

## Diseño de un sistema de recomendación de libros y tesis basado en ontologías asociadas a tesauros: el caso de las bibliotecas de la UNMSM

### Design of a book and thesis recommendation system based on ontologies associated with thesaurus: the case of the libraries of the UNMSM

Denis G. Huamán Acuña<sup>1\*</sup>, Carlos A. Cánepa Pérez<sup>2</sup>

<sup>1</sup> Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Escuela Profesional de Ingeniería de Sistemas. Lima, Perú

\*Email: 12200134@unmsm.edu.pe

<sup>2</sup> Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Departamento Académico de Ciencias de la Computación. Lima, Perú

#### Resumen

El incremento en el número de libros, textos, tesis y otros hace que los sistemas tradicionales de búsqueda de literatura sobre algún tema en particular sean complejos y lentos, no siempre obteniendo buenos resultados. Peor aún si se trata de recomendar algún libro en particular, basado en el conocimiento de la calidad de su contenido. Este es el caso muy particular de las bibliotecas de la UNMSM, ya que no se cuenta con una ayuda automatizada para la recomendación de libros y tesis físicas.

Por ello se propone diseñar e implementar un sistema de recomendación de libros y tesis para las bibliotecas de la UNMSM mediante la técnica de recomendación basada en Ontologías asociada a Tesauros. Se diseñó un modelo ontológico para la recomendación de libros y un modelo de recomendación para las tesis porque el objetivo de la recomendación de ambas literaturas es diferente, mientras la recomendación de libros busca incrementar el conocimiento de los usuarios, la recomendación de tesis ayuda a seleccionar un tema de investigación o encontrar una referencia bibliográfica ajustada al tema de investigación del usuario.

**Palabras clave:** Sistemas de Recomendación; biblioteca; ontología; tesoro.

#### Abstract

The increase in the number of books, texts, theses and others makes traditional search systems of literature on a particular subject are complex and slow, not always obtaining good results. Even worse if it is about recommending a particular book, based on knowledge of the quality of its content. This is the very particular case of the libraries of the UNMSM, since there is no automated help for the recommendation of books and physical theses.

Therefore, it is proposed to design and implement a book and thesis recommendation system for UNMSM libraries through the recommendation technique based on Ontologies associated with Thesauri. An ontological model was designed for the recommendation of books and a model of recommendation for the theses because the objective of the recommendation of both literatures is different, while the book recommendation seeks to increase the knowledge of the users, the thesis recommendation helps to select a research topic or find a bibliographic reference adjusted to the topic of user research.

**Keywords:** Recommendation Systems; library; ontology; thesaurus.

#### Correspondencia:

Dirección: Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática. Calle Germán Amézaga N° 375, Ciudad Lima 1.

Recibido 20/06/2018 - aceptado 17/08/2018

#### Citar como:

Huamán D., Cánepa C. Diseño de un Sistema de Recomendación de Libros y Tesis basado en Ontologías asociadas a Tesauros: El Caso de las Bibliotecas de la UNMSM. Revista Peruana de Computación y Sistemas 2018 1(2):13-24. <http://dx.doi.org/10.15381/rpcs.v1i2.15379>

© Los autores. Este artículo es publicado por la Revista Peruana de Computación y Sistemas de la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional Mayor de San Marcos. Este es un artículo de acceso abierto, distribuido bajo los términos de la licencia Creative Commons Atribución - No Comercia\_Compartir Igual 4.0 Internacional. (<http://creativecommons.org/licenses/by-nc-sa/4.0/>) que permite el uso no comercial, distribución y reproducción en cualquier medio, siempre que la obra original sea debidamente citada.

### 1. Introducción

En la actualidad, el número de libros, en cada biblioteca universitaria es de gran tamaño que lo hacen muy difícil de gestionar en todos los aspectos, como se puede ver en la Figura 1. Esto consume mayor tiempo para realizar por ejemplo, la búsqueda de un libro o tesis. En algunos casos, no se encuentra el ítem que cumpla con los requerimientos del usuario.

Diversas investigaciones se han realizado para solucionar el problema de la búsqueda dentro de un gran número de ejemplares. Incluso, algunas bibliotecas han optado por tener libros electrónicos. Por ejemplo, según [2] las Nuevas Tecnologías De Información y Comunicación (NTIC's) han ayudado a la Biblioteca María Cristina Niño de Michelsen en su apoyo a la comunidad académica, donde inicialmente se introdujo la tecnología para mejora de sus servicios, todo ello se encaminó para la automatización de algunos procesos administrativos. En la primera etapa fue importante para el desarrollo tecnológico de la biblioteca mas no para el desarrollo de nuevos servicios de información. Por lo tanto un problema evidente en dicha biblioteca es el atraso en la prestación de servicios de referencia electrónicos,

lo que no permitió poner a disposición de los alumnos obras de consulta digitales y otras herramientas que permitan a los alumnos localizar información puntual requerida, además de algún tipo de guía para el manejo de las fuentes. Para solucionar este problema se diseñó un servicio de referencia digital para la biblioteca en mención para apoyar a los usuarios no preferenciales y ampliar los canales de interacción y comunicación entre ellos y la biblioteca, basándose en las recomendaciones para el servicio de referencia digital propuesta por la Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA). Con ello, el número de consultas aumentó y el proyecto fue viable para usuarios remotos y presenciales. Además para demostrar ello se realizó una encuesta tomando una muestra de 20 personas, y entre los principales resultados se tuvo lo siguiente:

La Figura 2 muestra que las personas aún no utilizan mucho el Servicio de Referencia Digital (SRD), pero por lo menos lo han utilizado una vez.

La Figura 3 muestra que la respuesta obtenida por el Servicio de Referencia Digital (SRD) es adecuada en un 91%, y esto ayuda mucho en la validación de dicho proyecto.

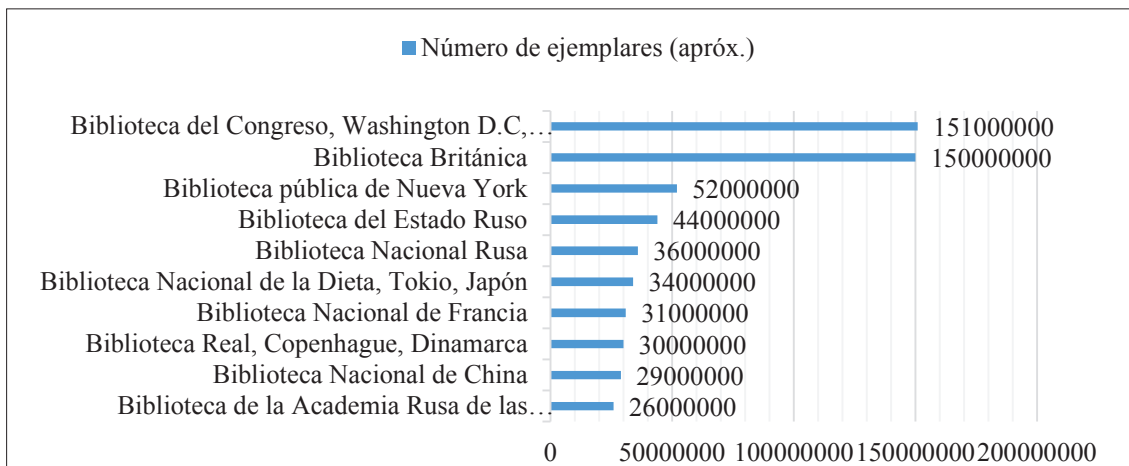


Figura 1. Número de ejemplares de las principales bibliotecas del mundo [1]

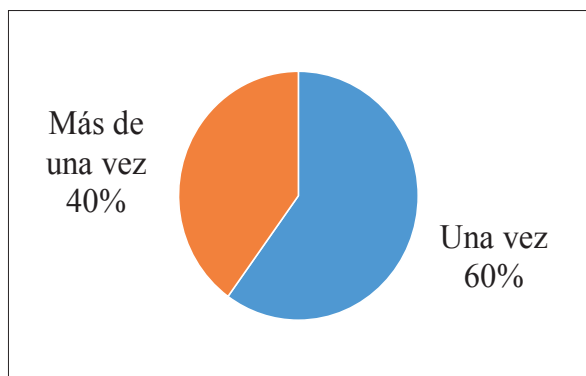


Figura 2. Frecuencia de Uso del SRD [2]

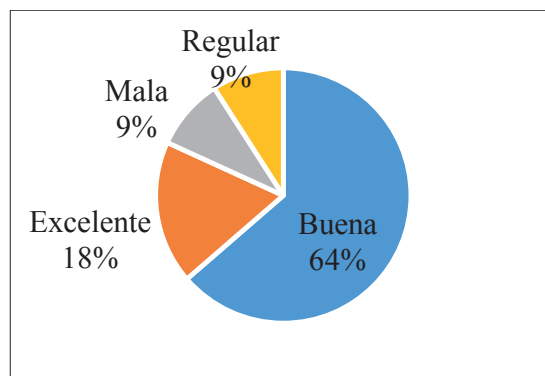


Figura 3. Calificación de la respuesta obtenida [2]

Otro de los trabajos de investigación fue realizado en la Universidad de Tsukuba, en Japón como indica [3], con el fin de mostrar la efectividad de usar el récord de préstamos de libros e información de contenido de los libros para los sistemas de recomendación de libros, para ello participaron 32 estudiantes que pertenecen a la Universidad de Tsukuba con una muestra de libros de la misma universidad. Se obtuvieron 2'324,418 récords de préstamos de la biblioteca de la universidad que van del 2 de Enero del 2006 al 31 de Marzo del 2012, de todos esos libros, 999,630 fueron revisados por estudiantes de pregrado, 1'294,012 libros por estudiantes graduados, y 30,776 fueron revisados por otro tipos de patrones. Finalmente se demostró que considerar el código Nippon Decimal Clasification (NDC), el título del libro y el récord de préstamo del mismo tuvo una mayor proporción de recomendación.

Por otro lado, según indica [4] los servicios de referencias en China se originaron desde el siglo XX. En la década de 1920 Tsing Hua University en Beijing fundó el primer departamento de referencia moderno en China. A finales de los años 1970 y principios de los años 1980 cuando finalizó los tumultos por los movimientos políticos, China proporcionó una serie de reglamentos relacionados a los servicios de bibliotecas, algunos de ellos están relacionados a los servicios de las bibliotecas académicas tales como Work Regulations for Libraries in Higher Education in the People's Republic of China

(1981) y Regulations for Libraries in Regular Institutions in Higher Education (1987).

Como caso específico están las bibliotecas de la UNMSM, por ejemplo, la demanda de recursos de información para el año 2016 en la Facultad de Medicina fue un total de 90,070 consultas como muestra el Apéndice 1 (véase Tabla 3). Entre todas las facultades existen 461,264 consultas para dicho año. La demanda de ítems de bibliotecas, tanto entre libros y tesis, es significativamente grande. Por lo tanto, el usuario debe tener la seguridad que el ítem consultado debe cumplir con sus requerimientos.

El problema a resolver es el siguiente: Los usuarios de las bibliotecas de la UNMSM no cuentan con una ayuda automatizada para poder seleccionar un libro o tesis más adecuada en función a un requerimiento específico. El presente trabajo de investigación ayudará a los alumnos, docentes y egresados de la UNMSM a contar con dicha herramienta automatizada.

Como un caso de estudio se aplicará el sistema de recomendación de libros y tesis en la Biblioteca de la Facultad de Ingeniería de Sistemas e Informática, esto ayudará a los estudiantes de dicha facultad para poder seleccionar el libro y tesis más adecuada para un tema determinado. Posteriormente y con la experiencia ganada, se replicará en las demás facultades de la UNMSM.

**Tabla 3.** Consultas a los recursos de información por tipo de usuario según biblioteca, 2016 [15]

Biblioteca	Total	Tipo de usuario				
		Pregrado	Postgrado	Egresado/ Graduado	Docente	Administrativo
<b>Total</b>	<b>461264</b>	<b>388384</b>	<b>43769</b>	<b>20002</b>	<b>7580</b>	<b>1529</b>
Letras y Ciencias Humanas	11594	10892	376	0	308	18
Derecho y Ciencia Política	22960	21821	920	0	160	59
Medicina	90070	85882	2204	1092	769	123
Ciencias Matemáticas	11709	11681	8	0	16	4
Ciencias Biológicas	35903	34972	794	0	137	0
Ciencias Económicas	23295	23229	4	0	57	5
Farmacología y Bioquímica	3304	2955	164	0	176	9
Odontología	1367	1069	235	0	57	6
Educación	3492	1936	1262	0	274	20
Medicina Veterinaria	9572	9252	215	0	102	3
Química e Ingeniería Química	14431	14276	62	0	93	0
Ciencias Administrativas	18306	16769	977	0	363	197
Ciencias Contables	46758	25414	19056	0	1841	447
Ciencias Físicas	5601	5481	73	0	47	0
Ciencias Sociales	16227	15705	375	0	147	0
Ing. Geol. Min. Metal. y Geo.	2556	1978	211	0	292	75
Ingeniería Industrial	25036	24307	370	0	222	137
Psicología	27580	16890	10412	0	264	14
Ing. de Sistemas e Informática	371	290	81	0	0	0
Biblioteca Central	90281	63307	5835	18608	2158	373
Museo de Historia Natural	851	278	135	302	97	39

El resto de este trabajo está organizado de la siguiente manera. En la sección 2 se muestra el Estado del Arte, donde se justifica la elección de la técnica seleccionada y se describe algunos casos de éxito con otras técnicas para un problema similar. En la sección 2, 3 y 4 se detalla la base teórica del diseño de la solución. En la sección 5 se explica la técnica seleccionada y los casos de éxito, donde se explica algunos trabajos para solucionar problemas similares. La sección 6 explica la propuesta de solución para el presente trabajo. Finalmente se da algunas conclusiones a las que se llegó durante el proceso de diseño de la solución.

## 2. Sistemas de recomendación

### 2.1. Definición

Según [5], “un sistema de recomendación es un conjunto de técnicas de recuperación de información que intenta descubrir el interés de los usuarios por determinados objetos, con la finalidad de ofrecerles un conjunto de objetos afines, relacionados a su perfil, en los que podría estar interesado”.

### 2.2. Fases del proceso de recomendación

**2.2.1. Fase de recolección de información:** Según [6], en dicha fase se recopila la información relevante del usuario para generar un perfil o modelo de usuario para las tareas de predicción, donde se incluye los atributos o datos del usuario, el comportamiento o el contenido de los recursos a los que accede el usuario.

**2.2.2. Fase de aprendizaje:** Según [6], se aplica un algoritmo de aprendizaje para filtrar y explotar las características de los usuarios de las votaciones que se reunieron en la fase de recolección de información.

**2.2.3. Fase de predicción o recomendación:** Según [6], esta fase recomienda o predice que tipo de ítems el usuario puede preferir, esto puede ser hecho basado en el conjunto de datos que se recogieron en la primera fase el cual puede estar basado en memoria o basado en modelo o a través de observaciones del sistema de las actividades del usuario.

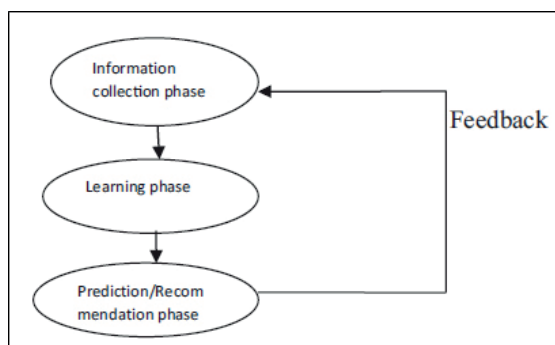


Figura 4. Fases de la recomendación [6]

En la Figura 4 se puede visualizar las fases del proceso de recomendación que se ha descrito partiendo de la recolección de los datos, seguidamente de la

fase de aprendizaje, para luego realizar la recomendación del correspondiente ítem para el usuario, finalmente esta recomendación servirá para retroalimentar en la fase de recolección de información.

### 2.3. Técnicas de filtrado de recomendación

**2.3.1. Filtrado basado en contenido:** Según [6], esta técnica es un algoritmo de dominio dependiente y se enfatiza más en el análisis de los atributos de los ítems para generar predicciones. La técnica es exitosa al tratarse de documentos como páginas web, publicaciones y noticias. La recomendación se basa en los perfiles de usuarios utilizando las características extraídas del contenido de los ítems que el usuario ha evaluado en el pasado. En la Tabla 1 se puede visualizar las ventajas y desventajas principales de esta técnica.

Tabla 1. Ventajas y desventajas de las técnicas de Filtro Basado en Contenido [6]

Ventajas	Desventajas
Esta técnica supera los problemas del filtrado colaborativo.	Las técnicas de filtrado basado en contenido son dependientes en los metadatos de los ítems.
Tiene la capacidad de recomendar nuevos ítems incluso si no hay calificaciones proporcionados por los usuarios.	Su eficacia depende de la disponibilidad de datos descriptivos.
Ajusta su recomendación en un corto espacio de tiempo cuando el usuario cambia sus preferencias.	Los usuarios sólo pueden conseguir recomendaciones similares a los ya definidos en su perfil.

**2.3.2. Filtrado colaborativo:** Según [6], es una técnica de predicción por contenido que no es descrito fácil y adecuadamente por los metadatos. Trabaja con la construcción de una base de datos (matriz usuario-ítem) de preferencias por ítems de usuarios. Cada usuario construye un grupo llamado vecindad. Un usuario obtiene recomendaciones de ítems que no ha clasificado pero que ya fueron clasificados por usuarios que están en esta vecindad. La predicción es un valor numérico  $R_{ij}$  que expresan un puntaje predicho del ítem  $j$  por el usuario  $i$ , mientras que la Recomendación ( $R$ ) es una lista de los primero  $N$  ítems que al usuario le interesará como se muestra en la Figura 5. Esta técnica puede ser dividida en dos categorías: basado en memoria y basado en un modelo.

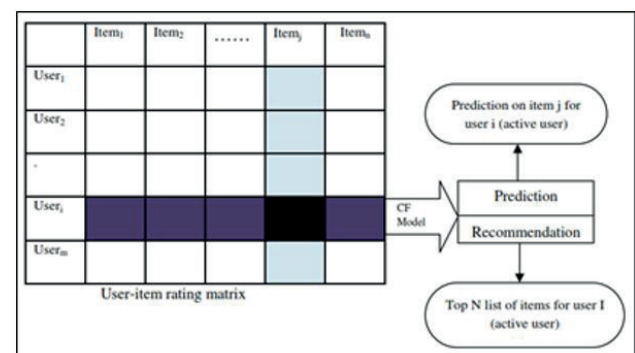


Figura 5. Proceso del filtrado colaborativo [6]

### 3. Ontologías

#### 3.1. Definición

Según [7], “una ontología es la especificación explícita de una conceptualización”.

#### 3.2. Componentes de la ontología

Según [7], una ontología consta de un conjunto de atributos que describen conceptos propios o heredados, cuyos componentes son los siguientes:

**3.2.1. Conceptos:** Es cualquier cosa a cerca de la cual se puede afirmar algo, puede ser tangible o intangible. En la Figura 6 se muestra un ejemplo de una ontología representado mediante un diagrama UML, donde los conceptos están representados por los rectángulos y cada uno contiene una lista de atributos.

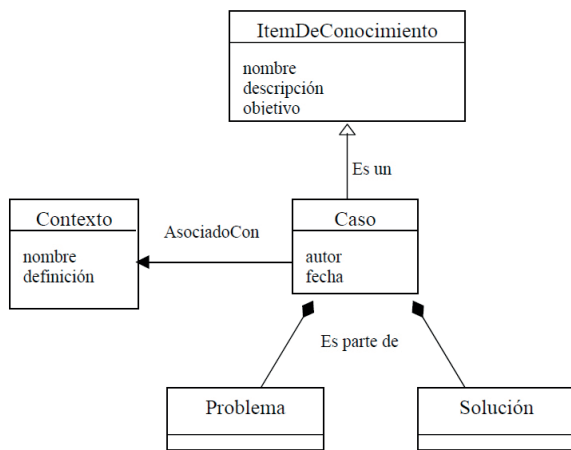


Figura 6. Conceptos, atributos y relaciones en una ontología [7]

**3.2.2. Relaciones:** Según [7], representan el tipo de interacción entre conceptos de un dominio. De manera formal, se representa de la siguiente forma  $R: C_1 \times C_2 \times \dots \times C_n$ , es decir, son definidos como subconjuntos del producto cartesiano de  $n$  conjuntos. En la Figura 6, la relación **AsociadoCon** vincula los conceptos **Caso** y **Contexto**, esto significa, cada Caso se asocia con el Contexto al que se aplica.

**3.2.3. Funciones:** Según [7], son un tipo especial de relaciones donde el  $n$ -ésimo elemento de la relación es único para los  $n-1$  anteriores. De manera formal, se define como  $F: C_1 \times C_2 \times \dots \times C_{n-1} \times C_n$ . En la Figura 6, la relación **AsociadoCon** es una función porque para cada Caso, el Contexto en el que ocurrió es único.

**3.2.4. Axiomas:** Según [7], modelan verdades que siempre se cumplan en la realidad que se ha modelado.

Los axiomas son estructurales o no estructurales. El primero establece condiciones relacionadas a las jerarquías de la ontología, conceptos y atributos

definidos, por ejemplo, en la Figura 6 “Un Problema no puede a la vez ser una solución”. El segundo establece relaciones entre los atributos de un concepto. Por ejemplo, en la Figura 6 “El atributo nombre de caso no puede ser nulo”.

**3.2.5. Instancias:** Según [7], representan elementos específicos en el dominio de la ontología. Por ejemplo, una instancia del Caso sería Demanda de libros en la universidad.

### 4. Tesauros

#### 4.1. Definición

Según [8], “los tesauros son intermediarios en un proceso de información que ayuda al usuario a representar sus necesidades de información mediante el mismo sistema de descriptores que, previamente, sirvió para indizar o representar el contenido de los documentos”.

#### 4.2. Composición general

Generalmente, según [9], los tesauros pueden contener:

- Lista de términos ordenados de forma alfabética y temática.
- Lista de sinónimos de esos términos.
- Jerarquía o relaciones entre términos.
- Breves definiciones de los términos.
- Conjuntos de reglas para usar el tesoro.

#### 4.3. Elementos

Los elementos de los tesauros que se muestran en [9] son los siguientes:

- Unidades lexicales: Son grupos de descriptores que contienen términos de indización por algunos campos o temas o clase de términos.
- Descriptores: Son expresiones o palabras del lenguaje natural que el tesoro utiliza para designar conceptos representativos del documento.
- No descriptores: Son sinónimos de los descriptores o términos que designan conceptos afines a los que cubren los descriptores.

### 5. Estado del Arte

#### 5.1. Caso de éxito:

**5.1.1. Diseño de un Sistema de Recomendación en Repositorios de Objetos de Aprendizaje Basado en la Percepción del Usuario - Caso RODAS [10]:** Este trabajo fue realizado por: Manuel Fernando Caro Piñeres, Jaime Hernández y Jaime Hernández. Dicho trabajo fue elaborado para profesores y es-

tudiantes de la Facultad de Educación y Ciencias Humanas de la Universidad de Córdoba. Se realizó un sistema de recomendación (SR) de objetos de aprendizaje (OA) en repositorios. Además se validó el SR utilizando el repositorio RODAS. El SR está basado en filtrado colaborativo con una adaptación del algoritmo k-vecinos. El sistema utiliza la percepción de usabilidad y utilidad que usuario tiene de los OA. Para validar el sistema se trabajó con un Repositorio de Objetos de Aprendizaje (ROA), muy utilizado por los docentes de educación básica primaria y básica secundaria adscritos a la Secretaría de Educación Municipal de Montería – Córdoba, denominado RODAS. Para validar el sistema se seleccionaron 10 usuarios representados como  $U = \{u1, u2, u3, u4, u5, u6, u7, u8, u9, u10\}$ , el tiempo de la validación fue de 5 días, se tuvo en cuenta los criterios de percepción de usabilidad, accesibilidad y pertinencia que los usuarios tienen acerca de los OA que descargan del repositorio, la valoración que le dan los usuarios varía de 1 a 5.

En la Tabla 2, la columna NOR-V representa el número de OA recomendados que fue descargado y valorado por el usuario, la columna NOR – 4-5 representa el número de objetos recomendados y que fueron valorados por el usuario entre 4 y 5, de forma similar con la columna NOR – 3-4, la columna, la columna % 4 a 5 representa el porcentaje del número de OA, de la columna NOR-V, que fueron valorados con votos entre 4 a 5, la columna % 3 a 4 representa el porcentaje del número de OA que fueron valorados con votos entre 3 a 4. En la Tabla 2, el 88.3 % de las votaciones dadas a los OA fue alta (de 4 a 5 puntos), ello indica que las recomendaciones fueron acordes a las necesidades del usuario y el algoritmo propuesto genera una recomendación acertada.

Según [10], se concluyó que el sistema de recomendación cumplió con las expectativas de recomendaciones de OA relacionadas con la usabilidad, accesibilidad y pertinencia. Además el algoritmo propuesto es adecuado para detectar las percepciones de los usuarios vecinos y ofrecerlas con eficacia a los nuevos usuarios, con un ahorro de tiempo y alto nivel de garantía de calidad y confiabilidad.

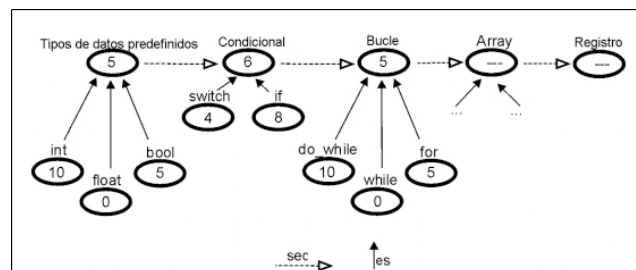
**Tabla 2.** Cuadro de aceptación de objetos recomendados [10]

Usuarios	NOR-V	NOR – 4-5	% 4 a 5	NOR – 3-4	% 3 a 4
u1	12	10	88.3	2	16.7
u2	5	4	80	1	20
u3	14	11	78.6	3	21.4
u4	6	6	100	0	0
u5	5	5	100	0	0
u6	21	18	85.7	3	14.3
u7	8	8	100	0	0
u8	9	7	77.8	2	22.2
u9	11	10	90.9	1	9.09
u10	15	13	86.7	2	13.3

5.1.2. *Personalización en Recomendadores Basados en Contenido y su Aplicación a Repositorios de Objetos de Aprendizaje [11]:* Este trabajo fue realizado por A. Ruiz-Iniesta, G. Jiménez-Díaz y M. Gómez-Albarrán. Este trabajo describe un enfoque novedoso que fomenta una recomendación basada en el contenido personalizado de objetos de aprendizajes (LO). Este enfoque da más prioridad a aquellos LO que son más similares a los objetivos de aprendizaje a corto plazo del estudiante y a la vez tener una utilidad pedagógica. El trabajo presenta una recomendación basada en contenido. La recomendación se compone de dos pasos: la etapa de recuperación y ordenación. En la primera etapa se busca los LO que coinciden con la búsqueda del usuario y con su perfil, luego son filtrados y aquellos que cubren los conceptos de la ontología “listos para ser explorados” por el estudiante serán considerados en la segunda etapa; se dice que un concepto está “listo para ser explorado” por un estudiante si de acuerdo a su perfil cumple con cualquiera de las dos condiciones: es un concepto ya explorado por el estudiante o es un concepto que el estudiante aún no ha explorado pero puede descubrirlo, es decir si un concepto  $c_1$  precede a un concepto  $c_2$ , un estudiante puede descubrir  $c_2$  si el nivel de competencia del estudiante alcanzado en  $c_1$  supera cierto umbral de progreso. En la segunda etapa, ordena los LO recuperados de acuerdo a la calidad asignada a cada uno de ellos.

La Figura 7, muestra parte de la ontología del campo Programación propuesta por este trabajo, donde se refleja que el concepto Bucle puede ser explorado luego que se tenga un buen nivel de competencia en el concepto Condicional que puede estar conformado por los términos switch e if.

En conclusión, dicho trabajo describe una aproximación novedosa en recomendación basada en contenidos de LOs que fomentan grandes niveles de personalización, además es aplicable a cualquier otro dominio educativo.



**Figura 7.** Ontología y esquematización de los niveles de competencia de un estudiante [11]

5.2. *Técnica seleccionada*

Como se ha descrito en los casos de éxito, algunas técnicas son efectivas con ítems electrónicos, pero el presente trabajo sólo contempla ítems físicos, por lo

tanto, la técnica que más se ajusta es la técnica de recomendación basada en contenido del punto 5.1.2. Para considerar dicha técnica hay que realizar algunas modificaciones, porque el contenido de un libro se ve reflejado en los descriptores (en algunos casos tesauros), por lo tanto los conceptos que se mencionan en el punto 5.1.2 se remplazarían por los tesauros y descriptores.

Por lo tanto, la técnica seleccionada es la Recomendación basada Ontología asociada a Tesauros por las siguientes razones:

- Genera una recomendación más confiable que las demás.
- Caracteriza las propiedades de los documentos.
- Los ítems sobre los cuales se va a trabajar son libros y tesis, material que se encuentra en físico, por lo que no se puede realizar una búsqueda clásica de palabras que contienen éstos.
- Los tesauros y descriptores reflejan los temas que contienen los libros y tesis, por lo tanto, el usuario siempre tendrá el ítem que necesita.
- Supera varios problemas comunes de las técnicas clásicas de los sistemas de recomendación.

## 6. Propuesta

### 6.1. Modelo ontológico seleccionado para recomendación de libros

Para realizar la recomendación de libros se ha planteado el modelo ontológico de la Figura 8.

El modelo ontológico plantea que todo libro tiene descriptores o tesauros que los identifican y estos a su vez pueden pertenecer a otros libros del mismo tema de investigación. Además estos descriptores pueden ser

pre-requisitos de otros, es decir, para conocer un término (descriptor) debemos conocer previamente otros. Por ejemplo, en un curso de matemáticas, no podemos conocer el término *Derivada* sin antes conocer el concepto *Límite*. De esta forma cada descriptor o tesoro vincula a otros descriptores y/o tesauros. La finalidad de ello es aumentar la base de conocimiento del usuario y que se pueda compartir con otros usuarios. Por lo tanto, si un usuario conoce *Límites*, entonces está preparado para conocer *Derivadas*, por lo tanto el sistema propuesto recomendará libros basándose en ello.

Además, el sistema propuesto recomendará el libro de acuerdo al curso de interés del usuario. Además un usuario puede realizar uno o muchos pedidos de libros y esto ayuda a retroalimentar a la base de conocimiento del usuario y genera más confianza para que otros usuario realicen pedidos de los mismos libros que ha cumplido con los requerimientos de los usuarios anteriores, incluso en el proceso de recomendación se toma en cuenta los descriptores que el usuario ya conoce y de esta manera se realiza una recomendación más personalizada.

### 6.2. Modelo de recomendación de Tesis

Se ha determinado realizar la recomendación de tesis tomando en cuenta los siguientes otros criterios:

- El investigador tiene dos principales necesidades para su tarea: búsqueda de un nuevo tema de investigación o referencias que le ayuden a su investigación.
- De acuerdo al tema de interés del investigador se le puede recomendar trabajos futuros que vayan de acuerdo a la rama a la que se oriente.
- De acuerdo al tema que el investigador esté desarrollando, se le puede recomendar referencias bibliográficas relacionadas a este.

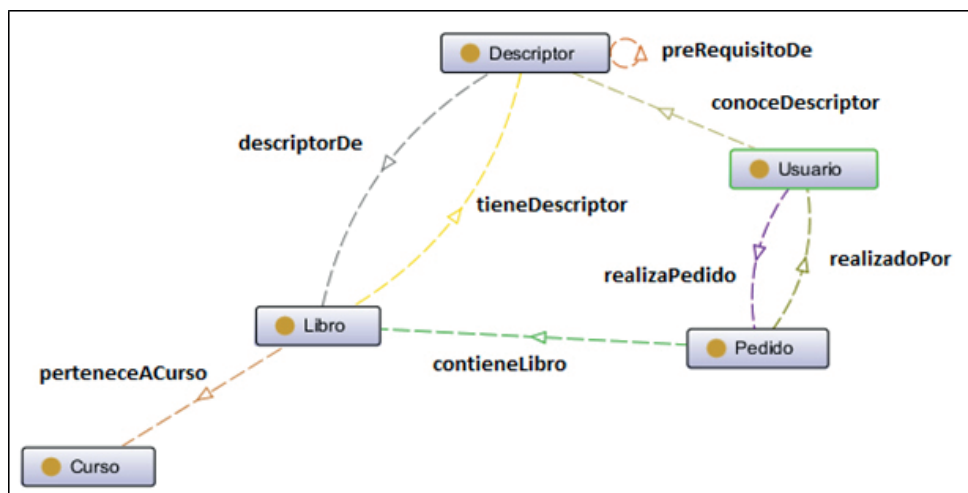


Figura 8. Modelo ontológico seleccionado [Fuente propia]

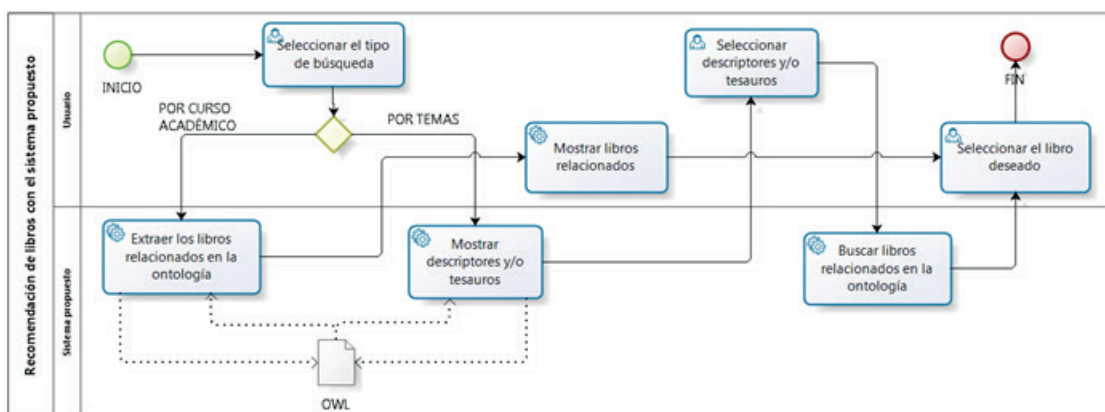


Figura 9. Propuesta de proceso de recomendación de libros [Fuente propia]

### 6.3. Sistema de recomendación propuesto de búsqueda de libros y tesis

6.3.1. *Proceso de recomendación de libros:* En la Figura 9 se muestra el nuevo proceso con el sistema de recomendación propuesto para la búsqueda y recomendación de libros. El proceso se inicia cuando el usuario selecciona el tipo de búsqueda, por curso o tema(s) específico(s). En caso sea un determinado curso, se procederá a buscar los libros que están relacionados a dicho curso en la ontología (OWL). Por otro lado, si el tipo de búsqueda es por temas específicos, el sistema mostrará una relación de descriptores y tesauros que el usuario seleccionará dependiendo de su requerimiento y de su conocimiento previo, seguidamente el sistema extraerá los libros relacionados en la ontología. El sistema mostrará al usuario los libros relacionados, que estarán clasificados de acuerdo a los siguientes criterios:

- Si los usuarios encontraron el tema buscado.
- El libro es teórico y/o práctico.
- El libro es entendible para el lector,
- Los libros de mayor relevancia son los que cumplen los requerimientos del usuario y además vinculan temas relacionados a los seleccionados por el usuario para que aumente su base de conocimiento,
- Los libros de mediana y menor relevancia son los que cumplen exactamente los requerimientos del usuario sin vincular temas relacionados.

Finalmente, el usuario procederá a seleccionar el ítem que más le convenga o que cumpla con sus requerimientos.

Claramente se puede ver que en este proceso se realiza la búsqueda y recomendación con el conocimiento previo del usuario y con el contenido principal de los ítems, porque el interés principal del usuario es el contenido del libro.

Por lo tanto, el libro seleccionado cumplirá con los requerimientos específicos del usuario, y será de gran ayuda para el mismo.

6.3.2. *Proceso de recomendación de tesis:* La Figura 10 muestra el proceso de recomendación de tesis con el sistema propuesto. El proceso inicia cuando el usuario selecciona el tipo de búsqueda a realizar, que puede ser por tema de tesis o por referencia. En el caso que el usuario realice una búsqueda por tema de tesis, es decir, para que el usuario encuentre su tema de tesis, el sistema solicitará los temas de interés al usuario, el usuario seleccionará los temas de interés y el sistema buscará en la ontología OWL las tesis relacionadas a los temas de interés seleccionados. Finalmente, el sistema mostrará los trabajos futuros de las tesis relacionadas. De esta manera, el usuario se ahorrará el trabajo de buscar ítem por ítem sin saber cuál es aquella que puede ayudarlo en encontrar su tema de tesis.

En el caso que el usuario realice una búsqueda por referencia, es decir, el usuario desea encontrar referencias que apoyen o sustenten su proyecto de tesis, el usuario seleccionará el tema de tesis y el filtro de años para buscar su referencia bibliográfica, el cual es de 5 a 10 años. Seguidamente el sistema realizará la búsqueda de tesis por el filtro seleccionado. Finalmente, el sistema mostrará las tesis relacionadas.

Claramente, se puede ver que este proceso ahorra el trabajo al usuario de buscar ítem por ítem para encontrar una tesis que ayude a sustentar su proyecto de tesis o de investigación, porque la búsqueda es más directa, sólo centrándose en el tema de tesis que realiza el usuario.

### 6.4. Diseño del OWL

6.4.1. *Herramienta para consulta del archivo OWL:* Según [12], el SPARQL (SPARQL Protocol and RDF Query Language) se utiliza para consultar diversas fuentes de datos si los datos están almacenados de forma nativa como RDF o vistas RDF.



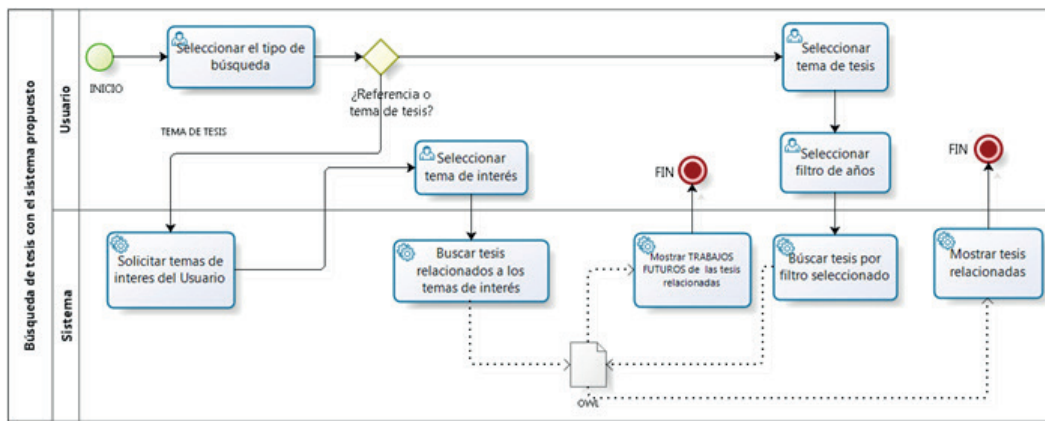


Figura 10. Propuesta de Proceso de recomendación de tesis [Fuente propia]

Según [13], la W3C conformó un grupo de trabajo para publicar el Web Ontology Language (OWL), una familia de lenguajes que extendía la funcionalidad presente en el Marco de Descripción de Recursos (RDF por sus siglas en inglés), la cual fue aceptada como una Recomendación de la W3C el 10 de Febrero de 2004.

Por lo tanto, la herramienta de consulta del OWL utilizado en el presente trabajo de investigación será el SPARQL.

6.4.2. Creación de instancias del OWL: La creación de las instancias del OWL será de forma manual y automática, la primera mediante un registro en el sistema, la segunda por medio de un archivo en formato Excel 97-2003 (.xls). La persona encargada de la biblioteca estará a cargo de ello.

6.4.3. Creación de relaciones entre descriptores: Teniendo en cuenta que los descriptores y/o tesauros se encuentran en los libros, además el contenido del libro está reflejado en el índice del mismo, dicho índice será el punto de partida para la creación de los descriptores y/o tesauros. Los conceptos del índice del libro están ordenados de acuerdo a una secuencia de aprendizaje. La Figura 11 muestra un fragmento del índice de un libro, donde se aprecia que el descriptor “Límite de funciones” se explica antes del descriptor “Derivadas de funciones” porque es necesario entender el primer descriptor antes de conocer el segundo, es decir están relacionados.

Índice	
1. Límites de funciones.....	2
2. Derivadas de funciones.....	3

Figura 11. Ejemplo de índice de libro [Fuente propia]

De acuerdo a la Figura 8, existe una relación *preRequisitoDe* que va de la entidad Descriptor a sí misma, dicha relación se ve reflejada en la Figura 13 entre los descriptores anotados con el nombre *LIMITE\_DE\_FUNCIONES* (Límite de funciones)

y *DERIVADA\_DE\_FUNCIONES* (Derivada de funciones), donde el primero es prerequisite del segundo, según la Figura 11. Pueden existir más de un prerequisite para un descriptor y viceversa.

En caso, el descriptor pueda pertenecer a diferentes campos de estudio, por ejemplo, el descriptor *Límites*, perteneciente al campo de la matemática y de la geografía, el sistema propuesto no permitirá el registro del descriptor más de una vez, por lo tanto, se sugiere que el usuario ingrese un nombre similar que los diferencie, por ejemplo, *Límites de funciones* y *Límite geográfico*.

En la recomendación del libro se vinculará un descriptor por cada uno seleccionado por el usuario, por ejemplo, si el usuario selecciona el descriptor *LIMITE\_DE\_FUNCIONES* el sistema tendrá en cuenta dicho descriptor y los descriptores que se desprendan de este; según la Figura 12, el sistema tendrá en cuenta el descriptor *DERIVADA\_DE\_FUNCIONES*. De esta forma el usuario tendrá una recomendación que le permitirá aprender más conceptos de lo que él requiere.

Pueden existir muchas relaciones *preRequisitoDe*, por ejemplo podría existir otra relación que vaya desde *DERIVADA\_DE\_FUNCIONES* hacia otro descriptor. Incluso, el descriptor *LIMITE\_DE\_FUNCIONES* puede ser prerequisite de otro descriptor.

6.4.4. Creación de relación entre libros, descriptores y cursos: La Figura 12 muestra que los descriptores *LIMITE\_DE\_FUNCIONES* y *DERIVADA\_DE\_FUNCIONES* pertenecen al libro etiquetado con el nombre *L\_TOPICOS\_DE\_CALCULO* (libro Tópicos de Cálculo), esto ayuda a la recomendación del libro a partir de los descriptores seleccionados por el usuario; de esta forma el sistema recomendará libros que cumplan con los requerimientos del usuario.

La Figura 12 muestra que el libro *L\_TOPICOS\_DE\_CALCULO* está relacionado al curso etiquetado con el nombre *C\_ANALISIS\_MATEMATICO* por medio de la relación *perteneceACurso*, esto

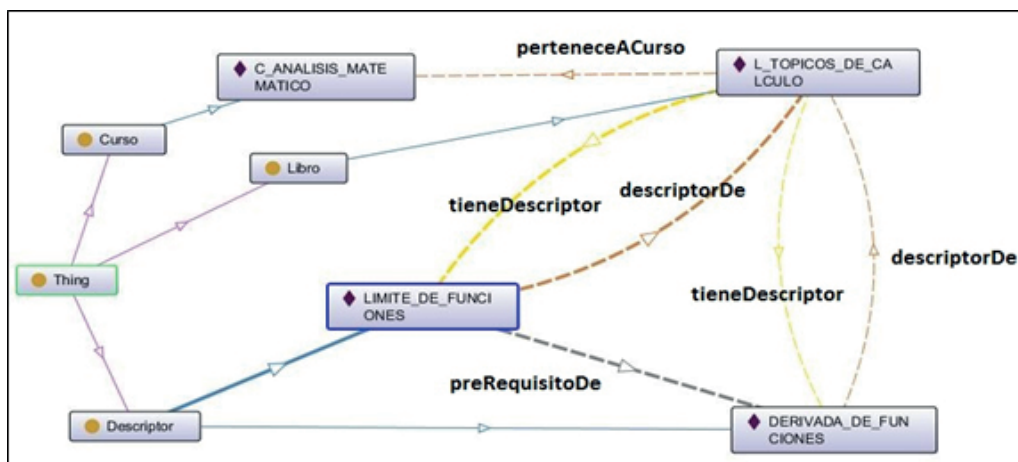


Figura 12. Representación gráfica de relación entre libro, descriptores y cursos [Fuente propia]

ayuda a la recomendación del libro por medio del curso.

6.5. Evaluación de rendimiento del sistema propuesto

Según [14], las métricas de decisión evalúan la efectividad de un sistema de recomendación, es decir, la frecuencia en que el sistema efectúa recomendaciones correctas. Por ello se asume que el proceso de recomendación es binario: o el elemento recomendado agrada al usuario o no le agrada.

Una de estas métricas es “Precision and Recall”. **Precision** es la probabilidad de que un elemento seleccionado sea relevante y **Recall** la probabilidad que se seleccione un elemento relevante.

Lo que se propone realizar para el sistema propuesto es definir un sistema de calificación de los ítems que fueron seleccionados por el usuario, el usuario cuando realice la devolución del ítem pedido tendrá que calificar el ítem como condición necesaria. Los puntajes se encontrarán en el rango de 0 a 10; se definirá un umbral de 6, de modo que los ítems que superen dicho umbral, se considerará que la recomendación agradó al usuario, sino la recomendación no agradó al usuario. A partir de ello, aplicaremos la métrica de “Precision and Recall”. La Figura 13 muestra el proceso de evaluación de un libro de acuerdo a determinado criterio.

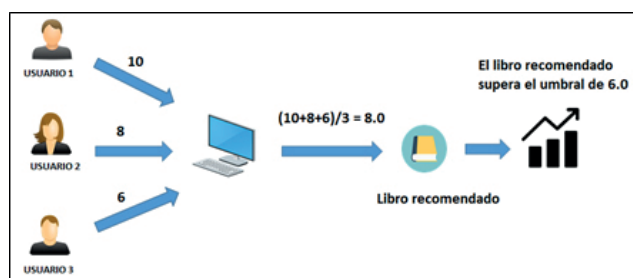


Figura 13. Flujo de evaluación para un determinado criterio [Fuente propia]

7. Conclusiones y trabajos futuros

7.1. Conclusiones

Como conclusiones del presente artículo, se tienen las siguientes:

- El modelo ontológico seleccionado en la recomendación de libros es efectivo porque ayuda a almacenar su contenido en los descriptores (tesauros en algunos casos) y a su vez almacena la base del conocimiento del usuario que irá aumentando.
- El modelo seleccionado de recomendación de tesis es efectivo porque está centrado en lo que es importante para el investigador que son las referencias bibliográficas y posibles temas de investigación mediante los trabajos futuros sugeridos.
- La implementación de un sistema de recomendación de libros es diferente a uno de recomendación de tesis, porque el objetivo de la recomendación de libros es aumentar el conocimiento del usuario, mientras que el de la recomendación de tesis es ayudar al usuario a seleccionar un tema de investigación adecuado y contar con una referencia bibliográfica ajustada a su tema de investigación.

7.2. Trabajos futuros

- Como trabajo futuro, el sistema podrá captar las opiniones de los lectores y la calificación que hacen de los textos o tesis leídas.
- Como trabajo futuro, el sistema podrá extenderse con más métricas de recomendación para tomar en cuenta.
- Como trabajo futuro, podrá realizarse la comunicación del sistema propuesto con el sistema actual de biblioteca en el tema de los pedidos y la comunicación con las demás bibliotecas de

la UNMSM y sobre todo la Biblioteca Central para que la recomendación sea más personalizada en cualquier facultad de la UNMSM.

## 8. Referencias

- [1] Algarabía, Top 10: Las bibliotecas más grandes del mundo, <http://algarabia.com/top-10/top-10-las-bibliotecas-mas-grandes-del-mundo/>, Octubre 2017
- [2] Sandra Naranjo, Melissa Palacios, Diseño de un Servicio de Referencia Digital en la Biblioteca María Cristina Niño Michelsen del Politécnico Gran Colombiano, 2008, Pontificia Universidad Javeriana, Facultad de Comunicación y Lenguaje – Carrera de Ciencias de la Información, Bogotá D.C. <http://hdl.handle.net/10554/5432>
- [3] Keita T., Nobuya T., Sho S., et al.: 'Book Recommendation Based on Library Loan Records and Bibliographic Information', 2014, Procedia – Social and Behavioral Sciences 147 (2014) pp. 478-486. DOI: <https://doi.org/10.1016/j.sbspro.2014.07.142>
- [4] Jingzhen Xie, LiliSun: 'Exploring Chinese Students' Perspective on Reference Services at Chinese Academic Libraries: A Case Study Approach', The Journal of Academic Librarianship 41 (2015), pp. 228–235. DOI: <https://doi.org/10.1016/j.acalib.2015.04.002>
- [5] Edward Núñez, Sistemas de Recomendación de Contenidos para Libros Inteligentes, Tesis Doctoral, Universidad de Oviedo, Sistemas y servicios informáticos para Internet, 2012. <http://di002.edv.uniovi.es/~cueva/investigacion/tesis/Tesis-Edward.pdf>
- [6] F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh: 'Recommendation systems: Principles, methods and evaluation', Egyptian Informatics Journal, 2015, 16, pp. 261–273. DOI: <https://doi.org/10.1016/j.eij.2015.06.005>
- [7] Dr. Luis Olsina, Dr. Gustavo Rossi, Memoria Organizacional Basada en Ontologías y Casos para un Sistema de Recomendación en Aseguramiento de Calidad, 2010, Facultad de Informática, Universidad Nacional de La Plata – Argentina. [http://postgrado.info.unlp.edu.ar/wp-content/uploads/2014/07/Martin\\_Maria\\_de\\_los\\_Angeles.pdf](http://postgrado.info.unlp.edu.ar/wp-content/uploads/2014/07/Martin_Maria_de_los_Angeles.pdf)
- [8] Lluís Codina, Rafael Pedraza-Jiménez: 'Tesauros y ontologías en sistemas de información documental', El profesional de la información, 2014, septiembre-octubre, v. 20, n. 5, pp. 555-563. DOI: <http://dx.doi.org/10.3145/epi.2011.sep.10>
- [9] María Silvestrini, 'TESAURO: RESUMEN EXPLICATIVO', 2007, Universidad Interamericana de Puerto Rico, Recinto de Ponce, Centro de Acceso a la Información, Oficina Desarrollo Destrezas de Información. <http://ponce.inter.edu/cai/manuales/TESAURO.pdf>
- [10] Manuel F. Caro-Piñeres, Jaime Hernández, Jovani A. Jiménez-Builes: 'Diseño de un Sistema de Recomendación en Repositorios de Objetos de Aprendizaje Basado en la Percepción del Usuario: Caso RODAS', Ciencia e Ingeniería Neogranadina., 2011, Vol. 21-1, pp. 51-72. <https://revistas.unimilitar.edu.co/index.php/rcin/article/view/270/1909>
- [11] A. Ruiz-Iniesta, G. Jiménez-Díaz y M. Gómez-Albarrán: 'Personalización en Recomendadores Basados en Contenido y su Aplicación a Repositorios de Objetos de Aprendizaje', IEEE-RITA Vol. 5, 1, 2010, pp. 33-38. [gaia.fdi.ucm.es/aigaion2/index.php/attachments/single/72](http://gaia.fdi.ucm.es/aigaion2/index.php/attachments/single/72)
- [12] SPARQL Query Language for RDF, W3C Recommendation 15 January 2008. <https://www.w3.org/TR/rdf-sparql-query/>
- [13] David E. Narváez: 'Las Ontologías y el Lenguaje OWL 2', Prisma Tecnológico, 2010, Vol. 2, pp. 31-34.
- [14] Sergio M. Galán Nieto, 'Filtrado Colaborativo y Sistemas de Recomendación', 2007, Universidad Carlos III de Madrid. <https://slidex.tips/downloadFile/filtrado-colaborativo-y-sistemas-de-recomendacion>
- [15] UNMSM Compendio Estadístico 2017. <http://ogpl.unmsm.edu.pe/compendios/Compendio2017/Cap%209-2017/9.32.xlsx>

