



UNIVERSIDAD DEL AZUAY
FACULTAD DE CIENCIAS DE LA ADMINISTRACIÓN.
ESCUELA DE INGENIERÍA DE SISTEMAS Y TELEMÁTICA

ANÁLISIS DE LENGUA MEDIANTE MINERÍA DE TEXTO.

TRABAJO DE GRADUACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO EN SISTEMAS Y TELEMÁTICA.

AUTOR:

JUAN FERNANDO LIMA SIGUA.

DIRECTOR:

MGST. ASTUDILLO RODRÍGUEZ CATALINA

CUENCA, ECUADOR

2017

Dedicatoria

Este trabajo va dedicado especialmente para mis padres: Martha y Rubén, quienes han sabido guiarme en los momentos más emotivos de mi vida.

A mis padrinos María y Manuel quienes me han brindado su apoyo de manera incondicional en todo momento.

Finalmente, a toda mi familia, seres queridos, amigos y tutores quienes han sido parte de mi vida.

Agradecimientos

El camino de la educación no tiene fin pues nadie es dueño del conocimiento absoluto, empieza desde que uno tiene conciencia y termina cuando uno deja este mundo, pero las personas que uno encuentra al recorrerlo siempre brindan con total amabilidad lo que ellos han aprendido en su camino, por eso quiero agradecer a todos los profesores de la carrera de Ingeniería de Sistemas y Telemática de la Universidad del Azuay ya que sin su guía no hubiera avanzado tanto en mis formación profesional y personal. Finalmente, quiero agradecer a mi madre Marthita que siempre creyó en mí, a mi familia y amigos que siempre han sido un gran apoyo en cualquier momento de declive, muchas gracias.

Índice de contenidos

Dedicatoria	ii
Agradecimientos	iii
Índice de contenidos.....	iv
Índice de tablas.....	vi
Índice de ilustraciones	viii
Resumen	x
Abstract.....	xi
Introducción	1
Capítulo 1: Indagación exploratoria.....	3
1.1 Sistema gestor de base de datos.....	4
1.2 Diseño de la base de datos	6
1.3 Mapas digitales	7
1.4 Google Charts API	9
1.5 Oracle Application Express	10
1.6 Minería de texto aplicada a la detección de idioma y fonemas	11
1.6.1 Clasificación de los algoritmos de minería de datos.....	11
1.6.2 Algoritmos en Oracle Database para minería de texto.....	12
1.6.3 Algoritmo GNU Aspell aplicado a detección de idioma	13
1.7 Conclusiones del capítulo	13
Capítulo 2: Especificación de requerimientos del software	15
2.1 Aplicando el estándar IEEE-830	15
2.1.1 Introducción	15
2.1.2 Descripción global	17
2.1.3 Los requisitos específicos	19

2.2	Diagrama de casos de uso del software	24
2.3	Metodología para el desarrollo del software – Scrum	25
2.3.1	Asignación de roles	25
2.3.2	Listado de requerimientos del usuario.....	25
2.3.3	Historias de usuario	26
2.3.4	Sprint y listado del producto	33
2.3.5	Planificación del Scrum	34
2.4	Conclusiones del capítulo	37
Capítulo 3:	Diseño de la aplicación web	38
3.1	Diseño del servicio web para clasificación de texto	38
3.1.1	Fases de limpieza.....	39
3.1.2	Aplicación de minería de texto	42
3.2	Diseño y normalización de la base de datos	44
3.3	Diseño del modelo Cliente-servidor	46
3.4	Conclusiones del capítulo	47
Capítulo 4:	Software Cartografía etnolingüística y análisis de lengua.	48
4.1	Configuración del espacio de trabajo.....	48
4.1.1	Configuración para modelo de tres capas con Glassfish	48
4.1.2	Configuración de Oracle APEX.....	52
4.2	Aplicación web - Cartografía etnolingüística	55
4.2.1	Convenciones en la programación del software	56
4.2.2	Elementos generales para interfaz de usuario.....	57
4.2.3	Aplicación para usuario Administrador de contenido	60
4.2.4	Aplicación para usuario Visitante	81
4.3	Conclusiones del capítulo	85
Capítulo 5:	Pruebas de integridad y presentación de los datos	87
5.1	Pruebas semánticas	87
5.2	Pruebas sintácticas	89

5.3	Pruebas de interfaz.....	90
5.4	Pruebas de usabilidad	91
5.5	Cambios para pruebas fallidas	92
5.6	Conclusiones del capítulo	93
	Conclusiones del trabajo.....	94
	Referencias.....	96
	Anexos.....	99
	Anexo 1: Guía de instalación de Oracle Database 11gR2.....	99
	Anexo 2: Guía de actualización de Oracle Apex 5.0.4.....	103
	Anexo 3: Guía de Instalación de Java JDK 1.7.0	105
	Anexo 4: Exportado de análisis de comunidad	106

Índice de tablas

Tabla 1.1:	Comparación entre SGBD populares.....	5
Tabla 1.2:	Comparación entre mapas digitales	8
Tabla 1.3:	Algoritmos de minería de texto en Oracle Database	12
Tabla 2.1:	Conceptos, acrónimos y abreviaturas.....	17
Tabla 2.2:	Scrum - Asignación de roles.....	25
Tabla 2.3:	Scrum - Historia de usuario 1	26
Tabla 2.4:	Scrum - Historia de usuario 2	26
Tabla 2.5:	Scrum - Historia de usuario 3	26
Tabla 2.6:	Scrum - Historia de usuario 4	27
Tabla 2.7:	Scrum - Historia de usuario 5	27
Tabla 2.8:	Scrum - Historia de usuario 6	27
Tabla 2.9:	Scrum - Historia de usuario 7	27
Tabla 2.10:	Scrum - Historia de usuario 8	28
Tabla 2.11:	Scrum - Historia de usuario 9	28
Tabla 2.12:	Scrum - Historia de usuario 10	28
Tabla 2.13:	Scrum - Historia de usuario 11	28

Tabla 2.14: Scrum - Historia de usuario 12	29
Tabla 2.15: Scrum - Historia de usuario 13	29
Tabla 2.16: Scrum - Historia de usuario 14	29
Tabla 2.17: Scrum - Historia de usuario 15	29
Tabla 2.18: Scrum - Historia de usuario 16	30
Tabla 2.19: Scrum - Historia de usuario 17	30
Tabla 2.20: Scrum - Historia de usuario 18	30
Tabla 2.21: Scrum - Historia de usuario 19	30
Tabla 2.22: Scrum - Historia de usuario 20	31
Tabla 2.23: Scrum - Historia de usuario 21	31
Tabla 2.24: Scrum - Historia de usuario 22	31
Tabla 2.25: Scrum - Historia de usuario 23	31
Tabla 2.26: Scrum - Historia de usuario 24	32
Tabla 2.27: Scrum - Historia de usuario 25	32
Tabla 2.28: Scrum - Historia de usuario 26	32
Tabla 2.29: Scrum - Historia de usuario 27	32
Tabla 2.30: Scrum - Historia de usuario 28	33
Tabla 2.31: Sprint y listado del producto	33
Tabla 2.32: Listado de responsables	34
Tabla 2.33: Listado del producto y responsables.....	35
Tabla 2.34: Planificación de reuniones de revisión.....	36
Tabla 4.1: Convenciones para variables en PLSQL	56
Tabla 4.2: Convenciones para componentes en Oracle APEX	56
Tabla 4.3: Convenciones para colecciones en Oracle APEX	57
Tabla 4.4: Convenciones para páginas web en Oracle APEX.....	57
Tabla 4.5: Elementos de interfaz de usuario	59
Tabla 4.6: Reporte de comunidades.....	62
Tabla 4.7: Mantenimiento de comunidad.....	63
Tabla 4.8: Reporte de eventos.....	64
Tabla 4.9: Mantenimiento de evento	66
Tabla 4.10: Reporte de modelos de estudio	66
Tabla 4.11: Mantenimiento de modelo de estudio.....	68
Tabla 4.12: Vista en árbol puntos de análisis	69
Tabla 4.13: Mantenimiento de punto de análisis.....	70

Tabla 4.14: Reporte de análisis de comunidades.....	72
Tabla 4.15: Mantenimiento análisis comunidad - Vista Jerárquica	73
Tabla 4.16: Mantenimiento análisis comunidad - Editor de texto	74
Tabla 4.17: Mantenimiento análisis comunidad - Clasificador de textos.....	75
Tabla 4.18: Reporte multimedia comunidad.....	76
Tabla 4.19: Mantenimiento multimedia comunidad.....	77
Tabla 4.20: Reporte comunidad contraste.....	78
Tabla 4.21: Mantenimiento comunidad contraste.....	80
Tabla 4.22: Reporte exportar análisis comunidad	80
Tabla 5.1: Prueba semántica – Administración.....	88
Tabla 5.2: Prueba semántica – Visitante	88
Tabla 5.3: Prueba sintáctica – Administración	89
Tabla 5.4: Prueba sintáctica – Visitante.....	89
Tabla 5.5: Prueba de mecanismos de interfaz - Administración	90
Tabla 5.6: Prueba de mecanismos de interfaz – Visitante.....	90
Tabla 5.7: Prueba de usabilidad - Administración	91
Tabla 5.8: Prueba de usabilidad - Visitante	91
Tabla 5.9: Cambios en pruebas fallidas	92

Índice de ilustraciones

Ilustración 1.1: Ejemplo Google Chart API	9
Ilustración 1.2: Oracle APEX Application Builder.....	10
Ilustración 2.1: Dependencia entre software de desarrollo	19
Ilustración 2.2: Diseño de inicio de sesión	20
Ilustración 2.3: Diseño de mapa digital en la aplicación	20
Ilustración 2.4: Diseño para reportes de datos.....	20
Ilustración 2.5: Diseño para mantenimientos.....	21
Ilustración 2.6: Diagrama de casos de uso.....	24
Ilustración 3.1: Servicio web - Diagrama de clases	39
Ilustración 3.2: Identificación del dato mediante un conjunto de datos	40
Ilustración 3.3: Transformación de texto.....	41
Ilustración 3.4: Clasificación lógica de texto.....	42

Ilustración 3.5: Minería de texto en la aplicación web	43
Ilustración 3.6: Cuantificación de textos	43
Ilustración 3.7: Modelo Entidad-Relación	44
Ilustración 3.8: Tuning de base de datos - Tablespaces	45
Ilustración 3.9: Tuning de base de datos- Asignación de memoria.....	45
Ilustración 3.10: Modelo de tres capas.....	47
Ilustración 4.1: Servidor de aplicaciones Glassfish	50
Ilustración 4.2: ORDS sobre servidor de aplicaciones Glassfish.....	51
Ilustración 4.3: Administrador de servicios de Oracle APEX	53
Ilustración 4.4: Crear entorno de trabajo – Identificadores del entorno.....	53
Ilustración 4.5: Crear entorno de trabajo - Identificadores de esquema	54
Ilustración 4.6: Crear entorno de trabajo - Identificar el administrador	54
Ilustración 4.7: Referencias de servicios web de Oracle APEX	55
Ilustración 4.8: Crear referencia WSDL.....	55
Ilustración 4.9: Cartografía etnolingüística - Inicio de Sesión	60
Ilustración 4.10: Cartografía etnolingüística – Página de Inicio.....	61
Ilustración 4.11: Aplicación de presentación – Comunidades.....	81
Ilustración 4.12: Aplicación de presentación - Introducción comunidad	82
Ilustración 4.13: Aplicación de presentación - Análisis etnolingüístico.....	82
Ilustración 4.14: Aplicación de presentación - Cuantificador de idiomas	83
Ilustración 4.15: Aplicación de presentación – Fotografías comunidad	83
Ilustración 4.16: Aplicación de presentación – Grabaciones comunidad.....	84
Ilustración 4.17: Aplicación de presentación – Videos comunidad	84
Ilustración 4.18: Aplicación de presentación - Conclusiones comunidad	85

Resumen

En la actualidad publicar información importante de cualquier ámbito al público es un suceso común mediante páginas web. Una aplicación que permita gestionar la información lingüística de comunidades, analizar el contenido idiomático de los textos proporcionados, visualizar contenido multimedia y los publicar en la web, es factible desarrollarlo. La solución presentada en este documento es crear una aplicación web mediante técnicas de minería de texto clasificar, cuantificar y presentar la información de los textos ingresados. Todo esto para brindar apoyo a personas dedicadas al estudio del comportamiento del idioma en los grupos sociolingüísticos.

Palabras clave: minería de texto, aplicación web sociolingüística, clasificación de textos.

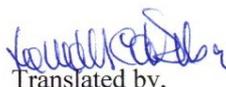
Abstract

ABSTRACT

At present, publishing important information from any field to public access through web pages is a common event. Therefore, it is feasible to develop an application that enables to manage the linguistic information of communities, analyze the idiomatic content of the provided texts, visualize multimedia content and publish it in the web. The procedure presented in this paper is to create a web application using text mining techniques so as to classify, quantify, and present the information of the texts entered. The aim is to support people dedicated to the study of language behavior in sociolinguistic groups.

Keywords: Text Mining, Sociolinguistic Web Application, Text Classification.




Translated by,
Lic. Lourdes Crespo

Introducción

En la actualidad las aplicaciones web se encuentran en la cúpula de las herramientas para la administración y presentación de información digital, ahora bien, el desarrollo de una aplicación web de calidad depende de varios procesos de análisis como: el sistema gestor de base de datos a usar, el lenguaje de programación para la construcción, el personal involucrado en el desarrollo, de qué manera se va a administrar el proceso de desarrollo, y como usar todo esto para cumplir de manera eficiente los objetivos para los que se va a construir la aplicación web.

En el presente trabajo, se considera la inclusión de minería de texto, los procesos de minería de texto consisten en derivar información de textos no estructurados para generar nuevo conocimiento, en este proyecto se va a generar procesos de clasificación de textos en base a los idiomas español y quechua, con el fin de obtener conocimiento acerca del comportamiento lingüístico en los diferentes grupos sociales.

El siguiente punto será generar bocetos o diseños para que el usuario final los apruebe y se sienta a gusto con las interfaces futuras de la aplicación web, estos bocetos permiten generar documentación que refleje el estado del proceso de desarrollo desde su inicio.

Los procesos de ingeniería para el desarrollo de software ya presentan algunos puntos importantes como: el diseño de una base de datos cuyo objetivo es estructurar los datos correctamente y así evitar problemas con la integridad de los datos de la aplicación, un diseño que represente el flujo de la aplicación web, es la manera eficaz de obtener los requisitos de software y hardware necesarios para el desarrollo y de la aplicación web. El modelo de capas de un sistema distribuido permite a un grupo de desarrolladores planear la gestión de todas las tareas de adquisición y adecuación de hardware o software para el entorno de trabajo, como dice un refrán "no se

puede cabalgar sin haber ensillado", una vez se encuentre listo el ambiente de desarrollo, el proceso de desarrollo puede iniciar adecuadamente.

Así también, el seguimiento del proceso de desarrollo de software es una tarea que debe llevarse a cabo fielmente con el objetivo de generar una aplicación web de calidad en el tiempo previsto. Scrum un marco de desarrollo ágil de proyectos que considera: las necesidades de los usuarios, los participantes y la planificación de desarrollo del software, siendo esta metodología la acogida para el presente proyecto. Scrum propone la inclusión del usuario final en el proceso de desarrollo de software, lo que disminuye la cantidad de los cambios solicitados por el usuario en la fase final.

Finalmente, un punto a tomar en cuenta es la aplicación de pruebas al software desarrollado, se propone las pruebas de sintaxis, semántica y de interfaz. Pruebas como las de coherencia de datos son aplicadas en el proceso de desarrollo de software.

Capítulo 1: Indagación exploratoria

En este capítulo se presentan los conceptos básicos fundamentales para el presente trabajo de titulación. En primer lugar, se describe la noción de un sistema gestor de base de datos para usarlo como almacén y procesador de los datos de la aplicación, la selección de un adecuado sistema gestor de base de datos que contenga funciones específicas como: la minería de texto, extracción de datos que estén en formato XML, entre otras, es el primer objetivo de la indagación para este trabajo.

Por consiguiente, debido a la inclusión de la minería de texto como objetivo primordial, se pueden generar conflictos internos en la eficiencia de selección de datos cuando la aplicación web entre en operación, entonces, se necesita indagar en la adecuada forma de configurar y administrar el sistema gestor de base de datos, si no se toma en cuenta este punto, el usuario optaría por declinar de su uso debido a las ineficiencias.

Para generar una aplicación web amigable para el usuario final, se necesita indagar acerca de herramientas de desarrollo de software que se integren al proyecto y tengan una buena popularidad por parte de los desarrolladores, en este apartado se tendrá como objetivo cubrir dos herramientas básicas para la interfaz del usuario: la primera el uso de los mapas digitales para la ubicación geográfica de las comunidades, y la segunda, el uso de una herramienta de desarrollo que permita realizar gráficas estadísticas válidas para la presentación de información.

Finalmente, indagar en técnicas de minería de texto que sirvan de guía para este trabajo, pues la minería de texto no es un campo común cuando se trata de crear aplicaciones web. El objetivo de la indagación es el uso e integración de las herramientas junto con las técnicas antes mencionadas con el fin de generar un software totalmente funcional y amigable para el usuario final.

1.1 Sistema gestor de base de datos

A una base de datos se le puede considerar como una colección de datos relacionados. El SGDB (Sistema manejador de base de datos) es el software que interactúa con el software de usuario y con la base de datos, permite a los usuarios definir su base de datos mediante DDL (Lenguaje de Definición de Datos), permite también insertar, actualizar y eliminar los datos mediante DML (Lenguaje de Manipulación de Datos), el lenguaje más común para realizar esto es mediante SQL (Lenguaje de Consultas Estructurado), otra ventaja del SGDB es asegurar la base de datos mediante los controles de acceso, integridad y concurrencia (Thomas Connolly, 2005).

Una de las bases fundamentales para el desarrollo de una aplicación web, es la apropiada distribución y tratamiento de los datos que esta genera, por tanto, el diseño de una base de datos bien concebido es el cimiento principal sobre el cual se trabaja en la etapa de codificación de la aplicación. La elección de un sistema gestor de base de datos adecuado que cumpla con los requisitos de actuación para el correcto tratamiento de los datos de la aplicación es la incógnita previa al desarrollo de cualquier software.

Existen varios SGBD populares como los presentados por (DB-Engines, 2016) en su publicación, utilizando este referente y en base al conocimiento técnico, se toma como candidatos a: SQL Server, MySQL, Oracle Database, PostgreSQL, todos estos tienen una gran cantidad de herramientas para el tratamiento de los datos, ahora bien, es importante hacer una evaluación comparativa entre estos con el fin de obtener el producto que mejore y facilite el tratamiento de los datos según las necesidades futuras en la etapa de desarrollo del software.

Normalmente una comparativa entre productos SGDB basada en un estándar para medición de calidad de productos de software como la ISO 25010 se convertiría en un arduo trabajo, tomando en cuenta esto como punto de partida se va a comparar las características más importantes en un SGDB, centrando a la minería de texto como la función primordial en las herramientas de los SGDB, a continuación, se realiza una comparación según las distintas necesidades básicas en el desarrollo de la aplicación.

Una ventaja para la homologación de los diferentes SGBD, es que los proveedores de estos publican sus características y métodos propios para el manejo y procesamiento de los datos, toda esta información permite realizar una rápida homologación y selección bajo principales criterios propuestos, a continuación, en la tabla 1.1 se presenta una breve comparativa entre los SGBD junto a criterios básicos predefinidos para la selección.

Criterio	SQL Server	MySQL	Oracle Database	PostgreSQL
Consulta de texto	Si	Si	Si	Si
Expresiones regulares	Si	Si	Si	Si
Desarrollo rápido de aplicaciones web	No	No	Si	No
Minería de datos	Si	No	Si	No
Soporte a XML	Si	Si	Si	Si
Obtener versiones antiguas de datos	No	No	Si	No
Soporte JDBC y ODBC	Si	Si	Si	Si
Total de (Si)	5	4	7	4

Tabla 1.1: Comparación entre SGBD populares

Elaboración: Autor, basado en: (Microsoft Corporation, 2016); (MySQL, 2016); (Oracle Corporation, 2016); (Postgresql, 2016)

La comparativa entre los SGBD candidatos y sus diferentes herramientas de trabajo indica que las necesidades básicas y fundamentales son totalmente cubiertas por dos SGBD: SQL Server y Oracle Database, pero existe una característica que puede favorecer totalmente el desarrollo del proyecto y es el “Desarrollo rápido de aplicaciones web” propia de Oracle Database, esta característica de desarrollo rápido de aplicaciones es un marco de trabajo llamado Oracle APEX (Oracle Application Express) que permite diseñar y crear páginas web con completa funcionalidad dependiendo de las tablas del esquema, dando el valor agregado que comúnmente se busca en los proyectos de desarrollo de software, por ende, el SGBD elegido para el desarrollo es Oracle DataBase.

1.2 Diseño de la base de datos

El diseño de la base de datos es tomado como un proceso que lleva un análisis y modelamiento de un conjunto de requisitos para que una base de datos esté organizada (Thomas Connolly, 2005). Un buen diseño de la base de datos y la administración de datos proporcionado por un buen SGDB son el núcleo de diferentes software de alta calidad, ahora bien, aquí nace otra incógnita, ¿Basta con un SGDB y un diseño de la base de datos para obtener el máximo rendimiento?, en búsqueda de la respuesta (Basit Raza, 2016) en su publicación en la librería digital IEEE Xplore, propone algunos consejos a tener en cuenta tanto en la configuración como en la recolección de los datos que estarán presentes en la base de datos para obtener un excelente rendimiento en Oracle Database, el autor expone:

1. Auto optimización

Referente a monitorear continuamente y en buscar optimizar cuando se construye.

a) Optimizar las consultas, en Oracle Database, el tuning se lo puede desarrollar en cuatro pasos propuestos:

- Identificar las consultas de alta carga.
- Generar un plan de optimización.
- Mejorar el plan de ejecución de consultas (QEP).
- Generar un mejor QEP.

b) Componente para colección de estadísticas automáticos.

Oracle genera diferentes tipos de estadísticas como: de la base de datos, las del sistema operativo y de interpretación.

c) Monitor automático de diagnóstico de la base de datos

Las actividades incluyen *SQL tuning*, espacio de fragmentación de análisis, cuellos de botella, adaptar las estructuras de memoria, alta carga de ejecución de los *requests*.

d) Repositorio automático de carga de trabajo (AWR).

Es un *snapshot* que almacena temporalmente los datos bajo un periodo de tiempo.

2. Auto Configuración

Característica para configurar las metas definidas, el software debe ser capaz de auto-configurarse o adaptarse según el medio en el que se encuentre.

3. Auto recuperación

Característica para mantener el sistema en trabajo y en un estado operacional.

4. Auto protección

Es la capacidad de protegerse de objetos maliciosos y el acceso no autorizado. Oracle Database y Oracle APEX (Oracle Application Express) ya llegan con estas características, por ejemplo: cambio de contraseñas temporalmente y el bloqueo de cuentas por continuos accesos fallidos.

1.3 Mapas digitales

Un mapa es la representación geográfica total o parcial de la tierra sobre una superficie plana, dicha representación se encuentra basada en una escala. Los mapas digitales son utilizados para aplicaciones o sistemas que necesiten del uso de ubicaciones geográficas y como los datos son digitales son fáciles de actualizar y distribuir (R. Ohbuchi, 2002).

Un requerimiento del presente trabajo es la ubicación de las comunidades en un mapa digital, por tanto, la selección de software que permita realizar esta tarea es primordial debido a que ya se tiene contemplado el SGDB y el mapa digital debe ser capaz de presentar los datos que se encuentren almacenados en este. Como requerimiento del mapa digital es priorizar la adaptación a la web, por lo tanto, debe ser compatible con lenguajes de programación como: HTML, JavaScript o PHP.

La probabilidad de usar el marco de trabajo Oracle APEX debido al SGDB seleccionado en el apartado anterior es alta, por lo tanto, una prueba de adaptación de un mapa digital en Oracle APEX es útil en esta instancia de la indagación, la selección debe tomar en cuenta este nuevo requerimiento de compatibilidad con el SGDB.

En el medio existen herramientas populares como: ARGIS, Google Maps API y Here Maps, entre otros, igual que en la comparativa de los SGBD se busca cubrir las necesidades de la aplicación para obtener la mejor herramienta de desarrollo, mediante la tabla 1.2 se compara los diferentes mapas digitales:

Característica	ARGIS	Google Maps	Here Maps
Compatibilidad HTML	Si	Si	Si
Compatibilidad JavaScript	Si	Si	Si
Mapa mundial	Si	Si	Si
Licencia	Si	No	Si

Tabla 1.2: Comparación entre mapas digitales

Elaboración: Autor, basado en: (Argis, 2016); (Google Maps, 2016); (Here Maps, 2016).

Todas las herramientas de desarrollo candidatas cumplen los requisitos de compatibilidad con la aplicación a desarrollar, en este caso la elección es basada completamente en el costo de uso del software, Google Maps brinda gratuitamente hasta 25000 peticiones o "requests" por día, esta característica cubre los requerimientos de la aplicación.

Para profundizar el uso Google Maps API, el autor (Li, 2011) indica en su publicación, que esta herramienta permite una fácil adaptación de la cartografía digital con las bases de datos, además, permite la facilidad de controlar el mapa mediante eventos del navegador como: clicks, mouse over, etc., junto con la capacidad de presentar información real del mapa como calles, ciudades y países. Permite dibujar ciertas figuras como círculos y polígonos, e importar figuras en él. Todas estas características de Google Maps API permite realizar diversas combinaciones entre diferentes formas y colores.

Los criterios de selección propuestos en la tabla comparativa de mapas digitales y la descripción de (Li, 2011) en su publicación hacen de la herramienta Google Maps la mejor opción para integrarla al desarrollo del software.

1.4 Google Charts API

La librería digital Google Charts Api es una galería de gráficas estadísticas, simple de usar en aplicaciones web, esta librería es propia de la empresa Google, para uso personal brinda acceso gratuito y para uso empresarial un acceso pagado (Google Charts, 2016). El objetivo de la minería de texto es obtener información nueva de los textos que se encuentren almacenados, por tanto, utilizar Google Charts API para presentar la información estadística que se obtenga de la minería genera un valor agregado para el cliente en la aplicación web.

El autor (Zhu, 2012) en su publicación acerca del uso de gráficas, presenta la herramienta de Google Charts Tools para crear dibujos basados en la fuente de datos y agregarlos a una página web, esta herramienta puede generar hasta 12 tipos diferentes de gráficas en los que se encuentran los siguientes tipos: pie, charts, scattercharts, gauge charts, geocharts, tables, treemaps, combo charts, linecharts, bar, la ilustración 1.1 presenta un ejemplo de visualización de google chart API.

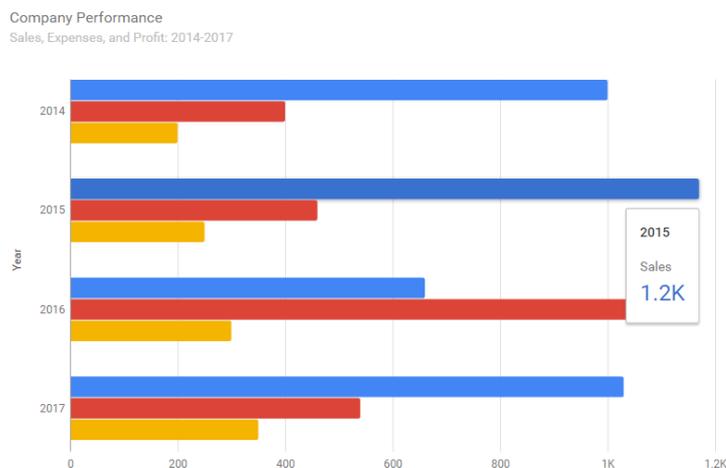


Ilustración 1.1: Ejemplo Google Chart API

Esta herramienta cumple con los objetivos de facilitar el desarrollo de la aplicación web y la presentación de la información obtenida en la minería de texto, además, el ambiente amigable que brinda para el usuario son la causa para seleccionarla como parte del desarrollo del software.

1.5 Oracle Application Express

Oracle APEX (Oracle Application Express), propuesto por la empresa Oracle como un marco de trabajo para el desarrollo rápido de aplicaciones web (Oracle Apex, 2016), este software presenta dos principales ventajas: la primera, contiene plantillas predefinidas para la administración y presentación de los datos, y la segunda, se encuentra integrado en Oracle Database. El autor (Wojciechowski, 2011) ha redactado un documento en el que genera un sistema transaccional y muestra pautas de las ventajas del marco de trabajo como:

1. *Application builder*, un espacio completamente dedicado a la creación y edición de aplicaciones web con bases de datos dinámicas.
2. *SQL Workshop*, un espacio dedicado a navegar por los objetos de la base de datos y ejecutar consultas SQL.
3. Interfaces para ingreso de datos, como está ligado directamente a Oracle Database facilita toda la administración de datos.
4. Múltiples desarrolladores pueden acceder al entorno de desarrollo y modificar la aplicación usando SQL, HTML y JavaScript, estas opciones se las puede distinguir en la ilustración 1.2.

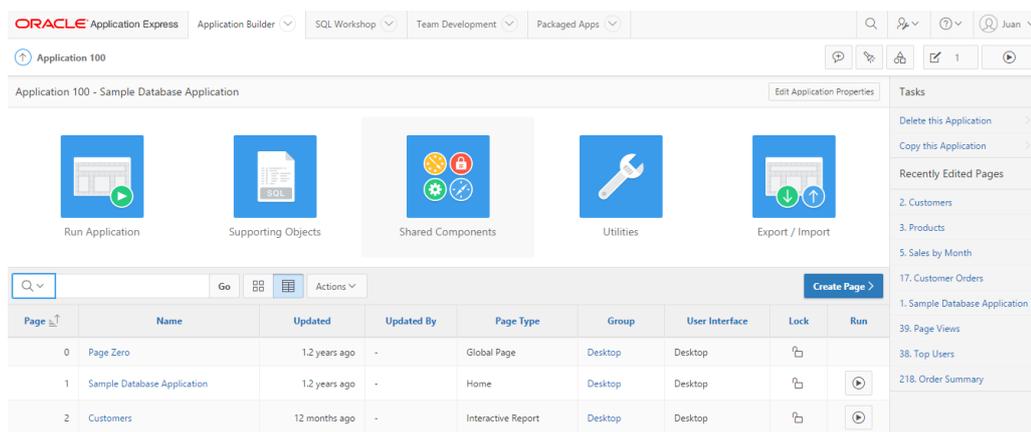


Ilustración 1.2: Oracle APEX Application Builder

Como se presenta en el inicio de este apartado, una herramienta para el desarrollo ágil de aplicaciones web, junto a las características enumeradas señala a Oracle Application Express como la mejor opción para el desarrollo del presente proyecto.

1.6 Minería de texto aplicada a la detección de idioma y fonemas

La minería de datos es un campo donde se ha dado grandes avances en los últimos años. La minería de textos un caso particular de minería de datos, su desarrollo destinado a las aplicaciones web y redes sociales, lo permite crear rápidamente repositorios de datos y basados en estos derivar nuevos conocimientos. Generalmente la minería de datos estructurados es administrada mediante un SGDB que facilita su ejecución, a diferencia de esta, en la minería de texto debido a la falta de estructuración normalmente los datos son administrados mediante un motor de búsqueda, esto hace que la minería de texto se torne en una tarea más compleja que la minería de datos numéricos (Charu Aggarwal, 2012).

En la aplicación web a desarrollar se necesita cuantificar las palabras en base al idioma de los textos que el usuario ingrese, a fin de obtener información numérica del comportamiento lingüístico, por ejemplo: en un texto ingresado identificar el porcentaje del idioma español.

1.6.1 Clasificación de los algoritmos de minería de datos

El autor (Lopez Perez C, 2007) en su libro indica que los algoritmos para minería de texto se pueden clasificar mediante su propósito general en los siguientes tipos:

- Modelos Predictivos: Para ejecutar una minería valida es necesario comprobar el modelo con el dato. Se describe un modelo en base a un conocimiento previo, lo que permite predecir el valor de un atributo que no es conocido.

En algunas ocasiones el modelo es la resultante de conocimiento obtenido antes y después de la aplicación de la minería, conocido como entrenamiento del modelo.

En este tipo de modelos se incluye: la regresión, series temporales, análisis de varianza, de covarianza, análisis discriminante, arboles de decisiones y redes neuronales.

- Modelos descriptivos: Estos son modelos que se crean a partir del reconocimiento de patrones. En estos algoritmos no se supone la existencia de variables de dependencia o independencia, no existe un modelo previo para los datos.

Se incluyen tipos de algoritmos como: *clustering*, segmentación y técnicas de asociación.

Este es el tipo de algoritmo en el que se desarrolla la minería de texto en el presente proyecto debido a la clasificación mediante reconocimiento de patrones basado en el algoritmo GNU Aspell.

- Modelos Auxiliares: Son algoritmos específicos para el apoyo, basados en técnicas estadísticas y consulta cuyo objetivo es la verificación.

1.6.2 Algoritmos en Oracle Database para minería de texto

Oracle Database propone siete tipos de algoritmos para minería de texto, los cuales se detallan en la tabla 1.3:

Algoritmo	Función de minería
Naive Bayes	Clasificación
Modelos lineales generalizados	Clasificación, regresión
Soporte con máquinas de vectores	Clasificación, regresión, detección de anomalías
Grupos-K	Agrupamiento
Matriz de factorización no negativa	Extracción de características
De anterior	Reglas de asociación
Longitud mínima de descripciones	Importancia del atributo

Tabla 1.3: Algoritmos de minería de texto en Oracle Database

Elaboración: (Oracle Help Center: Text mining, 2016)

La función de clasificación es considerada dentro de la minería de textos, en la clasificación de Naive Bayes por Oracle Database no permite clasificar textos en idiomas extras a los definidos, por ello, se tiene que realizar una clasificación asistida por otro algoritmo el cual es presentado en el siguiente apartado.

1.6.3 Algoritmo GNU Aspell aplicado a detección de idioma

En la actualidad existen algoritmos implementados para corrección ortográfica, una ventaja de este tipo de algoritmos es que se adaptan a varios idiomas, por medio de diccionarios y cuando se hace la verificación de un texto o palabra, un parámetro indica en que idioma se quiere hacer la verificación y un valor de retorno indica si el texto pertenece a dicho idioma ingresado.

El algoritmo presentado por (Atkinson, 2002) implantado en el software GNU Aspell sirve esencialmente para la verificación ortográfica en tiempo real, este algoritmo actualmente está implementado en tipos de software como navegadores y openOffice, trabaja con diccionarios disponibles en la web; el uso de este algoritmo permite saber si una palabra o texto pertenece a un idioma, la ventaja de este algoritmo es su adaptación a otros entornos de trabajo y su mejor eficiencia frente a otros algoritmos de este tipo como: Ispell, NetScape 4.0, Microsoft Word 97, todas sus ventajas y usos están presentados con detalle en el documento del autor.

En el presente proyecto el método aplicable en la minería de texto para clasificar las palabras en el idioma al cual pertenece utiliza el software Aspell y su algoritmo de verificación.

1.7 Conclusiones del capítulo

Al concluir este capítulo se puede afirmar que se realizó una selección de herramientas software para el desarrollo de la aplicación web, como:

- El sistema gestor de base de datos: Oracle Database.
- El mapa digital: Google Maps.
- Gráficas estadísticas: Google Charts API.
- El algoritmo de clasificación de textos.

Estas herramientas fueron seleccionadas con el fin de integrarlas y administrarlas mediante el marco de trabajo de desarrollo de software ágil Oracle APEX que se ejecuta con Oracle Database haciendo más sencilla la administración de los datos desde la aplicación web, Oracle APEX junto a las

técnicas de configuración y administración del SGBD generan una mejor optimización de la aplicación.

La clasificación mediante el algoritmo de GNU Aspell, genera una gran ventaja al disminuir el tiempo de desarrollo en otro algoritmo personal para obtener la precedencia idiomática de una palabra.

Capítulo 2: Especificación de requerimientos del software

A fin de obtener un producto de calidad es necesario especificar todos los requerimientos o especificaciones que el software debe contemplar para satisfacer las necesidades del cliente. Para llevar a cabo esta tarea se utiliza en modelo presentado por la IEEE (Instituto de Ingeniería Eléctrica y Electrónica), denominado como “Estándar IEEE-830-1998” que es aplicado por gran parte los desarrolladores de software para definir todos los requerimientos de software en producción.

2.1 Aplicando el estándar IEEE-830

En la plantilla del estándar se encuentran detallados múltiples versiones en el punto 3, por facilidad de comprensión en este documento se utiliza la plantilla de A.3 de SRS Sección 3 “Organizada por la clase del usuario”.

2.1.1 Introducción

La especificación de todos los requerimientos del software a desarrollar es el punto de partida para las empresas dedicadas a la producción de software, el objetivo es que el producto final cumpla con el propósito para el que fue creado, además, generar un software de calidad.

2.1.1.1 Propósito

- a) Identifica con claridad objetivos, herramientas y algunos estándares que conducen a un proceso adecuado de desarrollo del software.
- b) Aportar con un documento que contenga de una manera estructurada todos los procesos a ejecutar junto con las herramientas de uso en el proceso de desarrollo del software.

2.1.1.2 Ámbito del sistema

El software denominado "Análisis de lengua mediante minería de texto", es una aplicación web que brindará soporte en los siguientes aspectos:

- a) Visualizar la ubicación geográfica en el territorio ecuatoriano de los diferentes grupos "comunidades" a analizar.
- b) Facilitar la tabulación de la información de los diferentes grupos sociales en análisis.
- c) Presentar la información de los diferentes grupos sociales en análisis mediante contenido textual y multimedia.
- d) Realizar un análisis del comportamiento de la lengua en base a los textos que se hayan ingresado en el software mediante minería de texto.
- e) Presentar la información resultante de la minería de texto en tablas o gráficas que faciliten en comprensión de los datos obtenidos.
- f) Facilitar la administración de la información referente a comunidades, modelos y entrevistas.

2.1.1.3 Definición de acrónimos y abreviaturas

Para simplificar la comprensión de este documento es de vital importancia identificar algunos conceptos, acrónimos y abreviaturas que se utilizan en el desarrollo de este documento, la tabla 2.1 presenta en detalle estas palabras del documento.

Nombre	Abreviatura	Descripción
Instituto de Ingeniería Eléctrica y Electrónica	IEEE	Instituto sin fines de lucro, que promueve la innovación y excelencia de la tecnología.
Especificación de requisitos de software	ERS	Estándar propuesto por la IEEE para la especificación de atributos de software.
Lenguaje unificado de modelado	UML	Es un estándar internacional para crear documentación referente al desarrollo o comportamiento del software.
Sistema gestor de base de datos	SGDB	Uno o varios programas que administran una base de datos. Oracle Database: Es un SGDB junto a una instancia de base de datos, software propio de la empresa Oracle.
Oracle Application Express	Oracle APEX	Software para el desarrollo de aplicaciones web, es propio de la empresa Oracle.

Lenguaje Procedural/Estructura de lenguaje de consultas	PLSQL	Lenguaje de programación para Oracle Database.
Lenguaje de definición de servicio web	WSDL	Es un estándar para la utilización de software como servicio.
Integración de modelos de madurez de capacidades	CMMI	Es un estándar para garantizar la calidad en el software.
Real academia española	RAE	Organización dedicada a crear reglas normativas para el idioma español.
Google APIs		Interfaces para facilitar el desarrollo de aplicaciones web, propias de la empresa Google.

Tabla 2.1: Conceptos, acrónimos y abreviaturas

2.1.1.4 Visión general del documento

En esta especificación de requisitos se presenta las características del software, características de los usuarios, interfaces de comunicación con el servicio web y el prototipo de interfaces para el usuario final, además, se adjunta el diagrama de casos de uso para sintetizar todas las funciones de la aplicación web.

2.1.1.5 Descripción general

En el transcurso de este apartado se describe los siguientes factores: software para desarrollo, las técnicas de desarrollo que afectan directa o indirectamente al software y a sus requisitos, finalmente, los diseños de interfaz de usuario a aplicar en la aplicación web.

2.1.2 Descripción global

2.1.2.1 Perspectiva del producto

El producto software a desarrollar apoya a la investigación de Doctorado de la Mgst. Priscila Verdugo titulado "Hacia una Cartografía Sociolingüística entre el español y el Quichua en las provincias del sur del Ecuador: la lengua y la construcción de identidades", la cual propone "desarrollar un software que permita construir cartografías lingüísticas, en las cuales se podría evidenciar lo que está pasando con el español y el quichua por regiones geográficas, en cuanto a estructuras sintácticas, uso, actitud e identidad".

2.1.2.2 Funciones del producto

El software abarca funciones tanto en ingreso como de presentación de información referente a las comunidades en análisis:

1. Gestionar las comunidades y permitir la ubicación de estas en mapa regional ecuatoriano.
2. Gestionar los modelos de estudio que se aplicarán a una comunidad.
3. Gestionar los eventos de estudio, un evento de estudio es un escenario en el que se da el levantamiento de información.
4. Gestionar información relevante de cada comunidad.
5. Gestionar contenido multimedia referente a una comunidad con sus eventos y modelos de estudio.
6. Analizar el comportamiento del idioma en los textos que se encuentren almacenados en el software, basado en minería de texto.
7. Presentar los datos de las comunidades y los resultados del análisis mediante tablas o gráficas estadísticas.

2.1.2.3 Características del usuario

Tanto el usuario gestor de contenido como el usuario visitante deben tener características en conocimientos de informática para administrar el software:

1. Conocimiento acerca de software informático.
2. Conocimiento en lingüística e interpretación de estadísticas básicas.

2.1.2.4 Restricciones del software

De una manera macro se presenta las limitantes del software para los diseñadores y desarrolladores en el presente proyecto:

1. Compatibilidad de software de desarrollo: El marco de trabajo de desarrollo Oracle APEX presenta limitantes en cuanto a la interacción con otros lenguajes de programación de desarrollo haciendo de las nuevas fronteras de desarrollo difíciles de combinar con el sistema.
2. Consultas SQL en Oracle Database: Oracle Database permite nombrar una cabecera de columna con un límite de 30 caracteres, la opción es recortar caracteres.

2.1.2.5 Atención y dependencias

La aplicación estará desarrollada en el marco de trabajo Oracle APEX, por tanto, este debe estar instalado y configurado en los servidores, a su vez esta plataforma de desarrollo trabaja solo con Oracle Database. La ilustración 2.1 presenta su dependencia, eso indica que si en algún momento se da la necesidad cambiar el último nivel de este sistema (Oracle Database) se tendría que cambiar todos los requerimientos de desarrollo.



Ilustración 2.1: Dependencia entre software de desarrollo

2.1.3 Los requisitos específicos

Los requisitos específicos indican que técnicas de modelado que se aplicará en el desarrollo del software.

2.1.3.1 Requisitos de interfaz externa

La aplicación web requiere de técnicas de minería de texto, por lo que es necesaria la implementación de un servicio web externo que realice todas las clasificaciones de texto que se requieran, por tanto, es justo tener en cuenta la adecuación del software para que se conecte a un servicio web para facilitar la comprensión de los datos resultantes.

2.1.3.1.1 Interfaz de usuario

Se tiene definido que la interfaz de usuario será una adecuación de una de las plantillas proporcionadas por Oracle APEX, tomando en cuenta esto se generan los siguientes diseños para cada tipo de opción en el software.

a) Iniciar sesión: mediante ilustración 2.2 se presenta el diseño de la página para que el usuario pueda identificarse en la aplicación web.

Ilustración 2.2: Diseño de inicio de sesión

b) Mapas digitales: mediante la ilustración 2.3 se presenta el diseño del uso de mapas geográficos en la aplicación web.

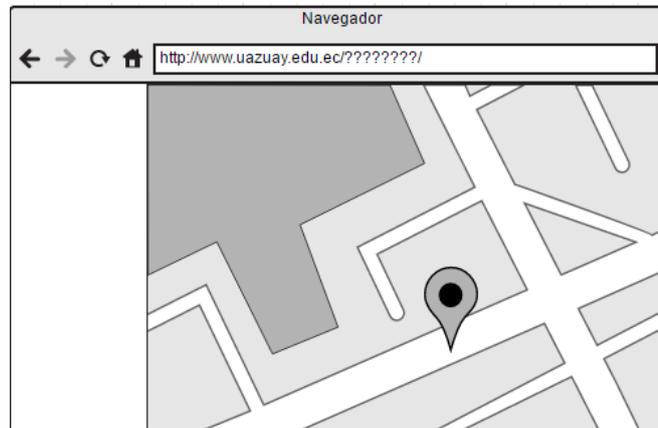


Ilustración 2.3: Diseño de mapa digital en la aplicación

c) Reportes: mediante la ilustración 2.4 se presenta el diseño para mostrar datos de una o varias entidades pertenecientes de la aplicación web.

Cartografía etnolingüística			
Comunidad	Head 1	Head 2	Head 3
Evento	Cell 1	Cell 2	Cell 3
Modelo	Cell 4	Cell 5	Cell 6
Punto Análisis	Cell 7	Cell 8	Cell 9
Info Comunidad	Cell 10	Cell 11	Cell 12
Administración			

Ilustración 2.4: Diseño para reportes de datos

- d) Mantenimientos: mediante la ilustración 2.4 se presenta el diseño de un mantenimiento que permite agregar, editar o eliminar datos referentes a una entidad de la aplicación web.

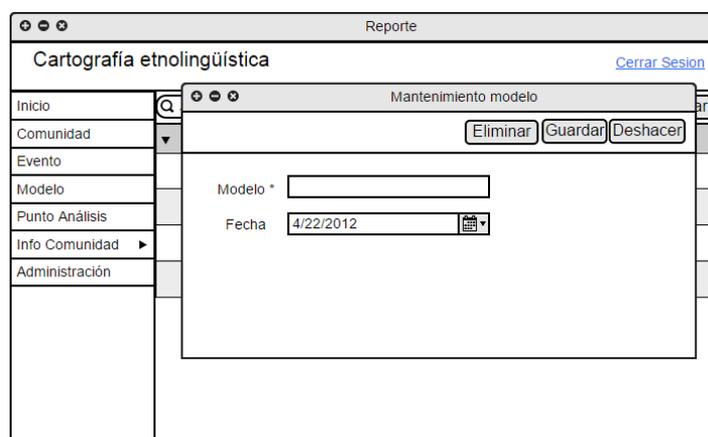


Ilustración 2.5: Diseño para mantenimientos

2.1.3.2 Listado de requisitos funcionales

2.1.3.2.1 Usuario administrador

- Gestionar comunidades: nombre, ubicación geográfica.
- Gestionar modelos de estudio: nombre del índice plantilla del análisis.
- Gestionar eventos de estudio: escenas en las que se realiza el análisis.
- Gestionar los puntos de análisis: puntos que pertenecen al índice del modelo de estudio.
- Gestionar el análisis de la comunidad: editar información de comunidad basada en el modelo de estudio, puntos de análisis y eventos de estudio.
- Gestionar notas acerca de comparaciones entre comunidades.
- Gestionar información multimedia de comunidad: información como fotografías, audio y video.
- Generar informes en archivos de formato pdf con información del análisis de las comunidades.
- Minería de texto: obtener de forma automática información cuantificada sobre los textos ingresados como: porcentaje de idioma detectado y un ranking de las palabras usadas.
- Además, puede visualizar todos los requisitos del usuario visitante.

2.1.3.2.2 Usuario visitante

- Visualizar geográficamente todas las comunidades analizadas.
- Visualizar información general de una comunidad.
- Visualizar el análisis investigativo de una comunidad.
- Visualizar datos de la minería de texto de una comunidad.
- Visualizar fotografías representativas de una comunidad.
- Visualizar grabaciones representativas de una comunidad.
- Visualizar videos representativos de una comunidad.
- Visualizar las conclusiones acerca de cada comunidad.

2.1.3.3 Interfaz de software

La aplicación debe estar conectada al algoritmo de clasificación de textos, por tanto, es necesario comunicar de alguna manera la aplicación web con el detector de texto, es así que se tiene:

Software base:

- Oracle APEX: marco de trabajo para desarrollo de la aplicación web.
- Lenguaje de programación Java: para implementar el algoritmo de clasificación de idioma.

Interfaz de comunicación de software:

- Servicio web, que permite conectar fácilmente aplicaciones orientadas a la web incluso si pertenecen a diferentes lenguajes de programación.

2.1.3.4 Requisitos de la actuación

La aplicación web a desarrollar contará con los siguientes requisitos de actuación en eventos básicos en un computador como:

- Eventos de mouse, los cuales funcionan como disparadores entre consultas y navegación entre las diferentes páginas de la aplicación.
- Eventos de teclado, para realizar los ingresos de información textual y así obtener un correcto funcionamiento de la aplicación.

2.1.3.5 Constreñimiento del plan

Según la RAE la palabra constreñimiento significa "una compulsión que se hace a alguien para que ejecute algo" (Real academia española, 2016). Por tanto, en este punto se identifica la necesidad del desarrollo de la aplicación web. Una aplicación web que pueda manejar información relevante al estudio de comunidades es un gran apoyo para personas dedicadas a buscar diferentes patrones de comportamiento idiomático de sus comunidades en estudio, dando así, la facilidad obtener nueva información de las comunidades en estudio permitiendo verificar las hipótesis de manera más fácil y organizada.

2.1.3.6 Atributos del sistema de software

En el modelo de calidad proporcionado por la ISO/IEC 25010 (ISO 25000 Standards, 2016) en su estándar detalla las características de calidad para el software. Considerado que este proyecto se enfoca en el desarrollo de una aplicación web, se busca centrar y fortalecer el atributo de Usabilidad en las siguientes sub-características:

- Capacidad para reconocer su adecuación: Capacidad del producto que permite al usuario entender si el software es adecuado para sus necesidades.
- Capacidad de aprendizaje: Capacidad del producto que permite al usuario aprender su aplicación.
- Capacidad para ser usado: Capacidad del producto que permite al usuario operarlo y controlarlo con facilidad.
- Protección contra errores de usuario: Capacidad del sistema para proteger a los usuarios de hacer errores.
- Estética de la interfaz de usuario: Capacidad de la interfaz de usuario de agradar y satisfacer la interacción con el usuario.

Estas sub-características definen las capacidades a tomar en cuenta en el proceso de desarrollo del software para que el usuario pueda entenderlo, aprenderlo, operarlo y sea capaz de evitar errores humanos.

2.2 Diagrama de casos de uso del software

Una forma fácil y eficaz de representar el funcionamiento global de un software en desarrollo es mediante el diagrama de casos de uso, el funcionamiento de la aplicación web a desarrollar se observa en la ilustración 2.6, este tipo de diagrama pertenece al lenguaje UML, que representa la interacción interna del software o representa la interacción con un usuario, la referencia de las diferentes notaciones del diagrama se puede encontrar en la página oficial de los diagramas UML (The Unified Modeling Language, 2016).

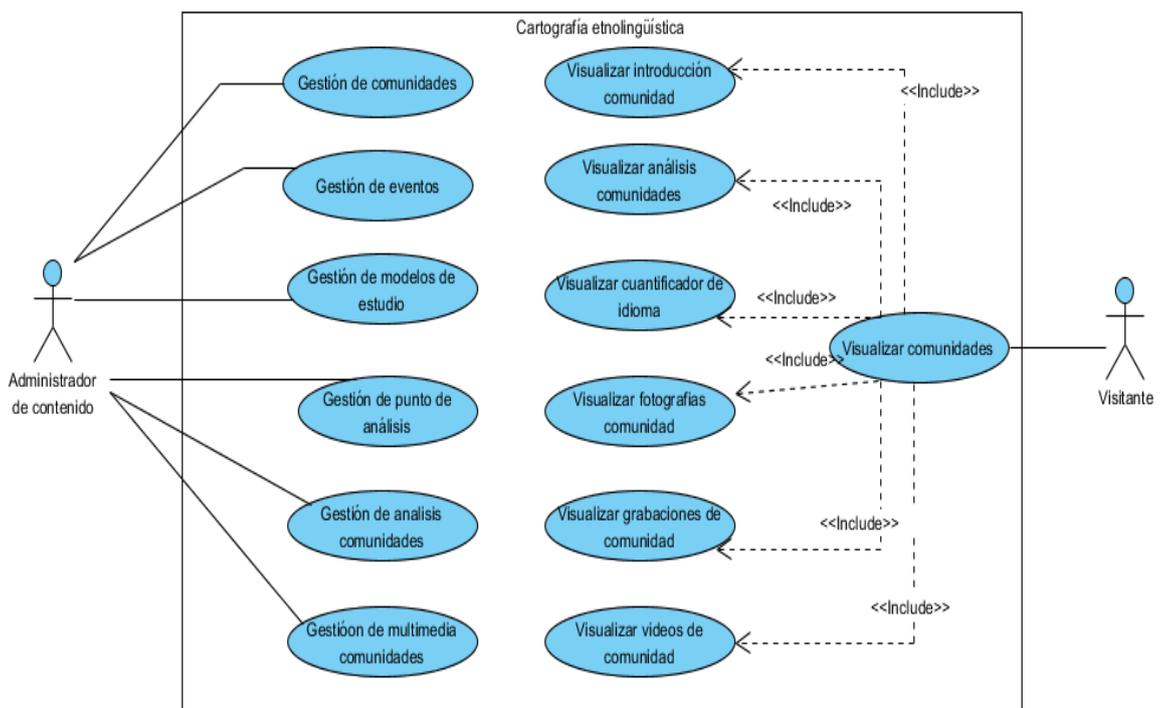


Ilustración 2.6: Diagrama de casos de uso

2.3 Metodología para el desarrollo del software – Scrum

Scrum es una metodología ágil y flexible generalmente aplicada en proyectos de desarrollo de software, esta metodología se enfoca en priorizar el desarrollo de las funciones más importantes para el cliente de forma colaborativa, por lo general se usa cuando el cliente quiere formar parte del grupo de trabajo (observando las implementaciones en el software), de esta manera se busca obtener la mayor satisfacción para el cliente.

2.3.1 Asignación de roles

En Scrum, la asignación de roles se identifica de forma simple a las personas y su labor en el proceso de desarrollo del software, la tabla 2.2 presenta los roles de Scrum para este proyecto de desarrollo.

Scrum	Nombres
Equipo de Scrum	Desarrollador, directores de proyecto.
Dueño del producto	Ciente.
Scrum Masters	Director de desarrollo, Director de minería de texto.
Equipo de desarrollo	Desarrollador del proyecto.

Tabla 2.2: Scrum - Asignación de roles

2.3.2 Listado de requerimientos del usuario

Una lista general de todos los requerimientos de un usuario en el software a desarrollar.

1. Ubicar geográficamente las comunidades en análisis del cliente.
2. Registrar la información obtenida acerca de una comunidad.
3. Almacenar archivos multimedia de una comunidad.
4. Analizar el estado de la lengua en cada comunidad.
5. Presentar todos los datos recopilados de manera sistematizada.
6. Acceder a toda su información en cualquier momento.

2.3.3 Historias de usuario

Las historias de usuario son tarjetas que identifican los objetivos en el proceso de desarrollo, las tablas en secuencia desde la tabla 2.3 hasta la tabla 2.30 son las historias de usuario a cumplir en desarrollo del proyecto.

2.3.3.1 Historias de usuario- Necesidades del software

núm. 1	Nombre: Especificar los diagramas de casos de uso	
Usuario: Desarrollador		
Modificación de historia número:		Iteración asignada: 1
Prioridad en negocio: Media (Alta / Media / Baja)		Tiempo estimado: 0.7 semanas
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)		Tiempo real: 0.7 semanas
Descripción: Para una mejor visión se emplea un diagrama de casos de uso de la aplicación.		
Observaciones:		

Tabla 2.3: Scrum - Historia de usuario 1

2.3.3.2 Historias de usuario- Diseñar y estandarizar la base de datos

núm. 2	Nombre: Crear y normalizar la base de datos	
Usuario: Desarrollador		
Modificación de historia número:		Iteración asignada: 2
Prioridad en negocio: Media (Alta / Media / Baja)		Tiempo estimado: 1 semanas
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)		Tiempo real: 1 semanas
Descripción: Se crea el diseño físico principal de la base de datos de la aplicación. Se aplica las diferentes formas de normalización.		
Observaciones:		

Tabla 2.4: Scrum - Historia de usuario 2

núm. 3	Nombre: Asignar los diferentes almacenes de datos	
Usuario: Desarrollador		
Modificación de historia número:		Iteración asignada: 2
Prioridad en negocio: Media (Alta / Media / Baja)		Tiempo estimado: 1 semana
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)		Tiempo real: 1 semana
Descripción: Un tuning de la base de datos genera ventajas en el rendimiento de la aplicación en bajo nivel, por tanto, la asignación de los diferentes almacenes de datos es un punto a tomar en cuenta.		
Observaciones:		

Tabla 2.5: Scrum - Historia de usuario 3

2.3.3.3 Historia de usuario - Espacio de trabajo para el desarrollo de software

núm. 4	Nombre: Configurar el espacio de trabajo: Oracle Database	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 3	
Prioridad en negocio: Baja (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Configurar los diferentes permisos y paquetes que van a ser usados en la aplicación.		
Observaciones:		

Tabla 2.6: Scrum - Historia de usuario 4

núm. 5	Nombre: Configurar el espacio de trabajo: Oracle APEX	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 3	
Prioridad en negocio: Baja (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Configurar los esquemas de utilización y espacio de trabajo de APEX.		
Observaciones:		

Tabla 2.7: Scrum - Historia de usuario 5

núm. 6	Nombre: Configurar el espacio de trabajo: API Google Maps	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 3	
Prioridad en negocio: Baja (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Configurar el usuario y las claves de acceso de Google Maps Api para la aplicación.		
Observaciones:		

Tabla 2.8: Scrum - Historia de usuario 6

núm. 7	Nombre: Configurar el espacio de trabajo: Procesos PLSQL	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 3	
Prioridad en negocio: Baja (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Configurar los procesos PLSQL para administración de triggers, secuencias y DML de tablas.		
Observaciones:		

Tabla 2.9: Scrum - Historia de usuario 7

2.3.3.4 Historias de usuario - Desarrollo del software

núm. 8	Nombre: Realizar el mantenimiento de comunidad	
Usuario: Desarrollador		
Modificación de historia número:		Iteración asignada: 4
Prioridad en negocio: Baja (Alta / Media / Baja)		Tiempo estimado: 1 semana
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)		Tiempo real: 1 semana
Descripción: Aplicación web, el mantenimiento para datos de comunidades.		
Observaciones:		

Tabla 2.10: Scrum - Historia de usuario 8

núm. 9	Nombre: Realizar el mantenimiento de Evento	
Usuario: Desarrollador		
Modificación de historia número:		Iteración asignada: 4
Prioridad en negocio: Baja (Alta / Media / Baja)		Tiempo estimado: 1 semana
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)		Tiempo real: 1 semana
Descripción: Realizar la aplicación- mantenimiento para datos de eventos.		
Observaciones:		

Tabla 2.11: Scrum - Historia de usuario 9

núm. 10	Nombre: Realizar el mantenimiento de Modelo	
Usuario: Desarrollador		
Modificación de historia número:		Iteración asignada: 4
Prioridad en negocio: Baja (Alta / Media / Baja)		Tiempo estimado: 1 semana
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)		Tiempo real: 1 semana
Descripción: Realizar la aplicación- mantenimiento para datos de los modelos de estudio.		
Observaciones:		

Tabla 2.12: Scrum - Historia de usuario 10

núm. 11	Nombre: Realizar el mantenimiento de Punto de análisis	
Usuario: Desarrollador		
Modificación de historia número:		Iteración asignada: 4
Prioridad en negocio: Baja (Alta / Media / Baja)		Tiempo estimado: 1 semana
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)		Tiempo real: 1 semana
Descripción: Realizar la aplicación- mantenimiento para datos de los puntos de análisis de un modelo de estudio.		
Observaciones:		

Tabla 2.13: Scrum - Historia de usuario 11

núm. 12	Nombre: Realizar el mantenimiento de Punto de análisis	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 4	
Prioridad en negocio: Baja (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Realizar la aplicación- el mantenimiento para datos de los puntos de análisis de un modelo de estudio.		
Observaciones:		

Tabla 2.14: Scrum - Historia de usuario 12

núm. 13	Nombre: Realizar el mantenimiento de Opciones de usuario	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 4	
Prioridad en negocio: Baja (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Realizar la aplicación- mantenimiento para datos de las opciones de usuario en la interfaz final.		
Observaciones:		

Tabla 2.15: Scrum - Historia de usuario 13

núm. 14	Nombre: Realizar el mantenimiento de Información comunidad	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 4	
Prioridad en negocio: Baja (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Realizar la aplicación- mantenimiento de la información pertinente a una comunidad, la cual consta de la unión entre comunidad, modelo, evento y punto de análisis.		
Observaciones:		

Tabla 2.16: Scrum - Historia de usuario 14

núm. 15	Nombre: Realizar el mantenimiento de Multimedia comunidad	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 4	
Prioridad en Negocio: Baja (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Realizar la aplicación- mantenimiento de la información multimedia de las comunidades como: imágenes, audio y video.		
Observaciones:		

Tabla 2.17: Scrum - Historia de usuario 15

núm. 16	Nombre: Realizar procesos de base de datos en general	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 4	
Prioridad en negocio: Baja (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Para que algunos mantenimientos puedan trabajar de una manera correcta es necesario lanzar procesos almacenados que se ligen a los datos que se tratan por ejemplo a que colección pertenece.		
Observaciones:		

Tabla 2.18: Scrum - Historia de usuario 16

núm. 17	Nombre: Realizar procesos de minería de texto	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 4	
Prioridad en negocio: Naja (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Realizar los procesos y el diseño de la base de datos para la minería de texto en la aplicación.		
Observaciones:		

Tabla 2.19: Scrum - Historia de usuario 17

núm. 18	Nombre: Realizar reportes de administrador	
Usuario: Desarrollador		
Modificación de Historia Número:	Iteración asignada: 4	
Prioridad en negocio: Baja (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: La minería de texto representa información acerca de los datos que se encuentran en la base de datos, por tanto, presentarlos de manera interpretativo es el enfoque de los reportes de minería para el administrador.		
Observaciones:		

Tabla 2.20: Scrum - Historia de usuario 18

núm. 19	Nombre: Configurar el espacio de trabajo Google Charts API	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 4	
Prioridad en negocio: Baja (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Bajo (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Configurar el usuario y las claves de acceso de Google Maps Api para la aplicación.		
Observaciones:		

Tabla 2.21: Scrum - Historia de usuario 19

núm. 20	Nombre: Crear visualización de cartografía etnolingüística	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 4	
Prioridad en negocio: Media (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Media alta (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: La presentación de información para el usuario final (visitante) es fundamental por tanto realizar la presentación de la información de las localidades.		
Observaciones:		

Tabla 2.22: Scrum - Historia de usuario 20

núm. 21	Nombre: Crear procesos adecuados de minería de texto para detección de fonemas.	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 4	
Prioridad en negocio: Media (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Alto (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Realizar minería de texto para la detección de fonemas en un punto análisis seleccionado.		
Observaciones:		

Tabla 2.23: Scrum - Historia de usuario 21

núm. 22	Nombre: Crear procesos adecuados de minería de texto para selección de idioma	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 4	
Prioridad en negocio: Media (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Alta (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Realizar minería de texto para la detección de idioma aplicado en un punto de análisis seleccionado.		
Observaciones:		

Tabla 2.24: Scrum - Historia de usuario 22

núm. 23	Nombre: Crear proceso para análisis de estado de la lengua	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 4	
Prioridad en negocio: Media (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Media (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Crear procesos para análisis de estado de lengua (son análisis manuales).		
Observaciones:		

Tabla 2.25: Scrum - Historia de usuario 23

2.3.3.5 Historias de usuario - Pruebas del software

núm. 24	Nombre: Realizar pruebas usuario final- coherencia	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 5	
Prioridad en negocio: Media (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Media (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Realizar pruebas de coherencia de los datos en la aplicación.		
Observaciones:		

Tabla 2.26: Scrum - Historia de usuario 24

núm. 25	Nombre: Realizar pruebas usuario final- integridad	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 5	
Prioridad en negocio: Media (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Media (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Realizar pruebas de integridad de los datos en la aplicación.		
Observaciones:		

Tabla 2.27: Scrum - Historia de usuario 25

núm. 26	Nombre: Realizar pruebas usuario final- presentación	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 5	
Prioridad en negocio: Media (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Media (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Realizar pruebas de presentación de los datos en la aplicación.		
Observaciones:		

Tabla 2.28: Scrum - Historia de usuario 26

núm. 27	Nombre: Realizar retroalimentación de las pruebas	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 5	
Prioridad en negocio: Media (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Media (Alto / Medio / Bajo)	Tiempo real: 1 semana	
Descripción: Buscar cambios en alguna retroalimentación dadas por las pruebas.		
Observaciones:		

Tabla 2.29: Scrum - Historia de usuario 27

núm. 28	Nombre: Implementar cambios de resultados de pruebas	
Usuario: Desarrollador		
Modificación de historia número:	Iteración asignada: 5	
Prioridad en negocio: Media (Alta / Media / Baja)	Tiempo estimado: 1 semana	
Riesgo en desarrollo: Media (Alto / Medio / Bajo)	Tiempo real: 1 semana	

Descripción: Realizar los cambios solicitadores en las pruebas de aplicación.
Observaciones:

Tabla 2.30: Scrum - Historia de usuario 28

2.3.4 Sprint y listado del producto

En el listado del producto presentado mediante la tabla 2.31 se indica todos los objetivos a cumplir segmentados por cada sprint, esta información permite presentar de una manera simple el proceso de desarrollo del software.

ID	Descripción
1. Sistematizar la información acerca de aplicaciones web de geomática y oracle text	
1	Buscar aplicaciones basadas en geomatica
2	Buscar aplicaciones basadas en minería de texto
3	Buscar aplicaciones basadas Google APIs
4	Probar herramientas de desarrollo bases de datos y minería de texto
5	Probar herramienta Google Maps APIs
2. Especificar los requerimientos de la aplicación. (SRS)	
6	Especificar introducción y vision global
7	Especificar los requisitos específicos
8	Especificar los diagramas de casos de uso
3. Historia de usuario: Diseñar y estandarizar la base de datos e interfaces adecuadas - erwin.	
9	Crear y normalizar la base de datos
10	Asignar los diferentes almacenes de datos
4. Desarrollo de Historias de usuario: Administración (Desarrollo del sistema)	
11	Configurar el espacio de trabajo: Oracle Database
12	Configurar el espacio de trabajo: Oracle Application Express (APEX)
13	Configurar el espacio de trabajo: API Google Maps
14	Configurar el espacio de trabajo: Adaptar procesos PL SQL
15	Realizar el mantenimiento de Localidad
16	Realizar el mantenimiento de Evento
17	Realizar el mantenimiento de Modelo
18	Realizar el mantenimiento de Punto de análisis
19	Realizar el mantenimiento de Opciones de usuario
20	Realizar el mantenimiento de Información localidad
21	Realizar el mantenimiento de Multimedia localidad
22	Realizar procesos de base de datos en general
23	Realizar procesos de minería de texto
24	Realizar reportes de administrador
5. Historias de usuario: Usuario final (Desarrollo del sistema)	
25	Configurar el espacio de trabajo: Google Charts API
26	Realizar la visualización de la cartografía etnolingüística
27	Crear procesos adecuados de minería de texto para detección de fonemas
28	Crear procesos adecuados de minería de texto para selección de idioma
29	Crear proceso para análisis de estado de la lengua
6. Aplicar pruebas de coherencia, integridad y presentación de los datos.	
30	Realizar pruebas usuario final- coherencia
31	Realizar pruebas usuario final- integridad
32	Realizar pruebas usuario final- presentación
33	Realizar retroalimentación de las pruebas
34	Implementar cambios de resultados de pruebas

Tabla 2.31: Sprint y listado del producto

2.3.5 Planificación del Scrum

La planificación del Scrum indica como ejecutar el desarrollo del software tanto en los procesos como en las personas y su ejecución en el proyecto.

2.3.5.1 Sprint existentes en el proyecto

En el listado del producto se presenta un total de 6 Sprint:

1. Sistematizar la información acerca de aplicaciones web de geomántica y minería de textos.
2. Especificar los requerimientos de la aplicación.
3. Diseñar y estandarizar la base de datos e interfaces adecuadas.
4. Desarrollo del sistema – Administración
5. Desarrollo del sistema – Usuario final
6. Aplicar pruebas de integridad y presentación de los datos.

2.3.5.2 Listado de productos en cada sprint y responsables

Una vez generado el listado del producto, mediante la tabla 2.33 se propone los responsables de cada sprint o de cada historia de usuario a desarrollar, el responsable es capaz de generar los entregables del sprint.

Para simplificar la asignación de los responsables de cada historia de usuario y su respectivo sprint, se detalla la tabla 2.32 con los responsables sus acrónimos y cargos en el proyecto de desarrollo.

Nombre	Acrónimo	Cargo
Verdugo Priscila, Mgst	PV	Director del proyecto y cliente.
Astudillo Catalina, Mgst	CA	Director de desarrollo.
Orellana Marcos, Mgst	MO	Director de Minería de texto.
Lima Fernando	FL	Desarrollador.

Tabla 2.32: Listado de responsables

ID	Descripción	Estimación (Semanas)	Responsable (s)
1. Sistematizar la información acerca de aplicaciones web de geomática y oracle text			
1	Buscar aplicaciones basadas en geomática	1	FL
2	Buscar aplicaciones basadas en minería de texto	1	FL
3	Buscar aplicaciones basadas Google APIs	1	FL
4	Probar herramientas de desarrollo bases de datos y minería de texto	1	FL
5	Probar herramienta Google Maps APIs	1	FL
2. Especificar los requerimientos de la aplicación (SRS)			
6	Especificar introducción y visión global	0,6	FL,CA
7	Especificar los requisitos específicos	0,6	FL,CA
8	Especificar los diagramas de casos de uso	0,6	FL,CA
3. Historia de usuario: Diseñar y estandarizar la base de datos e interfaces adecuadas - erwin.			
9	Crear y normalizar la base de datos	0,5	FL
10	Asignar los diferentes almacenes de datos	0,5	FL
4. Desarrollo de historias de usuario: Desarrollo del software - Administrador			
11	Configurar el espacio de trabajo: Oracle Database	0,2	FL
12	Configurar el espacio de trabajo: Oracle Application Express (APEX)	0,2	FL
13	Configurar el espacio de trabajo: API Google Maps	0,2	FL
14	Configurar el espacio de trabajo: Adaptar procesos PLSQL	0,2	FL
15	Realizar el mantenimiento de Localidad	0,5	FL,CA
16	Realizar el mantenimiento de Evento	0,5	FL,CA
17	Realizar el mantenimiento de Modelo	0,5	FL,CA
18	Realizar el mantenimiento de Punto de análisis	0,5	FL,CA
19	Realizar el mantenimiento de Opciones de usuario	0,5	FL,CA
20	Realizar el mantenimiento de Información localidad	0,5	FL,CA
21	Realizar el mantenimiento de Multimedia localidad	1	FL,CA
22	Realizar procesos de base de datos en general	1	FL,CA
23	Realizar procesos de minería de texto	1	FL,MO
24	Realizar reportes de administrador	1	FL,CA
5. Desarrollo de historias de usuario: Desarrollo del software - Usuario visitante			
25	Configurar el espacio de trabajo: Google Charts API	0,5	FL
26	Realizar la visualización de la cartografía etnolingüística	0,5	FL
27	Crear procesos adecuados de minería de texto para detección de fonemas	0,5	FL,MO
28	Crear procesos adecuados de minería de texto para selección de idioma	2	FL,MO
29	Crear proceso para análisis de estado de la lengua	1	FL,CA
6. Aplicar pruebas de integridad y presentación de los datos.			
30	Realizar pruebas usuario final- coherencia	1	FL,CA,PV
31	Realizar pruebas usuario final- integridad	1	FL,CA,PV
32	Realizar pruebas usuario final- presentación	1	FL,CA,PV
33	Realizar retroalimentación de las pruebas	1	FL,CA
34	Implementar cambios de resultados de pruebas	1	FL,CA

Tabla 2.33: Listado del producto y responsables

2.3.5.3 Objetivo de cada sprint

Luego de la revisión y aprobación de cada Sprint, se proponen los entregables al final de cada ciclo de la siguiente forma:

1. Sistematizar la información acerca de aplicaciones web de geomática y minería de textos.
Entregable: documentación de indagación, y breve resumen de herramientas o procedimientos para desarrollo del software.

2. Especificar los requerimientos de la aplicación.
Entregable: Especificación de requisitos del software.

3. Diseñar y estandarizar la base de datos e interfaces adecuadas.
Entregable: Modelo entidad/relación, modelo de capas y modelo del servicio web.

4. Desarrollo del sistema – Administración
Entregable: Aprobación de avances del software.

5. Desarrollo del sistema – Usuario final
Entregable: Aprobación de avances del software.

6. Aplicar pruebas de integridad y presentación de los datos.
Entregable: Documentación de pruebas y cambios aplicados en el software.

2.3.5.4 Planificar las reuniones para revisión del sprint

Las reuniones de planeación y revisión se efectúan según la tabla 2.34, aquí se analizan los avances hasta el momento, se recibe la aprobación o sugerencias del cliente y del director del proyecto.

Día	Hora	Planificación
Lunes	16:00	Revisión de avance.
Miércoles	16:00	Planificación tareas próximas.

Tabla 2.34: Planificación de reuniones de revisión

2.3.5.5 Retrospectiva del sprint

Cada reunión de revisión del Sprint genera observaciones que son aplicadas y consideradas en las revisiones consecuentes, estas observaciones pueden ser en las interfaces del usuario o en los procedimientos de desarrollo.

2.4 Conclusiones del capítulo

Al finalizar este capítulo se puede acotar que cada uno de los procesos en la planificación del desarrollo del software y saber cómo ejecutarlos, el uso del ERS permite identificar de forma clara los requisitos del software a desarrollar, estos requisitos se definen tanto por la parte funcional (las opciones del software) como por la parte técnica para el desarrollador (que lenguaje de programación se va a utilizar), estas especificaciones funcionales permiten al desarrollador indagar en métodos que posiblemente fueron usados en software de características similares.

Por otro lado, el marco de desarrollo Scrum, llamado metodología de desarrollo ágil, permite identificar de forma sencilla cuando ejecutar cada proceso de la planificación de forma organizada y que personaje interviene en dicho proceso para obtener el mejor rendimiento en la tarea de desarrollo del software.

Capítulo 3: Diseño de la aplicación web

Una vez seleccionadas las herramientas, descritos los requisitos del proyecto con la ERS y el marco de trabajo Scrum de para el seguimiento del desarrollo software, se procede a elaborar los diseños principales que servirán de base para al desarrollo: el diseño del servicio web con el clasificador de palabras para el apoyo al desarrollador, el diseño de la base de datos para la estructuración de los datos de la aplicación.

Un diseño lógico extra es necesario pues las aplicaciones desarrolladas en Oracle APEX al ingresar a producción son implementadas bajo una arquitectura de capas de un sistema distribuido, esta implementación garantiza una mayor funcionalidad y fácil mantenimiento de las aplicaciones o servidores, para visualizar el comportamiento entre los componentes del sistema distribuido es necesario generar un diseño de capas para la aplicación web.

3.1 Diseño del servicio web para clasificación de texto

En el apartado 1.6.1 se menciona la clasificación de modelos para minería de texto, se selecciona al Modelo descriptivo como parte de este trabajo debido a la característica del reconocimiento de patrones, en base a esta característica se procede a desarrollar la clasificación de texto en la presente sección.

El apartado 1.6.3 indica que para la clasificación de textos se necesita de un algoritmo de corrección ortográfica, GNU Aspell es una librería de corrección ortográfica que se encuentra implementada en varios lenguajes de programación. Basado en esta librería se va a implementar un servicio web que genere datos en formato XML y pueda ser leído por la aplicación web o cualquier cliente, a continuación, se detalla gráficamente en la ilustración 3.1 un diagrama de clases del servicio web: las entradas, los procedimientos y las salidas del servicio web a implementar.

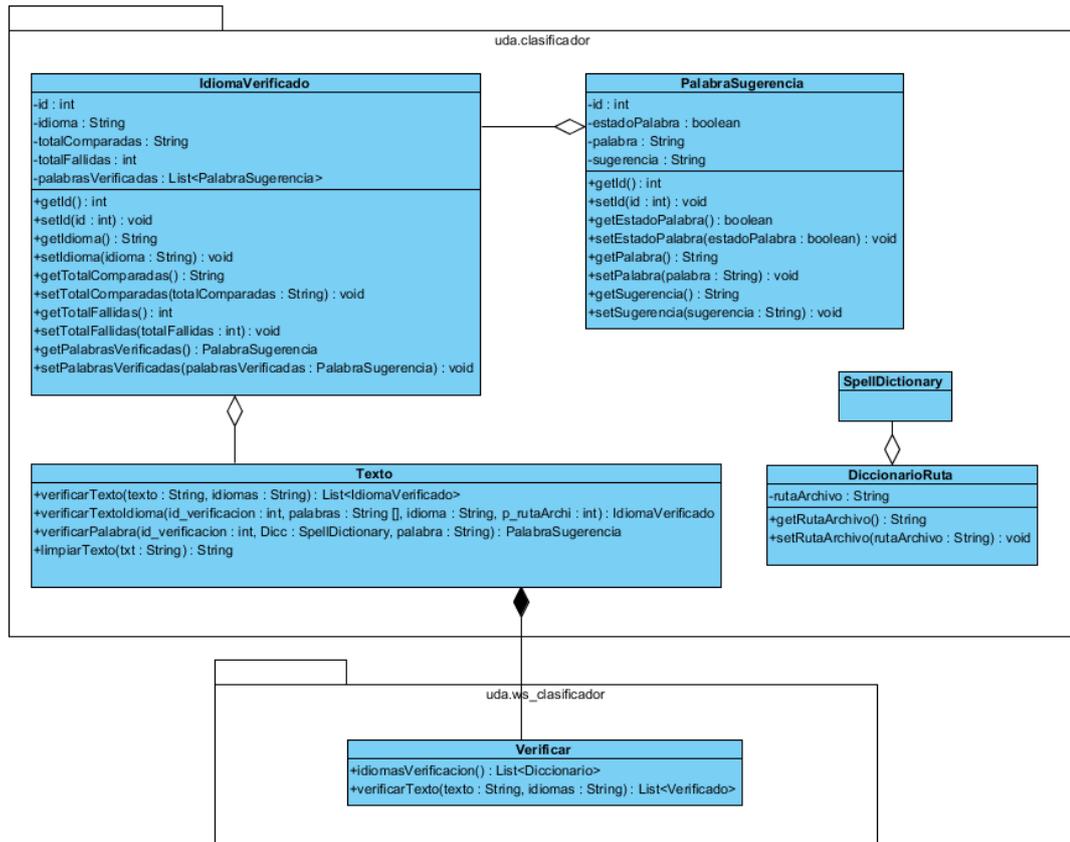


Ilustración 3.1: Servicio web - Diagrama de clases

El diagrama de clases, tiene como parámetros de entrada: el texto y el/los idiomas a verificar, y como salida una lista en la que especifica, si las palabras pertenecen o no a un idioma de una manera estructurada. Con esta clasificación el proceso de cuantificación de las palabras es sencillo de ejecutar mediante sentencias SQL.

3.1.1 Fases de limpieza

Según el autor (Lopez Perez C, 2007) indica que para evitar conflictos en la clasificación de idioma se debe considerar reglas de limpieza, debido a que los resultados que arroje la clasificación pueden no ser los esperados. A continuación, se detalla las fases de limpieza junto a otras características primordiales de programación del servicio web.

3.1.1.1 Análisis de datos

En el análisis de datos se determinan todas las propiedades y errores de la palabra.

- 1) *Data Profiling*: Analizar un grupo de características del dato con el fin de obtener propiedades de la misma.

En el presente proyecto, esta fase se ve reflejada con la búsqueda de patrones en las propiedades de cada palabra, por ejemplo: el pluralismo, singularidad, género o acentuación.

- 2) *Data mining*: Identificar los datos en un conjunto de datos.

En el presente proyecto, esta fase se define mediante la comparación de cada una de las palabras con el conjunto de palabras del diccionario presentado en la ilustración 3.2, esta fase junto a la anterior se ejecuta mediante el algoritmo GNU ASPELL.

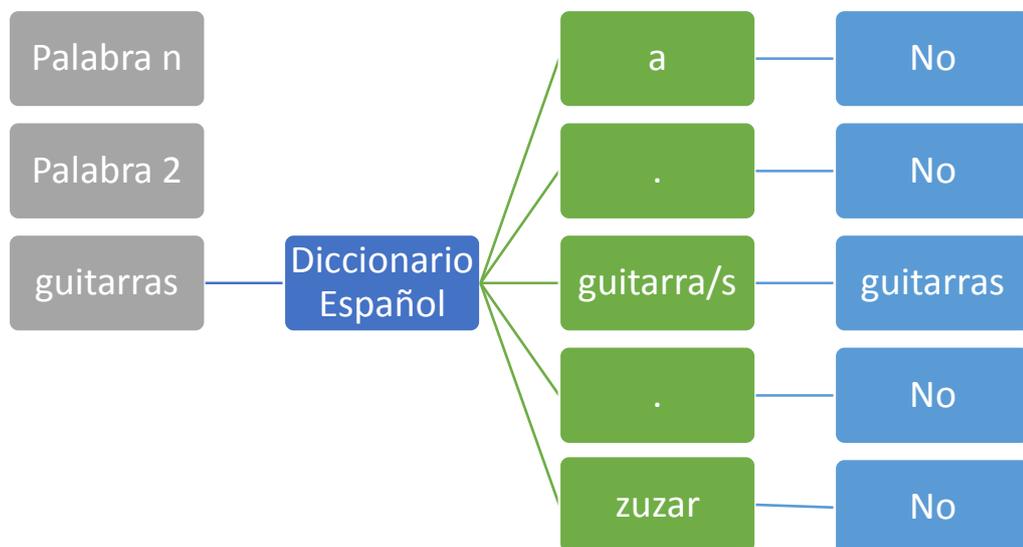


Ilustración 3.2: Identificación del dato mediante un conjunto de datos

3.1.1.2 Definición de transformaciones de datos y reglas de mapeo

Es el proceso de división de una frase a un grupo de palabras para posteriormente aplicar procesos individualmente.

Con el fin de facilitar la minería de cada una de las palabras, en la etapa de transformación de datos, se debe considerar reglas que permiten una transformación y ubicación de cada palabra en el texto (Lopez Perez C, 2007).

A continuación, se lista las reglas principales en la transformación de datos junto con un ejemplo de transformación de texto mediante la ilustración 3.3.

- Eliminar duplicados de una palabra.
- Convertir textos a minúsculas, debido a que se puede omitir en la sensibilidad del clasificador.
- Eliminar caracteres especiales.
- Eliminar caracteres numéricos.
- Dividir el texto resultante mediante el carácter especial "espacio" y almacenar en un arreglo.

Estas reglas son aplicadas mediante las expresiones regulares de los diferentes lenguajes de programación.

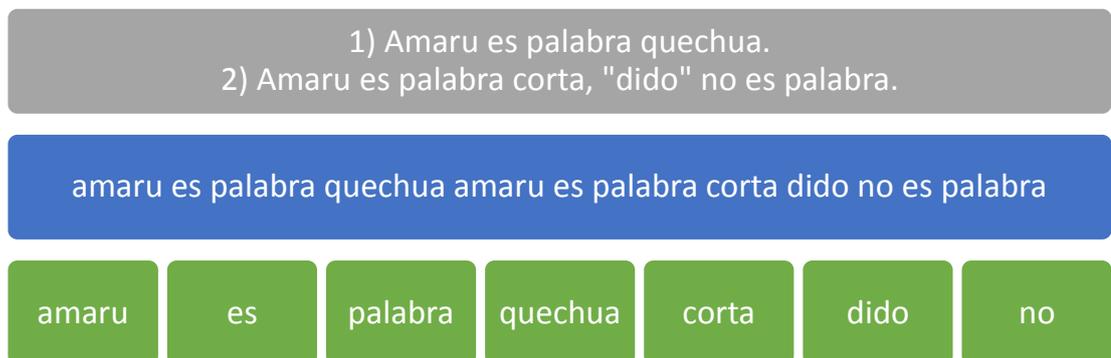


Ilustración 3.3: Transformación de texto

3.1.2 Aplicación de minería de texto

El análisis de la clasificación de los textos en la minería implica examinar el proceso en general para asegurar que la clasificación se cumpla adecuadamente. El posterior almacenamiento de la clasificación adecuada permite manipular la información de manera estructurada, la ilustración 3.4 se describe un ejemplo del comportamiento interno de la clasificación de textos.

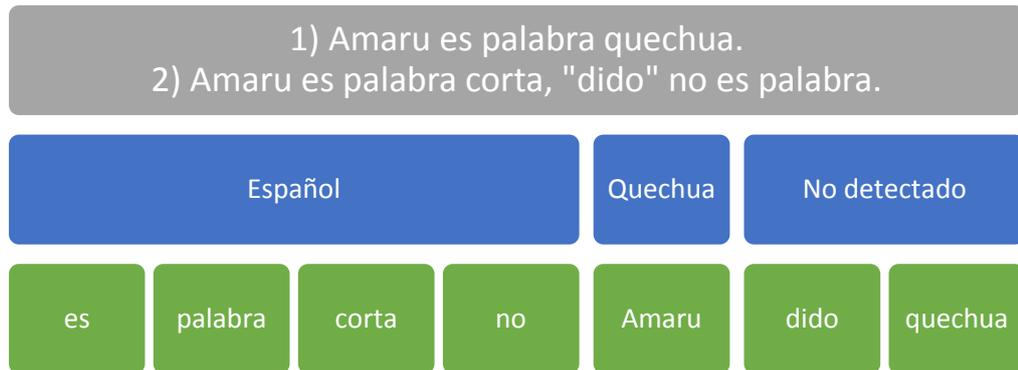


Ilustración 3.4: Clasificación lógica de texto

El comportamiento de la minería en la aplicación web es presentado mediante la ilustración 3.5, detalla los puntos importantes en la clasificación.

- El software discrimina palabras con longitud menor a tres caracteres, catalogados como conectores, los cuales no se consideran para la minería de texto.
- Se presenta las diferentes variantes en fonemas para las palabras que no han logrado ser clasificadas.
- Palabras cuya clasificación se encuentra en múltiples diccionarios por el software este brinda una alerta para ser clasificadas manualmente y recordar su clasificación para futuras incidencias.
- Finalmente, todas las palabras pueden ser clasificadas manualmente, considerado como entrenamiento para el software.

#	Idioma	Ver	Palabra	Idioma	Estadísticas	Fonemas y variantes
1	Español	<input checked="" type="checkbox"/>	amaru	Quechua_Peru	Si	
2	Quechua_Peru	<input checked="" type="checkbox"/>	corta	Español	Si	
3	No Detectado	<input checked="" type="checkbox"/>	dido	No Detectado	Si	Español : [nido,sido,mido,pido,fido,dado,dedo,dudo,dino,dilo,dimo] Quechua_Peru : []
1 - 3						
			es	Español	No	
			no	Español	No	
			palabra	Español	Si	
			quechua	No Detectado	No	Doble detección

Ilustración 3.5: Minería de texto en la aplicación web

Una vez completada la clasificación de los diferentes textos se almacenan de manera estructurada, esto que permite ejecutar diferentes tipos de cuantificación mediante sentencias SQL, la ilustración 3.6 presenta ejemplos de presentación de información de la minería de texto mediante gráficas estadísticas.



Ilustración 3.6: Cuantificación de textos

3.2 Diseño y normalización de la base de datos

Uno de los primeros pasos en el desarrollo de una aplicación ya sea de escritorio u orientada a la web es generar un diseño de la base de datos, en este caso como el SGDB es relacional es necesario el uso de un modelo entidad/relación, en cual se presenta en la ilustración 3.7, cabe recalcar que este modelo entidad/relación debe ser normalizado para obtener ventajas como: evitar la redundancia de los datos, evitar problemas en la actualización de los datos y los más importante, proteger la integridad de los datos de la aplicación.

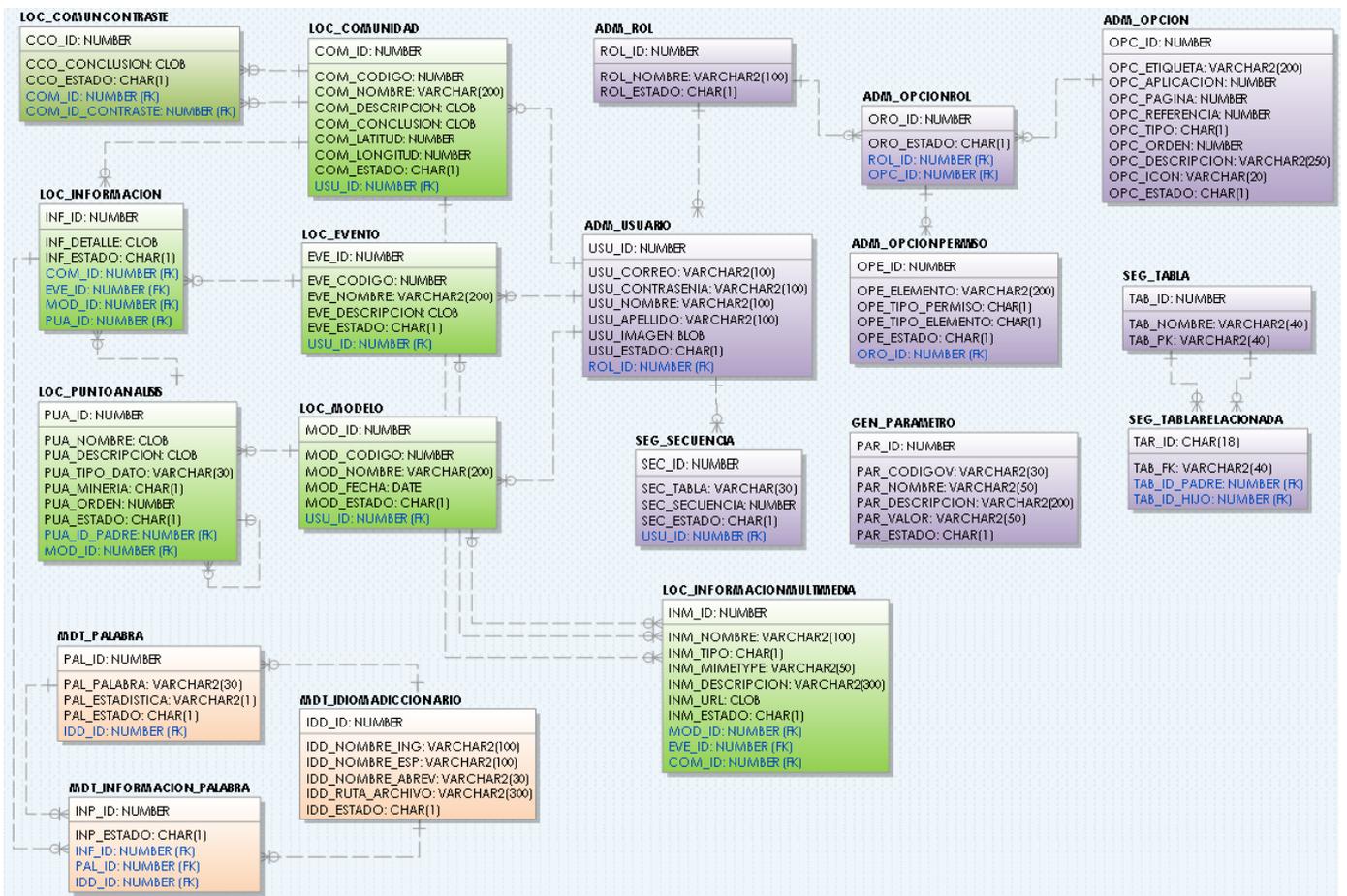


Ilustración 3.7: Modelo Entidad-Relación

En el apartado 1.2 de este documento se propone un tuning de la base de datos, el objetivo de este es mejorar los tiempos de respuesta del servidor mediante la adaptación de las estructuras de memoria. Para llevar a cabo esto es necesario: contar con tablespaces adecuados presentado en la ilustración 3.8 y dar la asignación de un porcentaje de memoria libre o *PCT free* en ellos como se observa en la ilustración 3.9, el software dedicado al modelado entidad/relación ya permite ejecutar este tipo tareas.

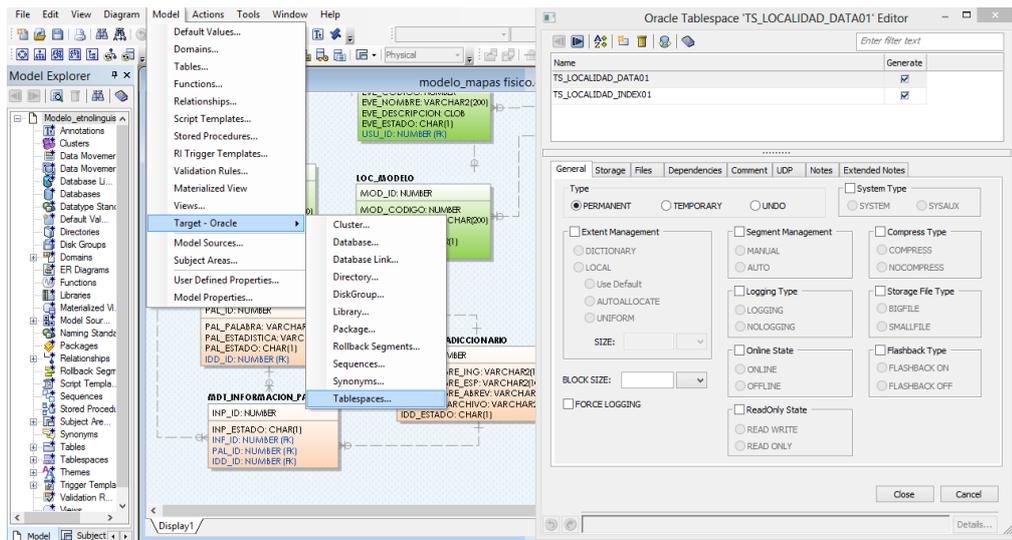


Ilustración 3.8: Tuning de base de datos - Tablespaces

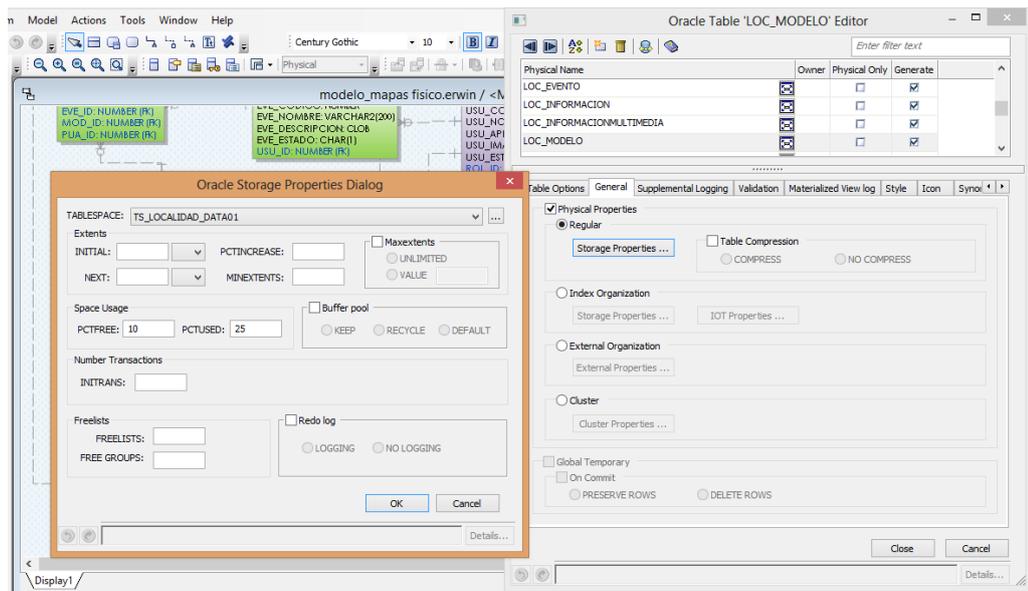


Ilustración 3.9: Tuning de base de datos- Asignación de memoria

3.3 Diseño del modelo Cliente-servidor

“El modelo Cliente/Servidor es un modelo de aplicación distribuida en el que las tareas se reparten entre los proveedores de recursos o servicios, llamados servidores, y los demandantes, llamados clientes” (Marini, 2012).

El funcionamiento es muy sencillo, pues un cliente realiza peticiones y luego recibe respuestas en base a su petición, el servidor procesa las peticiones de los clientes y genera respuestas (Departamento de Informática, 2016).

Según la tarea que un servidor ejecute para un cliente define su tipo, en este proyecto se detalla los tipos de servidores de uso junto con su función.

- Servidor de base de datos: las aplicaciones de los clientes envían consultas SQL al servidor y este retorna los datos resultantes de dicha sentencia.
- Servidor de aplicaciones: trabajan en base a peticiones a la World Wide Web, el cliente genera peticiones mediante el protocolo HTTP y este devuelve documentos.

Los bloques básicos en una arquitectura cliente/servidor (C/S) son: cliente, middleware y servidor.

- Cliente: Sistema operativo y navegador web con interfaz de usuario.
- Middleware: software que interactúa entre clientes y servidores.
- Servidor: ejecuta el software solicitado.

El modelo C/S de tres capas, estas definiciones están en detalle en el documento propuesto por (Departamento de Informática, 2016), indica que esta arquitectura añade una capa intermedia. El cliente implementa la lógica de presentación, los servidores de aplicaciones y de base de datos implementan la lógica del negocio, a continuación, se presenta en la ilustración 3.10 el sistema distribuido de este proyecto.



Ilustración 3.10: Modelo de tres capas

3.4 Conclusiones del capítulo

Al finalizar este capítulo, se cuenta con los diseños necesarios para la etapa de configuración del entorno de desarrollo y de codificación de la aplicación web. El diseño de la base de datos permite visualizar a breves rasgos el comportamiento del sistema y a la vez permite generar el código para crear los diferentes almacenes de datos. El diagrama de clases del servicio web permite visualizar su funcionamiento lógico y permite al programador preparar los códigos fuentes. El modelo de capas permite segmentar el sistema en partes totalmente independientes y a la vez visualizar su funcionamiento en cooperación. Estos modelos, diseños y diagramas son la base fundamental para proceder a un desarrollo del software definido.

Capítulo 4: Software Cartografía etnolingüística y análisis de lengua.

Este capítulo se enfoca en dos partes: la primera parte, la adecuación de los diferentes entornos de trabajo bajo un único sistema operativo para que el desarrollo del software se de en una manera efectiva, la segunda parte, destinada al desarrollo propio de la aplicación web "Cartografía etnolingüística" en la que se proponen prácticas para su codificación, junto con la materialización de algunos de los componentes visuales, componentes que han sido tomados desde el apartado 2.1.3 "Especificación de requisitos de software" ambas partes para generar un software de calidad.

4.1 Configuración del espacio de trabajo

Es necesario contar con la configuración adecuada de los diferentes entornos de desarrollo que apoyen a generar un software de calidad, a continuación, se presenta de una manera secuencial la configuración de entorno para el presente proyecto.

4.1.1 Configuración para modelo de tres capas con Glassfish

El servidor de aplicaciones Glassfish es un software de tipo middleware un software que interactúa entre clientes y servidores mejorando el comportamiento de un sistema distribuido, basado en el lenguaje de programación java y de propiedad de la Corporación Oracle, existe también su versión en código abierto para la comunidad (Glassfish, 2016).

Glassfish en este proyecto trabaja como contenedor común para ORDS (Oracle Rest Data Services) y para el servicio web clasificador de textos. Las configuraciones de espacio de trabajo son necesarias y propuestas en dos instancias: la configuración propia del servidor de aplicaciones Glassfish y consecuentemente la configuración de ORDS.

4.1.1.1 Configuración del servidor de aplicaciones Glassfish

La instalación y configuración de Glassfish es simple, como se detalla en el sitio de documentación oficial de Glassfish (Glassfish, 2016).

- 1) Prerrequisito: Instalación de Java JDK (Anexo 3)
- 2) Descargar el archivo zip de Glassfish mediante el origen de descarga "wget".

- 3) Crear el usuario y otorgar permisos para Glassfish

- `groupadd Glassfish`
- `useradd -s /bin/bash -g glassfish Glassfish`
- `chown -R glassfish: /opt/glassfish4/`
- `passwd Glassfish`

- 4) Generar un script de control de Glassfish

- `sudo nano /etc/init.d/glassfish`

```
GLASSFISH_HOME=/opt/glassfish4/glassfish/  
GLASSFISH_USER=Glassfish  
case $1 in  
start)  
    su $GLASSFISH_USER -c "$GLASSFISH_HOME/bin/asadmin  
start-domain domain1"  
    ;;  
stop)  
    su $GLASSFISH_USER -c "$GLASSFISH_HOME/bin/asadmin  
stop-domain domain1"  
    ;;  
restart)  
    su $GLASSFISH_USER -c "$GLASSFISH_HOME/bin/asadmin  
stop-domain domain1"  
    su $GLASSFISH_USER -c "$GLASSFISH_HOME/bin/asadmin  
start-domain domain1"  
    ;;  
esac  
exit 0
```

- 5) Asignar permisos de ejecución al archivo

- `chmod +x /etc/init.d/Glassfish`

- 6) Arrancar el servidor de aplicaciones

- `/etc/init.d/glassfish start`

- 7) Ingresar al servidor de aplicaciones mediante

- <http://localhost> o ipServidor:4848

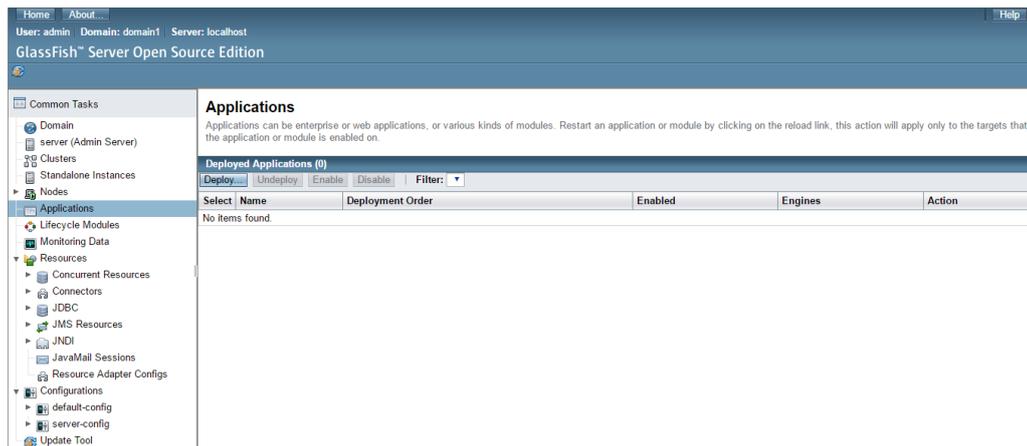


Ilustración 4.1: Servidor de aplicaciones Glassfish

Con esta configuración se puede acceder al servidor de aplicaciones Glassfish como se presenta en la ilustración 4.1, el servidor de aplicaciones funciona como contenedor del servicio web y Oracle ORDS que sirve para conectarse a las aplicaciones desarrolladas en Oracle APEX.

4.1.1.2 Configuración de Oracle Rest Data Services

ORDS (Oracle Rest Data Services) facilita el desarrollo de modernas interfaces REST para datos relacionales en Oracle Database, las APIs RESTful tienen como ventaja el acceso diferente al de los clientes convencionales ya que permite la escritura conformatele de SQL y PL/SQL, además puede ser implementado bajo un servidor de aplicaciones java (Oracle Blogs, 2016).

A continuación, se presenta la configuración de ORDS:

- 1) Prerrequisito: Instalación Oracle Database (Anexo 1)
- 2) Prerrequisito: Instalación de kit de desarrollo de Java (Anexo 3)
- 3) Prerrequisito: Actualización Oracle APEX (Anexo 2)
- 4) Desbloquear los usuarios en Oracle Database.
 - `alter user apex_public_user identified by <contraseña>;`
 - `alter user apex_public_user account unlock;`
 - `alter user apex_listener identified by <contraseña>;`
 - `alter user apex_listener account unlock;`
 - `alter user apex_rest_public_user identified by <contraseña>;`
 - `alter user apex_rest_public_user account unlock;`
 - `alter user ords_public_user identified by <contraseña>;`

- `alter user ords_public_user account unlock;`
- 5) Desactivar el Gateway de acceso local en Oracle Database.
- `exec dbms_xdb.setlistenerlocalaccess(l_access => true);`
 - `exec dbms_xdb.sethttpport(0);`
 - `commit;`
- 6) Descargar el archivo de ORDS de
- <http://www.oracle.com/technetwork/developer-tools/rest-data-services/downloads/index.html>
- 7) Configurar ORDS mediante Java
- `java -jar ords.war`
- 8) Llenar información relevante de la base de datos:
- Directorio temporal para archivos de configuración
 - Nombre o IP del servidor de base de datos
 - Puerto de la base de datos
 - Nombre de la instancia de la base de datos (SID)
 - Contraseña del usuario: `ords_public_user`
 - Contraseña del usuario: `apex_public_user`
 - Contraseña del usuario: `apex_listener`
 - Contraseña del usuario: `apex_rest_public_user`
 - Finalmente, la ruta del directorio de imágenes de APEX
- 9) Generar un archivo `.war` con el directorio de imágenes de apex
- `java -jar ords.war static /u01/app/apex/images/`

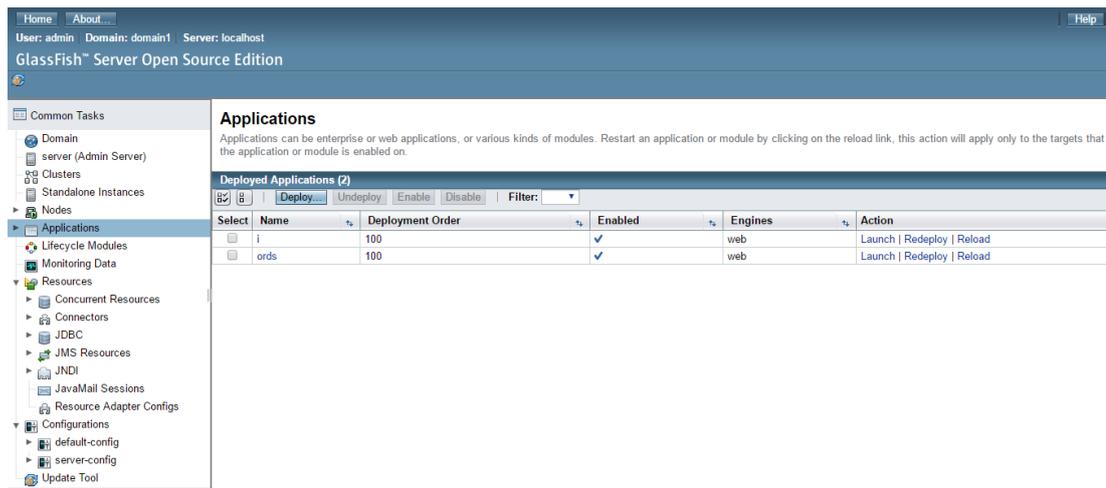


Ilustración 4.2: ORDS sobre servidor de aplicaciones Glassfish

Esta configuración genera dos archivos: ords.war y i.war, los mismos que pueden ser implementados en cualquier servidor de aplicaciones en este caso Glassfish como se puede observar en la ilustración 4.2.

4.1.2 Configuración de Oracle APEX

La configuración del entorno de trabajo en Oracle Apex brinda seguridad tanto para las aplicaciones en desarrollo como para los desarrolladores, esta seguridad se presenta debido a que encapsula bajo un mismo entorno a las aplicaciones y a los usuarios, la configuración de Oracle Apex se encuentra a más detalle en su sitio oficial (Oracle Workspace, 2016).

4.1.2.1 Configuración de Workspace

Una ventaja de Oracle APEX es permitir múltiples entornos de trabajos y cada uno de ellos puede tener varias aplicaciones en desarrollo, para obtener un entorno de trabajo es necesario configurarlo.

Antes de empezar con la configuración de trabajo se debe tener en cuenta algunas características:

- Crear o usar un esquema de la base de datos, un entorno de trabajo necesita estar ligado a un esquema de la base de datos, Oracle APEX permite crear un esquema al momento de crear el entorno de trabajo, pero tiene algunos inconvenientes pues no permite detallar parámetros como los *tablespaces* o las cuotas de almacenamiento.
- El encargado del entorno de trabajo, un entorno de trabajo permite la concurrencia de varios desarrolladores, para seguridad del entorno un usuario debe ser el encargado para administrar el acceso a los desarrolladores en el entorno.

Teniendo en cuenta estas características se procede a la configuración del entorno de trabajo.

- 1) Prerrequisito: Instalación de Oracle Database (Anexo 1).
- 2) Prerrequisito: Actualización de Oracle APEX (Anexo 2).
- 3) Prerrequisito: Configuración del servidor de aplicaciones Glassfish.
- 4) Prerrequisito: Configuración de Oracle Rest Data Services.
- 5) Crear el usuario en la base de datos con los parámetros deseados.

- Créate user <usuario> identified by <contraseña> default tablespace <tablespace> quota 10M on <tablespace> temporary tablespace <tablespace> quota 5M on system password expire;
- 6) Acudir al administrador de servicios de Oracle APEX (ilustración 4.3) cuyo usuario y contraseña fueron definidos en su proceso de actualización (Anexo 2).
- Ir a la URL: http://localhost:8080/ords/apex_admin

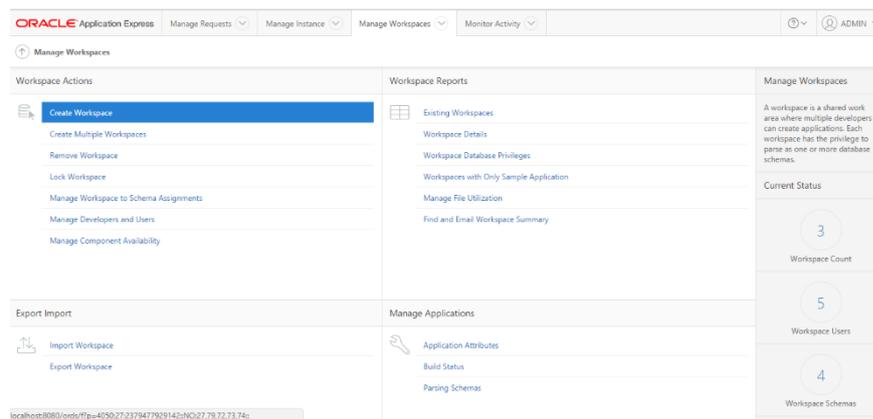


Ilustración 4.3: Administrador de servicios de Oracle APEX

- 7) Brindar los identificadores del entorno de trabajo (ilustración 4.4).

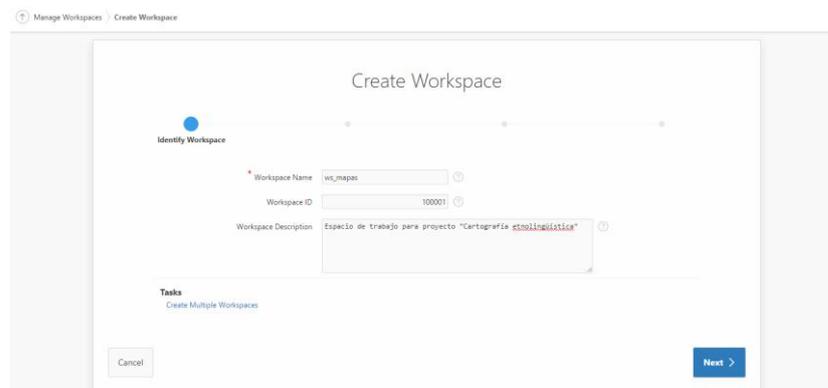


Ilustración 4.4: Crear entorno de trabajo – Identificadores del entorno

- 8) Proporcionar los identificadores del esquema (ilustración 4.5).

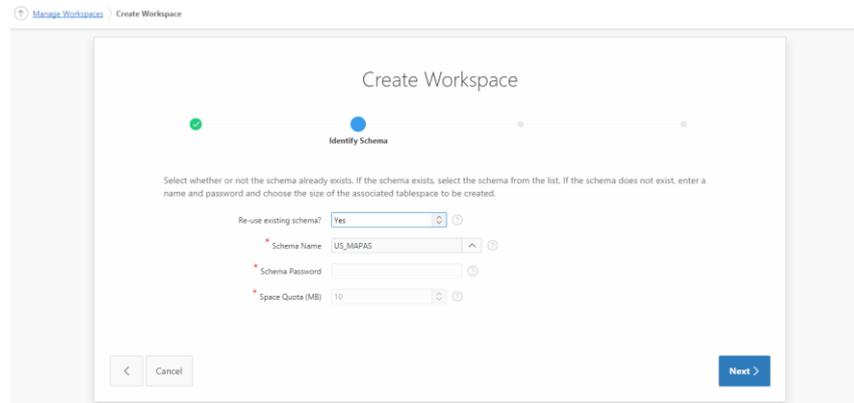


Ilustración 4.5: Crear entorno de trabajo - Identificadores de esquema

9) Identificar al administrador del espacio de trabajo (ilustración 4.6).

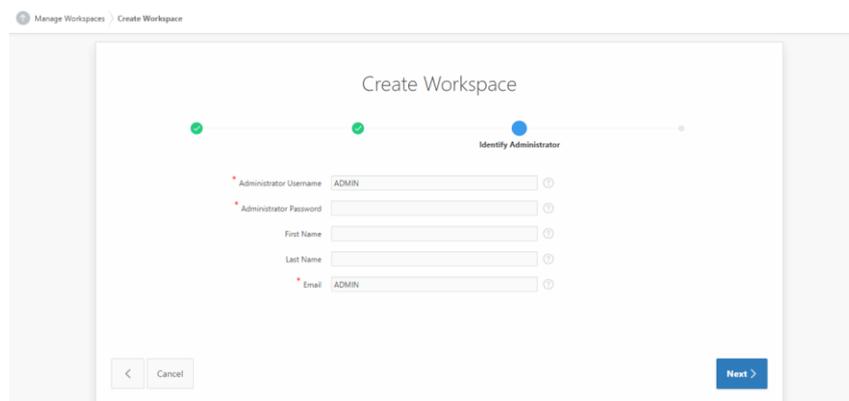


Ilustración 4.6: Crear entorno de trabajo - Identificar el administrador

10) Aceptar las condiciones de creación de entorno de trabajo.

4.1.2.2 Configuración de referencias de servicios web en Oracle APEX

Una vez creado el espacio de trabajo para el desarrollo de la aplicación web, se debe configurar la referencia al servicio web para el clasificador de texto, una ventaja de esto es que Oracle Apex detecta automáticamente la descripción WSDL y genera tablas temporales llamadas también colecciones con los datos de la descripción. La configuración de las referencias son las siguientes:

- 1) Prerrequisito: Crear el entorno de trabajo.
- 2) Ingresar al entorno de trabajo.
- 3) Ingresar a los componentes compartidos del espacio de trabajo.

4) Ingresar a las referencias de servicios web (ilustración 4.7).

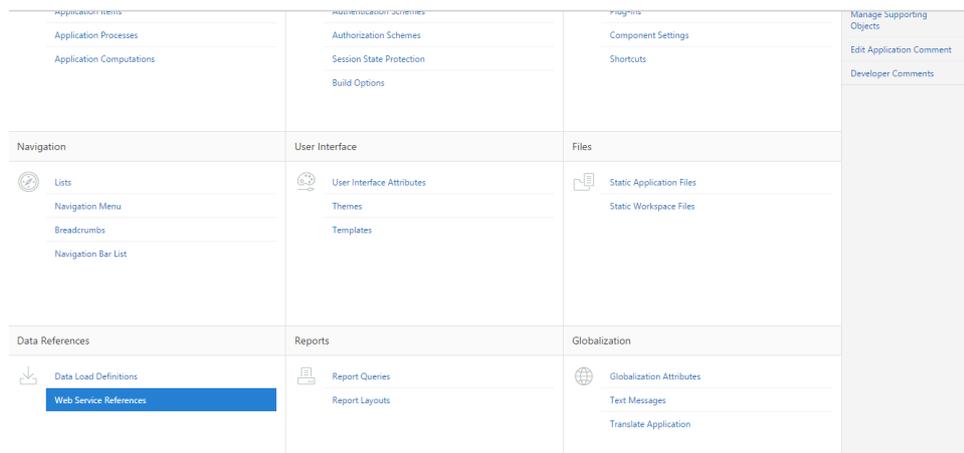


Ilustración 4.7: Referencias de servicios web de Oracle APEX

5) Seleccionar opción: crear y la opción Basado en WSDL (ilustración 4.8)

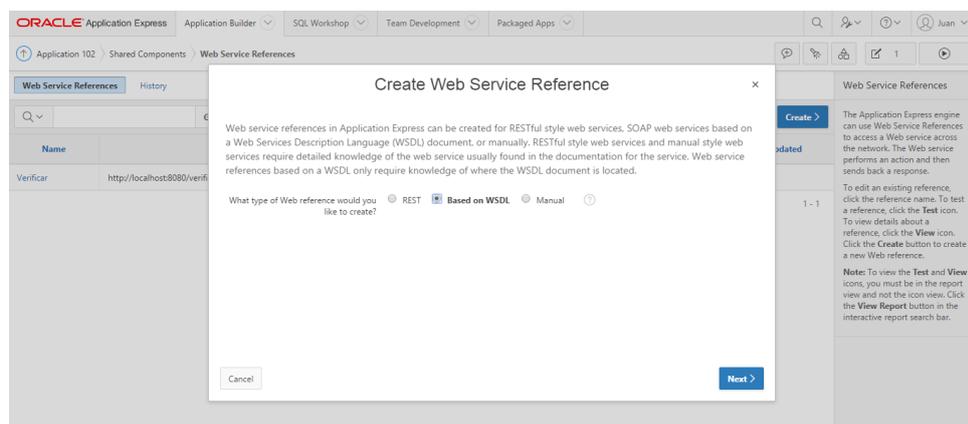


Ilustración 4.8: Crear referencia WSDL

4.2 Aplicación web - Cartografía etnolingüística

A continuación, se presenta el desarrollo de la aplicación web “Cartografía etnolingüística”, se propone algunas prácticas que guían el correcto desarrollo de la aplicación, los elementos generales de la aplicación web, y las instancias propuestas por la especificación de requisitos del software proporcionadas por el cliente.

4.2.1 Convenciones en la programación del software

Adoptar convenciones que faciliten la comprensión del código fuente es un correcto hábito de todo programador, es así que, en este punto desde la tabla 4.1 hasta la tabla 4.4 se redacta algunas convenciones en declaraciones de variables, nombres de tablas en la base de datos y nombres para las páginas web. Estas convenciones están utilizadas en el proyecto de desarrollo Análisis y Desarrollo del módulo de manufactura de un Sistema ERP Manufactura ERP (Paute, 2016).

Convenciones de variables en programación PL/SQL		
Tipo	Variable local	Variable global
Varchar	lv_nombreVariable	gv_nombreVariable
Number	ln_nombreVariable	gn_nombreVariable
Char	lc_nombreVariable	gc_nombreVariable
Blob	lb_nombreVariable	gb_nombreVariable
Date	ld_nombreVariable	gd_nombreVariable
Cursor	lc_nombreVariable	gc_nombreVariable
Otro	"l" + inicial tipo variable + guion bajo (_) + nombre variable.	"g" + inicial tipo variable + guion bajo (_) + nombre variable.

Tabla 4.1: Convenciones para variables en PLSQL

Convenciones en componentes en Oracle Apex	
Elemento	Formato
Elementos de aplicación (Application items)	AI_NOMBRE
Procesos de aplicación (Application processes)	AP_NOMBRE
Listas (Lists)	LI_NOMBRE
Lista de valores (List of values)	LOV_NOMBRE
Procesos (Processes)	PR_NOMBRE
Acciones dinámicas (Dynamic Actions)	DI_NOMBRE

Tabla 4.2: Convenciones para componentes en Oracle APEX

Convenciones en colecciones en Oracle Apex		
	Formato	Ejemplo
Nombre colección	"COL_" + NOMBRE TABLA	COL_MAQUINA
Variable colección	"lv_colPro"+ Nombre Tabla	lv_colProModelo
Procedimiento crear colección	"pr_creaCol" + Nombre Tabla	pr_creaColModelo
Procedimiento actualizar datos colección	"pr_actCol" + Nombre Tabla	pr_actColModelo

Tabla 4.3: Convenciones para colecciones en Oracle APEX

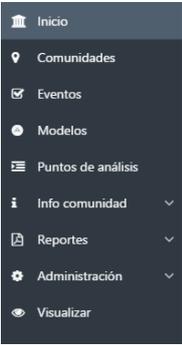
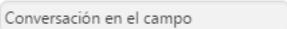
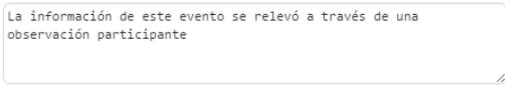
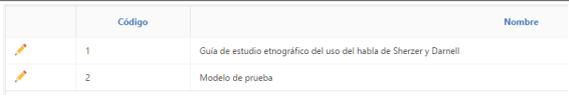
Convenciones en las páginas web en Oracle APEX		
Tipo página	Nombre página	Observación
Reporte	Tres primeras iniciales nombre módulo(loc) + tres primeras iniciales tipo página + nombre tabla	La primera letra de cada palabra en mayúscula a excepción de la primera palabra, debe ir todo en minúsculas. Ejemplo: locRepModelo
Mantenimiento	Tres primeras iniciales nombre módulo(loc) + tres primeras iniciales tipo página + nombre tabla	La primera letra de cada palabra en mayúscula a excepción de la primera palabra, debe ir todo en minúsculas. Ejemplo: locManModelo

Tabla 4.4: Convenciones para páginas web en Oracle APEX

4.2.2 Elementos generales para interfaz de usuario

Es importante describir los elementos generales para la GUI o interfaz de usuario, puesto que uno de los objetivos es que la aplicación web cumpla con la característica de usabilidad, la que tiene como objetivo mantener una buena presentación para el público (*front-end*).

A continuación, mediante la tabla 4.5 se describen los elementos de interfaz de usuario.

Elemento	Imagen
Interfaz	
Cabecera	
Barra de menú principal	
Botones	
Buscar	
Agregar	
Deshacer	
Guardar	
Eliminar	
Exportar a archivo pdf	
Enlaces	
Editar	
Elementos formulario	
Etiqueta	Código
Campo de texto	
Área de texto	
Casilla de verificación	<input checked="" type="checkbox"/>
Botón de radio	<input type="radio"/>
Lista de opciones	
Cuadrícula detalle	

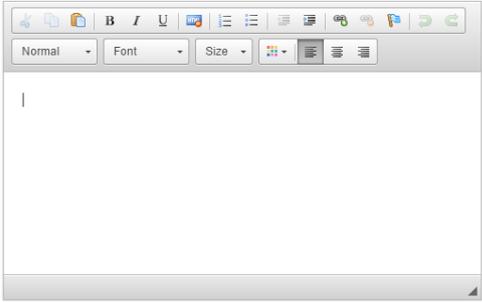
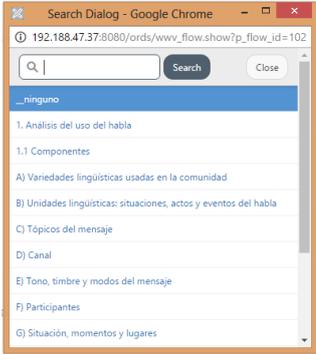
Vista en árbol	<ul style="list-style-type: none"> ▼ Guía de estudio etnográfico del uso del habla de Sherzer y Darnell <ul style="list-style-type: none"> ▼ 1. Análisis del uso del habla <ul style="list-style-type: none"> ▼ 1.1 Componentes <ul style="list-style-type: none"> A) Variedades lingüísticas usadas en la comunidad B) Unidades lingüísticas: situaciones, actos y eventos del habla
Selección de archivo	<div style="border: 1px solid gray; padding: 5px; display: inline-block;"> Seleccionar archivo Ningún archivo seleccionado </div>
Editor de textos	
Ventana emergente con lista de valores	
Mapa geográfico	

Tabla 4.5: Elementos de interfaz de usuario

4.2.3 Aplicación para usuario Administrador de contenido

En este apartado se describe cada una de las opciones (funciones) de la aplicación web "Cartografía etnolingüística" junto con los formularios y sus respectivos componentes.

4.2.3.1 Iniciar sesión

La pantalla de inicio de esta aplicación es común, el inicio de sesión requiere de dos campos básicos: Usuario y Password. Una vez llenada esta información el usuario debe acceder al botón "Entrar", seguido se realizan validaciones internas de autenticación del usuario para generar las opciones adecuadas según el rol que tenga el usuario y accede a la pantalla de inicio de la aplicación, Al presentar problemas de autenticación la aplicación genera el informe de error en la propia página.

The image shows a login form with a light gray header containing the text "Iniciar Sesión". Below the header, there are two input fields: "Usuario" with the text "jfernald" and "Password" with three dots indicating a masked password. A dark blue button with the text "Entrar" is located at the bottom right of the form.

Ilustración 4.9: Cartografía etnolingüística - Inicio de Sesión

4.2.3.2 Inicio

Posterior a la autenticación del usuario, la pantalla de inicio es la primera que se presenta al usuario final, su función es mostrar geográficamente las comunidades de estudio del usuario mediante los mapas de google.

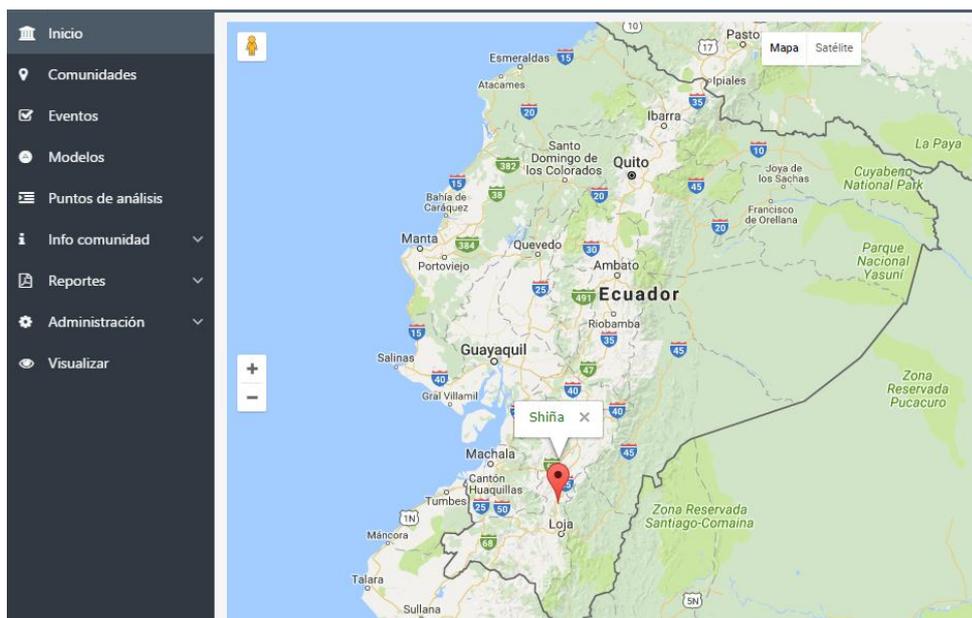


Ilustración 4.10: Cartografía etnolingüística – Página de Inicio

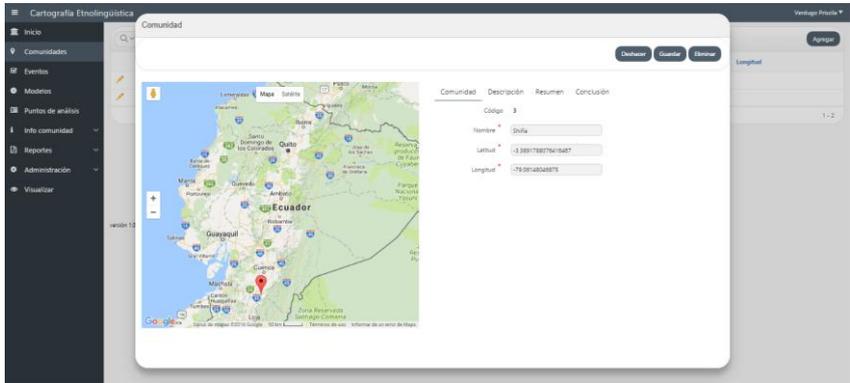
4.2.3.3 Comunidades

La opción de comunidades permite gestionar todas las comunidades objeto de estudio, esta opción está conformada por dos páginas, la primera una visualización del reporte en forma de tabla con los datos de las comunidades que han sido guardadas en la aplicación y la segunda un formulario modal que permite ingresar, modificar y eliminar el registro referente a la comunidad.

Objeto	Reporte de comunidades	
Composición	Contiene elementos básicos como barra lateral de menús, botón de agregar para nuevas comunidades y el reporte de todas las comunidades que el usuario tiene almacenado.	
Imagen		
Elementos del reporte	Buscar	Recibe cadenas de texto de búsqueda para filtrar datos de uso para el usuario.

	Botón: Agregar	Dirige al usuario hasta la pantalla del mantenimiento de comunidad, para agregar un nuevo registro.
	Enlace: Editar	Dirige al usuario hasta la pantalla del mantenimiento de comunidad, con la información del elemento seleccionado.
	Cuadrícula detalle	Listado de elementos existentes.

Tabla 4.6: Reporte de comunidades

Objeto	Mantenimiento de comunidades	
Composición	Una página modal que contiene elementos del formulario con los datos de una comunidad.	
Imagen		
Elementos del Formulario	Botón: Deshacer	Al hacer clic redirige a la página reporte de comunidades.
	Botón: Guardar	Al hacer clic, la aplicación verifica los datos ingresados según los requerimientos y los guarda, termina el proceso regresando al reporte de comunidades.
	Botón: Eliminar	<ul style="list-style-type: none"> Al hacer clic, el sistema visualiza un mensaje de confirmación con el texto "¿Está seguro de eliminar?", botón Aceptar para continuar de no ser así el usuario puede Cancelar la operación.

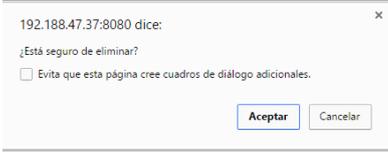
		<p>Al finalizar el proceso de eliminación la aplicación redirige al reporte de comunidades.</p>  <ul style="list-style-type: none"> Este botón estará visible en el formulario cuando el registro este en edición.
	Campo (lectura): Código	Número generado automáticamente que identifica el registro para el usuario.
	Campo de texto: Nombre	Nombre de la comunidad.
	Campo de texto (lectura): Latitud	Número para latitud geográfica de la comunidad y así poder presentarla con exactitud en un futuro, cambia según el puntero del cursor en el mapa.
	Campo de texto (lectura): Longitud	Número para longitud geográfica de la comunidad y así poder presentarla con exactitud en un futuro, cambia según el puntero del cursor en el mapa.
	Editor de texto: Descripción	Un editor de texto tipo office para ingresar información referente a la comunidad.
	Editor de texto: Resumen	Un editor de texto tipo office para ingresar un breve resumen de una comunidad, este se visualiza en el encabezado del mapa.
	Editor de texto: Conclusión	Un editor de texto tipo office para ingresar información referente a una conclusión exacta de una comunidad.
Indicaciones	Los campos del formulario marcados con asterisco (*) son de ingreso obligatorio para el usuario.	

Tabla 4.7: Mantenimiento de comunidad

4.2.3.4 Eventos

La opción de eventos permite gestionar los eventos (escenarios) en los cuales se da el estudio de una comunidad, esta opción está conformada por dos páginas, la primera una visualización del reporte en forma de tabla los datos esenciales de los eventos que han sido guardadas en la aplicación y la segunda un formulario modal que permite ingresar, modificar y eliminar el registro referente al evento.

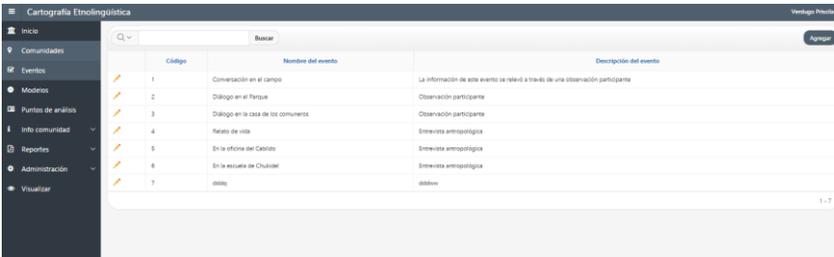
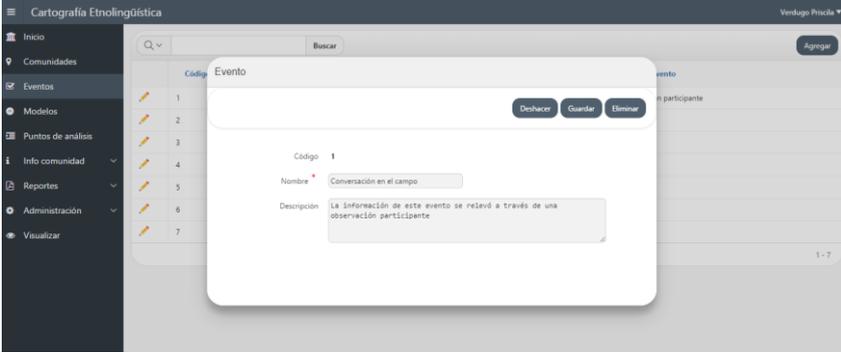
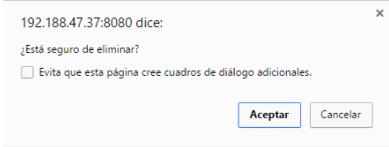
Objeto	Reporte de eventos	
Composición	Contiene elementos básicos como barra lateral de menús, botón de agregar para nuevos eventos y el reporte de todos los eventos que el usuario tiene almacenado.	
Imagen		
Elementos del reporte	Buscar	Recibe cadenas de texto de búsqueda para filtrar datos de uso para el usuario.
	Botón: Agregar	Dirige al usuario hasta la pantalla del mantenimiento de evento, para agregar un nuevo registro.
	Enlace: Editar	Dirige al usuario hasta la pantalla del mantenimiento del evento, con la información del elemento seleccionado.
	Cuadrícula detalle	Listado de elementos existentes.

Tabla 4.8: Reporte de eventos

Objeto	Mantenimiento de eventos	
Composición	Una página modal que contiene elementos del formulario referentes al evento de estudio.	
Imagen		
Elementos de Formulario	Botón: Deshacer	Al hacer clic redirige a la página reporte de eventos.
	Botón: Guardar	Al hacer clic, la aplicación verifica los datos ingresados según los requerimientos y guarda, termina el proceso regresando al reporte de eventos. Si son incorrectos la aplicación genera el informe para el usuario.
	Botón: Eliminar	<ul style="list-style-type: none"> Al hacer clic, el sistema visualiza un mensaje de confirmación con el texto “¿Está seguro de eliminar?”, botón Aceptar para continuar de no ser así el usuario puede Cancelar la operación. Al finalizar el proceso de eliminación la aplicación redirige al reporte de eventos.  <ul style="list-style-type: none"> Este botón estará visible en el formulario cuando el registro este en edición.
	Campo (lectura): Código	Número generado automáticamente que identifica el registro para el usuario.
	Campo de texto: Nombre	Nombre del evento.
Área de texto: Descripción	Un área de texto para ingresar información que distingue al evento.	

Indicaciones	Los campos del formulario marcados con asterisco (*) son de ingreso obligatorio para el usuario.
--------------	--

Tabla 4.9: Mantenimiento de evento

4.2.3.5 Modelos

La opción de modelos permite gestionar los modelo de estudio para una comunidad, un modelo de estudio es un nombre que representa una lista de puntos de análisis a tomar en cuenta en el estudio de la comunidad, esta opción está conformada por dos páginas, la primera una visualización del reporte en forma de tabla los datos esenciales de los modelos que han sido guardados en la aplicación y la segunda un formulario modal que permite ingresar, modificar y eliminar el registro referente al modelo.

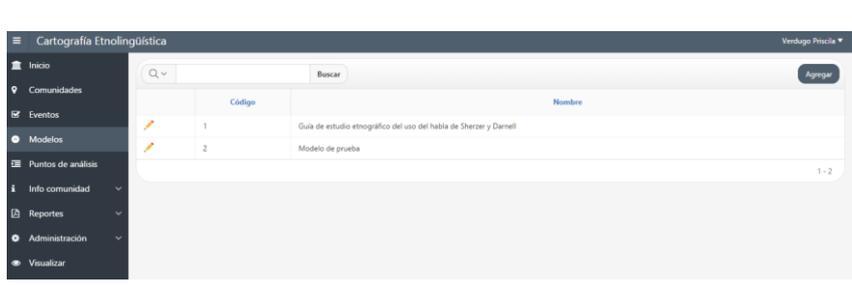
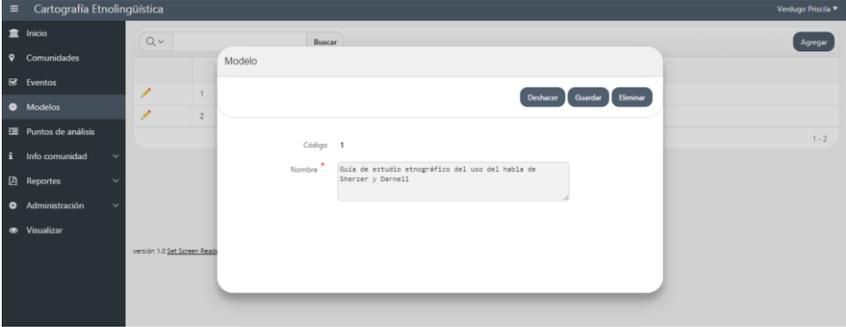
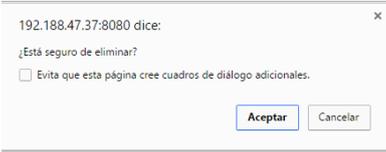
Objeto	Reporte de modelos	
Composición	Contiene elementos básicos como barra lateral de menús, botón de agregar para nuevos modelos y el reporte de todos los modelos que el usuario tiene almacenado.	
Imagen		
Elementos del reporte	Buscar	Recibe cadenas de texto de búsqueda para filtrar datos de uso para el usuario.
	Botón: Agregar	Dirige al usuario hasta la pantalla del mantenimiento de modelo, para agregar un nuevo registro.
	Enlace: Editar	Dirige al usuario hasta la pantalla del mantenimiento del modelo, con la información del elemento seleccionado.
	Cuadrícula detalle	Listado de elementos existentes.

Tabla 4.10: Reporte de modelos de estudio

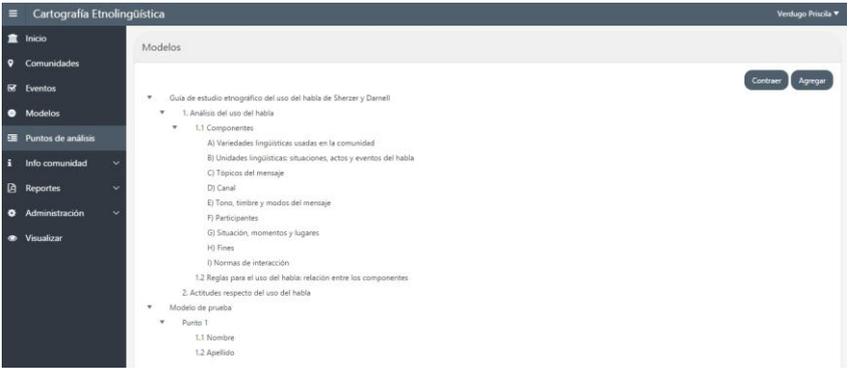
Objeto	Mantenimiento de modelos	
Composición	Una página modal que contiene elementos del formulario referentes al modelo de estudio.	
Imagen		
Elementos del formulario	Botón: Deshacer	Al hacer clic redirige a la página reporte de modelos.
	Botón: Guardar	Al hacer clic, la aplicación verifica los datos ingresados según los requerimientos y guarda, termina el proceso regresando al reporte de modelos, Si son incorrectos la aplicación genera el informe para el usuario.
	Botón: Eliminar	<ul style="list-style-type: none"> Al hacer clic, el sistema visualiza un mensaje de confirmación con el texto "¿Está seguro de eliminar?", botón Aceptar para continuar de no ser así el usuario puede Cancelar la operación. Al finalizar el proceso de eliminación la aplicación redirige al reporte de modelos.  <ul style="list-style-type: none"> Este botón estará visible en el formulario cuando el registro este en edición.
	Campo (lectura): Código	Número generado automáticamente que identifica el registro para el usuario.

	Campo de texto: Nombre	Nombre del modelo de estudio.
Indicaciones	Los campos del formulario marcados con asterisco (*) son de ingreso obligatorio para el usuario.	

Tabla 4.11: Mantenimiento de modelo de estudio

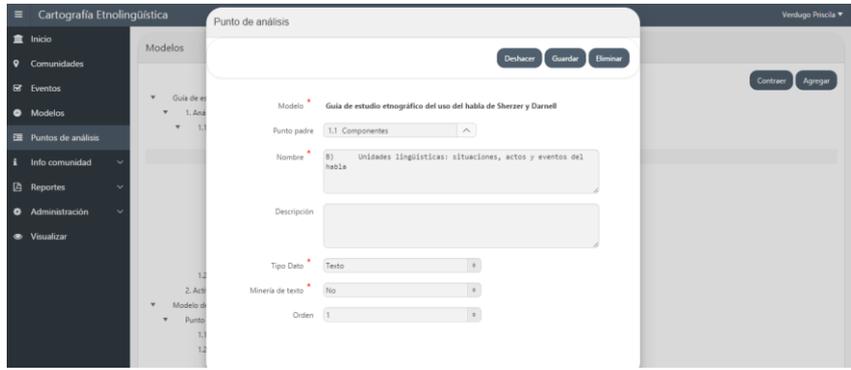
4.2.3.6 Puntos de análisis

La opción de puntos de análisis permite gestionar los puntos de análisis de un modelo, un punto de análisis es un dato relevante a tomar en cuenta en el estudio de una comunidad, esta opción está conformada por dos páginas, la primera una visualización en forma jerárquica con los nombres de los puntos de análisis que han sido guardadas en la aplicación y la segunda un formulario modal que permite ingresar, modificar y eliminar el registro referente al punto de análisis.

Objeto	Vista tipo árbol de puntos de análisis	
Composición	Contiene elementos básicos como barra lateral de menús, botón de agregar para nuevos puntos de análisis y una vista en árbol de los puntos de análisis que el usuario tiene almacenado.	
Imagen		
Elementos del reporte	Contraer	Contrae la vista al nivel más bajo.
	Expandir	Alarga la vista al nivel más alto.
	Botón: Agregar	Dirige al usuario hasta la pantalla del mantenimiento del punto de análisis, para agregar un nuevo registro.

	Nodo (doble click): Editar	Dirige al usuario hasta la pantalla del mantenimiento del punto de análisis, con la información del nodo seleccionado.
	Vista en árbol	Listado en forma jerárquica de elementos existentes.

Tabla 4.12: Vista en árbol puntos de análisis

Objeto	Mantenimiento de punto de análisis	
Composición	Una página modal que contiene elementos del formulario referentes al punto de análisis.	
Imagen		
Elementos del formulario	Botón: Deshacer	Al hacer clic redirige a la página de vista en árbol de puntos de análisis.
	Botón: Guardar	Al hacer clic, la aplicación verifica los datos ingresados según los requerimientos y guarda, termina el proceso regresando a la vista en árbol de puntos de análisis, Si son incorrectos la aplicación genera el informe para el usuario.
	Botón: Eliminar	<ul style="list-style-type: none"> Al hacer clic, el sistema visualiza un mensaje de confirmación con el texto "¿Está seguro de eliminar?", botón Aceptar para continuar de no ser así el usuario puede Cancelar la operación. Al finalizar el proceso de eliminación la aplicación redirige a la vista en árbol de puntos de análisis.

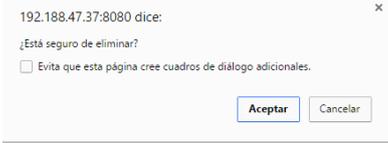
		 <ul style="list-style-type: none"> Este botón estará visible en el formulario cuando el registro este en edición.
	Lista de opciones: Modelo	Una lista con todos los modelos de estudio del usuario.
	Ventana emergente: Punto padre	Ventana emergente con lista de todos los puntos de análisis que pertenecen a la lista de opciones modelo.
	Área de texto: Nombre	Un nombre para el punto de análisis en el modelo.
	Área de texto: descripción	Una descripción para punto de análisis en mantenimiento.
	Lista de opciones: Tipo de dato	Opción para el tipo de dato que va a manejar el punto de análisis.
	Lista de opciones: Minería de texto	Opción que indica si este punto de análisis es considerado para minería de texto.
	Lista de opciones: Orden	Permite asignar una secuencia a los puntos de análisis en la vista tipo árbol.
Indicaciones	Los campos del formulario marcados con asterisco (*) son de ingreso obligatorio para el usuario.	

Tabla 4.13: Mantenimiento de punto de análisis

4.2.3.7 Info Comunidad - Submenú

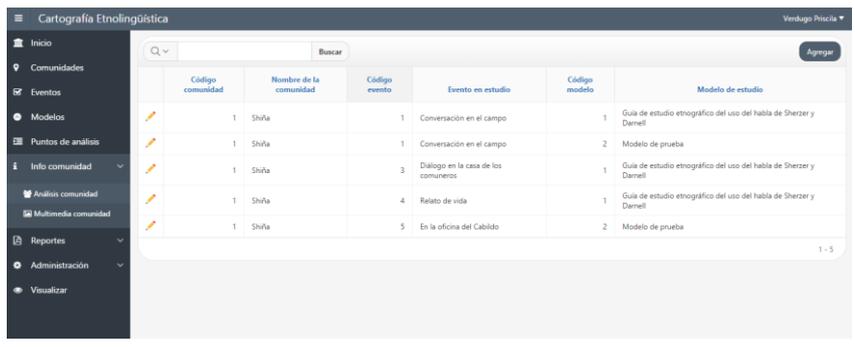
4.2.3.7.1 Análisis de comunidad

Una de las partes centrales de esta aplicación web es poder realizar un análisis de una comunidad mediante la intersección de: una comunidad, un modelo de estudio y un evento (escenario) de estudio, en este punto gestiona información relevante a los puntos de análisis según la intersección dada.

Según los textos proporcionados, la aplicación puede generar datos acerca de comportamiento lingüístico tanto en el idioma español y quechua, lo que no se encuentre detectado también será información como variante de la lengua.

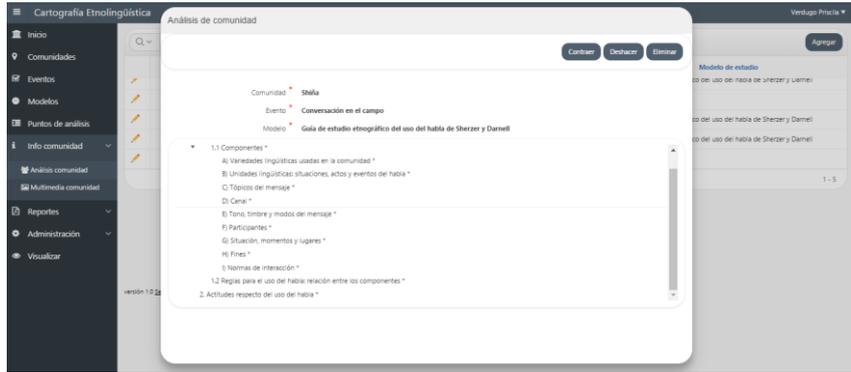
Esta opción está conformada por dos páginas maestras, la primera una visualización del reporte en forma de tabla los datos esenciales de los análisis que han sido guardados en la aplicación,

La segunda un formulario modal que permite ingresar, modificar y eliminar los registros referentes a cada análisis, esta segunda página contiene: en primera instancia, una vista en árbol que contiene los puntos de análisis de ese modelo, en segunda instancia, un editor de texto que permite ingresar o modificar información relevante al punto de análisis, y la tercera instancia el clasificador de texto el cual ejecuta la minería de texto usando el clasificador de palabras.

Objeto	Reporte de análisis de comunidades																																				
Composición	Contiene elementos básicos como barra lateral de menús, botón de agregar para nuevas comunidades y el reporte de todos los análisis que el usuario tiene almacenado.																																				
Imagen	 <p>The screenshot displays the 'Cartografía Etnolingüística' application interface. On the left is a dark sidebar menu with options: Inicio, Comunidades, Eventos, Modelos, Puntos de análisis, Info comunidad (selected), Análisis comunidad, Multimedia comunidad, Reportes, Administración, and Visualizar. The main content area features a search bar and a table with the following data:</p> <table border="1"><thead><tr><th>Código comunidad</th><th>Nombre de la comunidad</th><th>Código evento</th><th>Evento en estudio</th><th>Código modelo</th><th>Modelo de estudio</th></tr></thead><tbody><tr><td>1</td><td>Shiña</td><td>1</td><td>Conversación en el campo</td><td>1</td><td>Guía de estudio etnográfico del uso del habla de Sherzer y Danelli</td></tr><tr><td>1</td><td>Shiña</td><td>1</td><td>Conversación en el campo</td><td>2</td><td>Modelo de prueba</td></tr><tr><td>1</td><td>Shiña</td><td>3</td><td>Diálogo en la casa de los comuneros</td><td>1</td><td>Guía de estudio etnográfico del uso del habla de Sherzer y Danelli</td></tr><tr><td>1</td><td>Shiña</td><td>4</td><td>Relato de vida</td><td>1</td><td>Guía de estudio etnográfico del uso del habla de Sherzer y Danelli</td></tr><tr><td>1</td><td>Shiña</td><td>5</td><td>En la oficina del Cabildo</td><td>2</td><td>Modelo de prueba</td></tr></tbody></table> <p>At the bottom right of the table, there is a pagination indicator '1 - 5'.</p>	Código comunidad	Nombre de la comunidad	Código evento	Evento en estudio	Código modelo	Modelo de estudio	1	Shiña	1	Conversación en el campo	1	Guía de estudio etnográfico del uso del habla de Sherzer y Danelli	1	Shiña	1	Conversación en el campo	2	Modelo de prueba	1	Shiña	3	Diálogo en la casa de los comuneros	1	Guía de estudio etnográfico del uso del habla de Sherzer y Danelli	1	Shiña	4	Relato de vida	1	Guía de estudio etnográfico del uso del habla de Sherzer y Danelli	1	Shiña	5	En la oficina del Cabildo	2	Modelo de prueba
Código comunidad	Nombre de la comunidad	Código evento	Evento en estudio	Código modelo	Modelo de estudio																																
1	Shiña	1	Conversación en el campo	1	Guía de estudio etnográfico del uso del habla de Sherzer y Danelli																																
1	Shiña	1	Conversación en el campo	2	Modelo de prueba																																
1	Shiña	3	Diálogo en la casa de los comuneros	1	Guía de estudio etnográfico del uso del habla de Sherzer y Danelli																																
1	Shiña	4	Relato de vida	1	Guía de estudio etnográfico del uso del habla de Sherzer y Danelli																																
1	Shiña	5	En la oficina del Cabildo	2	Modelo de prueba																																

Elementos del reporte	Buscar	Recibe cadenas de texto de búsqueda para filtrar datos de uso para el usuario.
	Botón: Agregar	Dirige al usuario hasta la pantalla del mantenimiento análisis de comunidad, para agregar un nuevo registro.
	Enlace: Editar	Dirige al usuario hasta la pantalla del mantenimiento de análisis de comunidad, con la información del elemento seleccionado.
	Cuadrícula detalle	Listado de elementos existentes.

Tabla 4.14: Reporte de análisis de comunidades

Objeto	Mantenimiento análisis de comunidad – Vista jerárquica	
Composición	Una página modal que contiene elementos del formulario referentes al análisis de una comunidad.	
Imagen		
Elementos del formulario	Botón: Deshacer	Al hacer clic redirige a la página de reporte análisis de comunidades.
	Botón: generar Índice	Al hacer clic, la aplicación genera un índice con todos los puntos de análisis del modelo de estudio incluido en el análisis.
	Botón: Eliminar	<ul style="list-style-type: none"> Al hacer clic, el sistema visualiza un mensaje de confirmación con el texto “¿Está seguro de eliminar?”, botón Aceptar para continuar de no ser así el usuario puede Cancelar la operación.

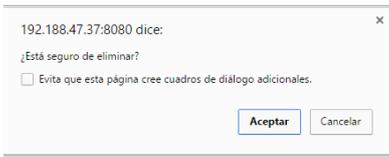
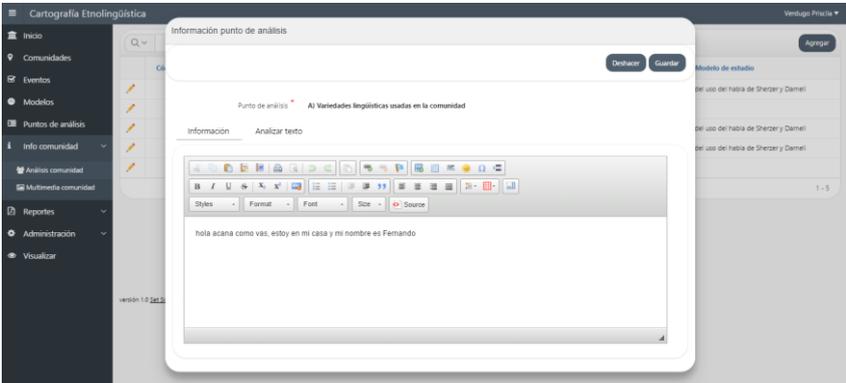
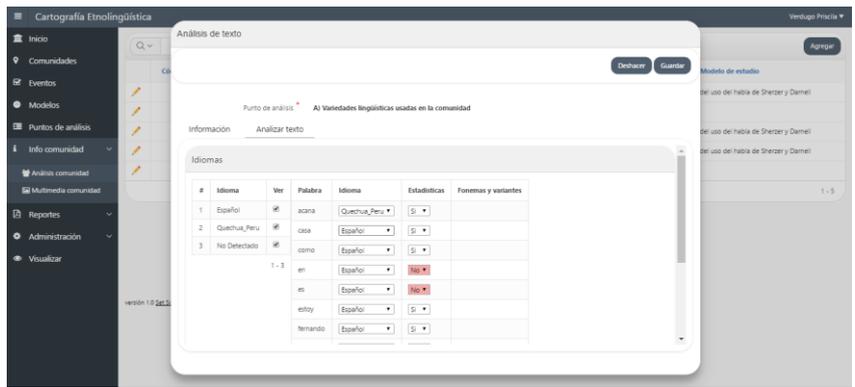
		<p>Al finalizar el proceso de eliminación la aplicación redirige al reporte de análisis de comunidades.</p>  <ul style="list-style-type: none"> Este botón estará visible en el formulario cuando el registro este en edición.
	Contraer	Contrae la vista al nivel más bajo.
	Expandir	Alarga la vista al nivel más alto.
Indicaciones	Los campos del formulario marcados con asterisco (*) son de ingreso obligatorio para el usuario.	

Tabla 4.15: Mantenimiento análisis comunidad - Vista Jerárquica

Objeto	Mantenimiento análisis de comunidad – Editor de texto	
Composición	Una página modal que contiene elementos del formulario referentes a la edición de un punto de análisis en el estudio de una comunidad.	
Imagen		
Elementos del formulario	Botón: Deshacer	Al hacer clic redirige a la página de vista jerárquica del mantenimiento del análisis de comunidad.
	Botón: Guardar	Almacena la información textual del punto de análisis en curso.
	Link: Información	Representa la página actual de trabajo.

	Link: Analizar texto	Redirección a la tercera instancia la cual es una opción de la minería de texto el clasificador de idioma para textos.
	Campo de texto (lectura): Punto de análisis	Visualiza el nombre del punto de análisis en el cual se está trabajando.
	Editor de texto: Descripción	Un editor de texto tipo office para ingresar información textual del punto de análisis en curso.

Tabla 4.16: Mantenimiento análisis comunidad - Editor de texto

Objeto	Mantenimiento análisis de comunidad – Clasificador de textos (Idioma)	
Composición	Una página modal que contiene elementos del formulario referentes a la minería de textos aplicando clasificadores de texto y así obtener su idioma de escritura.	
Imagen		
Elementos del formulario	Botón: Deshacer	Al hacer clic redirige a la página de vista jerárquica del mantenimiento del análisis de comunidad.
	Botón: Guardar	Almacena la información arrojada por la minería de texto.
	Link: Información	Redirección a la segunda instancia la cual es la opción de editor de texto.

Link: Analizar texto	Representa la página actual de trabajo.
Campo de texto (lectura): Punto de análisis	Visualiza el nombre del punto de análisis en el cual se está trabajando.
Cuadrícula detalle: Idioma	Listado de los idiomas existentes.
Cuadrícula detalle: Palabras	Listado de las palabras detectadas.
Casilla de verificación: Ver	Filtra las palabras que han sido detectadas en el registro en curso
Lista de opciones: Estadísticas	Permite realizar una inclusión manual de la palabra en la cuantificación final.

Tabla 4.17: Mantenimiento análisis comunidad - Clasificador de textos

4.2.3.7.2 Multimedia comunidad

En esta opción el usuario puede gestionar información multimedia (imágenes, audio y video) referente a cada comunidad, por condiciones de usabilidad para el usuario final todos los tipos de archivos se encuentran bajo un mantenimiento.

Esta opción está conformada por dos páginas, la primera una visualización del reporte en forma de tabla los datos esenciales de los archivos multimedia que han sido guardados en la aplicación y la segunda un formulario modal que permite ingresar, modificar y eliminar el registro referente a cada archivo.

Objeto	Reporte de multimedia comunidad
Composición	Contiene elementos básicos como barra lateral de menús, botón de agregar para nuevos modelos y el reporte de todos los archivos multimedia que el usuario tiene almacenado.

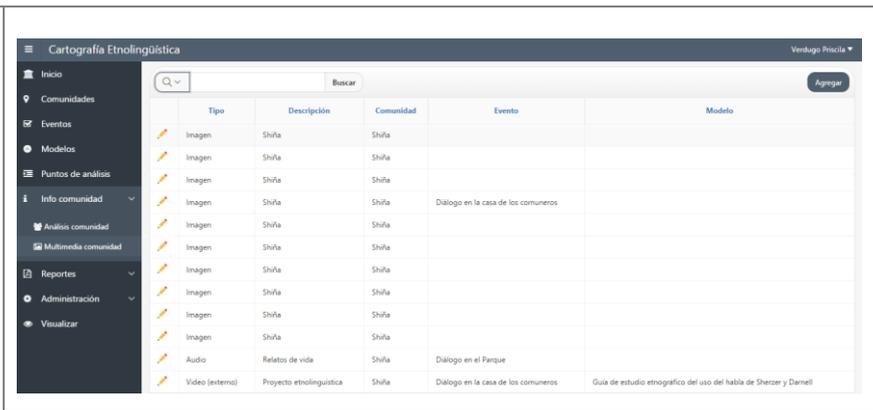
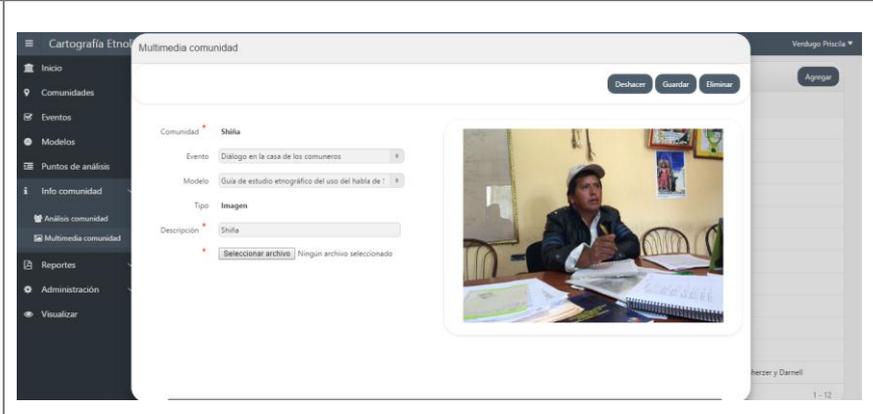
Imagen		
Elementos del reporte	<p>Buscar</p>	<p>Recibe cadenas de texto de búsqueda para filtrar datos de uso para el usuario.</p>
	<p>Botón: Agregar</p>	<p>Dirige al usuario hasta la pantalla del mantenimiento de multimedia comunidad, para agregar un nuevo registro.</p>
	<p>Enlace: Editar</p>	<p>Dirige al usuario hasta la pantalla del mantenimiento de multimedia comunidad, con la información del elemento seleccionado.</p>
	<p>Cuadrícula detalle</p>	<p>Listado de elementos existentes.</p>

Tabla 4.18: Reporte multimedia comunidad

Objeto	Mantenimiento multimedia comunidad	
Composición	Una página modal que contiene elementos del formulario referentes a un archivo multimedia de la comunidad.	
Imagen		
Elementos del formulario	<p>Botón: Deshacer</p>	<p>Al hacer clic redirige a la página reporte de multimedia comunidad.</p>
	<p>Botón: Guardar</p>	<p>Al hacer clic, la aplicación verifica los datos ingresados según los requerimientos y guarda,</p>

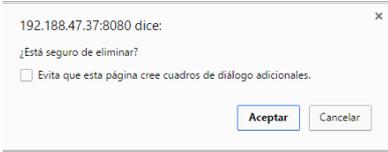
		termina el proceso regresando al reporte de multimedia comunidad, Si son incorrectos la aplicación genera el informe para el usuario.
	Botón: Eliminar	<ul style="list-style-type: none"> Al hacer clic, el sistema visualiza un mensaje de confirmación con el texto “¿Está seguro de eliminar?”, botón Aceptar para continuar de no ser así el usuario puede Cancelar la operación. Al finalizar el proceso de eliminación la aplicación redirige al reporte de multimedia comunidad.  <ul style="list-style-type: none"> Este botón estará visible en el formulario cuando el registro este en edición.
	Lista de opciones: Comunidad	Permite seleccionar la comunidad a la que se encuentra ligado el archivo.
	Lista de opciones: Evento	Permite seleccionar el evento al que se encuentra ligado el archivo.
	Lista de opciones: modelo	Permite seleccionar el modelo de estudio al que se encuentra ligado el archivo.
	Lista de opciones: Tipo	Permite seleccionar el tipo de archivo.
	Campo de texto: Descripción	Permite ingresar una breve descripción del archivo.
	Selector de archivo	Permite obtener un archivo del usuario.
Indicaciones	Los campos del formulario marcados con asterisco (*) son de ingreso obligatorio para el usuario.	

Tabla 4.19: Mantenimiento multimedia comunidad

4.2.3.7.3 Contraste de comunidades

Existen casos en los que es necesario almacenar notas acerca del contraste o comparación entre dos comunidades para ello es necesaria una función extra en la aplicación web.

Esta opción está conformada por dos páginas, la primera una visualización del reporte en forma de tabla los datos esenciales de los diferentes contrastes que han sido guardados en la aplicación y la segunda un formulario modal que permite ingresar, modificar y eliminar el registro referente a cada contraste.

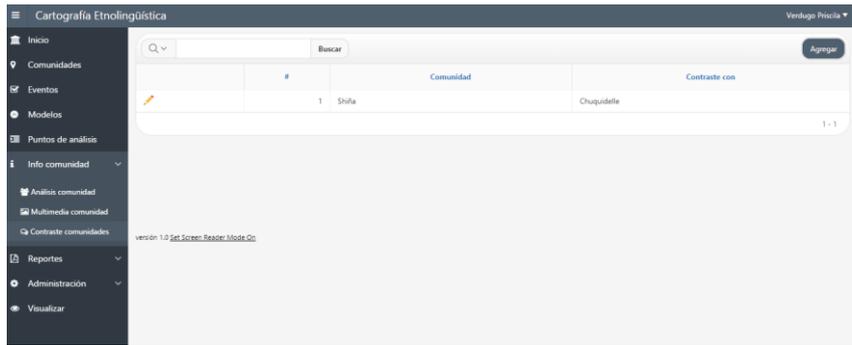
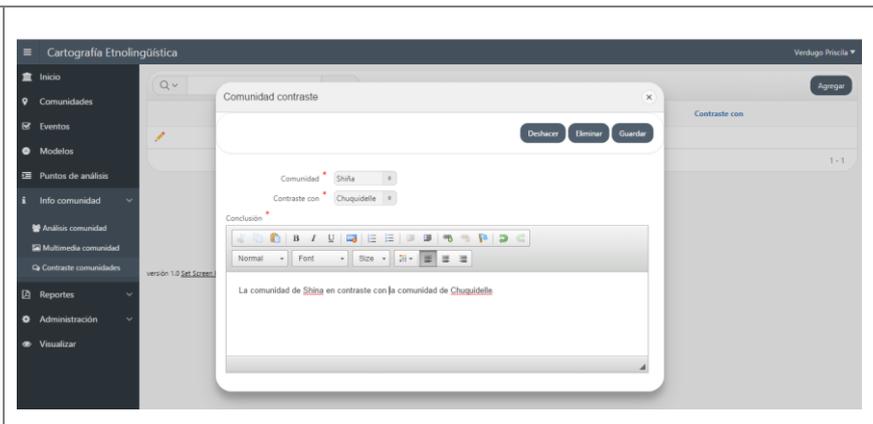
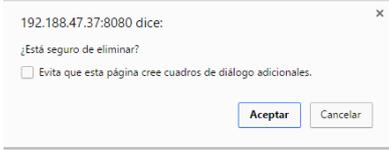
Objeto	Reporte de comunidad contraste	
Composición	Contiene elementos básicos como barra lateral de menús, botón de agregar para nuevos modelos y el reporte de todos registros de contrastes almacenados.	
Imagen		
Elementos del reporte	Buscar	Recibe cadenas de texto de búsqueda para filtrar datos de uso para el usuario.
	Botón: Agregar	Dirige al usuario hasta la pantalla del mantenimiento de contraste comunidad, para agregar un nuevo registro.
	Enlace: Editar	Dirige al usuario hasta la pantalla del mantenimiento de contraste comunidad, con la información del elemento seleccionado.
	Cuadrícula detalle	Listado de elementos existentes.

Tabla 4.20: Reporte comunidad contraste

Objeto	Mantenimiento comunidad contraste
Composición	Una página modal que contiene elementos del formulario referentes a un contraste entre dos comunidades.

<p>Imagen</p>		
<p>Elementos del formulario</p>	<p>Botón: Deshacer</p>	<p>Al hacer clic redirige a la página reporte de contraste de comunidades.</p>
	<p>Botón: Guardar</p>	<p>Al hacer clic, la aplicación verifica los datos ingresados según los requerimientos y guarda, termina el proceso regresando al reporte de contraste de comunidades, Si son incorrectos la aplicación genera el informe de errores para el usuario.</p>
	<p>Botón: Eliminar</p>	<ul style="list-style-type: none"> Al hacer clic, el sistema visualiza un mensaje de confirmación con el texto “¿Está seguro de eliminar?”, botón Aceptar para continuar de no ser así el usuario puede Cancelar la operación. Al finalizar el proceso de eliminación la aplicación redirige al reporte de contraste de comunidades.  <ul style="list-style-type: none"> Este botón estará visible en el formulario cuando el registro este en edición.
	<p>Lista de opciones: Comunidad</p>	<p>Permite seleccionar la comunidad en tema del contraste.</p>
	<p>Lista de opciones: Contraste con</p>	<p>Permite seleccionar la comunidad arbitraria del contraste.</p>

	Editor de texto: Conclusión	Permite ingresar la conclusión del contraste entre las comunidades.
Indicaciones	Los campos del formulario marcados con asterisco (*) son de ingreso obligatorio para el usuario.	

Tabla 4.21: Mantenimiento comunidad contraste

4.2.3.8 Reportes

Un requisito del usuario es poder exportar sus análisis de las diferentes comunidades, pues esta opción permite generar un archivo con extensión pdf, con información relevante a cualquier análisis creado por el usuario.

Presenta una única pantalla en forma de reporte tipo tabla con los datos referentes a los diferentes análisis realizados por el usuario.

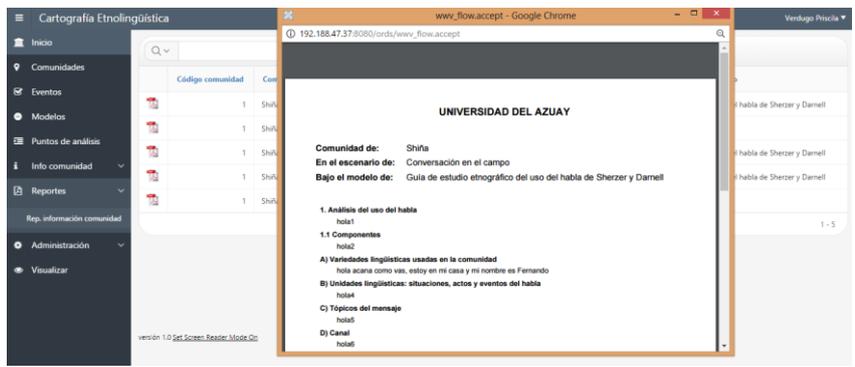
Objeto	Reporte de exportar análisis comunidad	
Composición	Contiene elementos básicos como barra lateral de menús, botón de agregar para nuevos modelos y el reporte de todos los archivos multimedia que el usuario tiene almacenado.	
Imagen		
Elementos del reporte	Buscar	Recibe cadenas de texto de búsqueda para filtrar datos de uso para el usuario.
	Link: pdf	Genera una ventana emergente con un archivo pdf para uso del usuario.
	Cuadrícula detalle	Listado de elementos existentes.

Tabla 4.22: Reporte exportar análisis comunidad

4.2.3.9 Administración

Una opción que permite al usuario manipular la presentación de las diferentes opciones de la aplicación web, permite también gestionar los usuarios a la aplicación, esta opción consta de un mantenimiento acerca de manejo de usuario y roles de la aplicación.

4.2.3.10 Visualizar

Esta opción permite redirigir al usuario a la aplicación del usuario visitante (presentación para el público).

Una vez en la aplicación del usuario visitante el administrador puede visualizar de manera simple todas las páginas (opciones) de presentación aun así puede regresar a la administrar son perder su sesión.

4.2.4 Aplicación para usuario Visitante

Para concluir con el desarrollo, la aplicación para los visitantes (buscadores de información), esta aplicación presenta de forma ordenada y amistosa para el usuario visitante todas las comunidades geográficamente, análisis de cada comunidad, datos resultantes de la minería de texto, fotografías, audios y videos referente una comunidad en específico.

4.2.4.1 Cartografía etnolingüística (inicio)

Un mapa geográfico que presenta de forma ordenada la ubicación de cada comunidad utilizando marcas o punteros sobre el mapa geográfico.

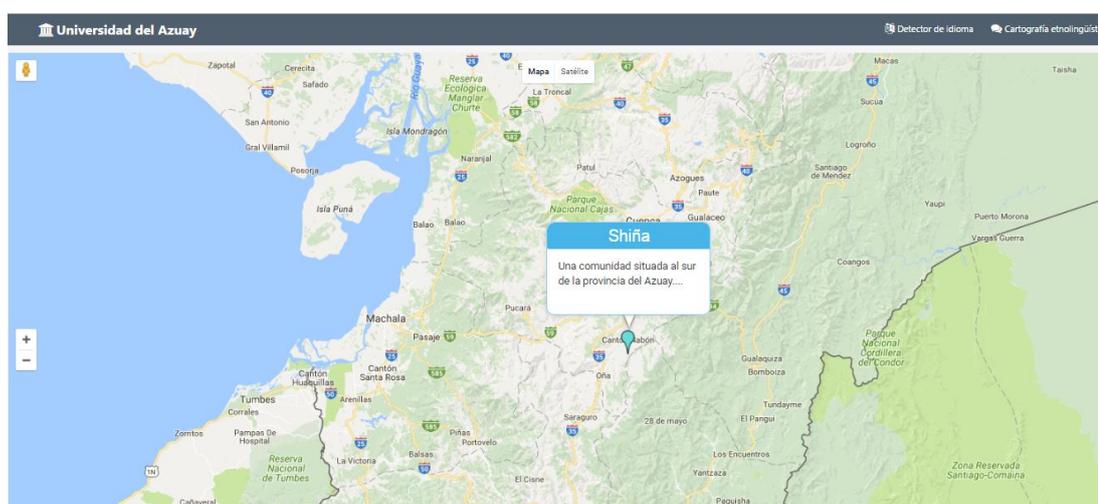


Ilustración 4.11: Aplicación de presentación – Comunidades

4.2.4.2 Introducción a la comunidad

Una página modal compuesta por una lista de links, un visor simple de imágenes y un campo de descripción, esta página contiene información relevante a la introducción de una comunidad.

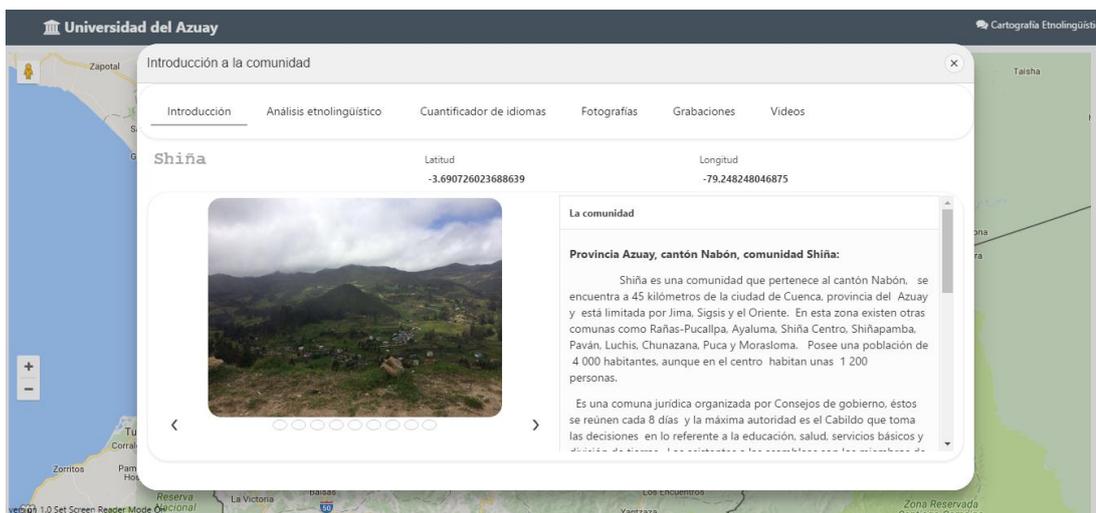


Ilustración 4.12: Aplicación de presentación - Introducción comunidad

4.2.4.3 Análisis etnolingüística

Una página modal compuesta por links, selectores de opciones: Modelo de estudio y Evento (estudio de), vista tipo árbol para los puntos de análisis de la comunidad y un detalle para presentar información del punto de análisis.

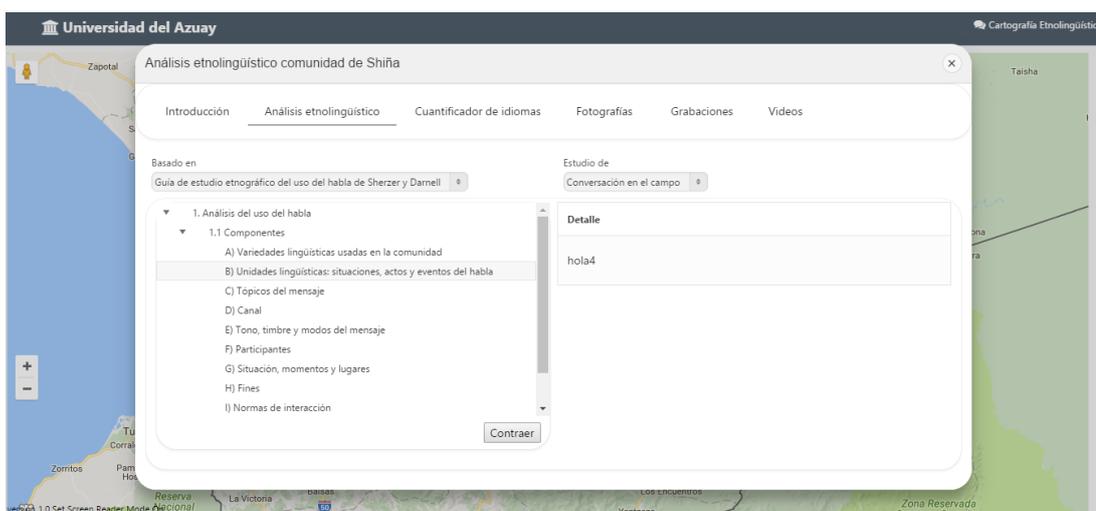


Ilustración 4.13: Aplicación de presentación - Análisis etnolingüístico

4.2.4.4 Cuantificador de idiomas

Una página modal con gráficas en HTML5 para presentar información numérica de las siguientes características:

- Gráfica tipo pastel: porcentaje de detección de cada idioma en los textos.
- Tabla: un ranking de las palabras más utilizadas en cada uno de los idiomas.
- Gráfica en barras: una comparación numérica de la cantidad de palabras entre dos idiomas.

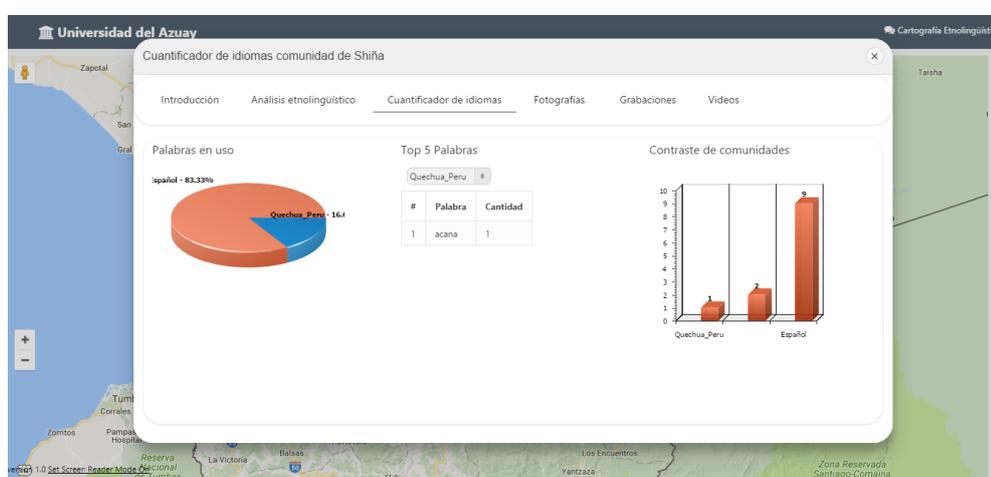


Ilustración 4.14: Aplicación de presentación - Cuantificador de idiomas

4.2.4.5 Fotografías

Una página modal donde se presenta una galería de las fotografías relevantes de una comunidad.

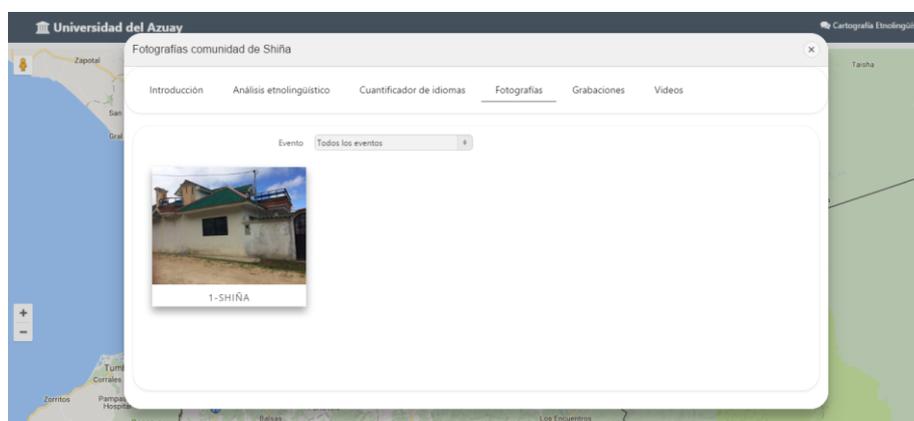


Ilustración 4.15: Aplicación de presentación – Fotografías comunidad

4.2.4.6 Grabaciones

Página modal con una galería de audios referentes a la comunidad.

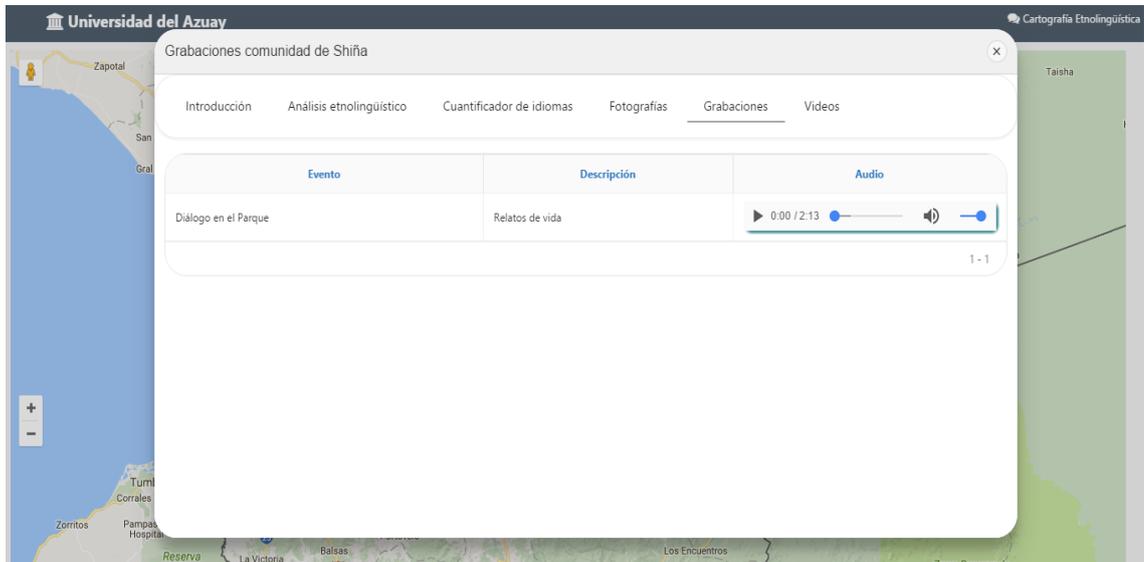


Ilustración 4.16: Aplicación de presentación – Grabaciones comunidad

4.2.4.7 Videos

Página modal con galería de videos de comunidades.

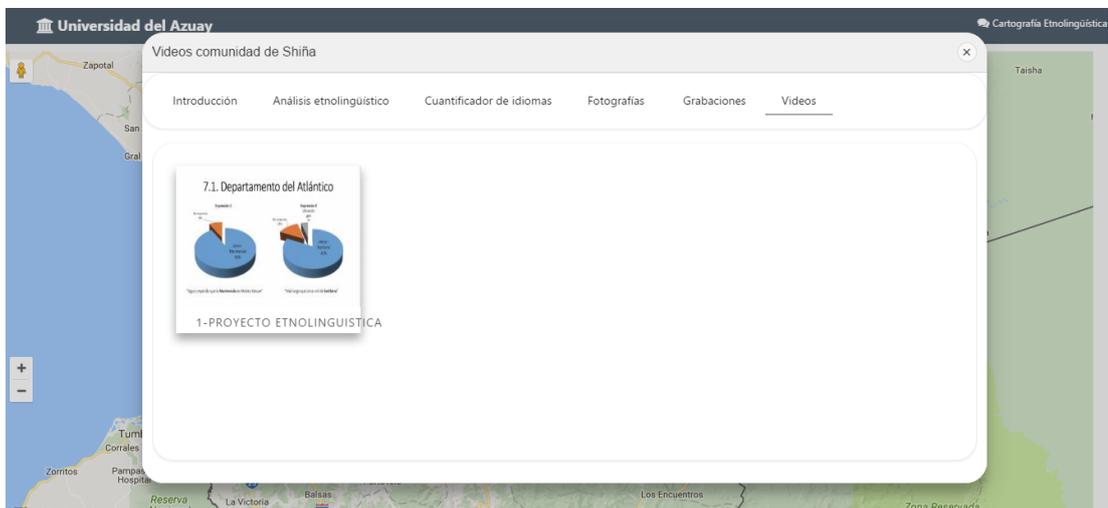


Ilustración 4.17: Aplicación de presentación – Videos comunidad

4.3.4.8 Conclusiones de comunidad

Página modal con campo visual de conclusiones de una comunidad, un selector de comunidades en contraste para presentar la conclusión acerca de una comparación entre dos comunidades.

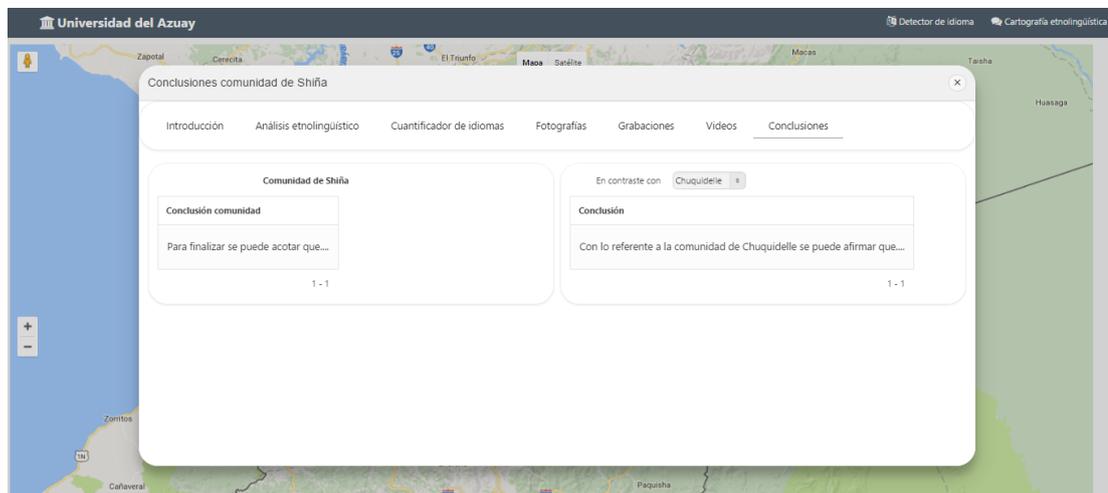


Ilustración 4.18: Aplicación de presentación - Conclusiones comunidad

4.3 Conclusiones del capítulo

Al terminar este capítulo, se obtiene el producto, una aplicación web para el manejo de información relacionada al estudio de comunidades cuyos alcances fueron definidos mediante el uso de los requisitos del software, su proceso de desarrollo administrado mediante la metodología de desarrollo de software "Scrum", los diseños propuestos en el capítulo 3 son la guía para que un programador emprenda con el proceso de codificación el cual está claramente definido mediante buenas prácticas que junto al uso de componentes comunes de interfaz de usuario hace que la aplicación web sea homogénea, la adecuación del entorno de trabajo para garantizar el proceso de codificación es un factor muy importante.

La aplicación desarrollada está lista para entrar a una etapa de pruebas aplicadas por el usuario final, estas pruebas generan resultados cualitativos positivos o negativos del proceso de desarrollo, de generarse resultados negativos se realizan cambios en el software para generar una positiva aceptación por el usuario, debido a que el proceso fue administrado mediante Scrum lo que implica que el usuario fue parte del equipo de trabajo no se esperan cambios fuertes, los futuros cambios serán acompañados de una retroalimentación del desarrollo para aplicarlos en este o en futuros proyectos.

Capítulo 5: Pruebas de integridad y presentación de los datos

La última etapa del proceso de desarrollo de software es la aplicación de pruebas, para ello se va a centrar en dos: pruebas de integridad que pretende que los datos administrados y presentados por el usuario sean los correspondientes a la información que representan, y las pruebas de presentación de los datos que significa que los datos sean presentados de una forma organizada y amistosa para el usuario final, los resultados positivos de estas pruebas prioritarias indican un buen proceso de desarrollo de software.

Las siguientes pruebas están tomadas de la metodología para el desarrollo de software IWeb (Pressman, 2010).

5.1 Pruebas semánticas

Busca la exactitud de la información presentada y la consistencia entre objetos de contenido y objetos relacionados. el examinador debe responder a las siguientes preguntas:

- ¿La información realmente es precisa (concisa y exacta)?
- ¿La plantilla del objeto de contenido es fácil de entender para el usuario?
- ¿La información anidada en un objeto de contenido se encuentra con facilidad?
- ¿La información presentada internamente es consistente con la información de otros objetos de contenido?
- ¿El contenido es ofensivo, engañoso o abre la puerta a pleitos?
- ¿El contenido infringe derechos de autor o marcas registradas?
- ¿El estilo estético del contenido entra en conflicto con el estilo estético de la interfaz global?

Componentes de administración	Correcto	Con cambio
Inicio	Si	
Comunidades (reporte)	Si	
Comunidad (mantenimiento)	Si	
Eventos (reporte)	Si	
Evento (mantenimiento)	Si	
Modelos (reporte)	Si	
Modelo (mantenimiento)	Si	
Puntos de análisis (árbol)	Si	
Punto de análisis (mantenimiento)	Si	
Info Comunidad (submenú)	Si	
Análisis comunidad (reporte)	Si	
Análisis comunidad (mantenimiento)	Si	
Información punto de análisis (mantenimiento)	Si	5.1.1.1
Analizar texto (mantenimiento)	Si	
Multimedia comunidades (reporte)	Si	
Multimedia comunidad (mantenimiento)	Si	
Contraste comunidades (reporte)	Si	
Contraste comunidad (mantenimiento)	Si	
Reportes (submenú)	Si	
Administración (submenú)	Si	
Usuario (reporte y mantenimiento)	Si	
Opción Rol (reporte y mantenimiento)	Si	

Tabla 5.1: Prueba semántica – Administración

Componentes de visitante	Correcto	Con cambios
Inicio	Si	5.1.2.1
Introducción a la comunidad	Si	
Análisis de comunidad	Si	
Cuantificador de idiomas	Si	
Fotografías	Si	
Grabaciones	Si	
Videos	Si	
Conclusiones	Si	

Tabla 5.2: Prueba semántica – Visitante

5.2 Pruebas sintácticas

Esta prueba está enfocada a verificar la ortografía y la gramática de cada una de las opciones de la aplicación.

Componentes de administración	Ortografía	Gramática	Cambio
Inicio	Si	Si	
Comunidades (reporte)	Si	Si	
Comunidad (mantenimiento)	Si	Si	
Eventos (reporte)	Si	Si	
Evento (mantenimiento)	Si	Si	
Modelos (reporte)	Si	Si	
Modelo (mantenimiento)	Si	Si	
Puntos de análisis (árbol)	Si	Si	
Punto de análisis (mantenimiento)	Si	Si	
Info Comunidad (submenú)	Si	Si	
Análisis comunidad (reporte)	Si	Si	
Análisis comunidad (mantenimiento)	Si	Si	
Información punto de análisis (mantenimiento)	Si	Si	
Analizar texto (mantenimiento)	Si	Si	
Multimedia comunidades (reporte)	Si	Si	
Multimedia comunidad (mantenimiento)	Si	Si	
Contraste comunidades (reporte)	Si	Si	
Contraste comunidad (mantenimiento)	Si	Si	
Reportes (submenú)	Si	Si	
Administración (submenú)	Si	Si	
Usuario (reporte y mantenimiento)	Si	Si	
Opción Rol (reporte y mantenimiento)	Si	Si	

Tabla 5.3: Prueba sintáctica – Administración

Componentes para visitante	Ortografía	Gramática	Observaciones
Inicio	Si	Si	
Introducción a la comunidad	Si	Si	
Análisis de comunidad	Si	Si	
Cuantificador de idiomas	Si	Si	
Fotografías	Si	Si	
Grabaciones	Si	Si	
Videos	Si	Si	
Conclusiones	Si	Si	

Tabla 5.4: Prueba sintáctica – Visitante

5.3 Pruebas de interfaz

Probar los mecanismos de interfaz que son usados por el usuario en la interacción con el sistema, como, por ejemplo: vínculos, formatos de páginas, y campos, validaciones, HTML dinámico y ventanas pop-up.

Componentes de administración	Navegadores		Con cambio
	Google Chrome	Mozilla	
Inicio	Si	Si	
Comunidades (reporte)	Si	Si	
Comunidad (mantenimiento)	Si	Si	
Eventos (reporte)	Si	Si	
Evento (mantenimiento)	Si	Si	
Modelos (reporte)	Si	Si	
Modelo (mantenimiento)	Si	Si	
Puntos de análisis (árbol)	Si	Si	
Punto de análisis (mantenimiento)	Si	Si	5.3.1.1 y 5.3.1.2
Info Comunidad (submenú)	Si	Si	
Análisis comunidad (reporte)	Si	Si	
Análisis comunidad (mantenimiento)	Si	Si	
Info. punto de análisis (mantenimiento)	Si	Si	5.3.1.3
Analizar texto (mantenimiento)	Si	Si	
Multimedia comunidades (reporte)	Si	Si	
Multimedia comunidad (mantenimiento)	Si	Si	
Contraste comunidades (reporte)	Si	Si	
Contraste comunidad (mantenimiento)	Si	Si	
Reportes (submenú)	Si	Si	
Administración (submenú)	Si	Si	
Usuario (reporte y mantenimiento)	Si	Si	
Opción Rol (reporte y mantenimiento)	Si	Si	

Tabla 5.5: Prueba de mecanismos de interfaz - Administración

Componentes de visitante	Navegadores		Con cambios
	Google Chrome	Mozilla	
Inicio	Si	Si	
Introducción a la comunidad	Si	Si	
Análisis de comunidad	Si	Si	5.3.2.1
Cuantificador de idiomas	Si	Si	
Fotografías	Si	Si	
Grabaciones	Si	Si	
Videos	Si	Si	
Conclusiones	Si	Si	

Tabla 5.6: Prueba de mecanismos de interfaz – Visitante

5.4 Pruebas de usabilidad

Permitir al usuario operarlo y controlarlo con facilidad.

Componentes de administración	Navegadores		Con cambio
	Google Chrome	Mozilla	
Inicio	Si	Si	
Comunidades (reporte)	Si	Si	
Comunidad (mantenimiento)	Si	Si	
Eventos (reporte)	Si	Si	
Evento (mantenimiento)	Si	Si	
Modelos (reporte)	Si	Si	
Modelo (mantenimiento)	Si	Si	
Puntos de análisis (árbol)	Si	Si	
Punto de análisis (mantenimiento)	Si	Si	
Info Comunidad (submenú)	Si	Si	
Análisis comunidad (reporte)	Si	Si	
Análisis comunidad (mantenimiento)	Si	Si	
Info. punto de análisis (mantenimiento)	Si	Si	5.4.1.1
Analizar texto (mantenimiento)	Si	Si	
Multimedia comunidades (reporte)	Si	Si	
Multimedia comunidad (mantenimiento)	Si	Si	
Contraste comunidades (reporte)	Si	Si	
Contraste comunidad (mantenimiento)	Si	Si	
Reportes (submenú)	Si	Si	
Administración (submenú)	Si	Si	
Usuario (reporte y mantenimiento)	Si	Si	
Opción Rol (reporte y mantenimiento)	Si	Si	

Tabla 5.7: Prueba de usabilidad - Administración

Componentes de visitante	Navegadores		Con cambio
	Google Chrome	Mozilla	
Inicio	Si	Si	
Introducción a la comunidad	Si	Si	
Análisis de comunidad	Si	Si	
Cuantificador de idiomas	Si	Si	
Fotografías	Si	Si	5.4.2.1
Grabaciones	Si	Si	
Videos	Si	Si	
Conclusiones	Si	Si	

Tabla 5.8: Prueba de usabilidad - Visitante

5.5 Cambios para pruebas fallidas

En su debido momento las pruebas de la aplicación resultaron fallidas, en este apartado se detalla todos los cambios que se realizaron en la aplicación para que sea aceptada por el usuario final.

Código	Opción	Autor	Observación	Solución
5.1.1.1	Comunidad (mantenimiento)	PV	No existe un lugar donde pueda presentar información de comunidad	Agregar campo de introducción comunidad.
5.1.2.1	Inicio (Visitante)	PV	Necesita poner sencillo resumen de cada comunidad	Agrandar la ventana de nombre comunidad en el mapa y agregar el resumen.
5.1.2.2	Introducción	PV	No existe un lugar donde pueda presentar información de comunidad	Crear página de presentación para la comunidad "Introducción".
5.3.1.1	Punto de análisis (mantenimiento)	CA	Botones Expandir/Contrair no cambian de posición al ejecutarlos.	Cambio de botones a un lugar común e inamovible.
5.3.1.2	Punto de análisis (mantenimiento)	PV	No le simpatiza la ventana popup para la selección de datos.	Cambiar todos los popups a selectores de opción.
5.3.1.3	Análisis comunidad (mantenimiento)	CA	Confunde la forma de crear un análisis comunidad.	Agregar un botón de crear árbol de análisis.
5.3.2.1	Análisis comunidad (mantenimiento)	CA	Botones Expandir/Contrair no cambian de posición al ejecutarlos.	Cambio de botones a un lugar común e inamovible.
5.4.1.1	Información punto de análisis (mantenimiento)	CA	Visualización de opciones "Información" y "Analizar texto" no es la adecuada	Cambio de estilo de botones que luzcan como links.
5.4.1.1	Fotografías	PV	Necesita visualizar imágenes relevantes de un evento de estudio.	Agregar un selector que filtre las imágenes por eventos.

Tabla 5.9: Cambios en pruebas fallidas

5.6 Conclusiones del capítulo

La aplicación de pruebas sobre el software es la mejor opción para verificar si el software es de calidad, si las pruebas resultantes por parte del usuario final son negativas no queda de otra que realizar una reestructuración atacando al punto fallido, algunas pruebas no son necesarias detallarlas pues son generadas y corregidas en el proceso de desarrollo, pero las pruebas con resultante negativo por parte del usuario final son las que generan una retroalimentación para no volver a cometer fallos que molesten al usuario en este o en otro proyecto de desarrollo de software.

Existe la gran ventaja de haber trabajado con la metodología de desarrollo Scrum la cual involucra al usuario final en el proceso de desarrollo, al tener la constante revisión del usuario final se enmiendan problemas que pueden ser fallidos en la etapa de pruebas, de no ser así, sería tedioso aplicar múltiples cambios en toda la aplicación. Los cambios graves que necesitan un largo tiempo para su resolución son prácticamente anulados. Con la aplicación de pruebas se finaliza el proceso de desarrollo de la aplicación web.

Conclusiones del trabajo

Al terminar con este trabajo de graduación, se afirma que la selección de las diferentes herramientas para el desarrollo de software: Oracle Database, Oracle Apex, Google Maps y GNU Aspell fueron las acertadas para el desarrollo de la aplicación web “Cartografía etnolingüística” que incluye minería de texto con el fin obtener conocimiento del comportamiento de la lengua en diferentes grupos lingüísticos. La integración de las herramientas antes mencionadas junto con los procedimientos aplicados para obtener el origen idiomático en la minería de texto causa un bajo coste de desarrollo y buscan generar un software de calidad.

Mediante el estándar IEEE-830 para especificación de requisitos de software se identificó claramente todas las necesidades a desarrollar conocidas también como “requisitos funcionales”, permitió también definir las interfaces de software y de usuario, esto dispuso al programador a la indagación de métodos usados en otras aplicaciones que faciliten la producción de este. La metodología de desarrollo de software “Scrum” permitió identificar de forma organizada, sencilla y puntual el proceso de desarrollo para obtener su máximo rendimiento.

El diseño del servicio web basado en un modelo descriptivo de minería de texto que trabaja a partir del reconocimiento de patrones, es el encargado de realizar la clasificación de los átomos de los textos, los resultados de esta clasificación almacenados de manera estructurada permiten la ejecución de diversas consultas SQL.

En la actualidad como apoyo al investigador de comportamiento de lengua están propuestas las siguientes consultas: las “Palabras en uso” una consulta dispuesta en una gráfica tipo pastel que permite visualizar el porcentaje de uso de cada idioma en los textos de cada comunidad y permitiendo identificar el comportamiento lingüísticos del grupo en estudio, el “Top 5

palabras" consulta dispuesta en una lista que apoya a los investigadores a visualizar el patrón de palabras repetidas en los diferentes eventos de estudio; Finalmente, "Contraste comunidades" que es una comparativa entre comunidades mediante una gráfica de barras de la cantidad de palabras pertenecientes a cada idioma. Estas son las tres consultas básicas propuestas, pero cabe recalcar que se puede agregar múltiples consultas debido a que la clasificación se encuentra almacenada estructuradamente.

El diseño del modelo entidad/relación permitió visualizar y estructurar el almacenamiento de los datos, tanto de la aplicación en general como de los resultados de la clasificación. El diseño de capas del sistema distribuido es necesario pues el servicio web de clasificación de textos es externo a la aplicación en desarrollo, por ende, es necesario adquirir la lógica del funcionamiento.

Previo a la etapa de codificación se dispuso a adecuar el entorno de desarrollo, además, establecer las convenciones de codificación en el proyecto que permitieron estructurar de manera efectiva la escritura de los códigos fuente. La codificación define la materialización de la aplicación y así cumplir con los objetivos obtenidos mediante el estándar IEEE-830 para la aplicación web, la implantación de la minería de texto para obtener conocimiento acerca del comportamiento de la lengua es el eje central de la codificación, pues los métodos de clasificación y cuantificación de palabras no deben ser un pretexto que ponga en riesgo el rendimiento de la aplicación web.

Previo a la puesta en operación, la etapa de pruebas, brinda la verificación de calidad del software centrándose en satisfacer la característica de usabilidad del modelo de calidad ISO25010 junto con sus sub-características más relevantes, la prueba de mecanismo de interfaz resalta la aceptación de la aplicación web por parte del usuario final.

Gestionar el desarrollo de una aplicación web que incluya minería de texto mediante procesos de ingeniería de software, garantiza la adecuada construcción y así presentar un software de buena calidad para el usuario final.

Referencias

- Argis. (15 de 06 de 2016). *Arcgis Features*. Obtenido de <https://www.arcgis.com/features/features.html>
- Atkinson, K. (01 de 11 de 2002). *GNU Aspell*. Obtenido de <http://ex.osaka-kyoiku.ac.jp/~fujii/dl/aspell.pdf>
- Basit Raza, A. M. (2016). Autonomicity in Oracle Database Management System . *IEEE Xplore*, 1-5.
- Charu Aggarwal, C. Z. (2012). *Text Mining Data*. New York: Springer Science.
- DB-Engines. (14 de 06 de 2016). *DB-Engines*. Obtenido de DB-Engines: <http://db-engines.com/en/ranking>
- Departamento de Informática. (05 de 07 de 2016). *Universidad de Valladolid*. Obtenido de Universidad de Valladolid: https://www.infor.uva.es/~fdiaz/sd/2005_06/doc/SD_TE02_20060305.pdf
- Glassfish. (05 de 12 de 2016). *Glassfish server documentation*. Obtenido de Glassfish server documentation: <https://glassfish.java.net/documentation.html>
- Google Charts. (04 de 12 de 2016). *Google Charts*. Obtenido de Google Charts: <https://developers.google.com/chart/>
- Google Maps. (15 de 06 de 2016). *Google Maps Feature*. Obtenido de Google Maps: <https://support.google.com/maps/answer/144349?hl=en>
- Here Maps. (15 de 06 de 2016). *Here Maps*. Obtenido de Here Maps: <https://help.here.com/android/maps/>
- ISO 25000 Standards. (01 de 07 de 2016). *ISO/IEC 25010*. Obtenido de ISO/IEC 25010: <http://www.iso25000.com/index.php/en/iso-25000-standards/iso-25010>

- Li, S. (2011). A Method for Building Thematic Map of GIS Based on Google Maps API. *IEEE Xplore*, 1-4.
- Lopez Perez C, S. G. (2007). *Minería de datos. Técnicas y herramientas*. Madrid: Thomson Ediciones Paraninfo S.A.
- Marini, E. (01 de 10 de 2012). *Modelo Cliente-Servidor*. Obtenido de Modelo Cliente-Servidor: <https://www.linuxito.com/docs/el-modelo-cliente-servidor.pdf>
- Microsoft Corporation. (14 de 06 de 2016). *Developer Reference for SQL Server 2014*. Obtenido de Developer Reference for SQL Server 2014: [https://msdn.microsoft.com/en-us/library/dd206988\(v=sql.120\).aspx](https://msdn.microsoft.com/en-us/library/dd206988(v=sql.120).aspx)
- MySQL. (14 de 06 de 2016). *MySQL 5.7 Reference Manual*. Obtenido de MySQL: <http://dev.mysql.com/doc/refman/5.7/en/>
- Oracle Apex. (15 de 06 de 2016). *Oracle Apex*. Obtenido de Oracle Corporation: <http://www.oracle.com/technetwork/testcontent/what-is-apex-099128.html>
- Oracle Base. (23 de 08 de 2016). *Oracle Base Installation Oracle 11gr2*. Obtenido de Oracle Base Installation Oracle 11gr2: <https://oracle-base.com/articles/11g/oracle-db-11gr2-installation-on-oracle-linux-5>
- Oracle Blogs. (01 de 12 de 2016). *Why Use REST*. Obtenido de Why Use REST: https://blogs.oracle.com/newgendbaccess/entry/why_you_should_turn_your
- Oracle Corporation. (13 de 06 de 2016). *Familia de productos Oracle Database 11g*. Obtenido de Oracle Database: <http://www.oracle.com/technetwork/es/database/enterprise-edition/documentation/productos-oracle-database-11g-2247590-esa.pdf>
- Oracle Help Center: Text mining. (01 de 01 de 2016). *Oracle Help Center: Text mining*. Obtenido de Oracle Help Center: Text mining: https://docs.oracle.com/cd/B28359_01/datamine.1111/b28129/text.htm

- Oracle Workspace. (05 de 12 de 2016). *Oracle Apex*. Obtenido de https://docs.oracle.com/cd/E11882_01/appdev.112/e12512/adm_create_wkspc.htm#AEADM239
- Paute, A. (2016). *Análisis y Desarrollo del módulo de manufactura de un Sistema ERP*. Cuenca.
- Postgresql. (12 de 06 de 2016). *Postgresql*. Obtenido de Postgresql Organization: <https://www.postgresql.org/docs/9.5/static/sql.html>
- Pressman, R. S. (2010). *Ingeniería del software*. Mexico DF: The McGraw-Hill.
- R. Ohbuchi, H. U. (2002). Robust watermarking of vector digital maps. *IEEE Xplore*, 1.
- Real academia española. (07 de 07 de 2016). *Real academia española*. Obtenido de Real academia española: <http://dle.rae.es/>
- Shin, S. (2015). Phoneme based realtime taboo words similarity browsing system of new words using multi-lingual taboo words databases in web environments . *IEEE Xplore*, 1-2.
- The Unified Modeling Language. (04 de 12 de 2016). *Unified Modeling Language*. Obtenido de Unified Modeling Language: <http://www.uml-diagrams.org/>
- Thomas Connolly, C. B. (2005). *Database Systems*. England: Pearson Education Limited.
- Tripod. (04 de 12 de 2016). *Minería de textos Tripod*. Obtenido de Sistemas Avanzados de Recuperación de la Información: <http://mineriadetextos.tripod.com/>
- Wojciechowski, T. (2011). Transaction system with reporting capability in a web based data warehouse application developed in Oracle Application Express. *IEEE Xplore*, 1-4.
- Zhu, Y. (2012). Introducing Google Chart Tools and Google Maps API in Data Visualization Courses. *IEEE Xplore*, 1-9.

Anexos

Anexo 1: Guía de instalación de Oracle Database 11gR2

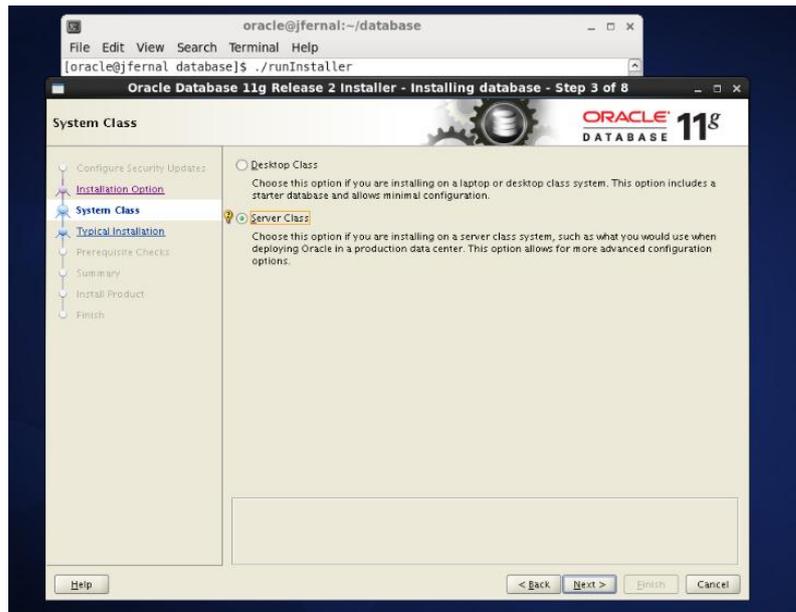
La siguiente guía de instalación de Oracle Database es para el sistema operativo CentOS 6.8, basado en (Oracle Base, 2016).

Configuración de prerequisites:

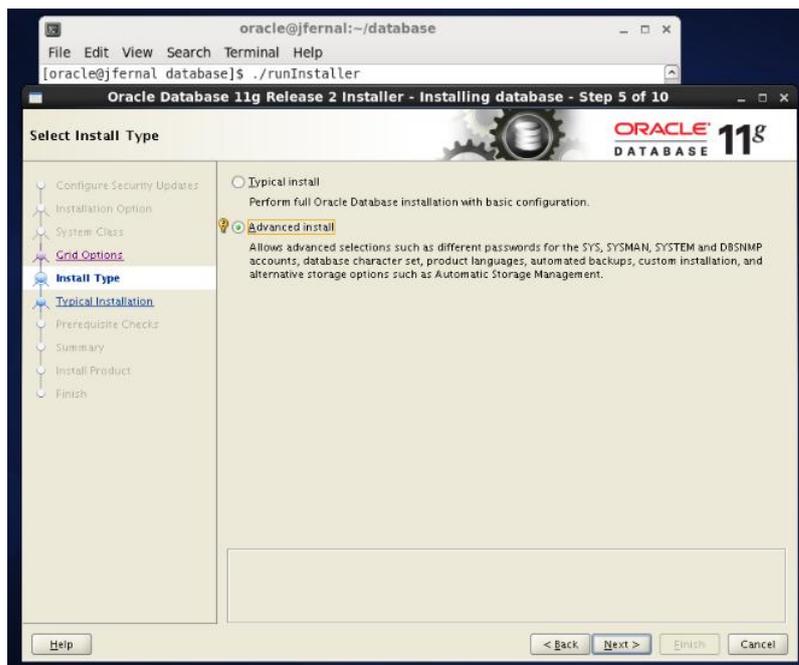
1. Obtener los archivos de Oracle Database 11gR2
 - <http://www.oracle.com/technetwork/database/enterprise-edition/downloads/index.html>
2. Descomprimir los archivos
 - `unzip linux_11gR2_database_1of2.zip`
 - `unzip linux_11gR2_database_2of2.zip`
3. Configurar el nombre del servidor y asegurar la ip del servidor
 - `nano /etc/hosts`
4. Configurar los nuevos grupos y usuarios par Oracle Database
 - `groupadd oinstall`
 - `groupadd dba`
 - `groupadd oper`
 - `groupadd asmadmin`
 - `useradd -g oinstall -G dba,oper,asmadmin Oracle`
 - `passwd Oracle`

Configuración de Oracle Database

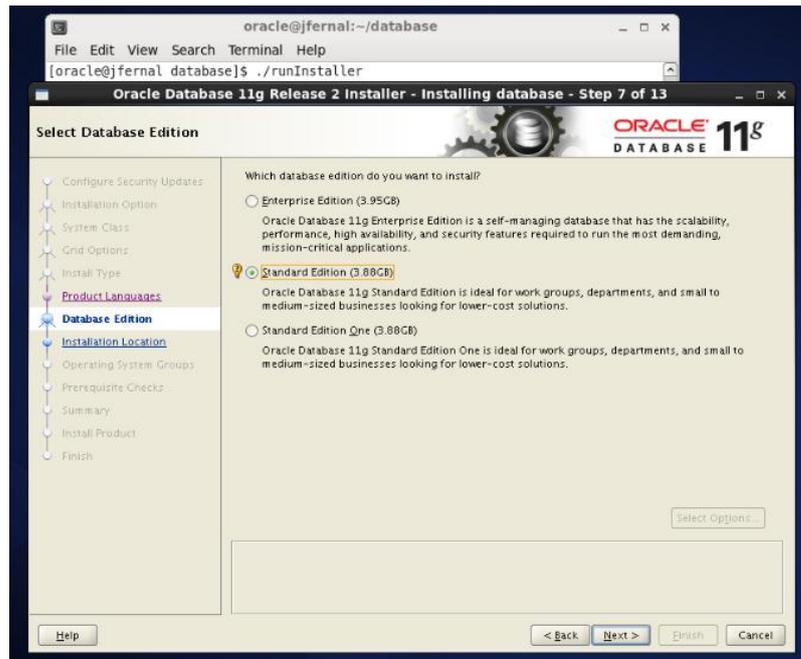
1. Ejecutar el instalador
 - `./runInstaler`
2. Configurar los diferentes parámetros
 - Clase de servidor



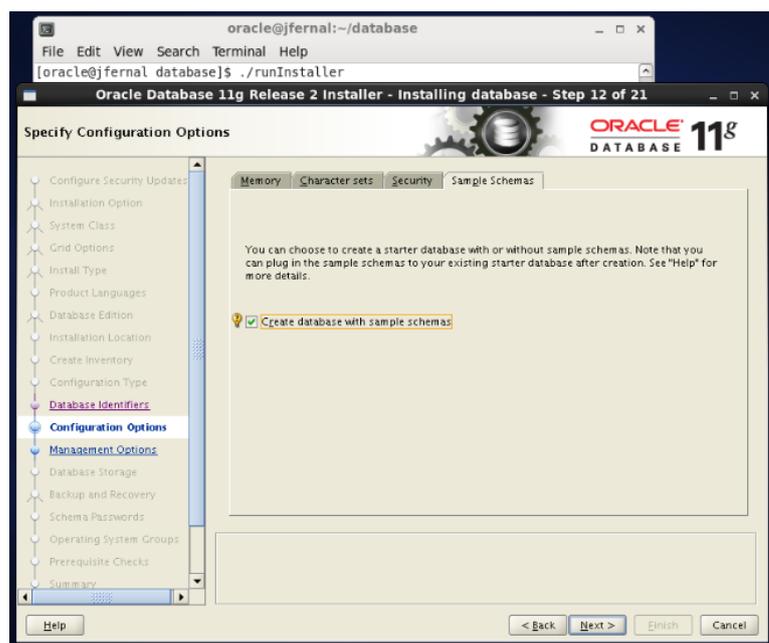
- Tipo de instalación



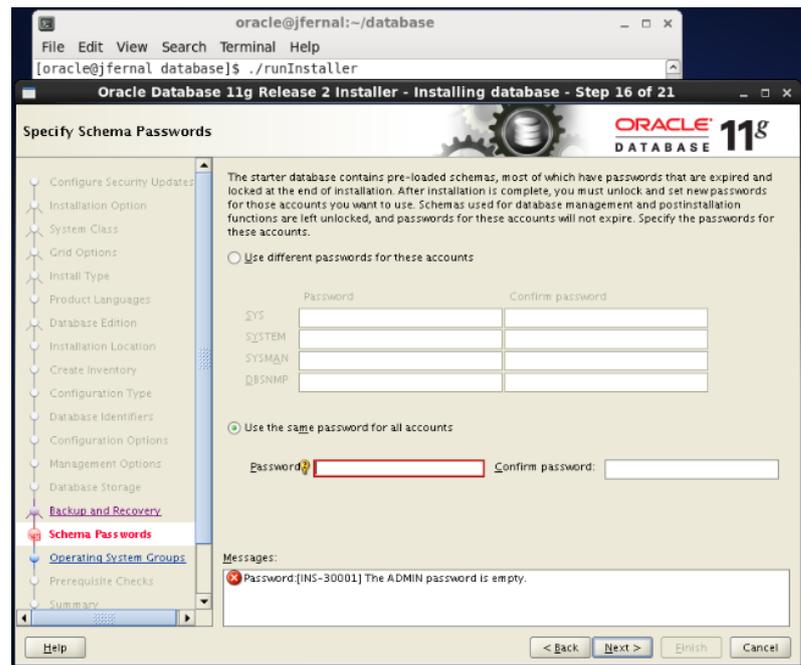
- La edición de la base de datos



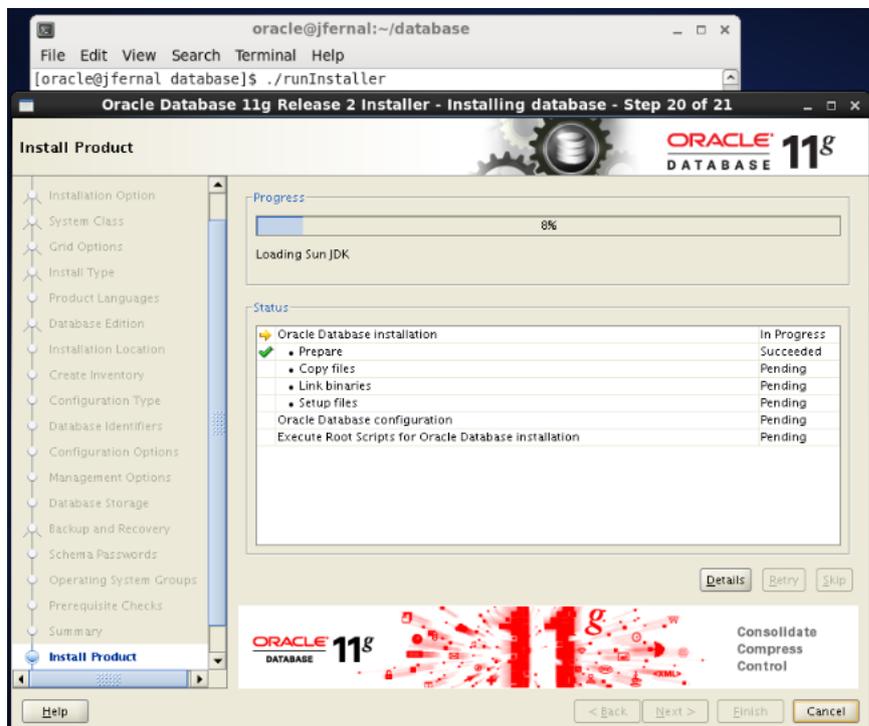
- Directorios para la configuración de Oracle Database
- El tipo de base de datos (Propósito general)
- Nombre de la base de datos
- Nombre identificador del servicio - SID
- Opciones de configuración
 - Memoria
 - Juego de caracteres
 - Seguridad
 - Esquemas de ejemplo



- Password de la base de datos



- Grupos SYSDBA y OSOPER
- Finalmente, instalación de la base con los parámetros



Anexo 2: Guía de actualización de Oracle Apex 5.0.4

Oracle Apex ya se encuentra instalado en cualquier versión de Oracle Database, pero el inconveniente es que se encuentra en una versión anticuada, para ello lo aconsejable es actualizar a la última versión estable, lo que permite ejecutar aplicaciones desarrolladas en cualquier versión anterior a esta de Oracle Apex 5.0.4, a continuación, se detalla el proceso de actualización que es aplicable a cualquier versión de sistema operativo.

1. Obtener los archivos de actualización de:
 - <http://www.oracle.com/technetwork/developer-tools/apex/downloads/download-085147.html>
2. Descomprimirlo en un directorio relevante:
 - `unzip apex_5.0.4.zip`
3. Mediante consola ejecutar los siguientes comandos
 1. Ingresar al Shell de Oracle e ingresar
 - `Sqlplus`
 2. Crear tablespace adecuados para APEX
 - `create tablespace ts_apex05 datafile
'/u01/app/oracle/oradata/istdb/df_apex5_01.dbf' size
100m autoextend on;`
 3. Instalar APEX en el tablespace
 - `@apexins.sql TS_APEX05 TS_APEX05 TEMP /i/`
 4. Ingresar clave de administrador de Apex
 - `@apxchpwd.sql`
 5. Ejecutar el gateway de Oracle
 - `@apex_epg_config.sql<directorio contiene folder Apex
nuevo>`
 6. Instalar las imágenes de Apex
 - `@apxldimg.sql <directorio contiene folder Apex nuevo>`
 7. Activar los usuarios para gestión de Apex
 - `alter user anonymous account unlock;`
 - `alter user xdb account unlock;`
 - `alter user apex_public_user account unlock;`
 - `alter user flows_files account unlock;`
 8. Habilitar el puerto para Apex

- `exec dbms_xdb.sethttpport(8080);`

9. Activar conexión local para Apex

- `exec dbms_xdb.setlistenerlocalaccess(l_access => false);`

10. Instalar los usuarios Rest

- `@apex_rest_config.sql`

11. Ejecutar la lista de control de acceso básica tomado de tutoriales básicos de Oracle (Oracle Base, 2016)

```

declare
  acl_path varchar2(4000);
begin
  -- look for the acl currently assigned to '*' and give apex_050000
  -- the "connect" privilege if apex_050000 does not have the
  privilege yet.

  select acl into acl_path from dba_network_acls
  where host = '*' and lower_port is null and upper_port is null;

  if dbms_network_acl_admin.check_privilege(acl_path, 'apex_050000',
    'connect') is null then
    dbms_network_acl_admin.add_privilege(acl_path,
    'apex_050000', true, 'connect');
  end if;

exception
  -- when no acl has been assigned to '*'.
  when no_data_found then
    dbms_network_acl_admin.create_acl('power_users.xml',
    'acl that lets power users to connect to everywhere',
    'apex_050000', true, 'connect');
    dbms_network_acl_admin.assign_acl('power_users.xml', '*');
end;
/
commit;

```

12. Ejecutar una prueba de acceso mediante la URL

- <http://localhost:8080/apex>

Anexo 3: Guía de Instalación de Java JDK 1.7.0

El kit de desarrollo de java puede instalarse en el sistema operativo CentOS 6.8 mediante dos opciones:

1. Mediante "yum"
 - Ejecutar el comando:
 - `yum install java-1.7.0-openjdk-devel`
2. Mediante el archivo tar.gz
 - Descargar el archivo del jdk comprimido de:
 - <http://www.oracle.com/technetwork/es/java/javase/downloads/jdk7-downloads-1880260.html>
 - Descomprimir el archivo con:
 - `tar -zxvf jdk-7u79-linux-i586.tar.gz`
 - Abrir el archivo de variables de entorno
 - `nano .bash_profile`
 - Agregar las siguientes líneas
 - `export JAVA_HOME=<directorio_descompimido_de_java>`
 - `export PATH=/usr/java/jdk1.7.0_21/bin:$PATH`

Finalmente, verificar la ejecución de un comando en java

```
File Edit View Search Terminal Help
[root@jfernal ~]# java -version
java version "1.7.0_121"
OpenJDK Runtime Environment (rhel-2.6.8.1.el6_8-i386 u121-b00)
OpenJDK Client VM (build 24.121-b00, mixed mode, sharing)
[root@jfernal ~]# █
```

Anexo 4: Exportado de análisis de comunidad

Un documento digital exportado de la aplicación con extensión pdf.

UNIVERSIDAD DEL AZUAY

Comunidad de: Shiña
En el escenario de: Evento 2: Diálogo en el parque
Bajo el modelo de: Guía de estudio etnográfico del uso del habla de Sherzer y Darnell

TRANSCRIPCIÓN DE AUDIO

Evento 2

1. Análisis del uso del habla

Las variedades lingüísticas de Shiña son el castellano y el quichua; la primera se utiliza dentro y fuera de

2. Actitudes respecto del uso del habla

Las actitudes de los pobladores de Shiña hacia la lengua es conflictiva ; por un lado, ellos creen que su

3. Adquisición de la competencia del habla

Los eventos de habla seleccionados evidenciaron ciertos elementos lingüísticos interesantes como

3.1 Existencia de una teoría nativa explícita respecto a la adquisición de la competencia del habla

Por otra parte, también notamos una mezcla entre el castellano y el quichua en el evento 4, segmento de

3.2 Trasmisión de las habilidades del habla

Otro aspecto interesante que observamos es el uso del diminutivo, en el evento 3, audio: relato de vida.

4. Uso del habla en la educación y el control social

Otra categoría de análisis que consideramos es la educación y el control social, en la comunidad de

5. Generalizaciones Tipológicas

Finalmente, en cuanto a las generalizaciones tipológicas, en los eventos del habla es posible encontrar

5.1 Pautas generales acerca de los usos del habla y el análisis de una cultura

Entonces, estos dos niveles son el explícito que representan las creencias sobre la lengua, en el caso de

5.3 Pautas del uso del habla características de tipos particulares de las comunidades del habla

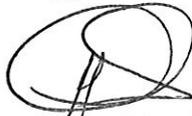
En el evento 4 presentado en este artículo es posible encontrar:

Doctora Jenny Ríos Coello, Secretaria de la Facultad de Ciencias de la Administración de la Universidad del Azuay

CERTIFICA:

Que, el Consejo de Facultad en sesión del 10 de junio de 2016, conoció la petición del estudiante **JUAN FERNANDO LIMA SIGUA** con código 62104, quien presenta su trabajo de titulación denominado: "**ANALISIS DE LENGUA MEDIANTE MINERIA DE TEXTO**", previa a la obtención del título de Ingeniero de Sistemas y Telemática. El Consejo de Facultad acoge el informe de la Junta Académica y aprueba el diseño. Designa como *Directora a la ingeniera Catalina Astudillo Rodríguez* y como miembros del *Tribunal Examinador a los ingenieros Francisco Salgado Arteaga y Marcos Orellana Cordero*. El peticionario tienen un plazo de **SEIS MESES** para presentar su trabajo de titulación, esto es hasta el 10 de diciembre de 2016.

Cuenca, junio 15 de 2016



Dra. Jenny Ríos Coello
Secretaria de la Facultad de
Ciencias de la Administración

UNIVERSIDAD DEL
AZUAY
FACULTAD DE
ADMINISTRACION
SECRETARIA



ACTA

SUSTENTACIÓN DE PROTOCOLO/DENUNCIA DEL TRABAJO DE TITULACIÓN

- 1.1 Nombre del estudiante: Juan Fernando Lima Sigua
- 1.2 Códigos: 62104
- 1.3 Director sugerido: Ing. Catalina Astudillo Rodríguez
- 1.4 Codirector (opcional): _____
- 1.5 Tribunal: Ing. Francisco Salgado Arteaga e Ing. Marcos Orellana Cordero
- 1.6 Título propuesto: *“ANALISIS DE LENGUA MEDIANTE MINERIA DE TEXTO”*
- 1.7 Resolución:

1.7.1 Aceptado sin modificaciones

1.7.2 Aceptado con las siguientes modificaciones:

1.7.3 No aceptado

- Justificación:

Tribunal

Ing. Catalina Astudillo Rodríguez

Ing. Francisco Salgado Arteaga

Ing. Marcos Orellana Cordero

Sr. Juan Fernando Lima Sigua

Dra. Jénny Ríos Coello
Secretaria de Facultad

Fecha de sustentación: Miércoles, 25 de mayo de 2016 a las 07h30 .



RÚBRICA PARA LA EVALUACIÓN DEL PROTOCOLO DE TRABAJO DE TITULACIÓN

- 1.1 Nombre del estudiante: Juan Fernando Lima Sigua
 1.2 Códigos: 62104
 1.3 Director sugerido: Ing. Catalina Astudillo Rodríguez
 1.4 Codirector (opcional):
 1.5 Título propuesto: "ANÁLISIS DE LENGUA MEDIANTE MINERÍA DE TEXTO"
 1.6 Revisores (tribunal): Ing. Francisco Salgado Arteaga e Ing. Marcos Orellana Cordero
 1.7 Recomendaciones generales de la revisión:

	Cumple totalmente	Cumple parcialmente	No cumple	Observaciones (*)
Línea de investigación				
1. ¿El contenido se enmarca en la línea de investigación seleccionada?	✓			
Título Propuesto				
2. ¿Es informativo?	✓			
3. ¿Es conciso?	✓			
Estado del arte				
4. ¿Identifica claramente el contexto histórico, científico, global y regional del tema del trabajo?	✓			
5. ¿Describe la teoría en la que se enmarca el trabajo	✓			
6. ¿Describe los trabajos relacionados más relevantes?	✓			
7. ¿Utiliza citas bibliográficas?	✓			
Problemática y/o pregunta de investigación				
8. ¿Presenta una descripción precisa y clara?	✓			
9. ¿Tiene relevancia profesional y social?	✓			
Hipótesis (opcional)				
10. ¿Se expresa de forma clara?				NA
11. ¿Es factible de verificación?				NA
Objetivo general				
12. ¿Concuerda con el problema formulado?	✓			
13. ¿Se encuentra redactado en tiempo verbal infinitivo?	✓			
14. ¿Se encuentra redactado en tiempo verbal infinitivo?	✓			

Objetivos específicos				
15.¿Concuerdan con el objetivo general?	✓			
16.¿Son comprobables cualitativa o cuantitativamente?	✓			
Metodología				
17.¿Se encuentran disponibles los datos y materiales mencionados?	✓			
18.¿Las actividades se presentan siguiendo una secuencia lógica?	✓			
19.¿Las actividades permitirán la consecución de los objetivos específicos planteados?	✓			
20.¿Los datos, materiales y actividades mencionadas son adecuados para resolver el problema formulado?	✓			
Resultados esperados				
21.¿Son relevantes para resolver o contribuir con el problema formulado?	✓			
22.¿Concuerdan 23.con los objetivos específicos?	/			
24.¿Se detalla la forma de presentación de los resultados?	/			
25.¿Los resultados esperados son consecuencia, en todos los casos, de las actividades mencionadas?	/			
Supuestos y riesgos				
26.¿Se mencionan los supuestos y riesgos más relevantes?	/			
27.¿Es conveniente llevar a cabo el trabajo dado los supuestos y riesgos mencionados?	/			
Presupuesto				
28.¿El presupuesto es razonable?	/			
29.¿Se consideran los rubros más relevantes?	/			
Cronograma				
30.¿Los plazos para las actividades son realistas?	/			
Referencias				
31.¿Se siguen las recomendaciones de normas internacionales para citar?	/			
Expresión escrita				
32.¿La redacción es clara y fácilmente comprensible?	/			
33.¿El texto se encuentra libre de faltas ortográficas?	/			

(*) Breve justificación, explicación o recomendación.

- Opcional cuando cumple totalmente,
- Obligatorio cuando cumple parcialmente y NO cumple.

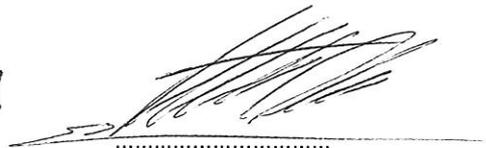
.....
.....
.....



.....
Ing. Catalina Astudillo Rodríguez



.....
Ing. Francisco Salgado Arteaga



.....
Ing. Marcos Orellana Cordero

Cuenca, 16 de diciembre de 2016.- Decano de la Facultad de Ciencias de la Administración, con autorización amplia y suficiente concedida por el Consejo de Facultad en sesión del 25 de febrero de 2016, conoce la petición del estudiante, **JUAN FERNANDO LIMA SIGUA** con código 62104, quien solicita prórroga para la presentación del trabajo de titulación: **“ANALISIS DE LENGUA MEDIANTE MINERIA DE TEXTO”**, presentado como requisito previo a la obtención del título de Ingeniero de Sistemas y Telemática, cuya fecha de presentación es el 10 de diciembre de 2016; *considerando el literal b) del Art. 6 del Instructivo para la Conformación y Funcionamiento de las Unidades de Titulación Especial en las Carreras de Grado de la Universidad del Azuay, aprobado por el Consejo Universitario el 05 de mayo de 2015, resuelve aprobar la solicitud y conceder una prórroga de seis meses adicionales, esto es hasta el 10 de junio de 2017.*

Cuenca, diciembre 16 de 2016.



Ing. Xavier Ortega Vásquez
Decano de la Facultad de
Ciencias de la Administración





CONVOCATORIA

En mi calidad de Decano de la Facultad de Ciencias de la Administración, convoco a los Miembros del Tribunal Examinador, a la sustentación del Protocolo del Trabajo de Titulación: **“ANALISIS DE LENGUA MEDIANTE MINERIA DE TEXTO”**, presentado por el estudiante Juan Fernando Lima Sigua con código 62104, previa a la obtención del grado de Ingeniero de Sistemas y Telemática, para el Miércoles, 25 de mayo de 2016 a las 07h30.

Cuenca, 18 de mayo de 2016

Ing. Xavier Ortega Vásquez
Decano de la Facultad

Ing. Catalina Astudillo Rodríguez

Ing. Francisco Salgado Arteaga

X

Ing. Marcos Orellana Cordero

Comunicado
20/05/2016
11:28AM



Oficio Nro. 055-2016-DIST-UDA

Cuenca, 10 de Mayo de 2016

Señor Ingeniero
Xavier Ortega Vázquez
DECANO DE LA FACULTAD DE CIENCIAS DE LA ADMINISTRACIÓN
Presente.-

De mis consideraciones:

La Junta Académica de la Escuela de Ingeniería de Sistemas y Telemática, reunida el día 10 de mayo del 2016, recibió el proyecto de tesis titulado "Análisis de lengua mediante minería de texto", presentado por el estudiante Juan Fernando Lima Sigua, estudiante de la Escuela de Ingeniería de Sistemas y Telemática, y revisado por la Ingeniera Catalina Astudillo, previo a la obtención del título de Ingeniero de Sistemas y Telemática.

Por lo expuesto, y de conformidad con el Reglamento de Graduación de la Facultad, recomienda como director y responsable de aplicar cualquier modificación al diseño del trabajo de graduación posterior a la Ingeniera Catalina Astudillo y como miembros del Tribunal a Francisco Salgado Ph.D. e Ingeniero Marcos Orellana.

Atentamente,

Ing. Marcos Orellana Cordero
Director Escuela de Ingeniería de Sistemas y Telemática
Universidad del Azuay

Oficio Nro. 075-2016-DIST-UDA

Cuenca, 1 de junio de 2016

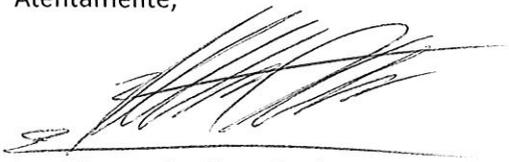
Señor Ingeniero
Xavier Ortega Vázquez
DECANO DE LA FACULTAD DE CIENCIAS DE LA ADMINISTRACIÓN
Presente.-

De mis consideraciones:

La Junta Académica de la Escuela de Ingeniería de Sistemas y Telemática, reunida el día 1 de junio del 2016, revisó la documentación del proyecto de tesis denominado "Análisis de lengua mediante minería de texto", presentado por Juan Fernando Lima, estudiante de la Escuela de Ingeniería de Sistemas y Telemática, y revisado por la Ing. Catalina Astudillo, previo a la obtención del título de Ingeniero de Sistemas y Telemática.

La Junta considera que la documentación cumple con las normas legales y reglamentarias de la Universidad y de la Facultad de Ciencias de la Administración y avala la aprobación por parte del tribunal designado, así por su digno intermedio, el conocimiento y aprobación por parte del Consejo de Facultad.

Atentamente,



Ing. Marcos Orellana Cordero
Director Escuela de Ingeniería de Sistemas y Telemática
Universidad del Azuay

DOCTORA JENNY RIOS COELLO SECRE-
TARIA DE LA FACULTAD DE CIENCIAS
DE LA ADMINISTRACION DE LA UNIVER-
SIDAD DEL AZUAY.

CERTIFICA:

Que, el señor **Juan Fernando Lima Sigua**, registrado con código 62104 alumno de la Escuela de Ingeniería de Sistemas y Telemática, tiene aprobado más del 80 % de su plan de estudios, le falta aprobar las siguientes materias: Metodología de la Investigación, Ingeniería de Software II, Calidad de Software, Proyectos Telemáticos, Sistema de Información Gerencial, Producción II, para egresar.

Cuenca, Marzo 21 del 2016



No. Derecho 046014
rgp.-

Cuenca, 6 de Mayo de 2016

Señor Ingeniero
Xavier Ortega Vázquez
DECANO DE LA FACULTAD DE CIENCIAS DE LA ADMINISTRACIÓN
Presente.-

De mis consideraciones:

Luego de revisar el diseño del trabajo de graduación denominado "Análisis de lengua mediante minería de texto", presentado por el señor Juan Fernando Lima, estudiante de la Escuela de Ingeniería de Sistemas y Telemática, considero que el documento cumple con las normas legales y reglamentarias de la Universidad y de la Facultad de Ciencias de la Administración, por lo que recomiendo su aprobación por parte del Consejo de Facultad.

Atentamente,



Ing. Catalina Astudillo
Director

Cuenca, 10 de Mayo de 2016

Ingeniero

Xavier Ortega

Decano de la Facultad de Ciencias de la Administración

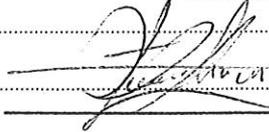
De mis consideraciones:

Yo, Juan Fernando Lima Sigua, estudiante de la carrera de Ingeniería de Sistemas y Telemática, con código 62104, solicito mediante la presente se sirva disponer el trámite para la aprobación del diseño del trabajo de titulación denominado "Análisis de lengua mediante minería de texto", previo a la obtención del título de Ingeniero de Sistemas y Telemática.

Adjuntamos el diseño correspondiente, que ha sido avalado por el Director propuesto y validado por la Junta Académica de la Escuela.

Con sentimientos de gratitud y estima.

Atentamente,



Juan Fernando Lima Sigua

Universidad del Azuay
Ingeniería de Sistemas y Telemática
Metodología de la Investigación
Guía para la presentación de la propuesta de proyecto de titulación (capstone project)

1. Datos generales

1.1. Nombre del estudiante: Lima Sigua Juan Fernando

1.1.1. Código: 62104

1.1.2. Contacto: teléfonos: 2885338, 0995718800, jfernal@outlook.com

1.2. Director sugerido: Astudillo Rodríguez, Catalina Mgst.

1.2.1. Contacto: cvastudillo@uazuay.edu.ec

1.3. Co-director sugerido: Orellana Cordero, Marcos Mgst.

1.3.1. Contacto: correo electrónico: marore@uazuay.edu.ec

1.4. Asesor metodológico: Salgado Arteaga, Francisco PhD

1.5. Tribunal designado: (de acuerdo a la normativa interna de cada Facultad).

1.6. Aprobación: fecha de Junta Académica y fecha de Consejo Facultad.

1.7. Línea de Investigación de la carrera: 1203 Informática de Computadoras

1.7.1. Código UNESCO: 1203.18 Sistemas de Información

1.7.2. Tipo de trabajo: Proyecto Integrador

1.8. Área de estudio: Bases de Datos, Geomática, Software, Lenguaje

1.9. Título propuesto: Análisis de lengua mediante minería de texto.

1.10. Subtítulo: Caso de estudio etnolingüística en el Azuay

1.11. Estado del proyecto: Nuevo proyecto de desarrollo.

2. Contenido

2.1. Motivación de la investigación:

Contribuir con software de apoyo para la tesis doctoral de la Mgst. Priscila Verdugo, adquirir los conocimientos de construcción de una aplicación con las características de geomática y elaborar un proyecto para la titulación de la carrera.

2.2. Problemática:

La búsqueda de las distintas etnias y su comportamiento lingüístico en la provincia de Azuay es un tema de investigación científico abordado por la Mgst Priscila Verdugo en su tesis Doctoral. El software con minería de texto que distinguirá los fonemas y el tipo de lenguaje usados en el apartado "historias de vida" de cada localidad, para mejorar la visualización de cada localidad, es necesario un mapa que situé las diferentes localidades con la información de la misma.

2.3. Resumen:

Una aplicación con capacidad de reconocer el cambio de fonemas, las distinciones del idioma, además, presentar datos e información de grupos sociales mediante el uso de mapas es una necesidad propuesta en la etnolingüística. Una base de datos mediante el uso de minería de texto junto con la geomática es la base fundamental que cubre la necesidad etnolingüística básica. Como resultado busca mejorar el análisis de los diferentes grupos sociales y su relación con el lenguaje.

2.4. Indagación exploratoria y base conceptual:

El diseño de la aplicación y la Base de Datos junto con la gestión proporcionada por el DBMS, es fundamental para cumplir con los objetivos planteados, por tanto, siguiendo las recomendaciones de propuestas por Raza, B [1]:

- Auto optimización
- a) Optimizar las consultas, en Oracle DBMS el tuning de la actuación en los 4 pasos propuestos.
 1. Identificar las consultas de alta carga.
 2. Generar un plan de optimización.
 3. Mejorar el plan de ejecución de consultas (QEP).
 4. Finalmente un generar un mejor QEP.
- b) Componente para colección de estadísticas automáticos.

Oracle genera diferentes tipos de estadísticas como: de la base de datos, las del sistema operativo y de interpretación.
- c) Monitor automático de diagnóstico de la base de datos
Las actividades incluyen SQL tuning, espacio de fragmentación de análisis, cuellos de botella, adaptar las estructuras de memoria, alta carga de ejecución de los requests.
- d) Repositorio automático de carga de trabajo (AWR).

Es un snapshot que almacena temporalmente



- Auto Configuración

La configuración es la característica para configurar las metas definidas.

- Auto recuperación

Es una característica para mantener el sistema en trabajo y en un estado operacional.

- Auto protección

Es la capacidad de protegerse de objetos maliciosos y el acceso no autorizado.

Otra necesidad para cumplir los objetivos de la aplicación es el uso de una herramienta que permita la administración de las localidades en la cartografía, un análisis del uso de las APIs de Google Maps hecho por Li, S [2] permiten una fácil adaptación de la cartografía digital con las bases de datos; además, se puede generar gráficos estadísticos usando las herramientas de Google Charts Tools en la que el usuario puede crear dibujos desde la fuente de datos y agregarlos a una página web, esta herramienta puede generar hasta 12 tipos diferentes de gráficos en los que se encuentran los siguientes tipos: pie, charts, scattercharts, gauge charts, geocharts, tables, treemaps, combo charts, linecharts, bar [3].

En cuanto al framework de desarrollo es el uso de Oracle Application Express la cual está basado en la integración con un navegador web. En Oracle Application Express OAE, Wojciechowski T [4], ha redactado un documento en el cual se genera un sistema transaccional y da algunas pautas hacer de las ventajas del framework como: un application builder, SQL Workshop, y administración, otra característica que brinda OAE es dar interfaces para ingreso de datos, como OAE está ligado directamente a Oracle Database facilita toda la administración de datos.

Saim Shin, D [5] brinda pautas para la detección de fonemas y comportamiento de las palabras en el lenguaje como:

1. Algoritmos de recuperación de palabras similares.

El algoritmo de la alineación de la secuencia se ha utilizado con el fin de medir la similitud de secuencias genéticas o para encontrar secuencias similares.

2. Clasificación de palabras similares

Se calculan su similitud mediante la comparación de la nueva palabra solicitada basa en sus pronunciaciones de fonemas

- a) Distancias unidimensionales: información de la distancia entre dos fonemas se expresan con los números enteros.
- b) Las distancias relativas: Las distancias de pronunciación entre fonemas se expresan relativamente.

La recuperación de esta información forma una luz para aclarar los objetivos y como solventarlos en el proceso de desarrollo del software:

2.5. Objetivo general:

Generar una aplicación web para ubicar geográficamente comunidades lingüísticas y analizar la dinámica de la lengua en cada comunidad mediante el uso de minería de texto.

2.6. Objetivos específicos:

1. Sistematizar información sobre cartografía digital, técnicas de minería de datos y análisis fonémico.
2. Especificar los requerimientos de la aplicación etnolingüística.
3. Diseñar y estandarizar la base de datos e interfaces adecuadas.
4. Desarrollar el sistema para la cartografía digital y análisis de lengua.
5. Aplicar pruebas de coherencia, integridad y presentación de los datos.

2.7. Metodología:

Scrum, una metodología ágil y flexible para gestionar el desarrollo de software.

2.8. Alcances y resultados esperados:

Al concluir con el proyecto se busca generar un software que pueda brindar soporte a personas dedicadas a la etnolingüística, el software mediante la minería de texto de los datos que se encuentren almacenados pueda distinguir los fonemas y el tipo de lenguaje usados en el apartado "historias de vida" que se encuentra en cada localidad.

Gráficos que presenten cuantitativamente los datos recabados por la minería de texto, mapas que muestren la ubicación de los grupos sociales y el uso de contenido multimedia facilitarían la comprensión de la información de cada grupo social a los usuarios que requieran información acerca de estos análisis.

2.9. Supuestos y riesgos:

Supuesto: La minería de texto es de alta complejidad.

Riesgo: La duración del proyecto se extiende.

Supuesto: El proyecto de desarrollo sea cancelado por parte de la Universidad de Azuay.

Riesgo: No tener un fundamento para el desarrollo del proyecto.

2.10. Presupuesto (Caso de materiales para obtener el producto final):

Materiales de oficina:

Computador:	\$ 700,00
Hojas de oficio:	\$ 20,00
Varios de oficina:	\$ 30,00
Total:	\$ 750,00

2.11. Financiamiento:

Propio.

2.12. Esquema tentativo:

1. Capítulo 1. Indagación exploratoria sobre aplicaciones web de geomática y minería de texto

- 1.1. Oracle Application Express
- 1.2. Oracle Database
- 1.3. Minería de texto en Oracle Database
- 1.4. Google Maps APIs
- 1.5. Integrar Google APIs a aplicaciones web

2. Capítulo 2. Especificación de los requerimientos de la aplicación (SRS)

- 2.1. Requisitos específicos
- 2.2. Diagrama de casos de uso

3. Capítulo 3. Diseño y estandarización de la base de datos e interfaces:

- 3.1. Creación y normalización la base de datos
- 3.2. Asignación de los almacenes de datos

4. Capítulo 4. Sistema para el análisis de lengua

- 4.1. Configuración de espacios de trabajo
- 4.2. Creación de mantenimientos
- 4.3. Creación de procesos almacenados
- 4.5. Creación de la visualización de la cartografía etnolingüística
- 4.6. Minería de texto

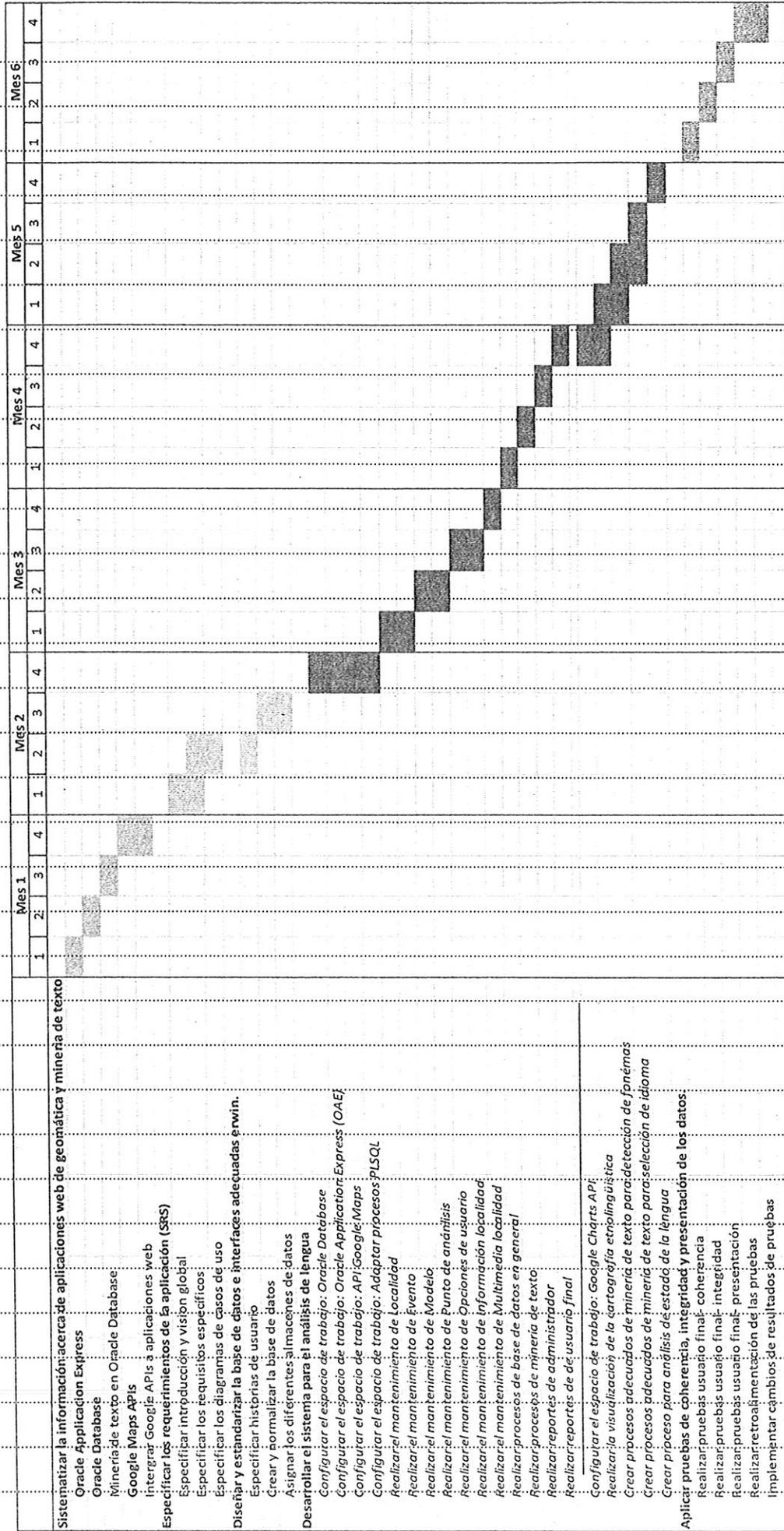
- 4.6.1. Realización la detección de fonemas
- 4.6.2. Realización de selección de idioma

5. Capítulo 5. Pruebas de coherencia, integridad y presentación de los datos:

- 5.1. Pruebas de usuario final
- 5.2. Retroalimentación y cambios de las pruebas

6. Capítulo 6. Conclusiones.

2.13. Cronograma:





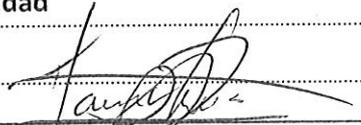
UNIVERSIDAD DEL
AZUAY

2.14. Referencias:

- [1] R. Pandita, R. P. Jetley, S. D. Sudarsan and L. Williams, "Discovering likely mappings between APIs using text mining," Source Code Analysis and Manipulation (SCAM), 2015 IEEE 15th International Working Conference on, Bremen, 2015, pp. 231-240.
- [2] S. Li, "A method for building thematic map of GIS based on Google Maps API," Geoinformatics, 2011 19th International Conference on, Shanghai, 2011, pp. 1-4.
- [3] Y. Zhu, "Introducing Google Chart Tools and Google Maps API in Data Visualization Courses," in IEEE Computer Graphics and Applications, vol. 32, no. 6, pp. 6-9, Nov.-Dec. 2012.
- [4] T. Wojciechowski, B. Sakowicz, D. Makowski and A. Napieralski, "Transaction system with reporting capability in a web-based data warehouse application developed in Oracle Application Express," CAD Systems in Microelectronics, 2009. CADSM 2009. 10th International Conference - The Experience of Designing and Application of, Lviv-Polyana, 2009, pp. 273-276.
- [5] S. Shin, D. Jang, J. S. Lee, D. H. Kim and S. Yoon, "Phoneme based realtime taboo words similarity browsing system of new words using multi-lingual taboo words databases in web environments." Information Networking (ICOIN), 2015 International Conference on, Cambodia, 2015, pp. 433-434.

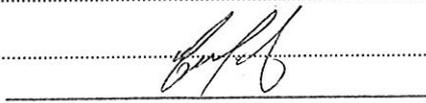
2.15. Anexos: No existen anexos en esta etapa.

2.16. Firma de responsabilidad



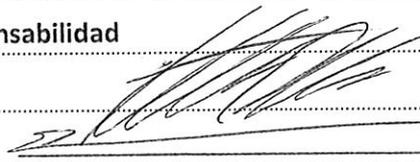
Irma Sigua Juan Fernando
Estudiante

2.17. Firma de responsabilidad



Astudillo Rodríguez, Catalina Mgst.
Director sugerido

2.18. Firma de responsabilidad



Orellana Cordero, Marcos Mgst.
Co-director sugerido

2.19. Fecha de entrega: Cuenca, mayo de 2016