

SEP

TNM

INSTITUTO TECNOLÓGICO DE CULIACÁN



Generación de un corpus de frases relacionadas con el aprendizaje
para el análisis de sentimientos

TESIS

PRESENTADA ANTE EL DEPARTAMENTO ACADÉMICO DE ESTUDIOS DE POSGRADO
DEL INSTITUTO TECNOLÓGICO DE CULIACÁN EN CUMPLIMIENTO PARCIAL DE LOS
REQUISITOS PARA OBTENER EL GRADO DE

MAESTRO EN CIENCIAS DE LA COMPUTACIÓN

POR:

Sandra Lucía Ramírez Ávila

LICENCIADA EN CIENCIAS COMPUTACIONALES

DIRECTOR DE TESIS:

Dra. María Lucía Barrón Estrada

CULIACÁN, SINALOA

Agosto 2018

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

Culiacán, Sin., 10 de Agosto del 2018

OFICIO: DEPI-336/VIII/2018

ASUNTO: **Autorización Impresión**

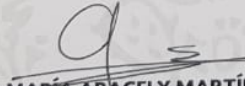
LIC. SANDRA LUCÍA RAMÍREZ ÁVILA
ESTUDIANTE DE LA MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN
PRESENTE.

Por medio de la presente y en virtud de que ha completado los requisitos para el examen de grado de la **Maestría en Ciencias de la Computación**, se concede autorización para la impresión de la tesis titulada: **"GENERACIÓN DE UN CORPUS DE FRASES RELACIONADAS CON EL APRENDIZAJE PARA EL ANÁLISIS DE SENTIMIENTOS"**, bajo la dirección del(a) **Dra. María Lucía Barrón Estrada**.

Sin otro particular reciba un cordial saludo.

ATENTAMENTE
"CON LA TÉCNICA AL PROGRESO"

SEP TecNM
Instituto Tecnológico
de Culiacán
División de Estudios
de Posgrado e Investigación


M.C. MARÍA ARACELY MARTÍNEZ AMAYA
JEFE(A) DE LA DIVISIÓN DE ESTUDIOS DE
POSGRADO E INVESTIGACIÓN

C.c.p. archivo

MAMA/lucy *

Dedicatoria

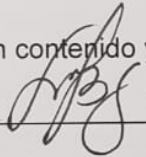
Dedicamos esta tesis de tesis a los padres María Elena y José Manuel Ramírez
y a quienes han sido un apoyo incondicional para lograr esta tesis.

**“GENERACIÓN DE UN CORPUS DE FRASES
RELACIONADAS CON EL APRENDIZAJE PARA EL
ANÁLISIS DE SENTIMIENTOS”**

Tesis presentada por:

LIC. SANDRA LUCÍA RAMÍREZ ÁVILA

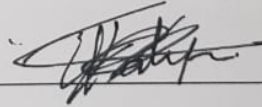
Aprobada en contenido y estilo por:



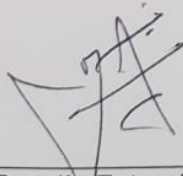
Dra. María Lucía Barrón Estrada
Director de Tesis



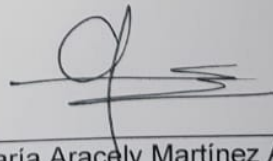
Dr. Ramón Zatarain Cabada
Secretario



M.C. Gloria Ekaterine Peralta Peñuñuri
Vocal -1



M.C. Rosalío Zatarain Cabada
Vocal -2



M.C. María Aracely Martínez Amaya
Jefe(a) de la División de Estudios de
Posgrado e Investigación

Dedicatoria

Dedico este trabajo de tesis a mis papás Norma Alicia Ávila F. y Jesús Salvador Ramirez G. quienes han sido un apoyo y motivación para superarme. Ellos que siempre estuvieron ahí para escucharme, aconsejarme, alentarme.

A mis hermanos Estefanía Ramirez A., Salvador Ramirez A. y Natalia Ramirez A... Ellos quienes se comunicaban para brindarme apoyo en todo momento, para hacerme sentir que no estaba sola y que podía lograr cualquier reto que me propusiera.

Agradecimientos

Quiero agradecer al Instituto Tecnológico de Culiacán por abrirme las puertas de estudiar un posgrado.

A CONACYT por otorgarme el apoyo económico para hacer posible la realización de estudios de posgrado.

A mi asesora Dra. María Lucía Barrón Estrada por compartir su conocimiento, por su apoyo incondicional durante mis estudios de maestría.

A mis profesores, Dr. Ramón Zatarain, Dr. Ricardo Quintero, Dr. Carlos Santillán y Dr. Héctor Rodríguez quienes compartieron su conocimiento durante mis estudios de maestría.

A mi familia que me apoyo a realizar una maestría fuera de mi área, ellos quienes me motivaron en todo momento y siempre estuvieron al pendiente de mí.

A mis compañeros y amigos Aldo Uriarte, Armando Guerra, Giovanni Montelongo y Miguel Montoya por su gran ayuda, motivación constante. Ellos quienes hicieron que mi estancia en la maestría fuera agradable.

A todas aquellas personas que conocí durante mi estancia en Sinaloa, ya que siempre tuvieron alguna palabra de apoyo para mí durante mis estudios de posgrado.

Declaración de autenticidad

Por la presente declaro que, salvo cuando se haga referencia específica al trabajo de otras personas, el contenido de esta tesis es original y no se ha presentado total o parcialmente para su consideración para cualquier otro título o grado en esta o cualquier otra Universidad. Esta tesis es resultado de mi propio trabajo y no incluye nada que sea resultado de algún trabajo realizado en colaboración, salvo que se indique específicamente en el texto.

Sandra Lucia Ramírez Ávila

Culiacán, Sinaloa

México, 2018

Resumen

Las emociones que un estudiante experimenta durante el proceso de aprendizaje son importantes para alcanzar las competencias y conocimientos que se proveen durante ese proceso; es por ello, que en el ámbito de la educación, muchos sistemas actuales involucran diferentes dispositivos electrónicos con sensores que permiten captar señales de los usuarios y utilizan esas señales para generar información importante que apoye a los usuarios en los procesos cognitivos.

Todas las personas experimentan emociones básicas como: miedo, alegría, tristeza, ira, etc, pero los estudiantes experimentan emociones secundarias que emergen durante las actividades de aprendizaje profundo como son: frustrado, aburrido, comprometido y emocionado; estas son conocidas como emociones centradas en el aprendizaje (Baker, 2010), (D’Mello S. &, 2012). Las emociones centradas en el aprendizaje juegan un papel importante en los procesos cognitivos de los estudiantes. Existen diferentes formas de reconocer emociones utilizando diferentes elementos como: imagen de la cara, voz, texto, expresiones corporales, frecuencia cardiaca, etc. Hoy en día muchos sistemas computacionales detectan emociones básicas usando el rostro del usuario que captan a través de la cámara, sin embargo en el ámbito educativo se requiere reconocer emociones centradas en el aprendizaje.

Este trabajo está enfocado en reconocer automáticamente emociones centradas en el aprendizaje a través del análisis de sentimientos que se aplica a una frase escrita por el estudiante. Para reconocer emociones en texto, fue necesario desarrollar un sistema analizador (clasificador) así como un corpus afectivo que permitiera entrenar al clasificador.

El corpus afectivo desarrollado contiene 9,009 frases en español escritas por estudiantes del área de desarrollo de software y más específicamente programación de computadoras. El clasificador alcanzó una precisión de 60.0% en la predicción. La integración del módulo reconocedor de emociones en texto en herramientas de aprendizaje como Sistemas Tutores Inteligentes o Entornos de Aprendizaje Inteligente ayudará a los estudiantes a tener un proceso de aprendizaje personalizado.

Palabras clave

Corpus de Frases en Español

Minería de Opiniones

Análisis de Sentimientos

Computación Afectiva

Detección de Emociones en Texto

Emociones Centradas en el Aprendizaje

Objetos de Aprendizaje

Sistema Tutor Inteligente

Ambiente de Aprendizaje Inteligente

Índice general

| | | |
|--------|---|----|
| 1. | INTRODUCCIÓN | 1 |
| 1.1. | DESCRIPCIÓN DEL PROBLEMA | 2 |
| 1.2. | JUSTIFICACIÓN | 3 |
| 1.3. | OBJETIVO GENERAL | 3 |
| 1.4. | OBJETIVOS ESPECÍFICOS | 3 |
| 1.5. | HIPÓTESIS | 4 |
| 1.6. | ORGANIZACIÓN DE LA TESIS | 4 |
| 2. | MARCO TEÓRICO | 5 |
| 2.1. | CORPUS DE FRASES EN ESPAÑOL | 5 |
| 2.2. | MINERÍA DE OPINIONES | 5 |
| 2.3. | ANÁLISIS DE SENTIMIENTOS | 6 |
| 2.4. | COMPUTACIÓN AFECTIVA | 6 |
| 2.4.1. | <i>¿Qué es una emoción?</i> | 7 |
| 2.4.2. | <i>Clasificación de la emoción</i> | 7 |
| 2.4.3. | <i>Técnicas de reconocimiento de emociones</i> | 8 |
| 2.5. | DETECCIÓN DE EMOCIONES EN TEXTO | 10 |
| 2.6. | EMOCIONES CENTRADAS EN EL APRENDIZAJE | 11 |
| 2.7. | OBJETOS DE APRENDIZAJE | 11 |
| 2.8. | SISTEMA TUTOR INTELIGENTE | 12 |
| 2.9. | ENTORNO DE APRENDIZAJE INTELIGENTE | 14 |
| 3. | ESTADO DEL ARTE | 15 |
| 3.1. | CORPUS DE TEXTO | 15 |
| 3.1.1. | <i>Corpus Productos de Amazon</i> | 15 |
| 3.1.2. | <i>Corpus TASS 2012</i> | 16 |
| 3.1.3. | <i>Corpus Muchocine</i> | 16 |
| 3.1.4. | <i>Corpus de críticas de cine</i> | 18 |
| 3.1.5. | <i>Corpus de valoraciones de usuarios de hotel</i> | 18 |
| 3.1.6. | <i>Minería de opinión en otros idiomas</i> | 19 |
| 3.2. | SISTEMAS RECONOCEDORES DE EMOCIONES EN TEXTO | 19 |
| 3.2.1. | <i>Facebook</i> | 20 |
| 3.2.2. | <i>Twitter</i> | 20 |
| 3.2.3. | <i>Altrabsheh, Cocea, and Fallahkhair (Altrabsheh N. M., 2014)</i> | 21 |
| 3.2.4. | <i>Munezero, Montero, Mozgovoy, & Sutinen (Munezero, 2013):</i> | 21 |
| 3.2.5. | <i>Tabla Comparativa</i> | 21 |
| 3.3. | CLASIFICADOR EVO DAG | 22 |
| 3.4. | PLATAFORMAS RECOLECTORAS DE OPINIONES EN TEXTO. | 22 |
| 3.4.1. | <i>Plataformas turísticas</i> | 23 |
| 3.4.2. | <i>Plataformas de entretenimiento</i> | 26 |
| 3.4.3. | <i>Plataformas de venta</i> | 27 |
| 3.4.4. | <i>Otras plataformas</i> | 28 |
| 4. | DESARROLLO DEL PROYECTO | 29 |
| 4.1. | METODOLOGÍA PARA LA CREACIÓN DE UN CORPUS | 29 |
| 4.2. | HERRAMIENTAS PARA LA OBTENCIÓN DE FRASES | 33 |

| | | |
|----------|--|----|
| 4.2.1. | <i>Sistema de Evaluación de Recursos Educativos (SERE)</i> | 33 |
| 4.2.1.1. | Metodología de desarrollo | 34 |
| 4.2.1.2. | Análisis del Sistema | 35 |
| 4.2.1.3. | Diseño | 41 |
| 4.2.1.4. | Implementación | 46 |
| 4.2.2. | <i>Obtención de frases desde los Cursos en línea</i> | 52 |
| 4.2.3. | <i>Obtención de frases desde Twitter</i> | 57 |
| 4.3. | MÓDULO DE ANÁLISIS DE SENTIMIENTOS. | 60 |
| 4.4. | SERVICIO. | 62 |
| 4.4.1 | <i>Metodología de desarrollo de SentiText</i> | 62 |
| 4.4.2 | <i>Análisis del Sistema</i> | 63 |
| 4.4.2.1. | Requisitos Funcionales y de Calidad | 63 |
| 4.4.2.2. | Restricciones | 65 |
| 4.4.2.3. | Actores | 65 |
| 4.4.2.4. | Casos de uso | 65 |
| 4.4.3. | <i>Diseño</i> | 67 |
| 4.4.3.1. | Diagrama de contexto | 67 |
| 4.4.3.2. | Arquetipos | 68 |
| 4.4.3.3. | Diferentes vistas del sistema | 69 |
| 4.4.4 | <i>Implementación</i> | 72 |
| 4.4.4.1. | Lenguajes y herramientas software | 72 |
| 4.4.4.2. | Interfaces del sistema | 72 |
| 5. | RESULTADOS | 75 |
| 5.1. | RESULTADOS DE LAS HERRAMIENTAS PARA LA OBTENCIÓN DE FRASES | 75 |
| 5.1.1. | <i>Sistema de Evaluación de Recursos Educativos (SERE).</i> | 75 |
| 5.1.2. | <i>Proceso manual</i> | 77 |
| 5.1.3. | <i>Extractor de información</i> | 79 |
| 5.2. | EVOLUCIÓN DEL CORPUS | 80 |
| 5.3. | EVALUACIÓN DEL MÓDULO DE ANÁLISIS DE SENTIMIENTO | 81 |
| 5.3.1. | <i>Corpus de frases etiquetadas con polaridad</i> | 81 |
| 5.3.2. | <i>Corpus de frases etiquetadas con emociones centradas en el aprendizaje.</i> | 82 |
| 5.4. | EVALUACIÓN DEL MÓDULO DE ANÁLISIS DE SENTIMIENTOS EvoDAG | 83 |
| 6. | CONCLUSIONES Y TRABAJO FUTURO | 85 |
| 6.1. | CONCLUSIONES | 86 |
| 6.2. | TRABAJO FUTURO | 88 |
| | BIBLIOGRAFÍA | 90 |
| | ANEXO | 94 |

Índice de Figuras

| | |
|--|----|
| FIGURA 2-1 NIVELES DE CLASIFICACIÓN EN AS | 6 |
| FIGURA 2-2 DIMENSIONES DE VALENCIA E INTENSIDAD SEGÚN PLUTCHIK (PLUTCHIK, 1980) | 8 |
| FIGURA 2-3 TÉCNICAS DE RECONOCIMIENTO DE EMOCIONES | 10 |
| FIGURA 2-4 EMOCIONES CENTRADAS EN EL APRENDIZAJE | 11 |
| FIGURA 2-5 DISCIPLINAS INVOLUCRADAS EN UN STI (WOOLF, 2009) | 12 |
| FIGURA 2-6 COMPONENTES BÁSICOS DE UN STI | 13 |
| FIGURA 3-1 CLASIFICACIONES DEL CORPUS TASS 2012 | 16 |
| FIGURA 3-2 CLASIFICACIÓN DEL CORPUS MUCHOCINE | 17 |
| FIGURA 3-3 CLASIFICACIÓN DEL CORPUS DE CRÍTICAS DE CINE | 18 |
| FIGURA 3-4 CLASIFICACIÓN DEL CORPUS HOTELES | 19 |
| FIGURA 3-5 PLATAFORMA TURÍSTICA HOTELES.COM | 23 |
| FIGURA 3-6 CATEGORÍAS DE HOTELES.COM | 24 |
| FIGURA 3-7 PLATAFORMA TURÍSTICA BESTDAY | 25 |
| FIGURA 3-8 PLATAFORMA VOLARIS | 26 |
| FIGURA 3-9 PLATAFORMAS DE ENTRETENIMIENTO | 26 |
| FIGURA 3-10 PLATAFORMA DE VENTAS AMAZON. | 27 |
| FIGURA 3-11 OTRAS PLATAFORMAS DE VENTAS. | 27 |
| FIGURA 4-1 METODOLOGÍA PARA LA CREACIÓN DE UN CORPUS DE FRASES. | 29 |
| FIGURA 4-2 CRÍTICAS EN LAS DIFERENTES PLATAFORMAS EDUCATIVAS EN LÍNEA. | 32 |
| FIGURA 4-3 CÓDIGO DE EXTRACCIÓN DE FRASES EN TWITTER. | 32 |
| FIGURA 4-4 FUNCIONAMIENTO DEL SISTEMA DE EVALUACIÓN DE RECURSOS EDUCATIVOS | 34 |
| FIGURA 4-5 METODOLOGÍA DE DESARROLLO DE SERE | 35 |
| FIGURA 4-6 ACTORES EN SERE | 38 |
| FIGURA 4-7 DIAGRAMA DE CASOS DE USO SERE | 38 |
| FIGURA 4-8 DIAGRAMA DE CONTEXTO SERE | 42 |
| FIGURA 4-9 ARQUETIPOS DE SERE | 43 |
| FIGURA 4-10 ARQUITECTURA CLIENTE-SERVIDOR SERE | 44 |
| FIGURA 4-11 VISTA LÓGICA DEL SISTEMA SERE | 45 |
| FIGURA 4-12 VISTA DE DESARROLLO SERE | 46 |
| FIGURA 4-13 PLATAFORMAS DE SERE | 46 |
| FIGURA 4-14 INTERFAZ PRINCIPAL DE SERE | 47 |
| FIGURA 4-15 INTERFAZ DE CURSOS SERE | 47 |
| FIGURA 4-16 INTERFAZ DEL TEMA FUNDAMENTOS DE PROGRAMACIÓN SERE | 48 |
| FIGURA 4-17 INTERFAZ CON EL CONTENIDO DEL TEMA EN FORMATO IMAGEN-TEXTO | 49 |
| FIGURA 4-18 INTERFAZ CON EL CONTENIDO DEL TEMA EN FORMATO VIDEO | 50 |
| FIGURA 4-19 EJERCICIOS DE PROGRAMACIÓN EN SERE | 51 |
| FIGURA 4-20 INTERFAZ DE UN EJERCICIO DE PROGRAMACIÓN | 51 |
| FIGURA 4-21 INTERFAZ DE CONSULTA | 52 |
| FIGURA 4-22 PLATAFORMA EDUCATIVA UDEMY | 53 |
| FIGURA 4-23 CATEGORÍAS DE UDEMY | 53 |
| FIGURA 4-24 RESEÑAS DE UN CURSO DE UDEMY | 54 |
| FIGURA 4-25 PLATAFORMA EDUCATIVA COURSERA | 55 |
| FIGURA 4-26 CATEGORÍA CIENCIAS DE LA COMPUTACIÓN - COURSERA | 56 |
| FIGURA 4-27 PLATAFORMA EDUCATIVA TEACHLR | 56 |
| FIGURA 4-28 INTERFAZ INICIAL PARA DESARROLLAR DE TWITTER | 57 |
| FIGURA 4-29 CREACIÓN DE UNA APLICACIÓN EN TWITTER | 58 |

| | |
|---|----|
| FIGURA 4-30 ACCESOS PARA TWITTER | 59 |
| FIGURA 4-31 ALGORITMO DEL MÓDULO DE ANÁLISIS DE SENTIMIENTOS..... | 61 |
| FIGURA 4-32 FUNCIONAMIENTO DE SENTITEXT..... | 62 |
| FIGURA 4-33 METODOLOGÍA DE CASCADA PARA EL DESARROLLO DE SENTITEXT | 63 |
| FIGURA 4-34 ACTORES EN SENTITEXT | 65 |
| FIGURA 4-35 DIAGRAMA DE CASOS DE USO DE SENTITEXT..... | 66 |
| FIGURA 4-36 DIAGRAMA DE CONTEXTO SENTITEXT | 68 |
| FIGURA 4-37 ARQUETIPOS DE SENTITEXT | 69 |
| FIGURA 4-38 ARQUITECTURA MVC SENTITEXT..... | 70 |
| FIGURA 4-39 VISTA LÓGICA DEL SISTEMA SENTITEXT | 70 |
| FIGURA 4-40 VISTA DE DESARROLLO DE SENTITEXT..... | 71 |
| FIGURA 4-41 LENGUAJES DE SENTITEXT..... | 72 |
| FIGURA 4-42 INTERFAZ PRINCIPAL SENTITEXT | 73 |
| FIGURA 4-43 INTERFAZ DE ANALIZADOR DE SENTIMIENTOS SENTITEXT | 73 |
| FIGURA 4-44 INTERFAZ DE INFORMACIÓN DE CORPUS SENTITEXT | 74 |
| FIGURA 5-1 RESULTADO DE LA OBTENCIÓN DE OPINIONES UTILIZANDO LOS DIFERENTES INSTRUMENTOS | 75 |
| FIGURA 5-2 OBTENCIÓN DE OPINIONES EN SERE..... | 76 |
| FIGURA 5-3 EVOLUCIÓN DEL CORPUS CON EMOCIONES CENTRADAS EN EL APRENDIZAJE. | 80 |
| FIGURA 5-4 DISTRIBUCIÓN DEL CORPUS POR EMOCIONES. | 81 |
| FIGURA 5-5 DISTRIBUCIÓN DEL CORPUS POR POLARIDAD. | 82 |
| FIGURA 5-6 FORMATO DE LAS ORACIONES DE LOS CLASIFICADORES TRADICIONALES. | 84 |
| FIGURA 5-7 FORMATO DE LAS ORACIONES UTILIZANDO EVODAG..... | 84 |

Índice de Tablas

| | |
|--|----|
| TABLA 3-1 COMPARACIÓN DE ANALIZADORES DE SENTIMIENTOS..... | 21 |
| TABLA 4-1 REQUERIMIENTOS FUNCIONALES SERE..... | 36 |
| TABLA 4-2 REQUERIMIENTOS DE CALIDAD DE SERE..... | 37 |
| TABLA 4-3 ACTORES EN SERE..... | 37 |
| TABLA 4-4 ENTIDADES EXTERNAS DE SERE | 42 |
| TABLA 4-5 ARQUETIPOS DEL SISTEMA SERE | 43 |
| TABLA 4-6 REQUERIMIENTOS FUNCIONALES SENTITEXT | 64 |
| TABLA 4-7 REQUERIMIENTOS DE CALIDAD SENTITEXT | 64 |
| TABLA 4-8 ACTORES EN SENTITEXT..... | 65 |
| TABLA 4-9 ENTIDADES DE SENTITEXT | 68 |
| TABLA 4-10 ARQUETIPOS DEL SISTEMA SENTITEXT..... | 69 |
| TABLA 5-1 FRASES RECOLECTADAS CON EL SISTEMA SERE..... | 77 |
| TABLA 5-2 FRASES EXTRAÍDAS DE DIFERENTES PLATAFORMAS EDUCATIVAS. | 77 |
| TABLA 5-3 FRASES PROPORCIONADAS CON EL DESARROLLO DE PROYECTOS. | 78 |
| TABLA 5-4 FRASES EXTRAÍDAS DE TWITTER. | 79 |
| TABLA 5-5 VALORES OBTENIDOS CON EL CLASIFICADOR UTILIZANDO EL CORPUS DE POLARIDAD..... | 82 |
| TABLA 5-6 VALORES OBTENIDOS CON EL CLASIFICADOR UTILIZANDO EL CORPUS DE EMOCIONES | 83 |

Capítulo 1

1. Introducción

Actualmente es indispensable el uso de las computadoras en la vida del ser humano esto debido a la gran evolución de la tecnología que ha invadido muchas actividades de sus vidas cotidianas. Además, con la expansión de las redes sociales donde los usuarios expresan sus opiniones se generó la necesidad de procesar automáticamente estas opiniones para obtener información relevante que pudiera ser utilizada para la toma de decisiones, surgiendo el área llamada Minería de Opiniones.

La minería de opiniones es importante porque tiene aplicaciones en diferentes áreas del conocimiento. Por ejemplo, en el ámbito educativo tradicionalmente los estudiantes interactúan con sus compañeros para intercambiar experiencias y colaborar en el desarrollo de proyectos o la resolución de problemas con el fin de aprender diversos temas o adquirir competencias. Este intercambio de opiniones y experiencias permite a los estudiantes colaborar en el desarrollo de soluciones de problemas y proyectos. Sin embargo, los estudiantes no siempre tienen el mismo nivel de conocimientos y a veces, su personalidad no les permite expresar sus ideas libremente. Actualmente con ayuda de la tecnología se han desarrollado diferentes herramientas como Sistemas Tutores Inteligentes (STI) y Entornos de Aprendizaje Inteligentes (EAI) en las cuales, con ayuda de la minería de opiniones será posible realizar el análisis de sentimientos y su resultado podrá ser utilizado para mejorar los materiales del curso, las estrategias de enseñanza y la personalización de contenidos en un STI, entre otras acciones.

El reconocimiento de emociones puede efectuarse utilizando diferentes dispositivos que permiten captar información (imágenes del rostro, frases textuales, voz, etc.) o señales biológicas (señales cerebrales, ritmo cardíaco, postura del cuerpo, etc.) de los usuarios con el fin de analizarlas y determinar su estado emocional. El análisis de frases de texto escritas por los usuarios ayuda a reconocer tendencias sobre los deseos de los usuarios, por ejemplo: las opiniones sobre una película de cine pueden ayudar a pronosticar la cantidad de personas que asistirán a las salas de cine a ver la función. Actualmente, la mayoría de las herramientas para el reconocimiento de emociones en el texto se enfocan en determinar la polaridad

(positiva o negativa) de las opiniones y trabajan con frases escritas en inglés (Ortigosa, 2014); además, las frases tratan sobre diferentes temas generales y no están enfocadas al área de aprendizaje.

Los STI o EAI generalmente adaptan su modelo de enseñanza a las necesidades cognitivas del estudiante, sin embargo, las emociones juegan un papel fundamental en el proceso de aprendizaje por lo que resulta necesario que los sistemas inteligentes para el aprendizaje adapten también su modelo de enseñanza a las necesidades afectivas del usuario. Para lograr esto, se requiere que los sistemas reconocedores de emociones sean capaces de reconocer, usando diferentes tipos de información o señales, emociones centradas en el aprendizaje dentro de un entorno educativo, con el objetivo de ayudar a los sistemas tutores inteligentes a personalizar la enseñanza a cada estudiante.

A continuación, se presenta la descripción del problema, una breve justificación sobre la importancia de este proyecto, se define el objetivo general, se establece la hipótesis; y finaliza con la organización del documento de tesis.

1.1. Descripción del problema

En el laboratorio de *Tecnologías Inteligentes Aplicadas a la Educación* del Instituto Tecnológico de Culiacán, se han desarrollado diversos sistemas tutores inteligentes y entornos inteligentes de aprendizaje como son: JavaSensei (González Hernández, 2015), FindErrorJava (Valencia Rodríguez, 2017), Aprende Java con Lucy (Sosa Ochoa, 2016), EasyLogic (Ríos Félix, 2016), CodeTraning (Lindor Valdez, 2016), entre otros, enfocados a la enseñanza del tema de desarrollo de programas con lenguaje Java. Algunos de estos sistemas incluyen el reconocimiento de emociones de los estudiantes a través de imágenes del rostro captadas por la cámara en tiempo real cuando el usuario utiliza el sistema. Sin embargo, no es posible detectar emociones en el texto que el usuario escribe para interactuar con los sistemas, debido a que no existe un reconocedor de emociones que reciba una frase y determine el estado emocional del estudiante.

Desarrollar este reconocedor, permitirá que los sistemas educativos mencionados sean capaces de reconocer emociones utilizando frases en español y no solo las imágenes del

rostro del usuario. Los sistemas que realizan análisis de sentimientos utilizan corpus que contienen grandes volúmenes de información (miles o millones de frases) que son necesarias para realizar el proceso de entrenamiento del sistema clasificador que determina la polaridad (positiva o negativa) del texto. El no contar con un corpus de frases en español inhibe el entrenamiento de un reconocedor de emociones usando frases textuales, la principal contribución de este trabajo es generar el corpus de frases para efectuar el análisis de sentimientos en entornos de aprendizaje.

1.2. Justificación

La creación de un corpus de frases en español enfocadas al aprendizaje contribuirá para que los sistemas tutores inteligentes detecten emociones a través del texto y realicen de forma más eficientes la adaptación del proceso de enseñanza con los estudiantes.

1.3. Objetivo general

Generar un corpus en español de las opiniones en texto relacionadas con el aprendizaje del tema de programación de computadoras utilizando diferentes sistemas, herramientas o plataformas. Posteriormente se utilizará este corpus para el entrenamiento de un clasificador de emociones en texto.

1.4. Objetivos específicos

- Diseñar y desarrollar un sistema que capture las opiniones de los estudiantes sobre diferentes objetos de aprendizaje.
- Generar un corpus de frases en español utilizando el sistema de captura de opiniones.
- Analizar y etiquetar las opiniones del corpus de frases relacionadas con las emociones centradas en el aprendizaje.
- Adaptar el reconocedor de emociones en texto para reconocer emociones enfocadas al aprendizaje.
- Evaluar el sistema reconocedor de emociones mediante diferentes algoritmos utilizando la técnica de validación cruzada para obtener una precisión de reconocimiento.

- Implementar un servicio web para utilizar en tiempo real el sistema reconocedor de emociones.

1.5. Hipótesis

En este trabajo se definió una hipótesis el cual es enfocada al uso del corpus en los clasificadores: El corpus de emociones centradas en el aprendizaje podrá ser usado para entrenar clasificadores de emociones obteniendo tasas de asertividad mayores de 70 % de éxito.

1.6. Organización de la tesis

En este documento de tesis, se describen los diferentes procesos ejecutados para la generación de un corpus de frases escritas en español las cuales están enfocadas al aprendizaje, además se describe el desarrollo de un módulo de análisis de sentimiento y un servicio web. Este documento se organiza en 5 capítulos principales, que se describen a continuación.

El capítulo 2 presenta el marco teórico, es decir, las bases teóricas utilizadas para diseñar y construir tanto el corpus de frases como el sistema reconocedor de emociones centradas en el aprendizaje.

En el capítulo 3, se expone el estado del arte de trabajos relacionados a este tema de investigación.

En el capítulo 4, se detalla el desarrollo del sistema que se utilizó para la construcción del corpus, del sistema reconocedor de emociones y del servicio web utilizado realizar un análisis de sentimiento en tiempo real.

El capítulo 5, muestra la evolución del corpus y su distribución de acuerdo con la polaridad y la emoción detectada. Además, se presentan los resultados obtenidos con los diferentes clasificadores utilizados.

Finalmente, en el capítulo 6 se presentan las conclusiones y trabajos futuros.

Capítulo 2

2. Marco teórico

En este capítulo se presenta una serie de elementos conceptuales que son fundamentales en el presente trabajo de tesis para su comprensión.

2.1. Corpus de frases en español

Un corpus es conocido también como corpora (latín) o base de datos; estos son definidos como un «conjunto de datos o textos de un mismo tipo que sirve de base a una investigación» (Real Academia Española, 2005). Para ser más concretos, un corpus es una colección de datos, los cuales pueden estar en diferentes formatos (texto, imágenes o audios, etc.) y representan una parte fundamental para el desarrollo de aplicaciones de procesamiento y reconocimiento de datos (Atkins, 1992).

2.2. Minería de opiniones

La minería de opiniones (MO) ha mostrado una alta tendencia en la investigación en los últimos años debido a la obtención de grandes escalas de opiniones y comentarios por parte de usuarios activos en Internet. Actualmente muchas de las empresas y organizaciones están interesadas en conocer las opiniones de sus trabajadores en las redes sociales para tomar decisiones con base a la retroalimentación. Adicionalmente a esto, la gran mayoría de trabajos de investigación involucran sistemas de MO en el idioma inglés (Vilares, 2013). Por este motivo, la comunidad científica está interesada en trabajos diferentes a este idioma.

Para la construcción de un sistema de MO se deben tener en cuenta varios aspectos. Primero la extracción de la opinión y luego la clasificación del sentimiento. Para la extracción se elige un conjunto de datos, que van desde redes sociales, hasta sitios web donde abundan opiniones y comentarios en línea (Henríquez, 2015). Para la clasificación, normalmente positiva o negativa, se utilizan en su gran mayoría técnicas basadas en aprendizaje de máquinas (ML

por sus siglas en inglés) y basadas en léxico (LEX). Las diferencias fundamentales radican en que la primera utiliza algoritmos para aprender y la segunda utiliza diccionarios y léxicos que ya vienen catalogados con su sentimiento.

2.3. Análisis de sentimientos

Análisis de sentimiento (AS) es el estudio computacional de las opiniones, actitudes y emociones de las personas hacia una entidad (Medhat, 2014). El AS identifica el sentimiento expresado en un texto y luego lo analiza. Por lo tanto, el objetivo de AS es encontrar opiniones, identificar los sentimientos que expresan y luego clasificar. En la Figura 2-1 se muestran, de acuerdo a Medhat, los tres niveles principales de clasificación en AS: nivel de documento, nivel de oración y nivel de aspecto que propone (Medhat, 2014)

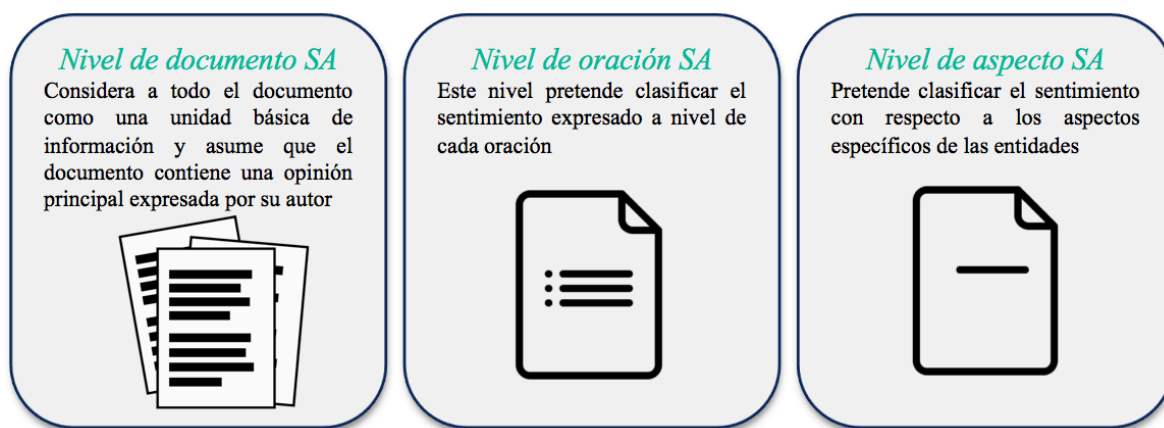


Figura 2-1 Niveles de clasificación en AS

El análisis de sentimiento combina técnicas de minería de textos y procesamiento de lenguaje natural (PLN), incluye una gran cantidad de tareas que han sido tratadas en mayor o menor medida (Pang B. y., 2008). Actualmente la mayoría de las investigaciones que realizan análisis de sentimiento utilizan textos escritos en inglés y se encargan de determinar la polaridad, es decir si el documento, oración o aspecto es positivo o negativo.

2.4. Computación Afectiva

La computación afectiva es el estudio de técnicas como reconocimiento, expresión y generación de emociones para que la computadora entienda y procese las emociones como

un ser humano (Picard, 2000). El objetivo de la computación afectiva es capturar la información afectiva y analizarla para mejorar la interacción entre las computadoras y los usuarios.

2.4.1. ¿Qué es una emoción?

El término emoción (Gross M., 2018) es más que la respuesta que damos ante un estímulo o evento determinado, como por ejemplo un recuerdo o una idea. Las emociones sencillamente ocurren, no es voluntario el acto de generarlas. Sin embargo, puede suceder que, a diferencia de una reacción, nos encontremos ante una predisposición para actuar, en cuyo caso decimos que estamos ante un "estado emocional" o más comúnmente, un "estado de ánimo".

Las emociones (Caruso D., 2005) tienen la característica de ser específicas y reactivas ante un determinado evento. Es decir, son una respuesta a acontecimiento. Por otra parte, los estados de ánimo son emociones que se instalan con permanencia en el tiempo, no dependen de acontecimiento alguno, sino que diversos acontecimientos pasados, traumas o experiencias, han influido para que la persona tenga una respuesta emocional.

El afecto es un proceso entre dos o más personas. Dar afecto implica realizar un trabajo no remunerado en beneficio de los demás.

2.4.2. Clasificación de la emoción

En la computación afectiva hay distintas formas de percibir emociones, por ejemplo, por tipo, como:

- **Emociones primarias:** son las emociones que se experimentan en respuesta a un estímulo. Paul Ekman (Ekman, 1992) define que las emociones básicas son seis: tristeza, felicidad, sorpresa, asco, miedo e ira. Todas ellas constituyen procesos de adaptación y, en teoría, existen en todos los seres humanos, independientemente de la cultura en la que se hayan desarrollado.
- **Emociones secundarias:** son emociones que aparecen después de las emociones primarias y dependen más de la situación. Por ejemplo: una persona puede sentir

miedo (emoción básica) y después transformarse en amenaza o enfado (emociones secundarias), dependiendo de la situación en que se encuentre.

Robert Plutchik (1980) presenta las emociones en un modelo estructural de las emociones como una analogía de una rueda de colores en donde agrupa las emociones según su similitud (Chóliz, 2005). En la Figura 2-2 se muestra el modelo de dimensiones de valencia e intensidad para describir una emoción con mayor precisión.

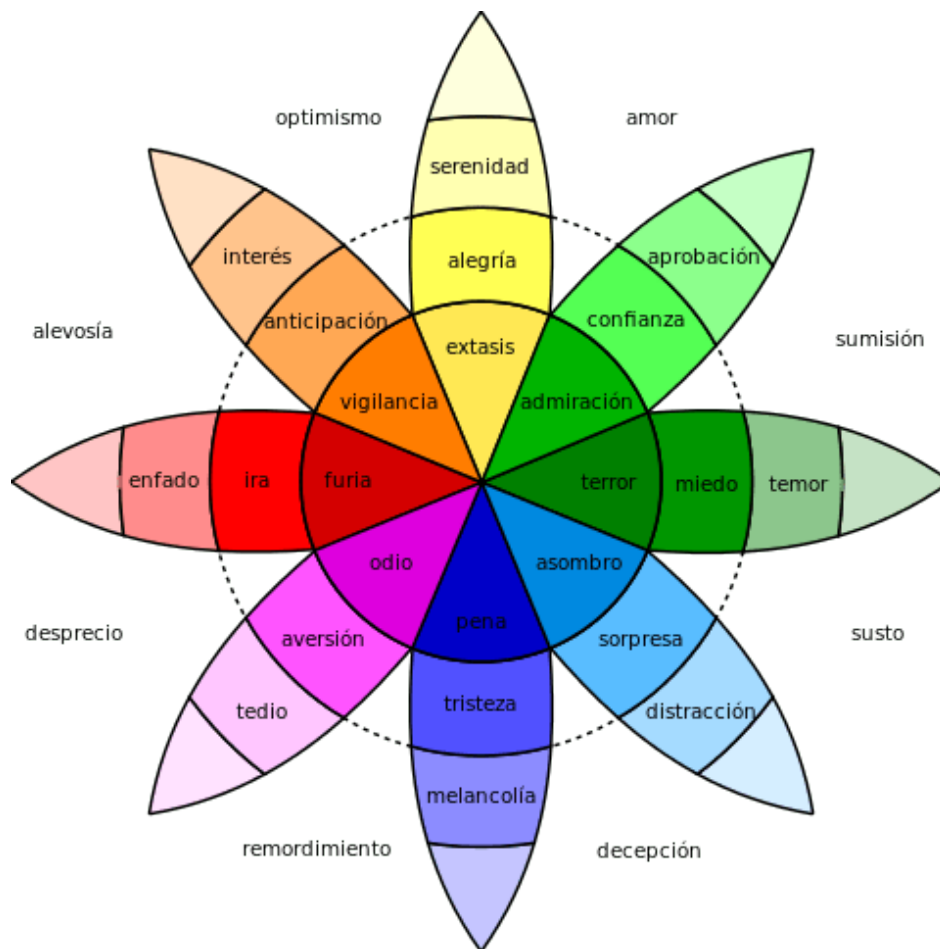


Figura 2-2 Dimensiones de valencia e intensidad según Plutchik (Plutchik, 1980)

2.4.3. Técnicas de reconocimiento de emociones

En la computación afectiva, una computadora debe ser capaz de reconocer emociones, expresar emociones, sentir emociones y poseer inteligencia emocional. Se han desarrollado diferentes técnicas para el reconocimiento del estado afectivo de una persona. Algunas técnicas toman como base señales biológicas producidas por el usuario (Arroyo, 2009),

mientras que otras se basan en información del contexto tal como el lenguaje corporal (Conati, 2009) o por medio de expresiones faciales tales como el movimiento de los ojos y los gestos musculares del rostro (D'Mello S. P., 2007).

Las técnicas que se utilizan para el reconocimiento de emociones son (Tao, 2005):

- **Procesamiento emocional del habla:** toman en cuenta el tono y velocidad de la voz.
- **Expresiones faciales:** utiliza el rostro humano y en su mayoría las técnicas descritas por (Ekman, 1992), que clasifica seis emociones básicas: tristeza, felicidad, sorpresa, asco, miedo e ira. Existen otros mecanismos que consideran otros estados afectivos no básicos, tales como la fatiga o dolor y estados mentales como concentración, desacuerdo, interés, frustración o inseguridad, que utilizan muestras de expresiones faciales deliberadas o actuadas (D'Mello S. K., 2010).
- **Gestos y movimiento del cuerpo:** considera la posición del cuerpo y sus cambios en el tiempo. En la actualidad, los trabajos de investigación se concentran en el seguimiento de la mano. Los movimientos gestuales con las manos transmiten varios y diversos significados, para mejorar el estado de ánimo o para comportarse como un lenguaje simbólico.
- **Señales cerebrales:** Esta técnica tiene sus fundamentos en la Electroencefalografía (EGG), que mide la actividad eléctrica cerebral. Considera la medición de la actividad cerebral para procesarla y obtener las características de interés de la señal y una vez obtenidas, utilizarlas para interactuar con el entorno.
- **Sistemas multimodales:** es una técnica reciente que integra las técnicas mencionadas anteriormente

En la Figura 2-3 se muestran diferentes técnicas que se utilizan para reconocer emociones, sin embargo, en este proyecto se manejó el reconocimiento de emociones en texto.



Figura 2-3 Técnicas de reconocimiento de emociones

2.5. Detección de emociones en texto

El análisis de sentimiento en texto es el proceso que establece la expresión emocional que hay atrás de unas palabras determinadas, si una frase contiene una opinión positiva o negativa sobre un producto, marca, institución, organización, empresa, evento o persona.

Existen diferentes métodos para determinar la emoción de un texto como: detección de palabras clave, afinidad léxica, procesamiento estadístico del lenguaje natural, entre otros.

- **Detección de palabras clave.** Este método identifica de palabras clave. el texto es marcado con emociones basándose en la presencia de palabras afectivas como feliz, triste o furioso. Ejemplos de este método son el Elliott's Affective Reasoner (Elliot, 1992) o el Affective Lexicon (Ortony, 1988).
- **Afinidad léxica.** Esta técnica no solo detecta las palabras afectivas, sino que asigna a palabras arbitrarias la probabilidad de ser afines a una determinada emoción. Estas probabilidades normalmente se obtienen de un corpus. Un ejemplo de este tipo de sistemas es el sistema de clasificación de críticas cinematográficas de Pang y Vaithyanathan (Pang B. L., 2002).
- **Procesamiento estadístico del lenguaje natural.** Este método consiste en alimentar un algoritmo de aprendizaje automático, con un amplio corpus de textos marcados con emociones.

2.6. Emociones centradas en el aprendizaje

Las emociones centradas en el aprendizaje son emociones secundarias que emergen durante actividades de aprendizaje profundo (Baker, 2010) (D'Mello S. K., 2010). Entre las emociones centradas en el aprendizaje se encuentran: frustrado, aburrido, comprometido o emocionado como se muestra en la Figura 2-4.



Figura 2-4 Emociones centradas en el aprendizaje

Las emociones centradas en el aprendizaje juegan un papel importante en los estudiantes, ya que afectan diferentes aspectos como mecanismos cognitivos y retención de información. (Pekrun R. , 1992) (Pekrun R. G., 2002).

2.7. Objetos de Aprendizaje

Un objeto de aprendizaje es cualquier elemento que se use en un ambiente educativo, sin embargo, en el ámbito del aprendizaje electrónico (e-learning) los objetos de aprendizaje son elementos electrónicos que cumplen ciertas características.

"Un Objeto de Aprendizaje es un conjunto de recursos digitales y reutilizable, con un propósito educativo y constituido por al menos tres componentes internos: contenidos, actividades de aprendizaje y elementos de contextualización. El Objeto de Aprendizaje debe tener una estructura de información externa (metadatos) que facilite su almacenamiento, identificación y recuperación" (Ministerio de Educación Nacional Colombiano, 2006).

2.8. Sistema Tutor Inteligente

Los Sistemas Tutores Inteligentes (STI) son programas de computadora que se implementan para simular el comportamiento de un tutor humano (Nwana, 1990). Un STI es similar a un profesor personal para el estudiante ya que es aquel que ofrece un proceso de aprendizaje personalizado, es decir, genera ejercicios, así como la explicación de la solución dependiendo de las necesidades de cada estudiante.

En la Figura 2-5 se presenta la intersección de las disciplinas que interactúan en un STI tales como ciencias computacionales, psicología y educación (Woolf, 2009).

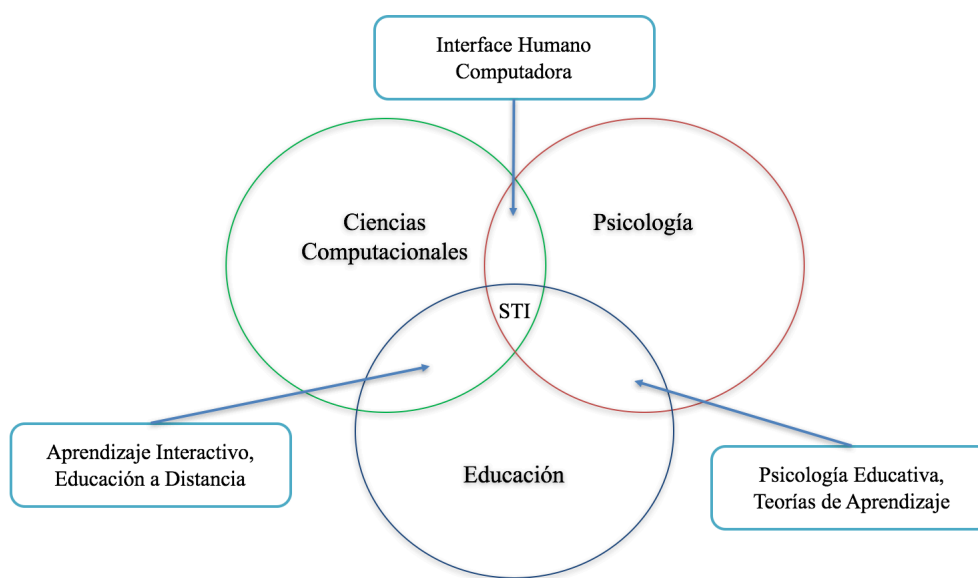


Figura 2-5 Disciplinas involucradas en un STI (Woolf, 2009)

Para finalizar en la Figura 2-6 se muestran los componentes básicos de un sistema tutor inteligente, el cual se compone por el modelo de dominio, el modelo pedagógico, el modelo de estudiante y la interfaz de usuario (Self, 1998) (Woolf, 2009).

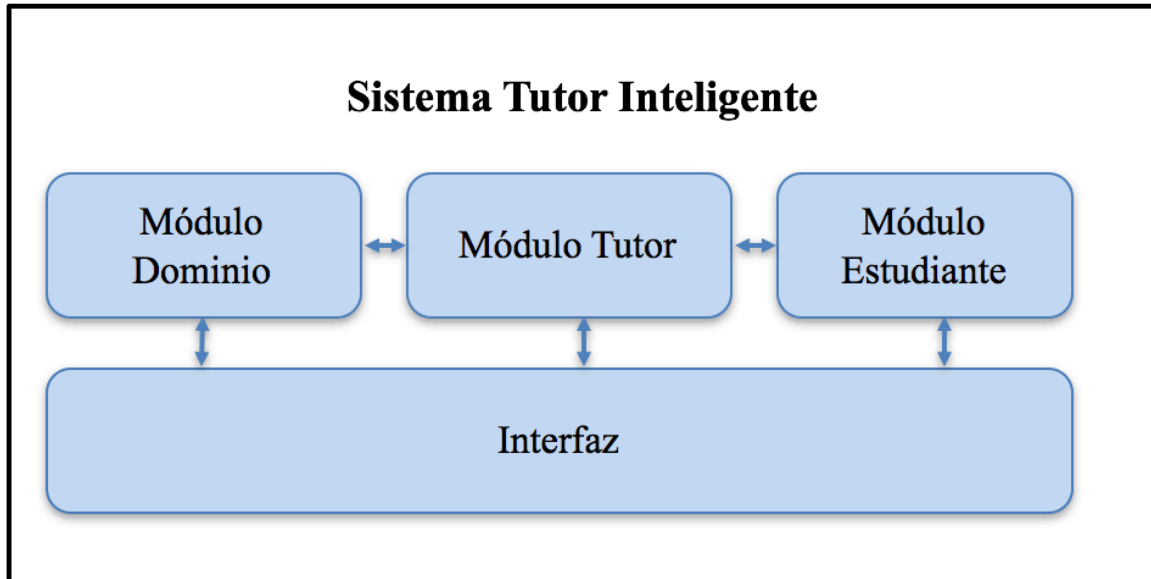


Figura 2-6 Componentes básicos de un STI

- **Módulo del Dominio:** es el área de estudio de uno o más expertos y este representa los conocimientos, las experiencias, las técnicas y las metodologías que son empleadas y perfeccionadas por estos expertos. A partir del dominio se elabora una representación computacionalmente viable, la cual es usada por el STI como objeto de enseñanza y es algo que el estudiante aspira a aprender. Así, el STI por medio de guías y supervisiones instruye al estudiante durante el proceso de aprendizaje (Woolf, 2009).
- **Módulo del Estudiante:** comprende el comportamiento de los estudiantes al utilizar el STI y para ello maneja al igual que el Módulo de Dominio una representación computacional de ese conocimiento. El módulo contiene información que proviene de distintas fuentes que el estudiante genera, como: respuestas, acciones en la interfaz gráfica, resultados de las acciones realizadas y resultados intermedios por mencionar algunos ejemplos. En esencia, el objetivo del módulo del estudiante es informar de varios aspectos como sus áreas de interés y sus estilos de aprendizaje preferidos (Greer, 1994) (Woolf, 2009).
- **Módulo del Tutor:** representa a un tutor humano y éste cuenta con dos funciones: la primera es representar el conocimiento sobre cómo realizar tutorías, tales como los

modelos de enseñanza. La segunda es la de proporcionar respuestas individualizadas durante su interacción con los estudiantes. El conocimiento del módulo del tutor es la representación de las estrategias de enseñanza, las intervenciones, las respuestas personalizadas y las técnicas de motivación pertinentes durante el proceso de enseñanza del estudiante. (Woolf, 2009). El módulo del tutor considera varios aspectos en la enseñanza, como son: tareas de aprendizaje, metas de aprendizaje, errores y obstáculos. Además de estos aspectos, el módulo del tutor considera otros parámetros que afecten al aprendizaje como: estados afectivos, prioridades del aprendizaje, prioridades del conocimiento y habilidades del estudiante.

2.9. Entorno de Aprendizaje Inteligente

Un Entorno de Aprendizaje Inteligente (EAI) es un sistema educacional inteligente el cual incorpora las características de los entornos de aprendizaje electrónicos (EAE) y los STI. A continuación, se definen dos características principales en un EIA:

- **Adaptabilidad Hipermedia:** La adaptabilidad hipermedia son técnicas con las cuales se modifica el contenido de los recursos hipermedia (por ejemplo, videos, texto y enlaces) de manera dinámica, la personalización se ajusta de acuerdo con los gustos y preferencias del usuario, los cuales se obtienen de su interacción con el sistema (Brusilovsky, 1998).
- **Modelo de Usuario:** El modelo del usuario es una representación individual de metas, preferencias, objetivos y conocimientos del estudiante. El modelo del usuario es usado principalmente como recurso para la obtención de preferencias dentro de las técnicas de adaptabilidad hipermedia (Brusilovsky, 2003).

Capítulo 3

3. Estado del arte

En este capítulo se presentan diferentes trabajos e investigaciones relacionados con este trabajo; es dividido en tres apartados el primero presenta los diferentes corpus que utilizan los algoritmos, en el segundo se mencionan algunos sistemas que implementan reconocimiento de emociones en texto, y finalmente se muestran algunas plataformas que recolectan opiniones de los usuarios.

3.1. Corpus de texto

Actualmente los analizadores de emociones en texto utilizan una colección de datos (oraciones o frases) para entrenar sus clasificadores. Sin embargo, la mayoría de estos corpus son frases escritas en inglés u otros idiomas.

En esta sección se presentan diferentes bases de datos de frases textuales, se menciona cómo están compuestas y su distribución, es decir cuantas frases se encuentran en cada etiqueta.

3.1.1. Corpus Productos de Amazon

Este corpus fue construido con la finalidad de realizar pruebas a los algoritmos utilizando documentos en un idioma diferente al inglés, se han recopilado reseñas de películas de Amazon (Denecke, 2008). Este corpus consta de 200 reseñas escritas en idioma alemán, de las cuales 100 son positivas y 100 son negativas.

Una opinión se etiqueta como positiva cuando el revisor de la película asigna cuatro o cinco estrellas y negativa si le asigna una o dos estrellas. Las revisiones con tres estrellas fueron eliminadas del corpus para esta evaluación.

Las reseñas de películas en alemán e inglés a veces son muy cortas y están escritas en lenguaje común, es decir pueden contener errores de escritura.

3.1.2. Corpus TASS 2012

Este corpus contiene más de 70,000 *tweets*, escritos en español por 200 personas como personalidades conocidas y celebridades de la política, la economía, la comunicación, los medios de comunicación y la cultura. Este corpus se desarrolló de noviembre de 2011 a marzo de 2012.

Cada mensaje de *Twitter* incluye un ID, fecha de la creación y la identificación del usuario. El contenido del mensaje real se puede obtener utilizando la API de *Twitter* con el ID.

En la Figura 3-1 se muestran las etiquetas que expresan estos *tweets*. Estas etiquetas se clasifican en sentimiento positivo, negativo, neutral, o ninguno.



Figura 3-1 Clasificaciones del corpus TASS 2012

Este corpus fue dividido en dos conjuntos: el primero es el conjunto de entrenamiento que usa un conjunto de *tweets* junto con las etiquetas correspondientes para que puedan capacitar y validar sus modelos de clasificación para el análisis de sentimientos; y el segundo es el conjunto de prueba el cual solo esta construido por los tweets, es decir, no contiene ninguna etiqueta y se usa para evaluar los resultados proporcionados por los diferentes sistemas (Villena Román, 2013).

3.1.3. Corpus Muchocine

Este corpus consta de críticas de cine escritas en español, está compuesto por 3,878 críticas recogidas de la web muchocine (Cruz, 2008). Las críticas que componen el corpus no están escritas por profesionales, sino por usuarios de la web. Esto puede aumentar la dificultad, ya que los textos pueden estar gramaticalmente incorrectos, es decir pudiendo aparecer faltas de ortografía o expresiones informales.

En la Figura 3-2 se muestra como las críticas están calificadas en un rango de 1 a 5. El 1 significa que la película fue muy mala, y el 5 fue una película muy buena. Las películas registradas con 3 se pueden catalogar como neutras, es decir que no se considera ni malas ni

buenas. Para la realización del estudio no se consideraron las críticas neutrales. Así, las cuatro clases restantes se reducen a dos, positiva o negativa. Las críticas con una valoración inferior a tres se consideran negativas, mientras que las valoradas con un cuatro o cinco, se etiquetan como críticas positivas.

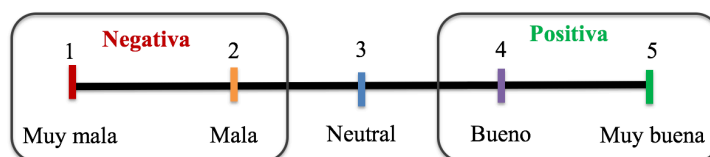


Figura 3-2 Clasificación del Corpus Muchocine

El número total de documentos que se tomaron para realizar los experimentos es de 2,625, de los cuales 1,274 se corresponden con críticas negativas, y 1,351 son críticas positivas. Para cada documento se tiene el nombre del autor que realiza la crítica, el título de la película, la puntuación asignada, la máxima nota que se puede asignar a una película, la fuente de los datos, un breve resumen a modo de titular y la crítica en sí. De todos esos campos, para la experimentación, un factor importante de interés es la puntuación de la película, el resumen de la crítica (Re), y la crítica desarrollada (Cr).

Las críticas que forman el corpus no tienen toda la misma extensión, y además no fueron redactadas por profesionales, por lo que lo más natural es que en muchas críticas la orientación de esta no sea muy clara. En cambio, los resúmenes al estar formados por una frase o en algunos casos dos, expresan con mayor claridad la impresión que a los autores ha causado la película. Debido a esto, se han generado tres corpus: uno formado por los resúmenes (Re), otro solamente por la crítica (Cr) y, por último, el constituido por los resúmenes y las críticas (Re_Cr).

Este corpus tiene el objetivo de abordar el problema desde una perspectiva multilingüe logrando que, independientemente del idioma utilizado, se permita la clasificación de la polaridad en un texto subjetivo.

3.1.4. Corpus de críticas de cine

Este corpus contiene opiniones de usuarios sobre películas cinematográficas (Peñalver Martínez, 2011). El número de opiniones relativas al mundo del cine está en constante crecimiento. Para poder manipular la gran cantidad de datos relacionados con las opiniones sobre películas, existe la creciente necesidad de utilizar herramientas de minería de opiniones que puedan evaluar algunas de las características relevantes de las películas.

Teniendo en cuenta las dificultades de desarrollar una nueva ontología desde cero, utilizaron la ontología sobre películas Movie Ontology. Esta ontología pretende proporcionar un vocabulario controlado para describir semánticamente los conceptos relacionados con las películas, tales como película, género, director o actor. Un total de 100 películas han sido agregadas en la ontología como instancias para los propósitos experimentales.

El corpus del experimento contiene 15,323 palabras y comprende 64 opiniones. En la Figura 3-3 se muestra la etiqueta de la clase origen (positivo y negativo) como línea base. Se encuentra distribuido 31 negativas y 33 positivas.



Figura 3-3 Clasificación del corpus de críticas de cine

3.1.5. Corpus de valoraciones de usuarios de hotel

Tripadvisor es un sitio Web que recolecta información turística acerca de la experiencia sobre el alojamiento de un hotel (Moreno Ortiz, 2010). Su sistema sigue los patrones estándar: el usuario selecciona un hotel, describe su experiencia mediante texto. En la Figura 3-4 se muestra la valoración del 1 al 5 (1 muy negativo y 5 muy positivo).



Figura 3-4 Clasificación del corpus hoteles

Tripadvisor también permite valorar determinados aspectos del hotel tales como limpieza, ubicación, habitaciones, servicio, relación calidad-precio y calidad del sueño.

3.1.6. Minería de opinión en otros idiomas

Existen otros corpus que se han creado recolectando opiniones, comentarios o reseñas acerca de un servicio o producto:

- Zhang y otros (Zhang, 2009) utilizan algoritmos basados en reglas y aprendizaje automático sobre dos corpus en chino. Uno de ellos trata comentarios sobre la eutanasia mientras que el segundo se construyó a partir de los comentarios en chino de seis tipos de productos de Amazon.
- Agić, Ljubešić y Tadić (Agić, 2010) presentan un corpus con textos en croata recopilado de mercados financieros y anotado manualmente.
- Ghorbel y Jacot (Ghorbel, 2010) utilizan aprendizaje automático sobre comentarios de películas en francés. Al igual que el trabajo de Denecke (2008) se traduce el corpus al inglés y se integra con SentiWordNet.
- Por otra parte, también existen algunos trabajos que utilizan un entorno multilingüe. Por ejemplo, Ahmad, Cheng y Almas (Ahmad, 2006) trabajan sobre un corpus con noticias financieras en inglés, chino y árabe.
- Boldrini y otros (Boldrini, 2009) presentan el corpus EmotiBlog que incluye comentarios sobre varios temas en tres idiomas: español, inglés e italiano.

3.2. Sistemas reconocedores de emociones en texto

Actualmente existen diversos medios de comunicación como Facebook, Twitter, Instagram, Snap, entre otros, los cuales ayudan a recolectar opiniones acerca de un tema específico y posteriormente utilizan esta información para saber la tendencia que hay en el momento.

3.2.1. Facebook

Facebook contiene un método de análisis de sentimientos que puede ser utilizado para el aprendizaje electrónico (e-learning). A partir de los mensajes escritos por los usuarios, se realizan las funciones para extraer información sobre la polaridad sentimental del mensaje (positiva, neutra o negativa) y modelar la polaridad del sentimiento habitual en los usuarios para poder detectar cambios emocionales significativos. Esto se ha implementado en SentBuk (Alvaro Ortigosa, 2014), el cual apoya a la detección del cambio emocional, el hallazgo emocional del amigo, la clasificación del usuario de acuerdo con sus mensajes y las estadísticas, entre otros. El método de clasificación implementado en SentBuk sigue un enfoque híbrido, es decir, combina técnicas basadas en el léxico y la máquina. Los resultados obtenidos a través de este enfoque muestran que es factible realizar análisis de sentimiento en Facebook con una precisión de 83.27%.

3.2.2. Twitter

Altrabsheh, Gaber, y Cocea (Altrabsheh N. &, 2013) presentan SA-E, donde exponen una arquitectura para analizar la retroalimentación de los estudiantes usando el análisis sentimental en Twitter. Este trabajo muestra el análisis de la minería de datos y el sentimiento educativo que incluye diferentes métodos para la retroalimentación de los estudiantes. Los estudiantes utilizaron *Twitter* para expresar sus opiniones sobre el material de clase, las opiniones de los estudiantes fueron utilizadas por el profesor para alterar el estilo de enseñanza de acuerdo con los resultados. La arquitectura esta constituido por cuatro pasos y este es implementado en el aula de la siguiente manera:

- Primero los estudiantes proporcionaron comentarios al conferencista a través de las redes sociales como Twitter.
- Después los comentarios de los estudiantes fueron preprocesados y analizados a través de clasificadores Naive Bayes y Support Vector Machine individualmente o combinados. Estas técnicas se escogieron porque se demostró que funcionan bien con revisiones y datos educativos.
- Los resultados se transmitirán a través de una aplicación donde el profesor puede decidir actuar de acuerdo con los resultados.

- Finalmente, después de obtener los resultados, se pueden aplicar métodos de verificación para asegurar que el análisis de sentimiento haya tenido éxito, además que los resultados se guardarán para su posterior estudio y se hará otro uso de los datos analizándolos durante largos períodos de tiempo.

3.2.3. Altrabsheh, Cocea, and Fallahkhair (Altrabsheh N. M., 2014)

Este trabajo realiza adaptación utilizando el conocimiento de los sentimientos de los usuarios. En la enseñanza los sentimientos de los estudiantes los utilizan para abordar problemas como la confusión y el aburrimiento que afectan la participación de los estudiantes.

Además, utilizan diferentes métodos (Naive Bayes, Complement Naive Bayes (CNB), Maximum Entropy y Support Vector Machine (SVM)) para aprender el sentimiento de los comentarios de los estudiantes. Estos métodos se entrenaron utilizando la retroalimentación de los estudiantes. Los dos clasificadores destacados por tener mejor percepción en el aprendizaje son SVM dando como resultado una precisión de 94%, y CNB de 84%.

3.2.4. Munezero, Montero, Mozgovoy, & Sutinen (Munezero, 2013):

Una problemática que se aborda en este trabajo es evitar el análisis manual de todos los diarios creados durante un curso ya que esto puede convertirse en una tarea compleja. Para evitar este problema presentan un sistema funcional para analizar y visualizar las emociones de los estudiantes expresados en los diarios de aprendizaje (instrumentos en los cuales los estudiantes reflexionan sobre su experiencia de aprendizaje). Este sistema permite, de forma automática extraer emociones y cambios de la experiencia de aprendizaje a lo largo del curso. Las emociones extraídas por el sistema se basan en las ocho emociones de Plutchik (Plutchik, 1980): alegría, tristeza, miedo, ira, anticipación, sorpresa, disgusto y confianza.

3.2.5. Tabla Comparativa

Para finalizar en la Tabla 3-1 se presenta una comparación de los diferentes analizadores de sentimientos descritos en esta sección.

Tabla 3-1 Comparación de Analizadores de sentimientos.

| Autor | Trabajo | Características | Tamaño |
|-------|---------|-----------------|--------|
|-------|---------|-----------------|--------|

| | | | |
|--|---|---|--|
| Álvaro Ortigosa | e-learning | Se extraen los mensajes escritos y se determina la polaridad sentimental (positiva, neutra o negativa) | Tiene un enfoque híbrido en el cual combina técnicas basadas en el léxico y la máquina obteniendo una precisión de 83.27% . |
| Altrabsheh, Gaber, y Cocea | SA-E | Expone una arquitectura para la obtención de mensajes utilizando la red social twitter. | Naive Bayes y Support Vector Machine individualmente o combinados. |
| Altrabsheh, Cocea, and Fallahkhair | Learning Sentiment from Students' Feedback for Real-Time Interventions in Classrooms | Realiza procesos de adaptación utilizando los sentimientos de los usuarios. | Naive Bayes, Complement Naive Bayes (CNB), Maximum Entropy y Support Vector Machine (SVM). Sin embargo, los dos clasificadores con mejor precisión fue SVM con 94% y CNB con 84% . |
| Munezero, Montero, Mozgovoy, & Sutinen | Exploiting sentiment analysis to track emotions in students' learning diaries. | Se creo un Sistema funcional que analiza y visualiza las emociones que los estudiantes expresan en sus días de aprendizaje. | Las emociones que extrae este sistema son las 8 emociones que menciona Plutchik . |

3.3. Clasificador EvoDAG

El reconocedor de emociones en texto tiene el nombre de EvoDAG (Evolving Directed Acyclic Graph) (Graff, 2016). EvoDAG es una librería en Python que implementa programación genética y trabaja con corpus de frases escritas en español cuyo objetivo es analizar exhaustivamente todas las combinaciones de las transformaciones de texto para descubrir características comunes.

3.4. Plataformas recolectoras de opiniones en texto.

Actualmente existen diferentes plataformas en internet que ofrecen diversos tipos de servicios: turísticos, entretenimiento, educativos, ventas, entre otros. Las empresas que ofrecen servicios y productos en línea están interesadas en conocer la experiencia de sus usuarios al utilizar sus servicios por lo cual les piden que agreguen una reseña junto con una valoración del servicio recibido. Esta información que proporcionan los usuarios es utilizada

por sistemas analizadores de sentimientos en texto con el fin de analizar la información y obtener el grado de satisfacción de los clientes para un producto o servicio en particular o hacia la empresa en general.

En esta sección se describen algunas plataformas recolectoras de opiniones.

3.4.1. Plataformas turísticas

Existen diferentes plataformas turísticas que se utilizan para hacer reservaciones cuando se va a viajar. Una vez que se concluye estos servicios se le pide al usuario que ingresen sus experiencias o sugerencias para posteriormente realizar una modificación con base a las opiniones de los usuarios. En la Figura 3-5 se muestra la plataforma Hoteles.com (Hoteles, 2018). Hoteles.com permite a los usuarios realizar consultas a través de la experiencia de los usuarios.

The screenshot shows the Hoteles.com search results page. On the left, there are several filter sections: 'Precio por noche' (Price per night) with a range from MXN 0 to MXN 10,000+; 'Categoría' (Category) with star ratings from 1 to 5; 'Evaluaciones de los huéspedes' (Guest ratings) with a range from 0 to 10; 'Pago y cancelación' (Payment and cancellation); 'Hoteles.com™ Rewards'; 'Zona' (Area); and 'Lugares de interés' (Points of interest) with a 2 km radius and checkboxes for 'Centro de la ciudad' and 'Cancún, Quintana Roo (CUN- Aeropuerto Internacional de)'. The main content area displays two hotel listings. The first is 'Suites Sina Cancún' with a rating of 8.2, 40 comments, and a price of MXN 2,542. The second is 'Hotel Green 16 Extended Stay' with a rating of 5.8, 135 comments, and a price of MXN 1,202. Each listing includes a photo of the hotel, a location map, and a 'Seleccionar' button.

Figura 3-5 Plataforma turística Hoteles.com

En la Figura 3-6 se muestra las categorías que los usuarios proporcionan cuando ingresan su experiencia. Esta plataforma ofrece diferentes formas de que los usuarios proporcionen una retroalimentación.

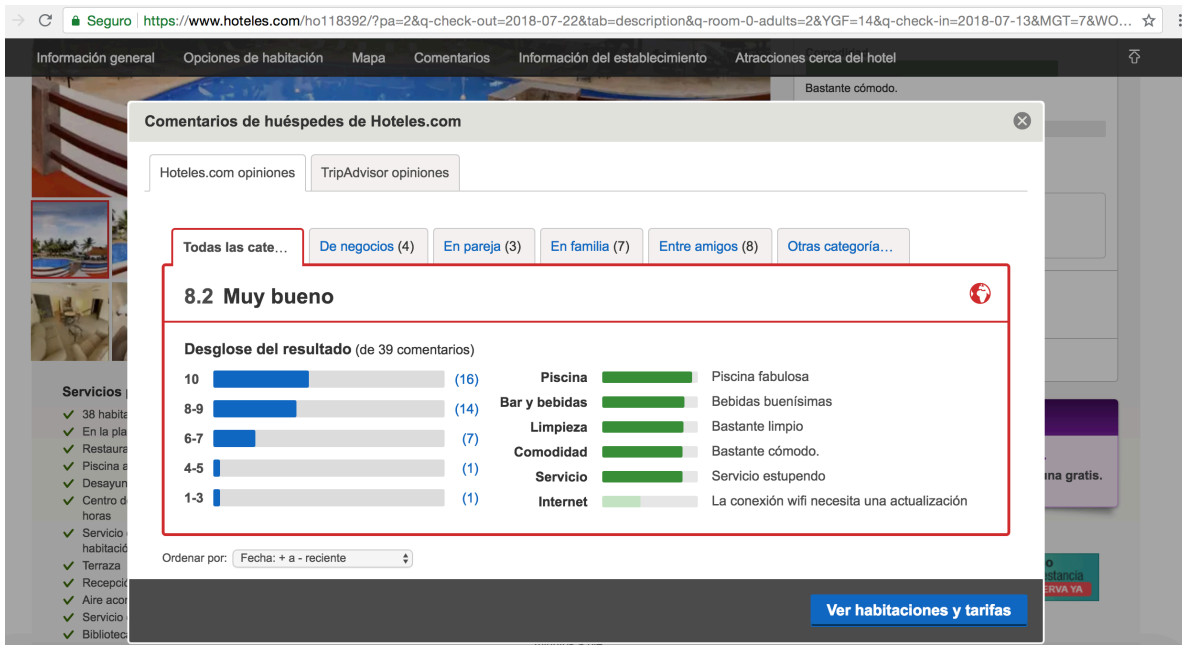


Figura 3-6 Categorías de Hoteles.com

Sin embargo, existen otras plataformas que ofrecen hospedaje como Bestday (BestDay, 2018) que ofrece servicios de hoteles, paquetes vacacionales, rentas vacacionales, vuelos, tours, entre otros servicios, estos servicios son evaluados. En la Figura 3-7 se muestra la evaluación de los tours que ofrece Bestday.

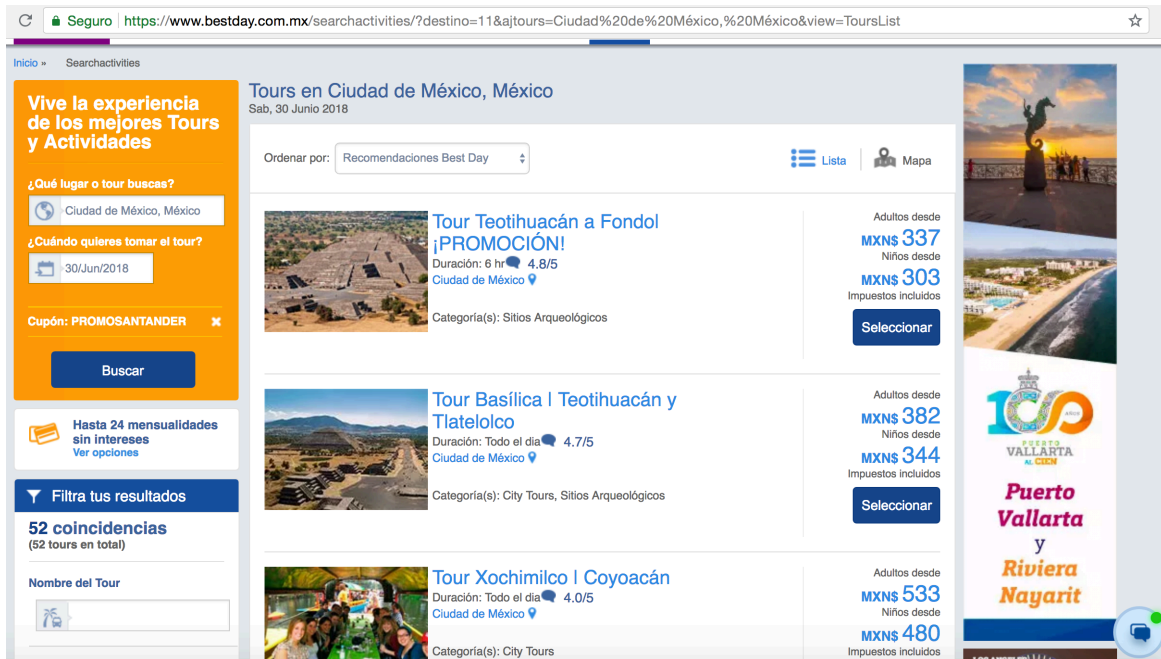


Figura 3-7 Plataforma turística Bestday

Volaris (Volaris, 2018) es una empresa que inicio ofreciendo servicio aéreo, posteriormente ofreció otros servicios como hoteles y renta de vehículos. En la Figura 3-8 se muestra el servicio de renta de autos. Al finalizar este servicio de renta de autos el usuario agrega una valoración, esta valoración es calculada por diferentes características como precio, limpieza, servicio del alquiler, funcionamiento, proceso a la hora de recoger y devolver el coche.

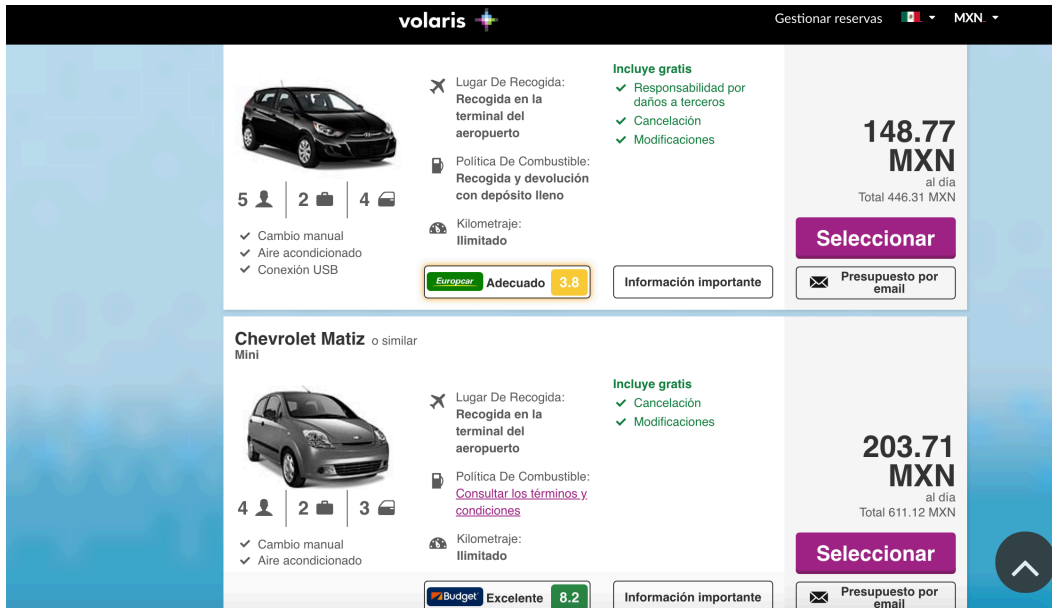


Figura 3-8 Plataforma Volaris

3.4.2. Plataformas de entretenimiento

Existen otras plataformas que se utilizan para el entretenimiento personal, entre estas se encuentra Netflix (Netflix, 2018) y YouTube (Youtube, 2018). En la Figura 3-9 se muestran estas plataformas el cual dan la opción de ingresar la opinión por medio de dos etiquetas (me gusta y no me gusta) con la finalidad de que los demás usuarios tengan una idea del contenido que observaran.

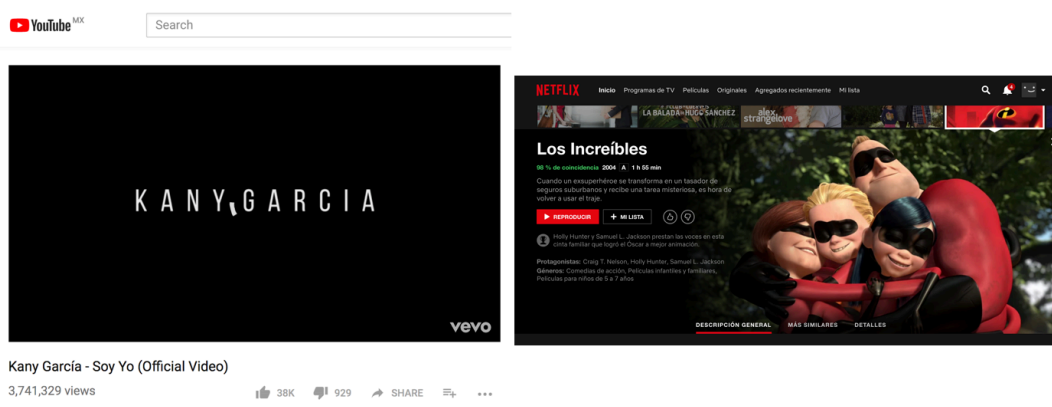


Figura 3-9 Plataformas de Entretenimiento

3.4.3. Plataformas de venta

Existen diferentes plataformas las cuales son populares por la venta de productos. Estas cuentan con un apartado donde los usuarios ingresen su opinión respecto al producto comprado, eso con la finalidad de que otros usuarios conozcan la evaluación del producto a través otros compradores. Esto se realiza una vez que el usuario ya tiene en sus manos el producto. En la Figura 3-10 se muestra la interfaz de Amazon (Amazon, 2018) en la parte superior derecha se encuentra la valoración del producto y si se ingresa a opiniones de clientes se desglosa un informe general. Este informe comenta un porcentaje de cada una de las opiniones.



Figura 3-10 Plataforma de ventas Amazon.

La Figura 3-11 muestra otras interfaces de ventas como google (Google Shopping, 2018) o mercado libre (Mercado Libre, 2018). En estas plataformas ofrecen los datos del producto como características, descripción del vendedor, fotografías y opiniones de los clientes.

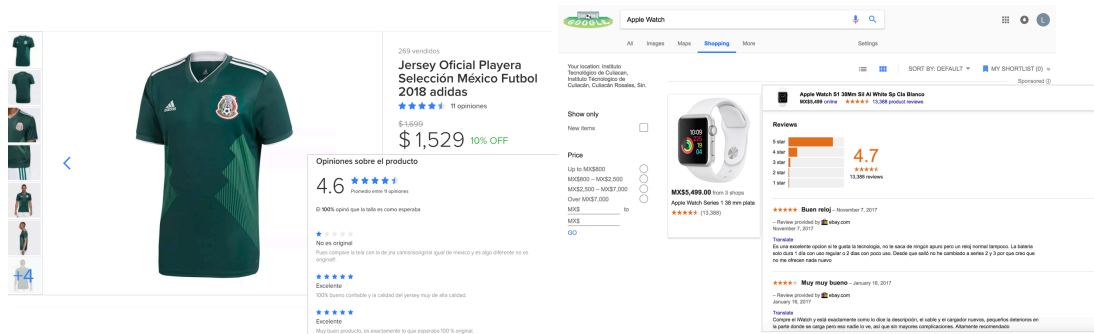


Figura 3-11 Otras plataformas de ventas.

3.4.4. Otras plataformas

Actualmente existen otras plataformas que ofrecen diferentes servicios y ayudan a la recolección de opiniones; ya que el usuario que lo utiliza tiene un apartado donde proporcionan sus opiniones con respecto al servicio. Existen diferentes plataformas que recolectan opiniones en áreas como: Comunicación (Facebook, Twitter), Noticias (El Norte, El debate), Educación (Udemy, Coursera, Teachlr), entre otras.

Capítulo 4

4. Desarrollo del proyecto

En esta sección se presenta la metodología diseñada para la creación de un corpus de frases en español. Las frases generadas están relacionadas con el aprendizaje en el área de la programación. Para la obtención de opiniones se utilizaron diferentes herramientas, plataformas y sistemas, entre ellas el Sistema de Evaluación de Recursos Educativos (SERE) para recolectar frases de estudiantes, algunas plataformas educativas en línea y la red social Twitter. Además, en este capítulo se describe el módulo de análisis de sentimientos que realiza la clasificación de las opiniones determinando emociones centradas en el aprendizaje, para finalizar, se describe el Servicio web que se desarrolló para evaluar una frase en tiempo real.

4.1. Metodología para la creación de un corpus

En esta sección se presenta la metodología que se diseñó para la creación del corpus de opiniones textuales en español para el área de programación de computadoras. En la Figura 4-1 se muestran las fases de la metodología.

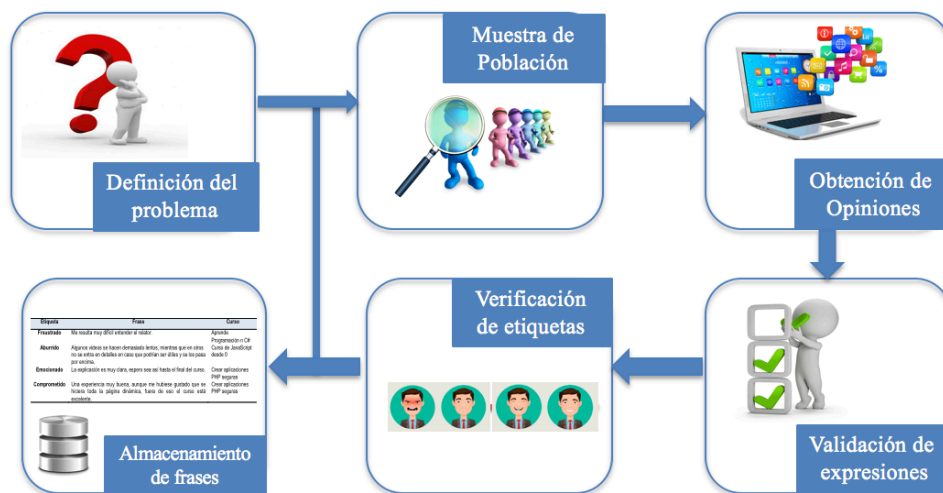


Figura 4-1 Metodología para la creación de un corpus de frases.

La metodología describe un proceso iterativo que permite incrementar el corpus de frases después de cada iteración.

A continuación, se explica cada fase de la metodología.

4.1.1. Definición del problema

Actualmente no existe un corpus de frases en español enfocadas al tema de aprendizaje de programación que permita realizar análisis de sentimientos para determinar el estado emocional del estudiante cuando utiliza un sistema educativo.

4.1.2 Muestra de población

En este trabajo, el interés se centra en determinar emociones relacionadas con el aprendizaje a partir de frases textuales en español, por lo que se estableció acumular frases escritas por estudiantes de diferentes escuelas, niveles y cursos, siempre y cuando estuvieran relacionados con la enseñanza del tema de programación de computadoras.

Los participantes seleccionados para expresar opiniones textuales en español, acerca de algunos recursos digitales educativos, fueron estudiantes inscritos en diferentes cursos con temas relacionados con la programación de computadoras. Algunos ejemplos son: estudiantes de las carreras de Ingeniería en Tecnologías de Información y Comunicaciones (ITIC) e Ingeniería en Sistemas Computacionales (ISC) del Tecnológico Nacional de México, campus Culiacán inscritos en el periodo agosto-diciembre 2017 en los cursos de Fundamentos de programación, Programación orientada a objetos, Estructura de datos, Programación II, Tecnologías de programación, entre otros y estudiantes registrados en cursos en línea en diversas plataformas educativas como son: Coursera (Coursera, 2018), Udemy (udemy, 2018) y Teachlr (teachlr, 2018). Además, las frases recolectadas de *Twitter* son de usuarios de una zona geográfica específica ya que estas fueron buscadas a través de una librería de Python.

4.1.3 Obtención de opiniones

El proceso para obtener las opiniones de los estudiantes se dividió en tres secciones de acuerdo con la fuente utilizada para obtener las frases que se incorporarían al corpus de opiniones. Las fuentes son: Sistemas SERE, desarrollo de proyectos prácticos, plataformas educativas digitales en línea y la red social Twitter. A continuación, se detalla cada una de las fuentes utilizadas.

Sistema de Evaluación de Recursos Educativos (SERE)

El sistema SERE fue diseñado para realizar la interacción con estudiantes con el objetivo de permitir que expresen sus opiniones acerca de algunos recursos educativos y así generar un corpus de opiniones en español enfocadas al aprendizaje del tema de programación. En el apartado 4.2.1 se presenta el desarrollo de este instrumento.

Desarrollo de proyectos prácticos.

Los estudiantes que participaron utilizando el sistema SERE, también participaron en la generación de opiniones en formato libre después de desarrollar la solución de un problema a través de un programa Java. A los estudiantes, se les solicitó que expresaran su opinión acerca de los aprendizajes logrados con el desarrollo de la solución del problema planteado. Las opiniones fueron capturadas y etiquetadas para ser incluidas en el corpus de opiniones.

Plataformas educativas digitales en línea

Las opiniones expresadas por estudiantes de los cursos en línea, se recolectaron con un proceso manual visitando diferentes plataformas educativas entre las que se encuentran: Udemy (udemy, 2018), Coursera (Coursera, 2018), y Teachlr (teachlr, 2018).

Las plataformas educativas en línea ofrecen distintos cursos que no son gratuitos; por lo que a los usuarios que compran un curso y están en el proceso de enseñanza les solicitan que ingresen una retroalimentación del curso. Estas retroalimentaciones se hacen públicas en una sección de la plataforma, con el fin de brindar información al público en general acerca de la calidad del curso a través de experiencias de los usuarios. Las opiniones publicadas en las plataformas y relacionadas con los temas de programación de computadoras fueron captadas y etiquetadas con una emoción centrada en el aprendizaje con el fin de incrementar el corpus. En la Figura 4-2 se muestran ejemplos de las tres plataformas educativas utilizadas para obtener opiniones de los usuarios.

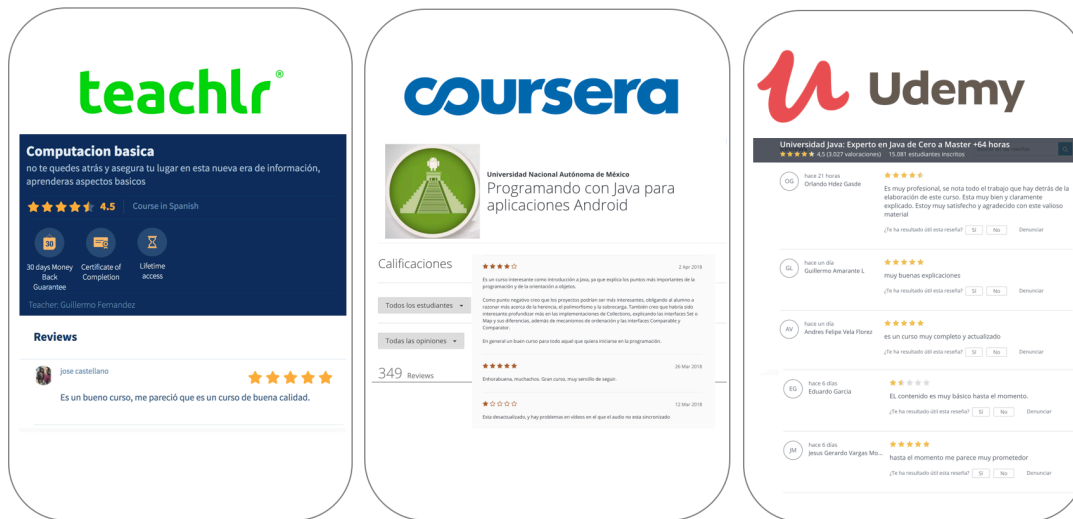
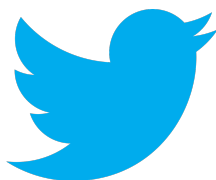


Figura 4-2 Críticas en las diferentes plataformas educativas en línea.

En el apartado 4.2.2 se muestra el proceso que se utilizó para la recolección de estas opiniones.

- *Extractor de información*

Este instrumento se utilizó para recolectar opiniones publicadas en la red social Twitter. A través de una librería de Python llamada *tweepy*, fue posible recolectar opiniones utilizando palabras clave contenidas en las frases publicadas. Estas frases también fueron etiquetadas con una emoción relacionada con el aprendizaje para incluirlas en el corpus. En la Figura 4-3. Se muestra el código ejecutado para buscar y seleccionar frases relacionadas a la programación.



```
print "Iniciando proceso mining-twitter..."
# This handles Twitter authentication and the connection to Twitter Streaming API
l = StdOutListener()
auth = OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
stream = Stream(auth, l)

locs = [-107.43488, 24.7885, -107.367493, 24.82118]
stream.filter(track=['programador', 'geek', 'informatico'], locations= locs, languages=["es"])
```

Figura 4-3 Código de extracción de frases en Twitter.

En el apartado 4.2.3 se presenta el proceso que se utilizó para la obtención de opiniones extraídas con la librería.

4.1.4. Validación de expresiones

La fase de validación de expresiones consiste en realizar un proceso de limpieza con el fin de garantizar que las frases generadas en la fase de *Obtención de frases* a través de los diferentes instrumentos fueran válidas, es decir que sean frases relacionadas a un objeto de aprendizaje del tema específico. Se revisó manualmente cada una de las frases y en este proceso se encontró que un 10% de frases eran incongruentes, realizando las correcciones necesarias. También se detectó que el 20% de las opiniones escritas correspondían a definiciones de términos básicos tales como algoritmo, computadora, entre otros, por lo que esas opiniones fueron anuladas ya que no expresaban emociones de los usuarios.

4.1.4. Verificación de etiquetas

Cada una de las frases generadas fue acompañada de una etiqueta que expresa la emoción del estudiante. El proceso de validación de etiquetas consistió en revisar manualmente que las etiquetas de las frases coincidan con el contenido emocional de los textos, considerando que frustrado y aburrido expresan una emoción negativa; en tanto que emocionado y comprometido expresan una emoción positiva.

Al finalizar el proceso de verificación de etiquetas, todas las frases y sus etiquetas fueron almacenadas en la base de datos para conformar el corpus de opiniones textuales en español para realizar análisis de sentimientos.

4.2. Herramientas para la obtención de frases

En esta sección se describen los diferentes instrumentos utilizados para la generación del corpus. En donde para el instrumento SERE se presenta el proceso de desarrollo, y para las plataformas educativas y la red social *Twitter* se muestra el proceso detallado que se utilizó para la obtención de opiniones.

4.2.1. Sistema de Evaluación de Recursos Educativos (SERE)

El Sistema de Evaluación de Recursos Educativos (SERE) fue implementado para generar un corpus de opiniones en español enfocadas en el aprendizaje, a través de la interacción de los usuarios del sistema (estudiantes de diferentes cursos) con el objetivo de permitir que

expresen sus opiniones y comentarios libremente acerca de los recursos educativos u objetos de aprendizaje de los temas de un curso.

En la Figura 4-4 se muestra como se planeó que el estudiante interactuara con el sistema usando diferentes dispositivos con acceso a Internet como son: Tablet, Smartphone, Laptop, o PC, con el fin de acceder al sistema y expresar sus opiniones.

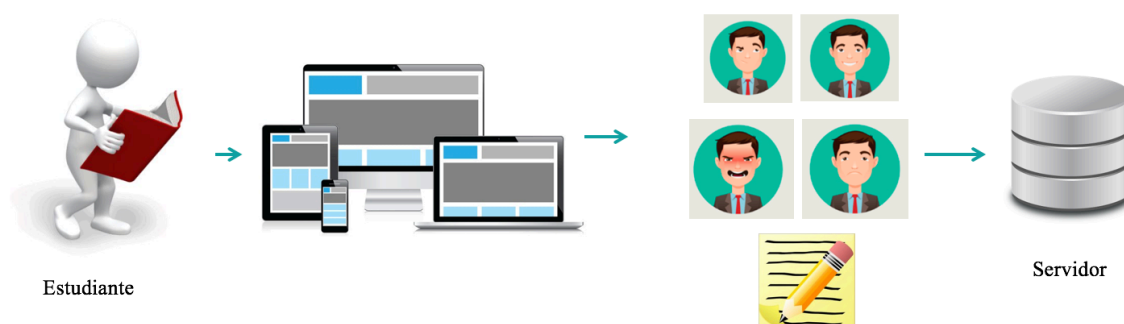


Figura 4-4 Funcionamiento del Sistema de Evaluación de Recursos Educativos

En esta sección se presenta el desarrollo de SERE, mostrando las diferentes etapas como son: la metodología, el análisis, diseño e implementación del sistema.

4.2.1.1. Metodología de desarrollo

Esta sección se describe el método incremental (Mills, 1980) el cual se basa en la filosofía de construir incrementando las funcionalidades del programa. Este modelo se utilizó para el desarrollo del sistema evaluaciones de recursos educativos (SERE) porque mediante este modelo se genera el software de forma rápida. Además de ser un modelo flexible, por los cambios que se realizan en cada incremento. Las fases que componen esta metodología se muestran en la Figura 4-5.

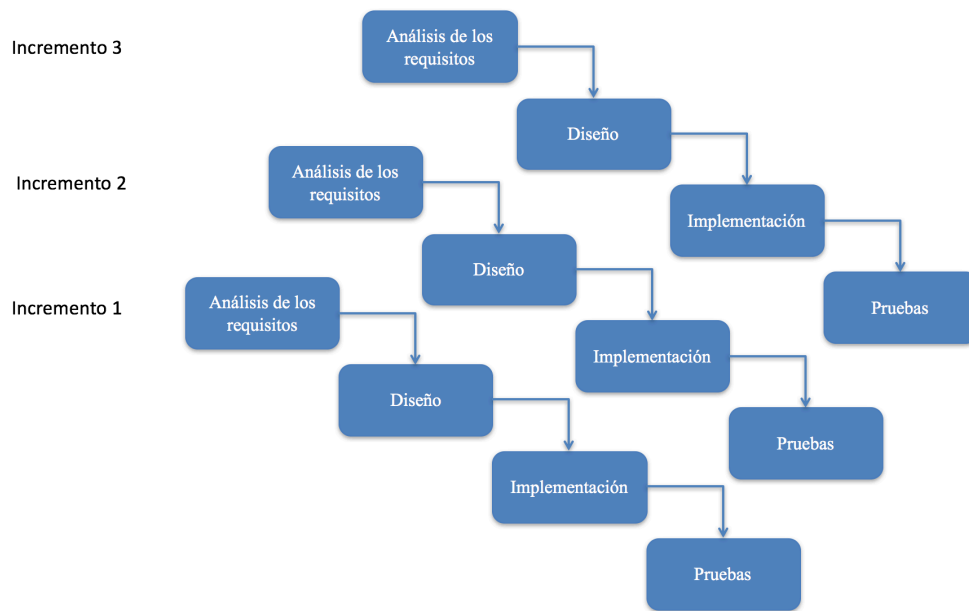


Figura 4-5 Metodología de desarrollo de SERE

Análisis de requisitos: En esta fase se realiza un análisis de las necesidades del cliente para determinar las características del software a desarrollar, y se especifica todo lo que debe hacer el sistema sin entrar en detalles técnicos.

Diseño: En esta etapa se describe la estructura interna del software, y las relaciones entre las entidades que lo componen.

Implementación: En esta fase se programan los requisitos especificados haciendo uso de las estructuras de datos diseñadas en la fase anterior. La programación es el proceso que lleva de la formulación de un problema de computación, a un programa que se ejecute produciendo los pasos necesarios para resolver dicho problema.

Pruebas: En esta fase como su propio nombre indica, una vez se termina la fase de implementación se verifica que todos los componentes del sistema funcionen correctamente y cumplen con los requisitos. El proceso de pruebas de este sistema SERE se presenta en el capítulo 5 de pruebas y resultados, donde se dan a conocer las diferentes pruebas que se realizaron con estudiantes y los resultados obtenidos.

4.2.1.2. Análisis del Sistema

En esta etapa se toman decisiones importantes para el desarrollo del proyecto software, ya que depende el rumbo del proyecto en las siguientes etapas.

Requisitos Funcionales y de Calidad

En esta etapa se clasifican los requerimientos funcionales y de calidad, donde se propusieron diferentes requisitos y se le asignaba una priorización en una escala alta, media y baja. En la Tabla 4-1 se muestran los 14 requerimientos funcionales con su respectiva descripción y prioridad.

Tabla 4-1 Requerimientos funcionales SERE

| No. | Descripción | Prioridad |
|--------------|--|--------------|
| RF-01 | El sistema debe permitir registrar usuarios, solicitándole sus datos personales: nombre, apellidos, materia y contraseña. | ALTA |
| RF-02 | El sistema debe solicitar usuario y contraseña para ingresar al sistema. | ALTA |
| RF-03 | El sistema debe mostrar un listado con los cursos dados de alta. | ALTA |
| RF-04 | El sistema debe permitir seleccionar a un usuario un objeto de aprendizaje que se quiera realizar. | ALTA |
| RF-05 | El sistema debe mostrar un listado con los temas de los cursos propuestos | ALTA |
| RF-06 | El sistema debe permitir al usuario guardar la opinión del usuario, el número de control, la fecha y hora una vez que terminen de ver el material. | ALTA |
| RF-07 | El sistema debe permitir adjuntar archivos para los proyectos de los estudiantes. | ALTA |
| RF-08 | El sistema debe almacenar en una Base de datos las opiniones proporcionadas por los estudiantes | ALTA |
| RF-09 | El sistema debe validar si ya se proporcionó una retroalimentación del material que seleccionó el usuario. | MEDIA |
| RF-10 | El sistema debe permitir agregar más de una opinión si el usuario así lo requiere. | MEDIA |
| RF-11 | El sistema debe permitir consultar el material ya interactuado. | BAJA |
| RF-12 | El sistema debe permitir generar una rúbrica por el material ya resuelto. | BAJA |
| RF-13 | El sistema debe registrar mediante una bitácora todas las acciones que se ejecuten. | BAJA |

Los atributos de calidad para el sistema de evaluación se clasifican en rendimiento, disponibilidad, portabilidad y usabilidad. Se aplicó la misma técnica de priorización que en

los requerimientos funcionales. Se clasificaron según el atributo de calidad. En la Tabla 4-2 se muestran los 7 requerimientos de calidad con su respectivo atributo, descripción y prioridad.

Tabla 4-2 Requerimientos de Calidad de SERE

| ID | Atributo | Requisito | Prioridad |
|--------------|-----------------|---|------------------|
| RC-01 | Rendimiento | El sistema deberá responder en máximo 3 segundos a cada petición de servicio | Alta |
| RC-02 | Rendimiento | El tiempo de espera del usuario para recibir respuesta de un proceso debe ser menor a 5 segundos. | Alta |
| RC-03 | Disponibilidad | El sistema deberá estar disponible 24/7 | Alta |
| RC-04 | Usabilidad | El sistema deberá ser fácil de usar para todas las personas que interactúan con él, sin que requiera una capacitación previa. | Media |
| RC-05 | Rendimiento | El tiempo esperado de carga de imágenes y documentos no excederá los 5 segundos. | Media |
| RC-06 | Portabilidad | El sistema deberá funcionar en cualquier navegador web. | Media |
| RC-07 | Usabilidad | El sistema cuenta con una interfaz gráfica amigable donde el usuario se puede adaptar máximo en 20 minutos. | Baja |

Restricciones

Durante el análisis de requerimientos se establecieron algunas restricciones, las cuales son:

- Todas las actividades del sistema se realizarán en línea a través del portal web.
- El usuario ingresará al sistema con su el usuario y contraseña.
- SERE solo estará disponible en el idioma español (México).

Actores

Los actores son aquellos que interactúan con el sistema. La Tabla 4-3 describe los de actores y su principal funcionalidad.

Tabla 4-3 Actores en SERE

| Actor | Descripción |
|-------------------|--|
| Estudiante | Es el usuario principal el cual proporciona una retroalimentación de los recursos educativos propuestos. |

En la Figura 4-6 se muestra gráficamente la relación donde el actor estudiante proporciona información sobre los objetos de aprendizaje propuestos.

Además, existe un administrador del Sistema que realiza modificaciones al sistema, administra la información proporcionada por los estudiantes y verifica que las frases ingresadas sean útiles para el módulo de análisis de sentimientos.



Figura 4-6 Actores en SERE

Casos de uso

Un caso de uso describe las actividades que deberán realizarse para llevar a cabo algún proceso mostrando la funcionalidad del sistema. Éstos se definieron con base al análisis de requerimientos y restricciones con la finalidad de establecer los mecanismos que le dan interacción al sistema. En la Figura 4-7 se muestra el diagrama general de casos de uso.

Este diagrama muestra a dos actores externos (Administrador de SERE y el Maestro) que realizan procesos manuales para el funcionamiento del Sistema de Evaluación de Recursos Educativos.

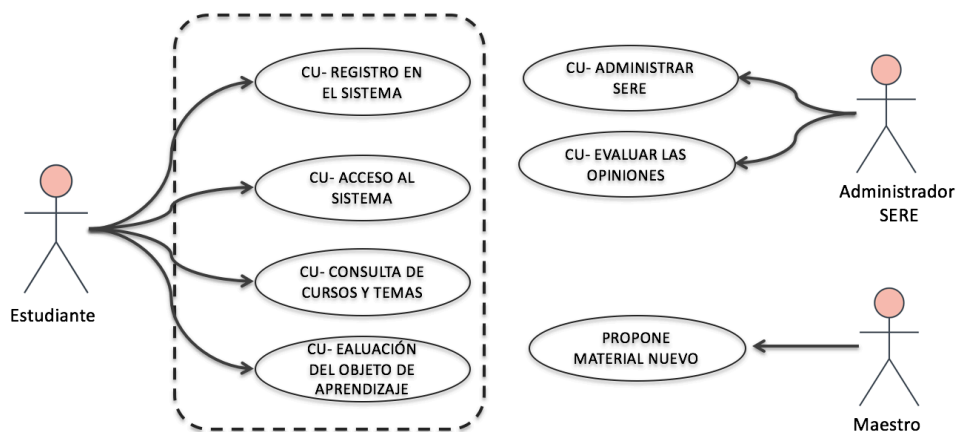


Figura 4-7 Diagrama de Casos de Uso SERE

A continuación, se presenta de forma detallada los casos de uso SERE.

| | |
|-----------------------|--|
| Caso de uso | CU-01.- REGISTRO EN EL SISTEMA |
| Actores | Estudiante |
| Función | Es un módulo de registros para el Sistema de Evaluaciones. |
| Descripción | El sistema deberá contar con una interfaz de registros para que los diferentes usuarios puedan darse de alta. |
| Precondición | El usuario podrá registrarse con el No. De control. |
| Referencia | <ul style="list-style-type: none"> El sistema debe permitir registrar usuarios, solicitándole sus datos personales: nombre, apellidos, materia y contraseña (RF-01). |
| Post Condición | El usuario se configuró satisfactoriamente y puede interactuar en el sistema. |

| | |
|-----------------------|--|
| Caso de uso | CU-02.- ACCESO AL SISTEMA |
| Actores | Estudiante |
| Función | Es un módulo para dar acceso al Sistema de Evaluaciones |
| Descripción | El sistema deberá contar con una interfaz de acceso donde los estudiantes ingresaran su No. De control y contraseña para poder interactuar con SERE. |
| Precondición | Deben ya estar registrados en el sistema |
| Referencia | <ul style="list-style-type: none"> El sistema debe solicitar usuario y contraseña para ingresar al sistema (RF-02). |
| Post Condición | El usuario podrá tener acceso al sistema. |

| | |
|-----------------------|--|
| Caso de uso | CU-03.- CONSULTA DE CURSOS Y TEMAS |
| Actores | Estudiante |
| Función | En este módulo se va a extraer la información de la Base de datos para mostrar el contenido. |
| Descripción | El sistema deberá contar con una interfaz donde los estudiantes podrán visualizar los cursos y temas disponibles y en el formato que se encuentran. |
| Precondición | Debe de acceder al SERE para poder tener acceso al contenido. |
| Referencia | <ul style="list-style-type: none"> • El sistema debe mostrar un listado con los cursos dados de alta (RF-03). • El sistema debe mostrar un listado con los temas de los cursos propuestos (RF-05) • El sistema debe permitir seleccionar a un usuario un objeto de aprendizaje que se quiera realizar (RF-04). |
| Post Condición | Una vez que seleccionen el curso y el tema que se desea consultar podrán evaluar el Objeto de aprendizaje |

| | |
|-----------------------|--|
| Caso de uso | CU-04.- EVALUACIÓN DEL OBJETO DE APRENDIZAJE |
| Actores | Estudiante |
| Función | En este módulo el estudiante va a consultar el Objeto de aprendizaje y posteriormente proporcionara una retroalimentación del material. |
| Descripción | El sistema deberá contar con diferentes interfaces donde se muestren distintos objetos de aprendizaje en diferentes formatos para que el estudiante tenga diferentes perspectivas de un tema en específico. |
| Precondición | Seleccionar el curso y tema que se desea evaluar. |
| Referencia | <ul style="list-style-type: none"> • El sistema debe permitir al usuario guardar la opinión del usuario, el número de control, la fecha y hora una vez que terminen de ver el material (RF-06). • El sistema debe permitir adjuntar archivos para los proyectos de los estudiantes (RF-07). • El sistema debe almacenar en una Base de datos las opiniones proporcionadas por los estudiantes (RF-08) • El sistema debe permitir agregar más de una opinión si el usuario así lo requiere (RF-10). • El sistema debe permitir consultar el material ya interactuado (RF-11). |
| Post Condición | La información proporcionada por el usuario se almacena en la base de datos. |

| | |
|-----------------------|---|
| Proceso manual | CU-05.- ADMINISTRAR SERE |
| Actores | Administrador SERE |
| Función | El administrador podrá realizar cambios, bajas y altos de objetos de aprendizaje. |
| Descripción | El proceso se realizará en el código viéndose reflejado en el sistema. |
| Precondición | Los cambios que se realicen serán por medio de propuestas por el maestro. |
| Post Condición | Se espera que con los cambios propuestos logremos balancear lo más posible el corpus. |

| | |
|-----------------------|--|
| Proceso manual | CU-06.- EVALUA LAS OPINIONES |
| Actores | Administrador SERE |
| Función | En este módulo se verifica que las opiniones ingresaras en SERE |
| Descripción | Se realiza un pre-proceso manual de limpieza para mandarlo al módulo de análisis de sentimiento. |

| | |
|-----------------------|--|
| Proceso manual | CU-07.- PROPONE MATERIAL |
| Actores | Maestro |
| Función | proponer objetos de aprendizaje para elevar y hacer que los estudiantes nos ayuden a proporcionar opiniones |
| Descripción | El objetivo principal de este módulo es que el maestro proponga temas que nos ayude a balancear el corpus para obtener una mejor precisión del módulo de análisis de sentimientos. |

4.2.1.3. Diseño

A continuación, se describe parte del diseño que se elaboró con base a lo que se definió en la etapa de análisis del sistema.

Diagrama de contexto

En esta sección se muestra el diagrama de contexto del Sistema de Evaluaciones de Recursos Educativos, el cual que tiene diferentes interfaces que ayudan a realizar la comunicación con otras entidades. En la Figura 4-8 señala la dependencia con la Base de Datos donde se encuentra alojada información de los estudiantes, las opiniones ingresadas, la información de los cursos, además cabe señalar que también existe relación de dependencia con el Repositorio ya que en este se encuentran imágenes y archivos que el estudiante sube como evidencia.

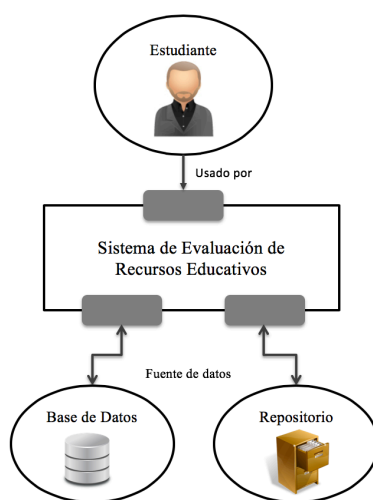


Figura 4-8 Diagrama de contexto SERE

SERE se encuentra ubicado en el centro del diagrama de contexto y a su alrededor aparecen diversas entidades con las que interactúa el sistema. Estas entidades se describen en la Tabla 4-4.

Tabla 4-4 Entidades Externas de SERE

| Entidad | Descripción |
|---------------|---|
| Estudiante | Es el usuario que proporciona una retroalimentación del objeto de aprendizaje interactuado. |
| Base de datos | Contiene las opiniones e información de los estudiantes, los cursos y los temas. |
| Repositorio | Carpetas que contienen las imágenes representativas de los cursos y los archivos de los proyectos encargados como evidencias. |

Arquetipos

Los arquetipos son entidades abstractas identificadas que se utilizan para representar el funcionamiento interno de SERE. Los arquetipos se determinaron con base en los requerimientos funcionales. En la Tabla 4-5 se describen los arquetipos y en la Figura 4-9 se muestra el flujo de las relaciones entre ellos.

Tabla 4-5 Arquetipos del Sistema SERE

| Arquetipo | Descripción |
|-------------------|---|
| Estudiante | Es el usuario que consulta y proporciona una retroalimentación de los Objetos de aprendizaje. |
| Opinión | Es una frase escrita por diferentes estudiantes donde describen sugerencias, observaciones, entre otras cosas del material proporcionado. |
| Curso | Son las entidades que contienen los temas a consultar, estos son proporcionados por el maestro. |
| Temas | Es información sobre un tema específico el cual se puede presentar en diferentes formatos como lo son imágenes, Texto, videos, programas. |

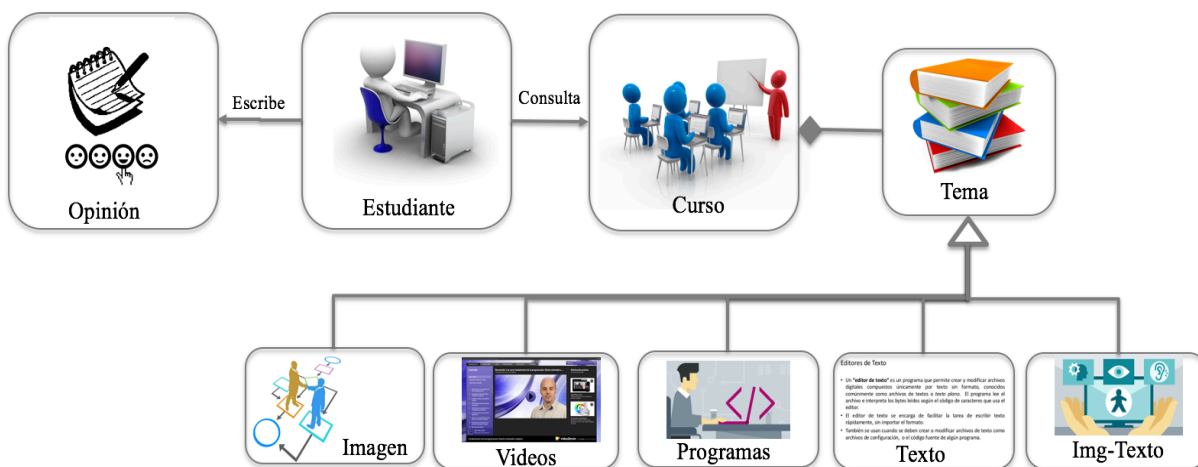


Figura 4-9 Arquetipos de SERE

Diferentes vistas del sistema

A continuación, se da una breve explicación de la arquitectura que se utilizó para el desarrollo del Sistema SERE.

Vista Física

La vista física considera los requerimientos no funcionales, tales como disponibilidad y portabilidad, entre otros. Satisface algunos requerimientos ya que los estudiantes pueden acceder al sistema SERE por medio de computadoras que cuenten con conexión a internet, por ser una plataforma web en la cual puede ser accesible desde cualquier computadora conectada a la web. En la Figura 4-10 se expone la arquitectura Cliente-Servidor de SERE.

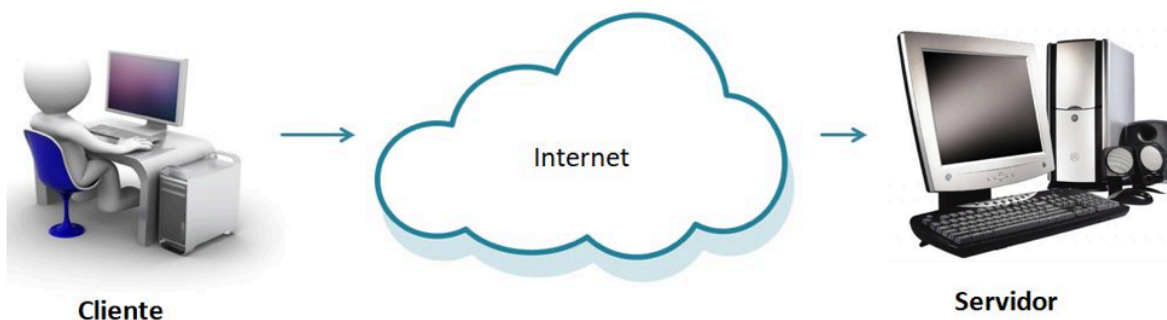


Figura 4-10 Arquitectura Cliente-Servidor SERE

Vista lógica

El modelo arquitectónico por capas permite el desacoplamiento de las partes que componen SERE, el desarrollo se lleva a cabo en niveles independientes con la finalidad de darle un orden al desarrollo. La arquitectura por capas de SERE se muestra en la Figura 4-11.

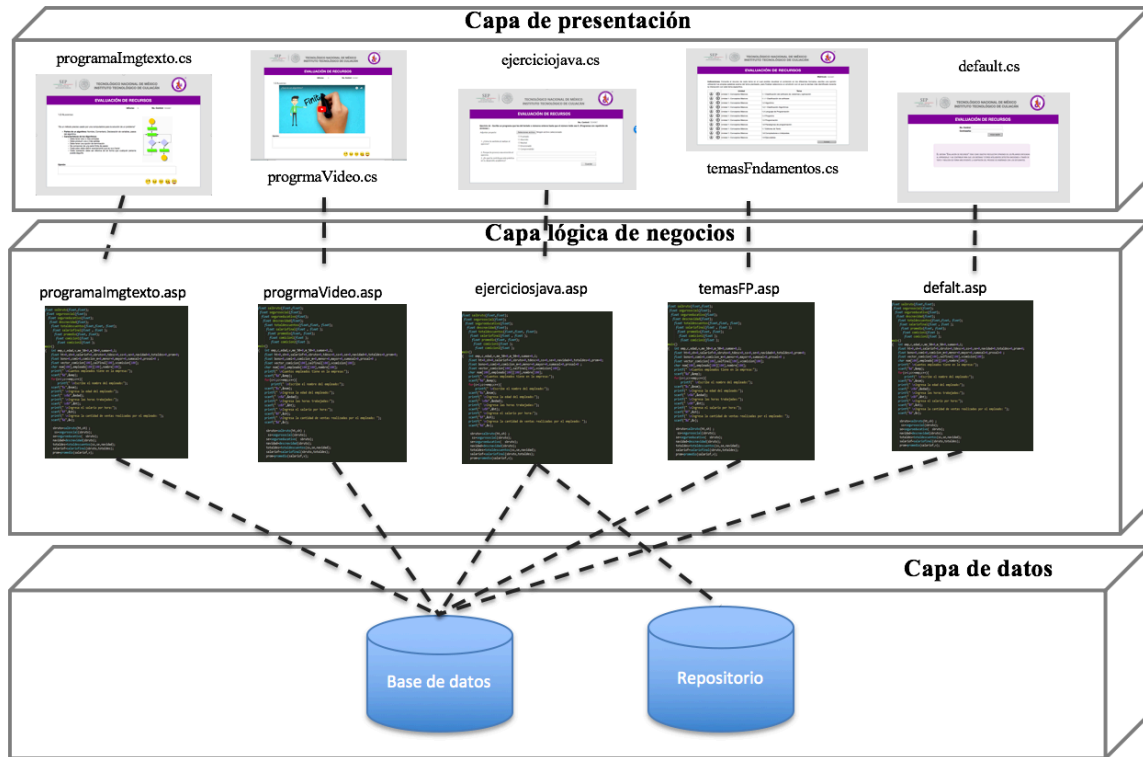


Figura 4-11 Vista Lógica del Sistema SERE

- **Capa de Presentación:** contiene las interfaces que se presentan al usuario. Es la responsable de gran parte de validaciones de formato en los datos de entrada y se comunica únicamente con la capa de dominio.
- **Capa lógica de negocios:** en esta capa es donde se encuentran los programas y también es conocida como capa de negocio. porque es aquí donde se establecen todas las reglas que deben cumplirse. Esta capa se comunica con la capa de presentación, para recibir las solicitudes y presentar los resultados, y con la capa de datos, para solicitar a la base de datos almacenar o extraer datos.
- **Capa de Datos:** es el contenedor de los datos, se encarga de la persistencia y los accesos a la información. Una de las características favorables de la programación en capas es la seguridad en los datos ya que los accesos se controlan desde la capa de dominio.

Vista de desarrollo

La arquitectura de desarrollo es representada por módulos y diagramas de subsistemas. En el desarrollo del sistema se consideraron tres partes: la primera va enfocada a la interfaz de usuario (esta interfaz puede ser cualquier dispositivo que tenga habilitado un navegador web), otra que va dirigida al servidor de aplicaciones y una tercera parte enfocada a un servidor de base de datos SQL Server. En la **Figura 4-12** se muestra la vista de desarrollo.

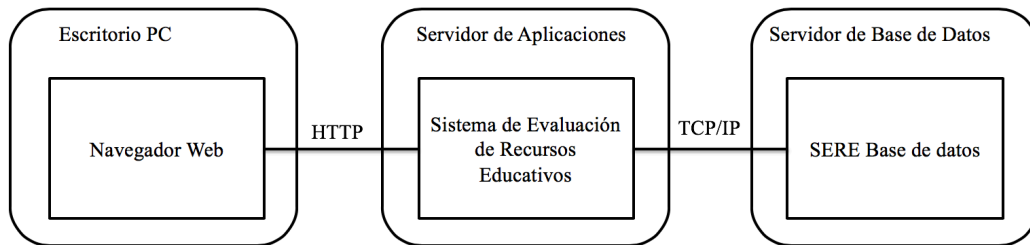


Figura 4-12 Vista de Desarrollo SERE

4.2.1.4. Implementación

SERE funciona en entorno web, por lo que se usaron herramientas adecuadas que facilitaron su construcción, éstas se explican a continuación.

Lenguajes y herramientas software

SERE es una aplicación Web que fue desarrollada para la plataforma .NET en Visual Studio 2013, y se utilizó la base de datos SQL Server 2008 R2 para almacenar la información de los cursos, recursos educativos, objetos de aprendizaje, usuarios, y opiniones emitidas por los estudiantes. SERE es alojado en la URL(<http://posgradoitc.ddns.net:8000/>). En la Figura 4-13 se muestran las plataformas que se utilizaron para el desarrollo de este sistema.

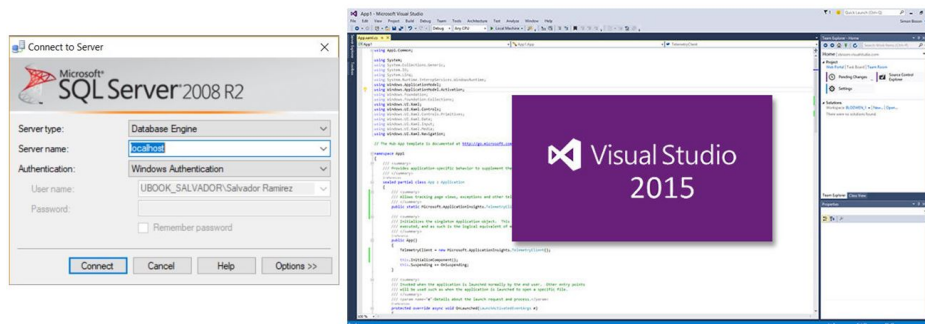


Figura 4-13 Plataformas de SERE

Interfaces del sistema

En esta sección se describen las diferentes interfaces desarrolladas para que el usuario interactúe con SERE. En la Figura 4-14 se muestra la interfaz inicial donde se presenta una breve introducción del objetivo de este sistema, además da acceso a los diferentes usuarios. En esta interfaz se valida que el usuario y la contraseña sean correctos, es decir que se encuentren registrados en la base de datos.

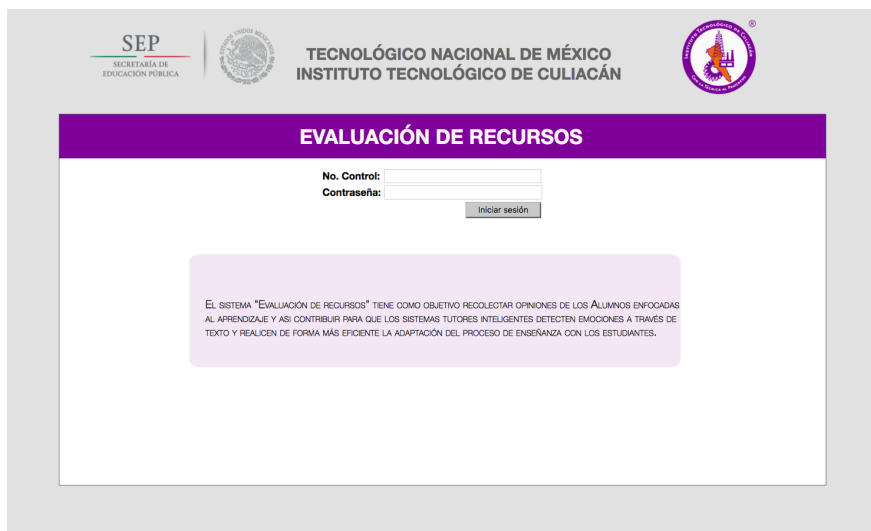


Figura 4-14 Interfaz principal de SERE

En la Figura 4-15 se presenta la platilla utilizada para desplegar los cursos incluidos en SERE. Estos cursos fueron propuestos por los maestros de las diferentes carreras.

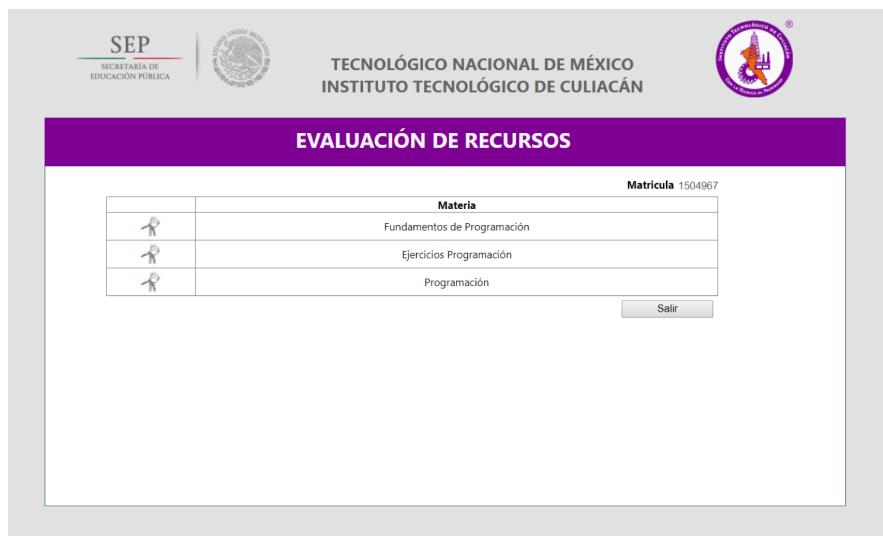


Figura 4-15 Interfaz de cursos SERE

En la Figura 4-16 se presenta la interfaz que contiene los temas del curso llamado Fundamentos de Programación. Estos temas están en dos formatos: Imagen-Texto y Video. Como se aprecia en la Figura 4-16, la interfaz esta dividida en tres secciones: la primera sección muestra el número de control del estudiante y las indicaciones; la segunda sección presenta una tabla con temas, donde aparecen los formatos disponibles, la unidad y el tema. La tercera sección presenta un botón para regresar a la interfaz donde se encuentran todos los cursos de SERE.

EVALUACIÓN DE RECURSOS

Matricula 1504967

1 **Indicaciones:** Consulta el recurso de cada tema en el cual podrás visualizar el contenido en los diferentes formatos, escribe una opinión utilizando tus propias palabras acerca del tema planteado, para finalizar selecciona un emoticón con el que te sientas más identificado durante la interacción con este tema específico.

| | Unidad | Tema |
|--|-----------------------------|--|
| | Unidad 1: Conceptos Básicos | 1.1 Clasificación del software de: sistemas y aplicación |
| | Unidad 1: Conceptos Básicos | 1.1.1 Clasificación de software |
| | Unidad 1: Conceptos Básicos | 1.2 Algoritmo |
| | Unidad 1: Conceptos Básicos | 1.2.1 Clasificación Algoritmos |
| | Unidad 1: Conceptos Básicos | 1.3 Lenguaje de Programación |
| | Unidad 1: Conceptos Básicos | 1.4 Programa |
| | Unidad 1: Conceptos Básicos | 1.5 Programación |
| | Unidad 1: Conceptos Básicos | 1.6 Paradigmas de programación |
| | Unidad 1: Conceptos Básicos | 1.7 Editores de Texto |
| | Unidad 1: Conceptos Básicos | 1.8 Compiladores e intérpretes |
| | Unidad 1: Conceptos Básicos | 1.9 Ejecutables |

2

3 **Cursos**

Figura 4-16 Interfaz del tema Fundamentos de programación SERE

En la Figura 4-17 se muestra la interfaz con el contenido del tema en formato Imagen-Texto. Esta interfaz se divide en tres secciones: en la primera sección se encuentran los datos del estudiante y del curso; en la segunda sección se muestra la información del tema en formato imagen-texto, y finalmente en la sección tres se presenta un cuadro para capturar la opinión del estudiante, así como un conjunto de cinco emoticones que se utilizan para seleccionar la emoción centrada en el aprendizaje que experimenta el alumno.

EVALUACIÓN DE RECURSOS

1
IdCurso: 3
No. Control: 1504967

1.2 ALGORITMO

"Es un método preciso usado por una computadora para la solución de un problema"

- **Partes de un algoritmo:** Nombre, Comentario, Declaración de variables, pasos del algoritmo.
- **Características de los Algoritmos:**
 - Debe tener cero o mas entradas
 - Debe producir una o más salidas
 - Debe tener una opción de terminación
 - Se componen de una serie finita de pasos
 - Cada paso debe definir exactamente que se va a hacer
 - Cada operación debe ser efectiva de tal forma que cualquier persona pueda seguirla

Opinión

3

Figura 4-17 Interfaz con el contenido del tema en formato Imagen-texto

La Figura 4-18 muestra la interfaz con el contenido del tema en formato video. Esta interfaz se divide en tres secciones: en la primera sección se despliega el número de control del estudiante y el Id del curso; en la segunda sección se encuentra el video del tema que fue seleccionado. Para finalizar, la última sección se utiliza para recolectar la opinión del estudiante; ésta se dividió en dos partes: un cuadro de texto donde el estudiante ingresará en formato texto una retroalimentación de la información proporcionada y un emoticón, el cual están relacionado con una emoción centrada en el aprendizaje (frustrado, aburrido, neutral, emocionado y comprometido).



Figura 4-18 Interfaz con el contenido del tema en formato video

Posteriormente se realizó una segunda versión de SERE con el fin de recolectar frases con emoción negativa (frustrado y aburrido). Para esto se decidió aumentar el grado de dificultad a los recursos educativos agregando ejercicios de programación, sin embargo, se agregaron más ejercicios tal como lo muestra en la Figura 4-19. Esta interfaz se divide en cuatro secciones: la primera muestra el número de control del estudiante y las indicaciones para realizar los ejercicios. La segunda sección, presenta las unidades de los ejercicios de programación. Una vez que se selecciona la unidad se desplegara la tercera sección en donde se encuentran los ejercicios que corresponden a la unidad seleccionada; para finalizar en la cuarta sección se muestran los botones de Consulta y Cursos, los cuales se utilizan para consultar material resuelto y regresar a la interfaz anterior respectivamente.



Figura 4-19 Ejercicios de programación en SERE

En la platilla que se muestra en la Figura 4-20 presenta una interfaz para resolver ejercicios de programación. Esta interfaz contiene dos secciones: en la primera se muestra el numero de control del estudiante y el planteamiento del problema. En la segunda sección se realizan diversas preguntas acerca de su experiencia durante el proceso de aprendizaje, además se le pide al estudiante que proporcione el ejercicio resuelto en un archivo .rar.

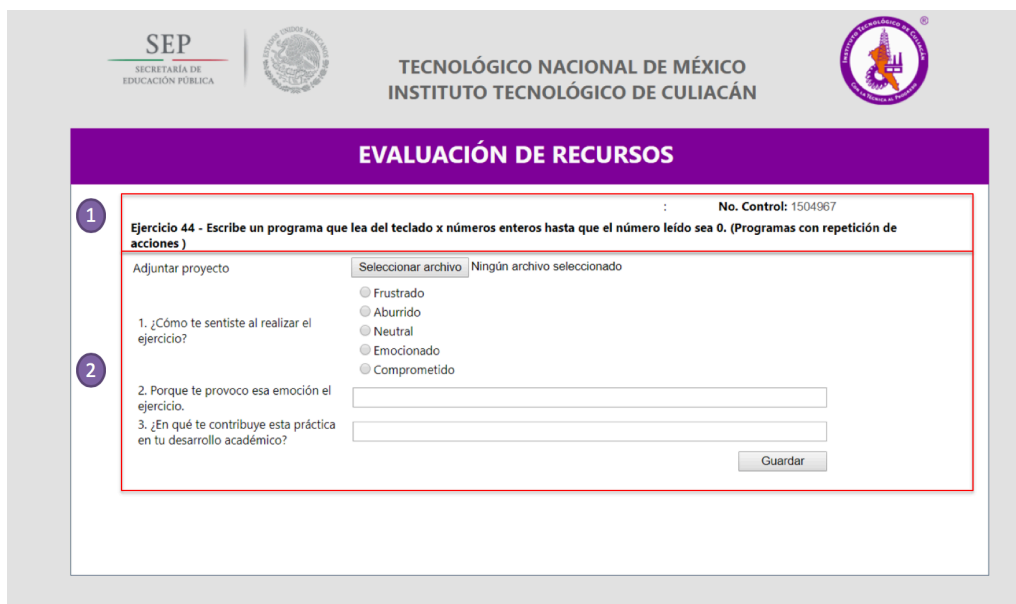


Figura 4-20 Interfaz de un ejercicio de programación.

Se realizó una tercera versión del sistema, donde se agrego una interfaz (Figura 4-21) la cual contiene una tabla con el número de ejercicio, la unidad y la fecha en el cual el estudiante resolvió el ejercicio, esto con la finalidad que el estudiante lleve un registro de los ejercicios ya resueltos.



Figura 4-21 Interfaz de Consulta

En el Anexo se incluye un CD con el código fuente del sistema.

4.2.2. Obtención de frases desde los Cursos en línea

El proceso de obtención de frases de este instrumento se inició a partir de enero 2018 con el fin de recolectar frases negativas (aburrido y frustrado), ya que la mayoría de las frases recolectadas por el instrumento SERE eran opiniones con emociones positivas (emocionado y comprometido).

Este instrumento consiste en recolectar frases de diferentes plataformas educativas en línea que ofrecen distintos cursos gratuitos. Para la creación de este corpus, se busco la categoría de desarrollo de software, sin embargo, ofrecen diferentes categorías en cada una de las plataformas. A continuación, se muestra el proceso que se utilizó en las tres diferentes plataformas.

Udemy es una plataforma educativa que cuenta con más de 100 cursos en español. En la Figura 4-22 se muestra la interfaz inicial de esta plataforma educativa.

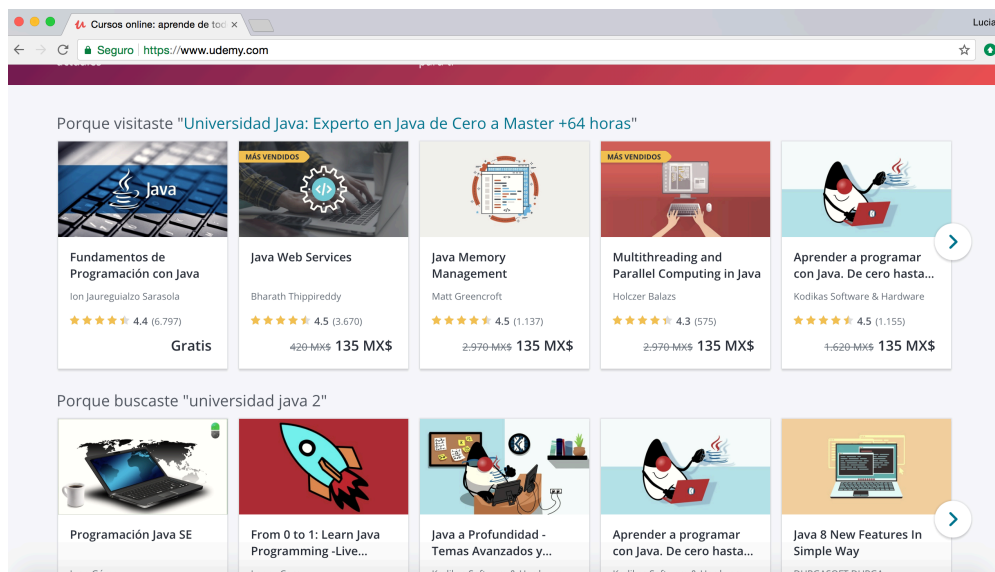


Figura 4-22 Plataforma educativa Udemy.

Udemy ofrece cursos en diferentes categorías, sin embargo, la categoría que se utilizó fue la de *Desarrollo* ya que se buscaban opiniones que estuvieran relacionadas con la programación. La Figura 4-23 muestra las categorías, los temas de estas categorías y algunos temas populares de categoría. La mayoría de las frases fueron extraídas de temas populares del desarrollo web ya que son los cursos más vendidos.

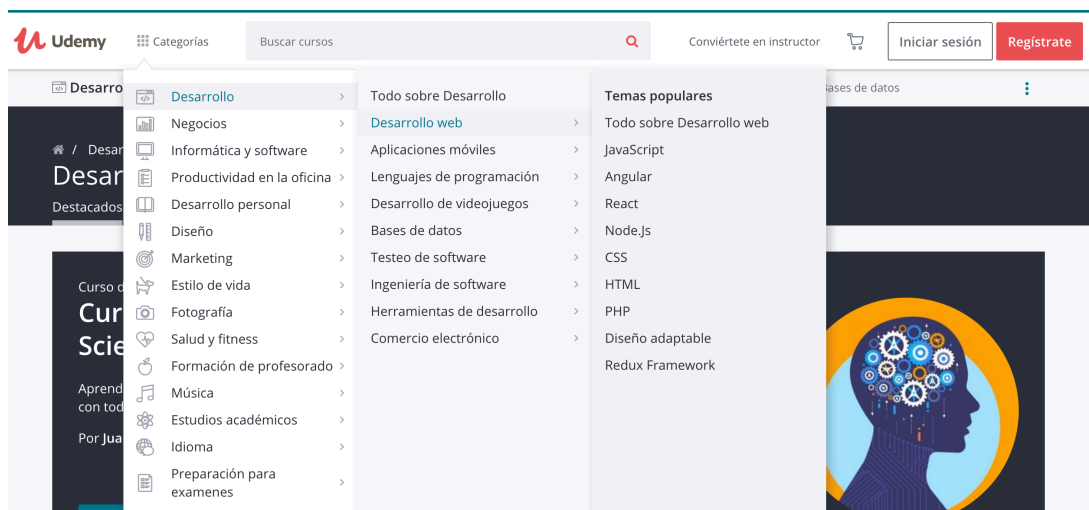


Figura 4-23 Categorías de Udemy

Cada uno de estos cursos cuenta con diferentes datos como: el programa del curso, los requisitos, la descripción, los cursos relacionados, información del instructor, la valoración de curso y las reseñas.

Estas reseñas son capturadas por aquellas personas que compraron el curso; se le pide al usuario tres veces durante todo el curso, que proporcione información sobre su experiencia de aprendizaje, además de evaluarla con una ponderación. Estas reseñas se hacen públicas, con el fin de brindar información al público en general acerca de la calidad del curso a través de diferentes experiencias de los usuarios que ya tienen este curso. En la Figura 4-24 se muestra un ejemplo de algunas reseñas del curso de PHP Y MYSQL, además se muestra la ponderación media.

The screenshot displays the course page for 'PHP y MYSQL: El Curso Completo, Practico y Desde 0!'. The course has a 4.6 average rating from 4,652 reviews and 12,493 students enrolled. A promotional offer for Mother's Day is active, showing a price of 135 MX\$ (95% discount from 2,570 MX\$) with a 7-hour countdown. The 'Incluye' section lists 15 hours of video, 2 supplementary resources, lifetime access, mobile and TV access, and a completion certificate. Three student reviews are visible, all with 5-star ratings. The first review by Cesar Kimura (2 months ago) mentions the course's practicality. The second by Gonzalo Amaral (3 months ago) praises the content but suggests adding more exercises. The third by Diego Guarrera (3 months ago) commends the clarity of the explanations.

| Calificación | Porcentaje |
|--------------|------------|
| ★★★★★ | 64 % |
| ★★★★☆ | 26 % |
| ★★★☆☆ | 7 % |
| ★★☆☆☆ | 2 % |
| ★☆☆☆☆ | 1 % |

| Usuario | Fecha | Calificación | Comentario |
|----------------|--------------|--------------|--|
| CK | hace 2 meses | ★★★★★ | Porque se un poco de programacion y hasta el momento me ha enseñado casi nada..., talvez mas adelante cuando mezcle html con php en algun ejemplo mas complicado... |
| Gonzalo Amaral | hace 3 meses | ★★★★★ | Me gustó el curso, un contenido completo. Quizás le agregaría alguna practica nueva, como por ejemplo, como crear un carrito de compras para una pagina de ventas. O como un usuario Admin, puede dar de altas a otros usuarios. Pero de verdad muy conforme con el contenido. |
| DG | hace 3 meses | ★★★★★ | Explica con claridad todo lo necesario para comenzar con programacion web y aparte los conceptos estan actualizados al paradigma de objetos, que es lo que se debe utilizar. No hay mucha |

Figura 4-24 reseñas de un curso de Udemy

Una vez extraídas las opiniones publicadas fueron capturadas y etiquetadas con una emoción centrada en el aprendizaje (frustrado, aburrido, emocionado y comprometido) con el fin de incrementar el corpus.

Algunas veces se realizaron filtros de las opiniones que estaban evaluadas con una ponderación baja, ya que estas plataformas se utilizaron con el objetivo de obtener opiniones negativas, puesto que se notó que los usuarios que compran un curso son personas que buscan que se cumpla con todos los requisitos que ofrecen y expresan su opinión cuando no les parece adecuado el material o la explicación del instructor, ya que su inversión no es lo que esperaban. Una vez que se tomaron las opiniones de cursos más destacados de Udey se continuó con las siguientes plataformas.

Coursera es una plataforma educativa que ofrece diferentes cursos y se utilizó un proceso similar para recolectar las opiniones de los usuarios. En la Figura 4-25 se muestra la interfaz inicial de la plataforma Coursera.



Figura 4-25 Plataforma educativa Coursera

La Figura 4-26 muestra del lado izquierdo las diferentes categorías que ofrece Coursera, y del lado derecho se muestran algunos cursos de la categoría ciencias de la computación, sin embargo el 50% de los cursos que ofrece esta plataforma se encuentran en el idioma inglés por lo que los usuarios que compran estos cursos, ingresan sus experiencias en el mismo idioma con el fin de que el tutor pueda leerlos, ya que es una retroalimentación para mejorar su material de aprendizaje. Otra gran desventaja es que cuando es un programa especializado como el que se muestra en la Figura 4-26 no se muestran las experiencias en los datos del curso.

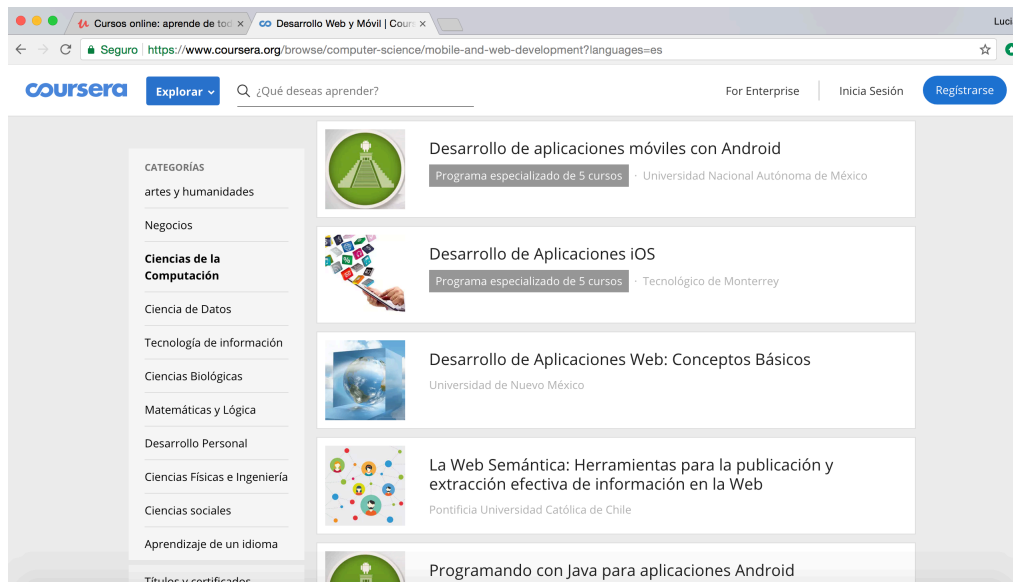


Figura 4-26 Categoría ciencias de la computación - Coursera

Para finalizar se utilizó una última plataforma educativa Teachlr; esta plataforma aún no es muy conocida a diferencia de las otras dos plataformas. Los cursos que ofrece son muy pocos además de que el 70% son cursos en inglés, por lo cual se utilizó muy poco para la extracción de opiniones. En la Figura 4-27 se muestra la interfaz de esta plataforma.

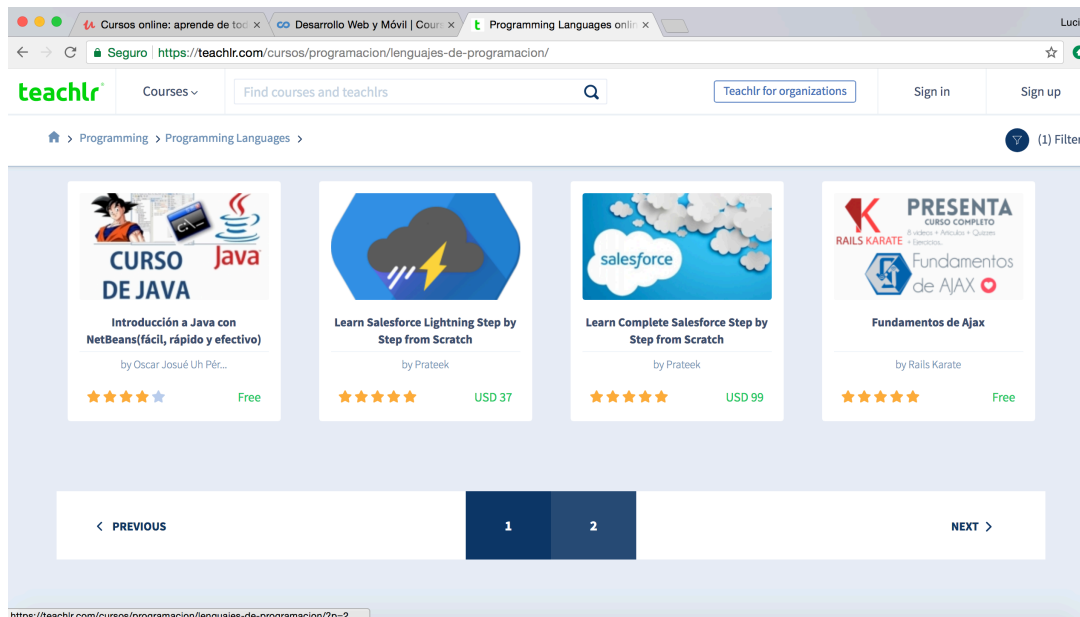


Figura 4-27 Plataforma educativa Teachlr

4.2.3. Obtención de frases desde Twitter

Este instrumento se utilizó en diferentes ocasiones para buscar opiniones a través de diferentes palabras claves. Para poder realizar este proceso se utilizó la librería *tweepy*; esta solo funciona con la versión 2.7 de Python. A continuación, se mencionan los pasos para obtener los accesos y así recolectar frases:

- Paso 1: Ingresa al siguiente link: <http://dev.twitter.com/apps> (requisito: tener un usuario y contraseña de Twitter). En la Figura 4-28 se muestra la interfaz de inicio donde pondrá ingresar al área de desarrollo de *Twitter* a través de tu cuenta.

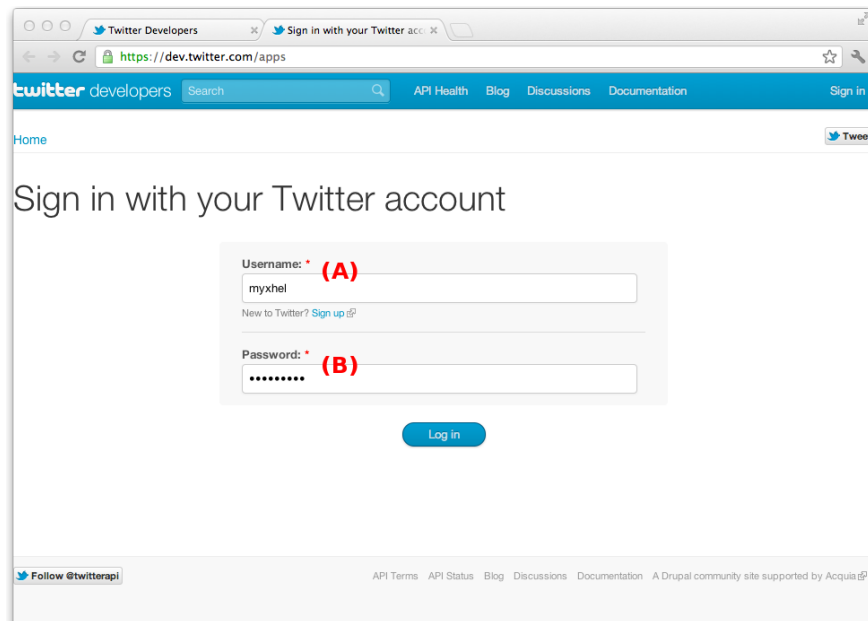


Figura 4-28 Interfaz inicial para desarrollar de Twitter

- Paso 2: Crea una aplicación. En la Figura 4-29 se muestra la interfaz donde se solicita ingresar los datos para crear una aplicación de Twitter. Estos datos son requeridos para generar accesos claves para poder extraer datos de Twitter.

Home → My applications

Create an application

Application Details (A)

Name: *
PyMG
Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description: *
This is a demo App for MundoGeek.com (B)
Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

WebSite: *
http://www.mundogeek.com (C)
Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens. (If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL:
(D)
Where should we return after successfully authenticating? For @Anywhere applications, only the domain specified in the callback will be used. OAuth 1.0a applications should explicitly specify their oauth_callback URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

Developer Rules Of The Road

Last Update - 1st of June 2011

Rules of the Road

Twitter maintains an open platform that supports the millions of people around the world who are sharing and discovering what's happening now. We want to empower our ecosystem partners to build valuable businesses around the information flowing through Twitter. At the same time, we aim to strike a balance between encouraging interesting development and protecting both Twitter's and users' rights.

So, we've come up with a set of Developer Rules of the Road ("Rules") that describe the policies and philosophy around what type of innovation is permitted with the content and information shared on Twitter.

The Rules will evolve along with our ecosystem as developers continue to innovate and find new, creative ways to use the Twitter API, so please check back periodically to see the most current version. Don't do anything prohibited by the Rules, but talk to us if you think we should make a change or give you an exception.

If you will eventually need more than 5 million user tokens for your projects, you will need to talk to us directly about access to the Twitter API.

I. Twitter Content

Yes, I agree (E)
By clicking the "I Agree" button, you acknowledge that you have read and understand this agreement and agree to be bound by its terms and conditions.

CAPTCHA

This question is for testing whether you are a human visitor and to prevent automated spam submissions.

senior, dreams (F)

reCAPTCHA™
stop spam.
read books.

Create your Twitter application

Figura 4-29 Creación de una aplicación en Twitter

En la Figura 4-30 se muestran los accesos que se ingresan en el código Python.

The screenshot shows the Twitter Developers interface for an application named 'PyMG'. The page is divided into several sections: 'Details', 'Settings', 'OAuth tool', '@Anywhere domains', 'Reset keys', and 'Delete'. The 'Settings' tab is active and contains the following information:

- Organization:** Information about the organization or company associated with your application. This information is optional. Fields for 'Organization' and 'Organization website' are both set to 'None'.
- OAuth settings:** Your application's OAuth settings. Keep the "Consumer secret" a secret. This key should never be human-readable in your application. Fields include:
 - Access level:** Read and write (A)
 - Consumer key:** [Redacted] (B)
 - Consumer secret:** [Redacted] (C)
 - Request token URL:** https://api.twitter.com/oauth/request_token
 - Authorize URL:** https://api.twitter.com/oauth/authorize
 - Access token URL:** https://api.twitter.com/oauth/access_token
 - Callback URL:** None
- Your access token:** Use the access token string as your "oauth_token" and the access token secret as your "oauth_token_secret" to sign requests with your own Twitter account. Do not share your oauth_token_secret with anyone. Fields include:
 - Access token:** [Redacted] (D)
 - Access token secret:** [Redacted] (E)
 - Access level:** Read and write (F)

At the bottom of the 'Your access token' section, there is a button labeled 'Recreate my access token'. The footer of the page includes a 'Follow @twitterapi' button and links for API Terms, API Status, Blog, Discussions, Documentation, and a note that it is a Drupal community site supported by Acquia.

Figura 4-30 accesos para Twitter

Finalmente, se tomó el código base para buscar frases publicadas por los usuarios. Estas frases fueron buscadas través de palabras claves como: programación, profesor, Python, curso, Java, entre otros, además de ser buscado por zona geográfica. Una vez realizada la

búsqueda se verificaba que las frases encontradas estuvieran relacionadas al ambiente de aprendizaje en el área de la programación, sin embargo, esta forma de obtener datos no se utilizó lo suficiente, porque de todas las frases encontradas se descartaba el 60% de las frases debido a que no cumplían con los objetivos. Las frases que fueron aprobadas para agregarse al corpus fueron etiquetadas con una emoción centrada en el aprendizaje (frustrado, aburrido, emocionado y comprometido).

4.3. Módulo de Análisis de Sentimientos.

En esta sección se describe del módulo de análisis de sentimientos el cual se encarga de determinar la polaridad (positiva, negativa) o la emoción centrada en el aprendizaje (aburrido, frustrado, emocionado y comprometido) de una frase, oración o documento.

Anteriormente el módulo de análisis de sentimiento (Barrón Estrada M. L., 2017) determinaba solamente la polaridad del texto, sin embargo, con la creación de este corpus con frases centradas en el aprendizaje se propuso adaptar este módulo para reconocer las emociones que se manejan en el corpus, por lo cual cada una de estas frases se encuentran etiquetadas con emociones como son: aburrido, frustrado, comprometido y emocionado.

Una vez adaptando el módulo de análisis de sentimientos se realizaron diversos entrenamientos para mejorar la precisión.

El módulo de análisis de sentimientos se desarrolló usando el clasificador Bernoulli Naive Bayes, debido a que es de los clasificadores más simples y comúnmente utilizados sin embargo es posible utilizar otros clasificadores como: Multinomial Naive Bayes, Support Vector Machine, Linear Support Vector Machine, Stochastic Gradient Descent Classifier, and K-Nearest Neighbors cambiando el código fuente.

Para probar el funcionamiento del módulo de análisis de sentimientos, se utilizó un corpus de frases etiquetadas con su respectiva emoción centrada en el aprendizaje las cuales fueron recolectadas de diversas fuentes como: SERE (Barrón Estrada M. C., 2017), Twitter, plataformas educativas y del corpus TASS (Villena Román, 2013).

Hay una serie de métricas que son utilizadas para estimar la calidad del algoritmo clasificador que es una parte importante del modelo de aprendizaje. El método más simple para calcular la efectividad de un clasificador es la medida de precisión que calcula el porcentaje del

documento de texto correctamente clasificado sobre el total de documentos a clasificar. Para obtener este valor se aplicó una técnica de validación cruzada con un conjunto de datos con un 90% para los datos de entrenamiento y un 10% para los datos de prueba. En la Figura 4-31 muestra en forma general el proceso que utiliza el módulo de análisis de sentimientos.

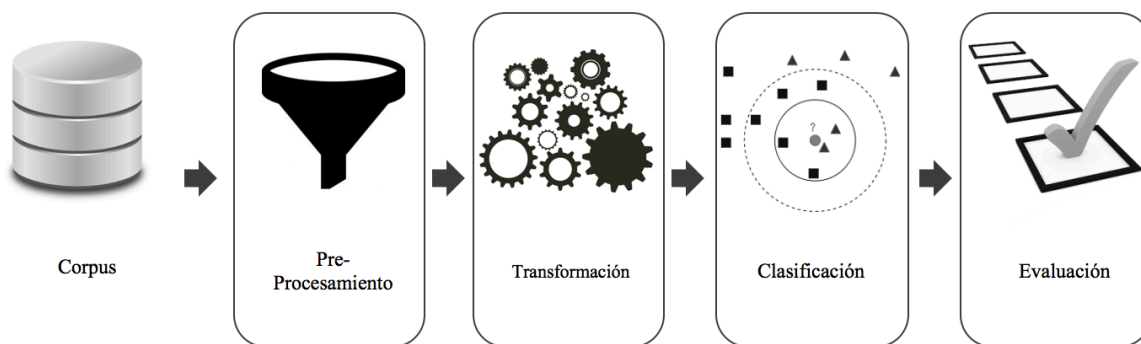


Figura 4-31 Algoritmo del Módulo de Análisis de Sentimientos.

- A. Corpus:** se requiere un corpus de opiniones textuales como entrada al módulo.
- B. Pre-pocesamiento:** para cada una de las entradas en el corpus, se realizan diferentes pasos para normalizar las frases. Los pasos de pre-procesamiento son:
 1. **Slang terms:** Trasladar los términos de jerga y emoticones a lo equivalente en texto.
 2. **Tokenizador:** Convertir las sentencias en palabras removiendo puntos y signos.
 3. **Stop-words:** Remover las palabras innecesarias.
 4. **Stemming:** Reducir las palabras a su palabra raíz.
- C. Transformación:** se genera una matriz TF-IDF que calcula la frecuencia de término (número de veces que aparece un término dado en un documento o conjunto de datos) y la frecuencia inversa de documento (número de documentos en los que aparece un término dado) para cada palabra en el corpus. Esto se conoce como extracción de características de ponderación.
- D. Clasificación:** define una función para predecir la etiqueta ingresada como entrada.
- E. Evaluación:** evalúa el modelo de la máquina de aprendizaje para predecir la emoción (aburrido, frustrado, emocionado y comprometido) de un texto de entrada; esto permite encontrar un modelo confiable.

4.4. Servicio.

SentiText es un servicio web que fue implementado para evaluar una opinión en texto y determinar su polaridad y emoción centrada en el aprendizaje utilizando diversas técnicas. La Figura 4-32 muestra como el usuario puede interactuar con la aplicación web usando diferentes dispositivos electrónicos, con el objetivo de ingresar una frase y que sea evaluada por el módulo de análisis de sentimiento en tiempo real y al mismo tiempo sean almacenadas en una base de datos.

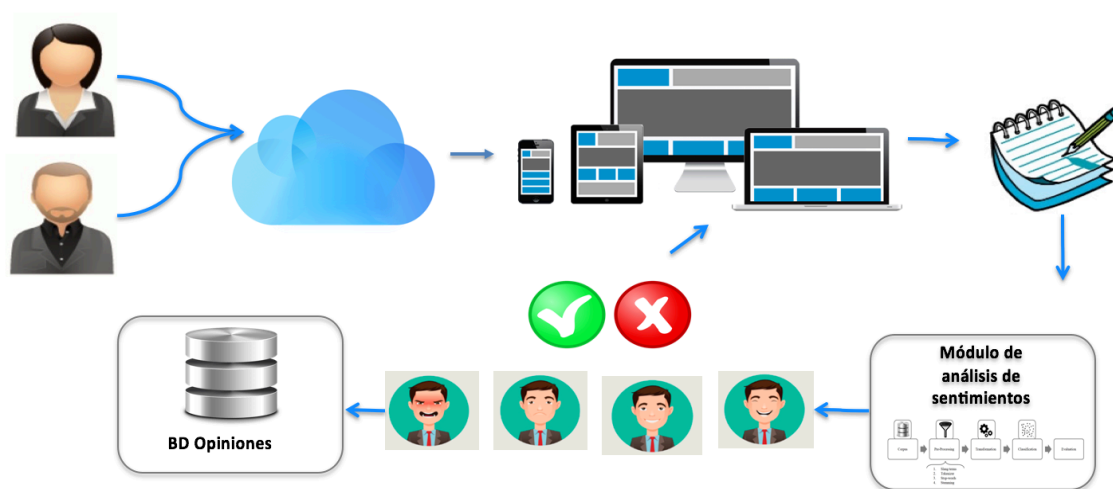


Figura 4-32 Funcionamiento de SentiText

4.4.1 Metodología de desarrollo de SentiText

Esta sección describe el método de cascada, en el cual consiste en una serie de etapas ejecutadas en forma sucesiva, la etapa siguiente empieza cuando termina la etapa anterior. Este modelo se utilizó para el desarrollo del sistema SentiText porque la documentación es muy exhaustiva, al ser un proyecto estructurado ayuda a comprender mejor el proyecto. Las fases que componen esta metodología se muestran en la Figura 4-33.

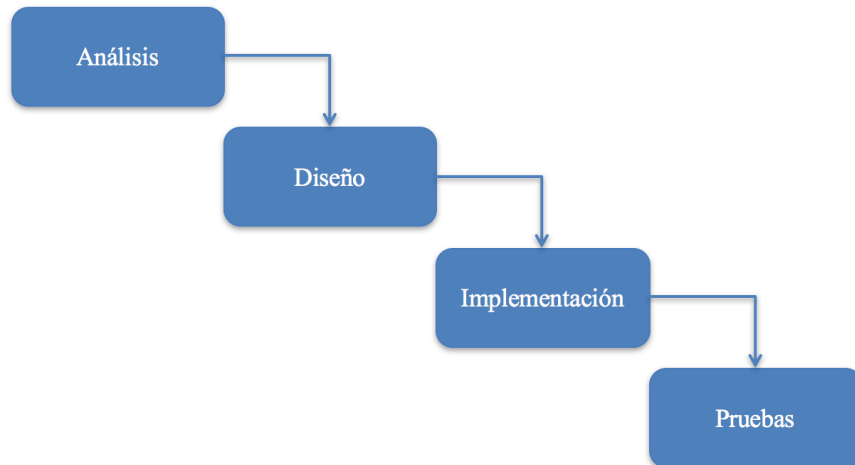


Figura 4-33 Metodología de cascada para el desarrollo de SentiText

- **Análisis:** En esta fase se analizan las necesidades del software para determinar qué objetivos debe cubrir esta aplicación Web.
- **Diseño:** En esta fase descompone y organiza el sistema en elementos que puedan elaborarse por separado, aprovechando las ventajas del desarrollo en equipo. Se define la arquitectura que conforman los componentes físicos y lógicos del sistema.
- **Implementación:** En esta fase se implementa el código fuente.
- **Pruebas:** En esta fase se verifica que los elementos, ya programados, funcionen correctamente y que cumple con los requisitos, antes de ser entregado al usuario final.

4.4.2 Análisis del Sistema

En esta etapa se consideran elementos importantes para el desarrollo de SentiText, como son los requisitos de funcionalidad y calidad, así como los actores con los que interactúa el sistema.

4.4.2.1 Requisitos Funcionales y de Calidad

En esta etapa se clasifican los requerimientos funcionales y de calidad haciendo un análisis detallado de la arquitectura, se selecciona diferentes requisitos, además se priorizaron con escalas (alta, media y baja). En la Tabla 4-6 se muestra los siete requerimientos funcionales con su respectiva descripción y prioridad.

Tabla 4-6 Requerimientos funcionales SentiText

| No. | Descripción | Prioridad |
|--------------|---|--------------|
| RF-01 | El sistema debe tener un apartado donde pueda el usuario ingresar su frase y este sea evaluado. | ALTA |
| RF-02 | El sistema debe poder proporcionar el corpus que este etiquetado con emociones centradas en el aprendizaje. | ALTA |
| RF-03 | El sistema debe poder proporcionar el corpus que este etiquetado con polaridad. | ALTA |
| RF-04 | El sistema debe almacenar todas las frases que son evaluadas por el Servicio SentiText. | ALTA |
| RF-05 | El sistema debe mostrar el resultado al usuario de la polaridad y emoción que contenga su frase | ALTA |
| RF-06 | El sistema debe mostrar información detallada de cómo se encuentra el corpus de emociones centradas en el aprendizaje, es decir el total de frases, como están distribuidas, como se obtuvieron entre otra información. | MEDIA |
| RF-07 | El sistema debe mostrar información detallada de cómo se encuentra el corpus de polaridades, es decir el total de frases, como están distribuidas, como se obtuvieron entre otra información. | MEDIA |

Además, se presentan los atributos de calidad para el sistema SentiText donde se clasifican en rendimiento, disponibilidad, portabilidad, y usabilidad. Se aplicó la misma técnica de priorización que en los requerimientos funcionales. Se clasificaron según el atributo de calidad. En la Tabla 4-7 se muestra los seis requerimientos de calidad con su respectivo atributo, descripción y prioridad.

Tabla 4-7 Requerimientos de Calidad SentiText

| ID | Atributo | Requisito | Prioridad |
|--------------|----------------|---|--------------|
| RC-01 | Rendimiento | El sistema deberá responder en máximo 3 segundos a cada petición de servicio | Alta |
| RC-02 | Disponibilidad | El sistema deberá estar disponible 24/7 | Alta |
| RC-04 | Rendimiento | El tiempo esperado de carga de imágenes y documentos no excederá los 5 segundos. | Media |
| RC-05 | Portabilidad | El sistema deberá funcionar en cualquier navegador web. | Media |
| RC-06 | Usabilidad | El sistema cuenta con una interfaz gráfica amigable donde el usuario se puede adaptar máximo en 10 minutos. | Baja |

4.4.2.2. Restricciones

Durante el análisis de requerimientos se establecieron algunas restricciones, las cuales son:

- Todas las actividades del sistema se realizarán en línea a través del portal web.
- SentiText solo estará disponible en el idioma español (México).

4.4.2.3. Actores

Los actores son aquellos que interactúan con el sistema por lo cual se identificó un conjunto de personas representativas que pueden interactuar con el sistema y las tareas que deben realizar. En la Tabla 4-8 se describe brevemente los tipos de actores y su principal funcionalidad.

Tabla 4-8 Actores en SentiText

| Actor | Descripción |
|---------|--|
| Usuario | Es el usuario que proporciona una frase que será evaluada por el clasificador. |

En la Figura 4-34 se muestra gráficamente la relación donde el actor usuario proporciona frases para ser evaluadas. El actor Administrador de SentiText realiza cambios en el sistema, además administra la información proporcionada por usuarios de forma que puedan verificar que si la frase ingresada puede ser utilizada para incorporarse al corpus.

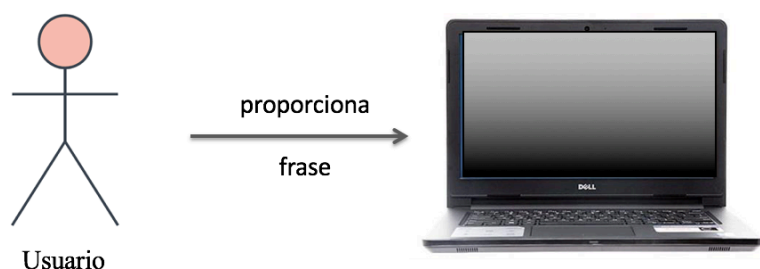


Figura 4-34 Actores en SentiText

4.4.2.4. Casos de uso

Un caso de uso describe las actividades que deberán realizarse para llevar a cabo algún proceso mostrando la funcionalidad principal del sistema. Éstos se definieron con base en el

análisis de requerimientos y restricciones con la finalidad de establecer los mecanismos que le dan interacción al sistema. En la Figura 4-35 se muestra el diagrama de casos de uso. Además, existe una entidad externa el cual realiza dos procesos manuales como administrar SentiText y Actualizar la información

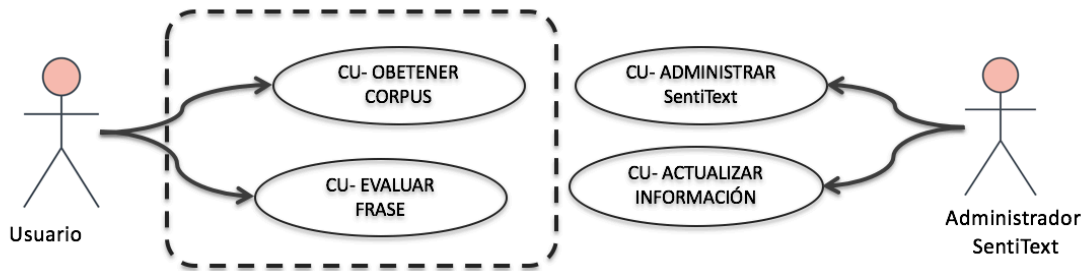


Figura 4-35 Diagrama de Casos de Uso de SentiText.

A continuación, se explican de forma más detallada los casos de usos que se desarrollaron para SentiText.

| | |
|--------------------|--|
| Caso de uso | CU-01.- OBTENER CORPUS |
| Actores | Usuario |
| Función | Es un módulo servirá para obtener los diferentes corpus con los que se cuentan y así puedan utilizar los datos que se recolectaron. |
| Descripción | El sistema deberá contar con un apartado donde tenga los diferentes corpus. |
| Referencia | <ul style="list-style-type: none"> El sistema debe poder proporcionar el corpus que este etiquetado con polaridad. (RF-03) El sistema debe poder proporcionar el corpus que este etiquetado con emociones centradas en el aprendizaje. (RF-02) |

| | |
|--------------------|---|
| Caso de uso | CU-02.- EVALUAR FRASE |
| Actores | Usuario |
| Función | Es un módulo donde el usuario podrá ingresar una frase que será evaluada por el clasificador. |
| Descripción | El sistema deberá contar con una interfaz de evaluación de frases. |
| Referencia | <ul style="list-style-type: none"> El sistema debe tener un apartado donde pueda el usuario ingresar su frase y esta sea evaluada. (RF-01) El sistema debe almacenar todas las frases que son evaluadas por el Servicio SentiText. (RF-04) El sistema debe retornar al usuario el resultado de polaridad y emoción detectada en la frase (RF-05) |

| | |
|-----------------------|--|
| Proceso manual | CU-03.- ADMINISTRAR SENTITEXT |
| Actores | Administrador SentiText |
| Función | En este módulo el administrador podrá |
| Descripción | El administrador deberá revisar que todo esté funcionando adecuadamente. |
| Referencia | <ul style="list-style-type: none"> • El sistema debe estar funcionando correctamente. |

| | |
|-----------------------|--|
| Proceso manual | CU-04.- ACTUALIZAR INFORMACIÓN |
| Actores | Administrador SentiText |
| Función | En este módulo el administrador deberá actualizar los datos proporcionados a los usuarios. |
| Descripción | Cada cierto periodo el administrador deberá actualizar la información del apartado donde se le proporcionará al usuario los corpus y la información de estos. |
| Referencia | <ul style="list-style-type: none"> • El sistema debe mostrar información detallada de cómo se encuentra el corpus de emociones centradas en el aprendizaje, es decir el total de frases, como están distribuidas, como se obtuvieron entre otra información. (RF-06) • El sistema debe mostrar información detallada de cómo se encuentra el corpus de polaridades, es decir el total de frases, como están distribuidas, como se obtuvieron entre otra información. (RF-07) |

4.4.3. Diseño

A continuación, se describe parte del Diseño que se elaboró con base en la etapa de análisis del sistema.

4.4.3.1. Diagrama de contexto

En esta sección se muestra el diagrama de contexto del Sistema SentiText el cual contiene una sola interfaz para los usuarios, ésta ayuda con la comunicación con otras entidades. La Figura 4-36 muestra la dependencia con la Base de Datos donde se encuentran alojadas las frases proporcionadas por los usuarios, además se tiene un repositorio que contiene el corpus con las frases etiquetadas

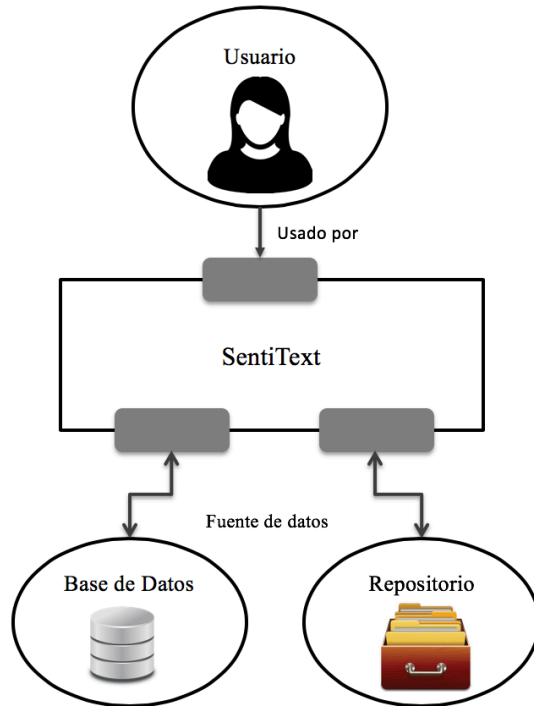


Figura 4-36 Diagrama de contexto SentiText

SentiText se encuentra ubicado en el centro del diagrama de contexto y a su alrededor aparecen diversas entidades con las que interactúa el sistema. Estas entidades se describen en la Tabla 4-9.

Tabla 4-9 Entidades de SentiText

| Actor | Descripción |
|----------------------|--|
| Usuario | Es el usuario que proporciona una frase para ser evaluada y almacenada en la base de datos. |
| Base de datos | Almacena las frases proporcionadas por los usuarios. |
| Repositorio | Contiene el corpus de polaridades y emociones las cuales serán proporcionados a los usuarios que la requieran. |

4.4.3.2. Arquetipos

Los arquetipos son entidades abstractas identificadas que indican el comportamiento de SentiText. En la Tabla 4-10 se describen los arquetipos y en la Figura 4-37 se muestran las relaciones entre los arquetipos.

Tabla 4-10 Arquetipos del Sistema SentiText

| Arquetipo | Descripción |
|---------------------|--|
| Clasificador | Es el artefacto que se encarga de determinar la polaridad (positiva, negativa) o la emoción centrada en el aprendizaje (aburrido, frustrado, emocionado y comprometido) de una frase. |
| Opinión | Es una frase escrita por el usuario. |
| Etiqueta | Es la emoción que contiene una frase; estas etiquetas pueden ser clasificadas en cuatro emociones centradas en el aprendizaje (aburrido, frustrado, emocionado y comprometido) y la polaridad (positivo y negativo). |
| Corpus | Es una base de datos que contiene las frases y etiquetas que se han recolectado y se usa para entrenar al clasificador. |

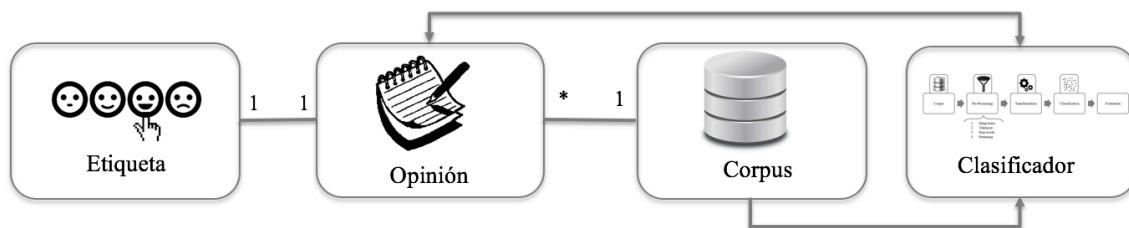


Figura 4-37 Arquetipos de SentiText

4.4.3.3. Diferentes vistas del sistema

A continuación, se da una breve explicación de la arquitectura que se utilizó para el desarrollo de SentiText.

Vista Física

La vista física toma en cuenta requerimientos no funcionales por lo que se utilizó el patrón de diseño llamado Modelo Vista Controlador (MVC) el cual separa los datos de una aplicación, la interfaz de usuario y la lógica de control en tres componentes distintos. En la Figura 4-38 se muestra la arquitectura de SentiText.

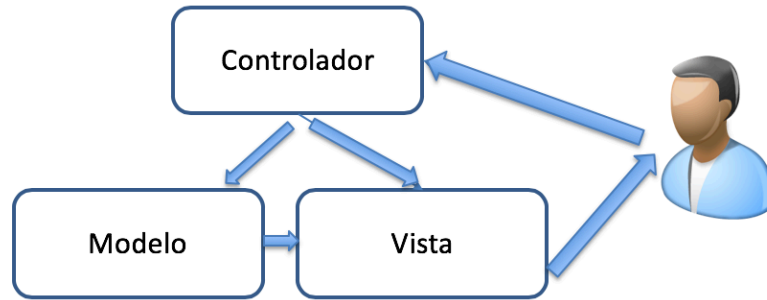


Figura 4-38 Arquitectura MVC SentiText

Vista lógica

En MVC se trata de un modelo muy maduro y que ha demostrado su validez a lo largo de los años en todo tipo de aplicaciones, y sobre multitud de lenguajes y plataformas de desarrollo. En la Figura 4-39 se muestra la arquitectura por capas del Sistema SentiText.

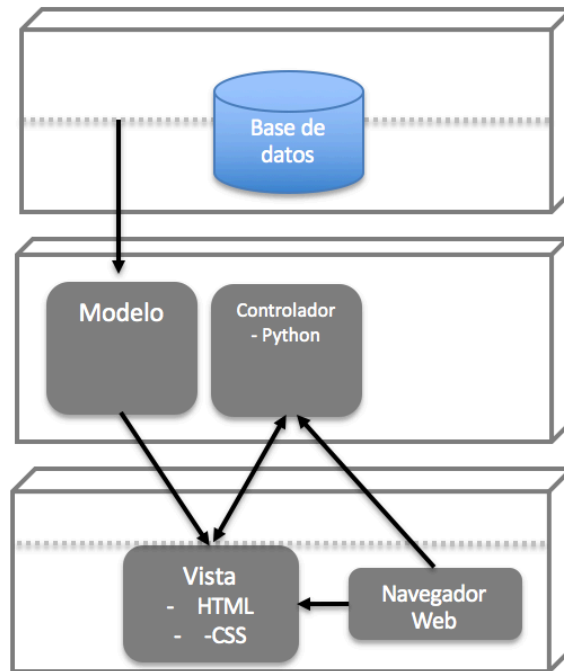


Figura 4-39 Vista Lógica del Sistema SentiText

- **Modelo** esta capa contiene una representación de los datos que maneja el sistema y sus mecanismos de persistencia, sin embargo, este proyecto utiliza esta capa para tener acceso insertar las frases que son evaluadas por el usuario en la base datos.
- **Vista**, contiene las interfaces que se le muestran al usuario; estas interfaces se muestran en la Figura 4-42, Figura 4-43, y Figura 4-44.
- **Controlador**, es aquel que actúa como intermediario entre el Modelo y la Vista, gestionando el flujo de información entre. En esta capa se encuentran dos módulos de análisis de sentimientos: uno de ellos utiliza un clasificador tradicional (Barrón Estrada M. C., 2017); y el otro módulo utiliza Deep Learning para determinar polaridad posteriormente utiliza un clasificador tradicional para las emociones centradas en el aprendizaje (Oramas B., 2018).

Vista de desarrollo

En el desarrollo del sistema se consideraron dos partes: la primera va enfocada a la parte del usuario la cual hace peticiones al servidor y la segunda la parte del servidor la cual arroja una respuesta al usuario pasando por el sistema. En la Figura 4-40 se muestra la vista de desarrollo.

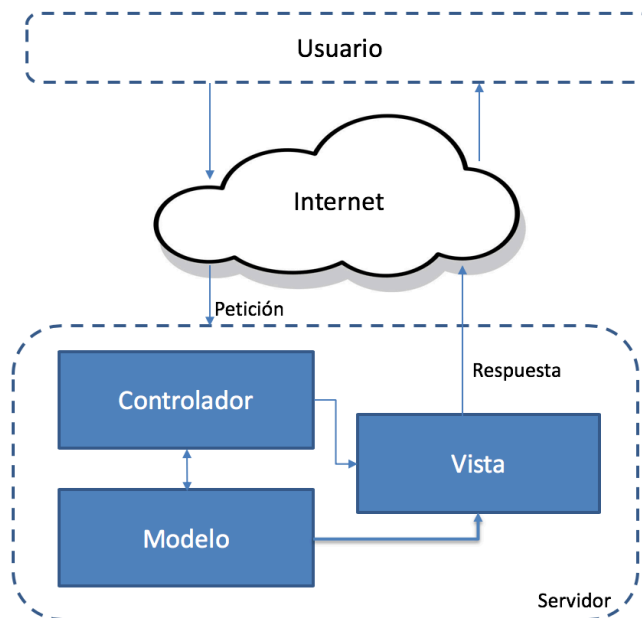


Figura 4-40 Vista de Desarrollo de SentiText

4.4.4 Implementación

SentiText funciona en entorno web, por lo cual se explica las herramientas que se utilizaron y el desarrollo de esta herramienta.

4.4.4.1. Lenguajes y herramientas software

SentiText es una aplicación web desarrollada en Flask. Flask es un micro Framework para Python basado en Werkzeug, Jinja 2 y concebido para facilitar el desarrollo de Aplicaciones Web bajo el patrón MVC. SentiText fue diseñado con dos diferentes objetivos: Evaluar la opinión en texto ingresada por el usuario y determinar su clasificación; y proporcionar más frases de texto a la base de datos que maneja etiquetas de polaridad (positivo y negativo) y de emociones centradas en el aprendizaje (aburrido, frustrado, emocionado y comprometido) que se utilizaron para el entrenamiento del módulo de análisis de sentimientos. En la Figura 4-41 se muestran el lenguaje que se utilizó para la elaboración de esta aplicación Web.



Figura 4-41 Lenguajes de SentiText.

4.4.4.2. Interfaces del sistema

En esta sección se describen las diferentes interfaces desarrolladas para que el usuario interactúe con SentiText.

SentiText se compone por diferentes interfaces principales: En la Figura 4-42 se muestra la interfaz inicial donde se da una breve introducción del objetivo de dicha página, además proporciona información para el uso de la información de los corpus.



Figura 4-42 Interfaz principal SentiText

En la Figura 4-43 se muestra la interfaz que el usuario utilizará para evaluar una frase. La frase se ingresa en el campo correspondiente y el usuario selecciona la forma de evaluar dicha frase (Deep learning o SVM); el botón *Evaluar* enviará la frase al módulo de análisis de sentimientos seleccionado para ser evaluada. La frase evaluada también será almacenada en la base de datos.

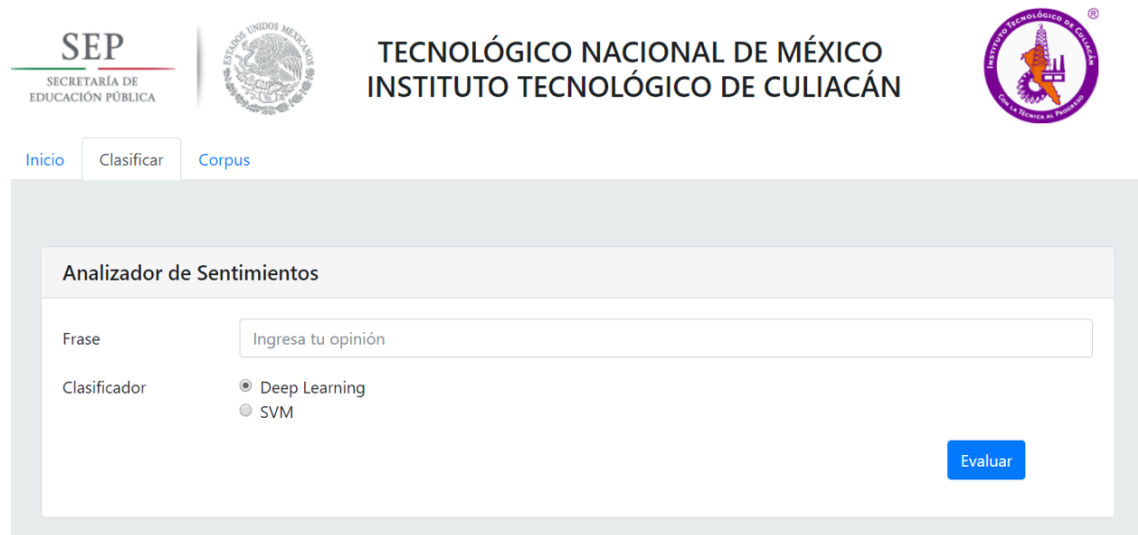


Figura 4-43 Interfaz de Analizador de sentimientos SentiText

En la Figura 4-44 se muestra la información correspondiente al corpus, es decir cómo se encuentran distribuidas las frases, el número total de frases. Además de proporcionar los corpus en los cuales se han trabajado para entrenar el módulo de Análisis de Sentimientos.

SEP
SECRETARÍA DE
EDUCACIÓN PÚBLICA

ESTADOS UNIDOS MEXICANOS

TECNOLÓGICO NACIONAL DE MÉXICO
INSTITUTO TECNOLÓGICO DE CULIACÁN

INSTITUTO TECNOLÓGICO DE CULIACÁN
CON LA TÉCNICA AL SERVIDOR

Inicio Clasificar Corpus

Los datasets que se encuentran disponibles en este sitio son para un uso estrictamente educativos; se prohíbe su uso comercial o con fines de lucro. El proceso de etiquetado se llevo a cabo en forma manual con un equipo de expertos de nuestro laboratorio.

El dataset SentiText tiene 13010 opiniones con la polaridad positiva-negativa. El dataset EduERAS tiene 7386 opinions marcadas con emociones centradas en el aprendizaje, tal como: frustrado (2082), aburrido (1361), neutral (945), emocionado (1697) y enganchado (1301).

Por favor citar, el siguiente artículo si usas alguno de nuestros datasets en tu proyecto de investigación: María Lucia Barrón-Estrada M.L. & Ramón Zatarain-Cabada R. & Oramas-Bustillos R. & Ramírez-Ávila S.L. (2017) Building a Corpus of Phrases Related to Learning for Sentiment Analysis

Información: roramas@gmail.com, luciabarron@gmail.com, sandra_ramirez_avila@hotmail.com

Emociones Polaridad

Figura 4-44 Interfaz de Información de corpus SentiText

Capítulo 5

5. Resultados

En esta sección se presentan los diferentes resultados como: la obtención de frases ingresadas en las diferentes plataformas, la evolución del corpus y los resultados del analizador de sentimientos utilizando los algoritmos de clasificación tradicionales y el clasificador desarrollado en el INFOTEC Aguascalientes.

5.1. Resultados de las herramientas para la obtención de frases

En esta sección se presentan los resultados obtenidos a través de los diferentes instrumentos que se utilizaron para la creación del corpus de frases en español y centradas en el aprendizaje. En la Figura 5-1 se muestra la distribución del corpus, es decir la cantidad de frases recolectadas con cada uno de los instrumentos.

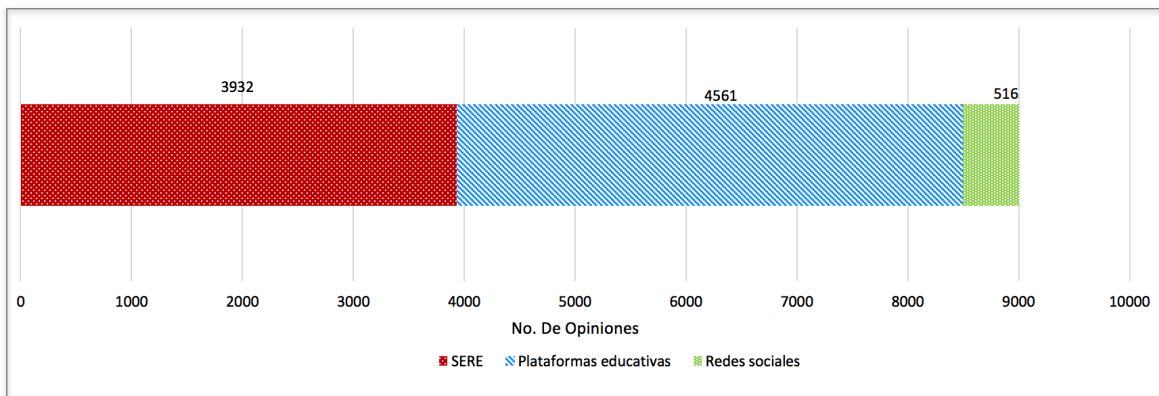


Figura 5-1 Resultado de la obtención de opiniones utilizando los diferentes instrumentos

5.1.1. Sistema de Evaluación de Recursos Educativos (SERE).

Este proceso inició en agosto del 2017, donde participaron 53 estudiantes (45 hombres y 8 mujeres) de la carrera de Ingeniería en Sistemas Computacionales del Instituto Tecnológico de Culiacán, utilizando SERE con los temas de la unidad 1 de la asignatura llamada Fundamentos de Programación, ejercicios de estructura secuencial, selección y repetitiva.

Después de estudiar cada subtema en formato de texto e imágenes o multimedia (video) o elaborar programas, el estudiante registró su opinión respecto al contenido del tema estudiado, además de etiquetar su opinión con un emoticón que representaba su estado emocional: aburrido, frustrado, emocionado o comprometido. En la Figura 5-2 se muestra como el estudiante interactúa con el sistema.

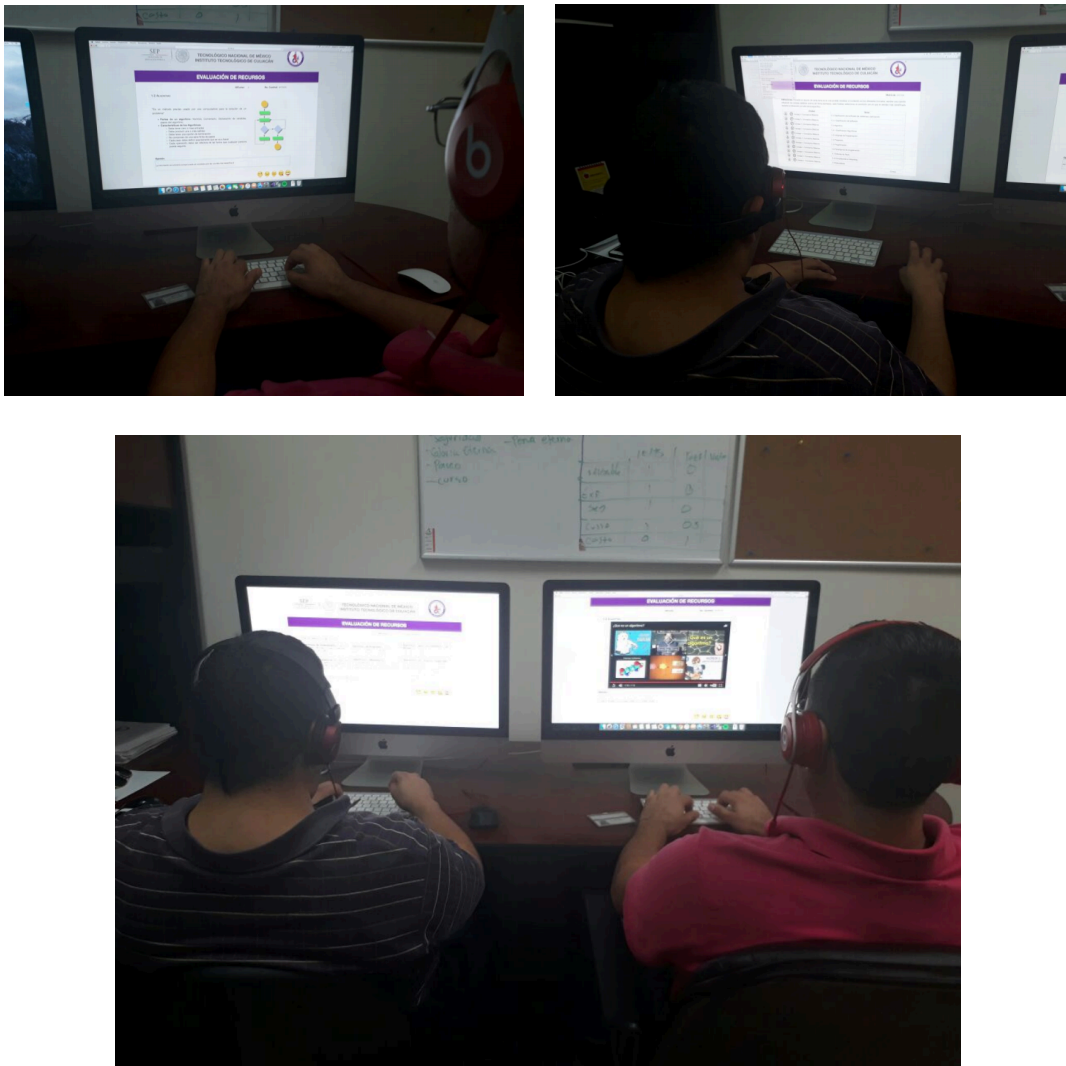


Figura 5-2 Obtención de opiniones en SERE

Actualmente el total de opiniones registradas por los estudiantes es de 3,932 opiniones, incluyendo las generadas en el desarrollo de proyectos prácticos. En la Tabla 5-1 se muestran algunos ejemplos de las opiniones registradas por los estudiantes que participaron en el estudio.

Tabla 5-1 Frases recolectadas con el Sistema SERE.

| Clave | Opinión en español | Formato | Evaluación |
|-------|--|------------------|--------------|
| E-15 | Me gusto bastante el vídeo | Video | Comprometido |
| E-64 | No me gusto que las voces fueran de España, además los gráficos de la animación están algo feos. | Video | Frustrado |
| E-67 | Quizás con un ejemplo quedaría más claro. | Imagen- Texto | Aburrido |
| E-71 | Hubiese sido mejor poner una tabla con sus diferencias y así compararlas y sea más diverso. | Imagen- Texto | Emocionado |

5.1.2. Proceso manual

Este proceso describe dos formas de obtener opiniones de los diferentes usuarios. La primera es ingresar a las diferentes plataformas educativas mencionadas anteriormente en la sección 4.2.2 Obtención de opiniones (plataformas educativas digitales en línea), para seleccionar los cursos que estén relacionados al ámbito de programación de computadoras, con el fin de extraer las experiencias que los usuarios ingresaron en la plataforma y posteriormente etiquetar estas frases con la emoción correspondiente a la experiencia del usuario. A la fecha actual, el total de opiniones registradas por el proceso manual es de 4,561 opiniones. En la se muestran algunos ejemplos de las frases recolectadas en las diferentes plataformas

| Etiqueta | Frase | Curso |
|--------------|--|--------------------------------|
| Frustrado | Me resulta muy difícil entender el relator. | Aprende Programación C# |
| Aburrido | Algunos videos se hacen demasiado lentos, mientras que en otros no se entra en detalles en caso de que podría ser útil y se los pasa por encima. | Curso de JavaScript desde 0 |
| Emocionado | La explicación es muy clara, espero sea así hasta el final del curso. | Crear aplicaciones PHP |
| Comprometido | Una experiencia muy buena, aunque me hubiese gustado que se hiciera toda la página dinámica, fuera de eso el curso está excelente. | Crear aplicaciones PHP seguras |

educativas.

Tabla 5-2 Frases extraídas de diferentes plataformas educativas.

| Etiqueta | Frase | Curso |
|-----------------|--|--------------------------------|
| Frustrado | Me resulta muy difícil entender el relator. | Aprende Programación C# |
| Aburrido | Algunos videos se hacen demasiado lentos, mientras que en otros no se entra en detalles en caso de que podría ser útil y se los pasa por encima. | Curso de JavaScript desde 0 |
| Emocionado | La explicación es muy clara, espero sea así hasta el final del curso. | Crear aplicaciones PHP |
| Comprometido | Una experiencia muy buena, aunque me hubiese gustado que se hiciera toda la página dinámica, fuera de eso el curso está excelente. | Crear aplicaciones PHP seguras |

La segunda forma de obtener opiniones manualmente se realizó con ayuda de profesores del Instituto Tecnológico de Culiacán. En esta forma, los profesores solicitaron a los estudiantes desarrollar un proyecto práctico para resolver un problema y posteriormente se aplicó una encuesta con el fin de obtener una retroalimentación por parte de los estudiantes; una vez obtenida esta retroalimentación, las frases se pasaban al proceso de limpieza y etiquetado para finalmente agregarlas al corpus. En la Tabla 5-3 Se muestran algunos ejemplos de la retroalimentación que los estudiantes proporcionaron acerca de sus proyectos del curso de programación orientada a objetos.

Tabla 5-3 Frases proporcionadas con el desarrollo de proyectos.

| Etiqueta | Frase |
|-----------------|---|
| Comprometido | Trabaje durante bastante tiempo en el tamaño del arreglo. |
| Emocionado | No fue complicado escribir algunas clases. |
| Frustrado | Tuve complicaciones al comunicarlo con la clase persona. |
| Aburrido | Fue muy sencillo de realizar las clases. |
| Comprometido | La clase persona la hicimos en equipo sin embargo elabore esta clase como practica yo sola. |

5.1.3. Extractor de información

El proceso para extraer información de la red social Twitter, se ejecutó con el fin de buscar y seleccionar frases relacionadas al tema de programación, utilizando palabras clave y zonas geográficas para la búsqueda. Fueron seleccionadas 512 frases en total, fueron etiquetadas con una emoción relacionada con el aprendizaje y almacenada en el corpus de opiniones. En la Tabla 5-4 se presentan algunas opiniones extraídas de *Twitter* usando palabras clave.

Tabla 5-4 Frases extraídas de Twitter.

| Etiqueta | Frase | Palabras clave |
|-----------------|---|-----------------------|
| Emocionado | ¡Este curso de programación me parece muy excelente!!! Muchísimas gracias por compartir sus conocimientos... | Curso Programación |
| Comprometido | Es un curso excelente. Ha sido de mucha ayuda y voy a continuar aprendiendo del curso avanzado. Se les agradece mucho. | Curso |
| Emocionado | Muy fácil de seguir, productivo, no se hace pesado y cumple a la perfección con los cursos presenciales de la rama programación, ¡incluso más! Me ha gustado mucho, felicitaciones. | Curso Programación |
| Aburrido | no entiendo muchas cosas de una clase abstracta porque me parece un tema aburrido | Clase abstracta |
| Aburrido | una de las cosas que más me da flojera es hacer la documentación de los programas | Programa |

5.2. Evolución del corpus

La metodología presentada en la sección 4.1, fue aplicada para la creación del corpus de frases en español etiquetadas con una emoción relacionada con el aprendizaje. En esta sección se describen los resultados alcanzados.

En cada una de las iteraciones de la metodología, se realizó un registro del número de frases adquiridas cada mes desde agosto 2017. En la Figura 5-3 se muestra la evolución del tamaño del corpus a través de los diferentes meses que duró su creación.

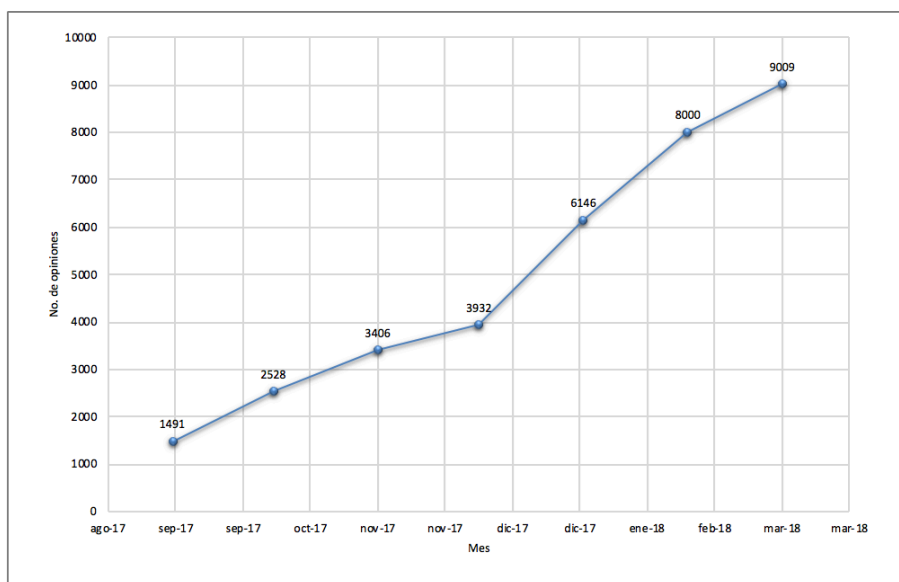


Figura 5-3 Evolución del corpus con emociones centradas en el aprendizaje.

Actualmente, el corpus contiene un total de 9,009 frases que se encuentran etiquetadas con cuatro diferentes emociones centradas en el aprendizaje (aburrido, frustrado, emocionado y comprometido). En la Figura 5-4 se muestra la distribución de las frases de acuerdo con la emoción con la que se encuentran etiquetadas en el corpus.

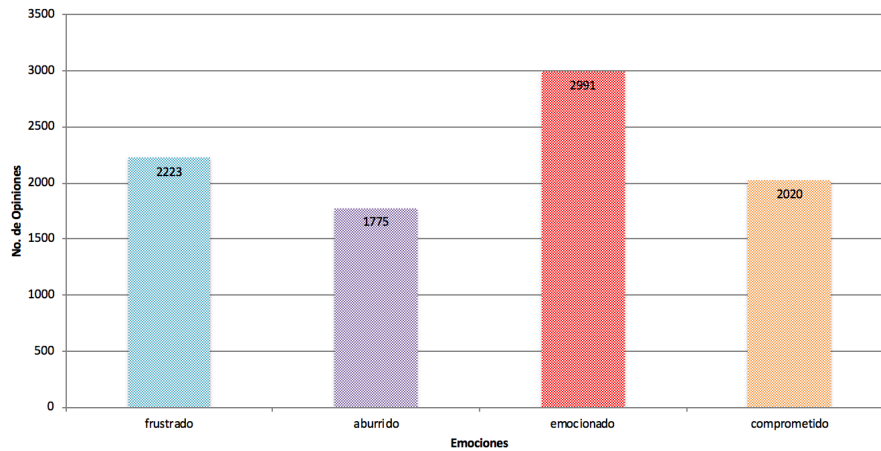


Figura 5-4 Distribución del corpus por emociones.

5.3. Evaluación del módulo de análisis de sentimiento

Se utilizaron una serie de métricas para estimar la calidad del algoritmo de clasificación que es parte del modelo de aprendizaje propuesto. En este caso, se usó el método más simple para calcular la efectividad de un clasificador, que es la medida de precisión. Esta calcula el porcentaje de documentos de texto correctamente clasificados sobre el total de documentos a clasificar. Para obtener este valor, se aplicó una técnica de validación cruzada con un conjunto de datos en español con un 90% para los datos de entrenamiento y un 10% para los datos de prueba.

La evaluación del módulo de análisis de sentimientos se divide en dos secciones, las cuales se explican en las secciones 5.3.1 y 5.3.2. Para clasificar cada una de estas secciones se utilizaron diferentes clasificadores como: Multinomial Naive Bayes, Support Vector Machine, Linear Support Vector Machine, Stochastic Gradient Descent Classifier, and K-Nearest Neighbors.

5.3.1. Corpus de frases etiquetadas con polaridad

Inicialmente se trabajó con el corpus de frases etiquetadas con polaridad, el cual contiene frases relacionadas a la educación. Sin embargo, no tiene un área específica, es decir trabaja con frases de estudiantes de diversas áreas como: psicología, enfermería, administración, entre otras. Actualmente el corpus contiene 14,958 frases, las cuales fueron recolectadas de diferentes formas y el 45% de estas se tomaron de SERE. La Figura 5-5 muestra la

distribución del corpus por polaridad, el cual contiene 8,161 opiniones positivas y 6,797 negativas.

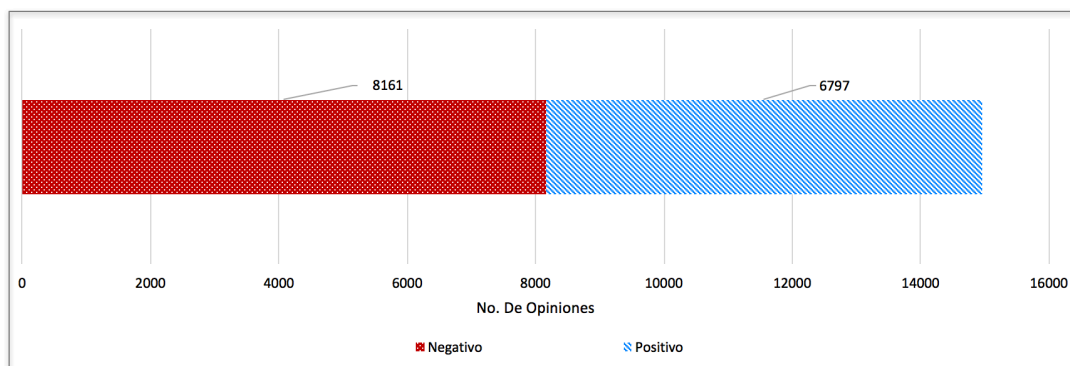


Figura 5-5 Distribución del corpus por Polaridad.

El clasificador con la precisión más alta fue el algoritmo ***Bernoulli Naïve Bayes*** con una exactitud (*accuracy*) del **83.56%**. En la Tabla 5-5 se presentan los valores obtenidos utilizando los diferentes clasificadores.

Tabla 5-5 Valores obtenidos con el clasificador utilizando el corpus de polaridad

| Clasificador | % Accuracy |
|---------------------|-------------------|
| Bernoulli NB | 83.56% |
| Multinomial NB | 75.31% |
| SVC | 75.79% |
| Linear SVC | 74.69% |
| SGDC Classifier | 76.69% |
| KNN | 68.46% |

5.3.2. Corpus de frases etiquetadas con emociones centradas en el aprendizaje.

Este corpus contiene 9,009 frases recolectadas por los diferentes instrumentos como se expuso en la sección 4.2. El módulo de análisis de sentimientos se adaptó para reconocer emociones centradas en el aprendizaje.

A continuación, se presentan los resultados generados por el clasificador utilizando el corpus con etiquetas de emociones centradas en el aprendizaje. La Tabla 5-6 muestra los resultados obtenidos por los algoritmos tradicionales.

Tabla 5-6 Valores obtenidos con el clasificador utilizando el corpus de emociones

| Algoritmo clasificador | % <i>Accuracy</i> |
|--|--------------------------|
| Linear Support Vector Machine | 60.29% |
| Stochastic Gradient Descent Classifier | 60.07% |
| Bernoulli Naive Bayes | 57.30% |
| Multinomial Naive Bayes | 55.30% |
| Support Vector Machine | 31.53% |

Para el caso del corpus de emociones centradas en el aprendizaje el clasificador con el *accuracy* más alto fue Linear Support Vector Machine con una exactitud (*accuracy*) del **60.29%**.

5.4. Evaluación del módulo de análisis de sentimientos EvoDAG

Este sistema fue desarrollado en INFOTEC durante una estancia de investigación, con la finalidad de que trabaje con frases etiquetadas con polaridad (positivo, negativo y neutral), además trabaja con diferentes corpus como SENTIPOLC 2014 – italiano, TASS 2015 - Spanish, SemEval 2015 – English, SemEval 2016 – English. Estos corpus están elaborados en diferentes idiomas como inglés, español, italiano, además se encuentran etiquetados con emociones básicas como: miedo, ira, tristeza, felicidad.

Una de las tareas desempeñadas durante la estancia fue revisar el funcionamiento de este sistema y la adaptación del corpus con la finalidad de probar otros clasificadores que trabajen con emociones centradas en el aprendizaje como frustrado, aburrido, emocionado y comprometido. Este proceso de adaptación consistió en cambiar el formato de las frases. En la Figura 5-6 se muestra el formato utilizado anteriormente para entrenar a los clasificadores tradicionales.

frustrado| pues todo bien pero estuvo complicado de una parte.
aburrido| Dura demasiado para dar una idea clara de lo que quiere transmitir.
comprometido| la manera de explicar es algo que queda muy claro, la velocidad utilizada es la adecuada.
emocionado| Porque me daba los valores tal como habida calculado.

Figura 5-6 Formato de las oraciones de los clasificadores tradicionales.

Sin embargo, para el clasificador EvoDAG se utilizó un formato JSON que contiene dos etiquetas llamadas “text” y “klass”, que representan la frase escrita por el usuario y la emoción de dicha frase respectivamente. En la Figura 5-7 se muestra un ejemplo del formato que utiliza el clasificador EvoDAG.

```
{"text": "pues todo bien pero estuvo complicado de una parte", "klass": "frustrado"}  
{"text": "Dura demasiado para dar una idea clara de lo que quiere transmitir", "klass": "aburrido"}  
{"text": "la manera de explicar es algo que queda muy claro, la velocidad utilizada es la adecuada", "klass": "comprometido"}  
{"text": "Porque me daba los valores tal como habida calculado.", "klass": "emocionado"}
```

Figura 5-7 Formato de las oraciones utilizando EVODAG

Una vez que se adaptó el corpus a la forma en que trabaja el sistema EvoMSA se evaluó el corpus de emociones centradas en el aprendizaje en el clasificador EVODAG. La técnica que se utilizó es validación cruzada en donde utiliza el 60% para los datos de entrenamiento y un 40% para los datos de prueba. El resultado de este clasificador obtuvo una precisión de 65.30%. Este resultado indica que EvoDAG es muy competitivo y supera estadísticamente a los otros clasificadores.

Capítulo 6

6. Conclusiones y trabajo futuro

La educación se reconoce como un derecho de los seres humanos desde el siglo XIX en México y el Estado está obligado a proveer educación básica a todos los ciudadanos; sin embargo en nuestro país existe un rezago educativo reconocido por el INEGI¹ que en sus informes estadísticos mostró que más de 4,700,000 mexicanos mayores de 15 años no saben leer ni escribir.

Además del reto de cobertura de la educación para llegar a la totalidad de la población, el Estado Mexicano enfrenta el reto de la calidad, ya que en diferentes pruebas nacionales e internacionales como ENLACE y PISA² queda constancia de las deficiencias de la enseñanza demostrando que gran parte de los estudiantes de nivel básico y medio (primaria y secundaria) no alcanzan niveles satisfactorios en las evaluaciones de sus aprendizajes; lo que indica que no son capaces de resolver problemas que los niños y jóvenes de otros países resuelven sin dificultad.

Actualmente, la tecnología ha irrumpido en las más diversas actividades que realizan las personas, quienes son capaces de utilizar una diversidad de dispositivos electrónicos como PC, laptops, tabletas y teléfonos inteligentes para realizar sus tareas más cotidianas como escuchar música, socializar, ordenar comida, tomar fotografías, o pedir un Uber.

Estos dispositivos también son utilizados en el área educativa para estudiar y aprender diversidad de temas a través de aplicaciones o sistemas computacionales que se encargan de administrar los contenidos de un curso de acuerdo a las necesidades de los estudiantes. Hoy en día, estos sistemas requieren conocer no solo el estado cognitivo de los estudiantes sino también su estado afectivo, ya que se ha demostrado que las emociones juegan un papel importante durante los procesos de aprendizaje de las personas.

¹ INEGI, 2015 <http://cuentame.inegi.org.mx/poblacion/analfabeta.aspx?tema=P>

² EL PAIS, 2016 https://elpais.com/internacional/2016/12/06/mexico/1481045534_791430.html

En este capítulo se presentan las conclusiones del trabajo de tesis desarrollado así como se mencionan algunos trabajos a futuro que se pueden realizar.

6.1. Conclusiones

Este trabajo generó principalmente cinco productos entre los que se encuentran: un corpus afectivo que contiene frases escritas en español y etiquetadas con polaridad y una emoción centrada en el aprendizaje, un sistema clasificador de frases que recibe como entrada una frase y genera como salida la emoción centrada en el aprendizaje que se detectó, una metodología iterativa para generar el corpus afectivo que consta de seis pasos, un sistema para la recolección de frases escritas por estudiantes al visualizar recursos educativos y un servicio web que sirve para establecer una comunicación entre un usuario y el sistema de clasificación de frases.

A continuación se describen las conclusiones para cada uno de estos productos.

El corpus afectivo contiene 9,009 frases escritas en español y etiquetadas con una emoción relacionada con el aprendizaje (aburrido, frustrado, emocionado y comprometido) así como su polaridad (positiva o negativa). Para el desarrollo de este corpus se diseñó y ejecutó una metodología con el fin de recolectar las opiniones de estudiantes de cursos relacionados con el aprendizaje de programación de computadoras; las opiniones escritas por estudiantes se recolectaron de diferentes plataformas educativas y fueron etiquetadas con polaridad y una emoción centrada en el aprendizaje que seleccionó el mismo usuario o el administrador del proyecto.

El desarrollo del corpus fue incremental, en una primera versión se consideraron cinco estados emocionales (aburrido, frustrado, neutral, emocionado y comprometido), sin embargo, se decidió eliminar las frases con etiqueta neutral debido a que los experimentos con el clasificador arrojaban una precisión de 50.25%, la cual no era satisfactoria porque esto quiere decir que el clasificador solamente acierta en un 50% y esto se considera inapropiado para usarse en los ambientes o sistemas tutores inteligentes. Un problema detectado en el corpus fue que estaba desbalanceado, esto quiere decir que contenía más opiniones positivas (emocionado y comprometido) que negativas (aburrido y frustrado).

La segunda versión del corpus, donde las frases con etiqueta neutral fueron removidas, se utilizó con el clasificador, y éste mejoró la precisión a un 60.30%, en esta versión del corpus, las frases con etiqueta *aburrido* eran menos que las otras por lo que se trató de generar más frases con esta etiqueta para balancear la cantidad de frases de cada una de la cuatro diferentes etiquetas.

En la tercera versión del corpus afectivo, se modificó el formato de almacenamiento de frases para que éste fuera utilizado con otro clasificador llamado EvoDAG con el cual se alcanzó una precisión de 68.80%. Este resultado se considera adecuado para implementar el análisis de sentimientos en un sistema tutor inteligente que reconozca el estado emocional del usuario utilizando frases escritas en español, y responda de acuerdo a las necesidades cognitivas y afectivas del estudiante.

El módulo de análisis de sentimientos se desarrolló tomando como base un clasificador que utiliza varios algoritmos tradicionales como: SVM, SVM linear, Multinomial Naïve Bayes, Bernoulli Naïve Baye, SGD y kNN para determinar la polaridad (positiva o negativa) de una frase, el cual alcanza una precisión de **83.50%** con el clasificador *Bernoulli Naïve Bayes*. El clasificador se adaptó para determinar la emoción centrada en el aprendizaje (aburrido, frustrado, emocionado y comprometido) en lugar de la polaridad de la frase. Este clasificador alcanzó una precisión de **60.30%** utilizando el algoritmo de clasificación *SVM linear*.

La metodología iterativa diseñada para generar el corpus, consiste de cinco pasos: seleccionar la muestra de la población, obtener opiniones etiquetadas de los estudiantes, validar las frases, verificar etiquetas de las frases y almacenar en la BD las frases etiquetadas. Esta metodología se ejecutó para generar las frases etiquetadas por los estudiantes y conformar el corpus afectivo, el cual se incrementó en cada iteración de la ejecución de la metodología. Para seleccionar la muestra de la población de estudiantes que generarían las frases, se utilizaron diferentes plataformas educativas como SERE y otras plataformas en línea. La validación de las frases se ejecutó con un proceso manual con el fin de verificar que las frases fueran opiniones válidas y no oraciones sin sentido o definiciones de conceptos; las frases que fueron clasificadas como inválidas se eliminaron de la BD. Para todas las frases válidas, se verificó que la etiqueta asociada a la frase por el estudiante fuera la correspondiente a la

emoción expresada en la frase. Todas las frases válidas y etiquetadas con una emoción relacionada con el aprendizaje fueron almacenadas conformando así el corpus afectivo.

El Sistema de Evaluación de Recursos Educativos (SERE) se desarrolló con el fin de mostrar a los estudiantes recursos educativos a través de diferentes interfaces y así recabar opiniones textuales escritas por estudiantes. Se crearon diferentes versiones de SERE con el fin de generar frases con diferentes etiquetas. La primera versión de SERE permitió generar 3,932 frases principalmente etiquetadas con las emociones comprometido, frustrado y neutral. En la segunda versión de SERE se incluyeron ejercicios de programación con el fin de generar frases negativas y balancear el corpus. La tercera versión de SERE incluyó ejercicios de programación con mayor nivel de dificultad para generar frases negativas con la etiqueta frustrado o aburrido.

Finalmente se desarrolló un servicio web con el fin de poner al alcance de cualquier usuario el analizador de sentimientos y las bases de datos afectivas. Este servicio web estuvo alojado en los servidores de Amazon que fueron contratados por un año y durante este tiempo, estuvo disponible para que un usuario ingresara un texto y obtuviera como resultado una emoción centrada en el aprendizaje así como su polaridad; además de poder descargar los corpus afectivos etiquetados con polaridad o con emoción centrada en el aprendizaje.

6.2. Trabajo Futuro

Para trabajos futuros, se considera la inclusión de ejercicios para desarrollo de programas Java en el sistema SERE por lo cual continuará disponible en la url <http://posgradoitc.ddns.net:8000> para que más participantes expresen sus opiniones acerca de los recursos educativos. Esto permitirá además crear un banco de problemas para el aprendizaje de la programación de computadoras, así como ampliar el número de opiniones en el corpus balanceando las ocurrencias de cada una de las etiquetas de emociones relacionadas con el aprendizaje con el fin de mejorar la precisión del clasificador.

Por otra parte, el corpus será utilizado posteriormente para que el administrador de un Ambiente Inteligente de Aprendizaje o de un Sistema Tutor Inteligente tome decisiones acerca de la pertinencia de los recursos educativos que contiene el sistema y proponga cambios o mejoras a los mismos. Esto ayudará a que los Sistemas Tutores Inteligentes

detecten emociones a través de texto y realicen de manera más eficiente el proceso de enseñanza con los estudiantes, ajustando el contenido a las necesidades cognitivas y afectivas particulares de cada uno de ellos.

Bibliografía

- Altrabsheh, N. &. (2013). SA-E: Sentiment Analysis for Education. *Frontiers in Artificial Intelligence and Applications*.
- Altrabsheh, N. M. (2014). "Learning Sentiment from Students' Feedback for Real-Time Interventions in Classrooms," in *Adaptive and Intelligent Systems: Third International Conference ICAIS 2014, Bournemouth. Springer International Publishing*, 40–49.
- Alvaro Ortigosa, J. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Comput. Hum. Behav.* (31), 527-541.
- Agić, Z. N. (2010). Towards Sentiment Analysis of Financial Texts in Croatian. In *Proceedings of Language Resources and Evaluation (LREC)*.
- Ahmad, C. y. (2006). Multi-lingual sentiment analysis of financial news streams. *Proceedings of Science*,.
- Akkila, A. N. (2017). Teaching the right letter pronunciation in reciting the holy Quran using intelligent tutoring system. (2(1)), 64–68.
- Amazon. (2018). Obtenido de <https://www.amazon.com.mx/>
- Arroyo, I. C. (2009). Emotion Sensors Go To School. *Paper presented at the AIED*.
- Atkins, S. C. (1992). Corpus design criteria. *Literary and Linguistic Computing*. (7(1)), 1–16.
- Baker, R. &. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive affective states during interactions with three different computer-based learning environments. *Int. J. Hum.-Comput. Stud.*(68), 223-241.
- Barrón Estrada, M. L. (2017). "Sentiment Analysis in an Affective Intelligent Tutoring System". *IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, 394-397.
- Barrón Estrada, M. C. (2017). Building a Corpus of Phrases Related to Learning for Sentiment Analysis. *Research in Computing Science* 146, 17-26.
- BestDay. (2018). Obtenido de <https://www.bestday.com.mx/Tours/>
- Boldrini, E. A.-B. (2009). Emotiblog: an annotation scheme for emotion detection and analysis in non- traditional textual genres. In *DMIN*. 491–497.
- Brusilovsky, P. (1998). Methods and techniques of adaptive hypermedia *Adaptive hypertext and hypermedia* . Springer, 1-43.
- Brusilovsky, P. (2003). Developing adaptive educational hypermedia systems: From design models to authoring tools *Authoring tools for advanced technology Learning Environments*. Springer, 377-409.
- Caruso D., S. P. (2005). El directivo emocionalmente inteligente. 65-142.
- Chóliz, M. (2005). Psicología de la emoción: el proceso emocional . www.uv.es/=choliz.

- Conati, C. &. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*. (19(3)), 267-303.
- Coursera*. (2018). (<https://www.coursera.org/>, Productor) Obtenido de Coursera.
- Cruz, F. J. (2008). Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. *Sociedad Española para el Procesamiento de Lenguaje Natural*. (41).
- Denecke, K. (2008). Using SentiWordNet for multilingual sentiment analysis. *IEEE Computer Society*, 507–512.
- D’Mello, S. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*. 22(2), 145-157.
- D’Mello, S. J. (June de 2008). AutoTutor detects and responds to learners affective and cognitive states. In *Workshop on emotional and cognitive issues at the international conference on intelligent tutoring systems*. pp. 306-308.
- D’Mello, S. K. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. .
- D’Mello, S. P. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*(4), 53-61.
- Elliot, C. (1992). The affective reasoner: A process model of emotions in a multi- agent system. *Northwestern University: Tesis doctoral, The Institute for Learning Sciences, Northwestern University*.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*. (6), 169-200.
- Feurer, M. K. (2015).). Efficient and robust automated machine learning. In *Advances in Neural Information Processing Systems*. 2962–2970.
- Ghorbel, H. y. (June de 2010). Sentiment analysis of French movie reviews. *Proceedings of the 4th international Workshop on Distributed Agent-based Retrieval Tools (DART 2010)*. Geneva.
- González Hernández, F. (2015). Ambiente de aprendizaje inteligente y afectivo para el lenguaje java, Tesis de Maestría, Instituto Tecnológico de Culiacán.
- Google Shopping*. (2018). Obtenido de <https://www.google.com/shopping>
- Graff, M. (November de 2016). Evodag: A semantic genetic programming python library. In *Power, Electronics and Computing (ROPEC)*. *IEEE* , 1-6.
- Greer, J. E. (1994). *Student modelling: the key to individualized knowledge-based instruction*: Springer Verlag.
- Gross M. (2018). *Las emociones y la inteligencia emocional*.
- Henríquez, M. C. (2015). Extracción de información desde la web para identificar acciones de un modelo de dominio en planificación automática. *Ingeniare. Revista chilena de ingeniería*(23(3)), 439-448.
- Hoteles*. (2018). Obtenido de <https://www.hoteles.com/>

- I. Guyon, I. C. (2016). A brief review of the ChaLearn AutoML challenge. *In Proc. of AutoML 2016@ICML*.
- Lindor Valdez, M. A. (2016). CODETRAINING: una herramienta de autor para un ambiente de aprendizaje gamificado.
- Medhat, W. H. (2014). Sentiment analysis algorithms and applications: A survey. . *Ain Shams Engineering Journal*(5(4)), 1093-1113.
- Mercado Libre*. (2018). Obtenido de <https://www.mercadolibre.com.mx/>
- Mills. (1980). Método Incremental.
- Ministerio de Educación Nacional Colombiano. (2006). Objetos Virtuales de Aprendizaje e Informativos.
- Moreno Ortiz, A. P. (2010). Análisis de Valoraciones de Usuario de Hoteles con Sentitext: un sistema de análisis de sentimiento independiente del dominio. Procesamiento del lenguaje natural. (45).
- Munezero, M. M. (November de 2013). Exploiting sentiment analysis to track emotions in students' learning diaries. In Proceedings of the 13th Koli Calling International Conference on Computing Education Research. 145-152.
- Netflix*. (2018). Obtenido de <https://www.netflix.com/>
- Nwana, H. S. (1990). Intelligent tutoring systems: an overview. *Artificial Intelligence Review*(4(4)), 251-277.
- Oramas B., R. (2018). Opinion Mining and Emotion Recognition in an Intelligent Learning Environment.
- Ortigosa, A. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*. (31), 527-541.
- Ortony, A. C. (1988). The cognitive structure of emotions. *New York: Cambridge University Press*.
- Plutchik, R. (1980). A general psychoevolutionary theory of emotion. *Emotion: Theory, research, and experience*. 3-33.
- Pang, B. L. (2002). Thumbs Up? Sentiment Classification Using Machine Learning Techniques. . *Conference on Empirical Methods in Natural Language Processing*.
- Pang, B. y. (2008). Opinion Mining and Sentiment Analysis. . *Foundations and trends in information retrieval*(2(1-2)), 1-135.
- Pedregosa, F. V. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Pekrun, R. (1992). The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. *Applied Psychology*. (41(4)), 359-376.
- Pekrun, R. G. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist*. (37(2)), 91-105.

- Peñalver Martínez, I. V. (2011). Minería de Opiniones basada en características guiada por Ontologías. *Procesamiento del Lenguaje Natural*. (46).
- Picard, R. W. (2000). *Affective computing*: MIT press.
- Real Academia Española. (2005). Real Academia Española. <http://dle.rae.es/?id=AwTBMcs>.
- Ríos Félix, J. M. (2016). Ambiente de aprendizaje afectivo para la enseñanza de lógica algorítmica y programación. En *Tesis de Maestría*.
- Self, J. (1998). The defining characteristics of intelligent tutoring systems research: ITSs care, precisely. *International Journal of Artificial Intelligence in Education (IJAIED)*, 350-364. .
- Sosa Ochoa, C. (2016). Agente pedagógico con estrategia “aprendiz cognoscitivo” para un ambiente de programación colaborativo. En *Tesis de Maestría*.
- Tao, J. &. (2005). *Affective computing: A review Affective computing and intelligent interaction*. Springer, 981-995.
- Teachlr*. (2018). Obtenido de teachlr: <https://teachlr.com/>
- Udemy*. (2018). Obtenido de udemy: <https://www.udemy.com/>
- Valencia Rodríguez, E. M. (2017). Ambiente de Aprendizaje de Depuración de Errores de Programación Basado en Ejemplos Erróneos,. En *Tesis de Maestría*.
- Vilares, D. (2013). Clasificación de polaridad en textos con opiniones en español mediante análisis sintáctico de dependencias. *Procesamiento del lenguaje natural*. 50, 13-20.
- Villena Román, J. L. (2013). TASS -Workshop on Sentiment Analysis at SEPLN TASS - Taller de Análisis de Sentimientos en la SELPLN. 37-44.
- Volaris*. (2018). Obtenido de <https://www.volaris.com/>
- Woolf, B. P. (2009). *Building Intelligent Interactive Tutors*. Estados Unidos: Morgan Kaufmann Publishers.
- Woolf, B. P. (2009). *Building Intelligent Interactive Tutors*. Amherst, Massachusetts, Estados Unidos: Morgan Kaufmann Publishers.
- Youtube*. (2018). Obtenido de <https://www.youtube.com/>
- Zhang, C. D. (2009). Sentiment analysis of chinese documents: From sentence to document level. *JASIST*. (60(12)), 2474– 2487.

Anexo

En esta sección se incluye en un CD el código fuente que corresponde a los diferentes productos generados:

1. Corpus afectivo, contiene las dos versiones de las 9,009 frases etiquetadas con los formatos para el algoritmo tradicional y para EvoDAG.
2. Sistema de Evaluación de Recursos Educativos, el cual se encuentra en la carpeta llamada SERE.
3. Módulo de análisis de sentimientos, que se encuentra en la carpeta Analizador.
4. Servicio Web, que se incluye en la carpeta Servicio.

En cada carpeta de código, se incluyen los archivos correspondientes a las interfaces de usuario utilizadas por cada sistema.