

**UNIVERSIDAD POLITÉCNICA SALESIANA**  
**SEDE CUENCA**

**CARRERA DE INGENIERÍA ELECTRÓNICA**



Proyecto Técnico con enfoque investigativo:

DISEÑO E IMPLEMENTACIÓN DE UN PROTOTIPO PARA  
ANÁLISIS DE COMENTARIOS PARA LA DE RECOMENDACIÓN  
DE CONTENIDO Y FORMACIÓN DE GRUPOS DE APRENDIZAJE  
EN UNA PLATAFORMA EDUCATIVA

TRABAJO DE TITULACIÓN PREVIO A LA  
OBTENCIÓN DEL TÍTULO DE INGENIERO  
ELECTRÓNICO

AUTOR:

Pinguil Sánchez Diego Alejandro

TUTOR:

Dr. Ing. Jack Bravo T.

CUENCA – ECUADOR

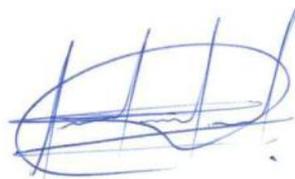
2017

## **CESIÓN DE DERECHOS DE AUTOR**

Yo, Diego Alejandro Pinguil Sánchez con número de cédula CI. 0302568456, manifiesto mi voluntad y cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del trabajo de titulación: **“DISEÑO E IMPLEMENTACIÓN DE UN PROTOTIPO PARA ANÁLISIS DE COMENTARIOS PARA LA DE RECOMENDACIÓN DE CONTENIDO Y FORMACIÓN DE GRUPOS DE APRENDIZAJE EN UNA PLATAFORMA EDUCATIVA”**, mismo que ha sido desarrollado para optar por el título de: Ingeniero Electrónico, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En aplicación a lo determinado en la Ley de Propiedad Intelectual, en mi condición de autor me reservo los derechos morales de la obra antes citada. En concordancia, suscribo este documento en el momento que hago entrega del trabajo final en formato impreso y digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, 11 de diciembre de 2017



Diego Alejandro Pinguil Sánchez.

CI: 0302568456

AUTOR

## CERTIFICACIÓN

Yo declaro que bajo mi **tutoría** fue desarrollado el trabajo de titulación: **“DISEÑO E IMPLEMENTACIÓN DE UN PROTOTIPO PARA ANÁLISIS DE COMENTARIOS PARA LA DE RECOMENDACIÓN DE CONTENIDO Y FORMACIÓN DE GRUPOS DE APRENDIZAJE EN UNA PLATAFORMA EDUCATIVA”**, realizado por Diego Alejandro Pinguil Sánchez, obteniendo el **Proyecto Técnico con enfoque investigativo** que cumple con todos los requisitos estipulados por la Universidad Politécnica Salesiana.

Cuenca, 11 de diciembre de 2017



Dr. Ing. Jack Fernando Bravo Torres.

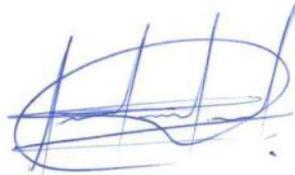
CI: 1103708697

**DIRECTOR DEL TRABAJO DE TITULACIÓN**

## **DECLARATORIA DE RESPONSABILIDAD**

Yo, Diego Alejandro Pinguil Sánchez con número de cédula CI. 0302568456, autor del trabajo de titulación “**DISEÑO E IMPLEMENTACIÓN DE UN PROTOTIPO PARA ANÁLISIS DE COMENTARIOS PARA LA DE RECOMENDACIÓN DE CONTENIDO Y FORMACIÓN DE GRUPOS DE APRENDIZAJE EN UNA PLATAFORMA EDUCATIVA**” certifico que el total contenido del **Proyecto Técnico con enfoque investigativo**, es de mi exclusiva responsabilidad y autoría

Cuenca, 11 de diciembre de 2017



Diego Alejandro Pinguil Sánchez.

CI: 0302568456

AUTOR

## **AGRADECIMIENTOS**

A mi familia, por todo el apoyo a lo largo de estos años, en especial, a mi madre Rosario.

A mis amigos, por compartir conmigo todo este tiempo, dándome ánimos y por su valiosa ayuda.

Y a mí tutor, Dr. Ing. Jack Bravo T, por facilitarme sus conocimientos y paciencia en el desarrollo de este trabajo.

*Diego Alejandro*

## **DEDICATORIAS**

A Dios, por permitirme llegar hasta este punto y haberme dado salud para lograr mis objetivos, además de su infinita bondad y amor.

A mi madre Rosario. Por su apoyo en todo momento, sus consejos, valores, y la motivación para ser una persona de bien, pero más que nada, por su infinito amor.

**Diego Alejandro**

# ÍNDICE GENERAL

AGRADECIMIENTOS .....	I
DEDICATORIAS .....	II
ÍNDICE GENERAL.....	III
ÍNDICE DE FIGURAS.....	VI
ÍNDICE DE TABLAS .....	VII
ABREVIATURAS .....	VIII
RESUMEN.....	IX
INTRODUCCIÓN .....	X
ANTECEDENTES DEL PROBLEMA DE ESTUDIO.....	XII
JUSTIFICACIÓN .....	XIII
OBJETIVOS .....	XIV
Objetivo General .....	XIV
Objetivos Específicos.....	XIV
Restricción.....	XIV
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA O ESTADO DEL ARTE .....	1
1.1. LA WEB SEMÁNTICA.....	1
1.1.1. World Wide Web Consortium .....	2
1.1.2. Arquitectura de la Web .....	3
1.1.3. Desafíos de la Web actual.....	4
1.1.4. Requisitos para una Web de datos efectiva.....	4
1.1.5. Definición de Web Semántica.....	5
1.2. ARQUITECTURA DE LA WEB SEMÁNTICA .....	5
1.3. ESTÁNDARES DE LA WEB SEMÁNTICA .....	8
1.3.1. El lenguaje RDF.....	8
1.3.2. La noción de URI.....	9
1.3.3. Especificación RDF .....	9
1.4. NOCIÓN DE RDF SCHEMA .....	10
1.4.1. Modelo de datos .....	10
1.4.2. Recomendación RDFS .....	11
1.4.3. Componentes básicos de RDFS .....	11
1.4.4. La noción de Clase .....	12
1.4.5. Noción de instancia de una clase .....	12
1.4.6. Propiedades en RDFS .....	12
1.4.7. Restricciones sobre las propiedades.....	13

1.4.8.	Técnicas de razonamiento en RDF Schema .....	13
1.4.9.	Reglas de inferencia en RDF Schema .....	14
1.4.9.1.	Las subclases .....	14
1.4.9.2.	Las subpropiedades .....	14
1.4.9.3.	Regla de inferencia sobre jerarquías de clases .....	15
1.4.9.4.	Regla de inferencia sobre jerarquías de propiedades .....	15
1.4.9.5.	Regla de inferencia para el dominio de una propiedad .....	15
1.5.	WEB ONTOLOGY LANGUAGE (OWL) .....	16
1.5.1.	Propiedades en OWL .....	17
1.5.2.	Restricciones en OWL .....	17
1.6.	E-LEARNING .....	18
1.6.1.	Componentes de e-Learning .....	18
1.7.	OPPIA .....	20
1.7.1.	Arquitectura de OPPIA .....	21
CAPÍTULO 2: MARCO METODOLÓGICO.....		23
2.1.	ÁMBITO DEL PROBLEMA.....	23
2.2.	DEFINICIÓN DEL PROBLEMA.....	23
2.3.	SISTEMA DE RECOMENDACIÓN .....	23
2.3.1.	Filtro colaborativo.....	23
2.3.2.	Similitud entre usuarios .....	24
2.3.3.	Calificación ponderada de un ítem.....	24
2.4.	PROCESAMIENTO DE TEXTO .....	25
2.5.	Herramientas de Procesamiento del Lenguaje Natural .....	26
2.5.1.	NLTK (Natural Language Toolkit).....	26
2.5.2.	Enfoque estadístico .....	26
2.5.3.	Naive Bayes .....	26
2.5.4.	Enfoque analítico .....	27
2.6.	ANÁLISIS DE SENTIMIENTOS .....	27
2.7.	ARQUITECTURA DEL SISTEMA .....	28
CAPÍTULO 3: IMPLEMENTACIÓN Y ANÁLISIS DE RESULTADOS .....		30
3.1.	DISEÑO DEL SISTEMA .....	30
3.1.1.	Base de datos .....	31
3.1.2.	Servidor Web .....	32
3.1.3.	Interfaz del Usuario .....	32
3.2.	Análisis DE RESULTADOS.....	36
3.2.1.	Recomendación de contenido .....	36
3.2.2.	Formación DE GRUPOS .....	38

3.2.3. Etapa 1, Análisis individual.....	39
3.2.4. Etapa 2, Promedio Global.....	42
CAPÍTULO 4: CONCLUSIONES Y RECOMENDACIONES .....	44
4.1. CONCLUSIONES.....	44
4.2. RECOMENDACIONES .....	45
REFERENCIAS BIBLIOGRÁFICAS.....	46

## ÍNDICE DE FIGURAS

Figura 1.1. Arquitectura de la Web inicial y la Web única.....	3
Figura 1.2. Pila de la Web Semántica. ....	6
Figura 1.3. Características del estándar RDF.....	8
Figura 1.4. Representación de un grafo. ....	9
Figura 1.5. Triple RDF.....	10
Figura 1.6. Componentes básicos de RDFS.....	11
Figura 1.7. Restricción de dominio y rango en RDFS. ....	13
Figura 1.8. Jerarquías de clases en RDFS. ....	14
Figura 1.9. Jerarquías en OWL. ....	17
Figura 1.10. Modelo ADDIE para e-learning. ....	20
Figura 1.11. Arquitectura de la plataforma OPPIA.....	21
Figura 2.1 Análisis de Sentimientos en plataformas educativas.....	25
Figura 2.2 Análisis de Sentimientos en plataformas educativas.....	27
Figura 2.3 Arquitectura del sistema de recomendación y análisis de comentarios....	29
Figura 3.1 Diagrama del sistema para el análisis de cometarios.....	30
Figura 3.2 Ventana Home. ....	33
Figura 3.3 Ventana de Perfil de Usuario.....	34
Figura 3.4 Ventana de consulta de contenidos.....	34
Figura 3.5 Ventana de ítems y calificaciones.....	35
Figura 3.6 Ventana de ítem y comentarios. ....	35
Figura 3.7 Ventana de Recomendación de contenidos para pandrader. ....	37
Figura 3.8 Ventana de Recomendación de contenidos para fochoa]. ....	37
Figura 3.9 Ventana de Formación de grupos para pandrader. ....	38
Figura 3.10 Ventana de Formación de grupos para csarmiento.....	38

## ÍNDICE DE TABLAS

Tabla 3.1 Campos principales de la base de datos .....	31
Tabla 3.2 Detalle del campo status .....	31
Tabla 3.3. Tabla de resultados de calificaciones individuales de cometarios.....	40
Tabla 3.4. Tabla de resultados de calificaciones y sus promedios.....	41

## ABREVIATURAS

CSS	Cascading Stylesheets
GITEL	Grupo de Investigación en Telecomunicaciones y Telemática.
HTML	Hyper Text Markup Language.
HTTP	Hypertext Transfer Protocol
NLTK	Natural Language Toolkit
OPPIA	OPPortunistic Intelligent Ambient learning.
OWL	Web Ontology language.
RDF	Resource Description Framework.
RDFS	RDF Schema
SPARQL	Protocol and RDF Query Language
TIC	Tecnologías de la Información y Comunicación.
URI	Uniform Resource Identifier.
URL	Uniform Resource Locator.
WWW	World Wide Web.
W3C	World Wide Web Consortium.
XHTML	eXtensible HyperText Markup Language
XML	eXtensible Markup Language.

## **RESUMEN**

La Web actual está formada por distintos contenidos (páginas electrónicas, documentos, videos, imágenes) de manera masiva. Estos datos se publican y presentan de diversas formas, además solo las personas son capaces de interpretar esta información. Una aplicación no puede identificar e interpretar los datos que contiene un sitio Web.

El propósito de este proyecto es realizar el análisis y evaluación de comentarios para la recomendación de contenidos (libros, videos, artículos) y formación de grupos de aprendizaje (compañeros con gustos y preferencias similares) en una plataforma educativa, a través de la utilización de los estándares y tecnologías de la Web Semántica y el uso de PLN (Procesamiento del Lenguaje Natural), apoyado en la inteligencia artificial.

La plataforma permite realizar comentarios al momento de compartir una recomendación de contenido. Se implementa un algoritmo que usa herramientas de Procesamiento de Lenguaje Natural, como el NLTK (Natural Language ToolKit). La aplicación genera e infiere conocimiento para realizar una calificación que expresa el grado de satisfacción de un usuario (estudiante, tutor) al momento de efectuar un comentario en una recomendación compartida.

Los resultados de esta implementación proporcionan un parámetro que permiten conocer la aceptación de los recursos y materiales recomendados por el sistema. Hay mayor robustez y eficacia, debido a que el sistema cuenta con una retroalimentación basada en el interés y el grado de satisfacción que tienen los usuarios.

## INTRODUCCIÓN

El presente proyecto se refiere al análisis y evaluación de comentarios para la recomendación de contenidos y la formación de grupos de aprendizaje en una plataforma educativa. El objetivo principal es utilizar la Web Semántica, la cual se basa en dos ideas principales: añadir datos semánticos a los recursos (páginas Web, documentos, videos, imágenes y todo material multimedia) y crear agentes inteligentes capaces de entender e inferir conocimiento a partir de los datos que contienen los recursos mencionados [1, 2].

El desarrollo de esta implementación se basa en el interés de crear y añadir datos semánticos a los recursos (libros, videos, enlaces) con los que cuenta el sistema de recomendación y formación de grupos, desarrollado por el proyecto OPPIA (OPPortunistic Intelligent Ambient learning) perteneciente al Grupo de Investigación en Telecomunicaciones y Telemática (GITEL) de la Universidad Politécnica Salesiana (UPS), sede Cuenca.

El sistema de recomendación y formación de grupos desarrollado permite acceder y navegar a los usuarios de manera sencilla. Los usuarios pueden registrar y editar perfiles, realizar consultas de contenidos y grupos de trabajo además de realizar solicitudes de amistad y formar parte de grupos de trabajo. Al recibir un contenido o acceso a un grupo de trabajo la aplicación brinda la posibilidad de calificar estos recursos mediante una calificación de cinco estrellas y likes. Las calificaciones sirven de retroalimentación para buscar intereses de los usuarios, así el sistema puede realizar sugerencias de amistad, contenido o grupos [3].

En el marco de la implementación se desarrolló un algoritmo que utiliza las herramientas del NLTK (Natural Language ToolKit), para realizar un análisis de sentimientos. Este análisis brinda dos parámetros: la polaridad y la objetividad de un texto. Para medir el grado de satisfacción de una persona se utilizó la polaridad que expresa un comentario. Entonces se consigue calificar el recurso recomendado, dicha calificación utiliza el sistema como parámetro de retroalimentación. La aplicación conoce si la recomendación es útil o necesita ser reemplazada por otro recurso [4, 5].

Este documento se ha dividido en capítulos que agrupan las etapas del proceso de la implementación y el alcance de los objetivos propuestos para este proyecto. A continuación, se resume el contenido de cada capítulo del presente documento.

## **ANTECEDENTES DEL PROBLEMA DE ESTUDIO**

La evolución de las Tecnologías de Información y Comunicación (TIC), han cambiado el método de enseñanza y aprendizaje, eliminando barreras de tiempo, espacios físicos propios de la educación tradicional [3]. Esto ha brindado un nuevo escenario que busca mayor eficiencia y un manejo adecuado de los recursos (enlaces, libros, artículos, material multimedia) dentro de los sitios Web [3, 6].

En la actualidad, la Web ha evolucionado hacia la Web de datos, la cual se basa en la adición de información y ontologías a los recursos almacenados en la World Wide Web [1]. La nueva información describe contenidos, significado y la relación entre los recursos. Los sistemas informáticos ahora disponen de la Web Semántica que tiene lenguajes entendibles por una aplicación y las personas [6, 7].

Al disponer de estas tecnologías, el proyecto OPPIA ha desarrollado un sistema de recomendación de contenido y formación de grupos de trabajo, el cual dispone de recursos que necesitan ser evaluados además la aplicación debe conocer el grado de satisfacción y operabilidad de la plataforma [3].

Existen pocos proyectos que involucran los estándares de la Web Semántica para el análisis de sentimientos que expresa un usuario cuando realiza un comentario. En este proyecto se usa la etiquetación de información para conocer los recursos que dispone la aplicación enfocado en el análisis de sentimientos y el Procesamiento de Lenguaje Natural.

## JUSTIFICACIÓN

La importancia de etiquetar los datos que almacenan los servidores es que las aplicaciones conocen esta información y pueden inferir conocimiento procesable para sí mismas y las personas. Entonces navegar por la Web se vuelve eficiente, preciso y automático; mejorando las búsquedas, consultas y el almacenamiento de datos con mayor rapidez y exactitud [6, 7].

La contribución de este proyecto beneficiará a los estudiantes de la Universidad Politécnica Salesiana-Sede Cuenca, motivando un aprendizaje dinámico usando los recursos de la Web en base a las preferencias de cada usuario.

Este documento se compone de cinco capítulos que abarcan las fases de desarrollo e implementación de datos semánticos a la recomendación de contenidos y formación de grupos de trabajo. La temática de este documento empieza con el Capítulo 1, este contiene el estado del arte correspondiente a la Web Semántica y los entornos e-Learning. Capítulo 2, se detalla la metodología propuesta para el análisis de sentimientos expresados a través de los comentarios efectuados. Capítulo 3, corresponde a la etapa de implementación, además del análisis de resultados de los algoritmos propuestos. Capítulo 4, contiene las recomendaciones y conclusiones que se obtuvieron en el proyecto.

## **OBJETIVOS**

En base a la importancia que representa la creación de plataformas educativas eficientes a través del uso de tecnologías de la Web Semántica se han planteado los siguientes objetivos:

### **OBJETIVO GENERAL**

- Diseñar e implementar un sistema basado en tecnologías de la Web Semántica para la evaluación de comentarios realizados sobre recomendaciones de contenido y formación de grupos de trabajo en entornos e-learning o una plataforma educativa online.

### **OBJETIVOS ESPECÍFICOS**

- Analizar el estado del arte de entornos E-learning y Web Semántica.
- Definir y formalizar una metodología basada en la semántica para la evaluación de comentarios de contenidos en entornos e-learning.
- Diseñar y programar una base de datos.
- Implementar un sistema de recomendación de contenidos para obtener los comentarios respectivos en cada grupo de trabajo o contenido recomendado. El servidor permitirá:
  - Obtener y editar perfil de usuario.
  - Realizar consultas sobre contenidos o grupos de trabajo.
  - Formar parte de grupos de aprendizaje.
  - Realizar comentarios sobre contenidos recomendados y grupos de trabajo.
- Recolectar y analizar comentarios realizados en la plataforma educativa.
- Probar el sistema, depurar códigos y protocolos.

### **RESTRICCIÓN**

- La única restricción que se presenta es que los comentarios deben ser realizados en inglés, para su posterior análisis semántico.

# **CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA O ESTADO DEL ARTE**

Este capítulo tiene como finalidad describir las principales características de la Web Semántica, entornos e-learning y las herramientas básicas del Procesamiento del Lenguaje Natural.

En primer lugar, se realiza una introducción, la arquitectura, ventajas y los principales estándares de la Web Semántica. A continuación, se estudia los componentes de sistemas e-learning. Posteriormente se detalla el funcionamiento y recursos del Procesamiento del Lenguaje Natural.

## **1.1. LA WEB SEMÁNTICA**

En la actualidad Internet está saturado de información (páginas electrónicas, blogs, documentos, videos, imágenes), la cual proviene de distintas fuentes y se encuentra en muchos formatos, dificultando su interpretación y asimilación. La Web Semántica permite a las aplicaciones (exploradores, redes sociales, blogs): procesar esta gran cantidad de información y etiquetarla en base a su significado, todo esto de forma automática para así inferir conclusiones lógicas [1].

La Web Semántica agrega significados, categorías y referencias descriptivas a los datos que contiene Internet. Este método de inferir conocimiento se aproxima la forma de pensar de los seres humanos. Por lo tanto, se mejora la búsqueda y el procesamiento de los datos con lenguajes desarrollados por el W3C (World Wide Web

Consortium), esto beneficia la interrelación entre los seres humanos y las computadoras [6].

Los nuevos lenguajes y estándares en los que se basa la Web Semántica son: RDF (Resource Description Framework), es el lenguaje básico y general que especifica los distintos recursos y sus relaciones. RDFS (RDF Schema), este estándar especifica más las características que el lenguaje RDF. OWL (Web Ontology language) dispone mayor capacidad para expresar significado y semántica que los lenguajes RDF y RDFS [2]. Existen más estándares los cuales se basan en los anteriores y se van adaptando a áreas específicas del conocimiento. Todos estos son compatibles con todas las tecnologías actuales y brindan una plataforma abierta para el desarrollo de los protocolos y tecnologías actuales [8].

Un ejemplo de la aplicación de la Web Semántica son los exploradores, navegadores, tiendas online, redes sociales, plataformas de comercio entre otros; los cuales se adaptan cada vez más a nuestros gustos. Esto se debe a que nuestra actividad puede ser registrada para conocer nuestras preferencias, los lugares desde donde se realizan las búsquedas, información visitada previamente, información que otros usuarios como nosotros han visitado, grupos sociales con los mismos intereses y anuncios que nos podrían interesar [8].

### **1.1.1. WORLD WIDE WEB CONSORTIUM**

El organismo internacional encargado de definir y estandarizar las tecnologías usadas en la Web es el Consorcio Word Wide Web conocido con la abreviación W3C. Este consorcio es un conjunto de organizaciones internacionales distribuidas en todo el mundo y pertenecen a empresas de Hardware y Software, centros de investigación, universidades entre otras [8].

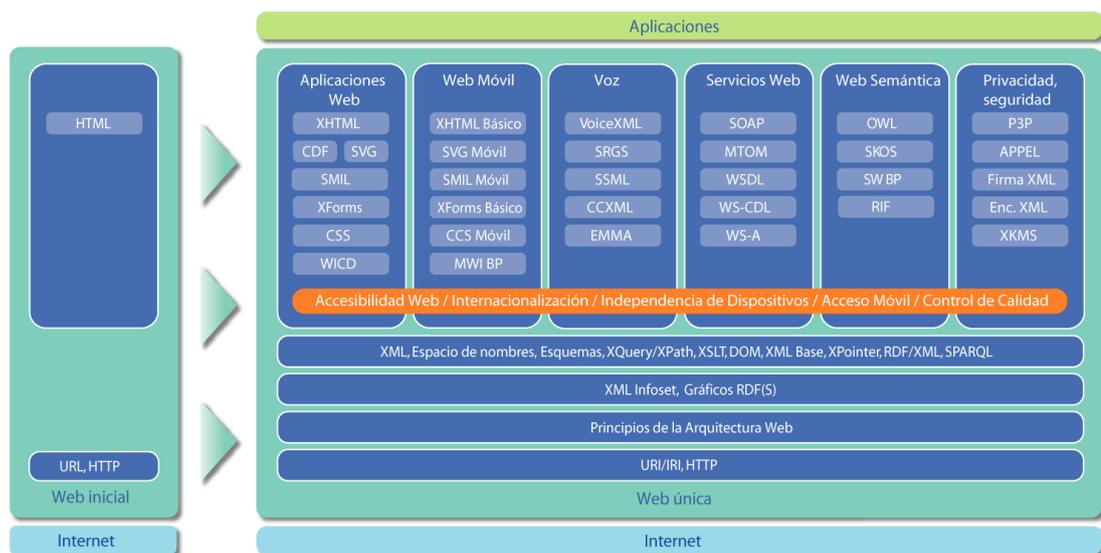
Las metas que busca el W3C son las siguientes. Primero, asegurar el acceso a Internet para todo el mundo desde cualquier dispositivo, destruyendo las barreras tecnológicas, limitaciones físicas y culturales. Segundo, el Internet ha sumado más dispositivos, como sensores, cámaras, televisiones, los cuales generan datos e información, una forma de llamar este acceso de los dispositivos es el Internet de las cosas. Tercero, la Web del conocimiento, busca que cada usuario pueda hacer un buen uso de los recursos. Finalmente, la seguridad que busca consolidar el desarrollo

respetando las normas internacionales, legales, comerciales y demás aspectos relacionados con tecnologías emergentes [8, 9].

Para garantizar el acceso universal a la red, el W3C desarrolla estándares, los cuales integran idiomas, navegadores, sistemas operativos y plataformas. También, impulsa la mayoría de recomendaciones para los lenguajes y tecnologías de uso frecuente en la creación de páginas electrónicas y aplicaciones [8, 9].

### 1.1.2. ARQUITECTURA DE LA WEB

La Web actual se constituye por datos, los cuales básicamente son páginas HTML (Hyper Text Markup Language) que observamos al abrir nuestro navegador. Los sitios digitales están formados por distintos contenidos y a la vez tiene enlaces que nos llevan a otras páginas, las cuales tienen un contenido distinto, pero mantiene la arquitectura de la Web inicial [2, 10].



**Figura 1.1.** Arquitectura de la Web inicial y la Web única.  
**Fuente:** [https://www.w3c.es/Consortio/img/pila\\_tecnologica.png](https://www.w3c.es/Consortio/img/pila_tecnologica.png)

La Figura 1.1, ilustra dos etapas marcadas: la Web inicial, y la Web única propuesta por el W3C. La Web inicial empezó con protocolos básicos URL (Uniform Resource Locator) y HTTP (Hypertext Transfer Protocol), así mismo el contenido de los datos se muestran mediante HTML (Hyper Text Markup Language). Sin embargo, en la actualidad la arquitectura se basa en áreas de interés y tecnologías en continuo desarrollo, representadas a través de capas. En el modelo existe un vínculo en base a las tecnologías anteriores y todas representan una Web única [10, 11].

Las cuatro primeras capas de la Web única representan los diferentes estándares, tecnologías y avances para el transporte, formato y dirección de los sitios electrónicos que se han implementado en la actualidad. La última capa se ha dividido en seis áreas que corresponden a los principales desarrollos por el W3C: Aplicaciones Web, Web Móvil, Voz, Servicios Web, Web Semántica y Privacidad [10, 11].

El área de la Web Semántica contiene estándares que se trataran con más detalle en las siguientes subsecciones. Todas estas recomendaciones se basan en RDF (Resource Description Framework). La banda naranja, representa las características que unen las seis áreas, en ella encontramos: Accesibilidad Web, Internacionalización, Independencia de Dispositivo, Acceso Móvil y Control de Calidad, cualidades que presentan las distintas tecnologías [11].

### **1.1.3. DESAFÍOS DE LA WEB ACTUAL**

Los desafíos que enfrenta la Web son: Heterogénea, cada sitio publica y presenta sus datos de diversas formas, asimismo no intercambian datos entre las organizaciones que los difunde. Es masiva, la cantidad de datos e información que se encuentra disponible es inmensa. Cambia muy rápido, a diario enormes volúmenes de datos se intercambian en la red. La información solo la interpretan los humanos, una aplicación no puede identificar los datos que almacena un servidor. Todo esto con el diseño actual es difícil de superar. Por consiguiente, la heterogeneidad de datos y su interpretación; son los principales desafíos que la Web Semántica intenta resolver. El propósito es que las máquinas interpreten la información de manera mucho más eficiente y automática [11, 12].

### **1.1.4. REQUISITOS PARA UNA WEB DE DATOS EFECTIVA**

Primero, se necesita un lenguaje que especifique los recursos y las relaciones entre los distintos contenidos publicados por los diferentes sitios Web. Para especificar recursos y sus relaciones, la computadora debe entender este lenguaje [11, 12].

Segundo, es necesario consultar los datos mediante un estándar que nos permita especificar lo que estamos buscando, además que de forma automática se extraiga la información necesaria. Para ello requerimos un lenguaje que sea procesable por una computadora, con esto una aplicación entenderá la consulta y será capaz de sacar

conclusiones a partir de los datos de forma automática frente a la consulta que se ejecute [11, 12].

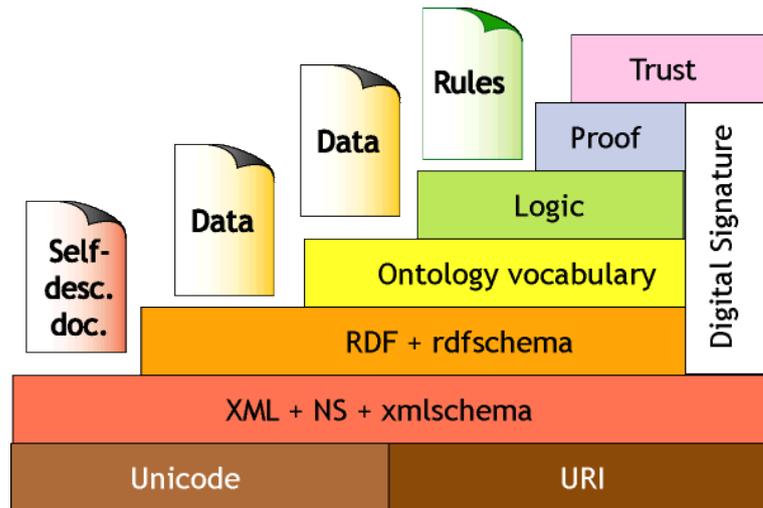
### **1.1.5. DEFINICIÓN DE WEB SEMÁNTICA**

La definición formal la dio Tim Berners-Lee en 2001 como: *"La Web Semántica es una extensión de la Web actual donde la información viene dotada de significado bien definido, y permite a las computadoras y a la gente trabajar en cooperación"* [1]. El W3C desarrolla un grupo de recomendaciones, con la cuales se pretende que las computadoras y las personas tengan la capacidad de interpretar los datos [1, 7].

Una recomendación es una representación formal y normalizada de una tecnología sujeta a las políticas de patentes del W3C que se utilizan de manera libre por desarrolladores e investigadores. Estos estándares son compatibles con todas las tecnologías ya desarrolladas y sirven de referencia en la construcción de una Web accesible, interoperable y eficiente, para el desarrollo de aplicaciones cada vez más robustas [7, 8]. Todo el desarrollo sirve para que las máquinas se entiendan entre sí y puedan interactuar con las personas. Las tecnologías de la Web Semántica permiten crear, almacenar, acceder y consultar datos almacenados, mediante la construcción de vocabularios y la posibilidad de crear reglas para manejar los datos y crear inferencias de forma automática [8, 9, 12].

## **1.2. ARQUITECTURA DE LA WEB SEMÁNTICA**

La Web Semántica se ha estructurado en distintas capas, que se usan para representar la información; por lo tanto, los usuarios podrán localizar y acceder a los datos que se requieran mediante una aplicación, de forma automática y exacta. Entonces, de manera general los recursos se identifican a través de URI (Uniform Resource Identifier) y el lenguaje RDF (Resource Description Framework) representa las relaciones de los recursos con expresiones denominadas triples o sujeto-predicado-objeto. Los lenguajes RDFS (RDF Schema) y OWL (Web Ontology Language) extienden la funcionalidad de RDF para clasificar la información [11, 13, 15].



**Figura 1.2.** Pila de la Web Semántica.

**Fuente:** <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide11-0.html>

La Figura 1.2 muestra la pirámide de estándares, en los cuales ha desarrollado la Web Semántica. En la parte inferior se ubican los componentes básicos, en el nivel más alto se encuentra el nivel de confianza. RDF, es el lenguaje básico y general para especificar los recursos de la Web y sus relaciones entre varios recursos, RDFS especifica mucho más el vocabulario RDF, SPARQL (Protocol and RDF Query Language) lenguaje usado en la consulta de datos y en la extracción de información de la Web. Finalmente tenemos OWL, este lenguaje identifica ontologías. Este conjunto de estándares especifica los recursos que dispone la Web, las relaciones entre datos y define lenguajes que permiten extraer la información de manera automática [11, 13, 14]. Todos los elementos de la pirámide en la Figura 1.2 se detallan a continuación:

En la primera capa observamos Unicode, que es un estándar universal para la codificación de caracteres y puede representar texto en diferentes idiomas. También en esta capa se encuentra el URI que es un identificador de un recurso en la Web. Un recurso básico es una página electrónica, pero se considera todos los elementos que constituyen la Web incluido los usuarios. El URI permite localizar recursos que pueden ser accedidos vía Internet [11, 13].

La segunda capa la conforman, XML (eXtensible Markup Language), NL (namespaces) y XML Schema. Estas tecnologías hacen posible que los agentes puedan comunicarse entre ellos. XML permite el intercambio de documentos mediante un formato general, es la base sintáctica para crear contenido Web. NL sirve para asociar las propiedades y atributos usados en XML siendo identificados con referencias URI.

XML Schema define elementos para elaborar documentos XML, además de su organización y sus atributos [11, 13, 15].

La tercera capa corresponde a RDF y RDF Schema que son los estándares básicos de la Web Semántica. El requisito principal de este lenguaje es que necesita ser entendido por una aplicación. Para ello surge el lenguaje RDF, que es el estándar del W3C y sirve para presentar recursos en la Web a través de triples. Los triples se constituyen por sujeto, predicado, y objeto. El sujeto señala un recurso, el predicado se representa y establece la relación del recurso o los atributos que tiene el recurso en cuestión. Finalmente, el objeto identifica la relación del recurso o su valor indicado por un literal. Estos tres componentes se identifican por URIs. Por su parte RDF Schema añade un vocabulario bien definido y se basa en RDF. Este estándar describe los datos y crea la información semántica [11, 13, 14].

La cuarta capa corresponde al lenguaje de las Ontologías u OWL. Esta recomendación es un vocabulario RDF, que permite dar un significado bien definido y estandarizado a las palabras o datos para catalogar y clasificar la información. A diferencia de los lenguajes anteriores, OWL brinda elementos avanzados para describir diversos dominios; además de la construcción de reglas avanzadas para describirlos. OWL incluye un sistema de razonamiento y podemos agregar nuevas clases y propiedades para describir los recursos [11, 13, 14].

La quinta capa es la capa lógica en la cual actúa cuando se tiene toda la información de recursos, propiedades y ontologías. Se establecen conjuntos de axiomas y reglas de inferencia para la deducción de nuevo conocimiento, a partir de información localizada y especificada en las capas XML, RDF y OWL de forma autónoma. De esta manera una aplicación sabrá qué es lo que se ha almacenado en la Web [11, 13, 15].

En la sexta capa se ubica la prueba, aquí se realiza la deducción a la que llega un sistema luego de aplicarle la lógica. Los datos se recolectan y se verifican que las inferencias sean las correctas. Esta información se almacena e intercambia en las distintas pruebas [11, 13, 15].

La séptima capa se trata de la confianza, para determinar si la información es fiable o no. Las firmas digitales son de gran importancia debido a que se basan en

métodos matemáticos y criptográficos que permiten dar fe de la información inferida. La mayoría de documentos tienen firmas digitales, las cuales necesitan ser identificadas para determinar la confianza que necesitan los datos [11, 13, 14].

### 1.3. ESTÁNDARES DE LA WEB SEMÁNTICA

El desarrollo hacia la Web Semántica se ha basado en las tecnologías que ha presentado el W3C. Cada estándar que se ha desarrollado especifica los datos de manera precisa, permitiendo que las búsquedas de información sean eficientes y exactas. Los estándares básicos que permiten representar los recursos para identificarlos y extraerlos se estudian a continuación.

#### 1.3.1. EL LENGUAJE RDF

RDF se basa en XML y se utiliza para especificar los recursos de la Web. Este lenguaje se describe en las recomendaciones proporcionadas por el W3C, los cuales describen aspectos sintácticos y se definen los vocabularios que emplea esta tecnología [11, 13, 16].

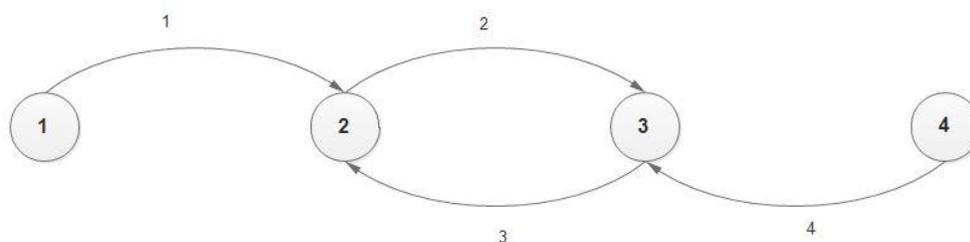


**Figura 1.3.** Características del estándar RDF.

**Fuente:** El Autor.

La Web necesita construir un lenguaje descrito en la Figura 1.3, que permita especificar los recursos que tenemos en la Web y las relaciones que abarcan entre los mismos. El requisito principal de este lenguaje es que necesita ser entendido por una aplicación. El W3C desarrolló el estándar RDF y sirve para representar los recursos (páginas electrónicas, documentos, enlaces, recursos multimedia) [11, 13, 16, 17].

Para representar este estándar, en primer lugar, RDF se basa en la utilización de grafos dirigidos, especificados y etiquetados. Segundo, una aplicación podría procesar de forma automática este lenguaje. En tercer lugar, esta tecnología se construye sobre los URIs y los literales [10, 13, 16].



**Figura 1.4.** Representación de un grafo.  
Fuente: El Autor.

La Figura 1.4 representa un grafo, el cual tiene tanto nodos como arcos los cuales están bien etiquetados. En esta figura podemos ver en círculos los nodos, que representan recursos en la Web. Mientras que las flechas son los arcos que también están etiquetados. En este caso tienen un número como etiqueta y ayudan a representar las relaciones que tenemos entre los recursos [10, 13, 16, 17].

### 1.3.2. LA NOCIÓN DE URI

El URI (Identificador de Recursos Uniforme) es un identificador de un recurso en la web. Estos son muchos, por lo cual se necesita reconocer la información de manera única y precisa. Todos los elementos en la Web tienen una dirección para acceder a los mismo. Desde la aparición de la Web se han usado los URL y los URN para identificar la información en la Web [13, 17].

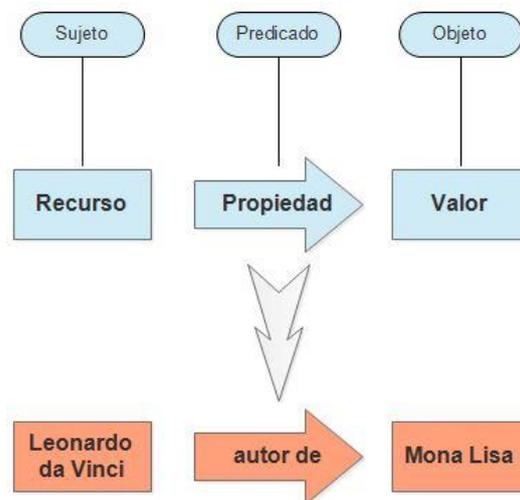
Los URI se conforman por tres componentes: Primero observamos http, siendo este el protocolo que permite el acceso a la información; segundo, tenemos la autoridad o la dirección Web, que nos permite acceder al dominio en donde se almacena cierto recurso; tercero vemos el camino del URI, que conduce al recurso e identifica el URI de nuestro recurso. Entonces esta dirección, desde el protocolo hasta cierto recurso almacenado en la Web, se conoce como URI [13, 17].

Los URI, identifican los recursos en nuestra web, y son una generalización de los URL. Los URL (Localizador de recursos uniforme), es un identificador de una página web. Es una dirección que nos permite acceder a una página web [13, 17].

### 1.3.3. ESPECIFICACIÓN RDF

Llamamos especificación RDF a la relación entre recursos. Esta a su vez se da mediante triples y se representa de forma gráfica. Ahora, consideramos un triple a la especificación o valor relacionado a un recurso en la Web [13, 17].

Los triples se constituyen por sujeto, predicado, y objeto. El sujeto señala un recurso que es identificado por una URI. El predicado se representado por un URI y establece la relación del recurso o los atributos que tiene el recurso en cuestión. Finalmente, el objeto identifica la relación del recurso o su valor indicado por un literal [13]. En la Figura 1.5, se representa un triple donde se identifica el sujeto como Leonardo da Vinci; a continuación, el predicado o la relación entre sujeto y objeto se denota como: es autor de. Finalmente tenemos el valor de la relación denominado objeto, que en la Figura 1.5 se representa como Mona Lisa [13, 17]. Entonces la información de Leonardo da Vinci es autor de la Mona Lisa queda representado en un lenguaje entendible para las personas y las máquinas.



**Figura 1.5.** Triple RDF.  
Fuente: El Autor.

## 1.4. NOCIÓN DE RDF SCHEMA

RDF Schema o RDFS es un estándar extensible para la descripción de ontologías y la clasificación de datos, es decir adiciona semántica como mecanismo para describir conjuntos de recursos y sus relaciones.

### 1.4.1. MODELO DE DATOS

El modelo de datos es un lenguaje que nos permite estructurar la información almacenada con sus respectivas restricciones que podrían estar presentes. Por ejemplo, se puede indicar datos acerca de personas. Cada elemento, en este caso una persona, quien tiene un nombre y cada nombre se denota por una cadena de letras o caracteres. Entonces resulta indispensable generar y emplear un modelo de datos que especifiquen

las propiedades que poseen o cumplen dichos datos en la web para ello el W3C sugiere el vocabulario RDF Schema cuya estructura se basa en RDF [11, 18].

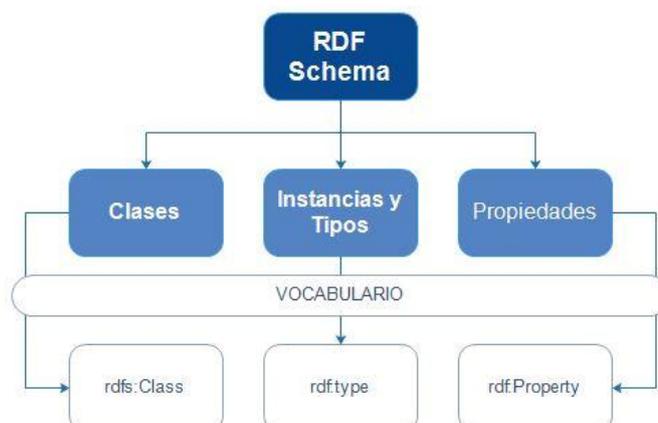
### 1.4.2. RECOMENDACIÓN RDFS

RDFS o RDF Schema usa URIs que representan prefijos o abreviaciones que mejoran y reducen la notación de propiedades; por lo tanto, RDF Schema es un vocabulario RDF. Cada palabra está bien definida usando los estándares y recomendaciones del W3C. Al desarrollar este vocabulario se consigue que cada término sea entendido por los humanos y de forma especial por las computadoras [11, 13].

El estándar RDFS tiene como objetivo proveer de elementos principales y generales para describir los datos de diversos dominios. Existen recursos comunes en la Web, entonces RDF Schema abstrae algunos de estos elementos comunes. Además, usa el vocabulario básico RDF para definir propiedades adicionales en cierto dominio [11, 13].

### 1.4.3. COMPONENTES BÁSICOS DE RDFS

Los elementos principales de RDFS para especificar los recursos en la Web se ven en la Figura 1.6. Primero, tenemos las clases y para especificar las clases se utiliza el prefijo `rdfs:class`. Segundo, existen las instancias y tipos cuya nomenclatura es representada mediante `rdf:type`. Finalmente este lenguaje contiene las propiedades y se denota con `rdf:Property`. Entonces, los componentes básicos de RDFS son: las clases, las propiedades, las instancias y tipos [13, 19, 20, 21].



**Figura 1.6.** Componentes básicos de RDFS.

**Fuente:** El Autor.

#### **1.4.4. LA NOCIÓN DE CLASE**

Una clase es una agrupación de elementos que mantienen propiedades comunes y poseen una representación real. Un ejemplo práctico es la clase “Perros”. Todos los perros son mamíferos, esta una característica común a los elementos de esta clase; luego cada elemento de la esta clase tiene una raza, un tamaño, un color entre otras características; existen propiedades que comparten ciertos elementos de esta clase, pero no todas las instancias de la clase en RDFS [19, 20, 21].

Para establecer una clase en RDFS se emplea el vocabulario RDF, primero utilizamos el prefijo que especifica la URL del recurso, esta es la base para construir los triples que van a indicar cuáles son las clases. En este caso tenemos un triple que nos indica las clases. Para ello utilizamos `rdf:type` para indicar que el recurso es de tipo `rdfs:Class`, vamos a tener la clase la cual reúne un conjunto de elementos que tienen características comunes [19, 20, 21].

#### **1.4.5. NOCIÓN DE INSTANCIA DE UNA CLASE**

Un elemento de una clase se denota como una instancia de la clase. Por ejemplo, si tenemos la clase de “Personas”; todos los seres humanos pertenecemos a este conjunto, esto es cada persona es un elemento de esta clase [19, 20, 21].

Las instancias emplean el vocabulario RDFS. Por ejemplo, la provincia de Azuay es una instancia de la clase “Provincias de Ecuador”. Para este ejemplo empleamos los prefijos `rdf:type` y `rdfs:Class`, para indicar la instancia y la clase respectivamente [19, 20, 21].

#### **1.4.6. PROPIEDADES EN RDFS**

Una propiedad indica el valor de un atributo, de un recurso o la relación que existe entre dos o más recursos. Para definir propiedades se emplea la palabra `Property` dentro del prefijo `rdf:Property`. Es decir, aquí asignamos los valores o las características de cada instancia [19, 20, 21].

### 1.4.7. RESTRICCIONES SOBRE LAS PROPIEDADES

Una restricción en el sujeto o el objeto en un triple RDF se emplea cuando la propiedad relaciona estos dos elementos con valores o relaciones que solo pueden cumplir un conjunto o una clase específica de sujeto u objeto [19, 20, 21].

RDFS establece 2 restricciones a las propiedades, `rdfs:domain` y `rdfs:range`. La primera es una restricción de dominio que delimita el tipo de sujeto de una relación. El segundo prefijo, define restricciones de rango que actúan en los objetos de una relación. La figura 1.7 indica que el sujeto del triple se restringe al conjunto de “Los seres humanos” de la misma manera que el objeto; estos dos recursos quedan restringidos mediante el predicado que indica que una persona conversa con una instancia del mismo dominio. Esta relación puede ocurrir solo entre los seres humanos o las personas ya que ni los animales ni las cosas pueden conversar, entonces es necesario restringir el dominio y el rango del triple [19, 20, 21].



**Figura 1.7.** Restricción de dominio y rango en RDFS.

**Fuente:** El Autor.

### 1.4.8. TÉCNICAS DE RAZONAMIENTO EN RDF SCHEMA.

Para generar y especificar el conocimiento con la recomendación RDFS, se puede especificar subconjuntos o jerarquías de clases, algo que la recomendación RDF no alcanza a determinar y tampoco puede crear reglas de inferencia para la generar conocimiento [19, 20, 21].

Por lo tanto, el propósito que busca RDFS es que a partir de los triples existentes que indican clases, subclases, tipos, instancias; una aplicación pueda razonar de manera automática y generar nuevos triples. Es decir, disponemos de un conjunto de triples que utiliza el significado de este vocabulario e inferimos nuevos, los mismos son considerados como parte del grafo RDF [19, 20, 21].

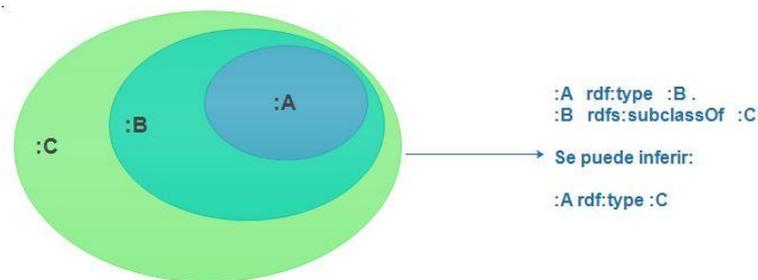
Este estándar dispone de manera principal: `rdfs:class`, `rdfs:type`, `rdf:property`, `rdfs:domain` y `rdf:range`, entre otros prefijos usados en diferentes áreas del conocimiento.[ 19, 20] Entonces, la forma de deducir información en RDFS se traduce en reglas específicas que generan datos, los cuales siempre crean nuevo conocimiento a partir de triples existentes de forma automática [21].

#### 1.4.9. REGLAS DE INFERENCIA EN RDF SCHEMA.

Las reglas de inferencia crean parámetros para inferir información sobre los recursos y establecer jerarquías que van actuar sobre las clases y propiedades para modelar nuestro dominio de los datos [19, 20, 21].

##### 1.4.9.1. LAS SUBCLASES

RDFS permite usar las subclases que sirven para representar un subconjunto de otro conjunto mayor, es decir es posible declarar una clase como una subclase de una clase mayor que contiene a la misma. Para esto utilizamos el URI `rdfs:subclassOf`, cada componente de este vocabulario está especificado por un URI y el prefijo `rdfs` [20, 21, 22].



**Figura 1.8.** Jerarquías de clases en RDFS.

**Fuente:** El Autor.

Las subclases en RDFS se pueden describir de la siguiente manera. El triple A es de tipo B y B es subclase de C entonces, se infiere que el triple A es de tipo C. Es importante señalar que las inferencias se producen de forma automática, que A es de tipo C. Gráficamente esto se puede mostrar de la Figura 1.8 [19, 20, 21].

##### 1.4.9.2. Las subpropiedades

RDFS ha creado los componentes para establecer que una propiedad puede ser una subpropiedad de otra. Para esto, las reglas de la Web Semántica ayudan a satisfacer

esta condición dentro de un dominio. La nomenclatura usa el prefijo `rdfs:subPropertyOf`, para determinar que existe la subpropiedad de un conjunto [19, 20, 21].

La relación de propiedades de triples a partir de otros triples que se establecen con anterioridad. Se traduce en una forma de inferencia: Tenemos los triples A, P, B, P es una subpropiedad de Q. Es decir, el sujeto A se encuentra relacionado con el objeto B a través de P, también conocemos que P es una subpropiedad de la propiedad Q, con esto se concluye que el sujeto A está relacionado con el objeto B mediante Q [19, 20, 21].

RDFS ha desarrollado distintas reglas de inferencia que generan nuevos triples a partir de triples ya creados. Las reglas se utilizan para producir nuevas relaciones de clase usando jerarquías de clases. Para inferir relaciones de propiedad se utiliza la jerarquía de propiedades. Además de crear información sobre las clases generales que contienen los objetos o instancias utilizando el dominio o el rango de los triples existentes [19, 20, 21].

A continuación, se analizan las principales reglas que se han desarrollado para el razonamiento RDF.

#### **1.4.9.3. Regla de inferencia en jerarquías de clases**

Esta propiedad de inferencia funciona cuando se tiene los triples: M subclase de N, N subclase de O, por consiguiente se infiere el nuevo triple que establece que M es subclase de O. Una aplicación será capaz de deducir este triple de manera automática [19, 20, 21].

#### **1.4.9.4. Regla de inferencia en jerarquías de propiedades**

Para establecer esta regla partimos de los triples: P es subpropiedad de Q, Q es subpropiedad de R, por lo tanto, se infiere que el triple P es subpropiedad de R. Como ya se ha venido mencionando esto ocurrirá de forma automática [19, 20, 21].

#### **1.4.9.5. Regla de inferencia en el dominio de una propiedad**

Esta regla establece que el sujeto A se relaciona con el objeto B mediante la propiedad P, asimismo, sabemos que el dominio de P es la clase C, cabe recalcar que

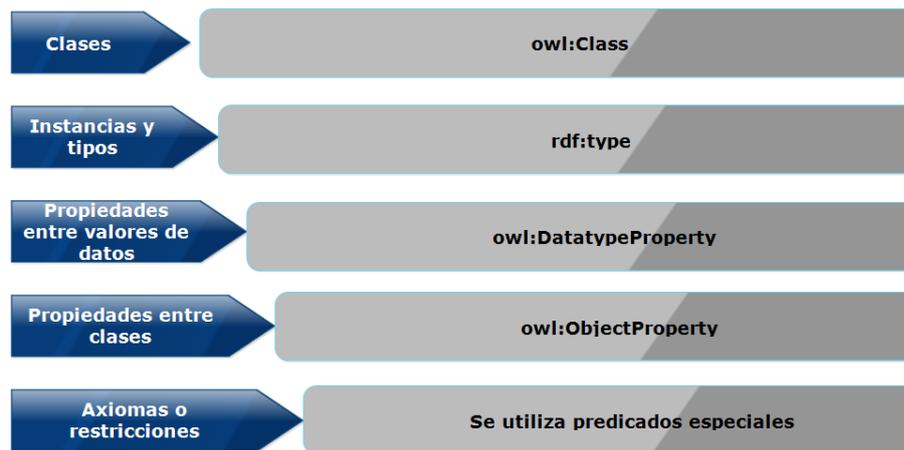
los sujetos que tenemos en esta relación deben ser de tipo C, por lo tanto, se infiere que A es de tipo C [19, 20, 21].

Esta regla de inferencia se compara al caso de la restricción de dominio del sujeto, y la regla de inferencia para el rango de una propiedad, resulta similar. Para aclarar esta regla sabemos que el sujeto A se relaciona con el objeto B mediante la propiedad P, y conocemos que el rango de la propiedad es B, por lo tanto, si los objetos que se encuentran relacionados a través de la propiedad P, deben ser de tipo D. Entonces se infiere que B es de tipo D [19, 20, 21].

### **1.5. WEB ONTOLOGY LANGUAGE (OWL)**

OWL es una recomendación del W3C que brinda una gran expresividad para establecer un modelo de datos, esta recomendación nos permite dar un significado bien definido y estandarizado a las palabras o datos [22, 23]. Los elementos principales que posee OWL son: las clases, las propiedades sobre clases y sobre valores de los datos, las instancias o individuos, y las restricciones o axiomas. Además, OWL a diferencia de los estándares anteriores brinda elementos avanzados para describir diversos dominios, como construir reglas avanzadas, incluir un sistema de razonamiento que detalla la información siendo posible declarar propiedades que satisfacen condiciones adicionales, como transitividad o simetría [24, 25].

OWL extiende a RDFS, este último tan solo dispone de clases, propiedades, tipos o instancias. Entonces este lenguaje dispone un nuevo término los individuos, que es un sinónimo de instancias, la característica importante de OWL es que permite definir axiomas y restricciones [23, 24]. La forma de representar estos componentes en OWL se representan en la Figura 1.9 y se definen de la siguiente forma: para definir las clases usamos el prefijo owl:Class, para establecer instancias y tipos, usamos el prefijo rdf:type, de la misma manera que en RDFS. Al momento de definir propiedades o valores de los datos, se emplea owl:DatatypeProperty. Para establecer propiedades entre clases, usamos owl:ObjectProperty, finalmente para definir axiomas o restricciones en las clases y propiedades, se usan predicados especiales los cuales se especiarán más adelante [25, 25, 26].



**Figura 1.9.** Jerarquías en OWL.

**Fuente:** El Autor.

Un axioma en OWL que no se puede utilizar en RDFS, permite especificar una clase que surge como la unión de otras dos clases. Entonces, en OWL usamos el componente owl:unionOf para establecer que una clase es la unión de otras dos clases [26].

### 1.5.1. PROPIEDADES EN OWL

Las propiedades en OWL describen relaciones entre clases o individuos. Las propiedades de objetos usan el prefijo owl:ObjectProperty y establecen relaciones entre determinadas clases. Por otro lado, las relaciones de tipo propiedades de datos, usan owl:DatatypeProperty y expresan las relaciones entre individuos y literales [26].

### 1.5.2. RESTRICCIONES EN OWL

Las restricciones y propiedades en el lenguaje OWL dan la capacidad para crear de una manera más precisa y específica, un modelo de datos [23, 24]. Al igual que RDFS, este lenguaje permite restringir las propiedades de los dominios y los rangos. Para aplicar restricciones en el sujeto del triple, se usa rdfs:domain y para restringir el objeto usamos rdfs:range [21, 26].

La propiedad transitiva se interpreta, si A se relaciona con B, a la vez B se relaciona con C, mediante la misma relación esta relación es transitiva. Entonces tenemos que A se relaciona con C. Esta propiedad se representa a través de rdf:type owl:TransitiveProperty. Si A se relaciona con B, y esta relación es simétrica, por lo tanto, B se relaciona con A. Esta propiedad se representa con rdf:type owl:SymmetricProperty [26].

OWL especifica restricciones más superiores sobre las propiedades. Por ejemplo, se puede establecer que una propiedad es inversa de otra o que dos propiedades son disjuntas [25]. En siendo OWL más específica que RDFS se puede establecer que dos clases son disjuntas usando el predicado owl:disjointwith. Con OWL podemos establecer que una instancia es similar a otra. Esta propiedad indica que dos instancias son la misma, para ello se utiliza el prefijo owl:sameAs. Esta restricción establece que una instancia en un grafo es la misma que otra instancia en un grafo remoto, o en cualquier base de datos [26].

## **1.6. E-LEARNING**

E-learning es un entorno de aprendizaje virtual que intenta combinar diferentes recursos, actividades educativas, tecnologías y estándares actuales para mejorar el acceso a la información y eliminar posibles limitaciones que presenta la educación tradicional [6, 26, 27].

Los sistemas online de educación se dividen en grupos: por la distribución de contenido, trabajos en grupo o colaboración, sistemas de gestión de contenido, sistemas de gestión de conocimiento y sistemas de gestión de contenidos para el conocimiento o aprendizaje [6, 26, 27].

Las características que necesitan disponer las plataformas son: Permitir el acceso mediante cualquier navegador mediante contraseña y diferentes estándares de seguridad. Los sistemas deben tener una de gestión que se adapte a las características y necesidades de los usuarios. Mantener los roles de estudiante, docente, tutor o administrador. Todos los accesos deben tener privilegios de acuerdo con el tipo de usuario. La plataforma debe ser interactiva y mantener una comunicación entre los diferentes usuarios [26, 27].

### **1.6.1. COMPONENTES DE E-LEARNING**

Los elementos que conforman los sistemas de aprendizaje online son principalmente: los recursos, el tutor y los estudiantes, aulas virtuales y actividades. Las características de e-learning son que disponen y hacen uso de los estándares actuales para generar contenido interactivo debido a que todo eso se realiza en línea asegurando el acceso a la información, gestionando los contenidos, planificando el estudio, administrando datos y asegurando la privacidad de estudiantes [6, 26, 27].

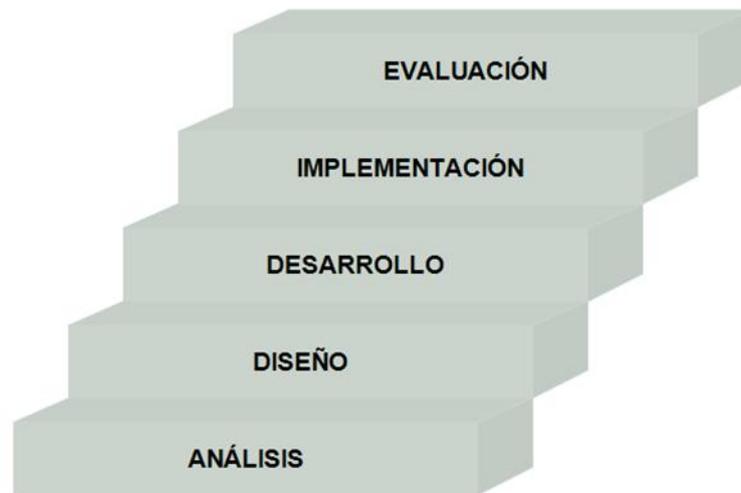
Los recursos que forman el conjunto de componentes son recursos simples tales como libros, tareas o lecciones [6, 7]. Además de esto estas plataformas disponen de simulaciones electrónicas, grupos de trabajo, recomendaciones de contenidos y demás material multimedia [26, 27].

Internet dispone de millones de datos accesibles para todos, razón por la cual una persona normal no podría tener la capacidad de discernimiento al momento de elegir el material adecuado para cierto tema o cátedra en especial. Los tutores son los encargados de recomendar el material idóneo a otro usuario en determinado tema [7, 26, 27]. En la actualidad el papel de tutor lo puede desempeñar una computadora que, a través de minería de datos, inteligencia artificial, filtros cooperativos; puede realizar recomendaciones de diferentes recursos, grupos de trabajo y todo el material multimedia [26, 27].

Todos estos sistemas se van adaptando a las preferencias, afinidades y métodos de aprendizaje de los usuarios. Las plataformas de e-learning se han vuelto eficientes, dinámicas, de código abierto, por lo tanto, los institutos educativos de manera especial las universidades ya tienen implementado sus plataformas, implementando una gran herramienta para la educación [26, 27].

Las aulas virtuales son espacios que se asemejan a la capacitación tradicional que está dirigido por un docente, pero facilita el intercambio de conocimiento, acceso a contenidos y materiales educativos y todas las herramientas que se puedan disponer [7, 29, 30].

El diseño pedagógico para la planificación de actividades y generación de contenidos se basa en muchos parámetros en la Figura 1.10 se observa el modelo ADDIE (Análisis Diseño Desarrollo Implementación y Evaluación), el cual sigue los pasos y es un sistema de diseño instruccional [7, 29, 30].



**Figura 1.10.** Modelo ADDIE para e-learning.  
**Fuente:** El Autor.

Este modelo empieza por el análisis, en esta etapa se estudia y determina las necesidades del usuario cuyo propósito es identificar los conocimientos, habilidades contextos de aprendizaje, entre otros. La etapa del diseño utiliza los datos de la etapa anterior y planifica estrategias pedagógicas y selecciona materiales para conseguir los objetivos generales y específicos de la plataforma. La etapa de desarrollo de contenido interactivo a su vez se subdivide en: desarrollo de contenidos, desarrollo de guion gráfico, desarrollo de programas pedagógicos; todas etapas utilizan la información generada y genera el conocimiento a ser utilizado por el alumno utilizando todos los recursos disponibles como: libros, multimedia, entre otros. La etapa de implementación es la aplicación del método educativo desarrollado en el nivel anterior, todos los recursos se etiquetan y se ponen a disposición del usuario. Finalmente, la etapa de evaluación permite determinar la reacción del estudiante, el logro de resultados y el impacto de la metodología [7, 29, 30].

### **1.7. OPPIA**

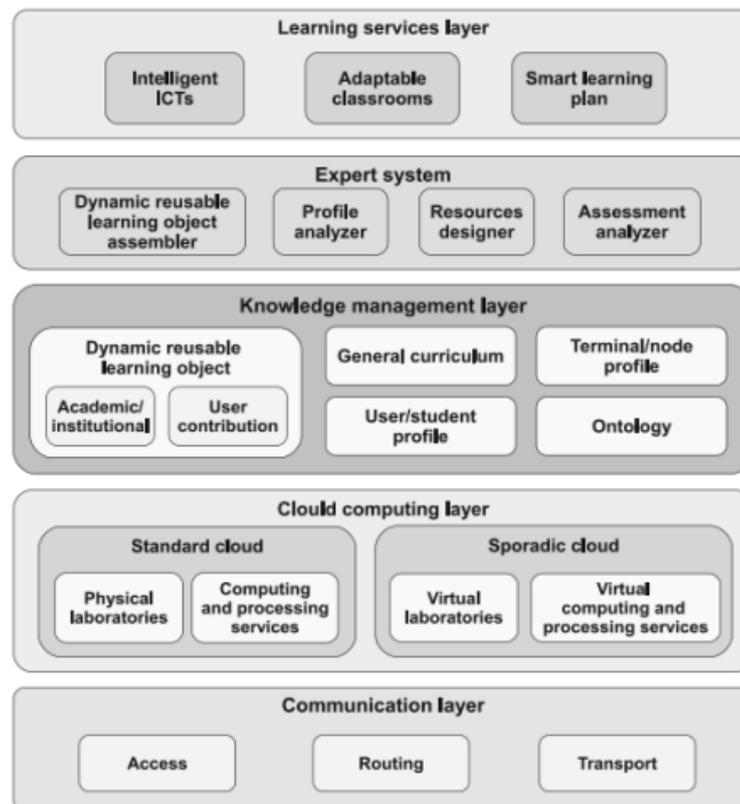
OPPIA es una plataforma de formación educativa online, que utiliza tecnologías y estándares que se han ido desarrollando para colaborar en el proceso de aprendizaje logrando el intercambio de recursos y conocimiento, con una interacción entre alumnos y profesores; superando barreras de tiempo y espacio, conectados de manera sencilla a través del Internet [3, 30, 31].

Los beneficios otorgados por OPPIA son: Estas aplicaciones emplean sistemas de recomendación basados en el conocimiento, emplea servicios adaptados de

aprendizaje, permite a los usuarios y dispositivos compartir recursos, utiliza minería de datos y estándares de la Web Semántica para gestionar de forma dinámica los contenidos y recursos educativos [3, 30, 31].

### 1.7.1. ARQUITECTURA DE OPPIA

Al igual que todas las tecnologías actuales OPPIA presenta cinco capas en su arquitectura [3, 30], véase la Figura 1.11, y se resumen a continuación:



**Figura 1.11.** *Arquitectura de la plataforma OPPIA.*

**Fuente:** Tomado de [30].

La Figura 1.11 presenta las siguientes capas: La primera capa corresponde a los protocolos de acceso, enrutamiento y transporte; los cuales aseguran la conexión y calidad de servicio a los usuarios. La segunda capa permite el intercambio de recursos e información dentro de la plataforma y crea redes de aprendizaje esporádicas entre los usuarios del sistema; además con Cloud computing se suprime capacidades de almacenamiento y procesamiento. La tercera capa corresponde al manejo del conocimiento, que usa sistemas de recomendación, inteligencia artificial, minería de datos; para formar la red del conocimiento, esta capa realiza un filtrado para determinar preferencias y gustos de los usuarios, así genera planes de estudio, actividades basadas

en ontologías. La cuarta capa selecciona diferentes contenidos (libros, enlaces, videos) para adaptarse al estilo de aprendizaje de cada usuario. Finalmente, la capa de servicios permite el acceso a los recursos de OPPIA mediante la Web, aplicaciones móviles con solo una conexión a Internet [3, 30].

# **CAPÍTULO 2: MARCO METODOLÓGICO**

## **2.1. ÁMBITO DEL PROBLEMA**

El análisis de comentarios se realizó mediante un algoritmo usando el Procesamiento de Lenguaje Natural.

## **2.2. DEFINICIÓN DEL PROBLEMA**

El caso de estudio elegido se ha dividido en dos problemas: Primero es obtener los datos del comentario desde la plataforma del sistema de recomendación. Segundo se trata del objetivo de este proyecto el cual es el análisis de comentarios una vez haya sido obtenido.

## **2.3. SISTEMA DE RECOMENDACIÓN**

El entorno de aprendizaje online desarrollado por OPPIA recomienda a los usuarios contenidos (libros, enlaces, videos) y la formación de grupos de trabajo, que se basan en el filtro colaborativo, la similitud entre los usuarios y gustos de los estudiantes. A continuación, se resumen los aspectos importantes de la recomendación de contenidos y formación de grupos de trabajo [3].

### **2.3.1. Filtro colaborativo**

Este método usa la historia de preferencias de los usuarios almacenadas en la base de datos realizados en el pasado y sugiere que en el futuro el usuario tendrá gustos

similares. Entonces el sistema da un parámetro de preferencias y realiza la recomendación, basado en el algoritmo del vecino más cercano cuyo objetivo es el de buscar a otros estudiantes con gustos similares a quien va a recibir la recomendación [3]. La aplicación obtiene la similitud luego realiza calificaciones ponderadas sobre un ítem q aún no se recomienda, dicha calificación debe superar un umbral para que el usuario reciba la recomendación. Un hecho importante es que el sistema considera que las preferencias del usuario se mantendrán estables [3, 32, 33, 36].

### 2.3.2. Similitud entre usuarios

Para determinar este parámetro la plataforma utiliza la correlación de Pearson [34], determinada por la ecuación 1:

$$\text{sim}(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} * \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}} \quad (3.1)$$

Donde:

- $U = \{u_1, \dots, u_n\}$ , indica el número de usuarios.
- $P = \{p_1, \dots, p_m\}$ , indica el número de ítems.
- $r_{i,j}$ , matriz de comentarios, con  $i \in 1 \dots, n, j \in 1 \dots, m$ .
- $\text{sim}(a, b)$ , es la similitud entre usuarios a y b.
- $\bar{r}_a$ , es la calificación media del usuario a.
- $\bar{r}_b$ , es la calificación media del usuario b.
- $\bar{r}_{a,p}$ , es la calificación del usuario a sobre el ítem p.
- $\bar{r}_{b,p}$ , es la calificación del usuario b sobre el ítem p.

Al momento de aplicar la ecuación 1 brinda los siguientes posibles valores: (0,+1) para indicar una correlación positiva, por lo tanto existe similitud entre los usuarios. (-1,0) indica una correlación negativa que indica que no hay similitud entre los usuarios [3, 34, 35, 36].

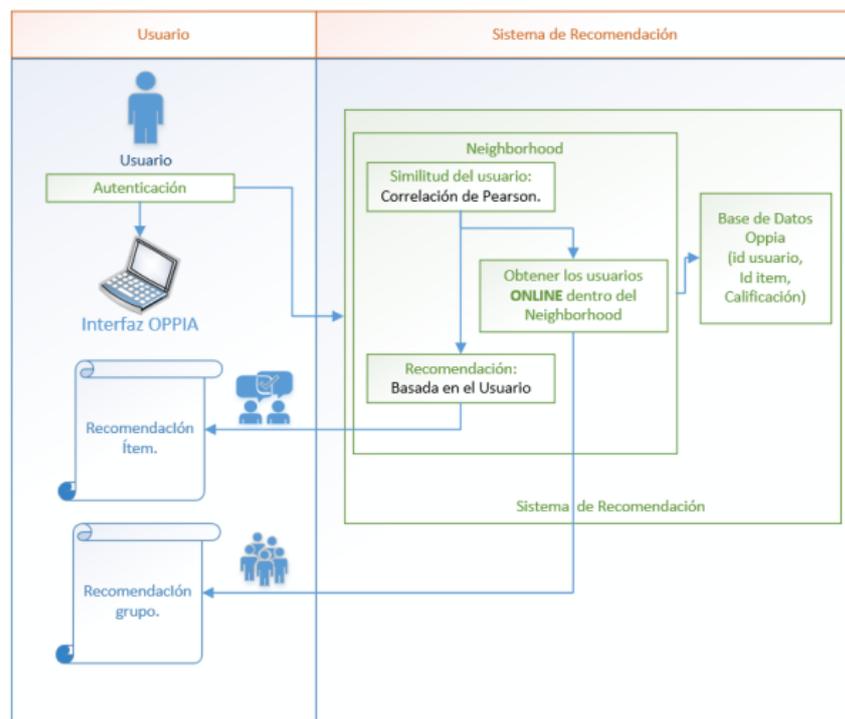
### 2.3.3. Calificación ponderada de un ítem

Para establecer la calificación de un ítem previo a la recomendación del mismo, el sistema utiliza la ecuación 2, cuyos parámetros son los mismos de la ecuación 1:

$$pred(a, b) = \frac{\sum_{b \in N} sim(a, b)(r_{b,p} - \bar{r}_b)}{\sum_{p \in P} sim(a, b)} \quad (3.2)$$

El umbral establecido para que un contenido sea recomendable hacia el usuario es de 3.5. Cuando la calificación supera este umbral la aplicación recomienda un ítem o recurso [3, 34, 35, 36].

Al haber establecido las ecuaciones en las que se basa el sistema de recomendación de contenidos y la formación de grupos de trabajo, el esquema de funcionamiento se muestra en la Figura 2.1.



**Figura 2.1** Análisis de Sentimientos en plataformas educativas.  
**Fuente:** Tomado de [3].

## 2.4. PROCESAMIENTO DE TEXTO

La etapa principal del proyecto se enfoca en el procesamiento de texto usando Web Semántica que es una rama de la inteligencia artificial. El lenguaje elegido debe adaptarse a la plataforma OPPIA para realizar el análisis semántico del comentario.

## **2.5. HERRAMIENTAS DE PROCESAMIENTO DEL LENGUAJE NATURAL**

Para elegir la herramienta se optó por usar la biblioteca NLTK (Natural Language Toolkit) desarrollada por el lenguaje de programación Python 3.6. Siendo esta la mejor opción para alcanzar los objetivos propuestos.

### **2.5.1. NLTK (NATURAL LANGUAGE TOOLKIT)**

Python es una herramienta de software libre, que ha desarrollado la biblioteca NLTK (Natural Language Toolkit) para el procesamiento de texto con la orientación hacia el análisis de sentimientos. En la actualidad existen alrededor de 50 recursos corporales y léxicos para el análisis de datos del lenguaje natural de los humanos. Esta librería se usa para clasificar y analizar un texto, análisis de sentimientos, etiquetar información. Los enfoques que utiliza esta librería son [4, 5].

### **2.5.2. ENFOQUE ESTADÍSTICO**

Este enfoque usa el aprendizaje automático basado en técnicas estadísticas para realizar estimaciones sobre conjuntos de datos desconocidos. Para el análisis se basa en características de elementos anteriores que ya se han determinado o que han sido clasificados de manera correcta [4, 5].

### **2.5.3. NAIVE BAYES**

El clasificador de Bayes es ampliamente utilizado para problemas de análisis de textos en ámbitos académicos e industriales. Las principales características son su simplicidad y eficiencia para una gran cantidad de aplicaciones. Naive Bayes se basa en el teorema de Bayes [4, 5]:

$$p(C_k|x) = \frac{p(C_k) p(x|C_k)}{p(x)} \quad (3.3)$$

La ecuación 3.3 determina que la probabilidad de un evento  $C_k$  dada una situación  $x$ , está determinada por el número de veces que se ha dado el evento  $C_k$  en esa situación  $x$ , dividido por el número de veces que se ha dado el evento  $x$  [4, 5].

El clasificador de Bayes toma en cuenta la contribución de cada palabra importante siendo estas los sustantivos, verbos y adjetivos, para generar una

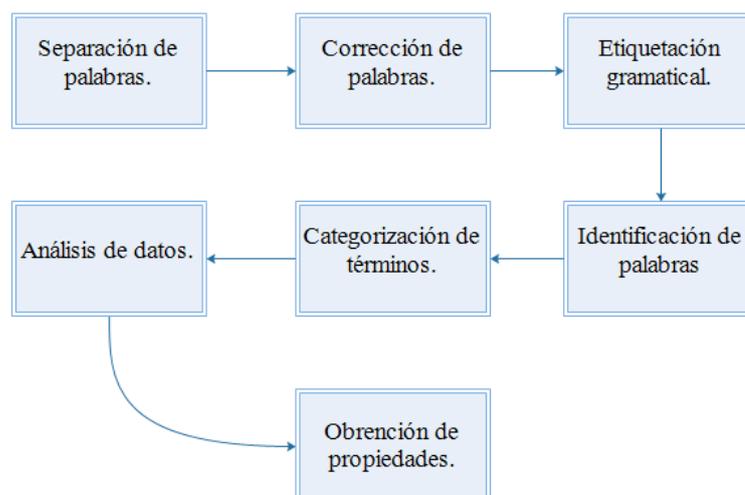
clasificación de manera independiente, de manera opuesta a correlación de varias características, para llevar a cabo una estimación más eficiente en la clasificación de sentimientos [4, 5].

#### 2.5.4. ENFOQUE ANALÍTICO

Este enfoque realiza un análisis morfosintáctico del texto, refiriéndose a la información sintáctica de oraciones y la morfología de las palabras encontradas en el texto. Este enfoque es tradicional, pero involucra mayor complejidad debido a que involucra gramática, sintaxis y semántica [4, 5]

#### 2.6. ANÁLISIS DE SENTIMIENTOS

Este análisis es una minería de opiniones, se basa en la extracción de información emitida por un usuario en una plataforma Web. Los datos son clasificados, etiquetados y analizados [5, 36].



**Figura 2.2** *Análisis de Sentimientos en plataformas educativas.*

**Fuente:** *El Autor.*

La Figura 2.2 muestra el proceso del análisis de sentimientos de textos que utiliza el corpus textblob, el cual es una librería de Python [4, 36]. Este proceso inicia con la separación de palabras, esta etapa contempla la extracción de sustantivos, verbos, adjetivos. En la corrección se eliminan las palabras que no aportan un nivel semántico como las preposiciones y artículos. En la Etiquetación gramatical, el algoritmo asigna valores de objetividad y polaridad a los términos, En las etapas de identificación y categorización de términos o palabras, la aplicación asigna valores referenciales de un nivel de sentimiento positivo o alegre en caso de obtener una

calificación de 0-1 de polaridad; en el caso contrario si se obtiene una calificación entre -1-0, se le califica como negativo o triste [4]. Finalmente, el análisis y obtención de propiedades que se encuentran en el texto analizado, el cual establece un nivel de polaridad entre 0-5 para calificar la recomendación en la plataforma de OPPIA mediante transformaciones efectuadas en el algoritmo.

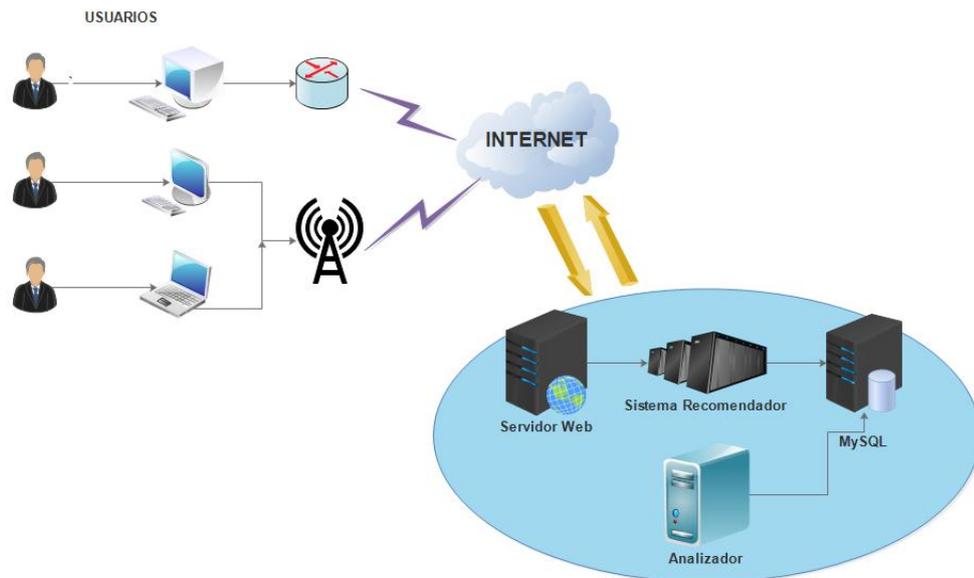
Para obtener un nivel de categorización entre positivo o negativo de un comentario realizado sobre una recomendación se usa la propiedad de polarización. Esta propiedad identifica las emociones y clasifica el texto en base a los sentimientos con los que fue escrito [4, 36].

Los posibles valores que se pueden enmarcan en tres categorías: Polaridad positiva, la cual demuestra que el usuario hizo el comentario con alegría, gratitud, bienestar, etc. Polaridad neutra, el usuario no demuestra emociones. Finalmente, la polaridad negativa, indica rechazo, tristeza u otro sentimiento negativo [4].

El análisis de sentimientos se ve afectado por la carga emocional de las palabras involucradas en el comentario y al igual que la polaridad las palabras tienen una carga positiva, negativa y neutra; todos estos valores van aportan un valor y se va usar un enfoque dependiendo del para dar la calificación final [4, 36].

## **2.7. ARQUITECTURA DEL SISTEMA**

La Figura 2.3 indica la arquitectura del sistema, una vez implementado el analizador de comentarios. Mediante una conexión a Internet el usuario puede acceder a la plataforma OPPIA desde cualquier lugar y espacio físico (universidad, casa). Una vez implementado los algoritmos y las interfaces las pruebas se realizarán en una misma red.



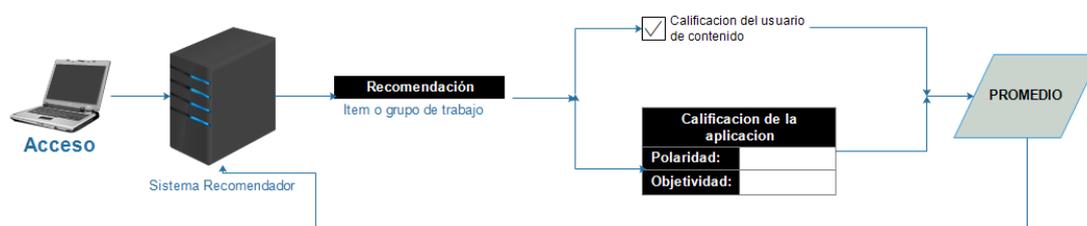
**Figura 2.3** *Arquitectura del sistema de recomendación y análisis de comentarios.*  
Fuente: **El Autor.**

Para el desarrollo del sistema se programó una base de datos en MYSQL, el cual almacena tablas que contiene la información de las calificaciones. Además, se usaron los lenguajes de programación de Python para el análisis de comentarios y PHP para mostrar los datos en una interfaz que se muestra a los usuarios.

# CAPÍTULO 3: IMPLEMENTACIÓN Y ANÁLISIS DE RESULTADOS

## 3.1. DISEÑO DEL SISTEMA

El diagrama de funcionamiento de todo el sistema se ve en la Figura 3.1, la plataforma funciona después de ingresar usuarios y claves de acceso, luego de que el sistema de recomendación actúe se podrá realizar 2 calificaciones: Primero, el usuario realiza la calificación sobre la recomendación mediante una calificación de cinco estrellas demostrando el nivel de satisfacción sobre la recomendación. Segundo, el sistema de análisis de sentimientos actúa después de que el usuario haya compartido un ítem expresando su satisfacción mediante un comentario, el sistema analizará el comentario y devolverá una calificación basada en la polaridad del comentario en una escala de 1-5. Finalmente, el sistema podrá estimar un promedio de todos los comentarios sobre un mismo recurso o ítem, esto para inferir el nivel global de satisfacción a través del cual el sistema hará una predicción del comportamiento de los usuarios y el nivel de satisfacción del ítem.



**Figura 3.1** Diagrama del sistema para el análisis de comentarios.  
Fuente: *El Autor*.

### 3.1.1. BASE DE DATOS

La base de datos fue configurada en MYSQL las tablas necesarias para la implementación de la plataforma se representan en la Tabla 3.1 [3]. Los datos del análisis de sentimientos se agregaron en dos columnas, como lo indica la Tabla 3.2, el campo comp\_calification, contiene las calificaciones realizadas por la aplicación sobre un ítem específico en un rango de 0 a 5. Por su parte el campo prom\_recursos, almacena el promedio que se realiza sobre un mismo ítem, los comentarios provienen de todos los usuarios que han revisado la recomendación, luego el sistema utiliza este promedio para evaluar el nivel de satisfacción y aceptación del ítem.

**Tabla 3.1 Campos principales de la base de datos**

<b>Tablas de la base de datos</b>	Blockedusers
	Friends
	Information
	Ítem
	Likes
	Notifications
	Photos
	profile
	Rating
	recomender_g
	recomender_u
	Status
	Timeline
	Useroptions
Users	

**Tabla 3.2 Detalle del campo status**

<b>Status</b>	id_s
	Osid

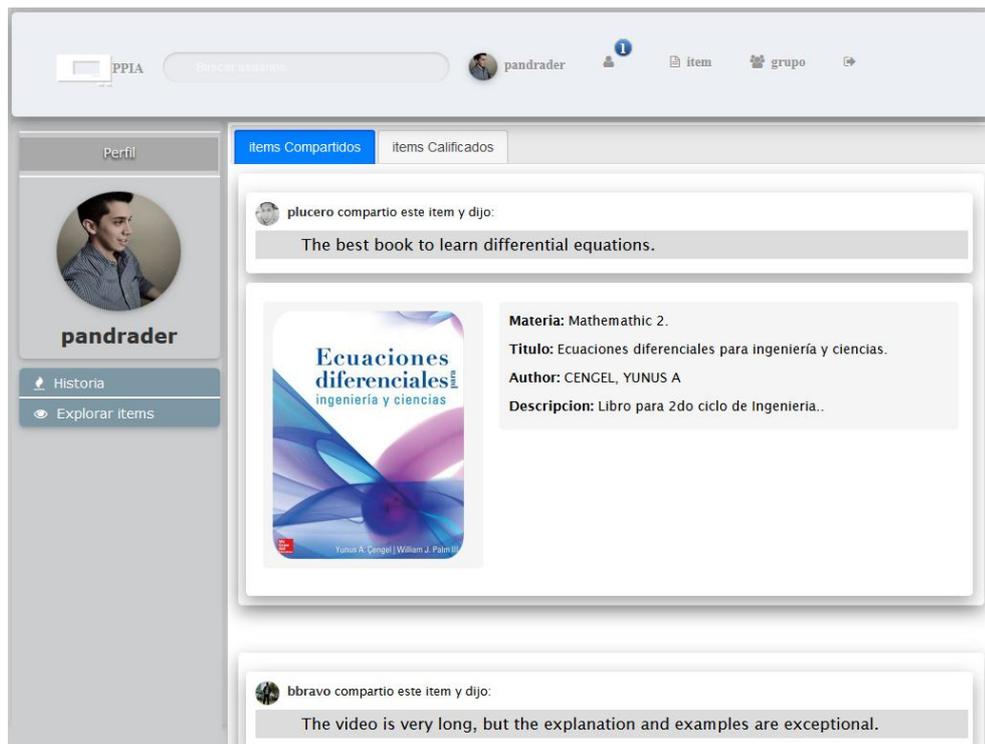
account_name
Autor
id_author
Type
Data
Postdate
Ítem
name_s
like_s
dislike_s
comp_calification
prom_recurso

### 3.1.2. SERVIDOR WEB

Para el presente proyecto se usó un servidor de código libre XAMPP, el cual maneja paquetes de Apache, MySQL, entre otros; estos servidores han sido implementados en nuestra plataforma. Este servidor realiza todas las configuraciones, es fácil su manipulación, pero el punto débil es la seguridad a este nivel nuestros datos son vulnerables.

### 3.1.3. INTERFAZ DEL USUARIO

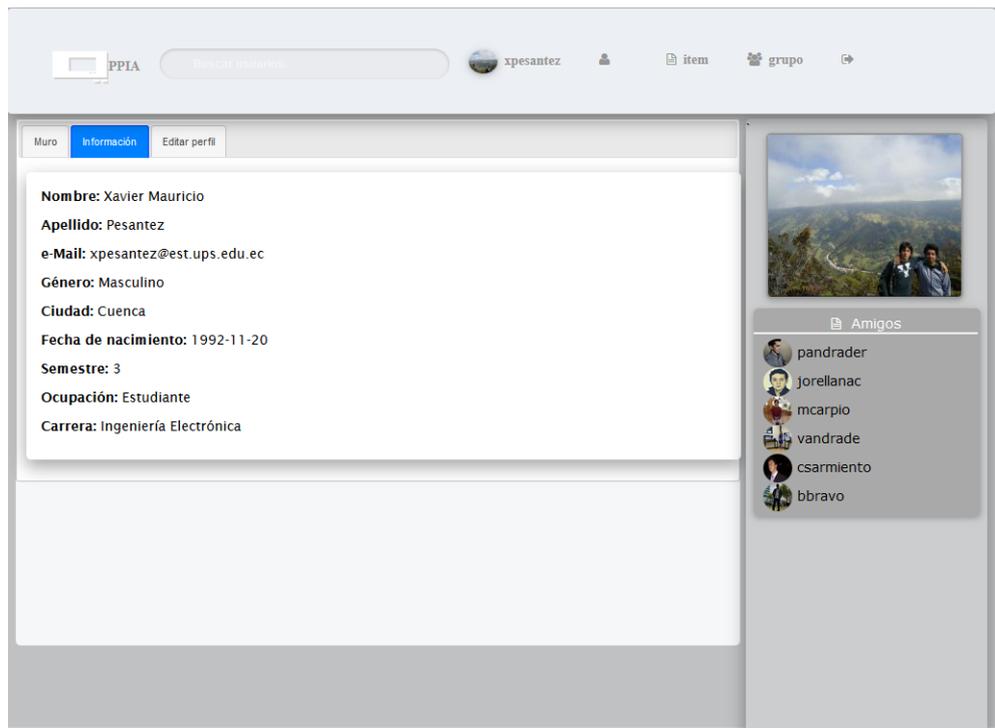
La interfaz del usuario presenta las siguientes ventanas luego de acceder a la plataforma [3]:



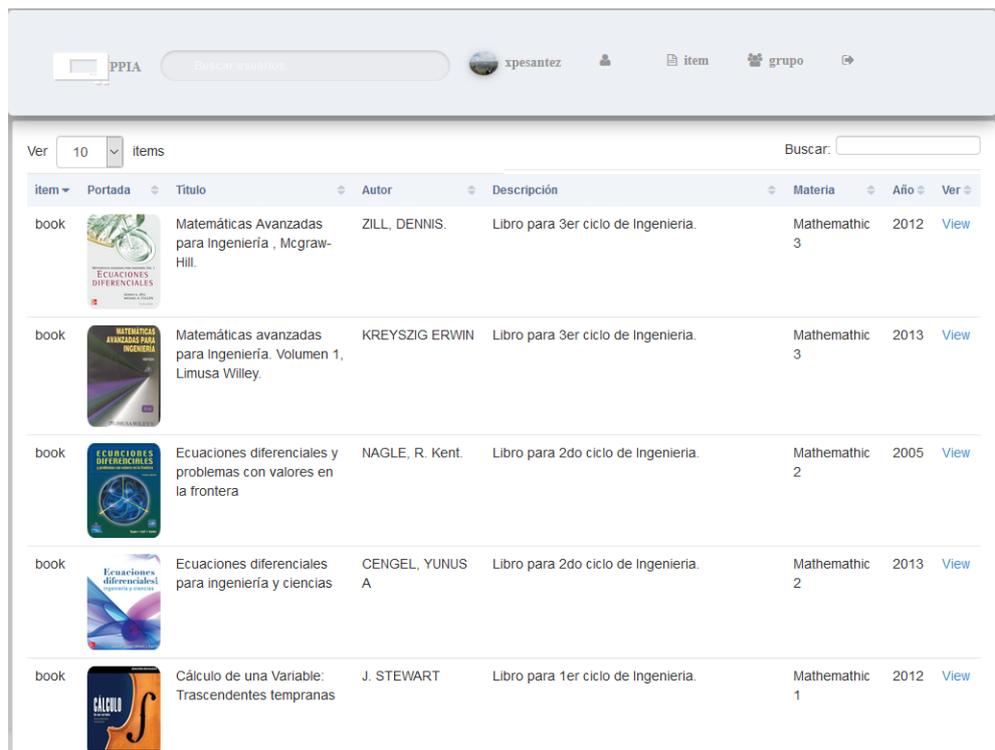
**Figura 3.2** Ventana Home.

Fuente: *El Autor*.

La Figura 3.2 muestra la ventana Home que se visualiza al momento de acceder a la plataforma OPPIA, la cual contiene información del usuario, pestañas que dan acceso a ítems compartidos e ítems calificados. La Figura 3.3 indica la información que el estudiante puede publicar en la plataforma OPPIA implementada, en la cual se muestra los nombres completos, e-mail, género, ciudad, fecha de nacimiento y datos de la carrera que estudia; además de los amigos que tiene aceptados en el sistema. Toda esta información puede ser cambiada [3, 30].



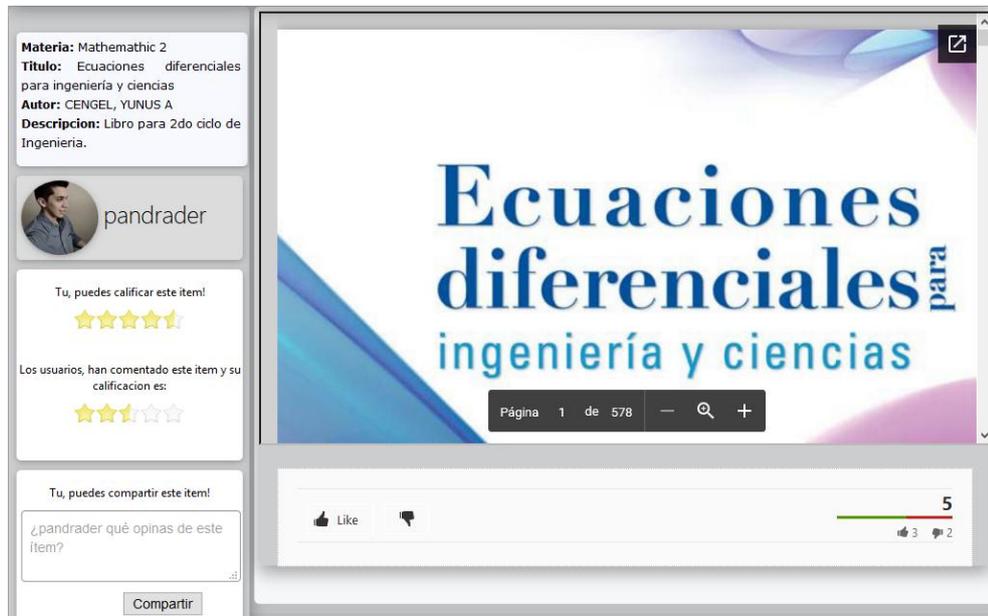
**Figura 3.3** Ventana de Perfil de Usuario.  
Fuente: *El Autor.*



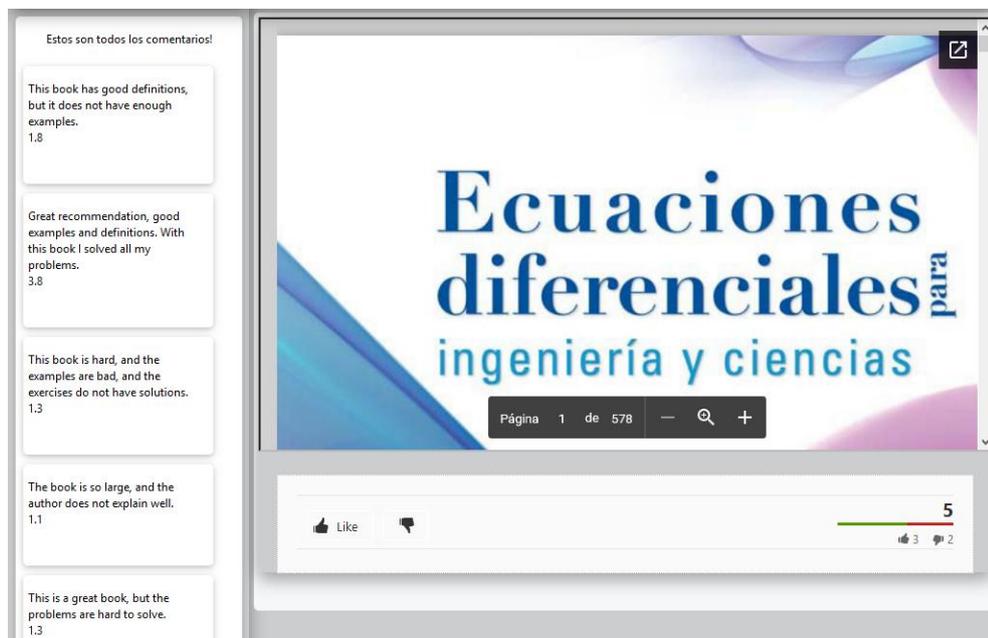
**Figura 3.4** Ventana de consulta de contenidos.  
Fuente: *El Autor.*

En la Figura 3.4, el sistema brinda la posibilidad de explorar los ítems colocados por un tutor para las diferentes cátedras, en las cuales el estudiante necesite ayuda. La Figura 3.5 muestra el ítem con la calificación del recurso,

en dos partes primero a la izquierda debajo de la información del ítem y el usuario en estrellas se muestra la calificación que realiza el usuario sobre la recomendación, enseguida y más abajo se muestra el promedio de calificaciones que va almacenado el ítem por n usuarios. Además de los comentarios que va teniendo el ítem en la Figura 3.6.



**Figura 3.5** Ventana de ítems y calificaciones.  
Fuente: *El Autor*.



**Figura 3.6** Ventana de ítem y comentarios.  
Fuente: *El Autor*.

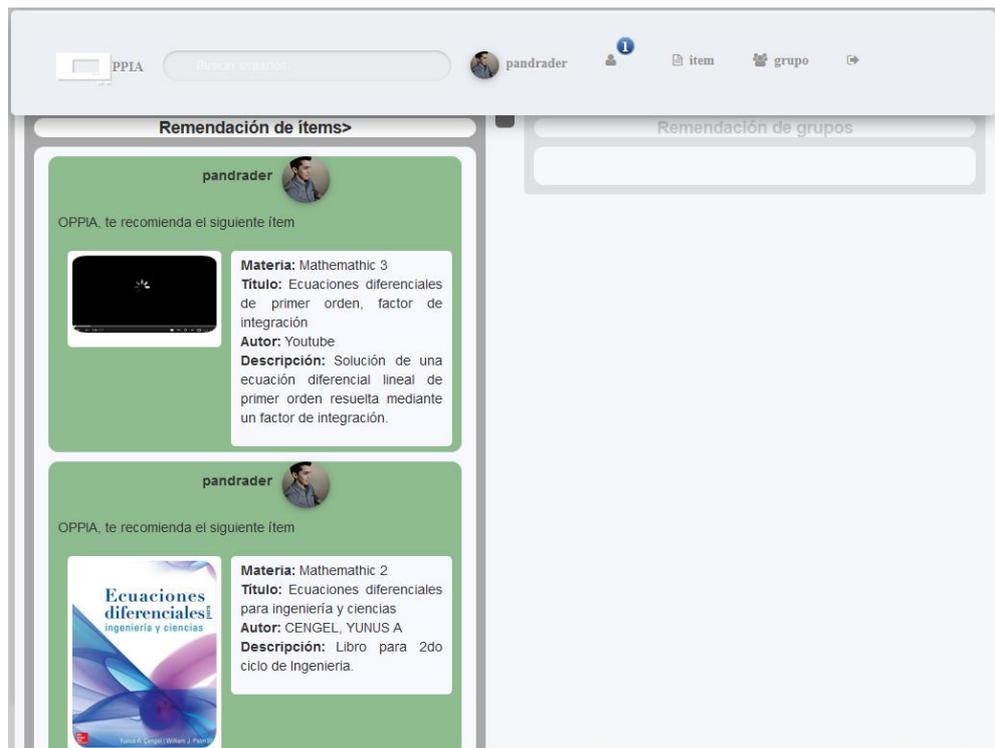
## **3.2. ANÁLISIS DE RESULTADOS**

En este capítulo muestra primero, la generación de contenido basada en las preferencias de los usuarios y la formación de grupos de trabajo entre usuarios con una similitud entre los mismos; segundo, el análisis de comentarios realizado sobre las recomendaciones compartidas, además de sus respectivas calificaciones y promedios.

Antes de realizar el estudio de los resultados, el usuario debe ingresar a la plataforma con su usuario y la respectiva contraseña, toda esta información se encuentra almacenada en la base de datos. El total de usuarios creados es 10, todos tienen las mismas prioridades dentro del sistema.

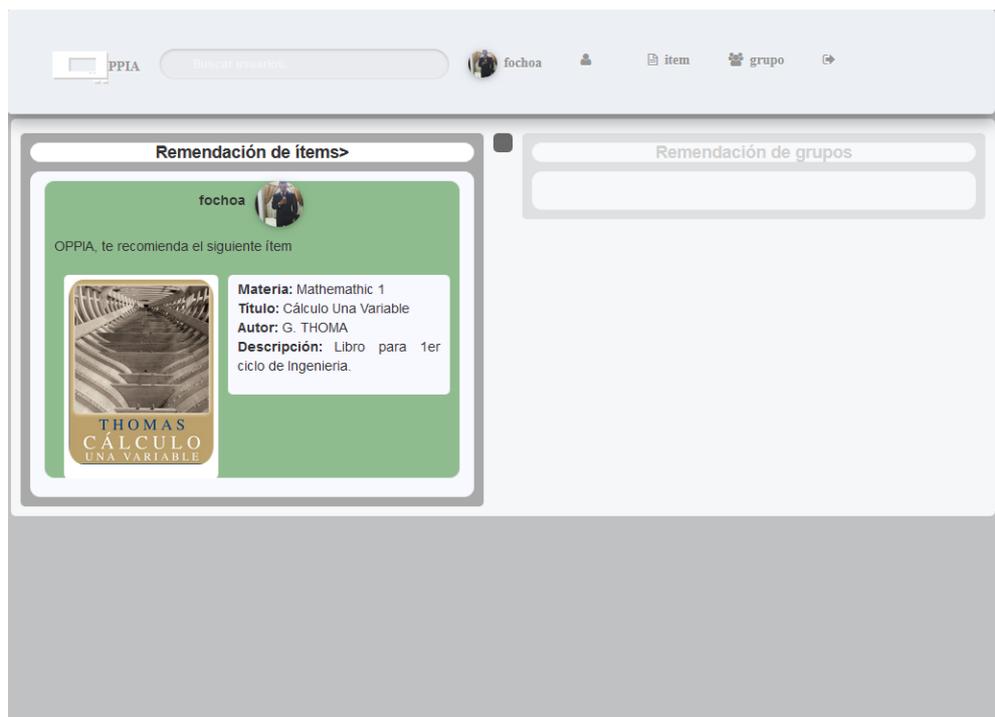
### **3.2.1. RECOMEDACIÓN DE CONTENIDO**

La recomendación de contenido en OPPIA se realiza en base a las calificaciones realizadas por los usuarios en un determinado ítem, los cálculos se realizan utilizando la ecuación 3.2 para determinar si el ítem alcanza el umbral de recomendación, establecido en 3.5. El sistema se probó con el ítem 4 (Ecuaciones diferenciales para ingeniería y ciencias), el cual en base a las calificaciones realizadas por los demás usuarios que ya han recibido la recomendación, se realiza los cálculos, el valor ponderado por el sistema es de 4.2; este número supera el umbral y la recomendación recibe es recibida por el usuario *pandrader*, ver la Figura 3.7.



**Figura 3.7** Ventana de Recomendación de contenidos para pandrader.  
Fuente: *El Autor*.

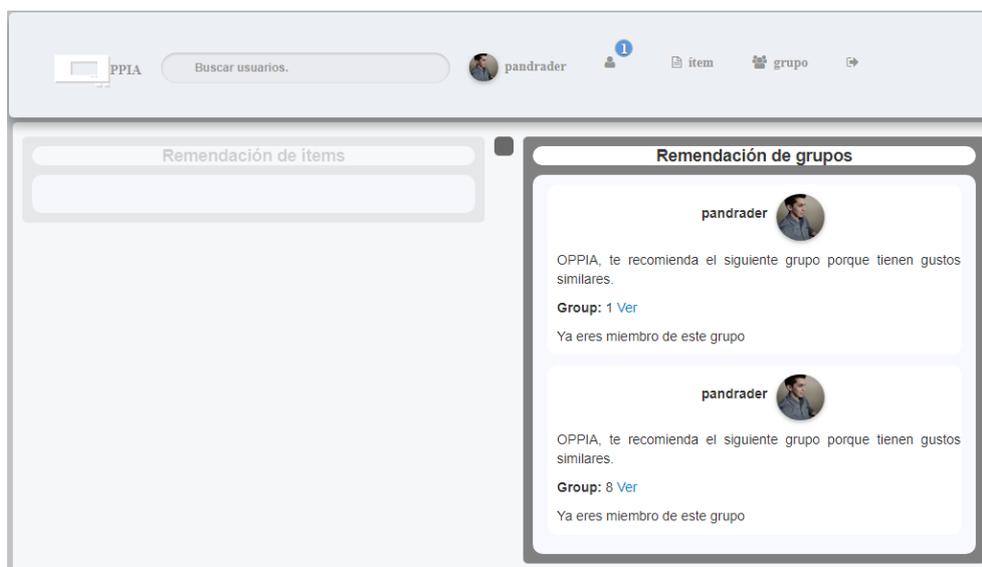
Por otra parte, el usuario *fochoa* que necesita una ayuda similar a *pandrader* recibe una calificación ponderada de 2.87 por consiguiente no recibe la recomendación como lo muestra la Figura 3.8. La diferencia de calificación se debe a las preferencias que registra la plataforma, es decir cada contenido se adapta al usuario.



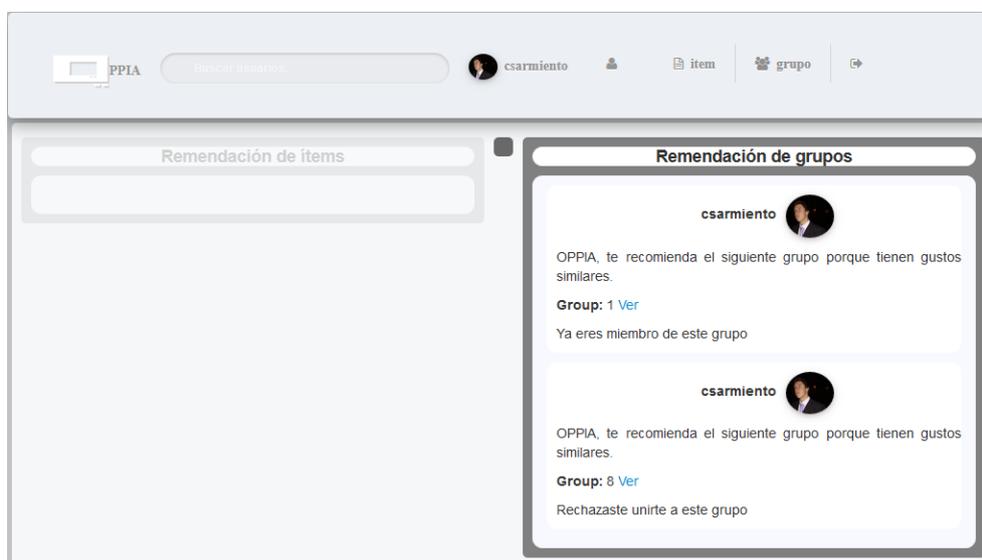
**Figura 3.8** Ventana de Recomendación de contenidos para fochoa].  
Fuente: *El Autor*.

### 3.2.2. FORMACIÓN DE GRUPOS

La formación de grupos se realiza con usuarios que estén online en OPPIA, los grupos se forman en base a la similitud entre usuarios el umbral en este caso es de 0.5 utilizando la ecuación 3.2. Entonces los usuarios que tienen un promedio mayor a 0.5 son *pandrade* y *csarmiento*, al encontrarse en línea, ambos reciben la solicitud de formación de grupo; así lo muestran las Figuras 3.9 y 3.10.



**Figura 3.9** Ventana de Formación de grupos para pandrader.  
Fuente: *El Autor*.



**Figura 3.10** Ventana de Formación de grupos para csarmiento.  
Fuente: *El Autor*.

Los datos obtenidos de los comentarios se efectuaron en dos etapas. La primera se realiza un análisis individual sobre los comentarios que se realizan en diferentes ítems que fueron recomendados por el sistema. La segunda etapa analiza el promedio

que se va aculando en el análisis de comentarios realizados sobre un mismo ítem, cuyo valor sirve de referencia acerca del nivel de aceptación de una recomendación, con lo cual dependiendo del valor que se vaya acumulando, la recomendación se evaluará para futuros usuarios que necesiten esa recomendación.

### **3.2.3. ETAPA 1, ANÁLISIS INDIVIDUAL**

En la Tabla 3.3 se resumen las calificaciones que dispone el sistema para la recomendación de contenido. Primero, en la columna de Calificación Usuario se muestra las calificaciones que puede hacer el usuario en una recomendación que realiza el sistema. En la siguiente columna Calificación Aplicación se muestra la calificación que realiza la aplicación usando el código implementado y corresponde al análisis de sentimientos realizada en el comentario correspondiente al ítem recomendado. Finalmente se obtiene un promedio entre ambas calificaciones para medir el grado real de satisfacción de un comentario.

El análisis de sentimientos basa su funcionamiento en palabras que realzan el significado de una oración como son los adjetivos y verbos, los sustantivos ayudan de manera subjetiva en el análisis general, finalmente los artículos y preposiciones son excluidos del análisis debido a q no aportan polaridad o subjetividad en el análisis de sentimiento.

El análisis fue realizado en inglés, por lo tanto, existen palabras denominadas conectores, estas palabras dan sentido a las oraciones como los conectores adversativos, estas palabras sirven para oponer una dificultad a la oración previa tal es el caso de *but*. Podemos encontrar conectores causales que dan una razón a un hecho expresado en la oración previa. Existen más conectores en inglés, pero no es tema de nuestro análisis. Con estos antecedentes se puede proceder con el análisis de los datos obtenidos en diferentes pruebas de la plataforma.

**Tabla 3.3 Tabla de resultados de calificaciones individuales de comentarios.**

<b>No.</b>	<b>Usuario</b>	<b>Comentario</b>	<b>Calificación Usuario</b>	<b>Calificación Aplicación</b>	<b>Promedio</b>
1	<i>Plucero</i>	This book has good definitions, but it does not have enough examples.	4,00	1,75	2,875
2	<i>Bbravo</i>	Great recommendation, good examples and definitions. With this book I solved all my problems.	4,00	3,75	3,875
3	<i>Bbravo</i>	This book is old, and the examples are bad, and the exercises do not have answers.	2,00	1,75	1,875
4	<i>Bbravo</i>	In my opinion, the multimedia is very useful to understand and solve mathematical topics, but this video is so large, and the author does not explain well.	2,40	1,01	1,705
5	<i>csarmiento</i>	The link is good, but it does not have examples or exercises.	3,50	3,50	3,50
6	<i>csarmiento</i>	Great video!!!! It is valuable, and the example is complete. It is satisfactory.	4,70	2,75	3,73
7	<i>csarmiento</i>	Good example to learn convolution.	4,00	3,50	3,75
8	<i>csarmiento</i>	Very good video, excellent and easy. The recommendations are excellent.	4,60	4,18	4,39
9	<i>Vandrade</i>	The video is very long, but the explanation and examples are exceptional.	3,80	1,50	2,65
10	<i>Vandrade</i>	The best book to learn differential equations.	5,00	5,00	5,00
11	<i>Mcarpio</i>	This book is very useful! Recommended for student that start with calculus.	3,50	2,44	2,97
12	<i>Fochoa</i>	This is a great book, but the problems are hard to solve.	2,50	1,27	1,89

**Tabla 3.4** Tabla de resultados de calificaciones y sus promedios.

<b>Item</b>	<b>Usuario</b>	<b>Comentario</b>	<b>Calificación Usuario</b>	<b>Calificación Aplicación</b>	<b>Promedio Acumulado</b>
4	<i><u>pablewz</u></i>	This book has good definitions, but it does not have enough examples. Very good!	4,00	1,80	1,80
4	<i><u>xpesantez</u></i>	Great recommendation, good examples and definitions. With this book I solved all my problems.	5,00	3,80	2,80
4	<i><u>jorellana</u></i>	This book is hard, and the examples are bad, and the exercises do not have solutions.	3,00	1,30	2,30
4	<i><u>fochoa</u></i>	The book is so large, and the author does not explain well.	2,00	1,10	2,00
4	<i><u>cortiz</u></i>	This is a great book, but the problems are hard to solve.	4,00	1,30	1,90
4	<i><u>mcarpio</u></i>	Great book!!!! It is valuable, and the example is complete. It is satisfactory.	5,00	2,80	2,00
4	<i><u>vandrade</u></i>	This recommendation is satisfactory.	4.90	2,50	2,00
4	<i><u>csarmiento</u></i>	Very good book, excellent and easy. The recommendations are excellent.	4.50	4,20	2,40
4	<i><u>bbravo</u></i>	The book is very long, but the explanation and examples are exceptional.	4.50	1,50	2,30
4	<i><u>plucero</u></i>	The best book to learn differential equations.	5,00	5,00	2,50

En la Tabla 3.3 se muestran 12 calificaciones de distintos ítems basadas en el análisis de sentimientos. Los comentarios No. 1, 3, 4, 9, 12; expresan un sentimiento de oposición por ello el sistema les da una calificación de poca satisfacción debido a que en el comentario los usuarios presentan un nivel neutro de aceptación de la recomendación además todos estos comentarios usan un conector común *but*.

Por otra parte, los comentarios que usan palabras de felicidad y aceptación son los comentarios No. 2, 5, 7, 8, 10. En estos comentarios los usuarios expresan felicidad y aceptación por lo tanto el análisis de sentimientos indica que la oración expresa un buen nivel de satisfacción entonces han sido calificados con valores altos entre 3-5. El comentario 9, expresa baja aceptación de la recomendación por lo tanto el sistema de análisis de sentimientos da una calificación baja debido a que encuentra un nivel de satisfacción bajo.

Todos los valores que se obtiene a partir del análisis de sentimientos se guardan en la base de datos programada para el análisis de la siguiente etapa.

#### **3.2.4. ETAPA 2, PROMEDIO GLOBAL**

En esta etapa de pruebas se procedió a realizar comentarios en un mismo ítem, para recopilar información de las calificaciones que efectúa la aplicación y el respectivo promedio que realiza después de cada comentario. Para el ítem 4 se efectuaron 10 comentarios, la aplicación por lo tanto suministró 10 calificaciones. Los comentarios y sus calificaciones se almacenan en la base de datos y luego de que un usuario realice el comentario respectivo busca todas las calificaciones del ítem y devuelve un promedio el cual expresa el nivel de satisfacción de todos los usuarios que han recibido la misma recomendación hasta ese momento, estos valores se pueden ver en la Tabla 3.4, en la columna Promedio Acumulado.

El sistema busca encontrar el nivel de satisfacción que expresan los usuarios a través de la calificación directa y la calificación que brinda la plataforma haciendo un análisis de sentimientos en el comentario. Por lo tanto, se ha buscado la manera de recopilar una calificación que exprese ese nivel de satisfacción promediando todos los comentarios realizados en el mismo ítem. Para determinar el nivel de aceptación se usa el promedio acumulado y el sistema se asegura, que la recomendación es la

adecuada o caso contrario realizará la recomendación de otro recurso para suplir las necesidades educativas de los usuarios.

# **CAPÍTULO 4: CONCLUSIONES Y RECOMENDACIONES**

A partir del análisis del estado de arte de sistemas de la Web Semántica, e-Learning y sistemas de recomendación de contenido y grupos de trabajo se determinó todos los estándares, lenguajes de programación y librerías. Además de los resultados obtenidos y analizados, después de ejecutar los algoritmos para el análisis de comentarios se lograron los objetivos planteados y se estableció las siguientes conclusiones y recomendaciones:

## **4.1. CONCLUSIONES**

- Al analizar el estado del arte respecto a sistemas e-learning, la Web Semántica aplicada a redes sociales educativas, además de los sistemas de NLP o Procesamiento del Lenguaje Natural, se determinaron los estándares, lenguajes de programación, algoritmos con estas tecnologías que se usaron para implementar el análisis de comentarios en nuestra plataforma educativa y obtener las calificaciones y promedios respectivos sobre los ítems evaluados.
- El algoritmo diseñado, implementado, probado y adaptado a la plataforma OPPIA cumple con el objetivo principal del proyecto que es analizar los comentarios y devolver información que expresa el nivel de satisfacción de los usuarios acerca de una recomendación realizada por el sistema, con ello el sistema mejora su eficiencia y conoce la aceptación de los recursos por parte de los usuarios mediante el promedio realizado en todos los comentarios y ahora el sistema no solo tiene la calificación que da el usuario, sino una que genera la aplicación y se promedia con todas las calificaciones.

- La plataforma se vuelve robusta y contribuye con el acceso a la información, rompe con los esquemas tradicionales de alumno-docente ya que utiliza tecnologías actuales para mejorar el servicio, todo esto bajo una interfaz fácil de utilizar y contribuye a la evolución de la educación.
- El uso de librerías de software libre permitió desarrollar un algoritmo que permite analizar los sentimientos que expresa un usuario en un comentario. Dichas librerías son el textblob y NLTK de Python 3.6. Con resultados se determinó la utilidad y eficiencia del algoritmo en el análisis de comentarios realizados sobre recursos del sistema implementado.
- El resultado obtenido por el análisis de comentarios es satisfactorio y ha hecho de nuestra plataforma un sistema robusto, debido a que los comentarios efectuados dan una calificación, la misma sirve para demostrar el nivel de aceptación y satisfacción de los usuarios sobre una recomendación específica y en base al historial de comentarios efectuados sobre el mismo recurso.

## **4.2. RECOMENDACIONES**

Al finalizar este proyecto se puede realizar las siguientes recomendaciones para futuros proyectos que usen el procesamiento del lenguaje natural:

- La librería textblob empleada en nuestra plataforma, dentro del análisis de sentimientos sirve para el procesamiento de contenidos como: la tokenización, extracción de raíces, eliminación de stopwords, corrección de texto para expresiones regulares y caracteres repetidos, sinónimos. Con ello, el uso de un lenguaje abierto brinda una gran variedad de usos y aplicaciones.

La Web actual contiene información masiva y heterogénea, entonces se vuelve necesario la implementación de la Web Semántica en las aplicaciones actuales para que así las máquinas y las personas tengan un lenguaje común y conozcan la información que disponen, a través de la especificación de los recursos en la Web.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] D. Castellanos Nieves et al., Sistema basado en tecnologías de la web semántica para evaluación en entornos de e-learning. Universidad de Murcia, 2007.
- [2] O. Menemencioglu and I. M. Orak, “A review on semantic web and recent trends in its applications,” in Semantic Computing (ICSC), 2014 IEEE International Conference on. IEEE, 2014, pp. 297–303.
- [3] P. E. Andrade Rea et al., “Diseño e implementación de un prototipo para un sistema de recomendación de contenido y formación de grupos de aprendizaje en una plataforma educativa,” B.S. thesis, 2017.
- [4] J. M. Hernández Hernández, “Análisis automático de textos en español utilizando nltk,” 2016.
- [5] M. Elhadad, “Natural language processing with python Steven Bird, Ewan Klein, and Edward Loper (University of Melbourne, University of Edinburgh, and bbn technologies) sebastopol, ca: O’reilly media, 2009,” 2010.
- [6] Alix Aguirre Andrade and Nelly Manasía Fernández. Web 2.0 y web semántica en los entornos virtuales de aprendizaje. *Multiciencias*, 9(3), 2009.
- [7] A. C. Alonso de Armiño and L. M. López, “Incorporación de semántica en plataformas para e-learning,” in III Congreso de Tecnología en Educación y Educación en Tecnología, 2008.

- [8] C. Gutierrez, "Como funciona la web," Santiago de Chile: Centro de Investigación de la Web Universidad de Chile, 2008.
- [9] G. E. Yahuita Quisbert, "Web semántica y representación del conocimiento con ontologías," *Revista del Postgrado en Informática*, p. 57, 2014.
- [10] D. Fernández Medina, "Estudio del modelo de representación XML/RDF," B.S. thesis, Universitat Oberta de Catalunya, 2010.
- [11] R. García Ruiz, "Xml y web semántica: estudio del impacto de las aplicaciones comerciales basadas en tecnologías de web semántica," B.S. thesis, Universitat Oberta de Catalunya, 2008.
- [12] C. Asmal, J. Fernando, and A. O. Vázquez Patiño, "Integración de la web 2.0 y la web semántica mediante mashups semánticos," B.S. thesis, 2010.
- [13] D. Fernández, G. López, and I. Jeder, *La naturaleza de la información semántica en la Web*. B-Universidad de Buenos Aires, 2009.
- [14] C. Bhatia and S. Jain, "Semantic web mining: Using ontology learning and grammatical rule inference technique," in *Process Automation, Control and Computing (PACC)*, 2011 International Conference on. IEEE, 2011, pp. 1–6.
- [15] Juan Antonio Pastor Sánchez. *Tecnologías de la web semántica*. Editorial UOC, 2011.
- [16] T. Segaran, C. Evans, and J. Taylor, *Programming the Semantic Web: Build Flexible Applications with Graph Data*. " O'Reilly Media, Inc.", 2009.
- [17] D Bernal, A Castro, and J González. *Web semántica, más de una década de su aparición*. *Revista PUENTE Científica*, 8(1), 2014.
- [18] Andreu Sulé, Miquel Centelles, Jorge Franganillo, and Jesús Gascón. *Aplicación del modelo de datos rdf en las colecciones digitales de bibliotecas, archivos y museos de españa*. *Revista española de Documentación Científica*, 39(1):121, 2016.
- [19] D. Beckett and B. McBride, "Rdf/xml syntax specification (revised)," W3C recommendation, vol. 10, no. 2.3, 2004.

- [20] D. Brickley, R. V. Guha, and B. McBride, “Rdf vocabulary description language 1.0: Rdf schema. W3C Recommendation (2004),” URL <http://www.w3.org/tr/2004/rec-rdf-schema-20040210>, 2004.
- [21] D. Brickley and R. Guha, “Rdf schema 1.1-W3C Recommendation 25 february 2014, W3C Recommendation,” 2015.
- [22] Silvana Castro and Marina Larraud. Recuperación de información bilingüe en la web semántica. 2007.
- [23] Alexander Maedche and Steffen Staab. Learning ontologies for the semantic web. In Proceedings of the Second International Conference on Semantic Web- Volume 40, pages 51–60. CEUR-WS.org, 2001.
- [24] G. Antoniou, E. Franconi, and F. Van Harmelen, “Introduction to semantic web ontology languages,” in Reasoning web. Springer, 2005, pp. 1–21.
- [25] Peter Mika. Flink: Semantic web technology for the extraction and analysis of social networks. Web Semantics: Science, Services and Agents on the World Wide Web, 3(2):211–223, 2005.
- [26] D. L. McGuinness, F. Van Harmelen et al., “Owl web ontology language overview,” W3C recommendation, vol. 10, no. 10, p. 2004, 2004.
- [27] B. GHIRARDINI, “Metodologías de e-learning: una guía para el diseño y desarrollo de cursos de aprendizaje empleando tecnologías de la información y las comunicaciones,” FAO, Roma (Italia), Tech. Rep., 2014.
- [28] Lluís Codina and Cristòfol Rovira. La web semántica. In Tendencias en documentación digital. Trea, 2006.
- [29] Alberto Garzón García et al. Prototipo de consulta de información biológica en la web de datos. 2013.
- [30] J. F. Bravo-Torres, V. E. Robles-Bykaev, M. L. Nores, E. F. OrdóñezMorales, Y. Blanco-Fernández, and A. Gil-Solla, “Oppia: A contextaware ubiquitous learning platform to exploit short-lived student networks for collaborative learning.” in CSEDU (1), 2016, pp. 494–498.

- [31] P. E. Vintimilla-Tapia, J. F. Bravo-Torres, P. L. Gallegos-Segovia, E. F. Ordoñez-Morales, M. López-Nores, and Y. Blanco-Fernández. Oppia: A multi-model platform for e-learning. In 2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON), pages 1–4, Aug 2017
- [32] P. Karampiperis, A. Koukourikos, and G. Stoitsis, “Collaborative filtering recommendation of educational content in social environments utilizing sentiment analysis techniques,” in *Recommender Systems for Technology Enhanced Learning*. Springer, 2014, pp. 3–23.
- [33] N. Manouselis, H. Drachsler, K. Verbert, and E. Duval, *Recommender systems for learning*. Springer Science & Business Media, 2012.
- [34] A. A. Gokhale, "Mining educational data: A focus on learning analytics," in 2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS), pp. 1-1, Dec 2015.
- [35] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, *Recommender systems: an introduction*. Cambridge University Press, 2010.
- [36] J. BUCAR and J. POVH, “Sentiment analysis in web text: An overview,” *Studies*, vol. 2412, p. 24.