



ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ingeniería en Electricidad y Computación

“CLASIFICACIÓN AUTOMÁTICA DE RESPUESTAS A FOROS DE DISCUSIÓN DE ACUERDO AL DOMINIO COGNITIVO DE LA TAXONOMÍA DE BLOOM EN SIDWEB 4, EMPLEANDO MINERÍA DE TEXTO Y ALGORITMOS DE APRENDIZAJE”

INFORME DE PROYECTO DE GRADUACIÓN

Previo a la obtención del título de:

**INGENIERO EN CIENCIAS COMPUTACIONALES
ESPECIALIZACIÓN SISTEMAS MULTIMEDIA**

Presentado por:

JHONNY VLADIMIR PINCAY NIEVES

GUAYAQUIL – ECUADOR

2013

AGRADECIMIENTO

Agradezco en primera instancia al Divino Creador quien guía cada uno de mis pasos y también por haberme dado la oportunidad de disfrutar de una familia maravillosa, son ellos quienes me han apoyado incondicionalmente a lo largo de todos estos años de estudio y gracias a ellos soy lo que soy.

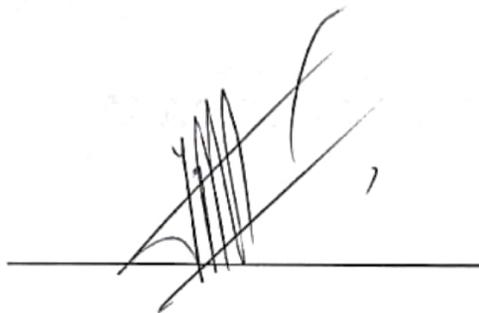
Un agradecimiento especial al Ph.D. Xavier Ochoa director de mi proyecto de graduación, por su paciencia, ayuda y el tiempo invertido en el mismo. Agradezco también a mis amigos con los que pude trabajar a lo largo de todas las asignaturas de la carrera y con quienes compartí gran parte de estos últimos cinco años.

DEDICATORIA

Dedico este trabajo a mis padres quienes me han respaldado incondicionalmente y se han sacrificado para que pueda alcanzar mis metas y cumplir mis sueños.

Dedico también este trabajo especialmente a mi Mami Cori quien a pesar de sus problemas de salud me cuidó desde muy pequeño y aunque ya no está en este mundo sé que aún lo sigue haciendo.

TRIBUNAL DE SUSTENTACIÓN

A handwritten signature in black ink, consisting of several vertical strokes followed by a diagonal stroke, positioned above a horizontal line.

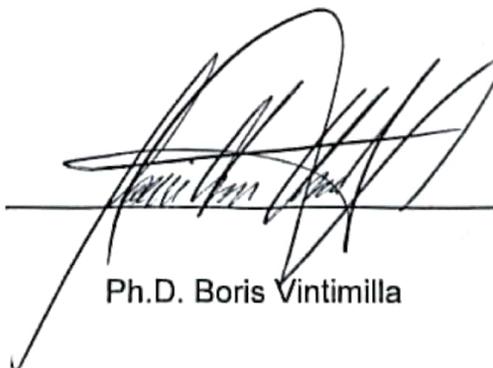
M.Sc. Miguel Yapur

PRESIDENTE

A handwritten signature in black ink, featuring a large, stylized initial 'X' followed by a cursive name, positioned above a horizontal line.

Ph.D. Xavier Ochoa

DIRECTOR DE PROYECTO DE GRADUACIÓN

A handwritten signature in black ink, with a large, stylized initial 'B' and a cursive name, positioned above a horizontal line.

Ph.D. Boris Vintimilla

MIEMBRO PRINCIPAL

RESUMEN

DECLARACIÓN EXPRESA

"La responsabilidad del contenido de este informe de proyecto de graduación me corresponde exclusivamente; y el patrimonio intelectual de la misma a la Escuela Superior Politécnica del Litoral".

(Reglamento de exámenes y títulos profesionales de la ESPOL)



Jhonny V. Pincay Nieves

RESUMEN

El presente trabajo presenta la implementación de un sistema de clasificación automática para categorizar las respuestas de estudiantes a foros de discusión de acuerdo al dominio cognitivo de la taxonomía de Bloom. Diversos estudios se han llevado a cabo en esta área, en esta ocasión la efectividad y desempeño de un clasificador Bayesiano, específicamente el clasificador Naïve Bayes, son analizados. En esta propuesta este clasificador es empleado conjuntamente con técnicas de minería de texto y un conjunto de datos previamente clasificado por expertos para llevar a cabo la categorización.

Este proyecto también contempló la implementación de una interfaz de usuario en la versión cuatro del Sistema Interactivo de Desarrollo para el Web – SIDWeb versión 4, con el objetivo que los profesores puedan interactuar con el sistema de clasificación automática, etiquetando a través de este las participaciones y contribuciones de estudiantes en los foros de discusión que se efectúan dentro de SIDWeb 4. También se proveen estadísticas de las participaciones de los estudiantes en las discusiones, así como también de las categorías dentro de la taxonomía que están alcanzando con la finalidad que los educadores puedan a simple vista tener una idea de si están logrando los objetivos de aprendizaje planteados, el

nivel al que están llegando sus estudiantes y si se encuentran discutiendo superficialmente o si están generando nuevo conocimiento.

Otro aspecto que se estudió fue el determinar si el sistema de clasificación puede compararse o considerarse como un codificador humano. Varias pruebas fueron llevadas a cabo con el objetivo de obtener métricas cuyos valores permitieron evaluar el desempeño del clasificador y calidad de los resultados. Los resultados obtenidos indican que usando la arquitectura propuesta es posible lograr los objetivos planteados, sin embargo la efectividad de la clasificación se ve afectada por la calidad de los datos de entrenamiento provocando que se logre buenos resultados para los niveles de la taxonomía de los que se tienen una cantidad considerable de ejemplos y malos resultados para los niveles que poseen pocos ejemplos etiquetados.

ÍNDICE GENERAL

| | |
|--|-----|
| AGRADECIMIENTO | I |
| DEDICATORIA | II |
| RESUMEN..... | V |
| ABREVIATURAS | X |
| ÍNDICE DE FIGURAS..... | XII |
| ÍNDICE DE TABLAS | XIV |
| INTRODUCCIÓN..... | XV |
| CAPÍTULO 1 | 1 |
| 1. JUSTIFICACIÓN Y ANÁLISIS DEL PROBLEMA | 1 |
| 1.1. ANTECEDENTES..... | 1 |
| 1.2. DESCRIPCIÓN DEL PROBLEMA | 3 |
| 1.3. JUSTIFICACIÓN | 5 |
| 1.4. PROPUESTA DE SOLUCIÓN..... | 6 |
| CAPÍTULO 2..... | 8 |
| 2. REVISIÓN DEL ESTADO DEL ARTE | 8 |
| 2.1. TAXONOMÍA DE BLOOM | 8 |
| 2.1.1. TAXONOMÍA DE BLOOM REVISADA..... | 11 |
| 2.1.2. DOMINIO COGNITIVO..... | 14 |
| 2.1.3. DOMINIO AFECTIVO..... | 17 |

| | | |
|-----------------|---|----|
| 2.1.4. | DOMINIO PSICOMOTOR | 18 |
| 2.2. | CLASIFICACIÓN DE TEXTO | 21 |
| 2.2.1. | PRE PROCESAMIENTO DE TEXTO..... | 24 |
| 2.2.2. | REPRESENTACIÓN VECTORIAL DE TEXTO | 27 |
| 2.2.3. | REDUCCIÓN DE CARACTERÍSTICAS | 30 |
| 2.3. | ALGORITMOS DE APRENDIZAJE | 35 |
| 2.3.1. | MÁQUINA DE SOPORTE DE VECTORES..... | 36 |
| 2.3.2. | CLASIFICADORES BAYESIANOS | 38 |
| 2.4. | ANÁLISIS COMPARATIVO DE SOLUCIONES EXISTENTES | 40 |
| CAPÍTULO 3..... | | 47 |
| 3. | ANÁLISIS Y DISEÑO DE LA SOLUCIÓN..... | 47 |
| 3.1. | ANÁLISIS Y DISEÑO DEL SISTEMA..... | 47 |
| 3.1.1. | ANÁLISIS Y DISEÑO DEL COMPONENTE DE APRENDIZAJE | |
| | 50 | |
| 3.1.2. | ANÁLISIS Y DISEÑO DEL COMPONENTE DE | |
| | CLASIFICACIÓN | 53 |
| 3.2. | DISEÑO DEL MÓDULO DE CLASIFICACIÓN EN SIDWeb 4..... | 55 |
| 3.2.1. | MODELO LÓGICO DEL MÓDULO DE CLASIFICACIÓN | 57 |
| 3.3. | DEFINICIÓN DEL CONJUNTO DE ENTRENAMIENTO | 58 |
| 3.4. | ALCANCE DEL DISEÑO | 64 |
| 3.5. | DISEÑO DE PRUEBAS..... | 65 |

| | |
|---|----|
| CAPÍTULO 4..... | 70 |
| 4. DESARROLLO Y ANÁLISIS DE RESULTADOS..... | 70 |
| 4.1. IMPLEMENTACIÓN | 70 |
| 4.2. PRUEBAS | 83 |
| 4.2.1. PRUEBAS DE FUNCIONALIDAD | 83 |
| 4.2.2. PRUEBAS DE USABILIDAD | 85 |
| 4.3. ANÁLISIS DE RESULTADOS | 86 |
| CONCLUSIONES | 89 |
| RECOMENDACIONES..... | 91 |
| BIBLIOGRAFÍA..... | 93 |

ABREVIATURAS

| | |
|--------------|---|
| CF-DF | Category Frequency – Document Frequency (en español Frecuencia de Categoría – Frecuencia de Documento). |
| DF | Document Frequency (en español Frecuencia de Documento). |
| HOTS | Higher Order Thinking Skills (en español Habilidades de Pensamiento de Orden Superior). |
| LMS | Learning Management System (en español Sistema de Administración de Aprendizaje). |
| LOTS | Lower Order Thinking Skills (en español Habilidades de Pensamiento de Orden Inferior). |
| MVC | Model View Controller (en español Modelo Vista Controlador). |
| PCA | Principal Component Analysis (en español Componente Principal de Análisis). |

| | |
|---------------|---|
| RDBMS | Relational Database Management System (en español Systema de Administración de Base de Datos Relacionales). |
| SVM | Support Vector Machine (en español Máquina de Soporte de Vectores). |
| TFxIDF | Term Frequency x Inverse Document Frequency. |
| XML | Extensible Markup Language (en español Lenguaje de Marcado Extensible). |

ÍNDICE DE FIGURAS

| | |
|--|----|
| Figura 2.1 Niveles de la Taxonomía de Bloom desde LOTS hasta HOTS.... | 10 |
| Figura 2.2 Cambios en la terminología de los nombres de cada nivel de la taxonomía de Bloom | 12 |
| Figura 2.3 Correspondencia entre los niveles de la dimensión del proceso cognitivo y la dimensión del conocimiento | 13 |
| Figura 2.4 Representación gráfica del flujo del proceso de clasificación de texto | 24 |
| Figura 2.5 Evolución de un texto al que se le ha aplicado pre procesamiento para eliminar información no relevante | 27 |
| Figura 2.6 Híper plano de un clasificador SVM binario | 37 |
| Figura 3.1 Esquema de la arquitectura del sistema de clasificación automática | 49 |
| Figura 3.2 Esquema de la estructura del componente de aprendizaje | 50 |
| Figura 3.3 Esquema de la estructura del subcomponente de pre procesamiento de texto y formación de vectores..... | 51 |
| Figura 3.4 Esquema de la estructura del componente de clasificación | 54 |
| Figura 3.5 Tabla discussion_entries con campo tag adicionado..... | 58 |
| Figura 4.1 Implementación del componente de aprendizaje en RapidMiner | 72 |

| | |
|--|----|
| Figura 4.2 Implementación del subcomponente de pre procesamiento de texto en RapidMiner | 73 |
| Figura 4.3 Características y pesos extraídos de cada elemento del conjunto de entrenamiento | 73 |
| Figura 4.4 Vista del desempeño de la clasificación por cada nivel de la taxonomía de Bloom | 74 |
| Figura 4.5 Implementación del componente de clasificación en RapidMiner | 74 |
| Figura 4.6 Resultado de ejecutar los procesos del componente de clasificación en RapidMiner | 75 |
| Figura 4.7 Estadísticas generales de un foro de discusión en SIDWeb 4..... | 80 |
| Figura 4.8 Tabla resumen y diagrama de pastel de las participaciones por estudiantes en un foro de discusión de SIDWeb 4 | 81 |
| Figura 4.9 Totales por nivel de la taxonomía de Bloom alcanzados en un foro de discusión en SIDWeb 4..... | 82 |
| Figura 4.10 Número de colaboraciones por nivel de un determinado estudiante en un foro de discusión en SIDWeb 4 | 82 |

ÍNDICE DE TABLAS

| | |
|--|----|
| Tabla 2.1 Principales características de soluciones existentes para clasificación automática de acuerdo a la Taxonomía de Bloom | 46 |
| Tabla 3.1 Número de respuestas etiquetadas por nivel del subconjunto de entrenamiento del año 2010 | 60 |
| Tabla 3.2 Número de respuestas etiquetadas por nivel del subconjunto de entrenamiento del año 2011 | 61 |
| Tabla 3.3 Número de respuestas etiquetadas por nivel del subconjunto de entrenamiento del año 2012 | 62 |
| Tabla 3.4 Número de respuestas etiquetadas por nivel del conjunto de entrenamiento | 63 |
| Tabla 3.5 Número de respuestas por nivel del conjunto de pruebas | 68 |
| Tabla 4.1 Resultados del conteo de los valores de A, B, C y D luego de aplicar el modelo de clasificación al conjunto de pruebas | 84 |
| Tabla 4.2 Valores resultantes de las métricas por nivel de la taxonomía | 84 |
| Tabla 4.3 Resultados de las pruebas de fiabilidad entre codificadores | 85 |

INTRODUCCIÓN

La taxonomía de Bloom es una taxonomía o clasificación de habilidades cognitivas que ha sido ampliamente empleada para evaluación y medición de objetivos educacionales. Es un concepto esencial que guía a los educadores en la elaboración de objetivos educacionales, preparación de currículums y creación de evaluaciones. Fue propuesta en 1956 por Benjamin Bloom, esta taxonomía divide los objetivos educacionales en tres dominios: cognitivo, psicomotor y afectivo.

El dominio cognitivo está relacionado con el conocimiento, habilidades y destrezas intelectuales de cada individuo, las cuales están clasificadas desde un nivel que corresponde a simplemente recordar o reconocer hasta niveles más altos que conllevan la creación de nuevo conocimiento. El dominio psicomotor se enfoca en las habilidades físicas, las cuales incluyen habilidades motoras refinadas, manipulación de herramientas o cualquier tipo de actividad que requiera de coordinación neuromuscular. Finalmente, el dominio afectivo se enfoca en las emociones, sentimientos, actitudes, el grado de aceptación o rechazo de un individuo.

Los beneficios de incluir la utilización de la taxonomía de Bloom en las prácticas educativas y de enseñanza son innegables, sin embargo el proceso de categorizar preguntas, opiniones y argumentos de estudiantes de acuerdo a esta taxonomía es tedioso, consume una gran cantidad de tiempo, dado que generalmente se lo realiza manualmente y no todos los educadores tienen la capacidad para hacerlo, razones por las cuales muchas veces se opta por no realizar la categorización.

En este trabajo se propone la implementación de un sistema de clasificación automática de las contribuciones en foros de discusión, empleando técnicas de minería de texto y el uso de un clasificador Bayesiano, también se busca medir que tan exacto es este clasificador en la tarea específica de la asignación de una categoría de la taxonomía de Bloom a un texto. El otro propósito de este estudio es determinar si los resultados proporcionados por el sistema de clasificación, son lo suficientemente buenos como para reemplazar a codificadores humanos.

La estructura del presente trabajo es la siguiente: en el capítulo uno se define el problema y se presenta una propuesta de solución, en el capítulo 2 se realiza una revisión del estado del arte, los análisis llevados a cabo para el diseño de la solución son presentados en el capítulo tres, detalles de la implementación de la solución y los resultados obtenidos se presentan en el capítulo cuatro, el documento finaliza con conclusiones y recomendaciones.

CAPÍTULO 1

1. JUSTIFICACIÓN Y ANÁLISIS DEL PROBLEMA

En este capítulo se revisan antecedentes y la descripción del problema que se espera resolver con su justificación y la propuesta de solución que incluye la implementación de un componente automático de clasificación e interfaz de usuario en el SIDWeb 4, sistema de administración de conocimiento empleado en la Escuela Superior Politécnica del Litoral - ESPOL.

1.1. ANTECEDENTES

La taxonomía de Bloom es una clasificación de objetivos educacionales empleada ampliamente por educadores en la planeación y diseño de objetivos de aprendizaje, preparación de planes de estudios y en el diseño de instrumentos de evaluación [8].

Tres dimensiones han sido identificadas cognitiva, psicomotora y afectiva. La dimensión cognitiva de la taxonomía de Bloom es aquella que implica el conocimiento y desarrollo de habilidades intelectuales de un individuo, las cuales son clasificadas desde un nivel bajo que corresponde a simplemente recordar o reconocer hasta niveles más altos que conllevan la creación de nuevo conocimiento [8] [12].

Los beneficios de incluir la utilización de la taxonomía de Bloom en las prácticas educacionales y de enseñanza son innegables, sin embargo el proceso de categorizar preguntas, opiniones y argumentos de estudiantes de acuerdo a esta taxonomía es tedioso y consume una gran cantidad de tiempo, dado que generalmente se lo realiza manualmente.

Lo expuesto anteriormente ha desembocado en que se busque automatizar este proceso de clasificación, por ello diversas soluciones han sido propuestas. Estas implementaciones hacen uso de distintas técnicas como la minería de texto, procesamiento de lenguaje natural, sistemas inteligentes con redes neuronales, etc. Lamentablemente la gran mayoría no son suficientemente precisas, no se encuentran disponibles para un uso masivo de usuarios y a fin de cuentas no constituyen una solución real a la problemática planteada [8].

En la ESPOL la práctica de la categorización en los niveles de la taxonomía de Bloom es casi nula, salvo investigaciones que se llevan a cabo en el Centro de Tecnologías de Información – CTI en donde se la realiza en ciertos cursos con fines experimentales.

ESPOL dispone de una herramienta llamada Sistema Interactivo de Desarrollo para el Web – SIDWeb, que es un sistema de administración de conocimiento empleado para facilitar la distribución de contenidos de los cursos, comunicación entre estudiantes y profesores fuera del aula, entre otras características, actualmente este sistema se encuentra en la versión 4.

Dentro de las funcionalidades de SIDWeb se encuentran los foros de discusión, que son el medio de recolección de datos empleado por CTI para efecto de las investigaciones que efectúa dicho centro en el campo educativo.

1.2. DESCRIPCIÓN DEL PROBLEMA

Para realizar una clasificación adecuada de acuerdo a la taxonomía de Bloom es requerido entrenamiento y cierto grado de experiencia por parte del educador, si esto no se cumple se tendrá como resultado una categorización errada y una identificación de nivel del estudiante incorrecta.

Aún para los educadores que tienen cierto grado de experiencia en la categorización resulta un trabajo tedioso y que demanda mucho tiempo, más aún si consideramos las participaciones de los estudiantes en un foro de discusión, por ende casi siempre la clasificación es evitada. Además gran parte del tiempo para tener una categorización acertada, es necesaria la participación de más de una persona en la clasificación y si las conclusiones previas a las que han llegado cada uno difieren, una discusión es requerida para finalmente llegar a una decisión final.

Las participaciones de estudiantes en foros de discusión constituyen un buen medio para determinar el nivel de desarrollo con respecto a un tema en el cual se encuentra un estudiante. SIDWeb 4 permite desarrollar foros de discusión en el cual el profesor propone un tema y los estudiantes discuten al respecto para llegar a conclusiones, sin embargo la actividad finaliza en este punto, sin que el docente realmente pueda determinar el nivel de comprensión y desarrollo del estudiante, ni establecer si las participaciones fueron realmente significativas y contribuyeron a construir conocimiento o simplemente la discusión se la llevó de una manera superficial.

El disponer de la información anteriormente descrita contribuirá para que los docentes ajusten la manera en la cual se encuentran

desarrollando sus clases, con la finalidad de lograr los resultados de aprendizaje que esperan.

1.3. JUSTIFICACIÓN

El proveer una herramienta de clasificación automática de respuestas a foros de discusión basado en la dimensión cognitiva de la taxonomía de Bloom y que además se encuentre disponible a través de un sistema web como lo es el SIDWeb, permitirá a los docentes de ESPOL conocer el grado de comprensión de los estudiantes de determinado tema, si se encuentran generando conocimiento a partir de lo aprendido y discutido y las habilidades que estos van adquiriendo.

Además facilitará a los docentes que no poseen experiencia en la categorización aprovechar los beneficios de realizar el etiquetado, así como también incorporar los resultados obtenidos como un aspecto más a considerar en el diseño de sus clases.

El sistema constituirá también un medio de recolección de datos que pueden ser empleados por investigadores para estudios futuros en el campo educativo.

1.4. PROPUESTA DE SOLUCIÓN

Luego de una revisión sobre la minería de texto como técnica para derivar información relevante a partir de una colección de texto, con la finalidad de extraer patrones, tendencias y asociaciones, es decir descubrir conocimiento que no es identificable a simple vista, se construirá un sistema que procesará un conjunto de respuestas de estudiantes a foros de discusión categorizados previamente de acuerdo al nivel cognitivo de la taxonomía de Bloom por expertos, con la finalidad de obtener una representación vectorial del texto llamada vectores de características en donde la información no significativa ha sido eliminada.

Los vectores de características servirán de entrada para un algoritmo de aprendizaje que como resultado arrojará un modelo el cual indicará cuales son las características que debe tener cierto texto para pertenecer a determinado nivel de la taxonomía de Bloom. El proceso de obtención de los vectores de características sumado al entrenamiento del algoritmo y la obtención del modelo constituirán el componente de aprendizaje.

A continuación se procederá a construir el componente de clasificación el cual consistirá en recibir determinado texto del cual no se conoce su categoría, procesarlo y evaluarlo empleando el modelo anteriormente

obtenido para finalmente obtener el nivel de la taxonomía de Bloom a la cual aquel texto pertenece.

Se implementará un módulo en la plataforma SIDWeb 4 que será capaz de comunicarse con el sistema de clasificación automática. A través de la interfaz de SIDWeb 4 un usuario podrá visualizar las etiquetas asignadas por el sistema de clasificación automática, el cual evalúa cada uno de los comentarios emitidos por los estudiantes y devolverá como resultado la categoría en la que más probablemente se encuentra determinada respuesta. El módulo poseerá un componente de reporte de resultados en el cual el usuario podrá observar en resumen las conclusiones a las que el sistema de clasificación automática ha llegado, así como también estadísticas generales del progreso de los estudiantes y las participaciones en los foros de discusión.

Se llevarán a cabo pruebas de funcionalidad a fin de determinar con que efectividad el sistema asigna la etiqueta correcta; cabe recalcar que si el sistema asigna etiquetas incorrectas la obtención de un mejor modelo de clasificación será necesaria, puesto que la efectividad del sistema depende en gran medida de la calidad del modelo empleado. Otras pruebas que se llevarán a cabo serán las de usabilidad.

CAPÍTULO 2

2. REVISIÓN DEL ESTADO DEL ARTE

Este capítulo presenta los fundamentos teóricos sobre los cuales se sustenta el diseño de la solución propuesta, además se examina en detalle las técnicas empleadas en la implementación de los componentes de aprendizaje y clasificación automática.

2.1. TAXONOMÍA DE BLOOM

La iniciativa de desarrollar un método de clasificación de las diferentes formas de pensar de cada ser humano surge de las discusiones que se presentaron en el año de 1948 en la convención de la asociación americana de psicología y condujeron a que un grupo de educadores liderados por Benjamin Bloom conciban un esquema para la categorización de objetivos y metas educacionales. Eventualmente este esquema se convirtió en una taxonomía de tres

dominios, donde cada dominio influye significativamente en el proceso de aprendizaje de un individuo [4]:

- Dominio cognitivo
- Dominio afectivo
- Dominio psicomotor

El dominio cognitivo se relaciona con las habilidades intelectuales de un individuo, mientras que el dominio afectivo se enfoca en emociones, actitudes y la forma en la que una persona reacciona emocionalmente, finalmente el dominio psicomotor corresponde a las habilidades físicas, manuales como por ejemplo la destreza al momento de manipular una herramienta [6][8].

En el año de 1956 ocho años después de que el grupo empezó a trabajar y las investigaciones del dominio cognitivo se completaron, se publicó un volumen técnico bajo el título *Taxonomy of Educational Objectives, The Classification of Educational Goals, Handbook I: Cognitive Domain*, en el cual su autor Benjamin Bloom desarrolla un sistema de codificación que permitiría a los educadores diseñar metas y objetivos de aprendizaje que tengan una organización jerárquica. Este sistema de codificación es conocido como la Taxonomía de Bloom y desde entonces ha sido bastamente empleado en el campo educativo [1].

La taxonomía o clasificación de Bloom es un modelo de múltiples niveles que permite la categorización del pensamiento de acuerdo a seis niveles de complejidad, los cuales son comprensión, aplicación, análisis, síntesis y evaluación siendo el nivel de comprensión el más bajo y el de evaluación el más alto [1] [8]. La taxonomía es jerárquica lo cual implica que un estudiante que se encuentra en el nivel de aplicación ha dominado los niveles de conocimiento y comprensión [5].



Figura 2.1 Niveles de la Taxonomía de Bloom desde LOTS hasta HOTS

Las categorías de la taxonomía de Bloom corresponden al proceso de desarrollo del pensamiento, el cual establece que no es posible entender un concepto si primero no se lo ha recordado; de manera

similar no es posible aplicar conocimiento o conceptos si previamente no se los ha entendido. Básicamente la taxonomía de Bloom es un camino continuo que parte desde las habilidades de pensamiento de orden inferior - LOTS¹ hasta las habilidades de pensamiento de orden superior - HOTS², el camino está compuesto por etapas que corresponden a los distintos niveles de la taxonomía de Bloom [5]. En la Figura 2.1 se presentan los niveles de la taxonomía de Bloom y como estos conducen desde el pensamiento de orden inferior hasta el de orden superior.

2.1.1. TAXONOMÍA DE BLOOM REVISADA

Considerando que la taxonomía de Bloom fue establecida en 1956 una revisión de la misma fue necesaria con el objetivo de volverla más relevante para los estudiantes y profesores del siglo 21. Es así como en la década de los noventas un ex alumno de Benjamin Bloom, Lorin Anderson lideró un grupo de expertos que buscaría reformar la taxonomía y adaptarla a las corrientes de aquel entonces [5]. Como consecuencia, en el año 2001 se publica la Taxonomía Revisada de Bloom, cuyos cambios principales se dieron en tres categorías: terminología, estructura y énfasis [2].

¹ LOTS por Lower Order Thinking Skills

² HOTS por Higher Order Thinking Skills

Los cambios de terminología entre las dos versiones pudiesen resultar los más significativos. Cada categoría de la taxonomía llevaba un nombre que es un sustantivo que describía a cada nivel, a partir de la versión revisada se cambia ese sustantivo a su forma verbal. También el nivel conocimiento fue renombrado a recordar, comprensión pasó a ser entender y síntesis corresponde ahora a crear [4] [5].

Con los cambios de terminología la taxonomía quedó establecida con los siguientes nombres de categoría: recordar, comprender, aplicar, analizar, evaluar y crear. En la Figura 2.2 se ilustran las diferencias entre la primera versión de la taxonomía y su versión revisada.

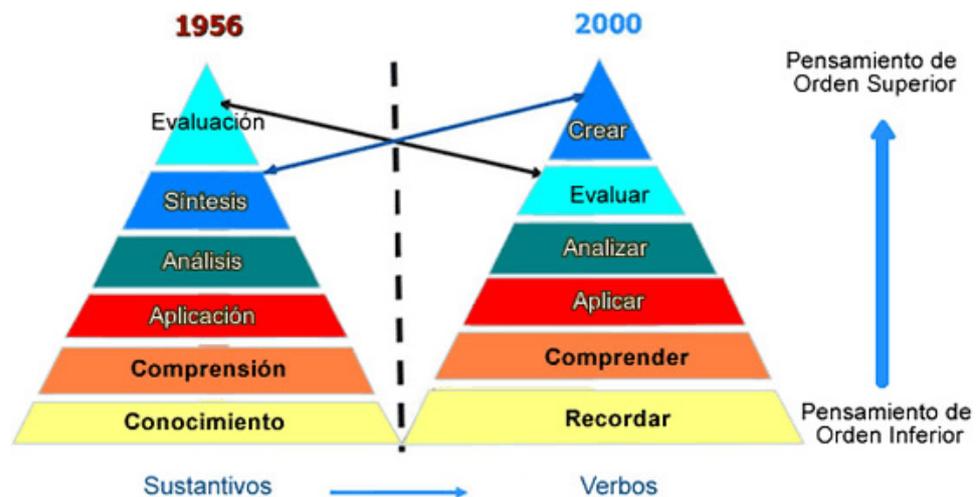


Figura 2.2 Cambios en la terminología de los nombres de cada nivel de la taxonomía de Bloom

En cuanto a los cambios estructurales, la taxonomía original estaba construida de manera unidireccional mientras que la taxonomía revisada tomó la forma de una tabla de dos dimensiones, siendo estas dimensiones la del conocimiento y la del proceso cognitivo [4].

La dimensión del conocimiento representa la acumulación de resultados de aprendizaje por parte de un estudiante y está dividida en conocimiento factual, conceptual, procedimental y meta cognitivo mientras que la dimensión del proceso cognitivo representa la manera de pensar y entender del individuo y se encuentra dividido en recordar, entender, aplicar, analizar, evaluar y crear [2].

En la Figura 2.3 se presenta la correspondencia entre los niveles de la dimensión de conocimiento y las dimensiones del proceso cognitivo.

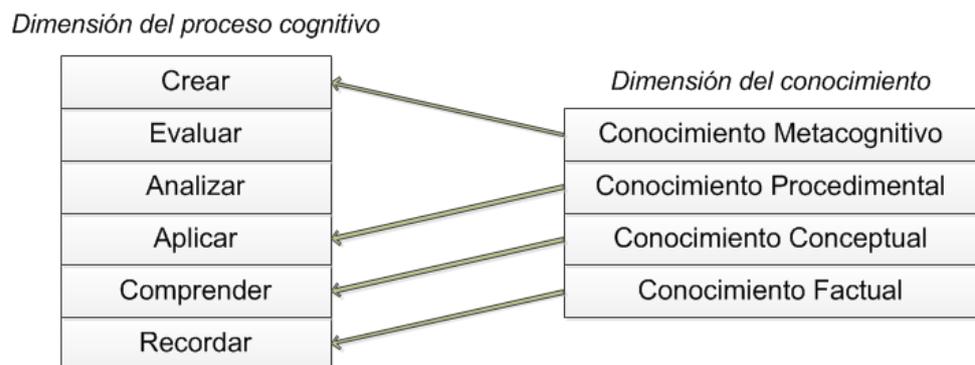


Figura 2.3 Correspondencia entre los niveles de la dimensión del proceso cognitivo y la dimensión del conocimiento

Debido a que la taxonomía de Bloom estaba siendo usada por una audiencia y grupos no previstos, la intención con la que se obtuvo la taxonomía revisada fue la de llegar a esa audiencia y a una mayor; esto condujo entre otras cosas, cambios en énfasis acerca de su uso como una herramienta más auténtica para la planeación de currículums, enseñanza y evaluaciones [4].

2.1.2. DOMINIO COGNITIVO

El dominio cognitivo de la taxonomía de Bloom involucra al conocimiento y al desarrollo de habilidades intelectuales [6]. Comprende una serie de destrezas intelectuales que incluyen la adquisición y uso del conocimiento en actividades que van desde simplemente recordar alguna definición hasta la habilidad de juzgar y evaluar material previamente aprendido [8].

Existen seis niveles de complejidad en este dominio. Estas categorías pueden considerarse como grados de dificultad, esto implica que el individuo debe haber dominado un nivel antes de poder avanzar al siguiente [6].

Estas seis categorías del nivel cognitivo corresponden a los niveles previamente mencionados en las dos secciones anteriores. Los niveles correspondientes a la taxonomía de Bloom publicada en 1956 son:

- **Conocimiento:** Operacionalmente llamada como recuperación de información [1]. Es la habilidad de recordar material previamente aprendido tales como características, hechos, términos y conceptos [8].
- **Comprensión:** Básicamente es la habilidad de entender [3]. Implica que un individuo es capaz de comprender definiciones, interpolaciones e interpretar instrucciones y problemas en sus propias palabras [6].
- **Aplicación:** Se refiere a la habilidad de usar material aprendido en una determinada situación [8]. Implica también usar un concepto en una nueva situación o el uso espontáneo de la abstracción [6].
- **Análisis:** Comprende la habilidad de separar partes componentes [3], es decir dividir determinado material aprendido en sus partes constituyentes con el fin de identificar las relaciones presentes en dichas partes o para reconocer como el material se encuentra organizado [8].
- **Síntesis:** Consiste en colocar distintas partes para formar un todo, haciendo énfasis en la creación de nuevo conocimiento [6]. Puede incluir también la obtención de algún patrón o

estructuras que inicialmente no se encontraban presentes claramente [8].

- **Evaluación:** Es la habilidad de juzgar el valor de una idea [3]. Implica usar determinado criterio para concluir que tan preciso, efectivo o satisfactorio resulta ser algún conocimiento [1].

Debido a las razones expuestas en la subsección 2.1.1, desde el año 2000 los niveles correspondientes al nivel cognitivo de la taxonomía de Bloom son:

- **Recordar:** Este nivel implica que un individuo es capaz de recordar, reconocer, recuperar, listar o localizar conocimiento relevante de la memoria a largo plazo [4].
- **Comprender:** Consiste en interpretar, resumir, inferir, parafrasear o clasificar información con el objetivo de construir significado [5].
- **Aplicar:** El encontrarse en esta categoría implica que una persona es capaz de realizar algún procedimiento a través de la ejecución o implementación de alguna definición o concepto [4].
- **Analizar:** Un individuo que se encuentra en este nivel básicamente es capaz de comparar, organizar, descomponer

material en sus partes constitutivas [5] con el objetivo de determinar la relación existente entre dichas partes [4].

- **Evaluar:** Consiste en juzgar o calificar conocimiento basado en criterios o estándares a través de revisiones o críticas [4].
- **Crear:** Corresponde al nivel más alto de la taxonomía, por ende es el más difícil de alcanzar y conlleva que una persona es capaz de reorganizar elementos en una nueva estructura a través de la generación, planeación o producción de nuevo conocimiento [4].

2.1.3. DOMINIO AFECTIVO

El dominio afectivo se enfoca en las emociones, sentimientos, actitudes el grado de aceptación o rechazo. Este dominio incluye la manera en la cual un individuo actúa con situaciones emocionales como sentimientos, valores, apreciaciones, actitudes, y motivaciones [8].

Al igual que el dominio cognitivo, en el dominio afectivo existen cinco niveles cada uno con un nivel de complejidad diferente los cuales son presentados a continuación desde el comportamiento más simple al más complejo [6]:

- **Recibir:** Implica que un individuo está prestando atención a alguna situación, que se encuentra dispuesto a escuchar.
- **Responder:** El encontrarse en este nivel significa que el individuo se encuentra atendiendo y reacciona ante algún fenómeno.
- **Valorar:** Es el valor que una persona asigna a un objeto en particular, fenómeno o comportamiento, es decir que tan significativo a su parecer resulta algo que ha aprendido.
- **Organización:** Consiste en organizar valores en prioridades contrastando diferentes valores, resolviendo conflictos entre ellos y creando un sistema de valores. Este nivel hace énfasis en las acciones de comparar y sintetizar valores.
- **Valores interiorizados:** Conocido también como caracterización. Un individuo que se encuentra en este nivel posee un sistema de valores que controla su comportamiento. Este comportamiento es persuasivo, consistente, predecible y más importante aún característico de la persona.

2.1.4. DOMINIO PSICOMOTOR

El dominio psicomotor se enfoca en las habilidades físicas, las cuales incluyen habilidades motoras refinadas, manipulación de

herramientas o cualquier tipo de actividad que requiera de coordinación neuromuscular. La fuerza y la velocidad frecuentemente necesaria en habilidades físicas o la adquisición de destrezas son capturadas en este nivel [8].

Existen un total de siete niveles en esta categoría los cuales son listados a continuación desde el más simple al más complejo [6]:

- **Percepción:** Es la habilidad de usar señales sensoriales para guiar el motor de actividad. Las señales van desde estimulación sensorial, selección de señales hasta la traducción de las mismas. Un ejemplo del uso de las señales sensoriales se presenta cuando se emplea comunicación no verbal.
- **Ajustar:** Que un individuo se encuentre en este nivel significa que se encuentra en disposición para actuar, esto incluye preparación mental, física y emocional.
- **Respuesta guiada:** Corresponde a las primeras etapas en el aprendizaje de una destreza compleja, que incluye imitación y prueba y error. En este nivel una persona perfecciona lo que va aprendiendo mediante la práctica.

- **Mecanismo:** Esta categoría corresponde a la etapa intermedia del proceso de adquirir una destreza compleja. Las respuestas aprendidas se vuelven habituales y los movimientos pueden ser ejecutados con confianza y algo de eficiencia.
- **Complejidad de respuesta:** Esta categoría implica que una persona ejecuta alguna acción aprendida sin vacilar y lo hace de manera casi automática. La competencia que se ha alcanzado en este nivel se manifiesta a través de una actuación rápida, acertada y altamente coordinada que requiere un mínimo de energía.
- **Adaptación:** Un individuo que es capaz de modificar comportamientos aprendidos y adaptarlos a requerimientos de situaciones especiales se encuentra en este nivel, básicamente consiste a que es capaz de responder a escenarios no esperados.
- **Originar:** Consiste en la creación de un nuevo movimiento o comportamiento para ajustarlo a una situación en particular o a un problema en específico, Los resultados de aprendizaje hacen énfasis en la creatividad basada en el desarrollo de

habilidades de alto nivel que tiene una persona que ha alcanzado esta categoría.

2.2. CLASIFICACIÓN DE TEXTO

La clasificación de texto, también conocida como categorización de texto, puede ser definida como la actividad de etiquetar textos que se encuentran en lenguaje natural con categorías temáticas obtenidas previamente de un conjunto predefinido [9]. Lo anteriormente mencionado significa que dado una serie de textos que se encuentran previamente clasificados en función de algún criterio, pueden ser empleados para extraer las reglas que determinan a que categoría pertenece cada uno de ellos y en función de estas reglas clasificar nuevos elementos que se encuentren escritos en lenguaje natural y cuya categoría es desconocida.

Formalmente, si d_i es un documento perteneciente a un conjunto de documentos D y $\{c_1, c_2, \dots, c_n\}$ es el conjunto de todas las categorías, entonces un clasificador de texto asigna una categoría c a un documento d_i . En otras palabras la categorización de texto es el proceso de clasificar un documento bajo una categoría predefinida [10].

Minería de texto son términos que se ha venido usando con bastante frecuencia para denotar las actividades de analizar grandes conjuntos de datos, detectar patrones y tratar de extraer información probablemente útil que no es obvia a simple vista. Bajo este enfoque, también es posible afirmar que la clasificación de texto es una instancia de la minería de texto [9].

La clasificación de texto incluye clasificación basada en texto y la clasificación basada en género. La categorización basada en texto clasifica documentos de acuerdo a los temas de los que estos tratan, también los textos pueden estar escritos en distintos géneros como por ejemplo artículos científicos, reportes, críticas de películas, publicidades, etc. El género está definido en la manera en la cual el texto fue creado, la forma en la cual fue editado, el lenguaje que emplea y la clase de audiencia a la cual se encuentra dirigido [10].

El proceso de concebir un clasificador de texto no difiere tanto de los procesos que conlleva el aprendizaje de máquina o automático, el principal inconveniente que se presenta en ambos casos es la representación del problema [8]. Un problema particular de la categorización de texto es el gran número de características, que son palabras únicas o frases que determinan a que categoría pertenece determinado texto, que pueden presentarse. Esto representa un gran

problema al momento de entrenar un algoritmo de aprendizaje y al aplicarlo a un problema de categorización de texto [10].

Considerando lo anteriormente descrito, un proceso de reducción de características es necesario con la finalidad de solucionar los inconvenientes que se presentan, mejorar el rendimiento y la escalabilidad del clasificador de texto. Sin embargo, antes de poder realizar el proceso de reducción de características es necesario obtener una representación vectorial de las características que facilite la tarea del entrenamiento del algoritmo de aprendizaje.

Estas representaciones vectoriales son llamadas vectores de características y son obtenidos a partir de los textos de un conjunto de entrenamiento previamente definido y a la vez es necesario aplicar un proceso de pre procesamiento a dichos textos, con la finalidad de eliminar información no relevante y así obtener un conjunto inicial de características [8].

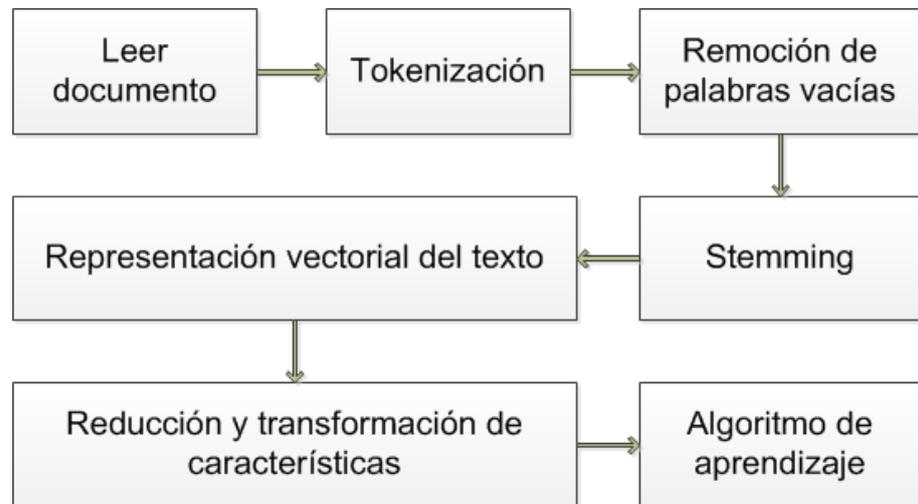


Figura 2.4 Representación gráfica del flujo del proceso de clasificación de texto

En la Figura 2.4 se muestra una representación del proceso de clasificación de texto en donde en resumen la primera etapa corresponde a obtener los textos, seguidamente se procede a aplicar procesos de pre procesamiento al texto leído para obtener una representación vectorial del mismo. Los vectores obtenidos pasan por una fase de selección o transformación para finalmente proceder a entrenar el algoritmo de aprendizaje.

2.2.1. PRE PROCESAMIENTO DE TEXTO

El componente de pre procesamiento de texto está compuesto por subprocesos que cumplen una tarea específica y que se ejecutan en orden secuencial, con el objetivo de reducir la información no

significativa de los textos y que no resulta relevante al momento de extraer características [11].

Usualmente los subprocesos que se aplican son los siguientes:

- **Tokenización:** La tokenización consiste en dividir largas cadenas de caracteres o frases en cadenas más simples como por ejemplo las palabras, esto con el objetivo de descubrir o identificar una característica.

Formalmente una palabra o token es definido como un grupo de caracteres que tienen un significado [8].

- **Remoción de palabras vacías:** La remoción de palabras vacías es otro paso muy común en el pre procesamiento de texto. Las palabras vacías son llamadas con más frecuencia *stop words* y se las define como palabras sin utilidad, por lo general son verbos auxiliares, conjunciones y artículos que no son indispensables en el entrenamiento del clasificador [10].

Por lo general las palabras vacías aparecen con una alta frecuencia en el texto lo cual constituye otra de las razones por las cuales pueden ser discriminadas. Con el objetivo de eliminar *stop words* de los textos se crean listas con estas palabras, sin embargo esto no constituye una tarea tan trivial

puesto que muchas veces el decidir si una palabra es o no vacía depende del dominio del problema que se está tratando [8].

- **Stemming:** *Stemming* es el proceso de reducir palabras a su forma canónica o raíz, la forma canónica de una palabra es la base o lema de esa palabra. Si tenemos las palabras bibliotecario y bibliotecas, su raíz o lema sería bibliotec [11].

Luego de aplicar *stemming* se reduce el *tamaño* de conjunto inicial de entrenamiento, removiendo palabras mal escritas o palabras que tienen la misma raíz y con ello se conservan las raíces de las palabras más comunes que pudiesen constituir una característica. El algoritmo más usado para *stemming* es el de Porter, otro algoritmo ampliamente empleado es el de *Snowball* [10].

Otros pasos dentro del pre procesamiento incluyen el cambiar todo el texto a minúsculas y remover signos de puntuación, estos pasos son llevados a cabo con el objetivo de reducir el trabajo de procesamiento que se debe realizar en las etapas siguientes [12].

En la Figura 2.5 se muestra los efectos de aplicar cada uno de los pasos de pre procesamiento de texto en la frase “*Yo considero que hoy en día las computadoras y el desarrollo de la tecnología en*

general han contribuido a que disfrutemos de una mejor calidad de vida, facilitando muchas tareas y acortando las distancias”.

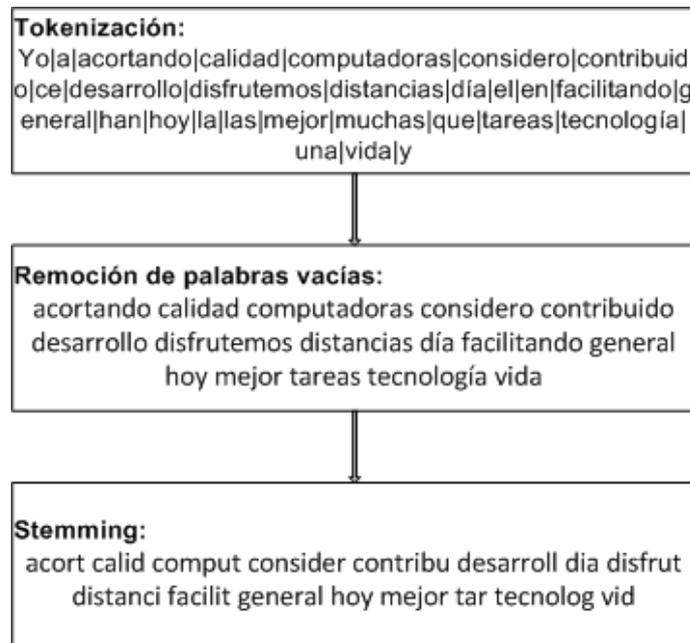


Figura 2.5 Evolución de un texto al que se le ha aplicado pre procesamiento para eliminar información no relevante

2.2.2. REPRESENTACIÓN VECTORIAL DE TEXTO

Como se mencionó anteriormente, uno de los principales problemas en la clasificación de texto es la representación del problema. El texto que representa el lenguaje natural no puede ser directamente interpretado por un clasificador o por un algoritmo que construye un modelo de clasificación, debido a esto un proceso de indexación que mapea un texto d_j en una representación compacta de su contenido debe ser aplicada uniformemente a los documentos que se emplean para entrenar y validar al clasificador.

Un texto d_j usualmente es representado como un vector de pesos de cada término $\vec{d}_j = \langle w_{1j}, \dots, w_{|T|j} \rangle$, donde T es el conjunto de términos o características que se presentan al menos una vez en al menos un documento y el peso w_{kj} representa cuál es la contribución del término en la semántica del documento d_j .

Diferentes enfoques se presentan en las siguientes cuestiones [9]:

- La definición de lo que un término o característica es.
- La manera de determinar el peso de un término.

En cuanto a la definición de término o característica, de manera general una característica puede ser un simple token, una frase o una estructura sintáctica mucho más compleja y en algunos casos también pueden constituir una característica cuantitativa en diferentes niveles lingüísticos. Las características pueden ser generales o locales, en la categorización de texto las características generales son siempre usadas pero con diferentes escalas de localidad.

Un aspecto muy importante de transformar un documento a un conjunto de tokens, es que se puede perder información lingüística como la secuencia de las palabras y dado que esto es muy importante para comprender el sentido o el mensaje que se desea

transmitir, tanto para los humanos como para un computador, muchas veces es necesario emplear frases como características en los vectores, aunque esto constituye sólo una solución parcial [13].

Otro aspecto a considerar es como calcular el peso que se le da a una característica, para empezar el peso debe mostrar el grado de información representado por el número de ocurrencias de la característica en el documento como mínimo. Un poco más complejo sería emplear un esquema en el cual el peso de la característica también representa información estadística de la ocurrencia del término en todo el conjunto de entrenamiento o en una base de conocimiento pre existente [13].

Usualmente los pesos son números que se encuentran entre cero y uno. También se emplean los pesos binarios lo cual significa que cero representa ausencia del término en un documento y uno significa la presencia del término. El emplear un esquema de pesos binario o no binario depende directamente del algoritmo de aprendizaje que se está usando [9].

Tomando en consideración lo anteriormente expuesto, un documento es usualmente representado como un arreglo de palabras, el conjunto de estas palabras obtenidas del conjunto de entrenamiento es llamado vocabulario o conjunto de características [10].

Luego del pre procesamiento de texto, cada documento es convertido en un conjunto de raíces correspondientes a las palabras del conjunto de entrenamiento, un conjunto inicial de características es formado con la unión de los conjuntos de raíces, siendo las raíces duplicadas eliminadas en el proceso con el objetivo de que una raíz sea única en el conjunto de características. A cada característica se le asigna un peso el cual es determinado de acuerdo a algún esquema, para finalmente convertir el documento en cuestión en un vector de características.

2.2.3. REDUCCIÓN DE CARACTERÍSTICAS

Un gran inconveniente que se presenta en la clasificación de texto es el gran número de vectores de características que se pueden tener, esto trae como consecuencia que se obtengan modelos de clasificación menos efectivos, que sea más complicado el proceso de entrenamiento y que se necesiten más recursos computacionales. Es por ello que antes de proceder a entrenar a un algoritmo de aprendizaje es necesario pasar por una etapa llamada reducción de dimensionalidad, cuyo efecto es reducir el espacio de vectores de características removiendo características que son consideradas irrelevantes para la clasificación [8] [9].

Dentro de los beneficios de realizar una reducción de características se encuentra una mejora considerable de la efectividad del clasificador, así como también la tendencia a reducir el *overfitting*. *Overfitting* es un fenómeno en el cual un clasificador se enfoca en las características contingentes en lugar de las características constitutivas de los datos de entrenamiento [10], esto implica que un clasificador sea bueno reclasificando datos con los que previamente ha sido entrenado pero muy malo clasificando datos completamente nuevos [9].

Dos métodos para la reducción de características son los más empleados:

- Selección de características
- Transformación de características

SELECCIÓN DE CARACTERÍSTICAS

El proceso de selección de características es llevado a cabo aplicando métodos de asignación de peso a características, con la finalidad de obtener un ranking de características y seleccionar las que se encuentran en las más altas posiciones. Dentro de los esquemas de asignación de peso se encuentran los siguientes [8]:

- **DF³**: El método DF reduce la dimensionalidad del espacio de características seleccionando términos basados en una técnica de ranking local. Primeramente, todos los documentos de entrenamiento pertenecientes a una misma categoría son puestos en un mismo grupo, a continuación todos los términos contenidos en los documentos de una misma categoría son puestos en subgrupos para así formar sub vocabularios.

A cada categoría le corresponde un sub vocabulario, en cada sub vocabulario los términos son ordenados en función del número de veces que aparecen en determinada categoría y se procede a seleccionar los que se encuentran en las posiciones más altas. Finalmente cada sub vocabulario filtrado se une con los de otras categorías para así obtener el conjunto reducido de características.

- **CF-DF⁴** : A diferencia del método DF, este método considera un valor de discriminación que es la frecuencia de categoría de un término al momento de realizar la selección, con el fin de distinguir entre documentos pertenecientes a diferentes categorías. La frecuencia de categoría de un término se define como el número de grupos en el cual el término o

³ DF por Document Frequency

⁴ CF-DF por Category Frequency-Document Frequency

característica aparece. Dos pasos son necesarios en el proceso de selección, primeramente un término es seleccionado si su categoría de frecuencia es menor que un umbral establecido; el segundo paso es seleccionar las características de acuerdo al método DF.

- **TFxIDF⁵**: Este método mide la importancia de un término en un documento determinada por el producto de la frecuencia de ocurrencia del término y la frecuencia inversa del documento.

La fórmula de este esquema está definida como:

$$IDF = \log \frac{N}{n} \quad (2.1)$$

Donde N es el número de documentos en el conjunto de entrenamiento y n es el número de documentos en el cual el término aparece. Cada una de las características es ordenada de acuerdo a este parámetro y solamente los términos con el valor más alto de TFxIDF son seleccionados para así reducir el conjunto de características.

TRANSFORMACIÓN DE CARACTERÍSTICAS

La transformación de características varía significativamente del proceso de selección, sin embargo el propósito es básicamente el

⁵ TFxIDF por Term Frequency x Inverse Document Frequency

mismo. Este enfoque no asigna peso a los términos para proceder a descartar a los que se consideran menos influyentes, sino que compacta el vocabulario basado en la concurrencia de las características.

El Componente Principal de Análisis, es un método muy empleado para la transformación de características. Su objetivo es descubrir una matriz de transformación de discriminaciones con el objetivo de reducir la complejidad de la tarea de clasificar sin afectar la efectividad del clasificador.

La matriz de covarianzas de los datos en PCA⁶ corresponde a la matriz de términos de un documento multiplicada por su transpuesta. Cada uno de los elementos de la matriz de covarianzas representan términos ocurrentes en los documentos.

Obteniendo los vectores propios de la matriz de covarianzas se obtienen “temas” o “conceptos semánticos” que son resultado de transformar simples características en unas más complejas. Este método ha demostrado tener muy buen rendimiento y provoca que los clasificadores alcancen una mayor efectividad [10].

⁶ PCA por Principal Component Analysis

2.3. ALGORITMOS DE APRENDIZAJE

Los algoritmos de aprendizajes son aquellos que toman como entrada datos empíricos y devuelven patrones o predicciones de características de los datos que recibió en primera instancia [9].

Una vez que se ha aplicado el proceso de reducción de características los documentos que componen el conjunto de entrenamiento son representados en una forma que puede ser usada como entrada de un algoritmo de aprendizaje automático y este a su vez puede empezar a clasificar el problema para finalmente obtener un modelo [8].

Existe una gran variedad en cuanto a algoritmos, clasificadores, técnicas empleadas para la clasificación, que por lo general difieren del enfoque adoptado, entre ellos tenemos: árboles de decisión, clasificador Naïve Bayes, reglas de inducción, redes neuronales, vecinos más cercanos y máquinas de soporte de vectores. Aunque distintas opciones se han propuesto, la clasificación automática de texto es un área que aún requiere de investigación debido a que el desempeño de los clasificadores actuales no está libre de fallos y su rendimiento aún requiere de perfeccionamiento [10].

Estudios actuales manifiestan que los clasificadores que muestran mejores resultados y una mayor efectividad son los clasificadores bayesianos [12] y las máquinas de soporte de vectores [13].

2.3.1. MÁQUINA DE SOPORTE DE VECTORES

Las máquinas de soporte de vectores o SVM⁷ es uno de los enfoques recientes que está tomando el campo del aprendizaje automático, es en muchos casos considerada una alternativa más viable en comparación con técnicas más maduras como redes neuronales y algoritmos de vecinos más cercanos.

SVM es un algoritmo de aprendizaje automático supervisado que es entrenado para separar dos conjuntos de datos, empleando ejemplos de ambos conjuntos. En su forma más simple, cuando se emplea las máquinas de soporte de vectores como clasificadores binarios, en el proceso de entrenamiento se construirá un hiper plano el cual actúa como una superficie de decisión entre los dos conjuntos de datos. Una vez que el entrenamiento está completo, nuevos datos pueden ser clasificados determinando donde se encuentra en relación al hiper plano [12]. En la Figura 2.6 se muestra un hiper plano obtenido de un clasificador SVM binario.

⁷ SVM por Support Vector Machines

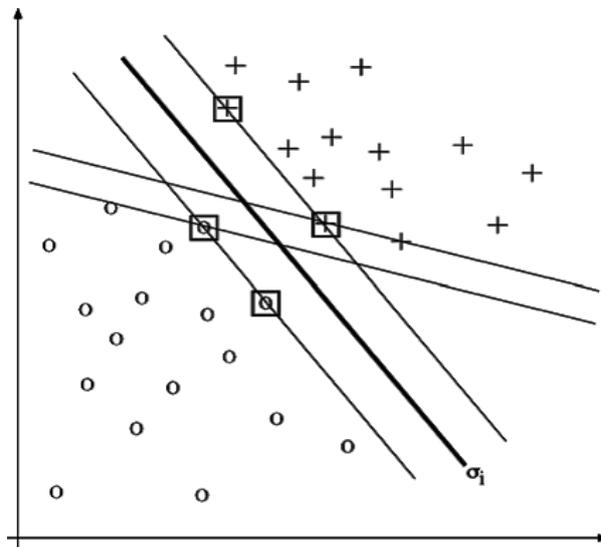


Figura 2.6 Híper plano de un clasificador SVM binario

Cuando las SVM son usadas en la tarea de la clasificación de texto, proveen una precisión excelente pero una pobre recuperación⁸ Una manera de mejorar el grado de recuperación de las máquinas de soporte de vectores es ajustando el umbral asociado a la máquina, esto podría contribuir a obtener mejores resultados [10].

Existen dos ventajas que ofrecen las máquinas de soporte de vectores en cuanto a la categorización de texto [9]:

- Por lo general la reducción de características no es necesaria, puesto que una SVM tiende a ser altamente robusta y es

⁸ Recuperación es más comúnmente llamada *recall*, es la probabilidad que si un texto aleatorio debería ser clasificado bajo una categoría c , esta decisión es tomada.

capaz de escalar para manejar dimensionalidades considerables.

- No es necesario un gran esfuerzo por parte de un humano o del computador para poner a punto un conjunto de validación, puesto que el escoger valores de parámetros predeterminados frecuentemente brinda una mayor efectividad

Las máquinas de soporte de vectores se ajustan muy bien al problema de la clasificación de texto debido a las características inherentes del texto como su alta dimensionalidad, la dispersión de los vectores y la mayoría de los problemas de clasificación de texto son linealmente separables [12].

2.3.2. CLASIFICADORES BAYESIANOS

Los clasificadores bayesianos son clasificadores probabilísticos, los cuales aplican el teorema de Bayes para determinar la probabilidad de que un documento representado por un vector $\vec{d}_j = \langle w_{1j}, \dots, w_{|T|j} \rangle$ de términos pertenece a una categoría c_i . El teorema de Bayes está dado por la siguiente ecuación:

$$P(c_i | \vec{d}_j) = \frac{P(c_i)P(\vec{d}_j | c_i)}{P(\vec{d}_j)} \quad (2.2)$$

En la ecuación el espacio de eventos es el espacio de documentos; $P(\vec{d}_j)$ es la probabilidad que un documento aleatoriamente escogido tenga al vector \vec{d}_j como su representación y $P(c_i)$ es la probabilidad que un documento aleatoriamente elegido pertenece a la categoría c_i . La estimación de $P(\vec{d}_j | c_i)$ resulta ser un inconveniente debido a que número de posibles vectores \vec{d}_j es muy elevado, por ello con el fin de apaciguar este problema se hace la asunción que cualquiera dos coordenadas del vector que representa al documento son estadísticamente independientes una de la otra.

Los clasificadores probabilísticos que usan la asunción antes descrita, son llamados clasificadores *Naïve Bayes*. Uno de los clasificadores *Naïve Bayes* más conocidos es el clasificador de independencia binaria, el cual resulta de emplear vectores de características con pesos binarios [9].

El clasificador *Naïve Bayes* es frecuentemente empleado en aplicaciones de clasificación de texto y experimentos debido a su simplicidad y efectividad. Sin embargo, su rendimiento se ve degradado dado que no modela el texto bien. Los problemas que muestra este clasificador pueden ser resueltos con simples

correcciones, además estudios recientes muestran que empleando redes Bayesianas con enfoque de árbol es posible manejar la clasificación de texto con miles de variables con una velocidad y efectividad aceptables [10].

2.4. ANÁLISIS COMPARATIVO DE SOLUCIONES EXISTENTES

El disponer de un sistema que clasifique automáticamente texto, opiniones, preguntas, de acuerdo a la taxonomía de Bloom es un tema del cual se han propuesto varias soluciones y enfoques. Uno de ellas es la sugerida por Nurazah Yusof y Chai Jing de la Universidad de Malasia, quienes propusieron construir un sistema que clasificara las preguntas empleadas en exámenes de acuerdo al dominio cognitivo de la taxonomía de Bloom.

Su solución empleaba una red neuronal como clasificador, la cual fue entrenada empleando un algoritmo de aprendizaje de escala conjugada. Uno de los mayores problemas que experimentaron fue la pobre escalabilidad de la red neuronal, por lo que fue necesaria la aplicación de varios métodos de reducción de características. En este estudio también determinaron que dadas las particularidades de su conjunto de entrenamiento, el método más efectivo de reducción de

características fue el DF dado que mantenía la precisión mientras mejoraba la velocidad de convergencia.

En cuanto a la efectividad obtenida fue de 65.9%, empleando para el entrenamiento un total de 192 preguntas y obteniendo un total de 605 características [8].

Una investigación similar llevada a cabo por Anwar Ali y Addin Osman de la Universidad de Najran tenía la misma finalidad que la investigación anteriormente expuesta, pero en este caso en lugar de emplear una red neuronal emplearon máquina de soporte de vectores. Dentro de los resultados que obtuvieron evidenciaron que usando SVM obtuvieron resultados satisfactorios con respecto a la efectividad y precisión de la clasificación, sin embargo el no disponer de un conjunto de entrenamiento grande provocaron que el valor resultante de recuperación sea bajo y con ello no les fue posible obtener resultados más concluyentes.

El conjunto de entrenamiento que emplearon estaba compuesto por 190 preguntas las cuales se encontraban uniformemente distribuidas entre los diferentes niveles de la taxonomía, a pesar de no ser un número elevado de ejemplos obtuvieron indicadores de efectividad de alrededor 87% lo cual es superior al obtenido a 65.9% que fue el resultado obtenido empleando una red neuronal como clasificador [12].

De manera similar Yi-Hsing Chang y Huan-Wen Chen de la Universidad de Taiwán proponen un sistema de inferencia para analizar de manera automática la calidad de ítems usados en pruebas basados en la taxonomía de Bloom. En su propuesta, la tarea de la clasificación era llevada a cabo por un sistema experto basado en reglas, en donde la determinación de las reglas de la clasificación constituyó una de las tareas más complicadas y cruciales.

Los mejores resultados que obtuvieron en cuanto a efectividad fueron del 51% y concluyen que si las palabras o verbos empleados en cada nivel no se encuentran en la base de conocimiento, el motor de inferencias no podrá efectuar la tarea con éxito [2].

Laurie Cutrone y Maiga Chang de la Universidad de Athabasca propusieron la arquitectura de un sistema que dentro de un LMS⁹ que sea capaz de calificar automáticamente las respuestas de estudiantes a preguntas que son catalogadas de abiertas, es decir preguntas en las que no hay una respuesta definida.

Si bien en este caso no se busca clasificar las respuestas de acuerdo a la taxonomía de Bloom, se llevan a cabo las mismas etapas del pre procesamiento de texto que se efectuaron en casos anteriores y se

⁹ LMS por Learning Management System, sistema de administración de conocimiento básicamente es una aplicación que facilita el compartir conocimientos y la comunicación entre estudiantes y profesores en un entorno de enseñanza – aprendizaje.

montó esta arquitectura en una aplicación web. Lamentablemente los resultados obtenidos no fueron suficientemente buenos y la funcionalidad del sistema se limitaba a calificar respuestas de no más de dos líneas [11].

Otra tentativa de clasificar preguntas de acuerdo a la taxonomía de Bloom fue llevada a cabo por Wen-Chih Chang y Ming-Shung Chung de la Universidad de Taiwán, quienes diseñaron un sistema online en el cual los profesores ingresaban las preguntas que deseaban hacer en un test y el sistema respondía en qué nivel de la taxonomía se encontraba dicha pregunta. Era posible también para los estudiantes rendir su examen a través del mismo sistema.

La arquitectura del sistema no contempló la inclusión de un clasificador automático, el determinar el nivel en el que se encontraba la pregunta dependía de un conjunto de palabras y verbos asociados a cada nivel los cuales se encontraban almacenados en una base de datos y para llegar a la conclusión era necesario comparar cada palabra de la pregunta de la cual se desea saber el nivel con las palabras que se encontraban en la base de datos y de esa forma el nivel del cual se tenía más palabras era la categoría que se le asignaba a la pregunta.

El mejor resultado que obtuvieron fue del 75% de efectividad para el nivel de conocimiento, mientras que para los otros niveles fue muy bajo, pues el nivel de efectividad fue de alrededor 20% [14].

Algo que tienen en común las propuestas mencionadas anteriormente, es que todas ellas emplean casi los mismos pasos para el pre procesamiento de texto y la reducción de características, mientras que la gran diferencia radica en el método de clasificación que emplean. Comparando estos pocos estudios es rescatable que el uso de algoritmos de aprendizaje para efectuar la clasificación da mejores resultados que simplemente emplear comparaciones contra una base de datos.

Además, solamente tomando en consideración los casos aquí expuestos, es notorio que el emplear máquinas de soportes de vectores provee una mayor efectividad que la obtenida empleando otros tipos de clasificadores como redes neuronales o sistemas expertos. Sin embargo, cabe recalcar que en ninguna de las soluciones propuestas se alcanza una efectividad del 100%, siendo las principales causas la no existencia de un clasificador totalmente efectivo y el no disponer de conjuntos de entrenamiento lo suficientemente grandes y buenos.

Esto último resulta ser un factor altamente influyente en el porcentaje de efectividad alcanzado por los sistemas, si no se tienen un número de ejemplos uniformemente distribuido entre los niveles de la taxonomía, los resultados serán buenos en unos casos, regulares en otros y pueden ser también totalmente malos.

También queda establecido el precedente de que es posible montar estas arquitecturas en aplicaciones web y sistemas online por lo que resulta viable pensar incluir un sistema de clasificación automática dentro de un LMS. Adicionalmente los casos expuestos buscaban en su mayoría categorizar solamente preguntas e ítems que se usarían en pruebas y exámenes y no textos de otra naturaleza.

Otro aspecto que vale la pena recalcar, es que todos estos sistemas se encuentran implementados para el idioma inglés y los algoritmos usados en pasos previos a la clasificación, como los algoritmos de stemming, se encuentran optimizados para este lenguaje por lo que implementar un sistema de este tipo para texto que se encuentra en idioma español constituye un aporte significativo en esta línea de investigación.

En la Tabla 2.1 se muestra a manera de resumen las principales características de las soluciones existentes en cuanto a clasificación automática de acuerdo a la Taxonomía de Bloom.

| Autores de la propuesta | Clasificador | Efectividad | Conjunto de Entrenamiento |
|---------------------------|--|-------------|--|
| Nurazah Yusof y Chai Jing | Red neuronal | 65.9% | 192 preguntas 605 características |
| Anwar Ali y Addin Osman | SVM | 87% | 190 preguntas uniformemente distribuidas entre los niveles de la taxonomía |
| Yi-Hsing y Huan-Wen Chen | Sistema experto | 51% | Base de conocimiento formada por palabras y verbos empleados en cada nivel de la taxonomía |
| Wen-Chih y Ming-Shung | Sin clasificador, uso de base de datos | 75% | Base de datos con verbos y palabras asociados a cada nivel de la taxonomía |

Tabla 2.1 Principales características de soluciones existentes para clasificación automática de acuerdo a la Taxonomía de Bloom

CAPÍTULO 3

3. ANÁLISIS Y DISEÑO DE LA SOLUCIÓN

En el presente capítulo se expone el análisis que se realizó para diseñar la solución propuesta, así como también los procedimientos, metodologías y métricas utilizadas en la evaluación del sistema. Se presentan descripciones de los módulos constituyentes del sistema y del conjunto de entrenamiento empleado, el capítulo finaliza con el diseño y plan de pruebas.

3.1. ANÁLISIS Y DISEÑO DEL SISTEMA

El sistema está constituido por dos partes principales:

- Interfaz de usuario y comunicación
- Motor de inferencia

La interfaz de usuario es el medio a través del cual los usuarios podrán hacer uso de las funcionalidades del sistema y ver los resultados de la clasificación de acuerdo al nivel cognitivo de la taxonomía de Bloom. Esta interfaz se la construyó en la plataforma SIDWeb 4 dentro del módulo de foros de discusión de esta plataforma.

SIDWeb 4 es capaz de comunicarse con el motor de inferencia, envía los textos que el usuario desea etiquetar, estos son categorizados y las respuestas devueltas a SIDWeb 4 son almacenadas en su base de datos, para finalmente ser mostrados al usuario.

El motor de inferencia se encuentra formado por dos componentes:

- Componente de aprendizaje
- Componente de clasificación

Para llevar a cabo la predicción de la etiqueta el motor de inferencia necesita de un modelo de clasificación, este modelo es producido previamente por el componente de aprendizaje haciendo uso de procesos de minería de texto, algoritmos de clasificación y un conjunto de entrenamiento previamente clasificado por expertos.

El componente de clasificación recibe el texto cuya etiqueta es desconocida, este es pre procesado y se le aplican las reglas obtenidas del modelo generado anteriormente, para finalmente

devolver la categoría a la que más probablemente pertenece dicho texto; esta es la respuesta que se entrega a SIDWeb 4 y este último es el responsable de mostrar los resultados al usuario.

Cabe recalcar que en la interfaz de usuario aparte de mostrar las etiquetas de las respuestas a los foros, también se presentan una serie de estadísticas y gráficos los cuales son obtenidos de la información almacenada en la base de datos de SIDWeb 4 y no es necesaria una comunicación constante con el motor de inferencia para su generación.

En la Figura 3.1 se muestra un esquema de la arquitectura del sistema previamente expuesta.

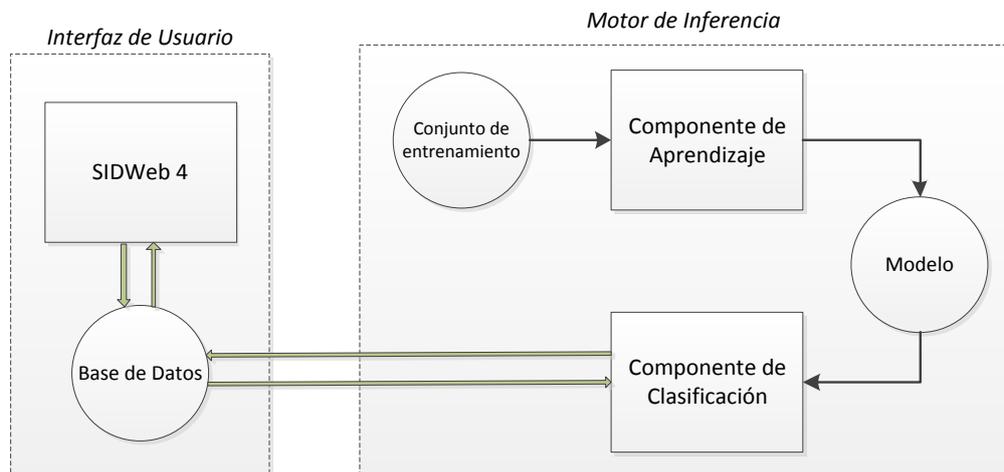


Figura 3.1 Esquema de la arquitectura del sistema de clasificación automática

3.1.1. ANÁLISIS Y DISEÑO DEL COMPONENTE DE APRENDIZAJE

Como se expuso en la sección anterior este componente tiene la finalidad de producir un modelo que permita clasificar las respuestas de los foros de discusión de acuerdo al dominio cognitivo de la taxonomía de Bloom. Un esquema de la estructura del componente de aprendizaje se muestra en la figura mostrada a continuación:

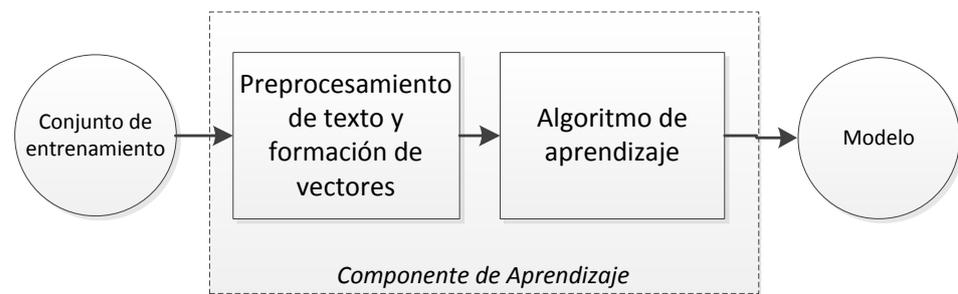


Figura 3.2 Esquema de la estructura del componente de aprendizaje

El componente de aprendizaje recibe como entrada un conjunto de entrenamiento, el cual será utilizado para construir el modelo. El conjunto de entrenamiento está constituido por respuestas generadas por estudiantes en foros de discusión y que han sido categorizadas de acuerdo a la taxonomía de Bloom previamente por expertos. La etiqueta de cada uno de los elementos del conjunto de entrenamiento es conocida al momento de ser procesados por el componente de aprendizaje.

Las respuestas del conjunto de entrenamiento deben pasar a través del subcomponente de pre procesamiento de texto y formación de vectores, el cual es responsable de representar el problema de tal forma que sea entendido por un algoritmo de aprendizaje. Una representación vectorial será la forma que tome el texto, los vectores estarán formados por características extraídas de cada categoría y cada una de esas características tendrá un valor asociado que indicará que tan influyente es una característica para determinada categoría.

Un esquema del subcomponente de pre procesamiento de texto y formación de vectores se muestra en la siguiente figura:

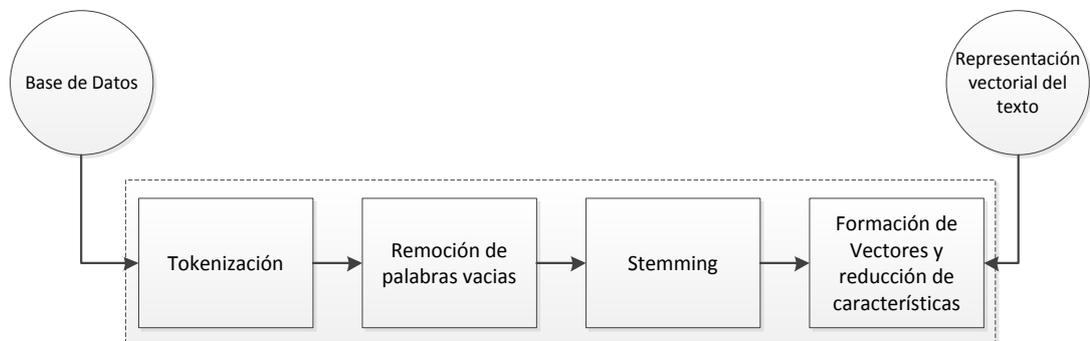


Figura 3.3 Esquema de la estructura del subcomponente de pre procesamiento de texto y formación de vectores

En la Figura 3.3 se observa que los elementos del conjunto de entrenamiento atraviesan varias etapas hasta obtener su representación vectorial. Es necesario eliminar toda información que no resulta relevante ni influyente, es por ello que un pre

procesamiento de los textos que forman el conjunto de entrenamiento es indispensable. El pre procesamiento de texto conlleva tokenización, eliminación de palabras vacías y stemming; mediante la tokenización se logra aislar cada una de las palabras que conforman los textos, para posteriormente eliminar aquellas palabras como intersecciones, conjunciones y verbos auxiliares que no afectan el mensaje que se desea transmitir y finalmente se extraen las raíces de las palabras consideradas como relevantes a través de un algoritmo de stemming.

En este punto se tienen listados de raíces de palabras por cada una de las categorías de la taxonomía, estas palabras pasan a formar los vectores de características. Debido a que el número de características que forman los vectores puede llegar a ser muy alto y por ende afectar el rendimiento y efectividad del sistema, un proceso de reducción de características es necesario.

El método de reducción de características a aplicar será DF, puesto que para este tipo de problemas resulta ser el que ofrece mejores resultados en cuanto a efectividad y tiempos de convergencia [7]. Una vez que se han extraído las características más representativas de cada uno de los vectores, estos se encuentran listos para pasar a ser la entrada del algoritmo de aprendizaje.

El algoritmo de aprendizaje o clasificador es el subcomponente que genera el modelo que se desea obtener. El algoritmo que se emplea para realizar la clasificación es el clasificador Naïve Bayes, el cual fue escogido debido a su adaptación a este tipo de problemas, simplicidad, efectividad y por sobre todo por la forma en la que vienen representados los datos.

El clasificador realiza una serie de procedimientos y validaciones para finalmente obtener un modelo de naturaleza probabilística, el cual está formado por las características extraídas del conjunto de entrenamiento y las probabilidades asociadas de cada característica por cada categoría que en conjunto son empleadas al momento de determinar el nivel al cual pertenece algún texto.

3.1.2. ANÁLISIS Y DISEÑO DEL COMPONENTE DE CLASIFICACIÓN

El componente de clasificación es el encargado de determinar cuál es el nivel dentro de la taxonomía al que pertenece algún texto que se desea clasificar. Este módulo recibe como entrada un texto del cual se desconoce su etiqueta, este texto se encuentra escrito en lenguaje natural razón por la cual antes de poder predecir la etiqueta del mismo es necesario aplicar operaciones de pre procesamiento.

Las operaciones de pre procesamiento que se aplican al texto son las mismas que se aplicaron a los elementos del conjunto de

entrenamiento y se lo realiza con el objetivo de eliminar palabras poco relevantes para el sentido del mensaje que se desea transmitir, palabras mal escritas y signos de puntuación, de esta forma se extraen las principales características del texto a clasificar y con ello se encuentra preparado para aplicar el modelo de clasificación.

El determinar si un texto pertenece o no a un nivel se lleva a cabo aplicando el modelo probabilístico aprendido, esto implica que dadas las características del texto se calcula la probabilidad de que pertenezca a una categoría o no. El resultado devuelto es el nivel de la taxonomía al que más probablemente pertenezca el texto.

Un esquema del componente de clasificación es mostrado en la Figura 3.4.

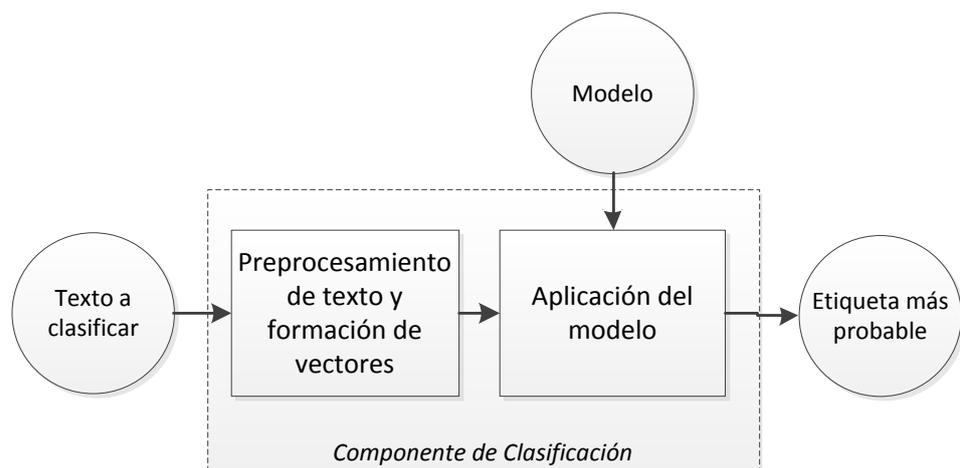


Figura 3.4 Esquema de la estructura del componente de clasificación

3.2. DISEÑO DEL MÓDULO DE CLASIFICACIÓN EN SIDWeb 4

El módulo de clasificación en SIDWeb 4 es el medio mediante el cual el usuario interactúa con el sistema de categorización automática, es decir, este es el módulo que contiene la interfaz de usuario y es el ente a través del cual los profesores pueden observar las respuestas a los foros de discusión, su respectiva etiqueta y visualizar las estadísticas del foro y de los niveles de la taxonomía de Bloom bajo los cuales se encuentran las respuestas emitidas.

Este módulo será también el responsable de iniciar la comunicación con el clasificador automático, de preparar las respuestas sin clasificar a ser enviadas y procesar las respuestas devueltas por el motor de inferencia. Cuando un usuario se disponga a visualizar las respuestas a un foro de discusión este módulo solicitará al motor de inferencia la etiqueta para las colaboraciones que no han sido clasificadas y serán mostradas junto a cada respuesta emitida sin necesidad de que el usuario ejecute alguna acción adicional.

Un aspecto importante a considerar son los tiempos de respuesta; en el caso que cada vez que un usuario ingresa a visualizar las

respuestas de un foro y se envían los requerimientos al sistema de clasificación, provocará que los tiempos requeridos para mostrar una respuesta al usuario tiendan a ser largos. Para solucionar este inconveniente, una vez que una respuesta ya tenga asignada una etiqueta esta será almacenada en la base de datos de SIDWeb 4 y de esta forma sólo será necesario clasificar aquellas respuestas que no dispongan de una etiqueta.

Otra de las ventajas de tener almacenadas las etiquetas en una tabla es que permite la elaboración de distintos gráficos estadísticos y la presentación de indicadores de la actividad de los estudiantes en el foro de discusión y con ello el profesor puede tener una visión más específica del desempeño de sus alumnos en el desarrollo de la tarea asignada.

Los gráficos que se presentan en este módulo son:

- Diagrama de pastel del número de contribuciones por estudiantes en un foro de discusión.
- Diagrama de pastel del número de contribuciones por nivel de la taxonomía de Bloom en un foro de discusión.
- Diagrama de barras del número de contribuciones de un estudiante por nivel de la taxonomía de Bloom en un foro de discusión.

3.2.1. MODELO LÓGICO DEL MÓDULO DE CLASIFICACIÓN

Como se expuso anteriormente debido a que los tiempos de respuesta del sistema pueden llegar a ser muy altos si se solicita constantemente que el motor de inferencia devuelva una etiqueta por cada respuesta, es necesario almacenar en la base de datos de SIDWeb 4 la etiqueta de las contribuciones de los estudiantes de tal forma que solo se solicita respuestas de las respuestas nuevas y no de las existentes.

Esto implica que es necesario modificar la estructura de la base de datos del SIDWeb 4 y lo más conveniente es que los cambios no representen un gran impacto que podrían llegar a afectar el desempeño de la aplicación. Por ello el único cambio en la estructura fue el de añadir un campo tipo texto que almacena la etiqueta, esta modificación se la realizó en la tabla que almacena las respuestas a los foros de discusión.

El campo es llamado *tag* y su valor es *null* en el caso que la respuesta no tenga una etiqueta asignada o el nombre del nivel de la taxonomía en el caso que ya posea una categoría. Este nuevo campo es el que se emplea principalmente en la obtención de los gráficos estadísticos que se proveen.

En la Figura 3.5 se muestra la modificación en la tabla *discussion_entries*, que es la que almacena las contribuciones de los estudiantes en los foros de discusión.

| discussion_entries | |
|--------------------|----------------------------|
| PK | <u>id</u> |
| | message |
| | discussion_topic_id |
| | user_id |
| | parent_id |
| | created_at |
| | updated_at |
| | attachment_id |
| | workflow_state |
| | deleted_at |
| | migration_id |
| | editor_id |
| | root_entry_id |
| | depth |
| | tag |

Figura 3.5 Tabla *discussion_entries* con campo *tag* adicionado

3.3. DEFINICIÓN DEL CONJUNTO DE ENTRENAMIENTO

El conjunto de entrenamiento empleado para obtener el modelo de clasificación está conformado por respuestas emitidas por estudiantes a diferentes temas de discusión planteados de acuerdo a una asignatura en particular, estas respuestas fueron recogidas desde el año 2010 hasta el año 2012 a través de los foros de discusión de SIDWeb 3.6 para los años 2010, 2011 y SIDWeb 4 para el año 2012. Las respuestas fueron etiquetadas por diferentes expertos.

Cabe recalcar que del total de respuestas recogidas, solo el 70% forman parte del conjunto de entrenamiento; el otro 30% pasa a formar

parte del conjunto de pruebas y este se encuentra definido en la sección de diseño de pruebas. Estos porcentajes se decidieron en función a otros estudios en donde la distribución mencionada ha dado los mejores resultados [8] [12].

Las respuestas a los foros de discusión recolectados fueron etiquetadas cada una por tres diferentes codificadores, quienes las categorizan de acuerdo a su criterio personal, comparan la etiqueta asignada por cada uno de ellos y después de una discusión y exposición de razones del porqué de la etiqueta asignada, llegan a un consenso y proceden a asignar el nivel de la taxonomía de Bloom que entre ellos consideran el más apropiado. Cabe recalcar que las respuestas fueron llevadas del idioma español al inglés para que los codificadores procedieran con sus análisis, debido a que no dominaban la lengua española.

Otro aspecto que es necesario recalcar, es que los codificadores fueron diferentes cada año, aunque el número de ellos y el procedimiento para proceder a asignar el nivel de la taxonomía de Bloom a las respuestas fue el mismo.

El subconjunto de entrenamiento uno fue recogido en el año 2010. Se planteó un caso de estudio relacionado a la seguridad informática a un grupo de estudiantes de maestría en seguridad informática de la

ESPOL; los estudiantes fueron separados en cinco grupos de cuatro estudiantes y cada grupo debía proveer una solución a la problemática planteada en el caso de estudio y debían alcanzar conclusiones en un lapso de 48 horas.

El total de respuestas por nivel del año 2010 se muestran en la Tabla 3.1.

| Nivel de la taxonomía de Bloom | Número de Respuestas |
|--------------------------------|----------------------|
| Crear | 0 |
| Evaluar | 8 |
| Analizar | 9 |
| Aplicar | 0 |
| Entender | 13 |
| Recordar | 11 |
| | 41 |

Tabla 3.1 Número de respuestas etiquetadas por nivel del subconjunto de entrenamiento del año 2010

El subconjunto de entrenamiento dos fue obtenido en el año 2011 y correspondieron a respuestas de estudiantes de la materia Ingeniería de Software I a un foro de discusión del cual debían entregar como resultado un documento de riesgos correspondientes a los proyectos que se encontraban desarrollando. La actividad fue realizada por treinta y cinco estudiantes y se los dividió en siete grupos de cinco miembros cada uno.

Un particular en la actividad realizada en el 2011, es que los estudiantes fueron capacitados sobre cómo clasificar sus respuestas de acuerdo de la taxonomía de Bloom y debían etiquetar cada una de sus contribuciones de acuerdo a lo que aprendieron. Posteriormente las etiquetas de los estudiantes fueron comparadas con las asignadas por los codificadores expertos con la finalidad de llegar a un consenso y determinar la categoría final.

| Nivel de la taxonomía de Bloom | Número de Respuestas |
|--------------------------------|----------------------|
| Crear | 0 |
| Evaluar | 2 |
| Analizar | 22 |
| Aplicar | 7 |
| Entender | 46 |
| Recordar | 30 |
| | 107 |

Tabla 3.2 Número de respuestas etiquetadas por nivel del subconjunto de entrenamiento del año 2011

En la Tabla 3.2 se muestra en detalle el número de respuestas por nivel de este subconjunto.

En el año 2012 las discusiones se llevaron a cabo en la asignatura Aplicaciones Multimedia Interactivas, se efectuaron dos discusiones siendo una de ellas acerca de decidir cuál es el mejor framework web dadas ciertas condiciones y la segunda decidir cuál es el mejor

framework para desarrollo de aplicaciones móviles. En esta ocasión los estudiantes debían adoptar distintos roles dentro de la discusión, los roles y la actividad que estos conllevaban son:

- *Moderador*: Generar preguntas críticas, explorar si se está haciendo un verdadero significado de lo que se coloca, solicitar participación de todos.
- *Detective*: Busca información externa, fuentes adicionales. Debe traer ideas nuevas a la discusión.
- *Evaluador*: Debe crear casos pruebas de que la solución funciona.
- *Resumidor*: Debe generar resúmenes de las contribuciones, debe iniciar y finalizar la discusión.

| Nivel de la taxonomía de Bloom | Número de Respuestas |
|--------------------------------|----------------------|
| Crear | 0 |
| Evaluar | 6 |
| Analizar | 32 |
| Aplicar | 0 |
| Entender | 155 |
| Recordar | 78 |
| | 271 |

Tabla 3.3 Número de respuestas etiquetadas por nivel del subconjunto de entrenamiento del año 2012

Los estudiantes fueron divididos en cuatro grupos de seis a ocho integrantes y cada equipo de trabajo debía obtener un documento final es donde exponían sus propuestas y explicaban la razón por la cual habían escogido determinada plataforma. El número de respuestas obtenidas por nivel se muestra en la Tabla 3.3.

El conjunto de entrenamiento final está formado de la unión de los subconjuntos de entrenamiento de los años 2010, 2011 y 2012, quedando establecido finalmente como se muestra en la Tabla 3.4.

| Nivel de la taxonomía de Bloom | Número de Respuestas |
|--------------------------------|----------------------|
| Crear | 0 |
| Evaluar | 17 |
| Analizar | 63 |
| Aplicar | 7 |
| Entender | 214 |
| Recordar | 119 |
| | 420 |

Tabla 3.4 Número de respuestas etiquetadas por nivel del conjunto de entrenamiento

Un aspecto que es necesario recalcar es que a pesar de que se recogieron datos de tres años distintos, el número de respuestas no se encuentra uniformemente distribuido entre todos los niveles y esto en muchos casos afecta el rendimiento de los clasificadores automáticos [2] [14].

3.4. ALCANCE DEL DISEÑO

El sistema desarrollado fue construido con el propósito de obtener un modelo de clasificación a partir de las decisiones tomadas por expertos y las predicciones y resultados devueltos por el sistema se basan exclusivamente de la información extraída del conjunto de entrenamiento.

Esto quiere decir que el funcionamiento y desempeño del sistema en cuanto a la asignación de las etiquetas es dependiente de lo que haya aprendido y lo que logró extraer de un conjunto de entrenamiento; si el conjunto de entrenamiento no es bueno entonces los resultados de la clasificación tampoco lo serán.

El sistema desarrollado no es evolutivo, esto quiere decir que el sistema realizará la clasificación siempre empleando el mismo modelo el cual no cambiará por si solo en el transcurso del tiempo, salvo en el caso que se vuelva a entrenar el sistema con un conjunto de entrenamiento diferente y se emplee el nuevo modelo obtenido para realizar la categorización.

Las etiquetas asignadas en SIDWeb 4 no cambiarán en el caso que el modelo de clasificación haya sido modificado, esto se debe a que una vez que una respuesta ha sido etiquetada dicha etiqueta es almacenada en la base de datos de SIDWeb 4 y en la siguiente

consulta ya no se envía a predecir la etiqueta nuevamente, sino que se muestra la almacenada localmente.

3.5. DISEÑO DE PRUEBAS

Las pruebas permiten evaluar la efectividad con la que el sistema realiza la clasificación de las respuestas a los foros de discusión. La forma en la que se evalúa la efectividad del sistema es a través de varias medidas que se presentan en una tabla de contingencia, la cual está formada por lo general de los siguientes valores [12]:

- *A*: El número de respuestas que el sistema asigna correctamente a la categoría, es decir, los verdaderos positivos.
- *B*: El número de respuestas que el sistema incorrectamente asigna a la categoría, estos constituyen los falsos positivos.
- *C*: El número de respuestas que pertenecen a la categoría pero que el sistema las asigna a otra, es decir los falsos negativos.
- *D*: El número de respuestas que el sistema correctamente no asigna a la categoría, lo cual constituye los verdaderos negativos.

Las medidas comúnmente empleadas para determinar la efectividad de un clasificador son las siguientes [8] [12]:

- **Exactitud:** Es la probabilidad de que un documento o respuesta d_x es clasificado bajo una categoría c_x , esta decisión es la correcta. Su valor se encuentra en el rango de cero a uno, siendo uno el mejor valor.

$$P = \frac{A}{A+B} \quad (3.1)$$

- **Recuperación:** Usualmente conocida como *recall*, es la probabilidad que si un texto aleatorio debería ser clasificado bajo una categoría c , esta decisión es tomada.

$$R = \frac{A}{A+C} \quad (3.2)$$

- **Precisión:** Esta es la medida más comúnmente empleada para medir la efectividad de un clasificador, sin embargo, los valores de precisión son mucho menos renuentes a variaciones en el número de los verdaderos positivos que las medidas de exactitud y reconocimiento. Esta métrica es más conocida como *accuracy*.

$$Acc = \frac{A+C}{A+B+C+D} \quad (3.3)$$

- F_β : Es la media armónica del reconocimiento y la exactitud. El valor de β depende de la importancia que se le da a la

exactitud y a la recuperación; si la exactitud es considerada más importante que el reconocimiento entonces el valor de β debe ser cero. Si la recuperación es más importante que la exactitud el valor de β debe ser llevado a infinito, por último si la exactitud y el reconocimiento son igual de importantes el valor de β debe ser igual a uno.

La ecuación de F_β para un valor de $\beta = 1$ está definida de la siguiente manera:

$$F_\beta = \frac{2RP}{R+P} \quad (3.4)$$

Para determinar los valores de A, B, C y D es necesario contar con un conjunto de prueba. El conjunto de pruebas o *testing set*, está formado por respuestas de las cuales se conoce el nivel de la taxonomía Bloom al que pertenecen y no formaron parte de conjunto de entrenamiento; la prueba consiste en que el clasificador categorice las respuestas del conjunto de pruebas y la respuesta devuelta por el clasificador es comparada con la etiqueta que realmente posee la respuesta, de esta forma se determinan los verdaderos positivos y negativos y los falsos positivos y negativos.

| Nivel de la taxonomía de Bloom | Número de Respuestas |
|--------------------------------|----------------------|
| Crear | 0 |
| Evaluar | 7 |
| Analizar | 27 |
| Aplicar | 3 |
| Entender | 92 |
| Recordar | 51 |
| | 180 |

Tabla 3.5 Número de respuestas por nivel del conjunto de pruebas

En la Tabla 3.5 se presentan el número de respuestas por nivel del conjunto de entrenamiento, estos números corresponden al 30% del total de respuestas etiquetadas.

Es necesario determinar también si el clasificador automático puede actuar como un codificador humano más, recordemos que el proceso de clasificación por lo general se lo realiza entre varias personas y la decisión que se toma es por consenso. Por ello se aplicará una prueba de fiabilidad entre codificadores o más conocida por su nombre en inglés *inter-rater reliability* con el estadístico Alfa de Krippendorff o *Krippendorff's Alpha* en inglés, el cual indica el grado de concordancia entre los codificadores.

El valor de alfa debe encontrarse entre 0.75 y 1 para concluir que los codificadores concuerdan en la mayoría de las etiquetas asignadas, mientras que si se obtiene un valor menor a 0.75 significa que las

decisiones emitidas por los codificadores no concuerdan y que la etiqueta resultado del consenso no es confiable.

CAPÍTULO 4

4. DESARROLLO Y ANÁLISIS DE RESULTADOS

En este capítulo se presentan las herramientas y el proceso empleado en la implementación de lo analizado y diseñado en el capítulo anterior. Se analizarán también los resultados de las pruebas con la finalidad de evaluar al sistema en lo funcional y su usabilidad.

4.1. IMPLEMENTACIÓN

Para implementar los componentes del motor de inferencia descrito en el capítulo 3 se emplearon las siguientes herramientas y lenguajes de programación:

- RapidMiner
- RapidMiner API
- Java

En cuanto a la implementación de la interfaz de usuario y de comunicación con el motor de inferencia se empleó lo siguiente:

- Ruby – Ruby On Rails
- MySQL
- Google Charts

A continuación se presenta la justificación de la selección de las herramientas mencionadas para la implementación de este proyecto.

RapidMiner

RapidMiner es un software para minería y análisis de datos desarrollado en el lenguaje de programación Java, permite la creación de procesos simples y complejos encadenando una gran variedad de operadores de extracción, procesamiento y visualización de datos a través de un entorno gráfico altamente intuitivo. Al estar programado en Java es posible correr los procesos creados en RapidMiner en una gran variedad de plataformas y sistemas operativos.

Otra de las bondades que ofrece RapidMiner es que todos los procesos pueden ser representados en formato XML¹⁰ lo cual facilita que dichos procesos sean usados por programas externos. Puede ser

¹⁰ XML por Extensible Markup Language, lenguaje de marcado que permite definir gramáticas de lenguajes específicos.

empleado a través de la interfaz gráfica, línea de comandos, procesos batch y también empleando sus bibliotecas.

En el presente proyecto RapidMiner se empleó para construir el motor de inferencia del sistema de clasificación. El componente de aprendizaje fue construido empleando la interfaz gráfica del software y una de las grandes ventajas de emplear esta herramienta es que todos los algoritmos que se necesitaron para la implementación (algoritmos para el pre procesamiento de texto, clasificador bayesiano) se encontraban ya programados y sus parámetros eran ajustables. El siguiente paso fue crear un proceso con los operadores que se definieron en el capítulo tres, proveer las respuestas etiquetadas por los expertos como entrada, correr dicho proceso, para finalmente obtener el modelo de clasificación.

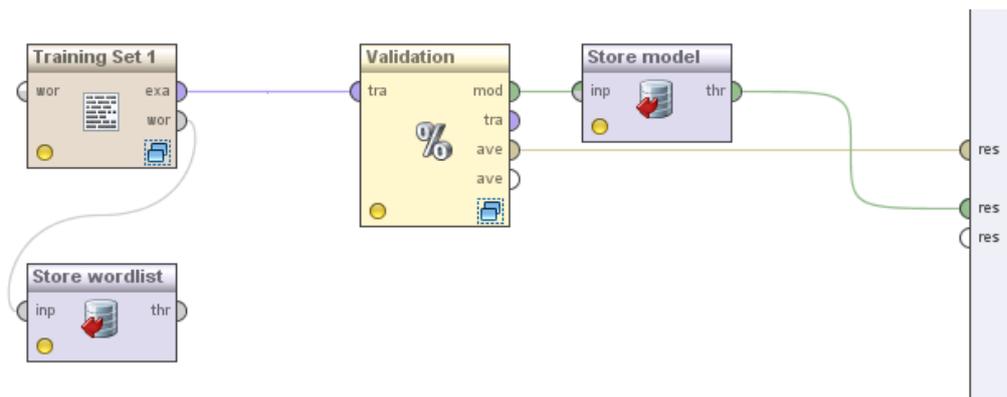


Figura 4.1 Implementación del componente de aprendizaje en RapidMiner

El componente de aprendizaje implementado con los operadores de RapidMiner luce como se muestra en la Figura 4.1.

El proceso que representa al sub componente de clasificación de texto y formación de vectores implementado en RapidMiner se muestra en la Figura 4.2.

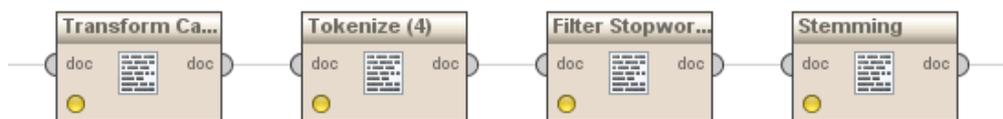


Figura 4.2 Implementación del subcomponente de pre procesamiento de texto en RapidMiner

Los resultados de ejecutar los procesos del componente de aprendizaje en RapidMiner se muestran en las siguientes figuras:

| R... | label | metadata_file | ... | acces | acerc | actual | acuerd | adapt | ademas | ahor | analiz | andro | anterior | apl |
|------|------------|------------------|-----------|-------|-------|--------|--------|-------|--------|------|--------|-------|----------|-------|
| 88 | analyzing | ani_f97_3_FM.bt | C O 0 | 0 | 0 | 0 | 0.334 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 89 | analyzing | ani_f9_1.bt | C S 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 90 | analyzing | un_f27_1.bt | C S 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 91 | applying | appl_f20_1.bt | C S 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 92 | applying | appl_f2_2.bt | C S 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 93 | applying | appl_f30_1.bt | C S 0 | 0 | 0 | 0 | 0.344 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 94 | applying | appl_f33_2.bt | C S 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 95 | applying | appl_f38_1.bt | C S 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 96 | applying | appl_f40_1.bt | C S 0 | 0 | 0 | 0 | 0 | 0 | 0.248 | 0 | 0.336 | 0 | 0 | 0 |
| 97 | applying | appl_f48_3.bt | C S 0 | 0 | 0 | 0 | 0 | 0 | 0.294 | 0 | 0 | 0 | 0 | 0 |
| 98 | applying | appl_f49_3.bt | C S 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 99 | applying | appl_f61_3.bt | C S 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.091 |
| 10 | applying | appl_f6_1.bt | C S 0 | 0 | 0 | 0 | 0.349 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 29 | evaluating | eva_f5_1.bt | C S 0 | 0 | 0 | 0 | 0.330 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 27 | evaluating | eval_f10_1_FM.bt | C O 0 | 0 | 0 | 0 | 0.129 | 0 | 0 | 0 | 0 | 0.238 | 0.132 | 0.103 |
| 27 | evaluating | eval_f17_5_Sl.bt | C S 0 | 0 | 0 | 0 | 0.320 | 0 | 0.285 | 0 | 0 | 0 | 0 | 0 |
| 27 | evaluating | eval_f19_2_Sl.bt | C S 0 | 0 | 0 | 0 | 0.235 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 27 | evaluating | eval_f20_3_FM.bt | C O 0 | 0 | 0 | 0 | 0 | 0 | 0.133 | 0 | 0 | 0.184 | 0 | 0.298 |
| 27 | evaluating | eval_f21_1_FW.bt | C O 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.134 |
| 27 | evaluating | eval_f24_2_Sl.bt | C S 0.279 | 0 | 0 | 0 | 0 | 0.288 | 0 | 0 | 0 | 0 | 0 | 0 |

Figura 4.3 Características y pesos extraídos de cada elemento del conjunto de entrenamiento

| accuracy: 32.83% +/- 8.69% (mikro: 32.83%) | | | | | | |
|--|----------------|---------------|------------------|-----------------|--------------------|-----------------|
| | true analyzing | true applying | true remembering | true evaluating | true understanding | class precision |
| pred. analyzing | 17 | 3 | 24 | 9 | 61 | 14.91% |
| pred. applying | 1 | 0 | 2 | 0 | 5 | 0.00% |
| pred. remembering | 12 | 2 | 62 | 2 | 56 | 46.27% |
| pred. evaluating | 12 | 2 | 32 | 0 | 66 | 0.00% |
| pred. understanding | 48 | 3 | 50 | 13 | 118 | 50.86% |
| class recall | 18.89% | 0.00% | 36.47% | 0.00% | 38.56% | |

Figura 4.4 Vista del desempeño de la clasificación por cada nivel de la taxonomía de Bloom

La Figura 4.3 muestra cuales fueron las características extraídas por cada elemento del conjunto de entrenamiento, así como también el peso asociado a cada una de ellas. Estas son las características que ayudan a predecir la etiqueta que tendrá un texto desconocido, siguiendo el procedimiento expuesto en el capítulo 3. La Figura 4.4 muestra un estimado de la exactitud y precisión del desempeño de la clasificación por cada nivel de la taxonomía.

El componente que realiza la predicción de la etiqueta también fue creado empleando la interfaz gráfica de RapidMiner, el proceso implementado se muestra en la Figura 4.5.

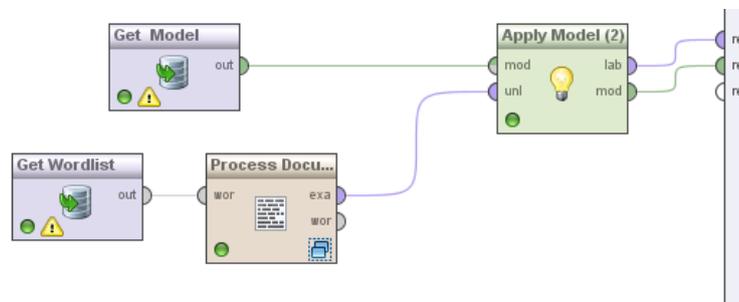


Figura 4.5 Implementación del componente de clasificación en RapidMiner

El resultado de ejecutar los procesos del componente de clasificación se muestra en la siguiente figura:

| ... | label | metadata_file | ... | confidence(... | confidence(... | confidence(... | confidence(... | confidence(... | prediction(label) |
|-----|--------|-------------------|-----|----------------|----------------|----------------|----------------|----------------|-------------------|
| 1 | unknow | anl_f12_1_FW.txt | C O | 1 | 0 | 0 | 0 | 0 | analyzing |
| 2 | unknow | anl_f16_3_FW.txt | C O | 1 | 0 | 0 | 0 | 0 | analyzing |
| 3 | unknow | anl_f23_3_FW.txt | C O | 0.000 | 0 | 0 | 0 | 1.000 | understanding |
| 4 | unknow | anl_f28_1_FW.txt | C O | 1 | 0 | 0 | 0 | 0 | analyzing |
| 5 | unknow | anl_f31_3_FW.txt | C O | 1.000 | 0 | 0 | 0 | 0.000 | analyzing |
| 6 | unknow | anl_f37_1_FW.txt | C O | 0 | 0 | 0 | 1 | 0 | evaluating |
| 7 | unknow | anl_f42_3_FM.txt | C O | 1 | 0 | 0 | 0 | 0 | analyzing |
| 8 | unknow | anl_f45_3_FM.txt | C O | 0.994 | 0 | 0 | 0 | 0.006 | analyzing |
| 9 | unknow | anl_f48_1_FM.txt | C O | 0.000 | 0 | 0 | 0 | 1.000 | understanding |
| 1 | unknow | anl_f49_3_FM.txt | C O | 0 | 0 | 0 | 1 | 0 | evaluating |
| 1 | unknow | anl_f58_3_FM.txt | C O | 0.475 | 0 | 0.001 | 0 | 0.524 | understanding |
| 1 | unknow | anl_f62_3_FM.txt | C O | 1.000 | 0 | 0 | 0 | 0.000 | analyzing |
| 1 | unknow | anl_f81_3_FM.txt | C O | 0 | 0 | 1.000 | 0 | 0.000 | remembering |
| 1 | unknow | anl_f95_3_FW.txt | C O | 0.000 | 0 | 1.000 | 0 | 0 | remembering |
| 1 | unknow | eval_f20_3_FM.txt | C O | 0 | 0 | 0 | 1 | 0 | evaluating |
| 1 | unknow | eval_f28_3_FW.txt | C O | 0 | 0 | 0 | 1 | 0 | evaluating |
| 1 | unknow | eval_f36_1_FW.txt | C O | 0 | 0 | 0 | 1 | 0 | evaluating |
| 1 | unknow | rem_f86_3_FW.txt | C O | 0.156 | 0 | 0 | 0 | 0.844 | understanding |
| 1 | unknow | rem_f01_1_FW.txt | C O | 0 | 0 | 1 | 0 | 0 | remembering |
| 2 | unknow | rem_f06_1_FM.txt | C O | 0 | 0 | 1 | 0 | 0 | remembering |
| 2 | unknow | rem_f07_1_FM.txt | C O | 0 | 0 | 1 | 0 | 0 | remembering |
| 2 | unknow | rem_f09_1_FW.txt | C O | 0 | 0 | 1 | 0 | 0 | remembering |
| 2 | unknow | rem_f108_3_FW.txt | C O | 0 | 0 | 1 | 0 | 0 | remembering |
| 2 | unknow | rem_f113_3_FW.txt | C O | 0 | 0 | 0 | 1 | 0 | evaluating |
| 2 | unknow | rem_f124_3_FW.txt | C O | 0 | 0 | 1 | 0 | 0 | remembering |
| 2 | unknow | rem_f127_3_FW.txt | C O | 0 | 0 | 0.999 | 0 | 0.001 | remembering |

Figura 4.6 Resultado de ejecutar los procesos del componente de clasificación en RapidMiner

La Figura 4.6 muestra los resultados luego de ejecutar los procesos del componente de clasificación en RapidMiner, los datos de entrada fueron los elementos del conjunto de pruebas. La columna *metadata_file* muestra el nombre del archivo a clasificar, las columnas *confidence* indican la confianza con la cual se realizó la clasificación y

finalmente la columna *prediction (label)* indica la etiqueta que asignó el sistema al texto contenido en el archivo a clasificar.

RadidMiner API

La interfaz gráfica de RapidMiner permite crear procesos de una manera rápida y sencilla, sin embargo si se desea emplear dichos procesos desde un programa o sistema exterior esto no es posible. Es por ello que RapidMiner dispone de un conjunto de bibliotecas que pueden ser empleadas para desarrollar los procesos desde un programa codificado en java.

Dado que el componente de clasificación del motor de inferencia debe ser empleado e invocado por SIDWeb 4, fue necesario implementar este componente en código java. RapidMiner ofrece la facilidad de exportar los procesos creados en la interfaz gráfica en XML y en este caso esto fue aprovechado, se construyó el componente de clasificación desde la interfaz de RapidMiner por las facilidades que esto conlleva y este proceso fue almacenado como un archivo XML.

El componente en XML es procesado por un programa escrito en java y se realizan las adecuaciones necesarias para que pueda ser empleado por SIDWeb 4. La funcionalidad de este programa es la de recibir las respuestas a etiquetar desde SIDWeb 4, procesarlas

empleando el componente de clasificación y devolver las categorías de la taxonomía de Bloom por cada respuesta recibida.

Java

Java es un lenguaje de programación orientado a objetos y multiplataforma. Se empleó java en la implementación del sistema debido a que RapidMiner está implementado en este lenguaje y las bibliotecas que provee son para esta plataforma. Usando este lenguaje se realizaron las adecuaciones necesarias al componente de clasificación generado en la interfaz de RapidMiner para que pueda ser usado desde SIDWeb 4.

Otra de las ventajas de usar java es que SIDWeb 4 corre sobre un servidor Linux y dada que java es multiplataforma, puede ser usado sin ningún problema sobre este sistema operativo.

Ruby – Ruby on Rails

Ruby es un lenguaje de programación dinámico y de código abierto enfocado en la simplicidad y productividad, es orientado a objetos e incorpora tanto la programación imperativa como la programación funcional [17].

Ruby On Rails es una plataforma de desarrollo de código abierto que se encuentra escrito en Ruby; las aplicaciones que se desarrollan en Rails siguen el paradigma de la arquitectura Modelo Vista Controlador – MVC¹¹. El objetivo principal de Ruby on Rails es que las aplicaciones sean desarrolladas escribiendo una menor cantidad de código en comparación con otras plataformas, que además requieran un mínimo de configuración y sean escalables [18].

El sistema de clasificación automática de respuestas a los foros de discusión tiene por finalidad ser una funcionalidad más de la plataforma SIDWeb 4 que brinde soporte a los procesos de enseñanza y aprendizaje de los cursos de ESPOL. SIDWeb 4 se encuentra implementado en Ruby on Rails y por ello la interfaz de usuario del sistema está implementado siguiendo la lógica de las aplicaciones desarrolladas bajo esta plataforma.

Rails además cuenta con una interfaz de consultas llamada *Active Records* y es esta la interfaz empleada para adquirir los datos necesarios de la base de datos para la creación de los gráficos y estadísticas que se muestran al usuario.

¹¹ MVC por Modelo Vista Controlador, es una abstracción de desarrollo de software que separa la interfaz gráfica de los datos y la lógica de negocios.

MySQL

MySQL es un gestor de base de datos relacionales de código abierto y tiene soporte multiplataforma. En este RDBMS¹² se encuentra instanciada la base de datos de SIDWeb 4 y como se expuso en el capítulo 3, en la tabla *discussion_entries* se alojarán las etiquetas de las respuestas en el campo *tag*.

Google Charts

Google Charts es una biblioteca JavaScript proveída por Google de manera gratuita, que permite la creación de gráficos, tablas y permite embeberlos en una página web. Google Charts permite la creación de una gran cantidad de gráficos estadísticos interactivos los cuales son personalizables, además es soportado por los navegadores web mayormente usados tanto para escritorio como para dispositivos móviles [19].

Google Charts es empleado en forma de código JavaScript, los gráficos y tablas son creados con mucha facilidad, además el código se escribe en las vistas lo cual implica que no se modifica código en los controladores y por ende no se afectan las funcionalidades de la aplicación.

¹² RDBMS por Relational Database Management System, gestor de base de datos relacionales en inglés.

Para crear los gráficos y llenar las tablas se consulta a la base de datos a través de los *Active Records* de Rails, las consultas son programadas de acuerdo a lo que se requiera y los datos devueltos son colocados en el formato que entiende Google Charts y con ello finalmente los gráficos son obtenidos.

A continuación se presentan imágenes que muestran la implementación de la interfaz de usuario del sistema de clasificación automática en SIDWeb 4 según las especificaciones expuestas en el capítulo 3, sección Diseño del Módulo de Clasificación en SIDWeb 4.

La Figura 4.7 muestra los datos generales de un foro de discusión en SIDWeb 4, estos datos incluyen el total de participaciones en el foro, su fecha de creación, cuando ocurrió la primera y última contribución, así como también los autores de las mismas.

Estadísticas del Foro Discusión sobre Frameworks Web

| Contribuciones | | Niveles de Bloom | Niveles de Bloom por Estudiante |
|---------------------------|--|------------------|---------------------------------|
| Datos generales del foro: | | | |
| Indicador | Valor | | |
| 1 Total respuestas | 198 | | |
| 2 Fecha de creación | Jun 18, 2012 en 11:55am | | |
| 3 Primer contribución | Jun 18, 2012 en 4:40pm por Manuel Alejandro Torres Campozano | | |
| 4 Última contribución | Jul 5, 2012 en 7:59pm por Jhonny Vladimir Pincay Nieves | | |

Figura 4.7 Estadísticas generales de un foro de discusión en SIDWeb 4

La Figura 4.8 muestra en resumen el número de participaciones por estudiantes dentro de un foro en particular, también se presenta un diagrama de pastel en el cual se puede apreciar gráficamente el porcentaje de colaboración de cada estudiante dentro del foro de discusión.

Participaciones por estudiantes

| | Usuario | Participaciones |
|---|------------------------------------|-----------------|
| 1 | Andres Alberto Gomez Bano | 24 |
| 2 | Jennifer Carolina Capelo Solorzano | 30 |
| 3 | Jhonny Vladimir Pincay Nieves | 49 |
| 4 | Johnny Omar Ramirez Malave | 28 |
| 5 | Juan Jose Plua Cuesta | 12 |
| 6 | Manuel Alejandro Torres Camposano | 22 |
| 7 | Maria Beatriz Gomez Bano | 31 |
| 8 | SIDWeb Admin | 1 |
| 9 | Xavier Antonio Ochoa Chehab | 1 |



Figura 4.8 Tabla resumen y diagrama de pastel de las participaciones por estudiantes en un foro de discusión de SIDWeb 4

La Figura 4.9 muestra las estadísticas del total de participaciones de un foro que han sido clasificadas bajo determinado nivel de la taxonomía de Bloom. Se presenta también un diagrama de pastel en el cual se aprecia porcentualmente la misma información de la tabla.

Estadísticas del Foro Discusión sobre Frameworks Web

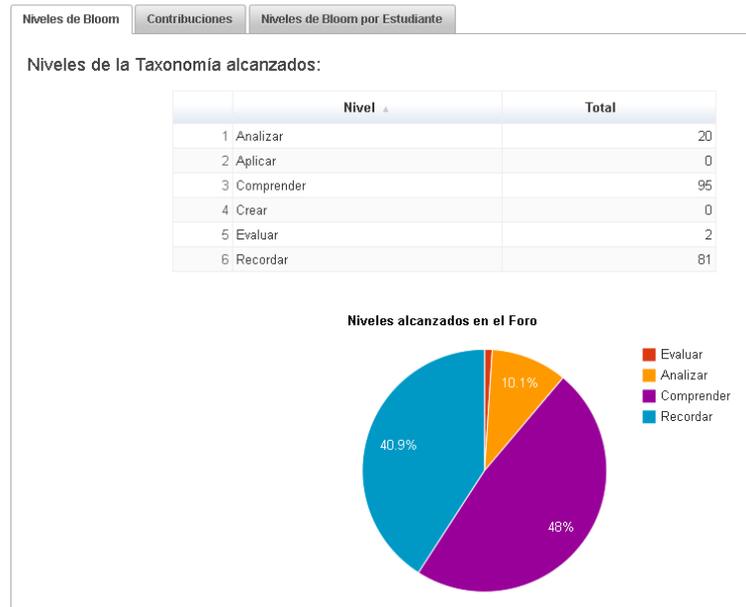


Figura 4.9 Totales por nivel de la taxonomía de Bloom alcanzados en un foro de discusión en SIDWeb 4

Niveles de la Taxonomía alcanzados por estudiante:

| Usuario ▲ | Participaciones |
|--------------------------------------|-----------------|
| 1 Andres Alberto Gomez Bano | 7 |
| 2 Jennifer Carolina Capelo Solorzano | 25 |
| 3 Jhonny Vladimir Pincay Nieves | 28 |
| 4 Johnny Omar Ramirez Malave | 22 |
| 5 Manuel Alejandro Torres Campozano | 20 |
| 6 Maria Beatriz Gomez Bano | 18 |

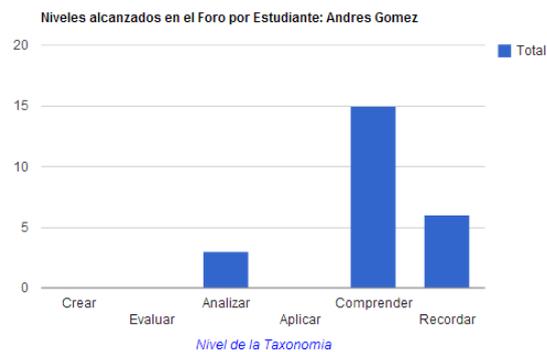


Figura 4.10 Número de colaboraciones por nivel de un determinado estudiante en un foro de discusión en SIDWeb 4

En la Figura 4.10 se muestra una tabla con el total de participaciones de un estudiante en el foro de discusión, así como también el número de colaboraciones por nivel de la taxonomía de Bloom del estudiante presentado en un diagrama de barras.

4.2. PRUEBAS

4.2.1. PRUEBAS DE FUNCIONALIDAD

Los resultados de las pruebas de funcionalidad permitieron determinar la efectividad con la que el sistema asigna una etiqueta correctamente a las respuestas emitidas por los estudiantes en los foros de discusión.

Las métricas empleadas para determinar si el sistema funciona o no fueron exactitud, recuperación, precisión y F_{β} , fueron calculadas empleando los valores del conteo de los verdaderos positivos representados por A, falsos positivos representados por B, falsos negativos representados por C y los verdaderos negativos representados por D. Las definiciones de las métricas y de los valores de A, B, C y D se encuentran expuestas en detalle en el capítulo 3, en la sección de Diseño de Pruebas.

El conteo de los valores de A, B, C y D se muestra en la Tabla 4.1. Los resultados son presentados para cada nivel de la taxonomía a

excepción del nivel de crear, del cual no se obtuvieron elementos que cayeran en esta categoría.

| Nivel de la taxonomía de Bloom | A | B | C | D |
|--------------------------------|----|----|----|-----|
| Evaluar | 7 | 37 | 0 | 135 |
| Analizar | 16 | 13 | 11 | 140 |
| Aplicar | 3 | 4 | 0 | 173 |
| Entender | 50 | 7 | 42 | 81 |
| Recordar | 28 | 15 | 23 | 114 |

Tabla 4.1 Resultados del conteo de los valores de A, B, C y D luego de aplicar el modelo de clasificación al conjunto de pruebas

| Nivel de la taxonomía de Bloom | Exactitud | Recuperación | Precisión | F_{β} |
|--------------------------------|-----------|--------------|-----------|-------------|
| Evaluar | 0.16 | 1.00 | 0.04 | 0.27 |
| Analizar | 0.55 | 0.59 | 0.15 | 0.57 |
| Aplicar | 0.43 | 1.00 | 0.02 | 0.60 |
| Entender | 0.88 | 0.54 | 0.51 | 0.67 |
| Recordar | 0.65 | 0.55 | 0.28 | 0.60 |

Tabla 4.2 Valores resultantes de las métricas por nivel de la taxonomía

Los valores de las métricas resultantes de la aplicación del modelo de clasificación a cada uno de los elementos del conjunto de pruebas se muestran en la Tabla 4.2. Estos valores son calculados aplicando las ecuaciones expuestas en el capítulo 3 y empleando los valores de A, B, C y D obtenidos previamente.

Con respecto a la prueba de fiabilidad entre codificadores, esta fue ejecutada empleando solo los elementos de las respuestas recolectadas en el 2012, puesto que para ejecutar esta prueba es necesario conocer la etiqueta asignada por cada codificador; el objetivo de esta prueba fue determinar si el sistema de inferencia puede actuar como un codificador humano adicional.

Es necesario indicar que hasta los codificadores más experimentados tienden a cometer errores, es así como varios test fueron ejecutados y los resultados obtenidos por cada uno se muestran en la Tabla 4.3.

| Prueba | Valor de Alfa de Krippendorff |
|--|-------------------------------|
| Fiabilidad de los tres codificadores humanos | 0.7189 |
| Fiabilidad de codificadores humanos y sistema automático | 0.53 |
| Fiabilidad sin humano uno | 0.579 |
| Fiabilidad sin humano dos | 0.418 |
| Fiabilidad sin humano tres | 0.418 |

Tabla 4.3 Resultados de las pruebas de fiabilidad entre codificadores

4.2.2. PRUEBAS DE USABILIDAD

Las pruebas de usabilidad fueron realizadas a seis profesores de la Facultad de Ingeniería en Electricidad y Computación de la ESPOL quienes debían calificar que tan complicada es la tarea de revisar las

estadísticas de los foros dentro de SIDWeb 4. Se empleó una escala del uno al cinco, donde uno corresponde a muy difícil y cinco corresponde a muy fácil.

Finalmente se calculó el promedio de las calificaciones brindadas por cada profesor arrojando como resultado un valor de 4.25, con lo cual es posible aseverar que la revisión de las estadísticas dentro de SIDWeb 4 es una tarea fácil, no complicada de realizar.

4.3. ANÁLISIS DE RESULTADOS

Los resultados de las cuatro métricas de exactitud, recuperación, precisión y F_{β} presentan diferencias significativas entre ellas; con respecto a exactitud el valor más alto es de 0.88 para el nivel de entender, mientras que el valor más bajo es de 0.16 para evaluar, esto significa que el 88% de las veces la clasificación para este nivel es hecha completa correctamente mientras que en el nivel de evaluar la predicción es errónea la mayoría de las veces. Estas diferencias entre los valores de las métricas muy seguramente son causadas porque el número de elementos del conjunto de entrenamiento no se encuentran uniformemente distribuidos a través de los niveles de la taxonomía, por ello los valores más altos de exactitud son para los niveles de entender, recordar y analizar, que son los niveles con mayor número de ejemplos.

Los resultados para precisión evidenciaron el mismo patrón que los valores de exactitud; el mejor valor para entender y el peor para evaluar, en general los resultados de precisión son más bajos que los de precisión y no son satisfactorios.

Los valores de recuperación están alrededor del 50%, esto significa que si un texto aleatorio externo al conjunto de pruebas es clasificado correctamente solo la mitad de las veces. Dado que los resultados de recuperación y exactitud son bajos, provocan que los valores de F también lo sean, lo cual en general indica que la efectividad del clasificador es del 67% para el nivel de entender y de 27% para evaluar, una vez más los resultados presentan esta forma debido a la estructura y distribución del conjunto de entrenamiento.

Comparando el mejor resultado obtenido que fue el de exactitud para el nivel de entender con los valores obtenidos por sistemas similares, debe recalcarse que 88% es el valor más alto entre todas las soluciones revisadas, considerando que el resultado obtenido por el sistema de [14] fue de 26% para el nivel equivalente de comprensión y 50% para el clasificador SVM de [12]. Si consideramos la exactitud alcanzada para el nivel de analizar, este es menor que el valor obtenido por el clasificador SVM el cual fue de 75% y más alto que 32% que fue el valor alcanzado por la propuesta de [14].

En general la efectividad lograda por el clasificador Naïve Bayes es inferior a los resultados obtenidos por el clasificador SVM, pero muestra un mejor desempeño que una red neuronal usada como clasificador [7] y que un sistema experto [2] en los niveles que fueron entrenados con un alto número de ejemplos. Aunque algunos resultados puede calificárselos de bueno, el sistema necesita mejoras y emplear un mejor conjunto de datos para el entrenamiento.

Con respecto a las pruebas de fiabilidad entre codificadores, el valor del estadístico alfa de Krippendorff para las etiquetas dadas por los tres codificadores humanos fue de 0.718, lo cual significa que existió un alto grado de acuerdo entre sus respuestas; añadiendo como un cuarto codificador al sistema automático de clasificación, el resultado fue de 0.53. Este valor sugiere que las conclusiones alcanzadas por los codificadores difieren bastante y que el codificador que está introduciendo ruido es el sistema automático, los tres últimos valores presentados en la Tabla 4.3 confirman esto puesto que excluyendo a un codificador humano diferente en cada prueba los valores permanecen bajos.

CONCLUSIONES

- 1) Los resultados obtenidos de las pruebas permitieron concluir que empleando la arquitectura propuesta en este trabajo para realizar la tarea de clasificar automáticamente texto de acuerdo a la taxonomía de Bloom es viable.

- 2) La efectividad y precisión de la clasificación dependen mucho de la calidad del conjunto de entrenamiento usado para generar el modelo de clasificación. Sin embargo este es un factor que aún requiere de mucha investigación puesto que a pesar de que existen ciertos lineamientos para crear datos de entrenamiento, para este tipo de problemas en particular aún no existen criterios definidos que permitan calificar a un conjunto de entrenamiento como bueno o malo.

- 3) Los valores resultantes para las distintas métricas sugieren que el

clasificador Naïve Bayes funciona relativamente bien cuando hay un número considerable de ejemplos para entrenarlo. De manera general, al menor número de ejemplos disponible menor será la precisión de la predicción.

- 4) A pesar de los bajos resultados obtenidos cabe destacar que el sistema proporciona una precisión bastante aceptable y que es comparable con los resultados obtenidos por sistemas similares en lo que respecta a los niveles de recordar, entender y analizar, por lo que los resultados devueltos por el sistema podrían ayudar a un profesor a tener una visión rápida y aproximada del nivel alcanzado por sus alumnos y si se está logrando o no un objetivo de aprendizaje.

- 5) Los resultados del estadístico alfa de Krippendorff indican que lamentablemente el sistema aún no está en la capacidad de sustituir a un codificador humano, sin embargo es importante recalcar que el proceso de clasificación realizado por personas es un trabajo que requiere de consenso y que hasta los codificadores más experimentados tienden a cometer errores. Considerando esto es posible afirmar que realizando las mejoras pertinentes, empleando esta misma arquitectura es posible alcanzar el objetivo de que el sistema pueda desempeñarse como un codificador humano más.

RECOMENDACIONES

- 1) Para futuras implementaciones o mejoras al presente proyecto es necesario disponer de un mejor conjunto de entrenamiento, de preferencia uno cuyos elementos se encuentren uniformemente distribuidos entre los niveles de la taxonomía de Bloom. De esta forma se mejorará sustancialmente la calidad del modelo de clasificación y con ello mejorará la efectividad del sistema.
- 2) En la actualidad no existe un clasificador óptimo para esta tipo de problemas, sin embargo es necesario que para implementaciones futuras se lleven a cabo pruebas con diferentes tipos de clasificadores o se realicen mejoras en los algoritmos a fin de escoger el que ofrezca una mayor precisión.
- 3) La etapa de reducción y transformación de características es

extremadamente importante para mejorar la efectividad del sistema, por ello se recomienda analizar cuidadosamente cual método es el más adecuado teniendo siempre en cuenta la forma de los datos del conjunto de entrenamiento.

- 4) Con respecto a las herramientas empleadas en la implementación del presente trabajo, RapidMiner demostró ser un software muy práctico y sencillo de usar al momento de construir procesos, sin embargo dado que los algoritmos vienen todos ya implementados no es fácil realizar cambios en los mismos en el caso que se desee adaptar un algoritmo a una situación especial. Con respecto a Google Charts, es una librería extremadamente sencilla de emplear y permite obtener una gran variedad de gráficos en poco tiempo. En general son herramientas muy versátiles que se recomienda sean consideradas al momento de desarrollar alguna solución.

BIBLIOGRAFÍA

[1] Marzano, R., Kendall, J., The New Taxonomy of Educational Objectives, Corwin 2nd edition, 2007

[2] Chang, Y., Chen, H., An Automatic Inference System for the Quality Analysis of Test Items Based on The Bloom's Revised Taxonomy, Proceedings of the Eight International Conference on Machine Learning and Cybernetics, <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=5212592&contentType=Conference+Publications>, 2009

[3] Khairuddin, N., Hashim, K., Application of Bloom's Taxonomy in Software Engineering Assessments, Proceedings of the 8th International Conference on Applied Computer Science, <http://www.wseas.us/e-library/conferences/2008/venice/acs/acs09.pdf>, 2006

[4] Forehand, M., Bloom's Taxonomy from Emerging Perspectives on Learning, Teaching and Technology, [http://textbookequity.com/oct/Textbooks/Orey Emerging Perspectives Learning.pdf](http://textbookequity.com/oct/Textbooks/Orey%20Emerging%20Perspectives%20Learning.pdf), 2010

[5] Churches, A., Bloom's Taxonomy Blooms Digitally, techLEARNING, <http://www.techlearning.com/studies-in-ed-tech/0020/blooms-taxonomy-blooms-digitally/44988> , 2008

[6] Knowledge Jump, Learning Domain or Bloom's Taxonomy, <http://www.nwlink.com/~donclark/hrd/bloom.html>, 2010

[7] Yusof, N., Hui, Ch., Determination of Bloom's Cognitive Level of Question Items using Artificial Neural Network, IEEE, <http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5687152> , 2010

[8] Hui, CH., Feature Reduction for neural Network in Determining the Bloom's Cognitive Level of Question Items, <http://eprints.utm.my/11449/3/ChaiJingHuiMFSKSM2009ABS.pdf>, 2009

[9] Sebastiani, F., Machine Learning in Automated Text Categorization, <http://nmis.isti.cnr.it/sebastiani/Publications/ACMCS02.pdf>, 2002

- [10] Ikonomakis, M., Kotsiantis, S., Tampakas, V., Text Classification using Machine Learning Techniques, WSEAS Transactions on Computers, http://www.infoautoclassification.org/public/articles/Ikonomakis-et.-al._Text-Classification-Using-Machine-Learning-Techniques.pdf, 2005
- [11] Cutrone, I., Chang, M., Automarking: Automatic Assessment of Open Questions, IEEE International Conference on Advanced Learning technologies, <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5571204&contentType=Conference+Publications&queryText%3Dautomatic+bloom%27s+taxonomy>, 2010
- [12] Yahya, A., Osman, A., Automatic Classification of Questions into Bloom's Cognitive Levels using Support Vector Machines, <http://www.nauss.edu.sa/acit/PDFs/f2914.pdf> , fecha de consulta agosto 2012
- [13] Liao, C., Alpha, S., Dixon, P., Feature Preparation in Text Categorization, Oracle Corporation, <http://www.oracle.com/technetwork/database/enterprise-edition/feature-preparation-130942.pdf> , fecha de consulta septiembre 2012
- [14] Chang, W., Chung, M., Automatic Applying Bloom's taxonomy to Classify and Analysis the Cognition Level of English Question Items, IEEE, <http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5420087>, 2009
- [15] Losada, I., Pareja, C., Velázquez, J., Testing-Based Automatic Grading: A Proposal from Bloom's Taxonomy, Eighth IEEE International Conference on Advanced Learning Technologies, <http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=4561849>, 2008
- [16] Zu, G., Ohyama, W., Wakabayashi, T., Kimura, F., Accuracy improvement of automatic text classification based on feature transformation, <http://miuse.mie-u.ac.jp/bitstream/10076/11098/3/40A12191.pdf>, fecha de consulta octubre de 2012
- [17] Ruby Visual Identity Team, Ruby es..., <http://www.ruby-lang.org/es/>, fecha de consulta noviembre de 2012
- [18] Heinemeier, Web development that doesn't hurt, <http://rubyonrails.org/>, fecha de consulta noviembre de 2012
- [19] Google Inc., Google Chart Tools, <https://developers.google.com/chart/>, fecha de consulta noviembre de 2012