



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

IDENTIFICACIÓN DE LAS TENDENCIAS DE RECLAMOS
PRESENTES EN RECLAMOS.CL Y QUE APUNTEN CONTRA
INSTITUCIONES DE EDUCACIÓN Y ORGANIZACIONES PÚBLICAS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL
INDUSTRIAL

DANIEL GUILLERMO BETH MADARIAGA

PROFESOR GUÍA:
JUAN D. VELÁSQUEZ SILVA

MIEMBROS DE LA COMISIÓN:
SEBASTIÁN A. RÍOS PÉREZ
ÁNGEL JIMÉNEZ MOLINA

SANTIAGO DE CHILE
JUNIO 2012

RESUMEN DE LA MEMORIA
PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL
POR: DANIEL GUILLERMO BETH MADARIAGA
FECHA: 19/06/2012
PROF. GUÍA: SR. JUAN D. VELÁSQUEZ SILVA

IDENTIFICACIÓN DE LAS TENDENCIAS DE RECLAMOS PRESENTES EN
RECLAMOS.CL Y QUE APUNTEN CONTRA INSTITUCIONES DE EDUCACIÓN Y
ORGANIZACIONES PÚBLICAS

En la siguiente memoria se busca corroborar, por medio de una experiencia práctica y aplicada, si a caso el uso de las técnicas de Web Opinion Mining (WOM) y de herramientas informáticas, permiten determinar las tendencias generales que pueden poseer un conjunto de opiniones presentes en la Web. Particularmente, los reclamos publicados en el sitio web Reclamos.cl, y que apuntan contra instituciones pertenecientes a las industrias nacionales de Educación y de Gobierno.

En ese sentido, los consumidores cada vez están utilizando más la Web para publicar en ella las apreciaciones positivas y negativas que poseen sobre lo que adquieren en el mercado, situación que hace de esta una mina de oro para diversas instituciones, especialmente para lo que es el identificar las fortalezas y las debilidades de los productos y los servicios que ofrecen, su imagen pública, entre varios otros aspectos.

Concretamente, el experimento se realiza a través de la confección y la ejecución de una aplicación informática que integra e implementa conceptos de WOM, tales como Knowledge Discovery from Data (KDD), a modo de marco metodológico para alcanzar el objetivo planteado, y Latent Dirichlet Allocation (LDA), para lo que es la detección de tópicos dentro de los contenidos de los reclamos abordados. También se hace uso de programación orientada a objetos, basada en el lenguaje Python, almacenamiento de datos en bases de datos relacionales, y se incorporan herramientas pre fabricadas con tal de simplificar la realización de ciertas tareas requeridas. La ejecución de la aplicación permitió descargar las páginas web en cuyo interior se encontraban los reclamos de interés para la realización experimento, detectando en ellas 6.460 de estos reclamos; los cuales estaban dirigidos hacia 245 instituciones, y cuya fecha de publicación fue entre el 13 de Julio de 2006 y el 5 de Diciembre de 2011.

Así también, la aplicación, mediante el uso de listas de palabras a descartar y de herramientas de lematización, procesó los contenidos de los reclamos, dejando en ellos sólo las versiones canónicas de las palabras que los constituían y que aportasen significado a estos.

Con ello, la aplicación llevó a cabo varios análisis LDA sobre estos contenidos, los que arbitrariamente se definieron para ser ejecutados por cada institución detectada, tanto sobre el conjunto total de sus reclamos, como en segmentos de estos agrupados por año de publicación, con tal de generar, por cada uno de estos análisis, resultados compuestos por 20 tópicos de 30 palabras cada uno. Con los resultados de los análisis LDA, y mediante una metodología de lectura e interpretación manual de las palabras que constituían cada uno de los conjuntos de tópicos obtenidos, se procedió a generar frases y oraciones que apuntasen a hilarlas, con tal de obtener una interpretación que reflejase la tendencia a la cual los reclamos, representados en estos resultados, apuntaban.

De esto se pudo concluir que es posible detectar las tendencias generales de los reclamos mediante el uso de las técnicas de WOM, pero con observaciones al respecto, pues al surgir la determinación de las tendencias desde un proceso de interpretación manual, se pueden generar subjetividades en torno al objeto al que apuntan dichas tendencias, ya sea por los intereses, las experiencias, entre otros, que posea la persona que realice el ejercicio de interpretación de los resultados.

A mi familia, a mi novia, a mis amigos y amigas, y a todas las personas que en todo este tiempo me privilegiaron con su compañía, cariño y apoyo. El fin llegó, pero es sólo el comienzo de una nueva aventura...

Agradecimientos

Han sido diez largos años, de muchas cosas, algunas increíbles y otras no tanto, pero al final, diez años de múltiples y diversas experiencias y aprendizajes.

Cómo olvidar todas esas reuniones, campañas, marchas y salidas con lo que fue la Juventud Socialista de Ingeniería y de la Universidad. Los trabajos en la FECH, en los plenos de federación, en la secretaría de comunicaciones y en el programa radial; las múltiples actividades realizadas con los centros de estudiantes, tanto de la facultad, como de otras; las capacitaciones y el manejo de las finanzas en los trabajos voluntarios de CAPSOCIAL, y en tantos otros proyectos, iniciativas e instancias en las que tuve la oportunidad y el privilegio de participar.

Pero eso no sería nada sin las personas que estuvieron presentes en cada uno de estos espacios. Fue gracias a ellas que todos estos momentos adquirieron significancia y sentido.

Quiero agradecer a mi familia, a su preocupación y paciencia por estar junto a mi todos estos años. Quizás no siempre fue de manera ideal, pero nunca retrocedieron y estuvieron conmigo hasta este momento final. También aprovecho de disculparme por no siempre estar abierto a sus ayudas y consejos, y por hacerlos esperar más de la cuenta para llegar a este hito que significa finalizar los estudios universitarios.

También quiero agradecer a todos aquellos maestros que pude conocer y que me brindaron de su sabiduría en este largo camino. Especialmente quiero agradecer a mis profesores Juan y Sebastián, quienes fueron capaces de entender mis inquietudes e intereses, y me otorgaron la oportunidad única de integrarme a su mundo intelectual. A uds. les debo el poder descubrir aspectos inéditos en torno a mi vocación, así como re encontrarme con el conocimiento que se puede acceder en las instancias universitarias.

A mi profesor Juan, le agradezco el haberme introducido e insertado al universo de la minería de datos aplicada a la Web, y el abrirme los ojos hacia esta área. Sin este empuje y motivación, el incluirme a trabajar con increíbles personas entendidas en esto, probablemente nunca me hubiese llegado a acercar a estas temáticas.

Mientras que a mi profesor Sebastián, el “Seba”, le agradezco el haber plantado la semilla inicial de lo que fue esta memoria. Sin esa idea y experiencia de haber hecho algo al respecto el verano del 2011, dudo que hubiese comprendido con mayor profundidad lo que pretendía realizar para este experimento.

A mis amigos y amigas, ya sean del colegio, la Universidad, o de cualquier otro espacio en donde nos pudimos haber conocido, les agradezco enormemente su compañía y apoyo en tantas cosas que vivimos juntos en esta etapa. Les agradezco sus risas y alegrías; sus palabras, conversaciones, atenciones y preocupaciones hacia mi persona. Realmente muchos de esos gestos fueron importantes y los guardo con cariño en los estantes de la vida y de sus experiencias.

Finalmente, quiero agradecer a la persona que marco un antes y un después en el transcurso de este proceso de aprendizaje y de desarrollo: a mi novia Francys. Sin ti, tus apoyos, paciencia, y por sobre todo, tu convicción de que era capaz de seguir hacia adelante, quizás no hubiese llegado a estar escribiendo estas líneas. Probablemente me habría retirado resignado, odiándome por no haberlo logrado y por no haber finalizado como correspondía esta etapa: cruzando por la puerta ancha por donde salen a la luz los “hijos del hambre y el frío; del rigor y el sufrimiento”, los ingenieros de la Universidad de Chile.

Índice general

1. Introducción	1
1.1. Hipótesis	3
1.2. Objetivos	3
1.2.1. Objetivo General	3
1.2.2. Objetivos Específicos	3
1.3. Metodología	4
1.4. Contribuciones	5
2. Marco Conceptual	7
2.1. La Web	7
2.1.1. Servidores Web	7
2.1.2. Sitios Web	8
2.1.3. Páginas Web	8
2.2. Bases de Datos Relacionales	13
2.2.1. Structured Query Language, SQL	13
2.3. Programación Orientada a Objetos, OOP	15
2.3.1. Python	16
2.4. Web Opinion Mining, WOM	17
2.4.1. Otros Campos Involucrados	17
2.4.2. Marco de Trabajo	18
2.4.3. Aplicaciones	18

2.4.4.	Knowledge Discovery from Data, KDD	18
2.4.5.	Problemas con el Idioma Español	20
2.4.6.	Latent Dirichlet Allocation, LDA	20
3.	Preparación del Experimento	22
3.1.	Caracterización del Entorno	22
3.1.1.	Estructura del Sitio	23
3.1.2.	Estructura y Contenido de las Páginas	25
3.1.3.	Rubros a Abordar	27
3.2.	Requerimientos Asociados	29
4.	Desarrollo de la Aplicación	30
4.1.	Módulos de Python a Utilizar	30
4.1.1.	Django	31
4.1.2.	Pattern	31
4.1.3.	Freeling	33
4.1.4.	Gensim	34
4.2.	Tablas de la Base de Datos	34
4.2.1.	Tablas de Fuentes	36
4.2.2.	Tablas de Reclamos.cl	37
4.2.3.	Tablas de Herramientas	41
4.2.4.	Tablas de Clasificaciones	42
4.3.	Módulos de la Aplicación	44
4.3.1.	Crawler y Spider	44
4.3.2.	Extractor, Almacenador y Tabulador	45
4.3.3.	Limpieza y Procesamiento de los Textos	46
4.3.4.	Análisis LDA	46
5.	Resultados y Análisis	48

5.1. Rubro de Educación	49
5.1.1. Las 10 Instituciones con más Reclamos y Recomendaciones	50
5.1.2. Departamento Universitario Obrero Campesino, DUOC	52
5.2. Rubro de Gobierno	67
5.2.1. Las 10 Instituciones con más Reclamos y Recomendaciones	68
5.2.2. Carabineros de Chile	70
6. Conclusiones	85
6.1. Objetivos Específicos	85
6.2. Objetivo General	86
6.3. Hipótesis	86
6.4. Comentarios y Apreciaciones	87
6.4.1. Resultados	87
6.4.2. Posibles Usos y Aplicaciones	87
Bibliografía	88
A. Consultas SQL de los Resultados	93
A.1. Total de Instituciones por Sub Rubro y Rubro	93
A.2. Total de Reclamos por Institución	93
A.3. Total de Recomendaciones por Institución	94
A.4. Total de Instituciones, Reclamos y Recomendaciones por Rubro y Sub Rubro	94
A.5. Instituciones que no Pertenecen a Ningún Sub Rubro	95
A.6. Detalles de los Reclamos y de las Recomendaciones, de las Instituciones por Rubro	95
A.7. Distribución de Reclamos, Recomendaciones, Respuestas y Comentarios por año y por Institución	97
A.8. Listado de Palabras de los Tópicos de LDA sobre Reclamos, por año y por Institución	100

Índice de cuadros

5.1. Cantidades totales de Instituciones, Reclamos y Recomendaciones existentes y realizados dentro de cada Sub rubro de Educación.	49
5.2. Las 10 Instituciones con mayor cantidad de Reclamos y Recomendaciones dentro del rubro de Educación.	50
5.3. Detalle de la cantidad total de Reclamos, Reclamos Respondidos, Respuestas y Comentarios realizados hacia las 10 Instituciones con más Reclamos y Recomendaciones del rubro de Educación.	51
5.4. Detalle de la cantidad total de Recomendaciones, Recomendaciones Respondidas, Respuestas y Comentarios realizados hacia las 10 Instituciones con más Reclamos y Recomendaciones del rubro de Educación.	51
5.5. Detalle cronológico de las cantidades totales de Reclamos, Reclamos Respondidos, Respuestas y Comentarios realizados hacia el Departamento Universitario Obrero Campesino DUOC.	52
5.6. Detalle cronológico de las cantidades totales de Recomendaciones, Recomendaciones Respondidas, Respuestas y Comentarios hechas hacia el Departamento Universitario Obrero Campesino DUOC.	52
5.7. Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2007.	54
5.8. Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2007.	55
5.9. Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2008.	56
5.10. Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2008.	57

5.11. Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2009.	58
5.12. Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2009.	59
5.13. Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2010.	60
5.14. Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2010.	61
5.15. Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2011.	62
5.16. Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2011.	63
5.17. Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC entre los años 2007 y 2011.	64
5.18. Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC entre los años 2007 y 2011.	65
5.19. Cantidades totales de Instituciones, Reclamos y Recomendaciones existentes y realizados dentro de cada Sub rubro de Gobierno.	67
5.20. Las 10 Instituciones con mayor cantidad de Reclamos y Recomendaciones dentro del rubro de Gobierno.	68
5.21. Detalle de la cantidad total de Reclamos, Reclamos Respondidos, Respuestas y Comentarios realizados hacia las 10 Instituciones con más Reclamos y Recomendaciones del rubro de Gobierno.	69
5.22. Detalle de la cantidad total de Recomendaciones, Recomendaciones Respondidas, Respuestas y Comentarios realizados hacia las 10 Instituciones con más Reclamos y Recomendaciones del rubro de Gobierno.	69
5.23. Detalle cronológico de las cantidades totales de Reclamos, Reclamos Respondidos, Respuestas y Comentarios realizados hacia Carabineros de Chile.	70
5.24. Detalle cronológico de las cantidades totales de Recomendaciones, Recomendaciones Respondidas, Respuestas y Comentarios hechas hacia Carabineros de Chile.	70

5.25. Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2007.	72
5.26. Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2007.	73
5.27. Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2008.	74
5.28. Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2008.	75
5.29. Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2009.	76
5.30. Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2009.	77
5.31. Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2010.	78
5.32. Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2010.	79
5.33. Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2011.	80
5.34. Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2011.	81
5.35. Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile entre los años 2007 y 2011.	82
5.36. Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile entre los años 2007 y 2011.	83

Índice de figuras

4.1. Modelo Entidad-Relación de todas las tablas que utilizará la aplicación para el almacenamiento y la consulta de los datos involucrados.	35
4.2. Modelo Entidad-Relación de las tablas de fuentes.	36
4.3. Modelo Entidad-Relación de las tablas de Reclamos.cl.	37
4.4. Modelo Entidad-Relación de las tablas de herramientas.	41
4.5. Modelo Entidad-Relación de las tablas de clasificaciones.	42

Capítulo 1

Introducción

Actualmente la World Wide Web, o simplemente la Web, se ha convertido en uno de los servicios de Internet más utilizados por las personas y las organizaciones alrededor del mundo, haciendo de esta un gran punto de encuentro, interacción, comunicación y por sobre todo, un gigantesco almacén de datos e informaciones de distintos tópicos e índoles [37, 50].

Esta realidad puede resultar enormemente atractiva para diversos actores, ya que en si, es una gran oportunidad para la obtención de información con variadas características y cualidades, la que podría ser útil y relevante para satisfacer las múltiples necesidades e intereses específicos que puedan poseer estos actores.

Pero esto a su vez presenta un problema y un desafío no menor: dada la gran cantidad de datos que actualmente se están almacenando en la Web, ¿Cómo es posible extraer la información relevante sin caer en un abuso de recursos?, sean estos de tipo temporal, humanos, financieros, entre otros.

Por otra parte, uno de los aspectos que es propio en el mundo de hoy, es que todos, sean quienes sean, son consumidores de productos y de servicios, y por ende, poseen alguna opinión o comentario respecto a lo que adquieren, ya sea a favor o en contra de este, lo que sumado al espacio de publicación que brinda la Web, ha hecho de esta una instancia en donde, estas opiniones y comentarios, pueden ser vertidas sin mayores complicaciones, censuras o tapujos.

Entonces, a partir de todo lo anterior: ¿Será posible extraer la información sobre las opiniones de productos y de servicios que se encuentran en la Web sin caer en un abuso de recursos?.

Ante esta pregunta, existen múltiples y variadas soluciones que se hacen cargo de responder y de solucionar el problema que subyace, siendo una de ellas: el uso de la minería de datos en sitios y páginas web, es decir, WOM¹.

Con WOM es posible extraer desde una gran cantidad de opiniones y de comentarios que se han publicado en la Web, la información que pueda resultar interesante para algún actor en particular, ya sea una persona, un grupo de ellas, o simplemente una organización. Todo esto sin caer en consumos excesivos de recursos e insumos, ya que WOM se apoya en el uso de máquinas y de sus capacidades de procesamiento para lograr esto.

¹Minería de Opiniones Web, o Web Opinion Mining en inglés.

Particularmente, el extraer este tipo información, puede ser relevante tanto para el que ofrece los productos y los servicios, como también para el que quiere adquirirlos, más aún si la información publicada es de carácter negativo, pues puede servir como una advertencia o un aviso respecto a las calidades, las experiencias de uso, entre otros, de aquellos productos o servicios involucrados o aludidos; cuestión que es fundamental al momento de tener que tomar una decisión al respecto, como por ejemplo, sacar de circulación un producto en particular, o simplemente decidir comprar o no un servicio.

En el caso de Chile, existen múltiples instancias en Internet y sitios web² que permiten a las personas criticar y quejarse contra productos y servicios que son o fueron adquiridos por ellos dentro del mercado nacional. Siendo uno de estos Reclamos.cl, que concentra una gran cantidad de quejas y de recomendaciones realizadas por consumidores finales respecto a los productos y a los servicios que han adquirido y que les han sido brindados por diversas instituciones de los distintos rubros y áreas de la industria nacional.

En este sitio, los consumidores finales, que en la práctica tienden a ser mayoritariamente personas naturales, pueden publicar libremente su queja o recomendación hacia cualquier institución que deseen, no necesariamente explicitando sus datos personales, como nombres y apellidos, lo cual les da un anonimato que quizás en otras instancias o medios no es igualmente posible.

Además, en el sitio, es posible navegar, buscar y filtrar por instituciones, rubro o sub rubro industrial, pero aún así no es posible obtener indicadores ni métricas estadísticas que realicen una síntesis de todos los reclamos y recomendaciones publicadas, salvo por la cantidad total de reclamos y recomendaciones asociadas a cada institución o rubro industrial, hecho que hace de la identificación de aspectos más transversales, tales como tendencias de los reclamos, correlaciones entre estos, sus frecuencias de publicación, por mencionar algunos ejemplos, algo difícil de obtener, por no decir casi imposible dada la actual propuesta de contenido existente en el sitio.

Frente a este desaprovechamiento de la información que se podría obtener de los reclamos y las recomendaciones presentes en Reclamos.cl, y que podría ser sumamente relevante para diversos actores interesados, tales como los mismos consumidores finales que visitan el sitio, el SERNAC³, o simplemente las instituciones que son aludidas en ellos, en la siguiente memoria se pretende identificar las tendencias generales de los reclamos publicados en el sitio, es decir, identificar de manera sintética el sentido semántico general al que apuntan los reclamos por cada institución dados los tópicos que los caracterizan, ya sea a nivel temporal global, como por año.

Concreta y específicamente se buscará, con las técnicas y las herramientas que WOM brinda para ello, así como también con las herramientas que aporta el desarrollo de aplicaciones informáticas, identificar las tendencias generales de los reclamos del sitio que apunten contra las instituciones pertenecientes a los rubros de educación y de servicios públicos.

Se abordarán los reclamos contra las instituciones de estos rubros, ya que también podría resultar interesante descubrir en los resultados de los análisis a realizar, si las manifestaciones sociales por una mejor educación ocurridas durante el año 2011 influyeron en las tendencias de los reclamos realizados hacia este conjunto de instituciones educacionales y públicas.

²Línea Directa del diario El Mercurio, Indignado.cl por mencionar algunos.

³Servicio Nacional del Consumidor.

1.1. Hipótesis

“El uso de las técnicas de WOM sobre los reclamos escritos en el sitio web Reclamos.cl permite identificar las tendencias generales de estos por cada institución, y en períodos de tiempo anual.”

1.2. Objetivos

Dada la hipótesis de investigación definida, es posible establecer tanto un objetivo general que apunte a permitir corroborarla, así como también objetivos específicos, que faciliten, paso a paso, alcanzar el objetivo general.

1.2.1. Objetivo General

“Identificar, mediante el uso de técnicas de WOM, las tendencias generales de los reclamos presentes en el sitio web Reclamos.cl y que apunten contra los servicios que brindan las instituciones de educación y del sector público, tanto a nivel global, como por año.”

1.2.2. Objetivos Específicos

Desglosando el objetivo general, es posible plantear los siguientes objetivos específicos:

1. Comprender los conceptos teóricos y el actual estado del arte de las técnicas y las herramientas que abarca WOM y que son necesarios para el experimento a realizar.
2. Establecer el conjunto de páginas web de Reclamos.cl que posean los reclamos y las recomendaciones hacia los productos y servicios que brindan las instituciones de educación y del sector público.
3. Extraer el contenido y sus meta datos anexos desde el conjunto de páginas web definido.
4. Definir el conjunto de palabras que no aportan valor al sentido de los contenidos extraídos, y suprimirlas de estos.
5. Transformar las palabras que aportan sentido a su forma canónica.
6. Procesar las palabras que aportan sentido y generar grupos de estas por cada institución abordada, tanto a nivel global, como por año.
7. Establecer frases o comentarios que sinteticen el sentido de los grupos y las palabras que los componen.
8. Confeccionar una aplicación informática que permita realizar todos los aspectos requeridos.

1.3. Metodología

Para alcanzar el objetivo general planteado, y corroborar el cumplimiento, o no, de la hipótesis de investigación, se procederá de la siguiente manera [47]:

1. **Identificación y Definición del Conjunto de Páginas Web:** Se identificará y establecerá el conjunto de páginas web del sitio Reclamos.cl que contengan los reclamos y las recomendaciones hacia los productos y servicios que brindan las instituciones de los rubros industriales de educación por un lado, y por otro, las del sector público.
2. **Construcción de la Aplicación Informática:** Una vez establecido el conjunto de páginas web correspondientes a estos rubros, se diseñará, confeccionará y construirá una aplicación informática que sea capaz de descargar y revisar cada una de estos documentos, de tal manera, que pueda extraer y procesar los contenidos que ahí se encuentren.

Dicha aplicación procederá a realizar los siguientes pasos:

- a) **Descarga de Páginas Web:** Lo primero que deberá realizar la aplicación será descargar el conjunto de páginas web a analizar y procesar. Para ello se le ingresará un listado que contenga las direcciones web de las páginas a abordar.
- b) **Extracción y Tabulación de los Datos:** Luego, una vez que se encuentren descargadas las páginas web requeridas, se le indicará a la aplicación los elementos de estas que contengan los datos relevantes para lograr los objetivos, con tal de que con ello la aplicación los identifique y los extraiga de cada una de las páginas web involucradas, almacenándolos de manera tabulada y ordenada al interior de tablas adecuadas para ello, las que se deberán encontrarse en una base de datos previamente construida.

Es importante señalar que los elementos a identificar dentro de cada página descargada, serán previamente establecidos en la aplicación mediante la revisión manual de algunos de estos documentos. En el caso de Reclamos.cl, el sitio posee la característica de que las páginas web que lo componen se encuentran basadas en plantillas cuyo HTML⁴ posee una estructura estándar y transversal. Esto quiere decir que las páginas que contienen cada uno de los reclamos se caracterizan por poseer una misma estructura y configuración de elementos en su HTML, diferenciándose entre ellas solamente por sus contenidos de texto individual (títulos, fechas, entre otros).

- c) **Borrado de Stopwords:** Habiendo terminado de extraer los datos relevantes, la aplicación deberá suprimir aquellas palabras que no aporten valor al sentido de los contenidos principales, es decir, deberá realizar un borrado de las stopwords presentes en los contenidos, tales como artículos, preposiciones, entre otros.

Para lograrlo, primero se confeccionará manualmente una lista de stopwords, la que deberá ser utilizada como referencia por la aplicación para ir borrándolas de los contenidos principales.

Dada la envergadura que podría tener todo el universo de contenidos recolectados, y las distintas formas en que se pueden presentar y escribir las palabras en el lenguaje escrito, incluyendo los errores ortográficos, quizás no sea posible suprimir todas las palabras que no aporten al contenido semántico, pudiendo quedar una que otra presente después de este proceso de limpieza.

⁴Lenguaje de Marcado de Hipertexto, o HyperText Markup Language en inglés.

- d) **Lematización de Palabras:** Cuando haya finalizado el borrado de las stopwords en los contenidos principales, la aplicación deberá lematizar las palabras, es decir, llevar las palabras restantes a sus formas canónicas. Vale mencionar que esta técnica no es completamente perfecta, pudiendo darse casos en que algunas de las palabras que no puedan ser llevadas a sus versiones canónicas, en donde se trabajará con la palabra original.
 - e) **Realización de Análisis LDA⁵:** Finalizada la etapa de lematización de las palabras que conforman los contenidos principales, la aplicación deberá reunir todos aquellos reclamos existentes por cada institución abordada y utilizar la herramienta estadística LDA, sobre el universo de palabras que estos posean, con tal de generar grupos, o más bien, en bolsas de palabras denominadas tópicos.
3. **Asignación de Etiquetas:** Con los tópicos elaborados, se procederá a realizar una lectura manual de estos, con tal asociarles alguna frase o comentario que refleje el sentido de las palabras que los conforman, y por ende, expresen su tendencia general.

En la lectura manual, se evitará considerar aquellas palabras que no aporten a los sentidos de los reclamos y que no fueron suprimidas cuando correspondía, así como también se tomarán como una sola palabra aquellas que posean la misma palabra canónica, pero que no fueron transformadas a ella en la etapa de lematización.

Luego, tomando las frases y comentarios elaborados, se debiese poder identificar la tendencia general de los reclamos que existen contra la institución y las demás instituciones abordadas, ya sea anualmente, como de manera global.

Habiendo realizado esto último, se habrá finalizado el proceso que permitiría identificar las tendencias generales de los reclamos contra las diversas instituciones consideradas, tanto a nivel general, como por año, logrando con ello alcanzar el objetivo general planteado, y corroborando el cumplimiento, o no, de la hipótesis de investigación planteada.

1.4. Contribuciones

Con la presente memoria se pretende aportar información sobre las técnicas, las herramientas y el actual estado del arte que existe en torno a WOM.

Asimismo, se aportará con un prototipo de una aplicación informática que integre diversos algoritmos cuyo trasfondo sea la implementación concreta de algunas de las técnicas de WOM sobre un conjunto de reclamos del sitio web Reclamos.cl.

En concordancia, también se contribuirá a entregar, al menos como una muestra, información sintética sobre las tendencias generales que presentan los reclamos publicados contra las instituciones abordadas, ya que como se ha mencionado, estos tipos de informaciones podrían resultar útiles y relevantes para aquellos actores que posean intereses sobre las prestaciones que otorgan estas organizaciones, ya sea a nivel de su calidad, el efecto negativo que pueden generar, entre varios otros más.

Pero fundamentalmente, con lo realizado y los resultados obtenidos, se apelará a dar muestra sobre los potenciales que podría tener una herramienta de este tipo para lo que es la generación

⁵Asignación Latente de Dirichlet, o Latent Dirichlet Allocation en inglés.

de informaciones de interés en diferentes ámbitos, como por ejemplo, la gestión de operaciones, la reducción de asimetrías de información, el marketing, el mejoramiento de las ventas y de los ingresos, la elaboración de políticas públicas y leyes que recojan de mejor manera la opinión de las personas, por sólo mencionar algunas [5].

Capítulo 2

Marco Conceptual

Para lograr los objetivos de manera adecuada y sistemática se hace necesario abordar previamente diferentes conceptos teóricos relacionados con la Web, los servidores web, los sitios y las páginas web, así como también, la minería de datos y WOM, el almacenamiento de datos en bases de datos relacionales, y el desarrollo de aplicaciones informáticas, todo ello con el propósito de interiorizarse en ellos, y así facilitar la comprensión de lo que se abordará y ejecutará en el experimento a realizar, en los resultados que se vayan a obtener de este, y en las conclusiones que se puedan generar al respecto.

2.1. La Web

La Web [4] se compone de un gran número de documentos electrónicos interconectados entre si denominados páginas web, las cuales se agrupan formando sitios web.

Estas páginas web son almacenadas en computadores que se encuentran físicamente distribuidos por todo el orbe, pero que están comunicados y conectados entre si por conexiones de distinta naturaleza formando lo que se denomina redes de conexión.

Internet es la mayor red de computadores del mundo, en donde la Web es uno de los servicios que funcionan dentro de ella.

2.1.1. Servidores Web

En los computadores que almacenan páginas web generalmente se ejecutan aplicaciones informáticas denominadas servidores web, las que pueden ser ejecutadas en cualquier tipo de computador, desde computadores personales, hasta computadores específicamente diseñados para ello.

Funcionan bajo el paradigma del Cliente/Servidor [27], en donde brindan las páginas web que le solicitan tanto personas, como otras máquinas o computadores, denominados clientes.

Los servidores web pueden funcionar de manera aislada, sin estar comunicados con otros

computadores, como también de manera conectada, es decir, comunicándose e intercambiando informaciones y datos con otras máquinas mediante una conexión física, las que de manera grupal constituyen redes de conexión.

Un servidor funcionando de manera aislada sólo es útil para asuntos de prueba, ya que por propósito fundamental, un servidor web apunta a entregar páginas web a personas y máquinas que accedan a él mediante las redes de conexión existentes entre ellos.

Básicamente, cuando una persona o máquina abre una página web, lo que realiza, es solicitarla, a través de las redes de conexión, a un servidor web que se encuentra ejecutándose en un computador conectado a la Web. Dicho servidor, cuando recibe la solicitud por una página en particular la busca dentro del o los computadores¹ en que podría encontrarse. Si la encuentra, se la envía al cliente por las redes de conexión para que este la abra. En cambio, si no la encuentra, le envía un error, indicándole al cliente que la página solicitada no pudo ser encontrada, y por lo tanto no existe.

Las solicitudes de páginas web que puedan realizarle un cliente a un servidor web se hace mediante la dirección web única y particular que posee cada página, denominada URL².

Una URL se compone de tres elementos básicos [36]:

`protocolo://maquina/objeto`

El protocolo utilizado en el ambiente web es el HTTP³. La máquina corresponde al nombre (dominio de Internet) o al número IP⁴ que posee el servidor web, mientras que el objeto es la ruta en la que se ubica la página web en el servidor.

2.1.2. Sitios Web

Un sitio web [19] consiste en un conjunto de páginas web interconectadas entre sí mediante enlaces [14], generalmente localizadas dentro un mismo computador o servidor web, y que comparten una serie de elementos en común, como por ejemplo, temáticas, estructuras, funcionalidades, objetivos, entre otros.

Los sitios web poseen una dirección web o URL única.

2.1.3. Páginas Web

Las páginas web [9] son documentos digitales que se se manifiestan en archivos, y al igual que los sitios, cada una posee una URL única, con la que es posible acceder a ella..

¹Un servidor web puede funcionar y acceder a uno o a varios computadores.

²Localizador de Recursos Uniforme, o Uniform Resource Locator en inglés.

³Protocolo de Transferencia de Hipertexto, o Hypertext Transfer Protocol en inglés.

⁴Protocolo de Internet, o Internet Protocol en inglés.

Se componen de diversos elementos que poseen diferentes naturalezas, pudiendo ser algunos de estos: elementos de texto, código de ejecución (o scripts), imágenes, sonidos y vídeos, entre otros, los que a su vez, pueden ser clasificados en distintas categorías según el tipo de información que posean [49]; siendo las más relevantes para esta memoria:

- **Contenido:** Corresponde a los elementos, tales como textos, imágenes, sonidos, vídeos que se encuentran en una página web, y que las personas al visitarlas pueden observarlos u oírlos.
- **Estructura:** Son aquellos elementos que indican la estructura y la organización que poseerá el contenido presente en una página web. En general, tiende a ser el HTML que constituye a las páginas web.

Es importante mencionar que la visualización de las páginas, y la de sus elementos, por parte de las personas y de las máquinas no es la misma. Mientras que las personas generalmente visualizan las páginas a través de un navegador o browser que despliega gráficamente los colores, las imágenes, los formatos visuales y las tipografías que puedan tener asociadas; las máquinas o computadores ven directamente el HTML que las define y constituye. Por lo que la realización de operaciones sobre estos documentos a través de computadores o máquinas requiere necesariamente conocer y entender cómo estos “visualizan” las páginas web; más todavía, si se pretende realizar análisis de los contenidos que puedan poseer, como es en el caso de WOM.

HyperText Markup Language, HTML

HTML [51] es el lenguaje de publicación en la Web. Constituye la esencia de la estructura presente en las páginas web, así como también define las características y las propiedades de los diferentes elementos que las componen, y como su nombre lo indica, se conforma a partir de una serie de etiquetas o tags.

Por otra parte, el W3C⁵, es el organismo internacional responsable de definir el HTML y sus tags. En base a esto, y a las nuevas necesidades que se han ido presentando en la Web, cada cierto tiempo, el W3C publica cambios sobre el HTML, sintetizados en distintas versiones de este lenguaje, siendo la versión 4.01 [39] (publicada en Diciembre de 1999), la oficial hasta el momento, aunque el W3C se encuentra definiendo la que vendría a ser su siguiente versión, la número 5 [32], que incluye una serie de cambios, entre ellos, el apuntar hacia páginas web más semánticas, es decir, páginas que puedan ser mejor comprendidas por las máquinas en lo que se refiere a su contenido, todo ello mediante la incorporación de tecnologías como RDF⁶ y OWL⁷ [48, 2].

Este hecho es sumamente importante desde el punto de vista de WOM, ya que, como se ha mencionado, una de las actividades que se realizan en esta área, es el poder lograr que las máquinas sean capaces de comprender los textos que abordan y procesan, y una iniciativa de este tipo, permite claramente mejorar aquello.

En relación a los tags que conforman al HTML, es posible definir mediante ellos, si un texto incluido en la página corresponderá a un título, un párrafo, un ítem de una lista, o un dato dentro de una celda perteneciente a una tabla, entre varios otros más.

⁵Consortio de la World Wide Web, o World Wide Web Consortium en inglés.

⁶Marco de Descripción de Recursos, o Resource Description Framework en inglés.

⁷Lenguaje Ontológico Web, o Web Ontology Language en inglés.

También a través de estos tags, es posible definir las imágenes, los vídeos, los sonidos, o los códigos de ejecución (scripts) que se adjuntarán a una página web en particular, así como sus respectivas características y propiedades, a través de explicitar ciertos atributos dentro de cada tag, como por ejemplo el ancho (width) y el alto (height) en píxeles que posee el elemento en cuestión, la clase (class) a la que pertenece, o el identificador único (id) que lo singulariza dentro de la página, entre otros.

A modo de ejemplo, a continuación el HTML, versión 4.01, de una página básica, explicitando los elementos que pueden resultar más relevantes a comprender [52]:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/strict.dtd">
<html>
  <head>
    <title>Página Web Básica de Ejemplo</title>
    <meta name="description" content="Esta es la descripción de una
    página web básica para ejemplificar el uso de HTML" />
    <meta name="keywords" content="Web, Página, HTML, Ejemplo" />
  </head>
  <body>
    <!-- Esto es un comentario en el HTML -->

    <h1>Título Principal del Contenido</h1>

    <p>Esto es un párrafo de ejemplo.</p>

    <h2>Título Secundario del Contenido</h2>

    <p>Esto es otro párrafo de ejemplo, junto con un
    <a href="http://www.uchile.cl" title="Este es un texto que aporta
    más información sobre el enlace">enlace</a> de ejemplo, y
    acompañado de una imagen.</p>

    <ol>
      <li>Ítem 1 lista ordenada</li>
      <li>Ítem 2 lista ordenada</li>
    </ol>

    <ul>
      <li>Ítem 1 lista no ordenada</li>
      <li>Ítem 2 lista no ordenada</li>
    </ul>

    <p>Lo siguiente es una tabla de 2 filas por 2 columnas, además de
    una fila que hace de encabezado de la tabla.</p>

    <table>
      <thead>
```

```

    <tr>
      <th>Encabezado Columna 1</th>
      <th>Encabezado Columna 2</th>
    </tr>
  </thead>
  <tbody>
    <tr>
      <td>Celda Fila 1, Columna 1</td>
      <td>Celda Fila 1, Columna 2</td>
    </tr>
    <tr>
      <td>Celda Fila 2, Columna 1</td>
      <td>Celda Fila 2, Columna 2</td>
    </tr>
  </tbody>
</table>

<div id="capa-ejemplo" class="ejemplos">
  <p>Esto es un párrafo dentro de una capa cuyo identificador es
  "capa-ejemplo" y que pertenece a la clase "ejemplos".</p>
</div>

<script language="javascript">
  alert("Esto es un sript");
</script>
</body>
</html>

```

Document Object Model, DOM

El DOM⁸ [33] es una convención definida por el W3C que permite representar e interactuar con los elementos que constituyen el HTML de una página web.

Como su nombre deja entrever, DOM establece que cada elemento del HTML, es decir, sus tags, pueden ser modelados como objetos, cuyas propiedades específicas y particulares correspondientes a los atributos, textos u otros elementos que posean o encierren.

Por ejemplo, en el caso de un enlace, el tag “a” correspondería al objeto en el DOM, mientras que sus atributos, tales como la URL a la que apunta o el texto al que encierra, vendrían a ser las propiedades del objeto.

Además, el DOM de una página, es decir, el conjunto de objetos que representan los tags que constituyen una página web, pueden ser estructurados como un árbol (árbol DOM) integrado por nodos y sub nodos. En donde, un nodo es aquel tag del HTML que contiene a otros tags, y un sub nodo es aquel tag contenido dentro de otro tag.

Basándose en el HTML de la página web de ejemplo descrito anteriormente, su árbol DOM

⁸Modelo de Objetos del Documento, o Document Object Model en inglés.

correspondiente sería:

```
+ html
+ head
  - title
  - meta
  - meta
+ body
  - h1
  - p
  - h2
  + p
    - a
  - img
  + ol
    - li
    - li
  + ul
    - li
    - li
  - p
  + table
    + thead
      + tr
        - th
        - th
    + tbody
      + tr
        - td
        - td
      + tr
        - td
        - td
  + div
    - p
  - script
```

Esta estructuración que posibilita DOM resulta sumamente poderosa para lo que es WOM y la extracción de datos desde páginas web, pues permite que el identificar cualquier objeto, y su tag correspondiente dentro de una página, sea algo simple y fácil de realizar.

Asimismo, DOM también facilita y simplifica el poder modificar el HTML que constituye una página web, pudiendo agregar, quitar, o sólo cambiar en parte, cualquier tag, o un conjunto de estos, que se encuentre dentro de las páginas, con lo cual, se hace posible por ejemplo, todo aquello referente a cargar elementos asincrónicamente en una página mediante el uso de Javascript y tecnologías como AJAX⁹ [53].

⁹Javascript y XML Asincrónico, o Asynchronous JavaScript and XML en inglés.

2.2. Bases de Datos Relacionales

Las bases de datos relaciones [21] son un paradigma del almacenamiento de datos representado por un conjunto de tablas que poseen una o varias relaciones entre si, y que en esencia cumplen con el modelo relacional planteado por Edgar Frank Codd [15], cuyas bases nacen de la teoría de conjuntos.

En el modelo relacional, como base se plantean lo que son las relaciones, que vendrían a ser abstracciones de entidades existentes en el mundo real.

Dichas relaciones se constituyen a partir de campos o atributos de la entidad que representan, y de registros o tuplas, que vendrían a ser instancias de la entidad. En donde, además, los vínculos que puedan existir entre una relación y otra se manifiestan por atributos comunes que compartan entre ellas.

Desde la perspectiva de una base de datos relacional, las relaciones del modelo relacional se manifiestan como las tablas que conforman la base de datos, cuyas filas corresponden a los registros, y sus columnas, a los campos de las relaciones que representan, siendo las columnas que posean en común estas tablas, la representación de los vínculos existentes entre una relación y otra.

Una tabla de una base de datos relacional puede relacionarse con una o varias otras tablas de distintas maneras, siendo estas:

- **Uno a Uno:** Un registro de una tabla se vincula a sólo un registro de otra tabla.
- **Uno a Varios:** Un registro de una tabla se vincula varios registros de otra tabla.
- **Varios a Muchos:** Varios registros de una tabla se vinculan con muchos registros de otra tabla, y a su vez, varios registros de la otra tabla, se vinculan con muchos registros de la tabla inicial.

Con respecto a la manipulación de los datos que se encuentren dentro de una relación en el modelo relacional, se utilizan las herramientas matemáticas que brindan el Álgebra y el Cálculo Relacional, que se encuentra representado en las bases de datos relaciones por el SQL¹⁰, el cual permite realizar una serie de operaciones sobre los datos existentes en las tablas de la base de datos, así como también en los datos que se desean almacenar en estas.

A nivel de aplicaciones informáticas, una base de datos relacional es implementada y gestionada por un BDMS¹¹, siendo algunos de estos: PostgreSQL, MySQL, SQLite, entre otros.

2.2.1. Structured Query Language, SQL

SQL es un lenguaje, cuya base lógica surge del Álgebra y del Cálculo Relacional utilizado en el modelo relacional, y permite, mediante una serie de instrucciones y comandos, gestionar datos dentro de un ambiente de bases de datos relacionales propiciado por un DBMS.

¹⁰Lenguaje de Consulta Estructurado o Structured Query Language en inglés.

¹¹Sistema de Gestión de Bases de Datos, o Database Management System en inglés.

Dentro de las múltiples instrucciones que aporta y soporta SQL [60], se encuentran:

- Inserción de nuevos datos en las tablas. Cuya instrucción en SQL sería [55]:

```
INSERT INTO tabla
VALUES (valor1, valor2, valor3, ..., valorN)
```

Que permite insertar los valores indicados en los campos correlativamente correspondientes, dentro de una nueva fila de la tabla “tabla” de la base de datos.

O también:

```
INSERT INTO tabla (campo1, campo2, campo3, ..., campoN)
VALUES (valor1, valor2, valor3, ..., valorN)
```

Que permite insertar los valores indicados en los campos nombrados, dentro de una nueva fila de la tabla “tabla” de la base de datos.

- Actualización del contenido de los datos ya existentes en las tablas, bajo alguna condición o criterio que cumplan los datos a actualizar. Cuya instrucción en SQL sería [55]:

```
UPDATE tabla
SET campo1=valor1, campo2=valor2, ...
WHERE campo_i=valor
```

Que actualiza los campos nombrados con los valores indicados en la fila de la tabla “tabla” de la base de datos, en donde su campo “campo_i” posea el valor “valor”.

- Borrado de datos presenten las tablas, también bajo alguna condición o criterio que cumplan los datos a borrar. Cuya instrucción en SQL sería [54]:

```
DELETE FROM tabla
WHERE campo_i=valor
```

Que borra el o los registros de la tabla “tabla” de la base de datos, cuyo campo “campo_i” posea el valor “valor”.

- Obtención o “consulta” de datos que se encuentren en las tablas, los que se pueden acotar o filtrar en base a algún criterio o condición que cumplan, y cuyas instrucciones en SQL serían [59]:

```
SELECT campo1, campo2, ...
FROM tabla
WHERE campo_i=valor
```

Que obtiene los valores de los campos nombrados pertenecientes a la, o las, filas de la tabla “tabla” de la base de datos, que cumplan la condición de que su campo “campo_i” posea el valor “valor”.

O también:

```
SELECT %  
FROM tabla  
WHERE campo_i=valor
```

Que obtiene los valores de todos los campos pertenecientes a la, o las, filas de la tabla “tabla” de la base de datos, que cumplan la condición de que su campo “campo_i” posea el valor “valor”.

Asimismo, esto puede ser realizado sobre los datos pertenecientes no sólo a una tabla en particular, sino sobre datos que pertenezcan a varias tablas de la base de datos y que se encuentren relacionadas entre si, pudiendo con ello lograr un “cruce” de estos. Para lo cual se utiliza la instrucción de SQL “JOIN” [56], siendo las más relevantes para esta memoria, las sub instrucciones “LEFT JOIN” y “RIGHT JOIN”.

La instrucción “LEFT JOIN” [57] sería algo del estilo:

```
SELECT tabla1.%, tabla2.%  
FROM tabla1  
LEFT JOIN tabla2 ON tabla1.campo_i = tabla2.campo_i  
WHERE tabla1.campo_j=valor
```

Que obtiene todos los campos y la, o las, filas de la tabla “tabla1” en cuyo campo “campo_j” posean el valor “valor”, junto con todos los campos de la tabla “tabla2”, pero sólo aquellas filas pertenecientes a esta última tabla que posean en su campo “campo_i” el mismo valor que poseen las filas de la tabla “tabla1” en su campo “campo_i”.

Análogamente, la instrucción “RIGHT JOIN” [58] de SQL sería:

```
SELECT tabla1.%, tabla2.%  
FROM tabla1  
RIGHT JOIN tabla2 ON tabla1.campo_i = tabla2.campo_i  
WHERE tabla1.campo_j=valor
```

Que, en primer lugar, obtiene todos los campos y las filas de la tabla “tabla2”, y luego, todos los campos de aquellas filas de la tabla “tabla1” que poseen el mismo valor en su campo “campo_i” que los valores que poseen las filas de la tabla “tabla2” en sus respectivos campos “campo_i”; para luego, finalmente, filtrar sobre todas las filas obtenidas en base a aquellas cuyos campo “campo_j” (proveniente de la tabla “tabla1”) posean el valor “valor”.

Por otra parte, muchas veces los DBMS incluyen, además de las instrucciones básicas de SQL, instrucciones de SQL propias de ellos, que varían entre los distintos DBMS existentes.

2.3. Programación Orientada a Objetos, OOP

La Programación Orientada a Objetos [16], o Object-Oriented Programming, OOP, en inglés, es un paradigma de la programación y del desarrollo de aplicaciones informáticas.

En OOP se trabaja con abstracciones de entidades concretas que se basan en estructuras de datos, y que se denominan objetos.

Estos objetos se componen, por un lado, de campos o variables en donde se almacenan datos particulares de estos, mientras que por otro, también se componen por métodos o funciones que permiten realizar operaciones sobre ellos mismos, otros objetos, o sobre sus propios datos constituyentes.

Los datos asociados a cada objeto le atribuyen propiedades y características particulares a estos, siendo denominada como una instancia de un objeto, a un objeto cuyos datos específicos han sido declarados.

Los métodos o funciones que se encuentran asociados a un objeto, son un conjunto de instrucciones ordenadas secuencialmente, es decir, son algoritmos que al ejecutarse realizan operaciones basadas en las instrucciones que los constituyen.

Actualmente existen múltiples lenguajes de programación, que poseen sus propias sintaxis e instrucciones, y que se enmarcan dentro de este paradigma de programación, siendo algunos de estos: C, Java, Python, entre otros.

2.3.1. Python

Python es un lenguaje de programación creado por Guido van Rossum a principios de los años 90 cuyo nombre está inspirado en el grupo de cómicos ingleses “Monty Python”. Es un lenguaje similar a Perl, pero con una sintaxis muy limpia y que favorece un código legible.

Se trata de un lenguaje interpretado o de script, con tipado dinámico, fuertemente tipado, multiplataforma y orientado a objetos. [20]

Al ser un lenguaje interpretado, la ejecución de una aplicación desarrollada en Python pasa por una aplicación intermedia, llamada intérprete, que se encarga de traducir las instrucciones presentes en el código fuente a un lenguaje intermedio denominado “bytecode” que las máquinas pueden entender y ejecutar.

El hecho de que Python sea un lenguaje de tipado dinámico permite no tener que declarar el tipo de datos que contendrá una variable, es decir, por ejemplo, no hay que declarar explícitamente que una variable será de tipo “booleana”, sino que este se detectará automáticamente a partir del dato que se le asigne a la variable, y cambiará de tipo, si se cambia el tipo del dato que contiene la variable.

En concreto, Python acepta varios tipos distintos de datos, que para efectos de esta memoria, los más relevantes son los datos de texto (string), booleanos (true o false) y numéricos, ya sean enteros (integer) o no (float) [46].

Al ser fuertemente tipado, Python impide por ejemplo, trabajar con variables ya declaradas y que sean de tipos de datos distintos, a no ser, que las variables se transformen a un mismo tipo de datos.

Por otra parte, la propiedad de que Python sea multiplataforma [43], permite que las aplicaciones desarrolladas en este lenguaje puedan ser portables, es decir, puedan ser utilizadas y ejecutadas, sin mayores complicaciones, en distintos ambientes de sistemas operativos, ya sean estos Windows, Solaris, Unix, Linux, Mac OS, o cualquier otro.

Por tratarse de un lenguaje orientado a objetos, Python posee intrínsecamente una serie de elementos y de estructuras que son usuales en este tipo de lenguajes [45], siendo las más relevantes para esta memoria:

- Colecciones, ya sean Listas, Tuplas o Diccionarios.
- Controles de Flujo, ya sean Sentencias Condicionales (“if”, “else” y “elif”) o Bucles (“while”, “for” e “in”).
- Métodos y Funciones, tanto las de creación propia, como las que vienen por defecto en el lenguaje.
- Importación de Módulos y Clases.
- Control y Gestión de Errores y Excepciones.

2.4. Web Opinion Mining, WOM

WOM es una aplicación de la minería de datos [30], particularmente, de la minería de datos realizada sobre contenidos de origen web [36], y aborda específicamente, opiniones y comentarios que han sido escritos y publicados por personas al interior de páginas web.

Detallando, la minería de datos es:

The analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner. [31]

En base a aquello, WOM específicamente utiliza las metodologías, las técnicas y las herramientas de la minería de datos, para procesar y analizar un gran conjunto de datos basados en textos, constituidos a partir del contenido presente dentro de páginas web [37], con el propósito de descubrir patrones o informaciones que puedan resultar relevantes para los intereses de algún actor en particular.

2.4.1. Otros Campos Involucrados

Además de enmarcarse en lo que vendría a ser la minería de datos aplicada a la Web, las técnicas, las herramientas y los recursos que se utilizan en WOM también provienen de otros campos o áreas del conocimiento [5], tales como:

- Extracción de Información (EI), o Information Extraction (IE) en inglés.
- Recuperación de Información (RI), o Information Retrieval (IR) en inglés.
- Procesamiento de Lenguaje Natural (PLN), o Natural Language Processing (NLP) en inglés.
- Máquinas Aprendiendo Idiomas, o Machine Language Learning (ML) en inglés.

2.4.2. Marco de Trabajo

A partir de complementar los conocimientos que aportan los distintos campos que se encuentran más allá de la minería de datos, y en los que también se apoya WOM, se hace posible construir un marco de trabajo [5] que se compone de los siguientes elementos:

- Extracción de Elementos.
- Extracción de Características.
- Sentimientos en las Características.
- Sentimientos en los Elementos.
- Comparación de los Elementos.
- Comparación de las Características.

2.4.3. Aplicaciones

A partir de todo lo anterior, algunas áreas o rubros industriales en las que se ha aplicado y utilizado WOM han sido las siguientes [5, 38]:

- Compras y Ventas
- Entretenimiento
- Gobierno
- Investigación y Desarrollo
- Marketing
- Educación

2.4.4. Knowledge Discovery from Data, KDD

Quizás uno de los grandes desafíos en WOM es poder generar conocimientos a partir de datos, ante lo cual se presenta como solución KDD [22].

KDD es una metodología utilizada en minería de datos, que plantea un procedimiento para lograr descubrir conocimientos, a partir de datos que se encuentran desordenados y que aparentemente no poseen informaciones relevantes. Todo ello mediante el uso de métodos y técnicas de minería de datos que detecten patrones y faciliten el realizar los descubrimientos.

En base a lo anterior, y para poder comprender adecuadamente las etapas propiamente tales de KDD, se hace necesario definir previamente lo que son los datos, la información y el conocimiento [18]:

- **Datos:** Corresponden a los elementos básicos de información, los que de manera aislada no presentan relevancia para lo que son las decisiones a tomar. En términos algo más técnicos, se podría decir que representan la unidad básica de información y la agrupación de estos genera un conjunto discreto de valores, como por ejemplo un listado de productos varios.
- **Información:** Es la manifestación del procesamiento de un conjunto de datos, en el que se les ha otorgado a estos significado, relevancia, categorización y contexto, así como también, se les ha corregido y condensado para que adquieran propósito y sentido.

Con la información ya es posible que un actor pueda tomar decisiones al respecto.

- **Conocimiento:** Es la consecuencia de la integración de la información junto con otros elementos, tales como experiencias, conocimientos y prácticas previas, entre otros. Para generar conocimiento se hace necesario la intervención e interpretación humana, idealmente de algún experto en la materia, ya que así podrá realizar aún mayores integraciones de otros elementos. Con el conocimiento es posible ejecutar y llevar a cabo acciones en el tiempo.

Como se ha mencionado, KDD se compone de diferentes etapas, siendo estas:

1. Definir el objetivo principal de KDD, considerando el campo o el contexto en donde este se desenvolverá y ejecutará, así como también el conocimiento que se desea identificar.
2. Seleccionar una fuente que aporte un conjunto de datos en donde se centrará y ejecutará el proceso de descubrimiento del conocimiento.
3. Limpiar y pre procesar los datos involucrados, manteniendo cuidado con cuestiones como la pérdida de datos y la reducción de datos irrelevantes (reducción de ruido).
4. Reducir y proyectar los datos, es decir, buscar aspectos que permitan representar a los datos de una manera más sintética por medio de aplicarles alguna especie de reducción dimensional o un método de transformación que no se escape del contexto definido por los objetivos.
5. Seleccionar alguna técnica de minería de datos adecuada para la consecución del objetivo planteado.
6. Seleccionar un análisis exploratorio, modelo e hipótesis que implique escoger los algoritmos adecuados para la búsqueda de patrones, modelos y parámetros en los datos.
7. Aplicar la técnica de minería de datos (previamente seleccionada) sobre los datos, con tal de buscar patrones en estos y generar información al respecto.
8. Interpretar los patrones encontrados y la información previamente generada, para así poder generar conocimientos.
9. Utilizar inmediatamente o almacenar para utilizar más tarde, el conocimiento generado.

Es importante señalar, que KDD es un proceso interactivo e iterativo en cada etapa, es decir, en cada momento se debe interactuar con este, revisando y evaluando lo logrado, tal que si se hace necesario realizar correcciones o mejoras, se vuelva a repetir la etapa, calibrando los parámetros particulares que la definen, en pos de ir ajustando todo el proceso en lo que más se pueda al ideal del objetivo planteado.

2.4.5. Problemas con el Idioma Español

Como se ha visto, WOM hace uso de técnicas, herramientas y recursos que provienen de distintos campos del conocimiento, siendo muchos de ellos, de los campos de la lingüística y de las tecnologías de información, como por ejemplo NLP y ML.

Asimismo, también se ha visto que gran parte de los procesamientos y análisis que se realizan en WOM guardan relación con la identificación de sentimientos en los textos y en los datos sobre los cuales se trabaja.

En base a estos antecedentes, y poniendo énfasis en la disponibilidad de herramientas y de recursos robustos para realizar este tipo de análisis directamente y de manera nativa sobre opiniones escritas en español, se presenta la situación de que la mayoría de los recursos disponibles se ha enfocado en el idioma inglés [17, 38], por lo que hace del análisis de sentimientos algo complejo de realizar nativamente en español, a no ser que se confeccionen soluciones por cuenta propia o que utilicen traductores al inglés [3].

2.4.6. Latent Dirichlet Allocation, LDA

Frente a las dificultades que presenta el poder realizar un análisis de sentimientos directa, fácilmente y de manera nativa sobre textos escritos en español, ya que no hay suficientes recursos, ni herramientas informáticas disponibles para ello; una alternativa que permitiría identificar al menos las tendencias de estos sería LDA [7, 34].

Asignación Latente de Dirichlet, LDA por sus siglas en inglés, es un modelo Bayesiano del tipo de modelamiento de tópicos [6], que genera de manera probabilística, colecciones de datos o bolsas de palabras denominadas tópicos, cuyas constituciones internas están dadas por palabras originadas del o los contenidos de texto tomados como base.

La idea fundamental que se encuentra detrás de LDA reside en que cada documento de texto, o conjunto de estos, y que haya sido considerado para realizarle este tipo de análisis, posee una distribución sobre un conjunto de tópicos (\mathcal{T}), donde, a su vez, cada uno de estos se encuentra modelado como una probabilidad distribuida sobre un sub conjunto de palabras ($v_i \in \mathcal{V}$), siendo dichas distribuciones, muestras de las distribuciones multinomiales de Dirichlet.

El proceso que se encuentra detrás de este análisis puede ser descrito de la siguiente manera [29]:

Sea un documento o un contenido de texto \mathbf{p} :

1. Las palabras que posea serán consideradas como eventos independientes entre sí. Para representar esto, sea $S \sim Poisson(\xi)$ el número de palabras que existen en un documento dado. La s -ésima palabra del documento es representada por w^s .
2. Sea β la distribución sobre las palabras de cada tópico, y sea θ una distribución multinomial sobre cada tópico por cada documento, donde $\theta \sim Dir(\alpha)$. La distribución Dirichlet es usada en estadísticas Bayesianas para estimar parámetros escondidos de distribuciones de categorías, lo que en este caso corresponderían a los tópicos.

3. Luego, para cada palabra $w^s \in \mathbf{p}$, sea $z_s \sim Multinomial(\theta)$ un vector donde z_s^p es la probabilidad de que el t3pico s se encuentre en el documento p .
4. Finalmente, la palabra w^s de $p(w^s|z_s, \beta)$, que es una probabilidad multinomial condicionada por el t3pico z_s , es escogida.

Donde el conjunto final de t3picos \mathcal{T} es construido por el mayor k n3mero de t3picos z_s de n palabras, siendo k y n definidos previamente en la configuraci3n base del an3lisis.

Para LDA, dados los par3metros de suavizaci3n β and α , y una distribuci3n conjunta de la mezcla de t3picos θ , la idea es determinar la distribuci3n de probabilidad para generar, a partir de un conjunto de t3picos \mathcal{T} , un documento compuesto por un conjunto S de palabras w , tal que ($\mathbf{p} = (w^1, \dots, w^S)$),

$$p(\theta, z, \mathbf{p}|\alpha, \beta) = p(\theta|\alpha) \prod_{s=1}^S p(z_s|\theta)p(w^s|z_s, \beta) \quad (2.1)$$

Donde $p(z_s|\theta)$ puede ser representada por una variable aleatoria θ_i , de tal manera que el t3pico z_s puede ser presentado en el documento i ($z_s^i = 1$). Una expresi3n final puede ser deducida por la integraci3n de la ecuaci3n anterior sobre la variable aleatoria θ y la suma de los t3picos $z \in \mathcal{T}$. Dado esto, la distribuci3n marginal de un mensaje puede ser definida como::

$$p(\mathbf{w}|\alpha, \beta) = \int p(\theta|\alpha) \left(\prod_{s=1}^S \sum_{z_s \in \mathcal{T}} p(z_s|\theta)p(w^s|z_s, \beta) \right) d\theta \quad (2.2)$$

Lo que de otra manera puede ser dicho como: la combinaci3n de los t3picos generados por LDA y de las palabras que los constituyen permiten reconstruir la mayor parte del o los contenidos de texto tomados como base del an3lisis, con lo cual es posible afirmar, que los t3picos y sus palabras constitutivas describen la esencia de estos contenidos.

La ventaja de este an3lisis es que no requiere que se definan previamente la, o las ideas que sintetizan el contenido de cada t3pico, sino que estas pueden ser descubiertas y definidas por alg3n actor a partir de las palabras que el algoritmo que realiza el an3lisis va incluyendo en los t3picos.

Capítulo 3

Preparación del Experimento

Para la realización del experimento se hace previamente necesario estudiar y describir el medio, o el entorno, en donde este será realizado.

En este caso, y como se ha mencionado con anterioridad, el experimento se realizará en las páginas web que conforman el sitio Reclamos.cl, siendo estas específicamente, las páginas que corresponden a los rubros de Educación y de Gobierno, como así también, las páginas de las instituciones que pertenecen a estas industrias, y las que contienen los reclamos y las recomendaciones dirigidas hacia estos organismos.

Además, a partir de lo anterior, la metodología de trabajo y los objetivos ya planteados, también se definirán los requerimientos que debe satisfacer la aplicación informática a desarrollar con tal que permita llevar a cabo el experimento y con ello determinar las tendencias generales de los reclamos que se pretenden abordar.

3.1. Caracterización del Entorno

El experimento será realizado en las páginas web que conforman el sitio Reclamos.cl, y cuya URL corresponde a *http://www.reclamos.cl*.

Reclamos.cl es un foro público destinado a la publicación libre de experiencias y opiniones en materias de consumo por parte de sus usuarios. Busca facilitar la colaboración y encuentro entre consumidores y organizaciones comprometidas con el mejoramiento continuo y el perfeccionamiento del mercado mediante el acceso a la información. [40]

Particularmente, el experimento se realizará sólo en aquellas páginas que comprenden los rubros industriales de Educación y de Gobierno, y considerará la estructura de interconexión que estas poseen entre sí, así como también su estructura interna y su contenido escrito, descartando aquellos elementos que no realicen un aporte relevante para la consecución de los objetivos planteados.

3.1.1. Estructura del Sitio

Reclamos.cl posee varias páginas web, las que pueden ser clasificadas y agrupadas en los siguientes tipos:

- **Página Inicio, o Portada:** Esta página es la que se muestra en la URL base del sitio. Es decir en *http://www.reclamos.cl*.
- **Páginas de Secciones:** Corresponden a distintos rubros de la industria nacional, siendo algunos de estos, además de los rubros de Educación y Gobierno, los de Telecomunicaciones, Retail, Transporte, entre otros más.
- **Páginas de Sub Secciones:** Representan sub divisiones de las secciones, o en términos de la industria, sub rubros de los rubros industriales. Por ejemplo, en el rubro de Telecomunicaciones es posible encontrar los sub rubros de: Telefonía Móvil y Carriers.
- **Páginas de Instituciones:** Son las que correspondientes a las instituciones que pertenecen a cada rubro y sub rubro.
- **Páginas de Reclamos y de Recomendaciones:** Son las páginas que contienen los reclamos y las recomendaciones dirigidas hacia las instituciones.

Además existen otros tipos de páginas dentro del sitio, pero que resultan no ser relevantes para el experimento, ya que no aportan elementos útiles para la consecución de los objetivos planteados.

Dentro de estos tipos de páginas no significativas se pueden encontrar: la página para el registro de las personas como usuarios del sitio, la página en donde las personas pueden ponerse en contacto con los administradores del sitio, la página en donde las personas puedan iniciar su sesión de usuario del sitio, entre algunas otras más.

Secciones o Rubros

Las secciones o rubros, se manifiestan en páginas web cuyas URL se caracterizan por poseer la palabra “sector” y un número identificador único. Por ejemplo, el rubro de Telecomunicaciones posee la URL *http://www.reclamos.cl/sector/1462*.

Estas páginas contienen diversos elementos en su contenido, siendo los más relevantes para el experimento, los tres tipos de conjuntos de enlaces que se encuentran en ellas:

- Enlaces hacia las páginas de las sub secciones, o sub rubros, de la sección.
- Enlaces hacia las páginas de las instituciones de la sección o rubro.
- Enlaces hacia las páginas que contienen los reclamos y las recomendaciones dirigidas hacia las instituciones pertenecientes al rubro.

También poseen enlaces hacia las páginas de otras secciones del sitio y hacia la portada de este, pero para este caso no son relevantes, ya que no se apunta a acceder a dichas páginas en el experimento.

Sub secciones o Sub rubros

Las sub secciones, o sub rubros, se manifiestan en páginas web que son similares a las páginas de las secciones en términos de los tipos de enlaces que se encuentran en ellas, salvo que no poseen enlaces que apunten hacia páginas de otras sub secciones.

Los enlaces que apuntan hacia las páginas de las instituciones, de sus reclamos y recomendaciones, sólo corresponden a aquellas que pertenecen al sub rubro.

Sus URL incluyen la palabra “sector”, por lo que, por esta característica, podrían ser confundidas con las páginas correspondientes a las de secciones. Además sus URL poseen un número identificador único correspondiente a la sub sección. Por ejemplo, la URL de la sub sección de Telecomunicaciones, Telefonía Móvil es <http://www.reclamos.cl/sector/1464>.

Instituciones

Las páginas de las instituciones sólo poseen enlaces a las páginas que contienen los reclamos y las recomendaciones dirigidas hacia ellas.

En su URL incluyen la palabra “empresa” junto con un slug¹ de su nombre. Por ejemplo, la URL de la institución Telefónica del Sur es http://www.reclamos.cl/empresa/telefonica_del_sur.

Sistema de Paginación de Secciones, Sub secciones e Instituciones

Como se ha mencionado, las páginas de secciones, sub secciones y de instituciones poseen conjuntos de enlaces que apuntan hacia las páginas de reclamos y de recomendaciones que engloban, pero no de manera continúa si es que estos son más de 50. En dichos casos, estos conjuntos de enlaces se dividen en grupos de 50 dentro de varias páginas distintas que pueden ser accedidas mediante un sistema de enlaces de paginación que se encuentra en la parte inferior de cada una.

Páginas de Reclamos y Recomendaciones

Las páginas que contienen a los reclamos y a las recomendaciones son una especie de fin del camino en el recorrido por los enlaces internos del sitio, ya que si bien presentan enlaces, estos se caracterizan en su mayoría por apuntar hacia las páginas ya abordadas, como lo han sido las páginas de las secciones, las de sub secciones o las de instituciones a las que pertenecen, o a las páginas que no son relevantes para el experimento, como lo es por ejemplo, la página de inicio de sesión de usuario del sitio, por mencionar una.

Otro aspecto que es importante detallar, guarda relación con las URL de este tipo de páginas. Mientras que las páginas que contienen reclamos poseen en su URL la palabra “reclamo”, las

¹Cadena de texto en letras minúsculas, cuyos espacios han sido reemplazados por guiones, y sus caracteres no alfanuméricos han sido suprimidos, o en el caso de las vocales con tildes y las eñes, han sido reemplazados por su carácter correspondiente sin tilde y por n, respectivamente.

páginas de recomendaciones, poseen en su URL la palabra “recomendación”. Además, ambos tipos de páginas también poseen en su URL el año y el mes en que fueron publicadas, junto con un slug parcial o completo de su título, que vendría a ser una especie de identificador único de cada página.

Ejemplos de lo último, para el caso de las páginas de reclamos, sería *http://www.reclamos.cl/reclamo/2012/mar/la_polar_repactacion*, y para las páginas de recomendaciones: *http://www.reclamos.cl/recomendacion/2012/mar/ripley_mala_gestion*.

3.1.2. Estructura y Contenido de las Páginas

Las diversas páginas que constituyen al sitio de Reclamos.cl se componen de diferentes contenidos y elementos de HTML, scripts² y CSS³, aplicados e implementados de diferentes maneras según el tipo de página, ya sean estas del tipo sección, sub sección, institución, reclamo o recomendación.

Para efectos del experimento, los contenidos y los elementos estructurales más relevantes de cada página serían en gran medida los que se encuentran en el HTML de estas.

Secciones, Sub Secciones e Instituciones

Las páginas que corresponden a las secciones, sub secciones e instituciones presentan una gran similitud entre ellas en términos de los elementos que las componen y de la estructura interna de su HTML.

Todos estos tipos de páginas presentan elementos dentro de sus HTML que contienen diferentes datos, pero que para el experimento, los más relevantes son aquellos que encierran a sus respectivos nombres y a las URL que apuntan hacia las páginas de secciones, sub secciones, instituciones, reclamos y recomendaciones.

Los elementos del HTML que contienen a los nombres de estas páginas corresponden a los títulos que se encuentran representados por tags “h1”.

Mientras que las URL que apuntan hacia las páginas de interés se encuentran como atributos “href” de los tags de enlaces “a”.

En el caso de las páginas de las secciones, los enlaces que poseen, y que apuntan hacia las páginas de las sub secciones y de instituciones se encuentran contenidos en tags de HTML para listas desordenadas, es decir, en un principio por tags del tipo “ul”, para delimitar la lista de ítemes, y luego, dentro de tags “li” que delimitan cada ítem perteneciente a la lista.

En las páginas de sub secciones, es similar a las de las secciones, salvo, como se mencionaba antes, estas páginas no poseen enlaces que apunten hacia más sub secciones, por lo que no incluyen códigos de HTML que encierren listados de este tipo de enlaces.

Las páginas de las instituciones no poseen listados de enlaces hacia páginas de otras insti-

²Fundamentalmente Javascript.

³Plantillas de Estilos en Cascada, o Cascading Style Sheets en inglés.

tuciones, secciones o sub secciones, por lo que esta forma de listas no se presentan.

Por otra parte, los listados de enlaces que apuntan hacia los reclamos y las recomendaciones que pertenecen a la sección, a la sub sección, o a la institución, y que aparecen en cada una de sus respectivas páginas, se encuentran contenidos en elementos del código HTML del tipo de tablas. Esto es, se encuentran encerrados por tags como “table”, “tbody”, “tr”, y “td”.

Reclamos y Recomendaciones

Las páginas que contienen a los reclamos, y a las recomendaciones, poseen una estructura similar en el HTML que las componen, independientemente del rubro o de la institución a la que se encuentran asociadas o dirigidas.

En su interior se encuentra sólo un reclamo o una recomendación, además de otros elementos, tales como fechas, títulos, comentarios, soluciones, enlaces hacia otras páginas del sitio, siendo estos últimos no relevantes para efectos de este experimento.

En la parte superior de estas páginas está el nombre de la institución al que están asociados los reclamos y las recomendaciones respectivamente. Este nombre se presenta como un enlace hacia la página de la institución, por lo que, en términos del HTML, posee el tag “a” y la URL de la página en el atributo “href”.

Luego, más abajo, lo sigue el título del reclamo o de la recomendación, que se compone, por una parte, del nombre de la institución, y por otra, del nombre de la queja o de la sugerencia. Este título se encuentra en el HTML de las páginas como un tag “h1”.

A continuación del título, se presenta la fecha en la que se publicó el reclamo o la recomendación, junto con su número identificador único. Estos contenidos se presentan en el HTML al interior de un tag de capa “div”, que posee como parámetro de clase (class) “node-info”.

Más abajo se encuentra el texto con el contenido principal del artículo que se ubica con el tag de HTML para las capas, “div”, cuyo atributo de clase es “node-content”, y en donde cada párrafo de este posee el tag “p”.

Junto con el contenido principal del reclamo, o de la recomendación, se encuentra el porcentaje de evaluación pública que ha recibido por parte de los usuarios que han accedido a la página. Este porcentaje es el promedio de todas las valoraciones “Aceptable” (50%), “Importante” (75%), “Muy importante” (100%) o “Inapropiado” (25%) que los mismos usuarios le han dado al contenido. Con respecto a su HTML, es importante destacar que es cargado asincrónicamente en las páginas mediante AJAX, y se encuentra en el tag “div” cuyo parámetro de clase es “score”.

Posteriormente, hay otros datos, de los cuales, los más relevantes para el experimento son el número IP del computador desde donde se escribió y publicó el reclamo o la recomendación, seguido de la cantidad de lecturas (visitas de usuarios) que ha recibido, cuyo número no siempre aparece explícitamente en las páginas, sino que se encuentra como comentario en el HTML, por lo que sólo es visible revisando el código HTML y no mediante navegadores.

A continuación de todo lo anterior, y descendiendo por este tipo de páginas, se encuentran los distintos comentarios que han escrito y publicado los usuarios respecto al reclamo o la recomenda-

ción. Estos comentarios pueden ser realizados por medio de tres sistemas distintos: el de Facebook⁴, el de Disqus⁵ o el sistema de comentarios propio del sitio, siendo este último el que será abordado para lo que es la extracción de los datos de interés, ya que los otros sistemas presentan desafíos técnicos de desarrollo informático que van más allá de los alcances de este experimento.

Los comentarios de este sistema se componen de una estructura común y transversal en las distintas páginas, ya sean estas de reclamos o recomendaciones; poseyendo el nombre de fantasía del usuario (seudónimo) que lo realizó, la fecha en que se publicó, la diferencia total entre los votos a favor y en contra que han hecho los demás usuarios sobre el comentario en particular, y el texto principal del comentario.

A nivel de HTML, todos los comentarios se definen mediante un tag “div” con un atributo de clase denominado “comment”, en cuyo interior se encuentran los tags que representan a los demás elementos que los conforman. El seudónimo del usuario posee un tag “div” con el atributo de clase “comment-author”. Las fechas de publicación se encuentran dentro del mismo tag que el seudónimo, pero poseen un tag “span” con un atributo de clase “comment-fecha”. Los resultados de los votos a favor y en contra que tiene cada comentario se encuentran en un tag “span” con el atributo de clase “cscore”. Mientras que el contenido de texto de cada comentario posee un tag “div” con el atributo de clase “content”, en donde además, cada párrafo de este se encuentra delimitado por el tag “p”.

Otro aspecto importante en directa relación con los reclamos y las recomendaciones, es el hecho de que algunos de estos se presentan como “solucionados” o “respondidos”. Es decir, el autor, el moderador del sitio o la institución aludida, publica un texto, a modo de responder lo que se ha manifestado en el reclamo o en la recomendación, indicando que la situación a la que se refería ha cambiado, por ende, se entiende como que su objeto ya no es el inicial.

En términos de la estructura dentro de la página que contienen estas soluciones o respuestas, se puede decir que estas se manifiestan como uno o varios cuadros que aparecen entre el título y el cuerpo de texto principal del reclamo o la recomendación, cuyos contenidos se constituyen por la fecha en que fueron publicadas y el texto de la respuesta.

El tag de HTML que representa a estos cuadros con soluciones y respuestas es “div” y posee como atributo de clase la palabra “respuesta”. Mientras que las fechas de publicación de las soluciones se encuentran en tags “div” con el atributo de clase “date”, y el texto principal se encuentra al interior de los tags “div” con el atributo de clase “content”, y en donde cada uno de los párrafos de están delimitados por tags “p”.

3.1.3. Rubros a Abordar

Como se ha mencionado con anterioridad, el experimento se llevará a cabo sobre los reclamos que se encuentran en las secciones que corresponden a los rubros de Educación y de Gobierno (u organizaciones públicas), y cuyas direcciones web son <http://www.reclamos.cl/sector/1473> y <http://www.reclamos.cl/sector/1478> respectivamente.

⁴<http://www.facebook.com>

⁵<http://www.disqus.com>

Rubro de Educación

En la sección que aborda el rubro de Educación se encuentran los enlaces a las páginas de sub rubros tales como: Universidades, Institutos Profesionales, Preuniversitarios, Jardines Infantiles, entre otros, como así también, enlaces directos a las páginas de las instituciones que pertenecen a este rubro, y a las páginas de reclamos y recomendaciones dirigidos hacia dichas instituciones.

Rubro de Gobierno

En la sección correspondiente al rubro de Gobierno, también es posible encontrar organismos que se clasifican en sub rubros tales como: Municipalidades, Poder Judicial, Fuerzas Armadas, Ministerios, Partidos Políticos, entre varios otros más.

Y al igual que en el rubro de Educación, también se encuentran enlaces directos a las instituciones del rubro, y a los reclamos y recomendaciones dirigidos hacia ellas.

3.2. Requerimientos Asociados

Para poder identificar las tendencias generales de los reclamos publicados en Reclamos.cl y que apuntan contra las instituciones de educación y del sector público se creará y desarrollará una aplicación informática que sea capaz de extraer y procesar el contenido de estos que se encuentre en sus respectivas páginas web.

Específicamente, la aplicación informática a desarrollar deberá ser capaz de:

1. Recibir una lista de direcciones web que represente el conjunto de páginas web de los reclamos a abordar.
2. Descargar el conjunto de páginas web indicado.
3. Analizar el contenido de cada página descargada, identificando los elementos que contienen los datos de interés para el experimento, y manteniendo la estructura de agrupación y relación entre las instituciones y sus reclamos correspondientes, los reclamos y sus comentarios correspondientes, y también, los reclamos y sus respuestas o soluciones correspondientes, sólo si estas últimas existen al interior de la página web de un reclamo.
4. De los elementos identificados, extraer sus datos correspondientes tales como:
 - Institución a la que va dirigida el reclamo.
 - Fecha en que se realizó el reclamo.
 - Identificador único del reclamo.
 - Porcentaje de evaluación pública del reclamo.
 - Contenido del texto principal de reclamo.
 - Número IP del autor del reclamo.
 - Número de lecturas que ha recibido el reclamo por parte de los usuarios y visitantes del sitio web.
 - Comentarios asociados al reclamo, con su respectivo:
 - Contenido de texto principal.
 - Autor.
 - Fecha de publicación.
 - Valoración pública.
 - Si es que existen, soluciones o respuestas dadas hacia el reclamo, con su respectivo:
 - Contenido de texto principal.
 - Fecha de publicación.
5. Pre procesar los datos extraídos, suprimiendo aquellas palabras que sean considerables con bajo o nulo aporte al contenido, y lematizando a sus formas canónicas las restantes.
6. Generar grupos de palabras por cada institución abordada a través de un análisis LDA sobre el conjunto de reclamos que poseen.

Capítulo 4

Desarrollo de la Aplicación

Para la consecución de los objetivos planteados se confeccionará y desarrollará una aplicación informática que satisfaga los requerimientos mencionados con anterioridad, mediante la ejecución y la realización de diversas tareas y funciones.

Estas funciones y tareas estarán estructuradas y agrupadas dentro diferentes unidades denominadas módulos. Por ejemplo, la tarea de descargar las páginas web requeridas se encontrará dentro de uno de estos módulos.

El propósito de estructurar la aplicación en módulos, será reflejar en ellos, en la medida de lo posible, las distintas etapas de KDD para lo que es el descubrimiento de conocimientos a partir de datos.

En base a lo anterior, los módulos podrán ser ejecutados de manera autónoma, con tal de poder interactuar con cada uno independientemente de los otros, y así poder realizar calibraciones y ajustes en caso de que sea necesario. Esto en el fondo reflejará la característica interactiva e iterativa planteada para el proceso de KDD.

En términos concretos, por cuestiones de conocimientos y de experiencias previas en el desarrollo de aplicaciones informáticas, por la disponibilidad de documentación y de soporte, y por la existencia de diversas herramientas fácilmente accesibles y descargables sin costos desde Internet, la aplicación será desarrollada en el lenguaje de programación orientada a objetos Python, mientras que los datos que esta extraiga serán almacenados en una base de datos relacional soportada por el DBMS PostgreSQL [28].

4.1. Módulos de Python a Utilizar

Entre las múltiples características favorables que posee Python y que motivan su elección como el lenguaje de programación para el desarrollo de la aplicación, están la existencia de una completa documentación sobre su uso e implementación [20, 45, 46, 43, 44, 42], como así también, la existencia en Internet, y de manera gratuita, de una gran variedad de módulos, bibliotecas y librerías con diversas funciones y métodos que pueden ser fácilmente integrados entre si.

Esta última característica de Python es sumamente importante para la construcción de la aplicación, ya que al pre existir ciertas funciones y métodos potencialmente útiles para satisfacer los requerimientos y lograr los objetivos, se puede evitar confeccionar desde cero aspectos que resulten complejos de implementar por cuenta propia, hecho que genera ahorros en tiempos y en esfuerzos con respecto al trabajo de programación y de desarrollo que implica construir la aplicación que se necesita para la realización del experimento.

Dentro de los módulos de Python seleccionados para utilizar en la aplicación, se pueden encontrar: Django, Pattern, Freeling y Gensim.

4.1.1. Django

Django [23] Es un módulo principalmente enfocado a la generación de aplicaciones web y se basa en el modelo de desarrollo de tres capas: Modelo, Vista y Controlador, o MVC [24].

Para efectos de la aplicación, el aspecto más relevante que posee, guarda relación con su API¹ para la construcción y la gestión de los modelos, y que facilita enormemente varias cosas al respecto, tales como: la conexión con bases de datos (entre ellas, las que funcionan en PostgreSQL), la creación, edición, borrado y consulta sobre tablas y registros almacenados dentro de estas, entre otros.

En específico, esta API permite a través de un conjunto de comandos y métodos, realizar modificaciones y gestiones sobre la base de datos, sin necesariamente tener que recurrir a instrucciones SQL complejas, que podrían resultar difíciles de desarrollar e implementar por cuenta propia, como por ejemplo, consultas o actualizaciones en cascada sobre tablas y registros relacionados.

Además, al estar enfocado al desarrollo de aplicaciones web, Django ofrece la posibilidad de construir un sitio web, adaptado plenamente a lo requerido, en donde puedan presentarse y mostrarse los datos e informaciones generadas a partir del proceso de minería, pero que para efectos de esta memoria, no será realizado.

4.1.2. Pattern

Pattern [10] es un módulo de Python cuyos métodos que lo componen se enfocan estrictamente en permitir realizar rápida y fácilmente web mining y text mining.

Pattern trae consigo una serie de herramientas y métodos que son utilizados en estas áreas, de los cuales, los más relevantes para la aplicación y esta memoria serían:

Crawler

Pattern incluye una serie de métodos enfocados a realizar crawling sobre sitios web, es decir, descargar los diferentes documentos y archivos que podrían encontrarse dentro de un sitio, tales como imágenes, documentos en formato PDF, páginas web, archivos comprimidos, entre otros [13].

¹Interfaz de Programación de Aplicaciones, o Application Programming Interface en inglés.

Spider

Los métodos de spider, o araña de Pattern permiten identificar recursivamente los distintos enlaces que posee cada página web presentes en un sitio, es decir, estos métodos permiten identificar los enlaces que posee una página web dada, y luego, ir hacia aquellos que resulten ser otras páginas web, para así volver a repetir el proceso.

Además, dada a la documentación que aporta Pattern al respecto [13], es posible modificar los métodos asociados a la “araña”, pudiendo adecuarlos a las necesidades específicas que se posean, tal que podrían confeccionarse arañas ad hoc para las distintas estructuras de interconexión de páginas que posea cada sitio web.

Parseo de Páginas Web

Quizás uno de los aspectos que puede resultar más complicado de desarrollar y de implementar a nivel de programación para web mining es la identificación de los elementos del HTML de cada página que contienen los datos relevantes a extraer.

En estos casos, lo que se tiende a realizar es desarrollar un conjunto de comandos y de métodos que lean el código HTML de cada página como si fuesen un texto plano, y por medio de métodos de manipulación de cadenas (strings en inglés) que actúan de manera iterativa y recursiva, identificar aquellos elementos y sus respectivos contenidos para su posterior extracción.

Esta forma de realizar la extracción de los datos puede resultar algo engorrosa y lenta de desarrollar, y en el caso de minería de datos, puede resultar hasta lenta en su ejecución, dado a que la gestión de grandes cadenas de texto no es muy eficiente en términos del consumo de los recursos del sistema (memoria y capacidad de procesamiento).

Ante ello Pattern incluye un conjunto de métodos que permiten gestionar los elementos de una página web como si fuesen objetos [13], por lo que la identificación de cada elemento, y la extracción de sus datos respectivos resulta algo sumamente rápido y trivial de realizar, por no decir, eficiente en el consumo de los recursos del sistema.

Para lograrlo, estos métodos de Pattern reconstruyen lo que es el DOM de cada página web a partir del texto que corresponde a su HTML.

Pudiendo acceder a los diferentes elementos de cada página web mediante estos métodos, es posible construir rápidamente un extractor de sus datos, lo que también abre la posibilidad de construir, también rápidamente, extractores para diferentes tipos de páginas, ya sean de un mismo sitio, o de otros.

Esto hace que finalmente la extracción de datos relevantes desde páginas web algo rápidamente replicable y escalable de realizar en cualquier otro sitio distinto al de Reclamos.cl, pero que para efectos de esta memoria no será considerado.

Operaciones sobre Textos

Pattern también incluye un conjunto de métodos para realizar una serie de operaciones sobre los textos de los datos extraídos [13], como por ejemplo suprimir espacios múltiples entre las palabras, suprimir segmentos de código HTML u otro que aún puedan estar presentes en los datos extraídos, entre otros métodos más.

Para el caso del idioma español, Pattern no ofrece muchos métodos que generen un real aporte a la aplicación y a la consecución de los objetivos, pero no por ello ofrece métodos menos interesantes de conocer.

En ese sentido, se destaca que Pattern posee un conjunto de métodos para el idioma inglés [11] que lo hacen sumamente atractivo y poderoso como herramienta. Por ejemplo, posee métodos que hacen uso de los recursos lingüísticos existentes en WordNet, tales como lematizadores, diccionarios de sinónimos y antónimos, analizadores de sentimentalidad, entre otros, siendo uno de los más interesantes, un traductor de español a inglés, que si bien no es perfecto, de todas maneras presenta la posibilidad de poder utilizar los métodos de Pattern orientados hacia el inglés con los datos extraídos en español una vez que estos ya han sido traducidos.

Análisis sobre Contenido Escrito

Otros métodos relevantes que posee Pattern son los relacionados con la realización de análisis sobre los textos extraídos y procesados [12].

Si bien estos métodos que incluye Pattern no serán utilizados para esta ocasión, ya que se escapan del marco para la consecución de los objetivos planteados, de todas maneras vale la pena mencionar para explicitar lo completa que resulta ser esta herramienta.

Dentro de estos métodos se pueden encontrar el que permite la realización del análisis de frecuencia de términos en conjuntos de documentos TF-IDF², de latencia semántica LSA en los documentos³, como también de agrupamiento o clustering K-Means, y de clasificación Naive-Bayes, entre otros más.

4.1.3. Freeling

Dado a que Pattern no aporta suficientes herramientas para el procesamiento de textos en español que sean efectivamente útiles para la consecución de los objetivos, se hace necesario considerar otros recursos, como es el caso de Freeling [25].

Freeling es una biblioteca enfocada al procesamiento de textos en español, que trae consigo una serie de métodos para tales propósitos [26].

Entre los métodos que integra se encuentran la tokenización y la tipificación de las palabras (en sustantivos, adjetivos, y otros), lematización, y varios más; siendo el método de lematización el

²Frecuencia del Término-Frecuencia Inversa del Documento, o Term Frequency-Inverse Document Frequency en inglés.

³Análisis de Latencia Semántica, o Latent Semantic Analysis en inglés.

más relevante para su uso en la aplicación, ya que permite transformar las palabras de los distintos textos a analizar a su forma canónica.

En términos de aplicaciones, Freeling no es una biblioteca nativa de Python, sino que se encuentra construida con el lenguaje de programación orientada a objetos C, pudiendo ser utilizada dentro de un ambiente de Python a través de un puente o marco (wrapper) de comunicación que trae consigo.

Para su uso, sólo basta entregarle los textos a procesar a alguno de los métodos que son posibles de utilizar desde Python, para que estos los lleven hacia las librerías de Freeling, estas los procesen, y luego devuelvan los resultados correspondientes.

4.1.4. Gensim

Al igual que la sección de Pattern enfocada hacia la realización de análisis sobre textos, Gensim [63] es un módulo que incluye una serie de métodos [62] para realizar estos tipos análisis, sólo que a diferencia de Pattern, este módulo integra otros métodos que son útiles para el experimento, como lo es el análisis LDA.

Para utilizarlo, en particular, para el uso del análisis LDA que integra, hay que llamar a la función que implementa los algoritmos que se encuentran detrás de este análisis, y entregarle como parámetros el conjunto de contenidos a procesar, el número de tópicos que se desean generar, y el número de palabras que estos deberán contener; para luego, recibir una lista, en donde cada ítem corresponderá a un tópico, y en cuyo interior se encontrarán, a modo de par ordenado, las palabras que constituyen cada tópico junto con su probabilidad o nivel de relevancia relativa al tópico.

4.2. Tablas de la Base de Datos

Antes de explicar en detalle la aplicación informática que se desarrollará para la realización del experimento, se hace necesario plantear y profundizar un aspecto que no es menor en todo este asunto, y que guarda relación con dónde se almacenarán los diferentes datos que la aplicación vaya a descargar, extraer y procesar.

Para ello, se constituirán una serie de tablas dentro de una base de datos relacional gestionada mediante el motor PostgreSQL y el lenguaje SQL.

Tanto en estas tablas, como en su estructura interna, se reflejarán los distintos elementos que la aplicación considerará en las páginas web que abordará, tal que con ello puedan ser almacenados y accedidos adecuadamente todos los datos que se descargarán, extraerán y procesarán.

En términos generales, todas las tablas que se crearán poseerán un campo denominado “id”, que contendrá un número que permitirá identificar e individualizar cada registro o fila de dentro de ellas, independientemente sobre qué trate el contenido almacenado en cada tabla.

Concretamente se crearán 13 tablas distintas que se detallan en el siguiente diagrama Entidad-Relación y secciones.

Modelo Entidad-Relación

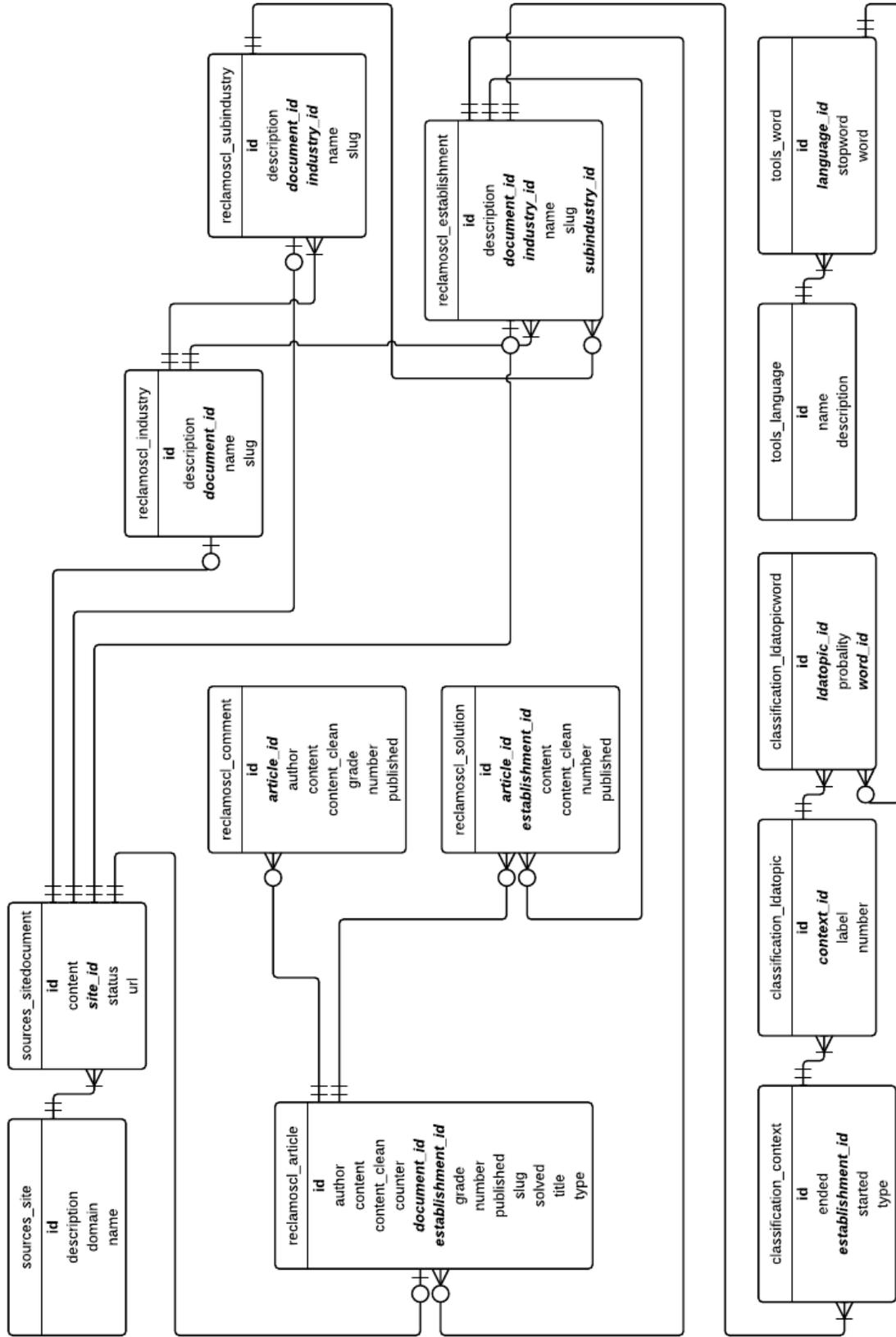


Figura 4.1: Modelo Entidad-Relación de todas las tablas que utilizará la aplicación para el almacenamiento y la consulta de los datos involucrados.

4.2.1. Tablas de Fuentes

Las tablas “fuente” corresponderán a las tablas que poseen en su nombre el prefijo “source_” y almacenarán principalmente las páginas web que se vayan a descargar del sitio de Reclamos.cl.

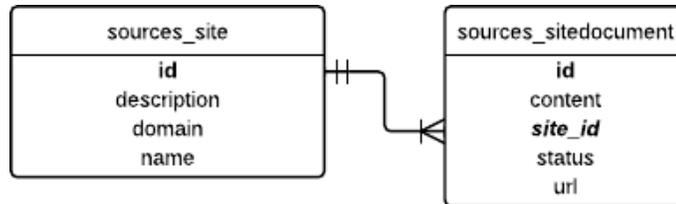


Figura 4.2: Modelo Entidad-Relación de las tablas de fuentes.

Tabla del Sitio

Particularmente, la tabla “source_site” almacenará los datos base del sitio, tal como su dominio, que en este caso es “reclamos.cl”, su nombre, “Reclamos.cl” y una descripción breve de este. Esto será ingresado manualmente en la aplicación y no se extraerá desde ninguna página web del sitio. Como se tratará de un sólo sitio, esta tabla contendrá un único registro.

Tabla de Documentos

La tabla “source_sitedocument” almacenará en cada uno de sus registros, y en los campos “url” y “content” respectivamente, la URL y el código HTML de cada página web que la aplicación descargue.

En caso de que la aplicación no pueda descargar una página en particular, esta lo indicará dejando vacío el campo “content” y colocando el valor booleano “False” en el campo “status”.

Cada página, independientemente si se descargó o no, será relacionada con el sitio al que pertenece mediante el campo “site_id”, el que contendrá al número identificador único del sitio que se encuentra como campo “id” en la tabla “source_site”, y puesto a que en el experimento se abordará sólo el sitio de Reclamos.cl, este valor será el mismo en todos los registros que se almacenen en esta tabla.

Es importante mencionar que en vez de almacenar los contenidos de las páginas descargadas al interior de una base de datos, perfectamente se podrían almacenar dentro de archivos en el disco duro de un computador. Pero esto podría dificultar, o hacer más engorrosa, la realización de ciertas operaciones sobre las páginas descargadas, tales como, operaciones de selección y de filtrado en base a sus nombres y contenidos, puesto que, a diferencia de una base de datos, este tipo de operaciones no es algo que se encuentre implementado por defecto en lo que es la gestión de archivos, por lo que se requeriría destinar tiempos y esfuerzos de programación adicionales para desarrollar un conjunto de funciones y de tareas que permitiesen llevar a cabo este tipo de operaciones de manera más expedita. En cambio, por la naturaleza funcional de las bases de datos, este tipo de operaciones ya se encuentran implementadas para que puedan ser ejecutadas de manera óptima y eficiente, por lo que

no habría que programar ni desarrollar nada extraordinario; sólo bastaría con invocarlas mediante algún comando para poder ejecutarlas.

4.2.2. Tablas de Reclamos.cl

Las tablas con el prefijo “reclamoscl_” almacenarán de manera estructurada, ordenada y tabulada, los datos que la aplicación extraiga de de los diferentes elementos de interés que conforman las páginas web que se han podido descargar del sitio y que se han almacenado en la tabla “source_sitedocument”, lo que en otras palabras sería decir, que en estas tablas se almacenarán los datos pre procesados por la aplicación.

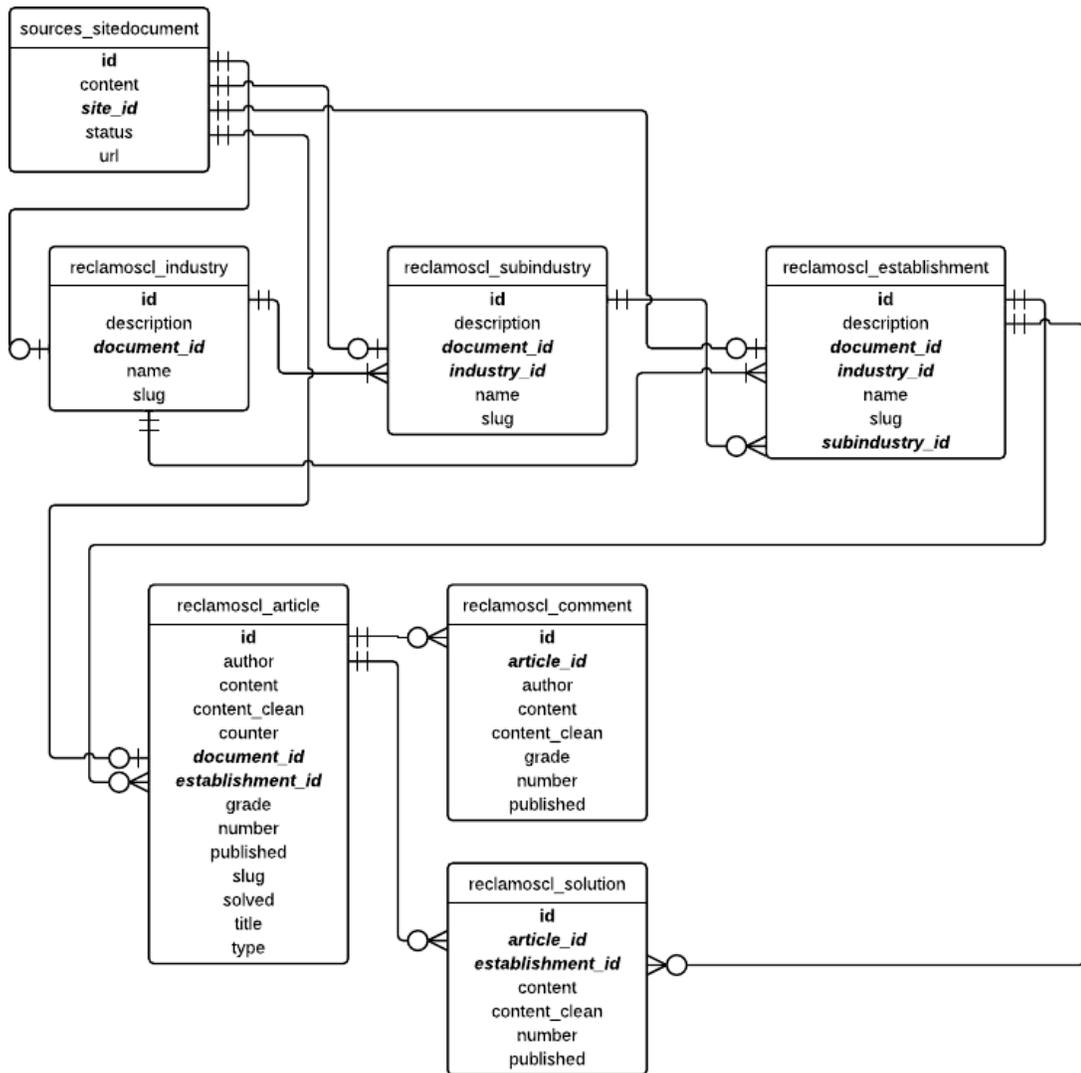


Figura 4.3: Modelo Entidad-Relación de las tablas de Reclamos.cl.

Tabla de Rubros

La tabla “reclamoscl.industry” almacenará en cada uno de sus registros los datos de cada rubro que sean posibles de extraer desde las páginas correspondientes al tipo de páginas de sección o de rubro, las que para este caso serían las páginas de los rubros de Educación y de Gobierno.

En el campo “name” se almacenará el nombre del rubro en letras minúsculas y reemplazando los tildes de las vocales y las letras eñes que pueda poseer por las vocales correspondientes sin tildes y por enes respectivamente.

El campo “slug” almacenará lo mismo del campo “name”, pero además se le reemplazará los espacios que posea por guiones “-” y se le suprimirá los caracteres no alfanuméricos que puedan encontrarse presentes.

En el campo “description” la aplicación no almacenará nada, pero de todas maneras se creará para el caso de se requiera incluir posteriormente algún comentario, descripción u observación respecto al rubro.

Cada rubro presente en esta tabla será relacionado con la página que contiene sus datos y que se encuentre en la tabla “source.sitedocument” mediante el campo “document_id”, el que contendrá al número identificador único que posea la página en la tabla de documentos, y que se ubica al interior del campo “id” de cada registro presente en ella.

Tabla de Sub rubros

La tabla “reclamoscl.subindustry” almacenará en cada uno de sus registros y campos los datos que se extraigan de las páginas correspondientes al tipo páginas de sub sección o de sub rubro, que en este caso serían las páginas de los sub rubros que son parte de los rubros de Educación y Gobierno respectivamente.

Esta tabla poseerá una estructura similar de campos y de relaciones que la tabla de rubros, salvo que contendrá un campo adicional que indicará la relación de pertenencia de cada sub rubro a un rubro. Este campo será denominado “industry_id”, y almacenará el número identificador único del rubro al que pertenece el sub rubro, el que se encuentra en la tabla de rubros dentro del campo “id” de cada registro o rubro almacenado en ella.

Tabla de Instituciones

La tabla “reclamoscl.establishment” almacenará en cada uno de sus registros y campos los datos que se extraigan de las páginas correspondientes al tipo páginas de instituciones.

Posee casi los mismos campos que la tabla de sub rubros y que la tabla de rubros, por lo que serán utilizados de manera análoga por la aplicación para almacenar los datos que extraiga sobre las instituciones.

El único campo que no aparece en las otras tablas y que se encuentra en esta es el campo “subindustry_id”, en el que la aplicación colocará el número identificador único del sub rubro al que pertenece la institución en cuestión, pero con la excepción, de que al presentarse alguna institución

que no pertenezca a un sub rubro en particular, lo llenará con el valor “null” en vez del número identificador del sub rubro.

Este número identificador del sub rubro al que pertenece cada institución (si es que pertenece a alguno) será extraído del campo “id” del registro correspondiente al sub rubro en la tabla de sub rubros.

Tabla de Artículos

La tabla “reclamoscl_article” corresponderá a la tabla en donde la aplicación almacenará los datos que extraiga de los diferentes elementos que constituyen las páginas web en las que se encuentran los reclamos y las recomendaciones, y que previamente han sido descargadas y guardadas en la tabla “source_sitedocument”.

Cada registro que poseerá esta tabla contendrá los datos de un reclamo o de una recomendación que la aplicación haya extraído desde su respectiva página web, y estarán organizados adecuadamente en distintos campos según corresponda.

En este caso también algunos campos serán similares a los de las tablas anteriores, tales como: “id”, “document_id”, y “slug”. El campo “name” que en las otras tablas almacenaba los nombres principales o más representativos de los tipos de páginas analizadas, ahora se llamará “title”, y almacenará de la misma manera el título de los reclamos o las recomendaciones.

Mientras que otros campos serán particulares y exclusivos de esta tabla, como los siguientes:

- **number:** Este es un campo de tipo numérico y contendrá el número identificador único que posee cada reclamo o recomendación según Reclamos.cl, y que la aplicación extraerá de las páginas al momento de analizarlas.
- **type:** En este campo la aplicación indicará si la página web analizada y sus respectivos datos extraídos corresponden a un reclamo o una recomendación, por lo que este campo será de tipo “booleano”, es decir, tendrá como valor “True”, si se trata de un reclamo, o un valor “False” si la página web analizada es una recomendación.
- **establishment_id:** Este campo almacenará el número identificador único presente en el campo “id” de la tabla de instituciones correspondiente al registro de la institución a la que pertenece el reclamo o la recomendación en cuestión.
- **published:** Este campo será de tipo “datetime” y almacenará la fecha en que fue realizado el reclamo o la recomendación y que la aplicación haya extraído de su página web correspondiente.
- **grade:** En este campo que será de tipo numérico, la aplicación almacenará la valoración pública que ha recibido el reclamo o la recomendación por parte de los usuarios que han votado al respecto dentro su página web correspondiente. Puesto que existen algunas páginas de reclamos o de recomendaciones que no presentan la opción de que los usuarios que las visitan puedan votar respecto a su contenido, por defecto este campo poseerá el valor “-1”, lo que permitirá diferenciar entre aquellos reclamos y recomendaciones que no han recibido votos, y por ende presentan una valoración de cero, y los reclamos o recomendaciones en donde es imposible que reciban una recomendación. Es importante recalcar que esta valoración pública

nunca es negativa, por lo que el de menos uno no generaría conflictos ni problemas con los valores posibles que podrían contener los datos que este campo almacenará.

- **content:** La aplicación almacenará en este el texto del cuerpo principal del reclamo o de la recomendación, previamente habiendo descartado cualquier tag de HTML que incluyan, ya sean estos de párrafos “p”, enlaces “a”, palabras en negritas “strong” o en cursiva “em”, o de cualquier otro tipo. Además se almacenará el texto extraído en letras minúsculas y habiendo reemplazado los tildes en las vocales y las letras eñes por sólo vocales y letras enes.
- **content_clean:** En la etapa de extracción de los datos desde las páginas web, la aplicación no almacenará ningún dato en este campo. Pero en la etapa de procesamiento de los datos extraídos, se almacenarán en este los datos que se encuentren en el campo “content” y a los que ya se les haya aplicado el borrado de stopwords y la lematización de sus palabras.
- **author:** En este campo se almacenarán los números IP correspondientes a los autores de los reclamos o las recomendaciones.
- **counter:** Será un campo de tipo numérico en el que la aplicación almacenará la cantidad de lecturas que posea cada reclamo o recomendación, y que esta haya podido extraer desde las páginas web correspondientes.
- **solved:** Es un campo de tipo “booleano” en donde la aplicación almacenará el valor “True” en caso de que el reclamo o la recomendación que se este analizando presente alguna respuesta o solución respecto a lo que plantea en su contenido de texto principal. En cambio, en caso de que no exista una respuesta o solución al respecto, la aplicación almacenará en este campo el valor “False”.

Tabla de Comentarios

La tabla “reclamoscl_comment” almacenará los datos correspondientes a los comentarios que posea cada reclamo o recomendación en su respectiva página web.

Particularmente poseerá campos similares a la tabla “reclamoscl_article”, pero con algunos cambios en la naturaleza del tipo de datos que almacenarán.

El campo “author” no almacenará un número IP como sucede en el mismo campo de la tabla que contiene los reclamos y las recomendaciones extraídas, sino que almacenará el seudónimo del autor del comentario, cuidando de que sea en letras minúsculas, sin tildes ni con letras eñes.

Mientras que el campo “grade”, a diferencia de su análogo en la tabla que contiene los datos de los reclamos y las recomendaciones, almacenará tanto números positivos como negativos, ya que en el caso de los comentarios, todos poseen la posibilidad de ser votados, partiendo con cero votos cada uno, por ende, no se dan situaciones de comentarios que no presenten el sistema de votación como sucede en el caso de los reclamos y las recomendaciones.

El campo “article_id” contendrá al número identificador único del reclamo o de la recomendación a la que pertenece cada comentario. Dicho número corresponderá al campo “id” del registro que contiene al reclamo o la recomendación asociada al comentario, y que se encuentra dentro de la tabla “reclamoscl_article”.

Tabla de Soluciones

La tabla “reclamoscl_solution” almacenará los datos que se extraigan, de cada una de las soluciones o respuestas dadas por las instituciones a cada reclamo o recomendación que han recibido en el sitio, siempre y cuando existan estas soluciones o respuestas en las páginas web correspondientes. En caso de no existir, la aplicación no almacenará ningún dato al respecto.

Esta tabla poseerá campos similares a los comentarios, siendo una de sus principal diferencias el reemplazo del campo “author” por el campo “establishment_id”, ya que las respuestas son fundamentalmente dadas por las instituciones aludidas en los reclamos o en las recomendaciones. Este campo contendrá el número identificador único de la institución y que se encontrará en su registro particular presente en la tabla de instituciones, en el campo “id”.

Además, esta tabla no poseerá el campo “grade” como sucede en la tabla de comentarios, ya que en este tipo de contenidos los usuarios no pueden realizar votaciones.

4.2.3. Tablas de Herramientas

Las tablas que cuyo prefijo en el nombre será “tools_” almacenarán fundamentalmente palabras, ya sean estas del tipo stopwords, o del contenido escrito de los distintos elementos presentes en las páginas web a analizar. Lo que en otras palabras quiere decir que estas tablas serán el diccionario de palabras la aplicación.

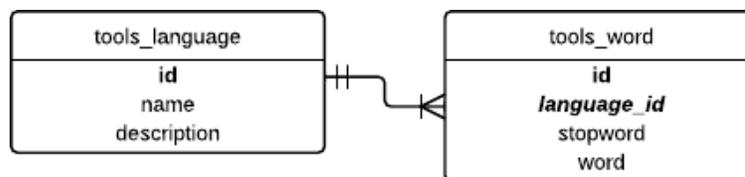


Figura 4.4: Modelo Entidad-Relación de las tablas de herramientas.

Tabla de Idiomas

La tabla “tools.language” almacenará en sus registros el nombre del idioma de cada diccionario almacenado, lo que para este caso en particular, será solamente el idioma español, por lo que en esta tabla sólo se tendrá un único registro correspondiente a dicho idioma, el que será ingresado a ella de manera manual.

Esta tabla poseerá los campos “id” que almacenará al número identificador único del idioma, “name” que contendrá al nombre del idioma, y que en este caso corresponderá a “spanish”, y también tendrá el campo “description”, en el que se podrá incluir una breve descripción del idioma si es que se hace necesario.

Tabla de Palabras

La tabla “tools_word” contendrá palabras sin repetir que sean tanto del tipo stopwords como las que no lo sean.

Esta tabla poseerá el campo “word” que almacenará las palabras en letras minúsculas, así como también el campo “stopword” que será de tipo booleano y tendrá el valor “True” si la palabra contenida en el registro es de tipo stopwords, o tendrá el valor “False” en caso contrario.

Además estará el campo “language_id” que contendrá al número identificador único del lenguaje al que corresponderá las palabras almacenadas en la tabla. Este número será obtenido del campo “id” de la tabla “tools_language”.

Inicialmente esta tabla se poblará con las palabras de tipo stopwords, lo que realizará la aplicación por medio de la extracción de estas desde un archivo de texto.

Luego, una vez que la aplicación realice el proceso de borrado de stopwords sobre el contenido escrito a analizar, esta almacenará en la tabla las palabras que queden en los contenidos, tal que puedan ser utilizadas para los procesos que siguen a continuación. Asimismo, estas palabras pueden ser marcadas como stopwords, para volver a realizar el proceso de borrado y suprimirlas de los contenidos de textos.

4.2.4. Tablas de Clasificaciones

Las tablas cuyo nombres poseerán el prefijo “classification_” serán las tablas que almacenarán los resultados que surjan de los diferentes análisis LDA que realizará la aplicación sobre los contenidos escritos y sus palabras que resten del proceso de borrado de stopwords.

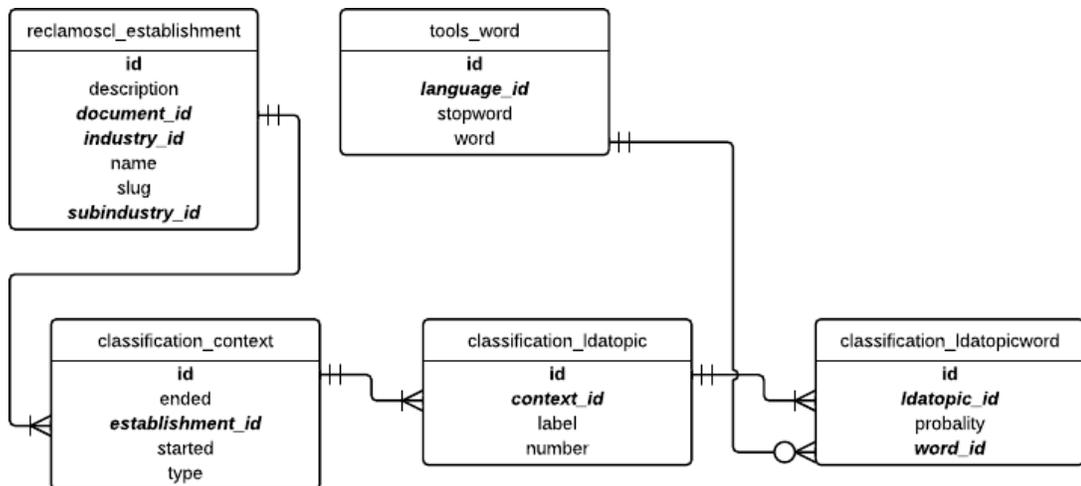


Figura 4.5: Modelo Entidad-Relación de las tablas de clasificaciones.

Tabla de Contextos

La tabla “classification_context” almacenará los contextos en los cuales se enmarcarán los análisis LDA que realizará la aplicación sobre el contenido escrito y sus palabras, ya sea tanto sobre los reclamos como también sobre las recomendaciones, siempre cuando este tipo de artículos se encuentren almacenados y tabulados en las tablas de Reclamos.cl de la base de datos.

Cada registro de la tabla contendrá uno de estos contextos, cuyos campos particulares serán:

- **establishment_id:** Este campo almacenará el número identificador único de la organización a la que estará asociado el análisis LDA que realice la aplicación. Este número será obtenido desde la tabla de instituciones.
- **stated y ended:** Estos campos serán de tipo “datetime” y almacenarán respectivamente las fechas de inicio y de término que demarcarán cronológicamente el conjunto de reclamos o recomendaciones que abordará el análisis LDA realizado aplicación.
- **type:** Será un campo de tipo “booleano” y almacenará el valor “True” si el análisis LDA que realizará la aplicación se ejecutará sobre reclamos, o “False” si el análisis se ejecutará sobre recomendaciones.

Tabla de Tópicos

La tabla “classification_ldatopic” almacenará por así decirlo, los meta datos de cada tópico o conjunto de palabras que genere cada análisis LDA ejecutado por la aplicación.

En términos concretos la tabla almacenará el número del tópico en el campo “number”, el contexto al que pertenece en el campo “context_id”, y además poseerá un campo denominado “label” para almacenar una frase, comentario o palabra que sintetice el sentido que pueda deducirse del conjunto de palabras que componen el tópico.

El campo “context_id” contendrá al número identificador único que posee el contexto al que pertenece el tópico. Este número será obtenido del registro correspondiente a dicho contexto y que se encontrará en el campo “id” de la tabla de contextos.

En el campo “label” inicialmente no se almacenará ningún dato, dejando la posibilidad de ser completado posteriormente de manera manual si es que se requiera.

Tabla de Palabras en Tópicos

La tabla “classification_ldatopicword” contendrá el conjunto de palabras, con su respectivas probabilidades, que pertenezcan a cada tópico generado por el análisis LDA que realice la aplicación sobre el conjunto de reclamos o de recomendaciones que se encuentren dentro de un determinado período de tiempo y que estén asociados a alguna institución en particular ya sea esta del rubro de Educación o el de Gobierno.

Concretamente esta tabla no almacenará a las palabras en sí, sino a su número identificador

único en el campo “word_id”, el que será obtenido desde el campo “id” perteneciente a la tabla de palabras.

Asimismo, cada registro de la tabla, también contendrá en el campo “topic_id” al número identificar único del tópico al que pertenece cada palabra, que será obtenido del campo “id” del registro correspondiente ubicado dentro de la tabla de tópicos.

El campo “probability” será un campo de tipo numérico y contendrá la probabilidad que el análisis LDA haya establecido para cada palabra.

4.3. Módulos de la Aplicación

Como se ha mencionado, la aplicación realizará una serie de tareas y funciones que permitirán satisfacer los requerimientos establecidos para así lograr la consecución de los objetivos planteados para esta memoria.

Todas estas serán realizadas, ya sea por medio del desarrollo y la programación de código en Python de creación propia, como también apoyándose en las herramientas, módulos y bibliotecas antes mencionadas.

Por ejemplo, todas las tareas que impliquen interacción con la base de datos, sus tablas, y los registros que se encuentren dentro de ellas, se realizará por medio del uso de Django y de PostgreSQL.

Pero más allá de esto, estas tareas y funciones, según su naturaleza funcional, pueden ser agrupadas en diferentes unidades internas de la aplicación denominadas módulos, los cuales se detallan a continuación:

4.3.1. Crawler y Spider

Este módulo deberá ingresar a cada página web que guarde relación con los rubros de Educación y Gobierno, identificar todos los enlaces que se encuentren presentes en ellas (función spider), descargar la página ingresada (función crawler), visitar las páginas apuntadas en los enlaces previamente identificados y volver realizar el proceso.

Particularmente la función spider será una modificación de la misma función que aporta el módulo de Python Pattern, y deberá construir una lista con los enlaces identificados en cada página que se ingrese, la que servirá como fuente de las próximas páginas a visitar. Posteriormente, esta función deberá marcar o eliminar de esta lista las páginas a las que ya se ha ingresado, con tal de evitar que la aplicación vuelva a ingresar a ellas, y con ello reducir la posibilidad de caer un bucle infinito.

Es importante señalar que esta lista de enlaces generada por la función spider, solo incluirá a aquellos que correspondan a paginas web del sitio de Reclamos.cl. Por lo tanto, no serán considerados en ella enlaces a otro tipo de documentos tales como: imágenes, documentos en formato PDF, páginas web externas al sitio, entre otros.

La función crawler tendrá como tarea específica la descarga de las páginas web a las que se ingresa, y será la que Pattern posee.

Luego, utilizando las herramientas que aporta Django para interactuar con la base de datos, todas las páginas descargadas serán almacenadas en la tabla de documentos de la base de datos. Concretamente se almacenará el código HTML de cada página y su URL asociada.

4.3.2. Extractor, Almacenador y Tabulador

Una vez finalizada la descarga y el almacenamiento en la base de datos de las páginas que guardan relación con los rubros de Educación y de Gobierno, se procederá a extraer y a tabular en las distintas tablas de Reclamos.cl los datos de interés que puedan contener cada una de estas.

Esto se realizará utilizando las herramientas de parseo de páginas web y las de gestión de datos en bases de datos que aportan los módulos de Python Pattern y Django respectivamente, así como también PostgreSQL.

A las páginas que sean del tipo rubro, se les extraerá nombre, y a partir de este se generará su slug correspondiente, luego se almacenarán como un registro en la tabla de rubros de la base de datos.

A las páginas que pertenezcan al tipo sub rubro, se les extraerá su nombre y el rubro al que pertenecen, para luego generar su slug correspondiente, y almacenar todos estos datos en la tabla de sub rubros existente en la base de datos.

Si las páginas son del tipo de instituciones, se extraerá su nombre, el rubro, y si se diese el caso, el sub rubro al que pertenecen. Con el nombre se generará el slug correspondiente, el que junto con los otros datos extraídos serán almacenados como un registro dentro de la tabla de instituciones de la base de datos.

En el caso de que las páginas sean del tipo de reclamo o de recomendación, se extraerán y almacenarán, en la tabla artículos de la base de datos, sus títulos, las fechas en que fue publicado el reclamo o la recomendación que contengan, el número identificador de estos, el porcentaje de evaluación pública que han recibido, el número IP del autor, el número de lecturas que poseen, la institución a la que están dirigidos, el contenido de texto principal (en letras minúsculas), y si corresponden precisamente a un reclamo o una recomendación. Además se identificará si poseen comentarios y respuestas, para realizar su extracción y posterior almacenamiento en las tablas de comentarios y soluciones respectivamente.

De existir respuestas o soluciones dentro de las páginas del tipo de reclamo o de recomendación, se procederá a extraer los datos correspondientes de cada una, tales como su contenido de texto principal en letras minúsculas, su número identificador, la institución que la genera, la fecha en que fue publicada, y al reclamo o la recomendación a la que se encuentran asociadas, para luego almacenarlas individualmente en los registros de la tabla de soluciones.

De la misma manera, si la página que contiene al reclamo o la recomendación presenta comentarios realizados por los usuarios, la aplicación procederá a extraer sus diferentes datos de interés, tal que luego los almacenará en la tabla de comentarios de las tablas de Reclamos.cl, cuidando siempre de que cada comentario sea ingresado dentro de un solo registro.

4.3.3. Limpieza y Procesamiento de los Textos

Cuando la aplicación haya completado el proceso de extracción de los datos de interés desde las páginas que se enmarcan dentro de los rubros de Educación y de Gobierno, así como también, finalice el almacenamiento de estos datos en las tablas correspondientes que se encuentran en la base de datos, esta deberá realizar el conjunto de tareas concernientes a la limpieza y al procesamiento de los contenidos de texto principales.

En primer lugar, la aplicación deberá obtener el contenido principal de cada reclamo o recomendación desde la tabla de artículos de la base de datos, para suprimirles todos los caracteres que no sean alfa numéricos.

Si es que no existen palabras del tipo stopword en la tabla de palabras de la base de datos, la aplicación deberá ingresarlas a ella a partir de un archivo de extensión "txt" que le será facilitado, y luego extraerlas desde ahí. Si ya existen, la aplicación sólo deberá extraerlas desde dicha tabla. En ambos casos, una vez que la aplicación extraiga las stopwords desde la tabla, y las colocara en una lista de Python para su posterior uso.

Con todo lo anterior completado, la aplicación deberá proceder a eliminar las stopwords que se encuentren presentes en los contenidos principales de los reclamos y de las recomendaciones, los que obtendrá desde la tabla de artículos existente en la base de datos.

Finalizada el borrado de stopwords de los contenidos principales de texto pertenecientes a los reclamos y a las recomendaciones almacenadas, la aplicación deberá reemplazar en las palabras restantes las vocales que posean tildes y las letras eñes por las mismas vocales sin tilde y por letras enes respectivamente.

Posteriormente, y utilizando las herramientas de procesamientos de textos que aporta la biblioteca Freeling, la aplicación deberá lematizar, o llevar a su forma canónica, las palabras que aún se encuentren en los contenidos principales de textos de los reclamos y la recomendaciones.

Ya con las stopwords eliminadas, la supresión de tildes y de letras eñes realizada, y las palabras transformadas a su forma canónica, la aplicación, por un lado, deberá individualizar estas palabras dentro cada contenido procesado, con tal de agregarlas a la tabla de palabras existente dentro la base de datos. Mientras que por otro lado, la aplicación deberá almacenar los contenidos que abarcan las palabras en el registro de la tabla de artículos correspondiente al reclamo o a la recomendación a la que pertenecen.

4.3.4. Análisis LDA

Una vez que los contenidos procesados de cada reclamo y recomendación hayan sido almacenados en la tabla de artículos, la aplicación deberá comenzar la realización de los análisis LDA sobre estos.

Para ello, primero deberá tomar por separado los conjuntos que abarcan el total de reclamos y el total de recomendaciones que posee cada una de las instituciones que fueron previamente detectadas en los rubros de Educación y de Gobierno.

Luego, la aplicación deberá identificar el rango cronológico que abarcan cada uno de estos

conjuntos, es decir, identificar la fecha de publicación del reclamo o de la recomendación más antigua, y la correspondiente a la más nuevas.

Con toda esta información la aplicación deberá crear primero, en la tabla de contextos de la base de datos, un registro que incluya la institución, si se trata de reclamos o de recomendaciones, y el rango de fechas que se abarcará, para luego comenzar a realizar el análisis LDA; tarea que ejecutará utilizando las herramientas que aporta para ello el módulo Gensim de Python.

Particularmente, dadas las características que posee el análisis LDA para su configuración base en Gensim, y a partir de la experiencia obtenida de la realización de pruebas preliminares para observar los resultados que se pueden obtener al ejecutarlo, se le indicará a la aplicación que al momento de realizar este análisis de manera definitiva sobre los conjuntos de reclamos y de recomendaciones detectados, genere a partir de estos, 20 tópicos de 30 palabras cada uno, los que serán almacenados en las tablas de tópicos y de palabras de tópicos según corresponda.

En relación a los tópicos y a las palabras que entregará el análisis LDA, es importante señalar que la aplicación deberá ser capaz de tomar cada una de estas palabras y reemplazarla por su número identificador único presente en la tabla de palabras de la base de datos, para luego almacenarlo adecuadamente, junto con su probabilidad respectiva, en la tabla de palabras en tópicos.

Con respecto a los rangos de tiempo que cubrirán los análisis LDA que realizará la aplicación, estos abarcarán tanto desde el primer, hasta el último reclamo y recomendación que ha recibido cada institución, como aquellos rangos correspondientes a cada año.

Capítulo 5

Resultados y Análisis

Como se ha mencionado con anterioridad, el experimento fue realizado por medio del desarrollo y la confección de una aplicación informática que ejecutó una serie de operaciones sobre el conjunto de páginas web de los tipos de secciones, sub secciones, instituciones, reclamos y recomendaciones pertenecientes al sitio Reclamos.cl y a los rubros industriales de Educación y de Gobierno que se encuentran al interior de este.

Dicho experimento arrojó una serie de resultados y datos que fueron almacenados por la aplicación en un conjunto de tablas ubicadas dentro de una base de datos relacional soportada por el motor PostgreSQL.

Dado lo anterior, para mostrar los resultados se confeccionaron una serie de consultas en SQL (ver Apéndice) que fueron ejecutadas sobre dichas tablas, con tal de extraer desde sus datos, tanto los resultados de los análisis realizados, como también indicadores y métricas que permitiesen ilustrar algunos aspectos sobre lo detectado en el experimento, todo ello cuidando de sintetizar lo más relevante, pues dada la cantidad de datos obtenidos, y en consecuencia, la gran cantidad de resultados generados, la extensión de esta memoria podría tornarse excesiva si es que son mostrados con todos sus detalles.

De ello, y gracias a la ejecución de las consultas y la sintetización de los resultados obtenidos, en total se pudieron detectar e identificar 74 y 171 instituciones pertenecientes a los rubros de Educación y de Gobierno, respectivamente. Las que en su conjunto poseían 6.460 reclamos y 472 recomendaciones en total, y que fueron realizadas entre el 13 de Julio de 2006 y el 5 de Diciembre de 2011, cantidades que se desglosan y profundizan a continuación.

5.1. Rubro de Educación

Particularmente, dentro del rubro de Educación se detectaron 4 sub rubros industriales, siendo estos: Instituto, Jardín Infantil, Preuniversitario y Universidad, los que poseen, respectivamente, las siguientes cantidades totales de instituciones, reclamos y recomendaciones:

Sub Rubro	Instituciones		Reclamos		Recomendaciones		Total	
Instituto	25	33,8 %	440	34,5 %	15	25,9 %	455	34,1 %
Jardín Infantil	2	2,7 %	7	0,5 %	0	0 %	7	0,5 %
Preuniversitario	3	4,1 %	50	3,9 %	2	3,4 %	52	3,9 %
Universidad	36	48,6 %	530	41,5 %	18	31 %	548	41 %
Ninguno	8	10,8 %	250	19,6 %	23	39,7 %	273	20,5 %
Totales	74	100 %	1.277	100 %	58	100 %	1.335	100 %

Cuadro 5.1: Cantidades totales de Instituciones, Reclamos y Recomendaciones existentes y realizados dentro de cada Sub rubro de Educación.

Como se puede apreciar, también se detectaron 8 instituciones que no pertenecían a ninguno de los sub rubros de Educación, tales como: Koe Corporation, Manpower S.A, entre otras.

5.1.1. Las 10 Instituciones con más Reclamos y Recomendaciones

A partir de los resultados anteriores, y en concordancia con sintetizarlos, se filtraron las 10 instituciones, de las 74 detectadas dentro del rubro de Educación, que presentaron el mayor número de reclamos y de recomendaciones.

Estas 10 instituciones filtradas fueron: Departamento Universitario Obrero Campesino (DUOC), Instituto Nacional de Capacitación Profesional (INACAP), Instituto Profesional AIEP (IP-AIEP), Universidad de las Americas (UA), Universidad Andrés Bello (UAB), Koe Corporation (KC), Universidad de Los Lagos (UL), Instituto Profesional de Chile (IP-Chile), Instituto Profesional Los Leones (IP-Leones) y la Universidad Tecnológica Metropolitana (UTEM).

Institución	Reclamos		Recomendaciones		Total	
DUOC	129	91,5 %	12	8,5 %	141	100 %
INACAP	104	96,3 %	4	3,7 %	108	100 %
IP-AIEP	104	99 %	1	1 %	105	100 %
UA	83	95,4 %	4	4,6 %	87	100 %
UAB	68	100 %	0	0 %	68	100 %
KC	65	90,3 %	7	9,7 %	72	100 %
UL	45	95,7 %	2	4,3 %	47	100 %
IP-Chile	43	95,6 %	2	4,4 %	45	100 %
IP-Leones	43	95,6 %	2	4,4 %	45	100 %
UTEM	34	100 %	0	0 %	34	100 %

Cuadro 5.2: Las 10 Instituciones con mayor cantidad de Reclamos y Recomendaciones dentro del rubro de Educación.

Detalle de los Reclamos y de las Recomendaciones

Profundizando en los reclamos y en las recomendaciones realizados hacia estas 10 instituciones, también fue posible identificar el número de comentarios que recibieron por parte de los usuarios que los leían, además de detectar si fueron respondidos o resueltos, así como también, el número de respuestas y de soluciones que fueron dadas al respecto, lo cual se detalla a continuación:

Institución	Reclamos	Respondidos	Respuestas	Comentarios	
DUOC	129	0	0 %	0	1.066
INACAP	104	0	0 %	0	591
IP-AIEP	104	0	0 %	0	641
UA	83	1	1,2 %	1	761
UAB	68	0	0 %	0	446
KC	65	1	1,5 %	1	622
UL	45	0	0 %	0	177
IP-Chile	43	0	0 %	0	627
IP-Leones	43	0	0 %	0	390
UTEM	34	0	0 %	0	241

Cuadro 5.3: Detalle de la cantidad total de Reclamos, Reclamos Respondidos, Respuestas y Comentarios realizados hacia las 10 Instituciones con más Reclamos y Recomendaciones del rubro de Educación.

Institución	Recomendaciones	Respondidas	Respuestas	Comentarios	
DUOC	12	0	0 %	0	92
INACAP	4	0	0 %	0	39
IP-AIEP	1	0	0 %	0	13
UA	4	0	0 %	0	26
UAB	0	0	0 %	0	0
KC	7	0	0 %	0	120
UL	2	0	0 %	0	2
IP-Chile	2	0	0 %	0	17
IP-Leones	2	0	0 %	0	3
UTEM	0	0	0 %	0	0

Cuadro 5.4: Detalle de la cantidad total de Recomendaciones, Recomendaciones Respondidas, Respuestas y Comentarios realizados hacia las 10 Instituciones con más Reclamos y Recomendaciones del rubro de Educación.

5.1.2. Departamento Universitario Obrero Campesino, DUOC

Ahondando en los resultados obtenidos para el DOUC, pues, sin pérdida de generalidad, para las otras instituciones resulta ser análogo; fue posible abordar detalles cronológicos con respecto a la publicación de los reclamos y de las recomendaciones dirigidas hacia la institución, así como también, fue posible, mediante la ejecución de un análisis LDA, generar los tópicos con las palabras que reconstruyen los reclamos, asunto fundamental para la consecución de los objetivos.

Detalle de los Reclamos y de las Recomendaciones

En las siguientes dos tablas se detallan las cantidades de reclamos y de recomendaciones que recibió la institución dentro de cada año, así como también, las cantidades de estos que fueron respondidos, las cantidades de respuestas dadas al respecto, y la cantidad de comentarios generados hacia ellos por parte de los usuarios que los leían.

Año	Reclamos		Respondidos		Respuestas		Comentarios	
2007	3	2,4 %	0	0 %	0	0 %	207	19,4 %
2008	15	11,6 %	0	0 %	0	0 %	169	15,9 %
2009	25	19,4 %	0	0 %	0	0 %	169	15,9 %
2010	43	33,3 %	0	0 %	0	0 %	351	32,9 %
2011	43	33,3 %	0	0 %	0	0 %	170	15,9 %
Totales	129	100 %	0	0 %	0	0 %	1.066	100 %

Cuadro 5.5: Detalle cronológico de las cantidades totales de Reclamos, Reclamos Respondidos, Respuestas y Comentarios realizados hacia el Departamento Universitario Obrero Campesino DUOC.

Año	Recomendaciones		Respondidas		Respuestas		Comentarios	
2007	0	0 %	0	0 %	0	0 %	0	0 %
2008	0	0 %	0	0 %	0	0 %	0	0 %
2009	0	0 %	0	0 %	0	0 %	0	0 %
2010	0	0 %	0	0 %	0	0 %	0	0 %
2011	12	100 %	0	0 %	0	0 %	92	100 %
Totales	12	100 %	0	0 %	0	0 %	92	100 %

Cuadro 5.6: Detalle cronológico de las cantidades totales de Recomendaciones, Recomendaciones Respondidas, Respuestas y Comentarios hechas hacia el Departamento Universitario Obrero Campesino DUOC.

Análisis LDA

Sobre el conjunto de reclamos dirigidos contra la institución se realizaron análisis LDA, los que consistieron en generar 20 tópicos, o grupos de palabras, compuestos por 30 palabras cada uno,

tanto a nivel general, abarcando el conjunto completo de reclamos publicados entre el año 2007 y el año 2011, así como también, por segmentos, abarcando los reclamos publicados durante cada uno de estos años.

Las tablas que se muestran a continuación, poseen los resultados de estos análisis LDA, estructurandese de manera tal que los tópicos generados corresponden a las columnas, y las palabras que los conforman, junto con sus respectivas probabilidades asociadas, corresponden a las filas.

Las probabilidades asociadas a cada palabra, en este caso sólo sirven para ilustrar la preponderancia, o relevancia, que posee cada palabra al interior de cada tópico.

Resultados Análisis LDA DOUC 2007 A

	Tópico 1		Tópico 2		Tópico 3		Tópico 4		Tópico 5		Tópico 6		Tópico 7		Tópico 8		Tópico 9		Tópico 10	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	carrera	0.024	feliz	0.006	rayssa	0.006	encuentro	0.006	no	0.041	no	0.026	rayssa	0.006	feliz	0.006	encuentro	0.006	feliz	0.006
2	ramo	0.023	encuentro	0.006	no	0.006	feliz	0.006	carrera	0.026	carrera	0.022	feliz	0.006	no	0.006	rayssa	0.006	encuentro	0.006
3	publicidad	0.022	ramo	0.006	feliz	0.006	rayssa	0.006	dtoc	0.021	ramo	0.016	ramo	0.006	rayssa	0.006	feliz	0.006	rayssa	0.006
4	tener	0.02	rayssa	0.006	insol	0.006	dtoc	0.006	clase	0.02	publicidad	0.017	dtoc	0.006	insol	0.006	no	0.006	insol	0.006
5	no	0.017	dtoc	0.006	ramo	0.006	carrera	0.006	ramo	0.018	tener	0.016	encuentro	0.006	ramo	0.006	pagar	0.006	ramo	0.006
6	pintar	0.013	insol	0.006	dtoc	0.006	ramo	0.006	ano	0.014	pintar	0.013	carrera	0.006	dtoc	0.006	ramo	0.006	no	0.006
7	historia	0.013	pagar	0.006	encuentro	0.006	insol	0.006	estudio	0.013	ano	0.012	no	0.006	abusivo	0.006	abusivo	0.006	dtoc	0.006
8	tecnología	0.013	mensual	0.006	realmente	0.006	abusivo	0.006	gran	0.012	tecnología	0.012	insol	0.006	encuentro	0.006	dtoc	0.006	abusivo	0.006
9	ano	0.012	no	0.006	semestre	0.006	no	0.006	actividad	0.012	dtoc	0.01	nombre	0.006	carrera	0.006	carrera	0.006	unico	0.006
10	computacion	0.012	carrera	0.006	abusivo	0.006	semestre	0.006	prestigio	0.011	historia	0.01	unico	0.006	mensual	0.006	unico	0.006	semestre	0.006
11	nuevo	0.012	semestre	0.006	pagar	0.006	mensual	0.006	entregar	0.011	nuevo	0.01	abusivo	0.006	unico	0.006	nombre	0.006	pagar	0.006
12	maqueta	0.011	nombre	0.006	mensual	0.006	pagar	0.006	profesor	0.011	enseñar	0.01	realmente	0.006	pagar	0.006	semestre	0.006	mensual	0.006
13	importancia	0.011	unico	0.006	nombre	0.006	alumno	0.006	nada	0.01	clase	0.009	pagar	0.006	semestre	0.006	insol	0.006	carrera	0.006
14	tecnica	0.011	abusivo	0.006	unico	0.006	realmente	0.006	enseñar	0.01	computacional	0.008	alumno	0.006	publicidad	0.006	mensual	0.006	realmente	0.006
15	retirar	0.01	alumno	0.006	carrera	0.006	permitir	0.006	sala	0.01	director	0.008	semestre	0.006	permitir	0.006	alumno	0.006	nombre	0.006
16	director	0.01	realmente	0.006	tener	0.006	unico	0.006	publicidad	0.01	digital	0.008	publicidad	0.006	alumno	0.006	realmente	0.006	permitir	0.006
17	malo	0.01	permitir	0.006	alumno	0.006	nombre	0.006	profe	0.01	exitola	0.008	mensual	0.006	nombre	0.006	publicidad	0.006	mi	0.006
18	grafo	0.01	mi	0.006	permitir	0.006	tener	0.006	trabajar	0.01	dinero	0.008	permitir	0.006	realmente	0.006	permitir	0.006	publicidad	0.006
19	rematar	0.01	clase	0.006	publicidad	0.006	mi	0.006	matricular	0.009	matricular	0.008	mi	0.006	mi	0.006	tener	0.006	alumno	0.006
20	debi	0.009	tener	0.006	ano	0.006	publicidad	0.006	hacer	0.009	hacer	0.008	tener	0.006	tener	0.006	clase	0.006	tener	0.006
21	publicitar	0.009	ano	0.006	historia	0.006	ano	0.006	tener	0.009	terminar	0.008	clase	0.006	ano	0.006	mi	0.006	clase	0.006
22	querer	0.009	publicidad	0.006	mi	0.006	clase	0.006	director	0.008	tecnica	0.008	ano	0.006	director	0.006	ano	0.006	pintar	0.006
23	entender	0.009	historia	0.006	enseñar	0.006	pintar	0.006	pagar	0.008	ilustracion	0.007	historia	0.006	clase	0.006	pintar	0.006	ano	0.006
24	similar	0.009	director	0.006	estudiar	0.006	historia	0.006	estudiar	0.008	debi	0.007	pintar	0.006	historia	0.006	estudiar	0.006	prestigio	0.006
25	estafar	0.009	estudiar	0.006	pintar	0.006	hacer	0.006	seccion	0.008	diseñador	0.007	enseñar	0.006	pintar	0.006	tecnología	0.006	estudiar	0.006
26	vendio	0.009	pintar	0.006	clase	0.006	director	0.006	suponer	0.007	persona	0.007	nuevo	0.006	tecnología	0.006	historia	0.006	estudio	0.006
27	ilustracion	0.009	gran	0.006	tecnología	0.006	tecnología	0.006	senor	0.007	estudiar	0.007	estudiar	0.006	matricular	0.006	enseñar	0.006	hacer	0.006
28	supuestamente	0.009	hacer	0.006	prestigio	0.006	matricular	0.006	solicitar	0.007	computacion	0.007	director	0.006	prestigio	0.006	hacer	0.006	director	0.006
29	titular	0.009	tecnología	0.006	hacer	0.006	tecnica	0.006	perder	0.007	importante	0.007	estudio	0.006	estudiar	0.006	director	0.006	gran	0.006
30	computacional	0.009	enseñar	0.006	tecnica	0.006	computacional	0.006	alumno	0.007	vendio	0.007	tecnología	0.006	profe	0.006	gran	0.006	historia	0.006

Cuadro 5.7: Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2007.

Resultados Análisis LDA DOUC 2007 B

	Tópico 11		Tópico 12		Tópico 13		Tópico 14		Tópico 15		Tópico 16		Tópico 17		Tópico 18		Tópico 19		Tópico 20	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	no	0,027	rayssa	0,006	ramo	0,006	duoc	0,039	feliz	0,006	no	0,039	ramo	0,006	ramo	0,006	rayssa	0,043	rayssa	0,006
2	clase	0,02	no	0,006	encuentro	0,006	no	0,03	rayssa	0,006	clase	0,028	rayssa	0,006	feliz	0,006	no	0,033	ramo	0,006
3	duoc	0,018	feliz	0,006	rayssa	0,006	insol	0,026	no	0,006	duoc	0,021	encuentro	0,006	carrera	0,006	tener	0,032	feliz	0,006
4	carrera	0,017	ramo	0,006	no	0,006	encuentro	0,026	carrera	0,006	carrera	0,019	insol	0,006	rayssa	0,006	publicidad	0,031	no	0,006
5	pagar	0,015	encuentro	0,006	insol	0,006	feliz	0,026	encuentro	0,006	profe	0,017	feliz	0,006	duoc	0,006	carrera	0,03	duoc	0,006
6	gran	0,014	insol	0,006	feliz	0,006	rayssa	0,026	ramo	0,006	prestigio	0,017	duoc	0,006	encuentro	0,006	nuevo	0,024	insol	0,006
7	trabajar	0,012	duoc	0,006	carrera	0,006	pagar	0,024	pagar	0,006	nada	0,017	carrera	0,006	insol	0,006	historia	0,023	encuentro	0,006
8	profe	0,012	unico	0,006	nombre	0,006	clase	0,021	insol	0,006	entregar	0,016	unico	0,006	no	0,006	tecnología	0,022	carrera	0,006
9	sala	0,012	abusivo	0,006	duoc	0,006	alumno	0,019	semestre	0,006	pagar	0,016	pagar	0,006	abusivo	0,006	pintar	0,021	abusivo	0,006
10	entregar	0,011	mensual	0,006	mensual	0,006	unico	0,019	abusivo	0,006	gran	0,015	mensual	0,006	pagar	0,006	ano	0,021	mensual	0,006
11	actividad	0,011	semestre	0,006	pagar	0,006	ramo	0,019	nombre	0,006	actividad	0,015	abusivo	0,006	mensual	0,006	convensar	0,013	pagar	0,006
12	prestigio	0,011	nombre	0,006	unico	0,006	carrera	0,017	mensual	0,006	profesor	0,015	no	0,006	permitir	0,006	dar	0,012	unico	0,006
13	profesor	0,011	alumno	0,006	alumno	0,006	trabajar	0,014	duoc	0,006	estudio	0,014	semestre	0,006	semestre	0,006	decir	0,012	nombre	0,006
14	estudio	0,01	realmente	0,006	abusivo	0,006	mi	0,013	permitir	0,006	sala	0,012	alumno	0,006	realmente	0,006	finalmente	0,012	realmente	0,006
15	nada	0,01	carrera	0,006	semestre	0,006	permitir	0,013	alumno	0,006	trabajar	0,01	realmente	0,006	decidi	0,012	semestre	0,006	0,006	0,006
16	plano	0,009	pagar	0,006	realmente	0,006	realmente	0,013	unico	0,006	junto	0,01	permitir	0,006	acusar	0,012	alumno	0,012	alumno	0,006
17	estudiar	0,009	mi	0,006	mi	0,006	nombre	0,013	realmente	0,006	ofrecer	0,009	ano	0,006	dos	0,012	mi	0,012	mi	0,006
18	pagar	0,008	publicidad	0,006	permitir	0,006	semestre	0,013	tener	0,006	calidad	0,009	nombre	0,006	publicidad	0,006	estudiar	0,012	permitir	0,006
19	servir	0,008	permitir	0,006	publicidad	0,006	mensual	0,013	clase	0,006	quedar	0,009	mi	0,006	clase	0,006	diario	0,012	publicidad	0,006
20	oeste	0,008	tener	0,006	tener	0,006	abusivo	0,013	publicidad	0,006	inútiles	0,009	tener	0,006	mi	0,006	q	0,012	ano	0,006
21	puestoq	0,008	ano	0,006	clase	0,006	sala	0,012	mi	0,006	pegar	0,009	publicidad	0,006	tecnología	0,006	resumir	0,012	tener	0,006
22	especial	0,008	pintar	0,006	ano	0,006	profesor	0,011	ano	0,006	curso	0,009	pintar	0,006	mouse	0,012	pintar	0,012	pintar	0,006
23	unico	0,008	clase	0,006	pintar	0,006	estudio	0,01	pintar	0,006	separar	0,009	tecnología	0,006	pintar	0,006	embellecer	0,012	tecnología	0,006
24	seccion	0,008	enseñar	0,006	hacer	0,006	nada	0,01	nuevo	0,006	hacer	0,009	clase	0,006	ano	0,006	leer	0,012	estudiar	0,006
25	perder	0,008	matricular	0,006	nuevo	0,006	actividad	0,01	historia	0,006	abrir	0,009	estudiar	0,006	director	0,006	sentar	0,012	director	0,006
26	mayoria	0,008	gran	0,006	prestigio	0,006	entregar	0,01	matricular	0,006	gra	0,009	actividad	0,006	nuevo	0,006	cuatro	0,012	clase	0,006
27	matricular	0,008	nuevo	0,006	estudio	0,006	profe	0,01	tecnología	0,006	menos	0,009	hacer	0,006	estudiar	0,006	duoc	0,012	enseñar	0,006
28	estero	0,008	historia	0,006	director	0,006	prestigio	0,01	director	0,006	reunión	0,009	nuevo	0,006	enseñar	0,006	hacer	0,012	trabajar	0,006
29	nisiquiera	0,008	hacer	0,006	estudiar	0,006	gran	0,009	enseñar	0,006	plano	0,009	gran	0,006	estudio	0,006	persona	0,012	gran	0,006
30	catalogar	0,008	computacion	0,006	actividad	0,006	musicar	0,007	estudiar	0,006	tec	0,009	profesor	0,006	hacer	0,006	matricular	0,012	nuevo	0,006

Cuadro 5.8: Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2007.

Resultados Análisis LDA DOUC 2008 A

Tópico 1		Tópico 2		Tópico 3		Tópico 4		Tópico 5		Tópico 6		Tópico 7		Tópico 8		Tópico 9		Tópico 10	
Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	no	0.032	no	0.034	profesor	0.015	no	0.023	no	0.027	no	0.037	no	0.050	no	0.048	alumno	0.017	
2	semestre	0.014	ir	0.016	diuoc	0.021	malo	0.011	tema	0.018	sede	0.013	carrera	0.018	diuoc	0.016	carrera	0.026	no
3	sede	0.014	diuoc	0.016	alumno	0.018	nada	0.011	contrato	0.014	semestre	0.010	problema	0.014	carrera	0.015	ano	0.023	insol
4	clase	0.013	decir	0.015	hacer	0.018	exámenes	0.011	problema	0.014	carrera	0.010	respuesta	0.014	ramo	0.012	profesor	0.019	nunca
5	hacer	0.013	poder	0.014	reemplazante	0.015	nota	0.011	cancelar	0.013	nuestro	0.009	semestre	0.013	tomar	0.012	ramo	0.018	clase
6	día	0.012	tema	0.012	enviar	0.013	dos	0.011	dicon	0.012	diuoc	0.009	profesor	0.012	dejar	0.011	inventar	0.014	querer
7	vez	0.010	llamar	0.011	profesor	0.012	impartir	0.010	tener	0.010	respuesta	0.009	ano	0.012	alumno	0.013	hacer	0.009	
8	ser	0.010	practicar	0.011	llevar	0.012	institución	0.010	estar	0.008	alumno	0.009	haber	0.010	sep	0.010	cuatro	0.013	0.008
9	compromiso	0.009	tener	0.011	respetar	0.008	semestre	0.009	solicitar	0.008	exámenes	0.008	exámenes	0.009	clase	0.010	titular	0.009	cambiar
10	tener	0.009	problema	0.011	cliente	0.008	construccion	0.009	diuoc	0.008	ramo	0.008	trabajo	0.009	profesor	0.010	lamentablemente	0.008	respeto
11	material	0.009	trabajar	0.010	sede	0.008	no	0.009	pasar	0.008	posible	0.007	levar	0.009	querer	0.009	estudiar	0.008	absolutamente
12	persona	0.009	cancelar	0.009	nunca	0.007	compañeros	0.009	dejar	0.007	llevar	0.007	jefe	0.009	uc	0.008	turismo	0.008	respuesta
13	diuoc	0.009	contrato	0.009	insol	0.007	sede	0.008	decir	0.007	problema	0.007	llevar	0.009	legar	0.008	compañeros	0.008	coordinar
14	institución	0.008	ano	0.009	enseñar	0.007	nuestro	0.008	trabajar	0.007	ano	0.007	sede	0.008	cambiar	0.008	problema	0.008	persona
15	uc	0.008	semestre	0.008	carrera	0.007	ramo	0.007	casa	0.007	profesor	0.007	estudiar	0.008	tener	0.007	ser	0.007	sap
16	exámenes	0.008	dejar	0.008	dos	0.006	falto	0.007	poder	0.006	reemplazante	0.007	siempre	0.008	haber	0.007	mismo	0.007	necesitar
17	pasar	0.008	casa	0.007	ano	0.006	trabajo	0.007	sentar	0.005	nota	0.007	posible	0.008	calidad	0.007	instituto	0.007	institución
18	nada	0.008	día	0.007	consertar	0.005	perla	0.007	enviar	0.005	jaramillo	0.006	victor	0.007	pasar	0.007	rellenar	0.007	actitud
19	profesor	0.007	dicon	0.007	avisar	0.005	cara	0.006	informar	0.005	jefe	0.006	institucion	0.007	estudiar	0.007	curricular	0.007	cita
20	calidad	0.007	tomar	0.006	perder	0.005	clase	0.006	estimar	0.005	vez	0.006	nota	0.007	jefe	0.006	esfuerzo	0.007	mil
21	muestro	0.007	habia	0.006	tener	0.005	transversal	0.006	mano	0.005	siempre	0.005	nada	0.007	semestre	0.006	malla	0.007	aludir
22	alumno	0.007	pedir	0.006	disponible	0.005	acreditar	0.006	sernac	0.005	claudilo	0.005	ingeniería	0.007	institucion	0.006	llenar	0.007	servir
23	nivel	0.007	marzo	0.006	valioso	0.005	prestigioso	0.006	incendio	0.005	coordinador	0.005	ramo	0.007	ano	0.006	pegar	0.007	tener
24	osa	0.006	estudiar	0.006	coordinar	0.005	diuoc	0.006	nadien	0.005	haber	0.005	coordinador	0.007	problema	0.006	suerte	0.007	disponible
25	andar	0.006	haber	0.006	tonito	0.005	negocio	0.006	rector	0.005	victor	0.005	solicitar	0.006	trabajar	0.005	valorar	0.007	quedar
26	cara	0.006	dar	0.006	convencer	0.005	dosente	0.006	capital	0.005	hacer	0.005	jaramillo	0.006	matricular	0.005	estudio	0.007	uc
27	posible	0.006	estar	0.005	point	0.005	aviso	0.006	empresa	0.005	dar	0.004	diuoc	0.006	exámenes	0.005	propio	0.007	llevar
28	director	0.006	pagar	0.005	persona	0.005	envia	0.006	estudiar	0.005	trabajo	0.004	cambiar	0.006	sorpresa	0.005	duro	0.006	tú
29	pedir	0.006	mal	0.005	institucion	0.005	lado	0.006	llegar	0.005	dos	0.004	tener	0.006	obviamente	0.005	ir	0.006	coordinar
30	ramo	0.006	matricular	0.005	esperar	0.005	debe	0.006	tomar	0.005	avisar	0.004	nunca	0.006	nada	0.005	enseñar	0.006	srta

Cuadro 5.9: Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2008.

Resultados Análisis LDA DOUC 2008 B

	Tópico 11		Tópico 12		Tópico 13		Tópico 14		Tópico 15		Tópico 16		Tópico 17		Tópico 18		Tópico 19		Tópico 20	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	no	0.013	no	0.032	no	0.060	no	0.002	no	0.035	no	0.017	semestre	0.045	no	0.011	no	0.036	no	0.033
2	practicar	0.010	clase	0.022	alumno	0.017	duoc	0.002	profesor	0.028	profesor	0.013	institucion	0.024	sede	0.009	semestre	0.022	profesor	0.015
3	llamar	0.009	profesor	0.014	duoc	0.015	carrera	0.002	carrera	0.023	carrera	0.010	no	0.024	clase	0.007	nada	0.018	ir	0.015
4	ir	0.009	ramo	0.014	hacer	0.014	semestre	0.002	llegar	0.019	llegar	0.009	carrera	0.022	ser	0.006	sede	0.014	duoc	0.014
5	dia	0.009	nada	0.013	carrera	0.011	alumno	0.002	alumno	0.018	ver	0.008	realizar	0.022	persona	0.006	profesor	0.013	carrera	0.013
6	poder	0.007	carrera	0.011	ramo	0.009	reclamar	0.002	reclamar	0.017	ano	0.008	reclamo	0.021	vez	0.006	hacer	0.012	practicar	0.012
7	decir	0.007	duoc	0.011	ano	0.009	clase	0.002	jefe	0.016	estudiar	0.008	sede	0.014	estudiar	0.006	muestro	0.011	ano	0.010
8	ofrecer	0.007	querer	0.010	profesor	0.008	profesor	0.002	estudiar	0.016	tener	0.007	exámenes	0.013	material	0.005	institucion	0.011	estudiar	0.010
9	habia	0.006	ano	0.010	pasar	0.008	ano	0.002	practicar	0.013	querer	0.007	nuestro	0.013	dia	0.005	duoc	0.011	problema	0.009
10	mal	0.006	llegar	0.009	enviar	0.008	jefe	0.002	calidad	0.012	nota	0.006	uc	0.013	compromiso	0.005	ano	0.010	dia	0.008
11	semestre	0.005	tomar	0.009	nunca	0.007	sede	0.002	instiituo	0.011	unico	0.006	alameda	0.012	haber	0.005	clase	0.010	llamar	0.008
12	problema	0.005	alumno	0.009	llamar	0.007	nuestro	0.002	encontrar	0.011	duoc	0.006	duoc	0.012	profesor	0.005	carrera	0.010	poder	0.008
13	duoc	0.005	resultar	0.007	ir	0.007	uc	0.002	malo	0.011	hito	0.006	dar	0.012	tener	0.005	ramo	0.010	tomar	0.007
14	dar	0.005	nunca	0.007	insol	0.007	hacer	0.002	momento	0.010	enseñar	0.006	alumno	0.011	cara	0.005	camilla	0.009	decir	0.007
15	trabajar	0.005	institucion	0.007	decir	0.007	exámenes	0.002	haber	0.010	falta	0.006	jefe	0.011	hora	0.004	exámenes	0.009	dar	0.007
16	tomar	0.005	peor	0.007	practicar	0.007	institucion	0.002	duoc	0.008	ingle	0.005	igual	0.011	dar.cuenta	0.004	problema	0.008	haber	0.007
17	marzo	0.005	sap	0.007	reemplazante	0.006	estudiar	0.002	semestre	0.007	ir	0.005	senores	0.011	revisar	0.004	alumno	0.008	tener	0.007
18	hablar	0.005	jefe	0.007	dia	0.006	pasar	0.002	mes	0.007	problema	0.005	ingenieria	0.011	calidad	0.004	ingenieria	0.008	matricular	0.006
19	pedir	0.004	sede	0.007	respuesta	0.006	companioneros	0.002	problema	0.007	nuestro	0.005	atencion	0.011	corredor	0.004	trabajo	0.008	calidad	0.006
20	ano	0.004	hacer	0.007	llevar	0.006	ser	0.002	pasar	0.007	realidad	0.005	razon	0.011	fojo	0.004	nuca	0.008	trabajo	0.006
21	matricular	0.004	tener	0.006	esperar	0.006	nada	0.002	despedir	0.006	haber	0.005	revisar	0.011	queria	0.004	luis	0.008	ramo	0.006
22	podia	0.004	semestre	0.006	tomar	0.006	dia	0.002	taller	0.006	alumno	0.005	despreocupacion	0.011	preguntar	0.004	nota	0.007	semestre	0.006
23	dejar	0.004	haber	0.006	matricular	0.006	problema	0.002	aconsejar	0.006	titular	0.005	otorgar	0.011	esperar	0.004	siempre	0.007	cambiar	0.006
24	insitituo	0.004	ingenieria	0.006	uc	0.005	nunca	0.002	certames	0.006	hijo	0.005	comentario	0.011	pasar	0.004	jefe	0.007	unico	0.006
25	paga	0.003	uc	0.006	problema	0.005	trabajo	0.002	respuesta	0.006	dejar	0.005	cuando	0.011	nivel	0.004	senor	0.007	llegar	0.006
26	uribe	0.003	senor	0.006	calidad	0.005	dejar	0.002	compromiso	0.006	calidad	0.005	particular	0.011	trabajar	0.004	posible	0.007	alumno	0.006
27	pasar	0.003	matricular	0.006	clase	0.005	reclamo	0.002	enviar	0.006	trabajar	0.005	confenmidad	0.011	charlatán	0.004	uc	0.006	pasar	0.005
28	boy	0.003	estudiar	0.006	dos	0.005	calidad	0.002	ano	0.006	asistente	0.005	servir	0.011	dichoso	0.004	impartir	0.006	hablar	0.005
29	enfotraba	0.003	calidad	0.005	dar	0.005	tener	0.002	considerar	0.006	examen	0.005	causar	0.011	comprometer	0.004	escuela	0.006	nada	0.005
30	perdar	0.003	academia	0.005	cambiar	0.005	tomar	0.002	generacion	0.006	institucion	0.005	prevencion	0.011	corresponder	0.004	resultar	0.005	trabajar	0.005

Cuadro 5.10: Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2008.

Resultados Análisis LDA DOUC 2009 A

	Tópico 1		Tópico 2		Tópico 3		Tópico 4		Tópico 5		Tópico 6		Tópico 7		Tópico 8		Tópico 9		Tópico 10	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	no	0.029	no	0.011	no	0.039	no	0.035	no	0.032	no	0.032	pago	0.018	no	0.007	no	0.016	no	0.029
2	problema	0.016	alumno	0.006	duoc	0.013	duoc	0.013	duoc	0.026	ramo	0.022	duoc	0.017	semestre	0.004	duoc	0.009	duoc	0.019
3	tomar	0.014	problema	0.005	ano	0.010	nunca	0.012	matricular	0.015	alumno	0.020	ser	0.014	practicar	0.003	alumno	0.008	institucion	0.013
4	alumno	0.014	sede	0.005	sede	0.010	pagar	0.010	pagar	0.012	tomar	0.016	mes	0.014	clase	0.003	sala	0.008	calidad	0.010
5	sede	0.013	duoc	0.005	matricular	0.009	sede	0.010	uc	0.012	jmaeb	0.010	no	0.013	nada	0.002	clase	0.007	alumno	0.008
6	carrera	0.012	ramo	0.004	pagar	0.008	matricular	0.009	poner	0.012	sistema	0.009	pagar	0.013	perder	0.002	casino	0.006	profesor	0.008
7	dias	0.010	sala	0.004	nunca	0.004	semestre	0.008	sede	0.011	duoc	0.009	alumno	0.013	poner	0.002	sede	0.005	carrera	0.007
8	horario	0.010	tomar	0.004	ramo	0.008	carrera	0.008	dar	0.010	momento	0.008	uc	0.012	ramo	0.002	precio	0.004	hacer	0.007
9	ramo	0.008	sistema	0.004	clase	0.007	ramo	0.008	ramo	0.010	semestre	0.007	atrazar	0.012	pedir	0.002	carlos	0.004	suu	0.006
10	sicol	0.007	clase	0.003	institucion	0.007	alumno	0.007	horario	0.010	paghar	0.007	funcionario	0.011	pasar	0.002	peso	0.004	educacion	0.006
11	sistema	0.006	casino	0.003	alumno	0.007	decir	0.007	institucion	0.008	exigir	0.007	quincena	0.011	examen	0.002	carrera	0.004	pagar	0.006
12	decir	0.006	institucion	0.003	dar	0.007	ano	0.007	alumno	0.008	nada	0.007	ayudante	0.010	institucion	0.002	institucion	0.004	carlos	0.006
13	dia	0.006	carrera	0.003	poder	0.007	clase	0.006	estudiar	0.008	sede	0.007	fecha	0.008	horario	0.002	san	0.004	sede	0.006
14	retomar	0.005	san	0.003	carrera	0.006	institucion	0.006	semestre	0.008	estar	0.006	nunca	0.007	problema	0.002	mal	0.004	matricular	0.006
15	dar	0.005	nuevo	0.003	uc	0.006	poder	0.006	nada	0.007	hacer	0.006	nombre	0.007	notar	0.002	ano	0.003	peso	0.005
16	deber	0.005	escoger	0.003	estudiar	0.006	dinero	0.006	profesor	0.007	pedir	0.006	contratar	0.007	respuesta	0.002	costar	0.003	muestro	0.005
17	corresponder	0.005	informacion	0.002	estudio	0.006	nada	0.006	perder	0.007	inscribir	0.006	sallo	0.007	matricular	0.002	pasar	0.003	ano	0.005
18	semestre	0.005	solicitar	0.002	horario	0.006	sala	0.006	pagar	0.006	gran	0.006	poner	0.007	nunca	0.002	poder	0.003	exigir	0.005
19	cupo	0.005	carlos	0.002	algo	0.003	uc	0.006	problema	0.006	haber	0.006	febrero	0.007	profesor	0.002	bla	0.003	equipo	0.005
20	suceder	0.005	deberia	0.002	semestre	0.005	calidad	0.005	notar	0.006	injusticia	0.006	boleta	0.007	paginar	0.002	tener	0.003	pasar	0.005
21	web	0.005	mes	0.002	nada	0.005	pasar	0.005	poder	0.006	encuentro	0.006	abuso	0.006	tenia	0.002	burocracia	0.003	parte	0.005
22	jefe	0.005	pagar	0.002	sala	0.005	pedir	0.005	hablar	0.006	problema	0.005	vespucio	0.006	cambiar	0.002	hacer	0.003	uc	0.005
23	ano	0.005	tener	0.002	pedir	0.005	ocupar	0.005	carrera	0.006	poder	0.005	plaza	0.006	dar	0.002	suponer	0.003	estudiar	0.005
24	nivel	0.004	ano	0.002	respuesta	0.004	firme	0.005	algo	0.006	otorgar	0.005	salgar	0.006	sede	0.002	menos	0.003	tener	0.005
25	elegir	0.004	pasar	0.002	san	0.004	estudio	0.005	practicar	0.006	pasar	0.005	antonio	0.006	dato	0.002	acceso	0.003	plata	0.005
26	gran	0.004	semestre	0.002	pasar	0.004	retomar	0.005	cambiar	0.005	pagar	0.005	central	0.006	bajo	0.002	ramo	0.003	poder	0.005
27	tande	0.004	persona	0.002	tomar	0.004	devolucion	0.005	sistema	0.005	explicion	0.005	siempre	0.006	informacion	0.002	reclamar	0.003	dar	0.004
28	estudio	0.004	horario	0.002	casino	0.004	estudiar	0.005	escribir	0.005	maldito	0.005	ejemplo	0.006	preguntar	0.002	computador	0.003	mal	0.004
29	duoc	0.004	obligacion	0.002	preguntar	0.004	hablar	0.004	caer	0.005	solucion	0.005	dias	0.006	paso	0.002	mejor	0.003	profesional	0.004
30	aparecer	0.004	costar	0.002	pago	0.004	dar	0.004	conocer	0.005	postular	0.005	dedo	0.006	alto	0.002	varios	0.003	haber	0.004

Cuadro 5.11: Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2009.

Resultados Análisis LDA DOUC 2009 B

	Tópico 11		Tópico 12		Tópico 13		Tópico 14		Tópico 15		Tópico 16		Tópico 17		Tópico 18		Tópico 19		Tópico 20	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	no	0.048	no	0.022	no	0.019	no	0.007	duoc	0.023	duoc	0.020	no	0.039	no	0.054	no	0.009	no	0.029
2	duoc	0.024	pagar	0.017	ramo	0.011	alumno	0.005	calidad	0.020	matricular	0.015	alumno	0.026	problema	0.015	duoc	0.007	sede	0.021
3	pagar	0.014	ramo	0.016	pedir	0.010	sala	0.004	no	0.016	publicar	0.011	problema	0.015	sala	0.014	casino	0.005	alumno	0.019
4	profesor	0.012	institucion	0.015	pagar	0.008	duoc	0.004	estoi	0.016	querer	0.010	haber	0.013	carrera	0.014	sala	0.005	retomar	0.014
5	semestre	0.012	problema	0.013	semestre	0.008	casino	0.004	hacer	0.010	ser	0.010	profesor	0.012	alumno	0.012	carrera	0.005	estudio	0.012
6	nunca	0.011	estudiante	0.012	nada	0.008	carrera	0.004	mensualmente	0.010	institucion	0.009	nivel	0.012	nada	0.012	alumno	0.005	carrera	0.012
7	sede	0.011	carrera	0.011	pasar	0.008	precio	0.004	plata	0.010	institucion	0.009	duoc	0.011	clase	0.010	sede	0.004	estudiar	0.011
8	calidad	0.010	ano	0.009	revisar	0.008	san	0.003	dar	0.010	educacion	0.008	sede	0.010	ramo	0.009	clase	0.004	ano	0.010
9	alumno	0.010	programa	0.009	hacer	0.007	sede	0.003	compromiso	0.009	estudiar	0.008	ano	0.010	esperar	0.009	costar	0.004	decir	0.010
10	hablar	0.010	alumno	0.008	clase	0.007	costar	0.003	porquepago	0.009	poder	0.008	ramo	0.010	credo	0.008	precio	0.004	duoc	0.009
11	matricular	0.008	semestre	0.008	institucion	0.006	clase	0.002	bajos	0.009	carrera	0.008	carrera	0.010	dia	0.008	burocracia	0.003	cupo	0.009
12	ramo	0.008	deber	0.008	mente	0.006	bla	0.002	molesto	0.009	no	0.007	sistema	0.009	duoc	0.008	san	0.003	estar	0.008
13	dinero	0.008	mail	0.008	lamentable	0.006	peso	0.002	confort	0.009	mensualidad	0.007	estudiar	0.009	horario	0.008	acceso	0.003	ramo	0.008
14	carrera	0.007	duoc	0.007	alumno	0.006	burocracia	0.002	mi	0.009	hablar	0.007	tomar	0.008	legar	0.008	carlos	0.003	institucion	0.007
15	clase	0.006	clase	0.007	momento	0.005	acceso	0.002	semestre	0.009	encontrar	0.006	diseño	0.008	profesor	0.008	peso	0.003	querer	0.007
16	esperar	0.006	1	0.007	mandar	0.005	carlos	0.002	prueba	0.009	mal	0.006	proyecto	0.008	1	0.007	ano	0.002	dicha	0.007
17	algo	0.006	peso	0.007	mes	0.005	institucion	0.002	osca	0.009	burocracia	0.006	institucion	0.007	institucion	0.007	institucion	0.002	habia	0.007
18	tecnico	0.006	respuesta	0.007	duoc	0.005	gastar	0.002	descaro	0.009	dia	0.006	persona	0.007	tomar	0.007	bla	0.002	febrero	0.007
19	tomar	0.005	agosto	0.007	sala	0.005	tener	0.002	razon	0.009	varios	0.006	hacer	0.006	sede	0.007	mal	0.002	preocupar	0.007
20	ano	0.005	terminar	0.007	tomar	0.005	costo	0.002	profe	0.008	inscripcion	0.006	pagar	0.006	tener	0.006	tener	0.002	iva	0.007
21	respuesta	0.005	verdad	0.006	carrera	0.004	administrativo	0.002	desilusionado	0.008	pesima	0.006	dar	0.006	ano	0.006	administrativo	0.002	ciar	0.006
22	educacion	0.005	sentar	0.006	correo	0.004	bolsa	0.002	jabon	0.008	cargo	0.006	tener	0.006	pagar	0.006	basura	0.002	semestre	0.006
23	pedir	0.005	convalidar	0.006	parecer	0.004	ano	0.002	tampoco	0.008	capacidad	0.006	nada	0.006	querer	0.006	varios	0.002	antonio	0.006
24	poder	0.005	amarar	0.006	problema	0.004	varios	0.002	hablar	0.008	buscar	0.006	dos	0.005	momento	0.006	concesion	0.002	vara	0.005
25	mal	0.005	paso	0.005	estudiante	0.004	mal	0.002	faltar	0.008	semestre	0.006	cantidad	0.005	nuevo	0.006	costo	0.002	calidad	0.005
26	autotoma	0.005	tomar	0.005	servicio	0.004	vez	0.002	q	0.008	coordinacion	0.006	mes	0.005	dar	0.005	ojo	0.002	tecnico	0.004
27	nada	0.005	tener	0.005	sacar	0.004	dar	0.002	s	0.008	coordinador	0.006	titular	0.005	nuestro	0.005	hacer	0.002	clase	0.004
28	institucion	0.005	estudiar	0.005	horario	0.004	reclamar	0.002	importante	0.008	obstaculos	0.006	titulos	0.005	sistema	0.005	cara	0.002	hora	0.004
29	interesar	0.005	entrar	0.005	peso	0.004	hacer	0.002	uc	0.008	resultar	0.006	reclamo	0.005	madie	0.005	impresión	0.002	encontrar	0.004
30	estudiar	0.004	pago	0.004	ver	0.004	suponer	0.002	reclamo	0.008	respuesta	0.006	calidad	0.005	poder	0.005	bolsa	0.002	posibilidad	0.004

Cuadro 5.12: Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2009.

Resultados Análisis LDA DOUC 2010 A

	Tópico 1		Tópico 2		Tópico 3		Tópico 4		Tópico 5		Tópico 6		Tópico 7		Tópico 8		Tópico 9		Tópico 10	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	no	0.020	no	0.012	no	0.046	ramo	0.022	ramo	0.023	ramo	0.025	no	0.038	no	0.046	no	0.023	no	0.055
2	carrera	0.019	carrera	0.009	carrera	0.029	tomar	0.013	no	0.019	tomar	0.020	carrera	0.020	profesor	0.017	clase	0.020	duoc	0.018
3	institucion	0.014	alumno	0.007	ramo	0.018	pa	0.013	carrera	0.017	duoc	0.016	nuestro	0.012	profesor	0.015	profesor	0.013	hijo	0.011
4	objetivo	0.010	ir	0.006	duoc	0.016	no	0.012	tomar	0.017	ano	0.014	sede	0.012	semestre	0.014	duoc	0.012	alumno	0.011
5	esperar	0.010	momento	0.005	tomar	0.014	sistema	0.010	semestre	0.015	sistema	0.013	ramo	0.012	duoc	0.014	dia	0.010	ramo	0.010
6	hora	0.010	nuestro	0.005	semestre	0.011	poder	0.008	institucion	0.015	pa	0.013	duoc	0.012	dar	0.013	institucion	0.008	ano	0.010
7	oportunidad	0.009	linear	0.005	asignatura	0.010	dia	0.008	alumno	0.015	no	0.013	clase	0.012	sede	0.010	tener	0.007	institucion	0.010
8	duoc	0.009	estudio	0.004	alumno	0.009	carrera	0.006	problema	0.015	universidad	0.013	alumno	0.010	ano	0.010	grafo	0.007	tener	0.009
9	lograr	0.009	duoc	0.004	querer	0.009	solucionar	0.006	sistema	0.012	semestre	0.012	ano	0.010	exagerar	0.010	exagerar	0.007	dar	0.009
10	prestigio	0.008	apoyo	0.004	ano	0.008	problema	0.005	sede	0.010	querer	0.011	q	0.008	tomar	0.010	semana	0.007	pagar	0.009
11	ser	0.007	malla	0.004	profesor	0.008	tres	0.005	duoc	0.009	alumno	0.011	querer	0.008	ramo	0.010	valor	0.007	semestre	0.009
12	querer	0.007	curricular	0.004	dia	0.008	coordinador	0.005	falta	0.007	catolica	0.009	semestre	0.008	hablar	0.009	dar	0.007	matricular	0.009
13	posibilidad	0.007	duda	0.004	nuestro	0.007	1	0.005	subse	0.006	institucion	0.009	1	0.008	decir	0.008	pasar	0.006	carrera	0.008
14	acudir	0.006	haber	0.004	clase	0.007	hablar	0.005	calidad	0.006	mil	0.009	problema	0.007	nada	0.008	enseñar	0.006	tomar	0.008
15	sede	0.006	credo	0.004	ing	0.007	el	0.005	director	0.006	alguien	0.008	tener	0.007	apoderar	0.007	interpretar	0.005	mes	0.007
16	hacer	0.006	tutor	0.003	estudiante	0.006	querer	0.004	realizar	0.006	nada	0.007	tomar	0.007	pagar	0.006	estudiante	0.005	pago	0.007
17	informar	0.006	horario	0.003	dar	0.006	alguien	0.004	ano	0.006	ud	0.007	matricular	0.006	mal	0.006	mencionar	0.005	profesor	0.007
18	lamentable	0.005	posibilidad	0.003	poder	0.005	sede	0.004	reuco	0.005	relacion	0.007	horario	0.006	docente	0.006	director	0.005	padre	0.006
19	hoy	0.005	contener	0.003	1	0.005	clase	0.004	circuito	0.005	dia	0.006	tres	0.006	llamar	0.006	ramo	0.005	estudiar	0.006
20	persona	0.005	entregar	0.003	el	0.005	persona	0.004	poder	0.005	pagar	0.006	suponer	0.006	examen	0.005	seguir	0.005	perjudicar	0.005
21	recibir	0.005	materia	0.003	tener	0.005	inscribir	0.004	matricular	0.005	poder	0.006	poder	0.006	dia	0.005	hacer	0.005	ser	0.005
22	dia	0.005	tema	0.003	sede	0.005	duoc	0.004	cambiar	0.005	llamar	0.006	dar	0.006	notar	0.005	pensar	0.004	dinero	0.005
23	senoria	0.005	tecnico	0.003	verdad	0.005	cardo	0.004	hacer	0.005	durante	0.006	cupo	0.005	respuesta	0.005	dinero	0.004	titular	0.005
24	mejorar	0.005	institucion	0.003	entregar	0.005	atender	0.003	dia	0.005	sede	0.006	asignatura	0.005	querer	0.005	parocer	0.004	presentar	0.005
25	resultado	0.005	estudiar	0.003	dinero	0.005	institucion	0.003	preguntar	0.005	cobrar	0.006	institucion	0.005	entender	0.005	practicar	0.004	uc	0.005
26	despues	0.005	ver	0.003	ser	0.004	seccion	0.003	compromiso	0.005	estudio	0.005	dia	0.004	telefono	0.005	recibir	0.004	obligar	0.005
27	encargado	0.005	documento	0.003	pagar	0.004	hora	0.003	hora	0.005	realizar	0.005	lado	0.004	asignatura	0.005	ser	0.004	universidad	0.005
28	credo	0.005	asignatura	0.003	solucion	0.004	tirar	0.003	ingenieria	0.004	sicol	0.005	lado	0.004	ayudar	0.005	parte	0.004	poder	0.005
29	salud	0.005	parte	0.003	uc	0.004	queson	0.003	respuesta	0.004	ejemplo	0.005	calidad	0.004	hacer	0.004	cuenta	0.004	muestro	0.004
30	ocupar	0.005	modificar	0.003	matricular	0.004	acercar	0.003	publicidad	0.004	cobranza	0.005	deber	0.004	secretario	0.004	reclamo	0.004	nada	0.004

Cuadro 5.13: Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2010.

Resultados Análisis LDA DOUC 2010 B

	Tópico 11		Tópico 12		Tópico 13		Tópico 14		Tópico 15		Tópico 16		Tópico 17		Tópico 18		Tópico 19		Tópico 20	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	problema	0.005	no	0.047	no	0.035	ano	0.022	no	0.036	no	0.041	no	0.051	no	0.023	no	0.026	no	0.001
2	sede	0.005	carrera	0.017	alumno	0.016	pagar	0.014	alumno	0.016	ramo	0.026	carrera	0.018	él	0.015	semestre	0.016	carrera	0.001
3	carrera	0.004	ramo	0.016	ramo	0.015	diuoc	0.014	decir	0.012	tomar	0.021	trabajar	0.015	curso	0.011	universidad	0.013	diuoc	0.001
4	rango	0.004	diuoc	0.015	diuoc	0.014	esperar	0.014	diuoc	0.012	sistema	0.013	ramo	0.013	dar	0.009	ramo	0.012	ramo	0.001
5	semestre	0.004	tomar	0.009	sede	0.014	no	0.013	pagar	0.011	diuoc	0.011	nada	0.012	decir	0.009	profesor	0.012	sede	0.001
6	circuito	0.004	profesor	0.008	carrera	0.011	decir	0.010	institucion	0.011	bueno	0.010	decir	0.012	diuoc	0.009	alumno	0.012	tomar	0.001
7	subse	0.004	pagar	0.008	problema	0.011	preguntar	0.009	ano	0.010	profesor	0.010	estudiar	0.012	apoderar	0.008	estudio	0.009	institucion	0.001
8	proyecto	0.003	sede	0.007	tomar	0.009	devolucion	0.008	tener	0.009	ser	0.008	tomar	0.011	llamar	0.006	diuoc	0.009	querer	0.001
9	falta	0.003	alumno	0.006	institucion	0.009	problema	0.008	esperar	0.008	momento	0.008	diuoc	0.010	pagar	0.006	catolica	0.009	semestre	0.001
10	director	0.003	uc	0.006	matricular	0.008	nada	0.007	dar	0.008	institucion	0.008	mensaje	0.010	nada	0.006	ano	0.008	problema	0.001
11	escuela	0.003	horario	0.006	ingresar	0.007	respuesta	0.007	nada	0.008	alumno	0.007	deber	0.009	notar	0.006	linear	0.008	alumno	0.001
12	prueba	0.003	curso	0.006	regular	0.007	tener	0.007	hacer	0.007	tener	0.007	problema	0.008	tener	0.006	estudiar	0.008	tener	0.001
13	momento	0.003	sistema	0.006	semestre	0.007	seguir	0.007	nuevo	0.007	calidad	0.007	tener	0.008	carrera	0.006	ser	0.008	ano	0.001
14	tecnico	0.003	semestre	0.006	tener	0.007	efectivo	0.007	preguntar	0.007	dar	0.006	hacer	0.007	uc	0.005	tomar	0.007	dia	0.001
15	ingenieria	0.003	entrar	0.006	sistema	0.006	devolucion	0.007	profesor	0.006	cuenta	0.006	estar	0.006	hacer	0.005	mes	0.007	sistema	0.001
16	no	0.003	problema	0.005	ser	0.006	estudio	0.007	clase	0.006	uc	0.006	ano	0.006	telefono	0.005	trabajo	0.007	clase	0.001
17	realizar	0.003	siempre	0.005	haber	0.006	universidad	0.007	carrera	0.006	servicio	0.006	sistema	0.006	sede	0.005	carrera	0.007	pagar	0.001
18	electricidad	0.003	pasar	0.005	algo	0.006	beca	0.006	credo	0.006	poder	0.006	menos	0.005	hablar	0.004	trabajar	0.006	dar	0.001
19	respuesta	0.003	bueno	0.005	validar	0.006	dar	0.006	estacionamiento	0.006	pagar	0.005	dar	0.005	reglamento	0.004	titular	0.006	profesor	0.001
20	electro	0.002	matricular	0.005	dar	0.005	estudiar	0.006	problema	0.005	administracion	0.005	claro	0.005	examen	0.004	decir	0.006	pasar	0.001
21	laboratorio	0.002	nada	0.005	horario	0.005	ser	0.006	dia	0.005	hacer	0.005	jefe	0.005	de	0.004	tener	0.006	decir	0.001
22	ramo	0.002	servicio	0.005	ano	0.005	contar	0.005	seguir	0.005	deber	0.005	ser	0.005	devolver	0.004	él	0.006	nada	0.001
23	enlace	0.002	resultar	0.005	pago	0.005	pedir	0.005	admission	0.005	carrera	0.005	munez	0.005	clase	0.004	hacer	0.005	respuesta	0.001
24	electronica	0.002	decir	0.005	realizar	0.005	algnien	0.004	lugar	0.005	estudio	0.005	despues	0.005	dia	0.004	vespertino	0.005	matricular	0.001
25	victor	0.002	clase	0.004	_	0.005	institucion	0.004	reclamo	0.005	paginar	0.005	tema	0.005	reclamo	0.004	educacion	0.005	hacer	0.001
26	senor	0.002	ser	0.004	seleccionar	0.004	llegada	0.004	algnien	0.004	sede	0.005	dejar	0.005	homecenter	0.004	calidad	0.005	ser	0.001
27	instrumento	0.002	demandada	0.004	paginar	0.004	catolica	0.004	beca	0.004	problema	0.005	él	0.005	valor	0.004	institucion	0.005	uc	0.001
28	presente	0.002	construccion	0.004	necesitar	0.004	milenio	0.004	pedir	0.004	ano	0.005	universidad	0.004	sodimac	0.004	totalmente	0.004	estudiar	0.001
29	estudiar	0.002	dar	0.004	favor	0.004	utilizar	0.004	edificio	0.004	dejar	0.004	hora	0.004	admission	0.004	fecha	0.004	poder	0.001
30	amarrar	0.002	haber	0.004	nada	0.004	plata	0.004	él	0.004	matricular	0.004	otro	0.004	cero	0.004	notar	0.004	gracia	0.001

Cuadro 5.14: Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2010.

Resultados Análisis LDA DOUC 2011 A

	Tópico 1		Tópico 2		Tópico 3		Tópico 4		Tópico 5		Tópico 6		Tópico 7		Tópico 8		Tópico 9		Tópico 10	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	no	0.038	semestre	0.009	no	0.029	no	0.041	debia	0.017	no	0.026	no	0.013	ano	0.016	no	0.014	no	0.014
2	semestre	0.022	problema	0.007	dato	0.019	dato	0.016	deuda	0.018	sede	0.014	materia	0.014	pagar	0.009	necesitar	0.016	notar	0.011
3	sede	0.013	no	0.007	matricular	0.010	comentar	0.012	dtuoc	0.018	solicitud	0.014	semestre	0.013	casa	0.008	no	0.014	semestre	0.011
4	dtuoc	0.013	tomar	0.007	pasar	0.010	dtuoc	0.011	pagar	0.013	no	0.014	clase	0.013	demorar	0.008	semestre	0.013	examen	0.010
5	ramo	0.012	haber	0.006	problema	0.009	solicitud	0.011	semestre	0.010	decir	0.012	profesor	0.013	central	0.008	estudiar	0.011	ano	0.007
6	problema	0.011	ramo	0.006	dtuoc	0.009	dia	0.010	gracia	0.008	ingenieria	0.012	ramo	0.012	valer	0.008	alumno	0.010	gente	0.007
7	carrera	0.010	ano	0.005	ramo	0.009	carrera	0.010	credo	0.008	preguntar	0.010	mes	0.012	vista	0.006	dia	0.010	ramo	0.007
8	matricular	0.009	dia	0.004	tomar	0.008	construccion	0.009	alumno	0.007	nada	0.009	tener	0.012	ano	0.006	gracia	0.010	derecho	0.007
9	ano	0.009	dtuoc	0.004	pagar	0.008	gestion	0.008	deber	0.007	joaquin	0.009	estudio	0.012	dtuoc	0.006	pagar	0.009	entrar	0.007
10	tomar	0.009	1	0.003	podia	0.008	sede	0.008	dar	0.007	dia	0.009	dos	0.011	tramitar	0.006	retirar	0.009	saber	0.007
11	tenia	0.008	nunca	0.003	osca	0.008	matricular	0.008	examen	0.007	pagar	0.009	entregar	0.010	cobranza	0.006	hoy	0.008	publicar	0.007
12	dia	0.007	institucion	0.003	proyecto	0.008	llamar	0.008	pago	0.006	aprobar	0.009	sede	0.010	salir	0.006	nunca	0.008	pasar	0.007
13	pasar	0.007	llamar	0.003	clase	0.008	enero	0.007	ano	0.006	escuela	0.009	carrera	0.009	pasar	0.005	documento	0.008	matricular	0.005
14	clase	0.007	carrera	0.003	dia	0.007	ano	0.007	esperar	0.006	contrato	0.009	cuestionar	0.008	diciembre	0.005	informacion	0.008	estudiar	0.005
15	titular	0.007	estudio	0.003	titular	0.007	problema	0.007	haber	0.006	matricular	0.008	enseñar	0.008	quedar	0.005	entregar	0.008	problema	0.005
16	dar	0.007	tenia	0.003	respuesta	0.007	ramo	0.007	comision	0.006	san	0.008	alumno	0.008	hoy	0.005	enfermeria	0.008	reprobe	0.005
17	pagar	0.006	solucionar	0.003	dar	0.006	indicar	0.006	dtuocuc	0.006	sistema	0.008	titular	0.008	querer	0.005	urgencia	0.008	solucionar	0.005
18	esperar	0.006	paghar	0.003	contabilidad	0.006	semana	0.006	cobrar	0.006	podia	0.008	semana	0.007	problema	0.005	curso	0.008	sede	0.004
19	proyecto	0.006	pagar	0.003	alumno	0.006	preocupar	0.006	solvenacia	0.006	clase	0.007	dtuoc	0.007	cobrar	0.004	entre	0.008	querer	0.004
20	estudiar	0.005	proyecto	0.003	sede	0.006	febrero	0.006	cursar	0.006	llegar	0.007	mitad	0.007	institucion	0.004	auxiliar	0.008	gracia	0.004
21	hacer	0.005	podia	0.003	nada	0.006	entendi	0.006	llamar	0.006	hoy	0.006	tomar	0.007	resultar	0.004	certificar	0.008	1	0.004
22	1	0.005	monit	0.003	carrera	0.005	tubieron	0.006	estudio	0.005	vespertino	0.006	pauta	0.007	tener	0.004	que	0.008	finalizar	0.004
23	hablar	0.005	deber	0.003	horario	0.005	preguntar	0.006	profesor	0.005	horario	0.006	examen	0.007	dinero	0.004	regular	0.008	prometer	0.004
24	algo	0.005	deuda	0.003	quedar	0.005	senora	0.006	profesor	0.005	dar	0.006	prueba	0.006	dia	0.004	saldo	0.008	parecer	0.004
25	podia	0.004	tener	0.003	ingenieria	0.004	inacap	0.006	cancelar	0.005	problema	0.006	haber	0.006	retanal	0.004	dinero	0.005	prueba	0.004
26	vespucio	0.004	puerto	0.003	estudiar	0.004	hablar	0.006	problema	0.005	vara	0.005	despues	0.006	aprobar	0.004	pagar	0.005	facil	0.004
27	falla	0.004	inscribir	0.003	haber	0.004	pasar	0.006	tenia	0.005	alumno	0.005	deuda	0.006	verdad	0.004	instituto	0.005	riesgo	0.004
28	estar	0.004	matricular	0.002	dias	0.004	cupo	0.006	tener	0.005	seguir	0.005	bueno	0.006	bei	0.004	plata	0.005	excelente	0.004
29	momento	0.004	titular	0.002	egresar	0.004	asegurar	0.006	total	0.005	antonio	0.005	cupo	0.005	ayer	0.004	cancelar	0.004	igual	0.004
30	nada	0.004	pasar	0.002	hoy	0.004	liliana	0.006	1	0.005	estar	0.005	horario	0.005	postular	0.003	hacer	0.004	menor	0.004

Cuadro 5.15: Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2011.

Resultados Análisis LDA DOUC 2011 B

	Tópico 11		Tópico 12		Tópico 13		Tópico 14		Tópico 15		Tópico 16		Tópico 17		Tópico 18		Tópico 19		Tópico 20	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	no	0.062	no	0.023	no	0.043	no	0.023	no	0.030	no	0.034	no	0.031	no	0.034	no	0.010	no	0.050
2	duoc	0.015	ramo	0.014	semestre	0.018	q	0.022	matricular	0.019	carrera	0.009	carrera	0.023	querer	0.019	dar	0.010	ramo	0.018
3	dar	0.013	dar	0.010	duoc	0.015	dar	0.016	pasar	0.019	dar	0.009	querer	0.021	duoc	0.017	q	0.010	clase	0.017
4	problema	0.012	clase	0.009	carrera	0.014	sede	0.011	ramo	0.017	problema	0.009	matricular	0.017	pagar	0.012	alumno	0.006	duoc	0.017
5	alumno	0.011	tomar	0.009	tomar	0.011	duoc	0.010	semestre	0.013	alumno	0.008	cambiar	0.017	ano	0.010	pagar	0.006	problema	0.013
6	hacer	0.009	esperar	0.009	problema	0.011	estudiar	0.010	despues	0.013	duoc	0.008	encuentro	0.016	dar	0.010	estudiar	0.006	semestre	0.010
7	fecha	0.008	solucion	0.008	dar	0.010	alumno	0.009	papel	0.009	dia	0.007	estudio	0.015	demorar	0.009	docente	0.006	alumno	0.010
8	nada	0.007	problema	0.008	ramo	0.009	clase	0.009	dias	0.009	solicitud	0.007	dinero	0.012	central	0.009	semestre	0.005	tomar	0.009
9	llamar	0.007	duoc	0.008	dia	0.009	insti-tucion	0.008	contrato	0.009	llamar	0.007	pagar	0.011	central	0.009	duoc	0.005	dar	0.009
10	estudiar	0.007	profesor	0.007	alumno	0.008	dejar	0.008	enero	0.009	ano	0.007	tec	0.011	hora	0.007	universidad	0.005	dia	0.008
11	pasar	0.006	haber	0.007	insti-tucion	0.008	vender	0.008	duoc	0.009	gracia	0.007	politica	0.011	semestre	0.007	vender	0.004	carrera	0.008
12	tenia	0.006	querer	0.007	estudiar	0.007	querer	0.008	examen	0.007	ramo	0.006	proximo	0.010	tramitar	0.007	mal	0.004	insti-tucion	0.008
13	cambiar	0.006	horario	0.007	deber	0.006	universidad	0.007	dia	0.007	querer	0.006	injusto	0.010	vista	0.007	catolica	0.004	matricular	0.007
14	pagar	0.006	deber	0.007	matricular	0.006	uc	0.005	notar	0.007	dejar	0.006	carrera	0.010	hoy	0.006	insti-tucion	0.004	ano	0.007
15	ramo	0.006	hacer	0.006	ano	0.006	bueno	0.005	tubo	0.007	insti-tucion	0.006	congelar	0.009	estudiar	0.006	carrera	0.004	ntunca	0.006
16	ano	0.006	tener	0.006	haber	0.006	mal	0.005	semestral	0.007	informacion	0.006	nada	0.008	cobranza	0.006	bonito	0.003	nada	0.006
17	carrera	0.006	respuesta	0.006	horario	0.006	nuevo	0.005	firmar	0.006	tomar	0.006	algo	0.008	tener	0.006	decir	0.003	sede	0.005
18	tener	0.005	respuesta	0.006	red	0.006	lucro	0.005	final	0.006	proceso	0.005	insti-tuto	0.008	diciembre	0.006	despues	0.003	esperar	0.005
19	trabajo	0.005	telefono	0.006	querer	0.006	haber	0.005	renunciar	0.006	tramitar	0.005	cancelar	0.008	entrar	0.006	profesor	0.003	asignatura	0.005
20	respuesta	0.005	comunicar	0.006	enero	0.005	enseñar	0.005	esperar	0.006	debia	0.005	duoc	0.007	problema	0.006	retirar	0.003	llamar	0.005
21	decir	0.005	fundacion	0.005	inscribir	0.005	ano	0.005	profesor	0.006	realizar	0.005	plata	0.007	lado	0.006	pasar	0.003	hacer	0.005
22	denda	0.005	parecer	0.005	empresa	0.005	terminar	0.004	deber	0.006	horario	0.005	tema	0.006	insti-tuto	0.006	titular	0.003	querer	0.005
23	momento	0.005	alumno	0.005	estar	0.005	tener	0.004	carrera	0.006	asignatura	0.005	ingenieria	0.006	valer	0.006	ano	0.003	profesor	0.005
24	sede	0.005	plaza	0.005	s	0.005	educacion	0.004	el	0.006	deber	0.005	insti-tucion	0.006	horario	0.006	clase	0.003	1	0.005
25	uc	0.005	momento	0.004	paghar	0.005	parecer	0.004	dinero	0.006	pedir	0.005	comunicacion	0.006	insti-tucion	0.005	el	0.003	pagar	0.005
26	clase	0.005	compartar	0.004	pagar	0.005	entrar	0.004	estudiar	0.006	matricular	0.005	unica	0.006	credo	0.005	ramo	0.003	haber	0.005
27	sistema	0.005	compañeros	0.004	esperar	0.004	carrera	0.004	alumno	0.006	reclamo	0.004	gustar	0.006	suponer	0.005	algo	0.003	pasar	0.005
28	pago	0.005	decir	0.004	nada	0.004	falso	0.004	gente	0.005	tener	0.004	estudiante	0.006	verdad	0.005	bueno	0.003	sistema	0.005
29	recibir	0.004	menos	0.004	tener	0.004	insti-tuto	0.004	derecho	0.005	pago	0.004	obtener	0.006	haber	0.005	solucion	0.003	deber	0.005
30	mal	0.004	llamar	0.004	llevar	0.004	solucion	0.004	trabajo	0.005	seguir	0.004	beneficio	0.006	obrero	0.005	pesima	0.003	examen	0.005

Cuadro 5.16: Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC durante el año 2011.

Resultados Análisis LDA DOUC 2007-2011 A

	Tópico 1		Tópico 2		Tópico 3		Tópico 4		Tópico 5		Tópico 6		Tópico 7		Tópico 8		Tópico 9		Tópico 10	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P								
1	no	0.041	no	0.021	no	0.051	no	0.040	no	0.046	ingeniería	0.001	no	0.021	no	0.038	no	0.020	no	0.049
2	duoc	0.016	duoc	0.015	institucion	0.014	duoc	0.016	duoc	0.018	debia	0.001	class	0.011	alumno	0.018	ramo	0.013	ramo	0.014
3	ramo	0.016	semestre	0.013	carrera	0.011	ramo	0.014	semestre	0.011	solicitud	0.001	duoc	0.009	semestre	0.013	ano	0.011	carrera	0.014
4	semestre	0.016	carrera	0.012	ramo	0.011	carrera	0.011	ano	0.011	preguntar	0.001	sap	0.009	duoc	0.012	duoc	0.011	profesor	0.012
5	alumno	0.014	sede	0.010	problema	0.011	semestre	0.009	nada	0.010	sede	0.001	semestre	0.008	class	0.011	esperar	0.010	duoc	0.012
6	ano	0.011	problema	0.009	ano	0.010	alumno	0.009	class	0.009	escuela	0.001	alumno	0.007	institucion	0.008	semestre	0.010	tomar	0.011
7	tomar	0.010	pasar	0.007	duoc	0.009	ano	0.008	dar	0.009	dia	0.001	nada	0.006	ano	0.008	tomar	0.009	alumno	0.010
8	carrera	0.009	ser	0.007	estudiar	0.009	dia	0.008	alumno	0.009	sistema	0.001	pasar	0.006	carrera	0.008	pagar	0.009	institucion	0.009
9	dar	0.008	class	0.006	bueno	0.008	sede	0.008	carrera	0.008	podia	0.001	tomar	0.005	dar	0.008	problema	0.007	ano	0.009
10	pagar	0.008	publicidad	0.006	semestre	0.007	tomar	0.007	profesor	0.008	san	0.001	estudiar	0.005	ramo	0.007	alumno	0.007	sede	0.008
11	sede	0.007	ramo	0.006	tomar	0.007	matricular	0.007	sala	0.007	nada	0.001	dia	0.005	nada	0.007	institucion	0.006	class	0.008
12	titular	0.007	calidad	0.005	nada	0.007	dar	0.007	hacer	0.007	no	0.001	querer	0.004	sede	0.007	dia	0.006	problema	0.007
13	querer	0.006	tener	0.005	pasar	0.007	problema	0.007	tener	0.007	pagar	0.001	hacer	0.004	pagar	0.007	estudiar	0.006	estudiar	0.007
14	institucion	0.006	dar	0.005	pagar	0.006	tener	0.006	q	0.007	decir	0.001	ramo	0.004	tener	0.007	pasar	0.005	semestre	0.007
15	problema	0.006	alumno	0.005	dar	0.006	pagar	0.006	pagar	0.007	joaquin	0.001	problema	0.004	profesor	0.006	querer	0.005	querer	0.007
16	estudio	0.005	dia	0.005	reclamo	0.006	profesor	0.006	institucion	0.007	contrato	0.001	carrera	0.004	tomar	0.006	tener	0.005	dar	0.006
17	tener	0.005	hablar	0.005	hacer	0.006	hacer	0.006	sede	0.006	aprobar	0.001	sistema	0.004	deuda	0.006	decir	0.005	pasar	0.006
18	universidad	0.005	estudiar	0.005	deber	0.006	institucion	0.006	ramo	0.006	electricidad	0.001	regular	0.004	deber	0.005	beca	0.005	tener	0.006
19	profesor	0.005	pagar	0.004	decir	0.006	estudiar	0.005	problema	0.005	nuevamente	0.001	saldo	0.004	uc	0.005	instituto	0.005	sistema	0.005
20	matricular	0.005	alguien	0.004	tener	0.006	class	0.005	horario	0.005	seguir	0.001	que	0.004	exámenes	0.005	respuesta	0.005	seguir	0.005
21	llamar	0.005	egresar	0.004	alumno	0.006	pasar	0.004	decir	0.005	algo	0.001	institucion	0.004	pasar	0.005	credo	0.005	realizar	0.005
22	poder	0.004	preguntar	0.004	momento	0.005	examen	0.004	estudio	0.005	matricular	0.001	matricular	0.004	problema	0.005	hacer	0.005	dia	0.004
23	respuesta	0.004	hacer	0.004	ser	0.005	hablar	0.004	hora	0.005	consultar	0.001	osea	0.004	respuesta	0.005	dar	0.005	poder	0.004
24	decir	0.004	compromiso	0.004	uc	0.005	querer	0.004	esperar	0.004	insistir	0.001	pagar	0.003	mestreo	0.005	tenia	0.004	entrar	0.004
25	pasar	0.004	esperar	0.004	matricular	0.005	dos	0.004	uc	0.004	mes.diciembre	0.001	academia	0.003	malo	0.005	sistema	0.004	hora	0.004
26	class	0.004	poder	0.004	profesor	0.004	estudio	0.004	munca	0.004	debi	0.001	mensualidad	0.003	hacer	0.004	carrera	0.004	nada	0.004
27	horario	0.004	persona	0.003	munca	0.004	dinero	0.004	jefe	0.004	encontro	0.001	profesor	0.003	estudiar	0.004	solucion	0.004	ver	0.004
28	sistema	0.004	vez	0.003	sede	0.004	ser	0.004	tomar	0.004	1	0.001	deber	0.003	dia	0.004	cancelar	0.004	calidad	0.004
29	nada	0.004	informacion	0.003	respuesta	0.004	nada	0.004	1	0.004	ano	0.001	trabajo	0.003	mes	0.004	poder	0.004	pedir	0.004
30	dia	0.004	agradeceria	0.003	credo	0.004	solicitud	0.004	tenia	0.004	dirigi	0.001	calidad	0.003	poder	0.004	mes	0.004	pagar	0.004

Cuadro 5.17: Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC entre los años 2007 y 2011.

Resultados Análisis LDA DOUC 2007-2011 B

	Tópico 11		Tópico 12		Tópico 13		Tópico 14		Tópico 15		Tópico 16		Tópico 17		Tópico 18		Tópico 19		Tópico 20	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	no	0.035	no	0.054	no	0.046	institucion	0.007	no	0.029	problema	0.007	no	0.035	no	0.018	no	0.017	no	0.026
2	tomar	0.015	carrera	0.018	dioc	0.017	no	0.006	carrera	0.020	ramo	0.006	ramo	0.031	dioc	0.018	sede	0.014	problema	0.011
3	problema	0.015	dioc	0.014	profesor	0.012	ramo	0.005	dioc	0.016	no	0.005	dioc	0.017	carrera	0.010	dioc	0.009	clase	0.011
4	ramo	0.014	profesor	0.011	alumno	0.011	tomar	0.005	matricular	0.011	tomar	0.005	clase	0.017	examen	0.009	dioc	0.007	carrera	0.011
5	dioc	0.012	alumno	0.009	carrera	0.009	sistema	0.005	pasar	0.009	alumno	0.003	carrera	0.015	clase	0.009	pagar	0.006	dioc	0.010
6	sede	0.012	ramo	0.009	pagar	0.009	alumno	0.005	semestre	0.008	sede	0.003	tomar	0.014	pagar	0.009	nada	0.006	ramo	0.010
7	dia	0.009	ano	0.008	sede	0.009	clase	0.005	contrato	0.007	realizar	0.003	matricular	0.013	ramo	0.008	san	0.006	alumno	0.008
8	carrera	0.009	haber	0.007	calidad	0.008	dioc	0.004	esperar	0.006	institucion	0.003	dia	0.011	matricular	0.008	carrera	0.006	dia	0.007
9	semestre	0.008	hacer	0.007	semestre	0.007	pasar	0.004	alumno	0.006	pedir	0.003	problema	0.009	alumno	0.007	matricular	0.006	pagar	0.007
10	matricular	0.008	nada	0.007	ramo	0.007	hora	0.003	problema	0.006	semestre	0.002	sistema	0.002	semestre	0.007	ano	0.005	tomar	0.006
11	decir	0.008	problema	0.007	dar	0.007	menos	0.003	enero	0.006	dar	0.002	alumno	0.008	institucion	0.007	estudiar	0.005	ano	0.006
12	ano	0.008	dar	0.007	clase	0.007	matricular	0.003	papel	0.006	carrera	0.002	horario	0.007	notar	0.006	decir	0.005	sede	0.006
13	hablar	0.007	sede	0.007	matricular	0.007	mayor	0.003	practicar	0.005	sistema	0.002	pa	0.006	haber	0.006	dar	0.004	sistema	0.006
14	dar	0.007	pagar	0.006	hacer	0.007	final	0.003	después	0.005	matricular	0.002	semestre	0.006	sede	0.005	pena	0.004	tener	0.005
15	pagar	0.006	institucion	0.006	nada	0.006	linear	0.003	ing	0.005	seguir	0.002	institucion	0.006	profesor	0.005	carlos	0.004	institucion	0.005
16	sistema	0.006	estudiar	0.006	esperar	0.006	cobranza	0.003	sede	0.005	pago	0.002	tener	0.006	hoy	0.005	publicidad	0.004	estudiar	0.005
17	alumno	0.006	semestre	0.006	querer	0.006	incomprensible	0.003	dar	0.005	momento	0.002	ano	0.006	uc	0.004	ingenieria	0.004	haber	0.005
18	haber	0.006	querer	0.006	tener	0.005	off_line	0.003	ramo	0.005	ingresar	0.002	1	0.006	semana	0.004	semestre	0.004	momento	0.005
19	poder	0.006	nuestro	0.005	querer	0.005	inconveniente	0.003	llamar	0.005	tener	0.002	poder	0.005	encuentro	0.004	vespucio	0.004	profesor	0.005
20	nada	0.005	tener	0.005	educacion	0.005	fin	0.003	dia	0.004	servicio	0.002	algo	0.005	valer	0.004	valer	0.004	semestre	0.004
21	institucion	0.005	tomar	0.005	nuestro	0.005	funcionar	0.003	ir	0.004	diocuc	0.002	dar	0.005	pasar	0.004	clase	0.004	quedar	0.004
22	1	0.005	cambiar	0.005	uc	0.004	superior	0.003	decir	0.004	respuesta	0.002	semana	0.004	feliz	0.004	contrato	0.004	estudiante	0.004
23	profesor	0.005	uc	0.004	estudiar	0.004	dar_cuenta	0.003	dinero	0.004	director	0.002	pagar	0.004	estudiar	0.004	dinero	0.004	perder	0.004
24	querer	0.005	ser	0.004	mes	0.004	educacion	0.003	tener	0.004	subside	0.002	nunca	0.004	insol	0.004	debia	0.004	hacer	0.004
25	solucionar	0.004	calidad	0.004	pasar	0.004	tener	0.003	ano	0.004	calidad	0.002	querer	0.004	estudio	0.004	alumno	0.004	solucionar	0.004
26	ingenieria	0.004	decir	0.004	dia	0.004	nunca	0.003	poder	0.004	molestia	0.002	esperar	0.004	credo	0.004	hacer	0.003	llamar	0.004
27	llamar	0.004	respuesta	0.004	horario	0.004	querer	0.003	tomar	0.004	ano	0.002	ser	0.004	uc	0.004	uc	0.003	programa	0.004
28	llegar	0.004	mes	0.004	hablar	0.004	calidad	0.003	nada	0.004	laboratorio	0.001	hablar	0.004	favor	0.004	algo	0.003	persona	0.004
29	tener	0.004	estudio	0.004	hora	0.004	cancelar	0.002	querer	0.004	reco	0.001	sede	0.004	alto	0.004	profesor	0.003	vez	0.004
30	terminar	0.004	matricular	0.004	entregar	0.004	compromiso	0.002	respuesta	0.004	falta	0.001	estudiar	0.004	dar	0.004	nuestro	0.003	estudio	0.004

Cuadro 5.18: Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra el Departamento Universitario Obrero Campesino DOUC entre los años 2007 y 2011.

Realizando un análisis, basado en la revisión y en la lectura manual de los tópicos y de las palabras generadas en cada período, la o las frases que pueden generarse y sintetizar el sentido que subyace en estos elementos son las siguientes:

■ **Período 2007**

“Existen problemas con respecto a la institución, sus carreras y sus ramos por estafas en la publicidad y en la infraestructura tecnológica disponible.”

“Hay quejas respecto a que los cobros abusivos tanto en matrículas como en aranceles mensuales y semestrales.”

“Se dan problemas con los directores y profesores.”

■ **Período 2008**

“Existen problemas con respecto a los profesores, los exámenes, las prácticas y el trabajar.”

■ **Período 2009**

“Existen problemas con respecto las sedes, la burocracia de los trámites que haya que realizar y los precios mensuales que se deben pagar.”

“También hay problemas con los ramos, los profesores y las clases.”

■ **Período 2010**

“Existen problemas con respecto a la institución, la universidad católica, los ramos, los profesores y las matrículas.”

■ **Período 2011**

“Existen problemas con las carreras de contabilidad e ingeniería, así como también con la matrícula, el congelamiento y la titulación, y el pago de las deudas.”

■ **Período 2007-2011**

“Existen problemas con la matrícula, las carreras, con la toma de ramos y sus horarios y con aspectos de la calidad.”

5.2. Rubro de Gobierno

Al igual que en el rubro de Educación, también en este rubro se detectaron sub rubros, pero 7 en vez de 4, siendo estos: Embajada, Gobierno, Justicia, Ministerio, Municipalidad, Partidos Políticos y Superintendencia, los que poseen, respectivamente, las siguientes cantidades totales de instituciones, reclamos y recomendaciones:

Sub Rubro	Instituciones		Reclamos		Recomendaciones		Total	
Embajada	1	0,6 %	0	0 %	0	0 %	0	0 %
Gobierno	1	0,6 %	3	0,1 %	0	0 %	3	0,1 %
Justicia	6	3,5 %	571	11 %	126	30,4 %	697	12,5 %
Ministerio	13	7,6 %	475	9,2 %	35	8,5 %	510	9,1 %
Municipalidad	95	55,6 %	1.129	21,8 %	87	21 %	1.216	21,7 %
Partidos Políticos	4	2,3 %	6	0,1 %	0	0 %	6	0,1 %
Superintendencia	4	2,3 %	98	1,9 %	9	2,2 %	107	1,9 %
Ninguno	47	27,5 %	2.901	55,9 %	157	37,9 %	3.058	54,6 %
Totales	171	100 %	5.183	100 %	414	100 %	5.597	100 %

Cuadro 5.19: Cantidades totales de Instituciones, Reclamos y Recomendaciones existentes y realizados dentro de cada Sub rubro de Gobierno.

Como se puede apreciar, se detectaron 47 instituciones que no pertenecían a ninguno de los sub rubros de Gobierno, tales como: FOSIS, COMPIN, Armada de Chile, entre otras, que si bien podrían clasificarse dentro de uno los sub rubros, en el sitio no aparecen así.

Las razones de esta, por así decirlo, falta de clasificación por sub rubros de algunas instituciones, pueden ser variadas, desde criterios de los administradores del sitio, hasta errores de ellos, pero para esta memoria no es relevante, ni de interés, conocer las causas de esto.

5.2.1. Las 10 Instituciones con más Reclamos y Recomendaciones

Realizando el mismo ejercicio de filtrar las 10 instituciones que poseían más reclamos, pero esta vez dentro del rubro de Gobierno, se obtuvieron las siguientes instituciones: Carabineros de Chile (Carabineros), Comisión Médica, Preventiva e Invalidez (COMPIN), Junta Nacional de Auxilio Escolar y Becas (JUNAEB), Juzgado de Familia (J-Familia), Servicio de Vivienda y Urbanización (SERVIU), Ministerio de Educación (MINEDUC), Juzgado de Policía Local (J-Policía), Municipalidad de Quilicura (M-Quilicura), Fondo Nacional de Salud (FONASA) y la Municipalidad de Maipú (M-Maipú).

Institución	Reclamos		Recomendaciones		Total	
Carabineros	1.008	94,9 %	54	5,1 %	1.062	100 %
COMPIN	532	94,3 %	32	5,7 %	564	100 %
JUNAEB	422	96,8 %	14	3,2 %	436	100 %
J-Familia	385	78,1 %	108	21,9 %	493	100 %
SERVIU	222	94,9 %	12	5,1 %	234	100 %
MINEDUC	211	95,4 %	10	4,5 %	221	100 %
J-Policía	126	92,6 %	10	7,4 %	136	100 %
M-Quilicura	103	91,2 %	10	8,8 %	113	100 %
FONASA	96	94,1 %	6	5,9 %	102	100 %
M-Maipú	85	96,6 %	3	3,4 %	88	100 %

Cuadro 5.20: Las 10 Instituciones con mayor cantidad de Reclamos y Recomendaciones dentro del rubro de Gobierno.

Detalle de los Reclamos y de las Recomendaciones

Profundizando en los reclamos y en las recomendaciones realizados hacia estas 10 instituciones, también fue posible detectar cuántos comentarios recibieron al respecto, si fueron respondidos o resueltos, y cuántas de estas respuestas y soluciones fueron dadas, cantidades que se detallan a continuación:

Institución	Reclamos	Respondidos	Respuestas	Comentarios	
Carabineros	1.008	1	0,1 %	1	11.175
COMPIN	532	0	0 %	0	3.483
JUNAEB	422	0	0 %	0	5.729
J-Familia	385	0	0 %	0	4.601
SERVIU	222	0	0 %	0	3.114
MINEDUC	211	0	0 %	0	2.148
J-Policía	126	0	0 %	0	1.329
M-Quilicura	103	0	0 %	0	772
FONASA	96	0	0 %	0	646
M-Maipú	85	0	0 %	0	735

Cuadro 5.21: Detalle de la cantidad total de Reclamos, Reclamos Respondidos, Respuestas y Comentarios realizados hacia las 10 Instituciones con más Reclamos y Recomendaciones del rubro de Gobierno.

Institución	Recomendaciones	Respondidas	Respuestas	Comentarios	
Carabineros	54	1	1,9 %	1	375
COMPIN	32	0	0 %	0	137
JUNAEB	14	0	0 %	0	62
J-Familia	108	0	0 %	0	958
SERVIU	12	0	0 %	0	144
MINEDUC	10	0	0 %	0	30
J-Policía	10	0	0 %	0	105
M-Quilicura	10	0	0 %	0	67
FONASA	6	0	0 %	0	99
M-Maipú	3	0	0 %	0	1

Cuadro 5.22: Detalle de la cantidad total de Recomendaciones, Recomendaciones Respondidas, Respuestas y Comentarios realizados hacia las 10 Instituciones con más Reclamos y Recomendaciones del rubro de Gobierno.

5.2.2. Carabineros de Chile

De manera análoga al DOUC, también se profundizó en los resultados y detalles a mostrar de la institución Carabineros de Chile.

Detalle de los Reclamos y de las Recomendaciones

A continuación, el detalle por año de las cantidades de reclamos y de recomendaciones, el número de estos que fueron respondidos, la cantidad de respuestas dadas al respecto, y la de comentarios dirigidos hacia estos reclamos y recomendaciones, que fueron generados por los usuarios que los leyeron:

Año	Reclamos		Respondidos		Respuestas		Comentarios	
2007	9	0,89 %	0	0 %	0	0 %	333	2,98 %
2008	50	4,96 %	0	0 %	0	0 %	1.303	11,66 %
2009	240	23,81 %	0	0 %	0	0 %	3.507	31,38 %
2010	314	31,15 %	1	100 %	1	100 %	2.980	26,67 %
2011	395	39,19 %	0	0 %	0	0 %	3.052	27,31 %
Totales	1.008	100 %	1	100 %	1	100 %	11.175	100 %

Cuadro 5.23: Detalle cronológico de las cantidades totales de Reclamos, Reclamos Respondidos, Respuestas y Comentarios realizados hacia Carabineros de Chile.

Año	Recomendaciones		Respondidas		Respuestas		Comentarios	
2007	0	0 %	0	0 %	0	0 %	0	0 %
2008	0	0 %	0	0 %	0	0 %	0	0 %
2009	0	0 %	0	0 %	0	0 %	0	0 %
2010	1	1,9 %	0	0 %	0	0 %	3	0,8 %
2011	53	98,1 %	0	0 %	0	0 %	372	99,2 %
Totales	54	100 %	0	0 %	0	0 %	375	100 %

Cuadro 5.24: Detalle cronológico de las cantidades totales de Recomendaciones, Recomendaciones Respondidas, Respuestas y Comentarios hechas hacia Carabineros de Chile.

Análisis LDA

Al igual que para el DOUC, también se realizó un análisis LDA para la institución de Carabineros de Chile, consistente en generar 20 tópicos de 30 palabras, cada uno sobre los contenidos de los reclamos que se hicieron contra la institución, tanto a nivel general, abarcando el conjunto completo de reclamos publicados entre el año 2007 y el año 2011, como también, por segmentos anuales, abarcando los reclamos publicados durante cada año.

Las tablas que se muestran a continuación poseen los resultados de estos análisis LDA, y se estructuran de manera tal, que los tópicos generados corresponden a las columnas, y las palabras que los conforman, junto con sus respectivas probabilidades asociadas, corresponden a las filas.

Las probabilidades asociadas a cada palabra, en este caso sólo sirven para ilustrar la preponderancia, o relevancia, que posee cada palabra al interior de cada tópico.

Resultados Análisis LDA Carabineros de Chile 2007 A

	Tópico 1		Tópico 2		Tópico 3		Tópico 4		Tópico 5		Tópico 6		Tópico 7		Tópico 8		Tópico 9		Tópico 10	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	no	0.010	chile	0.012	no	0.030	dinero	0.014	no	0.016	funcionario	0.010	no	0.012	parte	0.008	no	0.025	carabainero	0.020
2	funcionario	0.008	carabainero	0.010	carabainero	0.023	celular	0.012	muestro	0.015	comisaria	0.010	pasajero	0.007	carabainero	0.007	carabainero	0.017	no	0.019
3	carabainero	0.007	disparto	0.009	fiscalizar	0.009	funcionario	0.012	carabainero	0.015	dinero	0.010	conductor	0.007	calle	0.005	calle	0.010	pino	0.015
4	publicar	0.006	llamar	0.008	parte	0.014	carabainero	0.009	comisaria	0.012	estrada	0.009	decir	0.006	aviso	0.005	llamar	0.008	anotar	0.015
5	celular	0.006	haber	0.007	ser	0.011	estrada	0.008	producir	0.009	celular	0.008	colectivo	0.006	no	0.004	ciudad	0.008	choque	0.015
6	estrada	0.005	talcaluano	0.007	deber	0.010	chile	0.007	vecino	0.009	carabainero	0.008	colectivo	0.005	parecer	0.004	munoa	0.008	funcionario	0.012
7	decir	0.005	armar	0.005	os	0.010	comisaria	0.007	puerto	0.009	saez	0.006	tribunal	0.005	utilizar	0.004	hacer	0.007	decir	0.011
8	pasajero	0.005	molestar	0.005	aviso	0.010	habia	0.006	encuentro	0.009	santiago	0.005	medio	0.005	barato	0.004	decir	0.007	golpe	0.011
9	colectivo	0.005	lugar	0.005	existir	0.009	billettero	0.006	poblacion	0.009	chile	0.005	estar	0.005	chile	0.004	medio	0.007	claudio	0.010
10	dinero	0.005	insultar	0.005	quileura	0.009	publicar	0.006	esperar	0.009	detener	0.005	conducir	0.005	juez	0.004	matar	0.006	ley	0.010
11	chile	0.004	matute	0.005	guardia	0.009	santiago	0.006	fiscalizar	0.008	no	0.005	cuadra	0.005	molesto	0.004	hora	0.006	vercion	0.010
12	detener	0.004	dia	0.005	nada	0.009	saez	0.006	ta	0.006	habia	0.005	semaforo	0.005	abuso	0.004	conductor	0.006	frente	0.010
13	rojo	0.004	domingo	0.005	comma	0.008	institucion	0.005	favor	0.008	billettero	0.004	enseñar	0.005	caso	0.003	pasajero	0.006	c	0.010
14	comisaria	0.004	insulto	0.005	comisaria	0.008	detener	0.005	montt	0.008	investigacion	0.004	malb	0.005	palabra	0.003	estar	0.006	suponer	0.009
15	esperar	0.004	posterior	0.005	ciudadano	0.007	respuesta	0.005	ingresar	0.007	publicar	0.004	tipo	0.005	celebracion	0.003	seguridad	0.006	estrada	0.008
16	conductor	0.004	pañdilla	0.005	solicitar	0.007	fiscal	0.005	haber	0.007	ministerio	0.004	carabainero	0.005	querer	0.003	ciudadano	0.005	dinero	0.008
17	conducir	0.004	formar	0.005	hacer	0.007	huis	0.005	ubicar	0.007	obtener	0.004	trabajar	0.005	tocar	0.003	vecino	0.005	dia	0.007
18	santiago	0.004	lamentablemente	0.005	grande	0.007	investigacion	0.005	grande	0.006	institucion	0.004	publicar	0.005	actuo	0.003	dia	0.005	celular	0.007
19	denunciar	0.003	junto	0.005	chile	0.007	obtener	0.004	presencia	0.006	respuesta	0.004	rojo	0.005	previo	0.003	gritar	0.005	senor	0.007
20	iva	0.003	encontrar	0.005	seguridad	0.007	robar	0.004	ser	0.006	fiscal	0.004	documentacion	0.005	llamar	0.003	monnt	0.005	escribir	0.006
21	documentacion	0.003	significar	0.005	llamar	0.006	denunciar	0.004	pasar	0.006	imputar	0.004	iva	0.004	fe	0.003	colectivo	0.005	quitar	0.006
22	saez	0.003	calle	0.004	sera	0.006	documentacion	0.004	cerca	0.006	bio	0.004	funcionario	0.004	iva	0.003	colectivo	0.005	comisaria	0.006
23	tipo	0.003	dijero	0.004	documentacion	0.006	no	0.004	calle	0.006	robar	0.004	infraccion	0.004	ruido	0.003	reclamo	0.005	chile	0.006
24	querer	0.003	corrupto	0.004	nuestro	0.006	pasar	0.004	funcionario	0.005	sargento	0.004	denunciar	0.004	despues	0.003	mestro	0.005	justicia	0.006
25	semaforo	0.003	asistir	0.004	despues	0.006	actuar	0.004	guardia	0.005	pulsara	0.004	parte	0.004	entregar	0.003	personal	0.005	molesro	0.006
26	tenor	0.003	hebrtos	0.004	tienda	0.006	dicha	0.004	incluir	0.005	recibir	0.004	diario	0.003	colmo	0.003	semaforo	0.005	poner	0.006
27	tener	0.003	barrio	0.004	querer	0.006	ministerio	0.004	poder	0.005	fernando	0.003	cometer	0.003	cumpleaños	0.003	presencia	0.005	izquiero	0.006
28	estar	0.003	tambalear	0.004	villa	0.006	acaso	0.004	dia	0.005	esperar	0.003	tema	0.003	junio	0.003	grupo	0.004	reclamo	0.005
29	tribunal	0.003	vez	0.004	acceso	0.006	terminar	0.004	os	0.005	acaso	0.003	vehicular	0.003	deber	0.003	quedar	0.004	sorpresivo	0.005
30	trabajar	0.003	llamado	0.004	condominio	0.006	apellido	0.004	hogar	0.005	negativa	0.003	bajar	0.003	toque	0.003	documentacion	0.004	leida	0.005

Cuadro 5.25: Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2007.

Resultados Análisis LDA Carabineros de Chile 2007 B

	Tópico 11		Tópico 12		Tópico 13		Tópico 14		Tópico 15		Tópico 16		Tópico 17		Tópico 18		Tópico 19		Tópico 20	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	carabinero	0,002	no	0,002	carabinero	0,021	funcionario	0,019	carabinero	0,017	carabinero	0,033	no	0,005	carabinero	0,025	carabinero	0,002	funcionario	0,023
2	no	0,002	carabinero	0,002	calle	0,020	no	0,013	no	0,017	munca	0,016	pasajero	0,004	llamar	0,025	no	0,002	carabinero	0,021
3	calle	0,002	llamar	0,002	carabinero	0,020	llamar	0,010	carabinero	0,013	no	0,013	colectivo	0,004	chile	0,021	chile	0,002	celular	0,019
4	llamar	0,002	calle	0,002	ciudad	0,018	valdivia	0,009	dinero	0,011	matar	0,013	conductor	0,004	no	0,021	llamar	0,002	no	0,019
5	chile	0,002	chile	0,002	reclamo	0,014	seguridad	0,009	celular	0,011	hacer	0,010	tribunal	0,003	haber	0,016	funcionario	0,002	estrada	0,017
6	parte	0,002	parte	0,002	emergencia	0,012	matar	0,008	detener	0,009	vecino	0,009	rojo	0,003	talcahuano	0,015	celular	0,002	dinero	0,016
7	reclamo	0,002	ley	0,002	igual	0,008	gritar	0,008	chile	0,009	pedro	0,008	dejar	0,003	disparo	0,014	parte	0,002	comisaria	0,013
8	ley	0,002	ciudad	0,002	conducta	0,008	pedro	0,008	comisaria	0,009	joven	0,008	iva	0,003	fiscalizar	0,012	comisaria	0,002	publicar	0,013
9	fiscalizar	0,002	comisaria	0,002	subir	0,008	vecino	0,008	rojo	0,008	absolutamente	0,008	estar	0,003	os	0,009	calle	0,002	detener	0,011
10	comisaria	0,002	reclamo	0,002	forma	0,008	joven	0,008	decir	0,008	llamar	0,008	documentacion	0,003	ser	0,009	dinero	0,002	billeteo	0,011
11	ciudad	0,002	querer	0,002	personal	0,007	hora	0,008	publicar	0,008	manopla	0,008	semaforo	0,003	calle	0,008	estrada	0,002	chile	0,010
12	querer	0,002	dia	0,002	funcionario	0,007	grupo	0,008	estrada	0,007	ciudadano	0,008	infraccion	0,003	policia	0,008	dia	0,002	santiago	0,009
13	decir	0,002	fiscalizar	0,002	nuestro	0,007	hacer	0,008	santiago	0,007	valdivia	0,008	trabajar	0,003	intituciones	0,008	querer	0,002	seaz	0,008
14	dia	0,002	personal	0,002	asidero	0,006	ciudadano	0,007	seaz	0,007	seguridad	0,007	conducir	0,003	asaltar	0,008	ciudad	0,002	decir	0,008
15	haber	0,002	emergencia	0,002	augusto	0,006	manopla	0,007	estar	0,006	grupo	0,007	tener	0,003	parte	0,008	hacer	0,002	obtener	0,007
16	funcionario	0,002	haber	0,002	comisaria	0,006	no	0,007	conductor	0,006	comisaria	0,007	publicar	0,003	orden	0,008	nuestro	0,002	habia	0,007
17	dicha	0,002	despues	0,002	ejemplo	0,006	absolutamente	0,007	colectivo	0,006	dia	0,007	cuadra	0,003	colgar	0,008	ley	0,002	rojo	0,007
18	personal	0,002	dicha	0,002	pais	0,006	choque	0,007	dia	0,005	hora	0,007	funcionario	0,003	habian	0,008	dicha	0,002	fiscal	0,007
19	nuestro	0,002	nada	0,002	mi	0,006	dia	0,006	billeteo	0,005	gritar	0,006	medio	0,003	noche	0,008	seaz	0,002	ministerio	0,007
20	aviso	0,002	ser	0,002	calama	0,006	arotar	0,006	respuesta	0,005	decir	0,006	carabinero	0,003	bombero	0,008	decir	0,002	respuesta	0,006
21	nada	0,002	aviso	0,002	doble	0,006	quedar	0,006	ministerio	0,005	quedar	0,006	tomar	0,003	vez	0,008	haber	0,002	institucion	0,006
22	molesto	0,002	ciudadano	0,002	estacionar	0,006	suponer	0,005	institucion	0,005	funcionario	0,006	autoridad	0,003	reclamo	0,008	nada	0,002	investigacion	0,006
23	problema	0,002	decir	0,002	fila	0,006	manana	0,005	pasajero	0,005	munca	0,005	tipo	0,003	bastar	0,008	publicar	0,002	esperar	0,005
24	munca	0,002	hacer	0,002	respeto	0,006	palo	0,005	investigacion	0,005	diaz	0,005	ensenar	0,002	facil	0,008	fiscalizar	0,002	mayor	0,005
25	deber	0,002	pedro	0,002	entender	0,006	frente	0,005	documentacion	0,005	lejos	0,005	desprotejido	0,002	pedro	0,008	reclamo	0,002	dia	0,005
26	norte	0,002	molesto	0,002	clar	0,006	pino	0,005	medio	0,005	diciembre	0,005	valla	0,002	bomitando	0,008	santiago	0,002	s	0,005
27	despues	0,002	subir	0,002	transitar	0,006	jovenes	0,005	obtener	0,004	habian	0,005	recaltar	0,002	vivo	0,008	munca	0,002	anillo	0,005
28	sera	0,002	gracia	0,002	transitar	0,006	escuina	0,005	ensenar	0,004	mano	0,005	suvtieron	0,002	tambalear	0,008	detener	0,002	dos	0,005
29	ser	0,002	deber	0,002	dicha	0,006	decir	0,005	tener	0,004	portar	0,005	bulnes	0,002	san	0,008	molesto	0,002	denunciar	0,005
30	ministro	0,002	abuso	0,002	vial	0,006	bolivar	0,005	fiscal	0,004	fetal	0,005	serio	0,002	ayuda	0,008	pais	0,002	hais	0,005

Cuadro 5.26: Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2007.

Resultados Análisis LDA Carabineros de Chile 2008 A

	Tópico 1		Tópico 2		Tópico 3		Tópico 4		Tópico 5		Tópico 6		Tópico 7		Tópico 8		Tópico 9		Tópico 10	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	no	0,021	no	0,016	carabinero	0,029	carabinero	0,033	no	0,028	no	0,004	no	0,040	no	0,036	no	0,012	no	0,038
2	carabinero	0,015	carabinero	0,007	no	0,020	no	0,014	carabinero	0,015	carabinero	0,003	carabinero	0,016	carabinero	0,018	deber	0,012	carabinero	0,017
3	parte	0,010	hacer	0,006	dia	0,012	parte	0,010	parte	0,009	foto	0,002	auto	0,010	funcionario	0,010	encontrar	0,007	llamar	0,012
4	haber	0,009	auto	0,006	velocidad	0,008	persona	0,008	dia	0,008	parte	0,002	decir	0,009	lugar	0,009	ano	0,007	pasar	0,007
5	vehicular	0,008	traer	0,005	persona	0,006	casa	0,007	pasar	0,007	kmh	0,002	parte	0,009	calle	0,008	haber	0,007	legar	0,007
6	momento	0,008	uso	0,005	momento	0,006	padre	0,006	telefono	0,006	reclamo	0,002	hacer	0,008	reclamo	0,008	documento	0,007	haber	0,007
7	ser	0,008	reclamo	0,005	ver	0,006	estar	0,006	error	0,006	traer	0,002	estar	0,008	curso	0,007	pasar	0,006	dia	0,006
8	j	0,006	tener	0,005	violencia	0,006	velocidad	0,005	movil	0,006	urbano	0,002	nada	0,007	hacer	0,007	respuesta	0,006	sector	0,006
9	ley	0,006	dar	0,005	critero	0,006	indicar	0,004	decir	0,005	chile	0,001	decir	0,006	contra	0,007	medio	0,006	detener	0,006
10	pasar	0,006	hijo	0,004	key	0,005	habia	0,004	haber	0,005	zona	0,001	haber	0,006	institucion	0,007	reclamo	0,006	parte	0,006
11	virar	0,005	grado	0,004	agresor	0,005	dia	0,004	llamar	0,005	estacionar	0,001	pasar	0,006	infraccion	0,007	razon	0,006	decir	0,005
12	hacer	0,005	personal	0,004	procedimiento	0,005	hacer	0,004	bueno	0,005	droga	0,001	dar	0,006	decir	0,007	solucion	0,006	auto	0,005
13	avda	0,005	vehiculos	0,004	funcionario	0,005	auto	0,004	detener	0,004	vehiculos	0,001	persona	0,006	estacionar	0,006	rogar	0,006	dejar	0,005
14	infraccionar	0,005	funcionario	0,004	encontrar	0,005	deber	0,004	comisaria	0,004	colegio	0,001	paco	0,005	chile	0,006	nadie	0,006	dar	0,005
15	comisaria	0,005	siguiente	0,004	hacer	0,005	sacar	0,004	falta	0,004	pasar	0,001	calle	0,005	pasar	0,006	ambito	0,006	estar	0,005
16	civil	0,005	colegio	0,004	decir	0,005	policia	0,004	situacion	0,004	uso	0,001	llamar	0,005	persona	0,006	insistencia	0,006	lugar	0,005
17	proceder	0,004	ser	0,004	parte	0,005	realizar	0,004	cumplir	0,004	hijo	0,001	contra	0,005	siquiera	0,006	faras	0,006	denunciar	0,005
18	nogal	0,004	chile	0,004	lugar	0,005	detener	0,004	esperar	0,004	dar	0,001	casa	0,005	dato	0,005	judicial	0,006	cuadrante	0,004
19	nombre	0,004	estar	0,003	sucar	0,005	decir	0,004	policia	0,004	sector	0,001	poner	0,004	transitar	0,005	el	0,006	hacer	0,004
20	copiapo	0,004	ciudad	0,003	gente	0,005	foto	0,004	multar	0,004	2	0,001	tipo	0,004	documento	0,005	fotografias	0,006	jovenes	0,004
21	necesitar	0,004	esperar	0,003	paso	0,004	esperar	0,003	recurrir	0,004	hacer	0,001	realizar	0,004	siguiente	0,005	gracia	0,006	venir	0,004
22	contactar	0,004	haber	0,003	mujer	0,004	funcionario	0,003	funcionario	0,003	scr	0,001	lugar	0,004	parte	0,005	mano	0,006	funcionario	0,004
23	detener	0,004	decir	0,003	habian	0,004	momento	0,003	conducia	0,003	indicar	0,001	llevar	0,004	medicar	0,004	cerca	0,006	minuto	0,004
24	institucion	0,004	q	0,003	detener	0,004	andar	0,003	hacer	0,003	revision	0,001	llegar	0,004	permitir	0,004	detalle	0,006	sacar	0,004
25	justificar	0,004	parte	0,003	falta	0,004	llamar	0,003	estacionar	0,003	llevar	0,001	hora	0,003	personal	0,004	adjunto	0,006	calle	0,004
26	ano	0,003	contra	0,003	golpear	0,004	salir	0,003	auto	0,003	revisar	0,001	estacionar	0,003	entregar	0,004	incumplimiento	0,006	contra	0,004
27	cabo	0,003	vehicular	0,003	pasar	0,004	interior	0,003	hora	0,003	llegar	0,001	indicar	0,003	ley	0,004	jorge	0,006	hora	0,004
28	estar	0,003	estimar	0,003	identificar	0,004	controlar	0,003	recibir	0,003	carretera	0,001	parecer	0,003	indicar	0,004	obtener	0,006	munca	0,004
29	identificacion	0,003	llevar	0,003	hora	0,004	bueno	0,003	santo	0,003	pinera	0,001	dia	0,003	buscar	0,003	carabinero	0,005	el	0,004
30	publicar	0,003	encuentro	0,003	fisica	0,003	nada	0,003	deber	0,003	retar	0,001	dejar	0,003	policia	0,003	momento	0,003	fecha	0,004

Cuadro 5.27: Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2008.

Resultados Análisis LDA Carabineros de Chile 2009 A

	Tópico 1		Tópico 2		Tópico 3		Tópico 4		Tópico 5		Tópico 6		Tópico 7		Tópico 8		Tópico 9		Tópico 10	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	no	0.012	no	0.031	no	0.029	no	0.035	no	0.038	no	0.032	no	0.036	no	0.025	no	0.023	no	0.031
2	carabinero	0.009	carabinero	0.027	carabinero	0.024	carabinero	0.019	carabinero	0.022	carabinero	0.023	carabinero	0.023	carabinero	0.021	carabinero	0.018	carabinero	0.020
3	numerar	0.006	q	0.011	hijo	0.007	dar	0.010	q	0.009	parte	0.007	ley	0.006	parte	0.010	comisaria	0.007	parte	0.007
4	año	0.005	parte	0.009	decir	0.007	parte	0.009	funcionario	0.006	key	0.006	dar	0.006	hacer	0.009	dar	0.006	hacer	0.007
5	perro	0.005	auto	0.007	haber	0.007	día	0.007	hacer	0.006	vehicular	0.005	decir	0.006	auto	0.008	vehicular	0.005	comisaria	0.006
6	hijo	0.005	nada	0.006	ser	0.006	hijo	0.006	decir	0.006	estacionar	0.005	hijo	0.005	día	0.008	hacer	0.005	calle	0.006
7	placa	0.005	dar	0.006	hacer	0.006	año	0.006	dar	0.006	día	0.005	nada	0.005	detener	0.008	parte	0.005	lugar	0.005
8	dar	0.004	día	0.005	hacer	0.005	comisaria	0.005	pasar	0.005	lugar	0.005	año	0.005	pasar	0.006	año	0.004	haber	0.005
9	parte	0.004	institucion	0.005	hacer	0.006	año	0.006	dar	0.006	lugar	0.005	año	0.005	pasar	0.006	año	0.004	haber	0.005
10	pasar	0.004	detener	0.005	hacer	0.005	momento	0.005	día	0.005	comisaria	0.005	querer	0.005	dar	0.006	persona	0.004	infraccion	0.005
11	el	0.004	momento	0.005	hacer	0.005	momento	0.005	momento	0.005	momento	0.005	querer	0.005	dar	0.006	persona	0.004	infraccion	0.005
12	lugar	0.004	momento	0.005	hacer	0.005	momento	0.005	momento	0.005	momento	0.005	querer	0.005	dar	0.006	persona	0.004	infraccion	0.005
13	sacar	0.003	tener	0.004	haber	0.004	haber	0.004	haber	0.004	nada	0.005	tener	0.005	tener	0.005	habia	0.004	nuestro	0.004
14	haber	0.003	hacer	0.004	dar	0.004	haber	0.004	haber	0.004	nada	0.005	tener	0.005	tener	0.005	habia	0.004	nuestro	0.004
15	escribir	0.003	pasar	0.004	parte	0.004	detener	0.004	comisaria	0.004	hijo	0.004	esperar	0.004	decir	0.005	auto	0.003	encontrar	0.004
16	chile	0.003	llevar	0.004	forma	0.004	decir	0.004	chile	0.004	infraccion	0.004	mal	0.004	habia	0.005	haber	0.003	reclamo	0.004
17	comisaria	0.003	encontrar	0.004	llevar	0.004	llevar	0.004	llamar	0.004	pasar	0.004	encontrar	0.004	funcionario	0.004	nombre	0.003	detener	0.004
18	tres	0.003	querer	0.004	pasar	0.004	funcionario	0.004	madre	0.004	comisaria	0.004	llevar	0.004	amigo	0.004	realizar	0.003	llegar	0.003
19	cuidar	0.003	habia	0.004	encontrar	0.004	quedar	0.004	casa	0.004	detener	0.004	pasar	0.003	llegar	0.004	institucion	0.003	ley	0.003
20	propio	0.002	procedimiento	0.003	situacion	0.004	querer	0.004	querer	0.004	huz	0.003	deber	0.003	llevar	0.004	denunciar	0.003	deber	0.003
21	robar	0.002	hermano	0.003	ley	0.004	lugar	0.003	llevar	0.004	estar	0.003	tener	0.003	tener	0.003	institucion	0.003	tomar	0.003
22	delincuente	0.002	haber	0.003	cintron	0.004	nuestro	0.003	año	0.004	chile	0.003	comisaria	0.003	preguntar	0.004	hora	0.003	casa	0.003
23	sinistro	0.002	infraccion	0.003	chile	0.004	persona	0.003	estar	0.004	situacion	0.003	chile	0.003	ley	0.004	reclamo	0.003	denunciar	0.003
24	diestro	0.002	calle	0.003	comisaria	0.003	deber	0.003	momento	0.003	momento	0.003	denunciar	0.003	año	0.004	alguien	0.003	día	0.003
25	publicar	0.002	nuestro	0.003	lugar	0.003	llamar	0.003	hermano	0.003	dar	0.003	ser	0.003	vehicular	0.003	abuelo	0.003	chile	0.003
26	vidrio	0.002	pedir	0.003	vehicular	0.003	chile	0.003	institucion	0.003	decir	0.003	llegar	0.003	haber	0.003	llegar	0.003	momento	0.003
27	perdon	0.002	decir	0.003	nada	0.003	tener	0.003	llegar	0.003	vehiculos	0.003	respuesta	0.003	pedir	0.003	numerar	0.003	hijo	0.003
28	anotar	0.002	poder	0.003	deber	0.003	salir	0.003	hijo	0.003	casa	0.003	parte	0.003	estar	0.003	llevar	0.003	llevar	0.003
29	z	0.002	chile	0.003	trabajar	0.003	familia	0.003	calle	0.003	reclamo	0.003	estar	0.003	hora	0.003	nada	0.003	algo	0.003
30	esconder	0.002	responder	0.003	publicar	0.003	hora	0.003	detener	0.003	mal	0.003	persona	0.003	dejar	0.003	detener	0.003	perro	0.003

Cuadro 5.29: Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2009.

Resultados Análisis LDA Carabineros de Chile 2009 B

	Tópico 11		Tópico 12		Tópico 13		Tópico 14		Tópico 15		Tópico 16		Tópico 17		Tópico 18		Tópico 19		Tópico 20	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	no	0,015	no	0,033	no	0,031	carabinero	0,015	carabinero	0,028	no	0,035	no	0,025	no	0,050	no	0,020	no	0,027
2	decir	0,011	carabinero	0,016	carabinero	0,025	no	0,014	no	0,027	carabinero	0,017	carabinero	0,018	carabinero	0,027	carabinero	0,015	carabinero	0,019
3	carabinero	0,010	parte	0,009	hacer	0,007	hacer	0,006	hacer	0,009	parte	0,008	hacer	0,011	parte	0,009	ano	0,007	ley	0,007
4	ano	0,010	pasar	0,008	pasar	0,007	persona	0,005	dar	0,009	hacer	0,006	dia	0,008	decir	0,008	dar	0,006	hacer	0,007
5	hijo	0,007	haber	0,006	decir	0,007	haber	0,005	funcionario	0,008	muestro	0,005	funcionario	0,007	hacer	0,008	hijo	0,006	parte	0,007
6	hacer	0,006	comisaria	0,006	dia	0,006	entrar	0,005	decir	0,007	persona	0,005	parte	0,007	pasar	0,006	estar	0,006	vehicular	0,006
7	querer	0,004	hacer	0,006	dar	0,006	parir	0,005	pasar	0,006	vehicular	0,005	dar	0,006	dar	0,006	dar	0,004	pasar	0,006
8	después	0,004	decir	0,005	detener	0,006	hacer	0,005	haber	0,006	llevar	0,004	pasar	0,006	haber	0,005	haber	0,004	detener	0,005
9	haber	0,004	dia	0,005	dejar	0,005	disco	0,004	hijo	0,006	ley	0,004	decir	0,006	estar	0,005	funcionario	0,004	comisaria	0,005
10	tenia	0,004	dar	0,005	deber	0,005	decir	0,004	dia	0,005	dia	0,004	ano	0,005	deber	0,005	parte	0,004	q	0,004
11	dar	0,004	ano	0,005	habia	0,005	deber	0,004	estar	0,005	pasar	0,004	comisaria	0,005	detener	0,005	llegar	0,003	lugar	0,004
12	vehicular	0,004	llegar	0,005	calle	0,005	semafóro	0,004	parte	0,004	llegar	0,003	llamar	0,005	general	0,005	general	0,003	oficial	0,004
13	denunciar	0,004	detener	0,005	persona	0,005	chile	0,004	calle	0,004	momento	0,003	hora	0,004	llamar	0,005	documento	0,003	querer	0,004
14	pasar	0,004	calle	0,004	tomar	0,004	parte	0,004	vehicular	0,004	estar	0,003	auto	0,004	dia	0,005	hacer	0,003	ciudadano	0,004
15	perro	0,003	ver	0,004	estar	0,004	habia	0,004	hablar	0,004	tener	0,003	deber	0,004	persona	0,005	entrar	0,003	decir	0,004
16	situación	0,003	q	0,004	vehicular	0,004	casa	0,004	trabajo	0,004	denunciar	0,003	habia	0,004	habia	0,004	contra	0,003	reclamo	0,004
17	encantar	0,003	casa	0,004	parte	0,004	llegar	0,003	detener	0,004	decir	0,003	instiitucion	0,004	ano	0,004	denunciar	0,003	habia	0,004
18	parte	0,003	nada	0,004	momento	0,004	estadio	0,003	ano	0,004	auto	0,003	preguntar	0,004	llegar	0,004	ser	0,003	persona	0,004
19	chileno	0,003	persona	0,004	haber	0,004	realmente	0,003	comisaria	0,004	habia	0,003	persona	0,004	casa	0,004	dia	0,003	haber	0,004
20	persona	0,003	dejar	0,003	chile	0,004	dejar	0,003	chile	0,004	oficial	0,003	momento	0,004	ley	0,003	pasar	0,003	velocidad	0,004
21	algo	0,003	auto	0,003	reclamo	0,004	llevar	0,003	llegar	0,004	cursor	0,003	llevar	0,003	comisaria	0,003	petro	0,003	auto	0,004
22	edar	0,003	habia	0,003	ser	0,004	dar	0,003	lugar	0,004	estar	0,003	haber	0,003	hijo	0,003	muestro	0,003	dia	0,004
23	volver	0,003	zona	0,003	comisaria	0,003	poder	0,003	instiitucion	0,003	documento	0,003	funcionario	0,003	pasaje	0,003	detener	0,003	funcionario	0,004
24	nada	0,003	pedir	0,003	nada	0,003	detener	0,003	habia	0,003	lugar	0,003	ley	0,003	sacar	0,003	casa	0,003	encontrar	0,004
25	mujer	0,003	tomar	0,003	encontrar	0,003	calle	0,003	general	0,003	comisaria	0,003	dejar	0,003	niños	0,003	matar	0,002	tener	0,004
26	quitar	0,003	tener	0,003	tenia	0,003	esperar	0,003	deber	0,003	llamar	0,003	sacar	0,003	familia	0,003	comisaria	0,002	ir	0,003
27	niños	0,003	llevar	0,003	esperar	0,003	revisar	0,003	querer	0,003	luz	0,003	llegar	0,003	llevar	0,003	ver	0,002	llegar	0,003
28	deber	0,003	poder	0,003	llamar	0,003	llamar	0,003	tever	0,003	indicar	0,003	hablar	0,003	tenia	0,003	menstual	0,002	dejar	0,003
29	detener	0,003	hs	0,003	tener	0,003	sau	0,003	ley	0,003	dar	0,003	documento	0,003	funcionario	0,003	encontrar	0,002	dejar	0,003
30	domicilio	0,003	madre	0,003	funcionario	0,003	tenia	0,002	familia	0,003	esperar	0,003	calle	0,003	dejar	0,003	dias	0,002	dar	0,003

Cuadro 5.30: Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2009.

Resultados Análisis LDA Carabineros de Chile 2010 A

	Tópico 1		Tópico 2		Tópico 3		Tópico 4		Tópico 5		Tópico 6		Tópico 7		Tópico 8		Tópico 9		Tópico 10		
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	
1	no	0.036	no	0.037	no	0.030	no	0.028	no	0.020	no	0.034	no	0.036	no	0.033	no	0.013	carabinero	0.028	
2	carabinero	0.010	carabinero	0.025	carabinero	0.017	carabinero	0.018	carabinero	0.017	carabinero	0.017	carabinero	0.028	carabinero	0.023	carabinero	0.010	no	no	0.025
3	persona	0.007	decir	0.010	llamar	0.007	hacer	0.008	pasar	0.007	hacer	0.006	decir	0.008	persona	0.007	decir	0.005	pasar	0.007	
4	hacer	0.007	dar	0.009	agua	0.007	calle	0.006	detener	0.006	institucion	0.006	dar	0.008	dar	0.006	parte	0.004	llamar	0.006	
5	pasar	0.006	pasar	0.008	parte	0.006	parte	0.006	dar	0.006	parte	0.005	pasar	0.007	parte	0.006	pasar	0.004	pagar	0.005	
6	dar	0.005	hacer	0.007	haber	0.006	chile	0.005	hora	0.006	dar	0.005	dar	0.007	decir	0.005	derecho	0.004	dar	0.005	
7	casa	0.005	haber	0.006	decir	0.005	tener	0.004	ano	0.006	llevar	0.004	dia	0.006	hacer	0.005	haber	0.004	hacer	0.005	
8	detener	0.005	vehicular	0.005	hacer	0.005	dar	0.004	querer	0.006	chile	0.004	llegar	0.005	deber	0.005	delinciente	0.003	parte	0.005	
9	llamar	0.005	parte	0.005	persona	0.005	llevar	0.004	comisaria	0.004	persona	0.004	estar	0.005	llamar	0.005	pedir	0.003	cabo	0.005	
10	encontrar	0.004	llamar	0.005	pasar	0.005	detener	0.004	parte	0.005	tener	0.004	nuestro	0.005	pedir	0.005	key	0.003	deber	0.005	
11	parte	0.004	ver	0.004	deber	0.005	bus	0.004	llegar	0.005	trabajo	0.004	hacer	0.005	pasar	0.004	dar	0.003	detener	0.005	
12	decir	0.004	mal	0.004	lugar	0.004	cursar	0.004	persona	0.005	deber	0.004	nada	0.005	detener	0.004	esperar	0.003	estar	0.005	
13	seguir	0.004	institucion	0.004	llegar	0.004	funcionario	0.003	dia	0.005	dia	0.004	detener	0.005	institucion	0.004	detener	0.003	lugar	0.005	
14	querer	0.004	llevar	0.004	dar	0.004	momento	0.003	haber	0.005	lugar	0.004	deber	0.005	vehicular	0.004	persona	0.002	decir	0.004	
15	ser	0.004	hijo	0.004	hijo	0.004	persona	0.003	hacer	0.004	ser	0.003	persona	0.004	mestro	0.004	deber	0.002	dia	0.004	
16	dia	0.004	ano	0.004	menos	0.004	decir	0.003	reclamo	0.004	derecho	0.003	casa	0.004	esperar	0.004	hacer	0.002	llegar	0.004	
17	estar	0.004	dia	0.004	vehicular	0.003	pedir	0.003	reclamo	0.004	haber	0.003	infraccion	0.004	ser	0.004	agua	0.002	reclamo	0.004	
18	lugar	0.004	nada	0.004	ano	0.003	reclamo	0.003	llevar	0.004	ley	0.003	chile	0.004	ano	0.004	querer	0.002	casa	0.004	
19	calle	0.004	comisaria	0.004	querer	0.003	hora	0.003	noche	0.004	familia	0.003	vehicular	0.004	casa	0.004	estar	0.002	comisaria	0.004	
20	vehicular	0.004	llegar	0.004	estar	0.003	ser	0.003	ser	0.003	hijo	0.003	dejar	0.004	querer	0.003	cambiar	0.002	haber	0.004	
21	haber	0.004	seguir	0.004	estar	0.003	pasar	0.003	institucion	0.003	llamar	0.003	comisaria	0.004	llevar	0.003	documento	0.002	infraccion	0.003	
22	transitar	0.004	estar	0.004	comisaria	0.003	oficial	0.003	dejar	0.003	querer	0.003	calle	0.003	cabo	0.003	lugar	0.002	persona	0.003	
23	esperar	0.004	dejar	0.004	calle	0.003	q	0.003	decir	0.003	1	0.003	llamar	0.003	haber	0.003	pista	0.002	retirar	0.003	
24	comisaria	0.003	persona	0.003	forma	0.003	ley	0.003	deber	0.003	pedir	0.003	institucion	0.003	auto	0.003	mayor	0.002	momento	0.003	
25	nada	0.003	casa	0.003	publicar	0.003	institucion	0.003	ir	0.003	pagar	0.003	momento	0.003	q	0.003	sector	0.002	preocupar	0.003	
26	ver	0.003	detener	0.003	tratar	0.003	dia	0.003	mestro	0.003	esperar	0.003	tomar	0.003	oficial	0.003	huz	0.002	nada	0.003	
27	hoy	0.003	moto	0.003	habia	0.003	tenia	0.003	nada	0.003	causa	0.003	quedar	0.003	conductor	0.003	familia	0.002	situacion	0.003	
28	llevar	0.003	esperar	0.003	detener	0.003	deber	0.003	encontrar	0.003	encontrar	0.003	reclamo	0.003	procedimiento	0.003	derecha	0.002	hablar	0.003	
29	policia	0.003	sacar	0.003	pedir	0.003	nada	0.003	documentar	0.003	padre	0.003	oficial	0.003	sacar	0.003	nada	0.002	delinciente	0.003	
30	poder	0.003	cabo	0.003	dejar	0.003	estar	0.003	esperar	0.003	calle	0.003	ir	0.003	salir	0.003	san	0.002	tener	0.003	

Cuadro 5.31: Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2010.

Resultados Análisis LDA Carabineros de Chile 2010 B

	Tópico 11		Tópico 12		Tópico 13		Tópico 14		Tópico 15		Tópico 16		Tópico 17		Tópico 18		Tópico 19		Tópico 20	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	no	0,029	no	0,028	no	0,037	no	0,029	no	0,032	no	0,022	no	0,022	no	0,019	carabinero	0,034	no	0,022
2	carabinero	0,014	carabinero	0,027	carabinero	0,016	carabinero	0,023	carabinero	0,021	carabinero	0,021	carabinero	0,010	carabinero	0,012	no	0,031	carabinero	0,013
3	parte	0,008	parte	0,009	hacer	0,008	llamar	0,010	hacer	0,006	hacer	0,009	hacer	0,009	llamar	0,007	hacer	0,006	haber	0,009
4	q	0,008	dar	0,007	dia	0,007	decir	0,006	llegar	0,005	pasar	0,006	hacer	0,007	hacer	0,006	pasar	0,006	llamar	0,008
5	dia	0,006	deber	0,005	parte	0,006	hacer	0,006	querer	0,005	llevar	0,006	hacer	0,006	haber	0,005	llegar	0,006	dar	0,008
6	casa	0,005	ano	0,005	vehicular	0,005	dia	0,005	haber	0,005	deber	0,006	dar	0,006	corporacion	0,005	vehicular	0,006	hacer	0,007
7	dar	0,005	chile	0,005	dar	0,005	vehicular	0,005	decir	0,005	dia	0,005	vehicular	0,004	ir	0,004	dar	0,006	contestar	0,005
8	hacer	0,005	hacer	0,005	hermano	0,005	llegar	0,005	dia	0,005	haber	0,005	seguir	0,004	llegar	0,004	parte	0,006	hora	0,005
9	nuestro	0,005	detener	0,004	esperar	0,005	transitar	0,005	parte	0,005	institucion	0,004	llegar	0,004	pasar	0,004	deber	0,005	persona	0,005
10	persona	0,005	dia	0,004	nada	0,005	dar	0,004	encontrar	0,004	dar	0,005	deber	0,004	pasar	0,004	dia	0,005	pasar	0,005
11	llevar	0,005	vehicular	0,004	llegar	0,005	haber	0,004	dar	0,004	detener	0,005	parte	0,004	detener	0,003	hijo	0,005	casa	0,004
12	salir	0,005	hora	0,004	pasar	0,005	persona	0,004	qe	0,004	decir	0,004	hijo	0,004	calle	0,003	llamar	0,005	despues	0,004
13	procedimiento	0,004	haber	0,004	dejar	0,004	estacionar	0,004	detener	0,004	funcionario	0,004	funcionario	0,004	despues	0,003	comisaria	0,005	querer	0,003
14	llamar	0,004	esperar	0,004	persona	0,004	pasar	0,004	funcionario	0,004	persona	0,004	pasar	0,004	lugar	0,003	estar	0,004	dia	0,003
15	estar	0,004	institucion	0,004	llevar	0,004	tipo	0,004	vehicular	0,004	parte	0,004	paso	0,004	forma	0,003	llevar	0,004	deber	0,003
16	decir	0,004	llevar	0,004	calle	0,004	vehiculos	0,004	velocidad	0,004	querer	0,004	pagar	0,004	dos	0,003	detener	0,004	auto	0,003
17	hora	0,004	llegar	0,004	decir	0,004	lugar	0,003	pasar	0,004	llamar	0,004	tomar	0,004	persona	0,003	institucion	0,004	detener	0,003
18	deber	0,004	llamar	0,003	habia	0,004	lugar	0,003	persona	0,003	hijo	0,004	persona	0,003	hoy	0,003	nada	0,004	realizar	0,003
19	comisaria	0,004	situacion	0,003	momento	0,004	parte	0,003	poder	0,003	calle	0,004	encontrar	0,003	querer	0,003	decir	0,003	encontrar	0,003
20	hijo	0,003	pasar	0,003	q	0,004	denunciar	0,003	estar	0,003	ley	0,003	oficial	0,003	momento	0,003	calle	0,003	nuestro	0,003
21	cabo	0,003	tener	0,003	detener	0,004	despues	0,003	nuestro	0,003	tener	0,003	haber	0,003	dar	0,003	casa	0,003	habia	0,003
22	seguir	0,003	reclamo	0,003	ley	0,003	encontrar	0,003	llevar	0,003	el	0,003	dia	0,003	estacionar	0,003	encontrar	0,003	vehicular	0,003
23	mal	0,003	funcionario	0,003	pedir	0,003	detener	0,003	ser	0,003	mal	0,003	reclamo	0,003	auto	0,003	persona	0,003	lugar	0,003
24	habia	0,003	tenia	0,003	seguir	0,003	nada	0,003	dios	0,003	tratar	0,003	ley	0,003	contar	0,003	habia	0,003	tipo	0,003
25	nada	0,003	querer	0,003	llamar	0,003	documento	0,003	noche	0,003	llegar	0,003	nuestro	0,003	decir	0,003	documento	0,003	comisaria	0,002
26	ano	0,003	decir	0,003	querer	0,003	accidente	0,003	tomar	0,003	pedir	0,003	ano	0,003	chile	0,002	tener	0,003	cerca	0,002
27	llegar	0,003	tomar	0,003	nuestro	0,003	calle	0,003	dos	0,003	estar	0,003	habia	0,003	tipo	0,002	delincuente	0,003	cabo	0,002
28	pasar	0,003	nuestro	0,003	estar	0,003	derecho	0,003	bueno	0,003	estacionar	0,003	dos	0,003	publicar	0,002	lugar	0,003	decir	0,002
29	contra	0,003	estar	0,003	haber	0,003	esperar	0,003	zona	0,003	encontrar	0,003	contra	0,003	penetrecer	0,002	salir	0,003	pedir	0,002
30	ser	0,003	dejar	0,003	ser	0,003	llevar	0,003	momento	0,003	vehicular	0,003	comisaria	0,003	habia	0,002	infraccion	0,003	derecho	0,002

Cuadro 5.32: Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2010.

Resultados Análisis LDA Carabineros de Chile 2011 A

	Tópico 1		Tópico 2		Tópico 3		Tópico 4		Tópico 5		Tópico 6		Tópico 7		Tópico 8		Tópico 9		Tópico 10	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	carabinero	0,019	no	0,016	no	0,041	no	0,011	no	0,016	carabinero	0,020	carabinero	0,017	no	0,037	carabinero	0,019	carabinero	0,020
2	no	0,018	carabinero	0,016	carabinero	0,024	carabinero	0,007	carabinero	0,011	no	0,017	no	0,008	carabinero	0,030	no	0,016	no	0,017
3	hacer	0,006	hijo	0,009	hacer	0,007	persona	0,006	calle	0,009	institucion	0,006	ley	0,005	decir	0,008	parte	0,010	dar	0,009
4	parte	0,005	hacer	0,006	haber	0,007	comisaria	0,005	parte	0,008	hijo	0,006	calle	0,004	hacer	0,008	haber	0,009	oficial	0,006
5	institucion	0,005	tener	0,006	llamar	0,006	hacer	0,004	estacionar	0,008	auto	0,005	ano	0,004	parte	0,008	decir	0,009	comisaria	0,005
6	día	0,005	día	0,005	nada	0,006	chile	0,004	perro	0,006	comisaria	0,005	dar	0,004	dar	0,007	pasar	0,007	auto	0,005
7	decir	0,004	haber	0,005	dar	0,006	parte	0,003	estar	0,005	estacionar	0,005	vina	0,004	día	0,007	llegar	0,006	haber	0,004
8	chile	0,004	pagar	0,004	decir	0,005	dar	0,003	decir	0,004	general	0,005	funcionario	0,003	pasar	0,006	día	0,006	q	0,004
9	dar	0,004	ser	0,004	comisaria	0,005	solicitar	0,003	publicar	0,004	hacer	0,005	estacionar	0,003	haber	0,005	estar	0,006	día	0,004
10	persona	0,004	llamar	0,003	día	0,005	llevar	0,003	hacer	0,004	reclamo	0,005	situacion	0,003	detener	0,005	despues	0,005	llegar	0,004
11	pasar	0,004	conductor	0,003	pasar	0,004	día	0,003	detener	0,003	pasar	0,005	parte	0,003	comisaria	0,005	comisaria	0,005	hora	0,004
12	estar	0,004	tres	0,003	llegar	0,004	encontrar	0,003	nada	0,003	parte	0,004	día	0,003	deber	0,005	llevar	0,004	ano	0,004
13	haber	0,003	ano	0,003	funcionario	0,004	haber	0,003	lugar	0,003	ano	0,004	institucion	0,003	habia	0,005	nada	0,004	pasar	0,004
14	llegar	0,003	el	0,003	días	0,004	despues	0,003	ley	0,003	detener	0,004	comisaria	0,003	persona	0,005	esperar	0,004	estar	0,003
15	llamar	0,003	paso	0,003	deber	0,004	nuestro	0,003	reclamo	0,003	llegar	0,004	haber	0,003	querer	0,005	funcionario	0,004	san	0,003
16	protesta	0,003	reclamo	0,003	despues	0,004	caballo	0,002	auto	0,003	dar	0,004	mar	0,003	llamar	0,004	oficial	0,004	deber	0,003
17	comisaria	0,003	parte	0,003	ley	0,004	oficial	0,002	querer	0,003	ser	0,004	hijo	0,003	reclamo	0,004	problema	0,003	chile	0,003
18	infraccion	0,003	deber	0,003	parte	0,004	sacar	0,002	persona	0,003	persona	0,004	lado	0,003	estar	0,004	hijo	0,003	denunciar	0,003
19	hijo	0,003	llevar	0,003	denunciar	0,003	ano	0,002	ano	0,003	indicar	0,004	estacionar	0,003	funcionario	0,004	sacar	0,003	funcionario	0,003
20	general	0,003	problema	0,003	institucion	0,003	hijo	0,002	menos	0,003	mal	0,004	lugar	0,002	pedir	0,004	dar	0,003	ley	0,003
21	conocer	0,003	dar	0,003	tener	0,003	días	0,002	pasar	0,003	realizar	0,004	llegar	0,002	llegar	0,003	hacer	0,003	persona	0,003
22	habia	0,003	nada	0,003	chile	0,003	licencia	0,002	ver	0,003	mayor	0,004	auto	0,002	ano	0,003	hacer	0,003	tomar	0,003
23	nada	0,002	vivir	0,003	dejar	0,003	funcionario	0,002	vehicular	0,003	funcionario	0,003	policial	0,002	ser	0,003	querer	0,003	encontrar	0,003
24	hora	0,002	trabajo	0,003	ser	0,003	decir	0,002	uso	0,003	llamar	0,003	decir	0,002	institucion	0,003	llamar	0,003	hacer	0,003
25	ser	0,002	institucion	0,003	auto	0,003	publicar	0,002	haber	0,003	decir	0,003	dinero	0,002	chile	0,003	lugar	0,003	bajo	0,003
26	transitar	0,002	bajo	0,003	derecho	0,003	llegar	0,002	solicitar	0,002	malo	0,003	ver	0,002	dejar	0,003	dos	0,002	dejar	0,003
27	reclamo	0,002	decir	0,002	esperar	0,003	denunciar	0,002	nuestro	0,002	gordon	0,003	caballo	0,002	llevar	0,003	chile	0,002	detener	0,003
28	servicio	0,002	considerar	0,002	contra	0,003	realizar	0,002	pedir	0,002	hora	0,003	personal	0,002	nada	0,003	habia	0,002	tener	0,003
29	casa	0,002	cuadra	0,002	oficial	0,003	mandar	0,002	tomar	0,002	haber	0,003	publicar	0,002	encontrar	0,003	encontrar	0,002	despues	0,003
30	hablar	0,002	poder	0,002	gracia	0,003	animal	0,002	esperar	0,002	llevar	0,003	hacer	0,002	hora	0,003	familia	0,002	estacionar	0,002

Cuadro 5.33: Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile durante el año 2011.

Resultados Análisis LDA Carabineros de Chile 2011 B

	Tópico 11	Tópico 12	Tópico 13	Tópico 14	Tópico 15	Tópico 16	Tópico 17	Tópico 18	Tópico 19	Tópico 20
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	no	0,020	no	0,044	no	0,026	no	0,030	no	0,021
2	carabinero	0,015	carabinero	0,027	carabinero	0,015	carabinero	0,024	carabinero	0,014
3	funcionario	0,008	decir	0,008	carabinero	0,006	haber	0,008	decir	0,008
4	hacer	0,006	hacer	0,006	hacer	0,006	hacer	0,007	comisaria	0,006
5	llamar	0,006	llamar	0,006	llamar	0,004	llamar	0,006	hacer	0,005
6	dar	0,006	persona	0,006	dar	0,004	haber	0,006	legar	0,005
7	haber	0,005	chile	0,005	dar	0,004	trabajar	0,006	decir	0,005
8	servicio	0,004	ano	0,005	parte	0,004	llevar	0,005	hacer	0,005
9	comisaria	0,004	ano	0,004	ver	0,004	general	0,005	dejar	0,004
10	decir	0,004	haber	0,005	llevar	0,004	haber	0,005	funcionario	0,004
11	deber	0,004	institucion	0,004	llevar	0,004	general	0,004	institucion	0,004
12	auto	0,004	llevar	0,004	llevar	0,004	esperar	0,004	esperar	0,004
13	vehiculos	0,004	detener	0,004	llamar	0,004	esperar	0,004	esperar	0,004
14	querer	0,004	parte	0,004	dar	0,004	estar	0,004	chile	0,004
15	denunciar	0,004	ser	0,004	llevar	0,003	publicar	0,003	dejar	0,004
16	él	0,004	reclamo	0,004	dos	0,003	legar	0,004	habia	0,004
17	persona	0,003	pasar	0,004	ley	0,003	comuna	0,004	tomar	0,004
18	calle	0,003	casa	0,004	querer	0,003	ex	0,003	estacionar	0,004
19	oficial	0,003	deber	0,004	ano	0,003	persona	0,002	institucion	0,004
20	dejar	0,003	nada	0,004	funcionario	0,003	hacer	0,002	reclamo	0,003
21	dia	0,003	decir	0,004	dejar	0,003	el	0,002	habia	0,003
22	ser	0,003	comisaria	0,003	familia	0,003	hoy	0,002	despues	0,003
23	tener	0,003	funcionario	0,003	chile	0,003	dia	0,002	motivo	0,003
24	realizar	0,003	esperar	0,003	salir	0,003	alrededor	0,003	querer	0,003
25	parte	0,003	nuestro	0,003	preguntar	0,003	sun	0,003	querer	0,003
26	personal	0,003	hijo	0,003	él	0,003	dia	0,003	existir	0,003
27	ver	0,003	calle	0,003	parecer	0,003	parecer	0,003	dejar	0,003
28	dos	0,003	realizar	0,003	deber	0,003	deber	0,003	hoy	0,003
29	gracia	0,003	servicio	0,003	vehicular	0,003	vehicular	0,003	mayor	0,002
30	ni	0,003	trabajo	0,003	momento	0,003	momento	0,003	deber	0,002
									nombre	0,002
									persona	0,003
									mal	0,003
									deber	0,002
									preguntar	0,003
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002
									preguntar	0,003
									deber	0,002

Resultados Análisis LDA Carabineros de Chile 2007-2011 A

	Tópico 1		Tópico 2		Tópico 3		Tópico 4		Tópico 5		Tópico 6		Tópico 7		Tópico 8		Tópico 9		Tópico 10	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	no	0.043	no	0.022	no	0.035	no	0.015	no	0.024	no	0.037	no	0.040	no	0.035	no	0.028	no	0.028
2	carabinero	0.025	parte	0.009	carabinero	0.024	carabinero	0.011	carabinero	0.020	carabinero	0.018	carabinero	0.026	carabinero	0.020	carabinero	0.023	carabinero	0.025
3	dar	0.011	general	0.009	hacer	0.007	parte	0.008	chile	0.006	hacer	0.007	dar	0.009	pasar	0.009	parte	0.009	decir	0.007
4	decir	0.007	carabinero	0.008	parte	0.006	decir	0.007	persona	0.005	parte	0.006	hacer	0.008	dar	0.006	hacer	0.009	hacer	0.006
5	hacer	0.006	hacer	0.005	llamar	0.006	hacer	0.005	decir	0.005	dar	0.005	parte	0.008	haber	0.006	llamar	0.008	detener	0.006
6	día	0.006	ley	0.004	hijo	0.005	habia	0.005	haber	0.005	dejar	0.005	dejar	0.007	día	0.006	llegar	0.006	parte	0.005
7	comisaria	0.006	dar	0.004	dar	0.004	vehicular	0.005	hacer	0.005	día	0.005	hacer	0.006	hacer	0.005	dar	0.006	comisaria	0.005
8	haber	0.005	día	0.004	nada	0.004	pasar	0.005	dar	0.004	habia	0.005	decir	0.006	llegar	0.005	detener	0.006	día	0.005
9	pasar	0.005	detener	0.004	dejar	0.004	hijo	0.004	nuestro	0.004	pasar	0.005	funcionario	0.006	casa	0.005	comisaria	0.006	haber	0.004
10	nada	0.005	luz	0.003	decir	0.004	él	0.004	hijo	0.004	haber	0.005	persona	0.006	deber	0.005	deber	0.005	dar	0.004
11	llamar	0.005	momento	0.003	chile	0.004	ano	0.004	llamar	0.004	llamar	0.005	haber	0.005	persona	0.005	haber	0.005	llamar	0.004
12	hijo	0.004	ser	0.003	deber	0.004	llamar	0.004	día	0.004	auto	0.004	ano	0.005	funcionario	0.005	funcionario	0.005	pasar	0.004
13	chile	0.004	hijo	0.003	indicar	0.004	esperar	0.004	estar	0.003	detener	0.004	ley	0.004	querer	0.004	querer	0.004	habia	0.004
14	querer	0.004	haber	0.003	día	0.004	haber	0.004	despues	0.003	nada	0.004	estar	0.004	decir	0.004	habia	0.004	chile	0.004
15	nuestro	0.004	encontrar	0.003	haber	0.004	detener	0.003	querer	0.003	llevar	0.004	llevar	0.004	lugar	0.004	llevar	0.004	general	0.004
16	hora	0.004	sacar	0.003	persona	0.004	policia	0.003	llevar	0.003	llegar	0.004	deber	0.004	estar	0.004	decir	0.004	funcionario	0.004
17	institucion	0.004	calle	0.003	querer	0.003	institucion	0.003	podar	0.003	comisaria	0.004	llegar	0.004	ley	0.004	estar	0.004	lugar	0.003
18	reclamo	0.004	estar	0.003	comisaria	0.003	dejar	0.003	reclamo	0.003	persona	0.004	reclamo	0.004	esperar	0.004	pasar	0.004	vehicular	0.003
19	calle	0.004	chile	0.003	reclamo	0.003	querer	0.003	tener	0.003	decir	0.004	vehicular	0.004	habia	0.004	día	0.004	institucion	0.003
20	llevar	0.004	fecha	0.002	nuestro	0.003	llevar	0.003	contra	0.003	chile	0.004	chile	0.003	llevar	0.003	realizar	0.004	calle	0.003
21	deber	0.003	tomar	0.002	ano	0.003	tomar	0.003	parte	0.003	deber	0.004	ser	0.003	ser	0.003	calle	0.003	deber	0.003
22	denunciar	0.003	deber	0.002	tomar	0.003	dos	0.003	comisaria	0.003	tomar	0.004	detener	0.003	ano	0.003	ano	0.003	persona	0.003
23	hoy	0.003	ano	0.002	pasar	0.003	nuestro	0.003	esperar	0.003	trabajo	0.003	trabajo	0.003	hiz	0.003	policia	0.003	encontrar	0.003
24	llegar	0.003	orden	0.002	casa	0.003	conitar	0.003	estadio	0.002	mayor	0.003	querer	0.003	comisaria	0.003	hijo	0.003	tener	0.003
25	ano	0.003	pregunta	0.002	hora	0.003	funcionario	0.003	trabajar	0.002	ser	0.003	esperar	0.003	momento	0.003	tener	0.003	tenia	0.003
26	estar	0.003	significante	0.002	denunciar	0.003	día	0.002	ser	0.002	estar	0.003	hijo	0.003	esperar	0.003	pagar	0.003	denunciar	0.003
27	dejar	0.003	persona	0.002	institucion	0.003	reclamo	0.002	alto	0.002	ley	0.003	habia	0.003	quedar	0.003	reclamo	0.003	transitar	0.003
28	casa	0.003	pasar	0.002	seguir	0.003	tener	0.002	momento	0.002	funcionario	0.003	momento	0.003	commua	0.003	situacion	0.003	ano	0.003
29	persona	0.003	sr	0.002	ley	0.003	persona	0.002	perro	0.002	tener	0.003	nada	0.003	nada	0.003	pedir	0.003	nada	0.003
30	lugar	0.003	auto	0.002	detener	0.003	estacionar	0.002	detener	0.002	encontrar	0.003	pedir	0.003	nuestro	0.003	nuestro	0.003	esperar	0.003

Cuadro 5.35: Parte I de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile entre los años 2007 y 2011.

Resultados Análisis LDA Carabineros de Chile 2007-2011 B

	Tópico 11		Tópico 12		Tópico 13		Tópico 14		Tópico 15		Tópico 16		Tópico 17		Tópico 18		Tópico 19		Tópico 20	
	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P	Palabra	P
1	carabinero	0,029	no	0,024	no	0,018	no	0,031	no	0,032	carabinero	0,033	no	0,032	no	0,020	no	0,036	no	0,026
2	no	0,022	carabinero	0,012	carabinero	0,011	carabinero	0,015	carabinero	0,014	no	0,028	carabinero	0,020	carabinero	0,009	carabinero	0,031	carabinero	0,020
3	hacer	0,008	nuestro	0,007	parte	0,007	estacionar	0,006	parte	0,008	haber	0,007	decir	0,008	hacer	0,008	decir	0,008	llamar	0,008
4	haber	0,007	parte	0,005	ley	0,007	detener	0,006	dar	0,007	llamar	0,007	parte	0,007	dar	0,007	parte	0,007	persona	0,007
5	llegar	0,007	llamar	0,005	madre	0,004	pasar	0,005	decir	0,007	comisaria	0,005	hacer	0,008	detener	0,006	dia	0,006	decir	0,007
6	dar	0,006	persona	0,004	decir	0,004	parte	0,005	dia	0,007	llegar	0,005	pasar	0,006	funcionario	0,006	hacer	0,006	haber	0,006
7	dia	0,006	chile	0,004	llamar	0,004	hacer	0,005	hacer	0,006	hacer	0,005	llegar	0,006	institucion	0,005	pasar	0,006	pasar	0,005
8	decir	0,005	hacer	0,004	ano	0,004	decir	0,005	comisaria	0,006	casa	0,005	detener	0,006	aro	0,005	dar	0,005	dia	0,005
9	parte	0,005	pasar	0,004	comisaria	0,004	querer	0,005	pasar	0,005	pasar	0,004	dar	0,005	deber	0,004	detener	0,005	dar	0,005
10	pasar	0,005	trabajo	0,003	persona	0,003	llegar	0,005	ley	0,004	parte	0,004	dar	0,005	hijo	0,004	persona	0,005	parte	0,004
11	hijo	0,005	dar	0,003	dia	0,003	dia	0,005	tener	0,004	dia	0,004	comisaria	0,005	dia	0,004	llevar	0,005	vehicular	0,004
12	chile	0,004	encontrar	0,003	chile	0,003	estar	0,004	auto	0,004	decir	0,004	calle	0,004	querer	0,004	comisaria	0,005	hacer	0,004
13	momento	0,004	calle	0,003	pasar	0,003	dar	0,004	haber	0,004	ano	0,004	dia	0,004	decir	0,004	institucion	0,005	tener	0,004
14	estar	0,004	llegar	0,003	hacer	0,003	funcionario	0,004	estar	0,004	dar	0,004	nada	0,004	haber	0,003	llamar	0,004	lugar	0,004
15	persona	0,004	decir	0,003	querer	0,003	calle	0,004	persona	0,004	persona	0,004	vehicular	0,004	chile	0,003	deber	0,004	detener	0,004
16	llamar	0,004	deber	0,003	dar	0,003	ano	0,004	llamar	0,004	estar	0,004	estar	0,004	pasar	0,003	nada	0,004	deber	0,004
17	hora	0,004	estar	0,003	nuestro	0,002	lugar	0,004	esperar	0,004	hijo	0,004	haber	0,004	parte	0,003	vehicular	0,004	despues	0,004
18	funcionario	0,004	dia	0,003	joya	0,003	chile	0,004	institucion	0,004	habia	0,004	esperar	0,004	contra	0,003	q	0,004	hora	0,004
19	habia	0,004	ser	0,003	haber	0,002	casa	0,004	nuestro	0,003	chile	0,004	querer	0,003	auto	0,003	aro	0,004	ser	0,004
20	nada	0,004	tomar	0,002	cabo	0,002	persona	0,003	detener	0,003	detener	0,004	deber	0,003	ley	0,003	dejar	0,004	llevar	0,004
21	comisaria	0,004	comma	0,002	llegar	0,002	comisaria	0,003	nada	0,003	querer	0,003	hijo	0,003	general	0,003	calle	0,004	nada	0,004
22	ver	0,004	quedar	0,002	detener	0,002	institucion	0,003	lugar	0,003	calle	0,003	contra	0,003	llevar	0,003	tener	0,004	derecho	0,003
23	calle	0,003	ver	0,002	seguir	0,002	nada	0,003	casa	0,003	lugar	0,003	infraccion	0,003	publicar	0,003	encontrar	0,004	llegar	0,003
24	deber	0,003	casa	0,002	dejar	0,002	tomar	0,003	deber	0,003	reclamo	0,003	llevar	0,003	derecho	0,003	ser	0,003	hoy	0,003
25	ano	0,003	comisaria	0,002	tenia	0,002	reclamo	0,003	ano	0,003	deber	0,003	momento	0,003	reclamo	0,003	casa	0,003	estar	0,003
26	detener	0,003	reclamo	0,002	actualizar	0,002	dos	0,003	funcionario	0,003	dejar	0,003	lugar	0,003	comisaria	0,003	momento	0,003	tomar	0,003
27	encontrar	0,003	esperar	0,002	nombre	0,002	pedir	0,003	dejar	0,003	dejar	0,003	situacion	0,003	solicitar	0,003	llegar	0,003	casa	0,003
28	querer	0,003	haber	0,002	momento	0,002	haber	0,003	encontrar	0,003	dos	0,003	reclamo	0,003	habia	0,003	hora	0,003	nuestro	0,003
29	dos	0,003	situacion	0,002	llevar	0,002	vehicular	0,003	chile	0,003	padre	0,003	mal	0,003	llamar	0,003	funcionario	0,003	comisaria	0,003
30	institucion	0,003	pedir	0,002	procedimiento	0,002	quedar	0,003	dos	0,003	estacionar	0,003	ser	0,003	encontrar	0,003	esperar	0,003	encontrar	0,003

Cuadro 5.36: Parte II de los resultados obtenidos del análisis LDA ejecutado sobre los Reclamos realizados contra Carabineros de Chile entre los años 2007 y 2011.

Realizando un análisis, basado en la revisión y en la lectura manual de los tópicos y de las palabras generadas en cada período, la o las frases que pueden generarse y sintetizar el sentido que subyace en estos elementos son las siguientes:

■ **Período 2007**

“Existen problemas con respecto a los carabineros, la aplicación que realizan de las leyes, y también, en buena medida con respecto a multas por temas automovilísticos y de tránsito.”

■ **Período 2008**

“Existen problemas con respecto a los carabineros, y a partes tránsito.”

■ **Período 2009**

“Existen problemas con respecto a los carabineros, las comisarías, la aplicación de la ley y multas de carácter vehicular y de tránsito.”

■ **Período 2010**

“Existen problemas con respecto a la institución, los carabineros y la delincuencia.”

■ **Período 2011**

“Existen problemas con los carabineros, la aplicación que realizan de la ley, las multas y las comisarías.”

■ **Período 2007-2011**

“Existen problemas con los carabineros, las multas y las comisarías.”

Capítulo 6

Conclusiones

En el siguiente capítulo se abordará lo referente a las conclusiones propiamente tales del experimento realizado.

Se contrastará lo realizado con la consecución de los objetivos, tanto específicos, como el general, así como también se abordará la verificación de la hipótesis planteada.

Además, sin ánimo de concluir algo específico, pero sí de aportar algunas reflexiones adicionales, se comentarán ciertas apreciaciones y posibles causas y efectos sobre los resultados obtenidos, así como también, en base a la experiencia generada en el transcurso de la realización de esta memoria, se propondrán ideas acerca de posibles usos y aplicaciones que podrían hacerse con este tipo de técnicas.

6.1. Objetivos Específicos

Respecto a la consecución de los objetivos específicos, se puede decir que estos fueron logrados a cabalidad.

Se confeccionó un marco conceptual que aportase explicaciones y aclaraciones sobre distintos aspectos, técnicas, herramientas y recursos, que fueron utilizados, ya sea en la descripción del entorno en donde se iba a realizar el experimento, en el mismo experimento, y en los resultados obtenidos de este.

Asimismo, se confeccionó una aplicación informática que permitió lograr no sólo el objetivo específico que guardaba relación con ella, sino también, permitió lograr la extracción, pre procesamiento, y procesamiento de los datos de interés del sitio, así como los análisis LDA sobre estos, y la redacción de frases a partir de los resultados obtenidos.

6.2. Objetivo General

En base a las frases elaboradas a partir de los resultados obtenidos mediante los análisis LDA realizados sobre los reclamos contra las instituciones DOUC y Carabineros de Chile, y apelando a que no hay pérdida de generalidad, dado a que esto puede ser replicado análogamente para cualquier institución; y sumándolo al conocimiento teórico que se posee en torno al funcionamiento de LDA, hace posible afirmar que las técnicas de WOM aquí aplicadas permiten determinar las tendencias generales de los reclamos contra las distintas instituciones abordadas.

Por tanto, también se ha conseguido cumplir con el objetivo general planteado para esta memoria.

6.3. Hipótesis

Ahora, si bien resulta de cierta manera evidente que al cumplir el objetivo general se puede corroborar la hipótesis planteada, esto no resulta tan así, ya que al surgir las tendencias de los reclamos de un proceso de lectura e interpretación de las palabras que conforman los tópicos generados por los análisis LDA, abre la puerta a que las tendencias identificadas e inferidas se encuentren sujetas a los paradigmas, las experiencias, los intereses, u otros factores, que el actor que realiza las interpretaciones posea. Por lo tanto, el objeto específico al que apuntan las tendencias resulta no ser tan objetivo, ni puntualizado, sino más bien subjetivo y relativo al actor.

En otras palabras, la interpretación y la conclusión sobre los resultados obtenidos mediante los análisis LDA puede variar en función de aspectos inherentes al actor que realiza este ejercicio. Esto es, por ejemplo: la interpretación y la conclusión sobre los resultados obtenidos que llegue a realizar un funcionario del SERNAC, no será la misma a la que llegue un gerente del área de Ventas o de Marketing de alguna institución privada, ya que ambos poseen una serie de factores que pueden hacerles valorar de distinta manera las palabras que constituyen los resultados de los análisis de LDA. Cuestión que podría resultar similar si hubiese otro actor que fuese un consumidor final.

Ante ello, la conclusión final que se puede llegar a hacer sobre la hipótesis planteada, es que si bien se corrobora, esto no es absoluto, ya que se hace necesario hacer observaciones en torno a la objetividad y a la especificidad de las tendencias identificadas y logradas mediante este tipo de técnicas de WOM.

En base a esto, quizás el mayor aporte que se podría hacer con este tipo de análisis, es el hecho de haber logrado sintetizar, en espacios no excesivamente extensos y humanamente comprensibles, información que implícitamente representa grandes volúmenes de datos. Luego, las interpretaciones y conclusiones al respecto, así como la generación de conocimiento, quedan a juicio y voluntad de quien tome y revise dichas informaciones generadas.

6.4. Comentarios y Apreciaciones

6.4.1. Resultados

En relación a los resultados obtenidos, llama la atención el bajo número de respuestas o de soluciones que presentan tanto los reclamos como las recomendaciones publicados en el sitio web.

Esta situación se podría deber a varios factores, como por ejemplo, el hecho de que los reclamos o las recomendaciones publicados en el sitio no son de carácter vinculante.

Esto quiere decir, que no existe una obligación de responderlos por parte de las instituciones aludidas, o también, puede deberse a que existen otros medios por los cuales las instituciones en cuestión si responden, tales como las líneas de atención a clientes, o las denuncias a través del SERNAC, en el caso de las instituciones pertenecientes al sector privado; o las OIRS¹ y la CDC² para el caso de las instituciones públicas.

Por ende, el responder a través del sitio de Reclamos.cl, no resulte ser algo gravitante, desde el punto de vista de las instituciones.

Pero de todas maneras, si se considera que Reclamos.cl es el 318 avo sitio más visitado del país [1], el hecho de que las instituciones aludidas no respondan puede generar la impresión a los visitantes y a los usuarios del sitio, de que estas no toman en cuenta lo que han publicado hacia ellas, y por ende, podría generar una brecha de desconfianza entre estos últimos y las instituciones, cuestión que desde el punto de vista del marketing y de las ventas puede resultar negativo, tanto por la fidelización de los clientes, como por la compra y venta de productos o servicios, respectivamente.

6.4.2. Posibles Usos y Aplicaciones

Como se concluía respecto a la hipótesis, a pesar de que las tendencias que se puedan identificar mediante este tipo de análisis y técnicas no son puntualmente objetivas y dependen de quien realice la interpretación de los resultados, de todas maneras el aporte que puede generar este tipo de aplicaciones de WOM no deja de ser menor.

Por ejemplo, si se considera que la confección de la aplicación informática, la descarga de las páginas web, la extracción y la tabulación de los datos de interés que estas poseían, y su posterior procesamiento y análisis, no llevo más de dos meses de trabajo, e involucró a una sola persona con su computador personal, es simple poder visualizar que este tipo de iniciativas no debiese demandar de una gran inversión en tiempos, recursos e insumos, ya sean estos de tipo humano como también de tipo material.

Asimismo, abstrayéndose de los reclamos, y ampliándose a abarcar cualquier tipo de artículo, publicación, comentario u opinión que pueda encontrarse al interior de una página web, las aplicaciones de este tipo de técnicas pueden llegar a ser bastante diversas y variadas.

A modo de ejercitar la imaginación, por ejemplo, podría facilitarse la revisión histórica de

¹Oficina de Informaciones, Reclamos y Sugerencias

²Comisión Defensora Ciudadana

la jurisprudencia que presenta cualquier conjunto de casos que sean de una materia en particular, siempre y cuando los fallos y las resoluciones de estos se encontrasen en páginas web o en algún medio digital en donde una aplicación informática pudiese actuar, tal que con ello, sea posible identificar sus tendencias y concluir hacia que aspectos legales apuntan a concentrarse.

Otra aplicación podría ser el análisis de las ofertas que realiza la competencia dentro de un mercado en particular, con tal de identificar su tendencia, y en consecuencia, detectar oportunidades de negocios en aquellas áreas o sectores del mercado en donde no se han presentado, o propuesto, servicios y productos.

También podrían aplicarse en portales como Chile Proveedor, e identificar la tendencia histórica de las licitaciones realizadas dentro de ciertos períodos de tiempo.

Por otra parte, se podría aplicar en medios tipo redes sociales virtuales y con ello identificar los tópicos y palabras más recurrentes también dentro de períodos de tiempo, dentro de las cuales pueden surgir nombres de productos y servicios, pudiendo con ello identificar el grado de popularidad o recurrencia que poseen estos dentro de las conversaciones que se articulan dentro de estos medios.

Bibliografía

- [1] ALEXA INTERNET, INC. Alexa the Web Information Company, [en línea]. <<http://www.alexa.com/siteinfo/reclamos.cl>>. [consulta: 4 de Mayo de 2012].
- [2] GRIGORIS ANTONIOU and FRANK VAN HARMELEN. *A Semantic Web Primer*. Massachusetts Institute of Technology, MIT, 2004.
- [3] MIKHAIL BAUTIN, LOHIT VIJAYARENU, and STEVEN SKIENA. International Sentiment Analysis for News and Blogs. *Association for the Advancement of Arti[U+FB01]cial Intelligence*, pages 19–26, 2008.
- [4] TIM BERNERS-LEE, ROBERT CAILLIAU, ARI LUOTONEN, HENRIK FRYSTYK NIELSEN, and ARTHUR SECRET. The World-Wide Web. *Communications of ACM*, 37(8):76–82, 1994.
- [5] HAJI BINALI, VIDYYASAGAR POTDAR, and CHEN WU. A State Of The Art Opinion Mining And Its Application Domains. *IEEE Computer Society*, pages 1–6, 2009.
- [6] DAVID M. BLEI. Topic Models, [en línea]. <http://videlectures.net/mlss09uk_blei_tm/>. [consulta: 1 de Marzo de 2012].
- [7] DAVID M. BLEI, ANDREW Y. NG, and MICHAEL I. JORDAN. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(2003):993–1022, 2003.
- [8] ROBERT CERCÓS BROWNELL. Diseño y Construcción de un Web Warehouse para Almacenar Información extraída a partir de Datos originados en la Web. Memoria para Optar al Título de Ingeniero Civil Industrial, Universidad de Chile, Santiago, Chile, 2008.
- [9] BEN CALDWELL, MICHAEL COOPER, LORETTA GUARINO REID, and GREGG VANDERHEIDEN. Pautas de Accesibilidad de Contenido Web 2.0. Technical report, W3C, 2008. <<http://www.codexempla.org/traducciones/pautas-accesibilidad-contenido-web-2.0.htm>>.
- [10] CLIPS RESEARCH CENTER. Pattern — CLiPS, [en línea]. <<http://www.clips.ua.ac.be/pages/pattern>>. [consulta: Noviembre de 2011].
- [11] CLIPS RESEARCH CENTER. pattern.en. Technical report, CLiPS Research Center. <<http://www.clips.ua.ac.be/pages/pattern-en>>.
- [12] CLIPS RESEARCH CENTER. pattern.vector. Technical report, CLiPS Research Center. <<http://www.clips.ua.ac.be/pages/pattern-vector>>.
- [13] CLIPS RESEARCH CENTER. pattern.web. Technical report, CLiPS Research Center. <<http://www.clips.ua.ac.be/pages/pattern-web>>.

- [14] SOUMEN CHAKRABARTI, BYRON E. DOM, S. RAVI KUMAR, PRABHAKAR RAGHAVAN, SRIDHAR RAJAGOPALAN, ANDREW TOMKINS, DAVID GIBSON, and JON KLEINBERG. Mining the Web's Link Structure. *IEEE Computer*, 32(8):60–67, Agosto 1999.
- [15] EDGAR FRANK CODD. Relational Database: A Practical Foundation for Productivity. *Communications of ACM*, 25(2):109–117, 1982.
- [16] B.J. COX and A.J. NOVOBILSKI. *Object-Oriented Programming: An Evolutionary Approach*. Addison-Wesley, 2 edition, 1991.
- [17] EUGENIO MARTÍNEZ CÁMARA, MARÍA TERESA MARTÍN VALDIVIA, and LUIS ALFONSO UREÑA LÓPEZ. Opinion Classification Techniques Applied to a Spanish Corpus. 6716:169–176, 2011.
- [18] THOMAS H. DAVENPORT and LAWRENCE PRUSAK. *Working Knowledge: How Organizations Manage What They Know*. Harvard Business School Press, 1997.
- [19] THE FREE ONLINE DICTIONARY and THESAURUS AND ENCYCLOPEDIA. Website - Definition of Website, [en línea]. <<http://www.thefreedictionary.com/Website>>. [consulta: 04 de Mayo de 2012].
- [20] RAÚL GONZÁLEZ DUQUE. *Python Para Todos*.
- [21] RAMEZ ELMASRI and SHAMKANT B. NAVATHE. *Sistemas de Bases de Datos, Conceptos Fundamentales*. Addison-Wesley Iberoamericana, 2 edition, 2000.
- [22] USAMA FAYYAD, GREGORY PIATETSKY, and PADHRAIC SMYTH. The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of ACM*, 39(11):27–34, Noviembre 1996.
- [23] DJANGO SOFTWARE FOUNDATION. Django — The Web framework for perfectionists with deadlines, [en línea]. <<http://www.djangoproject.com>>.
- [24] DJANGO SOFTWARE FOUNDATION. *Django Documentation, Release 1.5*. Django Software Foundation, Mayo 2012.
- [25] FREELING. home - FreeLing Home Page, [en línea]. <<http://nlp.lsi.upc.edu/freeling/>>. [consulta: Diciembre de 2011].
- [26] FREELING. *FreeLing User Manual, 3.0*. FreeLing, Mayo 2012.
- [27] S. GALLAUGHER, J. y RAMANATHAN. Choosing a Client/Server Architecture: A Comparison of Two-Tier and Three-Tier Systems. *Information Systems Management Magazine*, 13(2):7–13, 1996.
- [28] POSTGRES GLOBAL DEVELOPMENT GROUP. PostgreSQL: Welcome, [en línea]. <<http://www.postgresql.org>>. [consulta: 9 de Octubre de 2011].
- [29] HÉCTOR IGNACIO ÁLVAREZ GÓMEZ. Detección de Miembros Clave en una Comunidad Virtual de Práctica Mediante Análisis de Redes Sociales y Minería de Datos Avanzada. Tesis para Optar al Grado de Magíster en Gestión de Operaciones, Memoria para Optar al Título de Ingeniero Civil Industrial, Universidad de Chile, Santiago, Chile, Octubre 2010.
- [30] JIAWEI HAN and MICHELINE KAMBER. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 2 edition, 2006.

- [31] DAVID HAND, HEIKKI MANNILA, and PADHRAIC SMYTH. *Principles of Data Mining*. Massachusetts Institute of Technology, MIT, 2001.
- [32] IAN HICKSON. HTML5. Technical report, W3C, 03 2012. Working Draft.
- [33] PHILIPPE LE HÉGARET, RAY WHITMER, and LAUREN WOOD. Document Object Model (DOM). Technical report, W3C, 06 2009. <<http://www.w3.org/DOM/>>.
- [34] MATTHEW L. JOCKERS. The LDA Buffet is Now Open; or, Latent Dirichlet Allocation for English Majors, [en línea]. <<https://www.stanford.edu/~mjockers/cgi-bin/drupal/node/61>>. [consulta: 1 de Marzo de 2012].
- [35] VÍCTOR LEONARDO REBOLLEDO LORCA. Plataforma para la Extracción y Almacenamiento del Conocimiento Extraído de los Web Data. Tesis para Optar al Grado de Magíster en Gestión de Operaciones, Memoria para Optar al Título de Ingeniero Civil Industrial, Universidad de Chile, Santiago, Chile, 2009.
- [36] ZDRAVKO MARKOV and DANIEL T. LAROSE. *Data Mining The Web, Uncovering Patterns in Web Content, Structure, and Usage*. John Wiley and Sons, Inc, 2007.
- [37] SANKAR K. PAL, VARUN TALWAR, and PABITRA MITRA. Web Mining in Soft Computing Framework: Relevance, State of the Art and Future Directions. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 13(5):1163–1177, Septiembre 2002.
- [38] BO PANG and LILLIAN LEE. Opinion Mining and Sentiment Analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, Enero 2008.
- [39] DAVE RAGGETT, ARNAUD LE HORS, and IAN JACOBS. Especificación HTML 4.01. Technical report, W3C, 12 1999. <<http://html.conclase.net/w3c/html401-es/cover.html>>.
- [40] RECLAMOS.CL. Quiénes Somos, [en línea]. <http://www.reclamos.cl/quienes_somos>. [consulta: 15 de Diciembre de 2011].
- [41] EDUARDO ANGEL MERLO REPETTO. Identificación de Comunidades de Copia en Instituciones Educativas Mediante el Análisis de Redes Sociales sobre Documentos Digitales. Memoria para Optar al Título de Ingeniero Civil Industrial, Universidad de Chile, Santiago, Chile, Septiembre 2010.
- [42] GUIDO VAN ROSSUM. Python HOWTOs, Release 2.7.3. Technical report, Python Software Foundation, Enero 2012. <<http://docs.python.org/howto/>>.
- [43] GUIDO VAN ROSSUM. Python Setup and Usage, Release 2.7.3. Technical report, Python Software Foundation, Enero 2012. <<http://docs.python.org/using/>>.
- [44] GUIDO VAN ROSSUM. Python Tutorial, Release 2.7.3. Technical report, Python Software Foundation, Enero 2012. <<http://docs.python.org/tutorial/>>.
- [45] GUIDO VAN ROSSUM. The Python Language Reference, Release 2.7.3. Technical report, Python Software Foundation, Enero 2012. <<http://docs.python.org/reference/>>.
- [46] GUIDO VAN ROSSUM. The Python Library Reference, Release 2.7.3. Technical report, Python Software Foundation, Enero 2012. <<http://docs.python.org/library/>>.
- [47] SEBASTIÁN A. RÍOS. Semantic Web Mining Techniques to Study Enhancements of Virtual Communities, 2009. FONDECYT, Número 11090188.

- [48] MANU SPORNY and SHANE MC CARROM. HTML+RDFa 1.1. Technical report, W3C, 06 2010. Working Draft, <<http://www.w3.org/TR/2010/WD-rdfa-in-html-20100624/>>.
- [49] JUAN D. VELÁSQUEZ, ALEJANDRO BASSI, HIROSHI YASUDA, and TERUMASA AOKI. Mining Web Data to Create Online Navigation Recommendations. *Procs. 4th IEEE Int. Conf. on Data Mining*, pages 551–554, Noviembre 2004.
- [50] JUAN D. VELÁSQUEZ and VASILE PALADE. *Adaptive Web site: A Knowledge Extraction from Web Data Approach*. IOS Press, 2008.
- [51] W3C. HTML, [en línea]. <<http://www.w3.org/community/webed/wiki/HTML>>. [consulta: 3 de Marzo de 2012].
- [52] W3SCHOOLS. HTML Tutorial, [en línea]. <<http://www.w3schools.com/html/>>. [consulta: 15 de Diciembre de 2011].
- [53] W3SCHOOLS. JavaScript and HTML DOM Reference, [en línea]. <<http://www.w3schools.com/jsref/>>. [consulta: 20 de Diciembre de 2011].
- [54] W3SCHOOLS. SQL DELETE Statement, [en línea]. <http://www.w3schools.com/sql/sql_delete.asp>. [consulta: 15 de Enero de 2012].
- [55] W3SCHOOLS. SQL INSERT INTO Statement, [en línea]. <http://www.w3schools.com/sql/sql_insert.asp>. [consulta: 15 de Enero de 2012].
- [56] W3SCHOOLS. SQL Joins, [en línea]. <http://www.w3schools.com/sql/sql_join.asp>. [consulta: 15 de Enero de 2012].
- [57] W3SCHOOLS. SQL LEFT JOIN Keyword, [en línea]. <http://www.w3schools.com/sql/sql_join_left.asp>. [consulta: 15 de Enero de 2012].
- [58] W3SCHOOLS. SQL RIGHT JOIN Keyword, [en línea]. <http://www.w3schools.com/sql/sql_join_right.asp>. [consulta: 15 de Enero de 2012].
- [59] W3SCHOOLS. SQL SELECT Statement, [en línea]. <http://www.w3schools.com/sql/sql_select.asp>. [consulta: 15 de Enero de 2012].
- [60] W3SCHOOLS. SQL Tutorial, [en línea]. <<http://www.w3schools.com/sql/>>. [consulta: 15 de Enero de 2012].
- [61] W3SCHOOLS. SQL UPDATE Statement, [en línea]. <http://www.w3schools.com/sql/sql_update.asp>. [consulta: 15 de Enero de 2012].
- [62] RADIM ŘEHŮŘEK. API Reference - gensim. Technical report. <<http://radimrehurek.com/gensim/apiref.html>>.
- [63] RADIM ŘEHŮŘEK. Gensim - Topic Modelling for Humans - gensim, [en línea]. <<http://radimrehurek.com/gensim/>>. [consulta: Febrero de 2012].

Apéndice A

Consultas SQL de los Resultados

A continuación se adjuntan las instrucciones en SQL de las consultas que permitieron extraer los datos mostrados en el capítulo de Resultados.

A.1. Total de Instituciones por Sub Rubro y Rubro

```
SELECT reclamation_industry.name,  
       reclamation_subindustry.name,  
       COUNT(reclamation_establishment.*) AS total_establishments  
FROM reclamation_establishment  
LEFT JOIN reclamation_industry  
ON reclamation_establishment.industry_id = reclamation_industry.id  
LEFT JOIN reclamation_subindustry  
ON reclamation_establishment.subindustry_id = reclamation_subindustry.id  
WHERE reclamation_industry.id IN (4, 5)  
GROUP BY reclamation_subindustry.name, reclamation_industry.name  
ORDER BY reclamation_industry.name, reclamation_subindustry.name;
```

A.2. Total de Reclamos por Institución

```
SELECT reclamation_establishment.id AS establishment_id,  
       COUNT(reclamation_article.*) AS total_reclamos  
FROM reclamation_establishment  
LEFT JOIN reclamation_article  
ON reclamation_establishment.id = reclamation_article.establishment_id  
WHERE reclamation_article.type = True GROUP BY reclamation_establishment.id;
```

A.3. Total de Recomendaciones por Institución

```
SELECT reamoscl_establishment.id AS establishment_id,
       COUNT(reamoscl_article.*) AS total_recomendaciones
FROM reamoscl_establishment
LEFT JOIN reamoscl_article
ON reamoscl_establishment.id = reamoscl_article.establishment_id
WHERE reamoscl_article.type = False
GROUP BY reamoscl_establishment.id;
```

A.4. Total de Instituciones, Reclamos y Recomendaciones por Rubro y Sub Rubro

```
SELECT reamoscl_industry.name AS industry,
       CASE WHEN reamoscl_subindustry.name IS NULL THEN 'ninguno'
       ELSE reamoscl_subindustry.name
       END AS subindustry,
       COUNT(reamoscl_establishment.id) AS total_establishments,
       CASE WHEN SUM(reamos.total_reamos) IS NULL THEN 0
       ELSE SUM(reamos.total_reamos)
       END AS total_reamos,
       CASE WHEN SUM(recomendaciones.total_recomendaciones) IS NULL THEN 0
       ELSE SUM(recomendaciones.total_recomendaciones)
       END AS total_recomendaciones,
       SUM(reamos.total_reamos)+SUM(recomendaciones.total_recomendaciones)
       AS total_articles
FROM reamoscl_establishment
LEFT JOIN (
  SELECT reamoscl_article.establishment_id AS establishment_id,
         COUNT(reamoscl_article.*) AS total_reamos
  FROM reamoscl_article
  WHERE reamoscl_article.type = True
  GROUP BY reamoscl_article.establishment_id
) AS reamos
ON reamoscl_establishment.id = reamos.establishment_id
LEFT JOIN (
  SELECT reamoscl_article.establishment_id AS establishment_id,
         COUNT(reamoscl_article.*) AS total_recomendaciones
  FROM reamoscl_article
  WHERE reamoscl_article.type = False
  GROUP BY reamoscl_article.establishment_id
) AS recomendaciones
ON reamoscl_establishment.id = recomendaciones.establishment_id
LEFT JOIN reamoscl_industry
ON reamoscl_establishment.industry_id = reamoscl_industry.id
LEFT JOIN reamoscl_subindustry
```

```

ON reamoscl_establishment.subindustry_id = reamoscl_subindustry.id
WHERE reamoscl_industry.id IN (4, 5)
GROUP BY reamoscl_subindustry.name, reamoscl_industry.name
ORDER BY reamoscl_industry.name, reamoscl_subindustry.name;

```

A.5. Instituciones que no Pertenecen a Ningún Sub Rubro

```

SELECT *
FROM reamoscl_establishment
WHERE reamoscl_establishment.subindustry_id IS NULL
AND reamoscl_establishment.industry_id IN (4, 5);

```

A.6. Detalles de los Reclamos y de las Recomendaciones, de las Instituciones por Rubro

```

SELECT reamoscl_establishment.name AS establishment,
       CASE WHEN reclamos.total_reclamos IS NULL THEN 0
       ELSE reclamos.total_reclamos
       END,
       CASE WHEN comentarios_reclamos.total_comentarios_reclamos IS NULL
       THEN 0
       ELSE comentarios_reclamos.total_comentarios_reclamos
       END,
       CASE WHEN reclamos_respondidos.total_reclamos_respondidos IS NULL
       THEN 0
       ELSE reclamos_respondidos.total_reclamos_respondidos
       END,
       CASE WHEN respuestas_reclamos.total_respuestas_reclamos IS NULL THEN 0
       ELSE respuestas_reclamos.total_respuestas_reclamos
       END,
       CASE WHEN recomendaciones.total_recomendaciones IS NULL THEN 0
       ELSE recomendaciones.total_recomendaciones
       END,
       CASE WHEN comentarios_recomendaciones.total_comentarios_recomendaciones IS NULL
       THEN 0
       ELSE comentarios_recomendaciones.total_comentarios_recomendaciones
       END,
       CASE WHEN recomendaciones_respondidas.total_recomendaciones_respondidas IS NULL
       THEN 0
       ELSE recomendaciones_respondidas.total_recomendaciones_respondidas
       END,
       CASE WHEN respuestas_recomendaciones.total_respuestas_recomendaciones IS NULL
       THEN 0

```

```

ELSE respuestas_recomendaciones.total_respuestas_recomendaciones
END
FROM reclamoscl_establishment
LEFT JOIN (
    SELECT reclamoscl_article.establishment_id,
           COUNT(*) AS total_reclamos
    FROM reclamoscl_article
    WHERE reclamoscl_article.type = True
    GROUP BY reclamoscl_article.establishment_id
) AS reclamos
ON reclamoscl_establishment.id = reclamos.establishment_id
LEFT JOIN (
    SELECT reclamoscl_article.establishment_id,
           COUNT(*) AS total_reclamos_respondidos
    FROM reclamoscl_article
    WHERE reclamoscl_article.type = True AND reclamoscl_article.solved = True
    GROUP BY reclamoscl_article.establishment_id
) AS reclamos_respondidos
ON reclamoscl_establishment.id = reclamos_respondidos.establishment_id
LEFT JOIN (
    SELECT reclamoscl_article.establishment_id,
           COUNT(reclamoscl_comment.*) AS total_comentarios_reclamos
    FROM reclamoscl_article
    LEFT JOIN reclamoscl_comment
    ON reclamoscl_article.id = reclamoscl_comment.article_id
    WHERE reclamoscl_article.type = True
    GROUP BY reclamoscl_article.establishment_id
) AS comentarios_reclamos
ON reclamoscl_establishment.id = comentarios_reclamos.establishment_id
LEFT JOIN (
    SELECT reclamoscl_article.establishment_id,
           COUNT(reclamoscl_solution.*) AS total_respuestas_reclamos
    FROM reclamoscl_article
    LEFT JOIN reclamoscl_solution
    ON reclamoscl_article.id = reclamoscl_solution.article_id
    WHERE reclamoscl_article.type = True
    GROUP BY reclamoscl_article.establishment_id
) AS respuestas_reclamos
ON reclamoscl_establishment.id = respuestas_reclamos.establishment_id
LEFT JOIN (
    SELECT reclamoscl_article.establishment_id,
           COUNT(*) AS total_recomendaciones
    FROM reclamoscl_article
    WHERE reclamoscl_article.type = False
    GROUP BY reclamoscl_article.establishment_id
) AS recomendaciones
ON reclamoscl_establishment.id = recomendaciones.establishment_id
LEFT JOIN (
    SELECT reclamoscl_article.establishment_id,
           COUNT(reclamoscl_comment.*) AS total_comentarios_recomendaciones

```

```

FROM reamoscl_article
LEFT JOIN reamoscl_comment
ON reamoscl_article.id = reamoscl_comment.article_id
WHERE reamoscl_article.type = False
GROUP BY reamoscl_article.establishment_id
) AS comentarios_recomendaciones
ON reamoscl_establishment.id = comentarios_recomendaciones.establishment_id
LEFT JOIN (
SELECT reamoscl_article.establishment_id,
COUNT(*) AS total_recomendaciones_respondidas
FROM reamoscl_article
WHERE reamoscl_article.type = False AND reamoscl_article.solved = True
GROUP BY reamoscl_article.establishment_id
) AS recomendaciones_respondidas
ON reamoscl_establishment.id = recomendaciones_respondidas.establishment_id
LEFT JOIN (
SELECT reamoscl_article.establishment_id,
COUNT(reamoscl_solution.*) AS total_respuestas_recomendaciones
FROM reamoscl_article
LEFT JOIN reamoscl_solution
ON reamoscl_article.id = reamoscl_solution.article_id
WHERE reamoscl_article.type = False
GROUP BY reamoscl_article.establishment_id
) AS respuestas_recomendaciones
ON reamoscl_establishment.id = respuestas_recomendaciones.establishment_id
WHERE reamoscl_establishment.industry_id = 4
ORDER BY total_reamos DESC, total_recomendaciones DESC;

```

A.7. Distribución de Reamos, Recomendaciones, Respuestas y Comentarios por año y por Institución

```

SELECT EXTRACT(YEAR FROM reamoscl_article.published) AS ano,
COUNT(reamos.id) AS total_reamos,
COUNT(reamos_respondidos.id)
AS total_reamos_respondidos,
SUM(respuestas_reamos.total_respuestas_reclamo)
AS total_respuestas_reamos,
SUM(comentarios_reamos.total_comentarios_reclamo)
AS total_comentarios_reamos,
COUNT(recomendaciones.id) AS total_recomendaciones,
COUNT(recomendaciones_respondidas.id)
AS total_recomendaciones_respondidas,
SUM(respuestas_recomendaciones.total_respuestas_recomendacion)
AS total_respuestas_recomendaciones,
SUM(comentarios_recomendaciones.total_comentarios_recomendacion)

```

```

        AS total_comentarios_recomendaciones
FROM reamoscl_article
LEFT JOIN (
    SELECT reamoscl_article.id
    FROM reamoscl_article
    WHERE reamoscl_article.type = True
) AS reamos
ON reamoscl_article.id = reamos.id
LEFT JOIN (
    SELECT reamoscl_article.id
    FROM reamoscl_article
    WHERE reamoscl_article.type = True AND reamoscl_article.solved = True
) AS reamos_respondidos
ON reamoscl_article.id = reamos_respondidos.id
LEFT JOIN (
    SELECT reamoscl_article.id,
        COUNT(reamoscl_solution.*) AS total_respuestas_reclamo
    FROM reamoscl_article
    LEFT JOIN reamoscl_solution
    ON reamoscl_article.id = reamoscl_solution.article_id
    WHERE reamoscl_article.type = True AND reamoscl_article.solved = True
    GROUP BY reamoscl_article.id
) AS respuestas_reamos
ON reamoscl_article.id = respuestas_reamos.id
LEFT JOIN (
    SELECT reamoscl_article.id,
        COUNT(reamoscl_comment.*) AS total_comentarios_reclamo
    FROM reamoscl_article
    LEFT JOIN reamoscl_comment
    ON reamoscl_article.id = reamoscl_comment.article_id
    WHERE reamoscl_article.type = True
    GROUP BY reamoscl_article.id
) AS comentarios_reamos
ON reamoscl_article.id = comentarios_reamos.id
LEFT JOIN (
    SELECT reamoscl_article.id
    FROM reamoscl_article
    WHERE reamoscl_article.type = False
) AS recomendaciones
ON reamoscl_article.id = recomendaciones.id
LEFT JOIN (
    SELECT reamoscl_article.id
    FROM reamoscl_article
    WHERE reamoscl_article.type = False AND reamoscl_article.solved = True
) AS recomendaciones_respondidas
ON reamoscl_article.id = recomendaciones_respondidas.id
LEFT JOIN (
    SELECT reamoscl_article.id,
        COUNT(reamoscl_solution.*) AS total_respuestas_recomendacion
    FROM reamoscl_article

```

```

LEFT JOIN reclamoscl_solution
ON reclamoscl_article.id = reclamoscl_solution.article_id
WHERE reclamoscl_article.type = False AND reclamoscl_article.solved = True
GROUP BY reclamoscl_article.id
) AS respuestas_recomendaciones
ON reclamoscl_article.id = respuestas_recomendaciones.id
LEFT JOIN (
SELECT reclamoscl_article.id,
COUNT(reclamoscl_comment.*) AS total_comentarios_recomendacion
FROM reclamoscl_article
LEFT JOIN reclamoscl_comment
ON reclamoscl_article.id = reclamoscl_comment.article_id
WHERE reclamoscl_article.type = False
GROUP BY reclamoscl_article.id
) AS comentarios_recomendaciones
ON reclamoscl_article.id = comentarios_recomendaciones.id
WHERE reclamoscl_article.establishment_id = 227
GROUP BY ano
ORDER BY ano;

```

A.8. Listado de Palabras de los Tópicos de LDA sobre Reclamos, por año y por Institución

```
SELECT classification_context.id AS context_id,  
       classification_context.establishment_id,  
       classification_context.type,  
       classification_context.started,  
       classification_context.ended,  
       classification_ldatopic.id AS ldatopic_id,  
       classification_ldatopic.number AS ldatopic_number,  
       tools_word.word AS palabra,  
       to_char(  
         round(  
           CAST (classification_ldatopicword.probability AS numeric),  
           3),  
         '0D999')  
       AS probabilidad  
FROM classification_ldatopicword  
LEFT JOIN tools_word  
ON classification_ldatopicword.word_id = tools_word.id  
LEFT JOIN classification_ldatopic  
ON classification_ldatopicword.ldatopic_id = classification_ldatopic.id  
LEFT JOIN classification_context  
ON classification_ldatopic.context_id = classification_context.id  
WHERE classification_context.establishment_id = 299  
AND classification_context.type = True  
AND classification_context.started = '2007-01-01 00:00:00'  
AND classification_context.ended = '2007-12-31 23:59:59'  
ORDER BY classification_ldatopic.number, classification_ldatopicword.probability DESC;
```