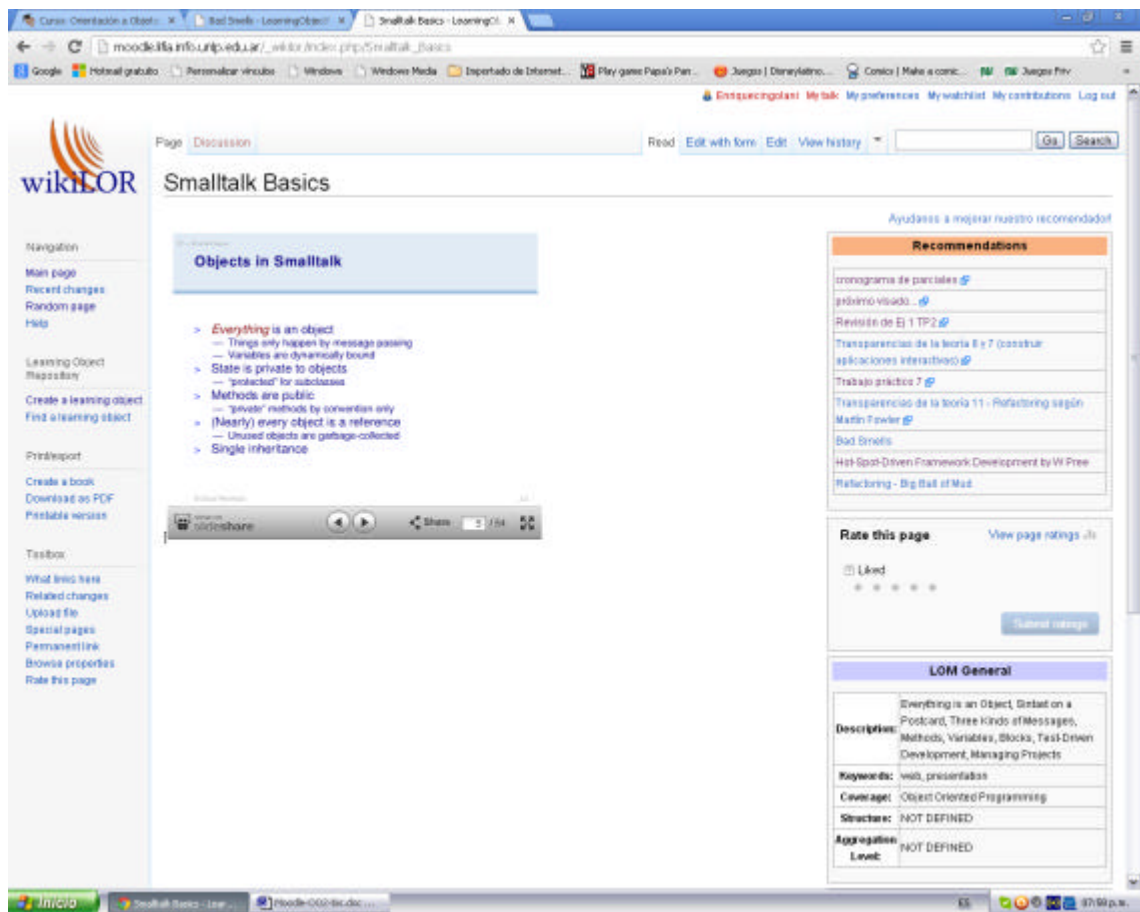
- cronograma de parciales
- próximo visado...
- Revisión de [] 1 TP2

Recomendaciones externas

- Bad Smells
- Hot-Spot-Driven Framework Development by W. Free
- Refactoring - Big Ball of Mud

Ilustración 15: Pantalla principal Moodle 01/07/2013

Pantalla repositorio externo Wikilor 01/07/2013 (Recomendaciones Externas desde la pantalla principal)



The screenshot displays the Wikilor external repository interface. The main content area is titled "Smalltalk Basics" and features a section "Objects in Smalltalk" with a list of bullet points:

- > **Everything is an object**
 - Things only happen by message passing
 - Variables are dynamically bound
- > **State is private to objects**
 - "protected" for subclasses
- > **Methods are public**
 - "private" methods by convention only
- > **(Nearly) every object is a reference**
 - Unused objects are garbage-collected
- > **Single inheritance**

Below the text is a media player showing a video titled "Smalltalk Basics".

On the right side, there is a "Recommendations" section with a list of external resources:

- Programa de prácticas
- programa Visado
- Revisión de Ej 1 TP2
- Transparencias de la lección 8 y 7 (construir aplicaciones interactivas)
- Trabajo práctico 7
- Transparencias de la lección 11 - Refactoring según Martin Fowler
- Bad Smells
- High-Speed-Driven Framework Development by W Free
- Refactoring - Big Ball of Mud

Below the recommendations is a "Rate this page" section with a "Submit ratings" button. At the bottom right, there is a "LOM General" section with the following details:

| | |
|---------------------------|--|
| Description: | Everything is an Object, Getset on a Postcard, Three Kinds of Messages, Methods, Variables, Blocks, Test-Driven Development, Managing Projects |
| Keywords: | web, presentaciones |
| Coverage: | Object Oriented Programming |
| Structure: | NOT DEFINED |
| Aggregation Level: | NOT DEFINED |

Ilustración 16: Pantalla repositorio externo Wikilor 01/07/2013

ANEXO C

Observaciones y Comentarios de Usuarios Correspondientes a la Encuesta a Usuarios

A continuación se presentan, a modo informativo, observaciones y comentarios que realizaron estudiantes que utilizaron el sistema recomendador y lo evaluaron a través de la encuesta a usuarios.

Observaciones de Usuarios

-Usuario A

Una vez dentro de alguna recomendaciones, por ejemplo me derivó a un paper dentro de wikilord, ahí dentro tenía mas recomendaciones a cosas de cursos a los que no tengo acceso.

-Usuario B

Me recomendaban links a las cuales no tengo acceso, por ejemplo: Notas de parciales como las siguientes:

Resultados de Parciales: 15 de Febrero del 2013

Material que faltaba

Resultados del parcial del 21 de Diciembre

Al tratar de descargar un archivo se muestra mensaje: "Render server error"

-Usuario C

El sistema me pareció correcto en su implementación y fácil de usar, pero la encuesta pide, además, la calificación de la información brindada, y eso no es fácil de realizar y creo que no tiene que ver con la concepción del sistema.

-Usuario D

El sistema solamente me mostraba recomendaciones locales no externas y no tenían nada que ver con el tema que estaba viendo <http://prntscr.com/1kvhvk>

-Usuario E

En general me parece que funciona bien, en cuanto a las teorías y prácticas que descargué, el sistema me sugería las siguientes, fechas de parciales, etc. Por lo que me pareció acertado en ese sentido.

En cuanto al contenido que fui navegando dentro de la wiki, busqué info sobre patrones y esperaba que el recomendador me sugiriera algún otro patrón para mirar o habiendo mirado la descripción de un patrón, el video asociado, pero no sucedió.

No noté diferencia en cuanto a los temas recomendados al navegar la wiki, pero si vi como se iba amoldando a medida que descargaba contenido del moodle.

En general, creo que funciona bien, para un alumno es importante tener a mano las siguientes prácticas, los resultados o muestras de parciales, creo que es lo más importante a la hora de realizar una cursada. Quizás por ese lado, se entiende que tenga menos prioridad el contenido de la wiki, ya que sería un "extra" a lo que es la cursada en sí (parciales, notas, material, etc).

-Usuario F

Buscando por ejemplo sobre el patrón adapter (hay dos links) ambos me presentan la misma información de recomendación (lo cual no es raro) pero ninguno de los dos me muestra al otro. Es decir: ninguno de los dos me recomienda el otro que habla precisamente sobre el mismo tema. En lo personal eso me chocó bastante. Otro ejemplo es el Design Patterns Video Tutorial Part 1 no me recomienda la parte 2 (el dos tampoco me recomienda el uno), lo mismo pasa con Debugger Overview.

Por otro lado no se si es casualidad (no revisé absolutamente todos), pero en todos los que me metí me recomendó el Trabajo Práctico 7 no se si es casualidad o si está bien.

-Usuario G

Primero que nada, considero que el cómo uno se pueda sentir con la información, y lo útil que le puede resultar el material para resolver una actividad o tarea, proyecto duda, esta dado por la calidad del material en si, y no afecta al sistema en cuestión. En general las recomendaciones fueron sin sentido ya que hizo muchas recomendaciones sobre actividades o archivos subidos a la plataforma de la cátedra y muy poco al contenido de repositorios como videos de youtube y artículos. en general, las recomendaciones siempre fueron las mismas, y no había relación, por ejemplo visados del primer parcialito, practicas 6 y 7, un articulo de refactoring y el libro de los 4. En cuanto al resultado tiene lógica que me recomiende tantas cosas de la plataforma, y tan pocas de los repositorios ya que las primeras fueron vistas por todos los compañeros pero yo me puse a buscar artículos sobre refactoring y sobre patrones, y las referencias fueron las mismas, variando solo una referencia si era el que estaba viendo.

En general, se podría mejorar si se recomienda los artículos de la plataforma basados en las fechas, por ejemplo, en el mes de mayo se recomendó los artículos con mas clicks del mes de mayo, pero se reduce su cantidad, a por ejemplo 4 y si estamos fuera del tiempo de cursada de la materia o en momentos que hay habilitada alguna autoevaluación o parcial, se recomienden los documentos relacionados a los parciales, sean de los temas tomados, o resultados de los mismos, o si estamos fuera del semestre de la cursada, las teorías y practicas mas bajadas, ya que es lo que el alumno iría a buscar mas frecuentemente pero mientras el tiempo que dure la cursada seria mas interesante orientar las recomendaciones a los temas que se están viendo en los últimos 30 días.

Priorizaría mucho más los contenidos de repositorios sobre los artículos de la plataforma como son prácticas y teorías ya que el alumno por lo general entra con la intención de buscar esos artículos, y no son difíciles de encontrar, viendo mas útil, que si el alumno entra a la plataforma en el mes que se está explicando patrones por dar un ejemplo, le recomienden artículos sobre el tema, el libro de los 4 y demás, ya que recomendarle al alumno lo mismo que entró a buscar y tiene en la página principal del curso no trae ventajas.

Dado que por lo general, por mes se dan no más de 3 practicas y 3 teorías, podría recomendarse 3 o 4 referencias a los más bajados en los últimos treinta días y poner dos recomendaciones a los temas mas vistos de los últimos treinta días.

En general el sistema no varió las recomendaciones incluso después de haber buscado varias cosas, y elementos fuertemente relacionados no eran referenciados como parte uno y parte dos del mismo elemento.

Lo que si, el sistema es sumamente sencillo, e intuitivo, es fácil darse cuenta y manipularlo, yo agregaría iconos a los links para distinguir entre videos, artículos, archivos descargables, material de la plataforma como prácticas y teorías y de esta forma, se puede ayudar al usuario sin tener que hacer click diciéndole que lo que va a abrir es un video o un artículo o un archivo para descargar, ya que cada usuario tiene preferencias distintas esto puede servir también a marcar tendencia en cuanto al formato de información que los usuarios prefieren por dar un ejemplo, algunos prefieren un video, otros una pista de audio, otros un texto, así como el idioma podría marcar tendencias como que el inglés es más o menos elegido que el español y de estas maneras, encontrar material mas útil.

Me gustó la división que hay en la plataforma del moodle en cuanto a recomendaciones de foros de discusiones, artículos externos y recursos, sin embargo, trataría la posibilidad de que sea configurable por el usuario en cuanto a que algunos capaz prefieren un sistema que les muestre discriminado las recomendaciones más importantes de cada tema, y a otros le interese que le muestre las recomendaciones más importante en general y que las pueda diferenciar fácilmente o anteponiendo una palabra identificatoria como por ejemplo "video" o un ícono.

Recalco el hecho de que creo que durante la cursada sería más importante que se recomienden pocos artículos de la plataforma, los más importantes del ultimo tiempo y más artículos externos y que cuando la cursada termine se recomienden artículos de la plataforma ya que es común que un alumno que se pone a preparar un final vuelva a buscar las teorías y prácticas si por ejemplo lo rinde al otro año.

ANEXO D

Observaciones y Comentarios de Expertos Correspondientes a la Encuesta a Expertos

A continuación se presentan, a modo informativo, observaciones y comentarios que realizaron docentes que evaluaron el sistema recomendador a través de la encuesta a expertos.

Observaciones de Expertos

-Experto A

- Item 1- En el primer instante le presenta al usuario material para ponerse la día y contenido extra (wiki). En el segundo instante se concentra en los parciales, porque es la temática prevalente en ese tiempo. Este último no es tan relevante en el aprendizaje, pero válido teniendo en cuenta el contexto.

- Item 2- Las recomendaciones resultan aplicables, pero es más que nada porque las fuentes de recomendaciones disponibles está acotado a los temas relevantes. No es tanto una característica del recomendador.

- Item 4- No se brinda mucha diversidad por lo mismo que en el punto 2. Las fuentes disponibles no son muy diversas.

-Item 9- Las recomendaciones no reflejan las preferencias del estudiante, sino su actividad en Moodle/Wikilor.

-Item 11- La mayoría son acertadas, pero en un caso al final del curso se presentó una discusión en el foro sobre la primera práctica.

-Experto B

Las recomendaciones quedaron bastante divididas, algunas creo que están muy bien, y hay otras que no tienen mucho sentido. No sabía como tomar el hecho de que recomiende resultados de exámenes.