



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

**ELECCIONES PRESIDENCIALES 2013:
UN ANÁLISIS SEMÁNTICO DE LA
PRENSA DIGITAL CHILENA**

DIEGO FELIPE ÁVILA ORELLANA

Proyecto de Tesis para optar al Grado de
Magíster en Procesamiento y Gestión de la Información

Profesor Supervisor:
JORGE GANA LEAY

Santiago de Chile, marzo, 2016.



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA
DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN

ELECCIONES PRESIDENCIALES 2013: UN ANÁLISIS SEMÁNTICO DE LA PRENSA DIGITAL CHILENA

DIEGO FELIPE ÁVILA ORELLANA

Proyecto de Tesis presentado a la Comisión integrada por:

JORGE GANA L.

DENIS PARRA S.

CÉSAR AGUILAR

JAIME NAVÓN C.

Para completar las exigencias del Grado de
Magíster en Procesamiento y Gestión de la Información

Santiago de Chile, marzo, 2016.

AGRADECIMIENTOS

Quiero agradecer por su colaboración en este trabajo de tesis a Jorge Gana, Denis Parra, César Aguilar y Karla Vidal.

INDICE GENERAL

Pág.

AGRADECIMIENTOS.....	I
INDICE DE TABLAS.....	IV
INDICE DE FIGURAS.....	V
RESUMEN	VI
ABSTRACT	VII
I. INTRODUCCIÓN	1
I.1 HIPÓTESIS Y OBJETIVOS	1
I.2 METODOLOGÍA	2
I.3 ORGANIZACIÓN DEL CONTENIDO	3
II. TENDENCIA EN MEDIOS DE COMUNICACIÓN	4
II.1 TENDENCIA EN MEDIOS DE COMUNICACIÓN.....	4
II.1.1 <i>Tipos de tendencia</i>	5
II.1.2 <i>Tendencia en medios de comunicación chilenos</i>	6
II.1.3 <i>Cobertura de las elecciones presidenciales chilenas</i>	6
II.2 ANÁLISIS AUTOMATIZADO DE ARTÍCULOS DE NOTICIAS.....	8
II.2.1 <i>Análisis de Semántica Latente</i>	9
III. ANÁLISIS DE SEMÁNTICA LATENTE.....	10
III.1. ORIGEN DE LSA EN LA TEORÍA DE ADQUISICIÓN DEL LENGUAJE	10
III.2. ORIGEN DE LSA EN RECUPERACIÓN DE INFORMACIÓN	11
III.3. PRINCIPIOS LINGÜÍSTICOS DE LSA.....	12
III.3.1. <i>Modelos sintagmáticos</i>	13
III.3.2. <i>Modelos paradigmáticos</i>	15
III.4. CONSIDERACIONES LINGÜÍSTICAS.....	16
III.5. MODELO VECTORIAL DE LSA	17
IV. IMPLEMENTACIÓN.....	20
IV.1. SELECCIÓN.....	20
IV.1.1. <i>Parsing</i>	21
IV.1.2. <i>Recuperación</i>	22
IV.1.3. <i>Búsqueda a lo ancho</i>	22
IV.2. PREPROCESAMIENTO	23
IV.3. MINERÍA DE DATOS.....	24
IV.4. VISUALIZACIÓN.....	27
IV.4.1. <i>Similitud de términos</i>	27
IV.4.2. <i>Dispersión de documentos</i>	29
V. RESULTADOS.....	32
V.1. COBERTURA	32
V.2. DISPERSIÓN DE DOCUMENTOS	33
V.1. ANÁLISIS DE SEMÁNTICA LATENTE	36
V.1.1. EMOL.....	38

V.1.2. LA TERCERA.....	39
V.1.3. LA NACIÓN	41
V.1.4. EL MOSTRADOR	43
VI. CONCLUSIONES	45
VI.1. OBJETIVO GENERAL.....	45
VI.2. OBJETIVOS ESPECÍFICOS.....	46
VI.3. TRABAJOS FUTUROS	47
BIBLIOGRAFIA.....	48
A N E X O S	51
ANEXO A : TABLAS LSA EMOL.....	52
ANEXO B : TABLAS LSA LA TERCERA	59
ANEXO C : TABLAS LSA LA NACIÓN.....	63
ANEXO D : TABLAS LSA EL MOSTRADOR.....	68

INDICE DE TABLAS

Pág.

Tabla III-1: Ocurrencias de palabras en cada documento	14
Tabla III-2: Ocurrencias de palabras frente a otras palabras.....	16
Tabla III-3: Matriz X - Términos por documentos	18
Tabla III-4: Matriz U - Componentes vectoriales de cada término en dimensiones d_0 y d_1	19
Tabla IV-1: Detalle de tema 0 para La Nación	26
Tabla V-1: Menciones de candidatos y votación final obtenida	32
Tabla V-2: Dispersión de documentos en medios analizados.....	34
Tabla V-3: Distribución de caracteres y párrafos.....	36

INDICE DE FIGURAS

Pág.

Figura III-1: Descomposición en valor singular	18
Figura III-2: Representación vectorial de los términos	19
Figura IV-1: Proceso KDD	20
Figura IV-2: Algoritmo de Búsqueda a lo ancho	22
Figura IV-3: Grafo Semántico	28
Figura IV-4: Visualización Web de Grafo Semántico	29
Figura IV-5: Visualización de dispersión de documentos	31
Figura V-1: Porcentaje de menciones de candidatos en medios y votación obtenida.....	33
Figura V-2: Dispersión de documentos por medio.	35
Figura V-3: Detalle de dispersión por medio (en sentido del reloj, desde arriba a la izquierda: EMOL, La Tercera, El Mostrador y La Nación).....	35
Figura V-4: Distribución de párrafos por documentos.....	37
Figura V-5: Campos semánticos de EMOL.	39
Figura V-6: Campos semánticos de La Tercera.	41
Figura V-7: Campos semánticos de La Nación.....	43
Figura V-8: Campos semánticos de El Mostrador.	44

RESUMEN

El presente trabajo realiza un análisis de la cobertura a las elecciones presidenciales chilenas en cuatro de los medios más importantes de prensa digital durante el periodo comprendido entre la inscripción de candidaturas y la primera vuelta de la elección presidencial utilizando el método computacional de Análisis de Semántica Latente (LSA), el cual entrega como resultado grupos de términos con significados relacionados en forma de campos semánticos. El objetivo de este estudio es comprobar, mediante el procesamiento masivo de los artículos de prensa usando LSA, si es posible determinar tendencias en la cobertura de prensa enfocándose en los campos semánticos resultantes. Los resultados encontrados muestran marcadas diferencias de contenido temático y volumen de artículos entre las coberturas de cada candidato a través de los medios estudiados.

ABSTRACT

The present work performs an analysis to the media coverage of the Chilean presidential elections in four of the most relevant digital newspapers during time of candidate inscription and the first lap of the presidential election by using Latent Semantic Analysis (LSA), which delivers as a result a group of the most meaningful related terms in form of semantic fields. The objective of this study is to confirm, by the use of LSA on a massive amount of newspaper articles, whether it is possible to determine the existence of bias on said media focusing on the resulting semantic fields.

Results found noticeable differences between content and volume of articles among the coverage of each candidate throughout the studied media.

I. INTRODUCCIÓN

El presente trabajo de investigación tiene como propósito analizar la producción de prensa digital durante el contexto de las elecciones presidenciales chilenas del año 2013 utilizando técnicas de minería de texto y procesamiento de lenguaje natural, en específico Análisis de Semántica Latente (LSA), con el fin de detectar la presencia de tendencia en la cobertura noticiosa de los candidatos presidenciales.

El análisis de tendencia se enmarca en las categorías de D'Alessio (Tendencia Estructural y Tendencia de Cobertura) y toma como corpus a la producción los medios electrónicos de La Tercera, El Mostrador, La Nación y EMOL durante la primera vuelta de la elección presidencial. Por el lado del LSA, su aplicación permite seleccionar términos relevantes, extraer su relación semántica y obtener temáticas emergentes con el fin de determinar diferencias de contenido en la cobertura a los distintos candidatos.

En el contexto de las candidaturas presidenciales de 2013, se detectan marcadas diferencias en el contenido de cobertura a distintos candidatos según medio de comunicación. Junto con esto se encuentran diferencias en la cantidad de artículos dedicados a cada candidato en cada uno de los medios estudiados.

I.1 Hipótesis y objetivos

Utilizando como punto de partida la propuesta que la tendencia partidista como

la más relevante en medios de distintas orientaciones políticas, se toma como hipótesis que existen tendencias en los principales medios de comunicación. Estas tendencias se deberían manifestar tanto en la cantidad de artículos sobre candidatos, con en el mismo contenido de estos artículos.

El objetivo principal planteado corresponde a responder la hipótesis de la existencia de tendencia en medios de comunicación, mientras que los objetivos específicos se listan a continuación.

- a) Generar un grafo semántico a partir de los términos más relevantes en los artículos de noticias para cada medio analizado.
- b) Representar visualmente la similitud de los artículos de noticias para cada medio analizado.
- c) Cuantificar los artículos dedicados a cada candidato por cada uno de los medios.

I.2 Metodología

Para cumplir los objetivos previamente señalados se utiliza una implementación de la metodología KDD (Knowledge Discovery in Databases).

El proceso se inicia con la selección y recuperan de datos de artículos de noticias, posteriormente se preprocesan, se transforman al modelo vectorial, se realiza minería de datos con LSA, obteniéndose modelos de similitud semántica y finalmente se interpretan los resultados.

I.3 Organización del contenido

La estructura de este documento presenta la investigación en cinco capítulos comenzando con el Capítulo II, en el cual se define conceptualmente la problemática de tendencia en medios de comunicación, la situación chilena y el estado del análisis automatizado de artículos de noticias por técnicas de Minería de Texto y Procesamiento de Lenguaje Natural.

El Capítulo III presenta el método de Análisis de Semántica Latente, técnica utilizada para el análisis de los datos recuperados, considerando su origen, principios lingüísticos y modelo vectorial.

El Capítulo IV describe la implementación del análisis, incluyendo la selección y recuperación de artículos, preprocesamiento, transformación de datos, minería de datos con LSA y la generación de las visualizaciones de grafos semánticos y dispersión de documentos.

El Capítulo V presenta los resultados, partiendo por un análisis cuantitativo de la cobertura de prensa a los candidatos presidenciales, continuando con un análisis semántico de los términos relevantes y temas emergentes en artículos de noticias para cada medio de comunicación, para terminar con una comparación entre los documentos por cada medio.

Finalmente, el Capítulo VI se analizan las conclusiones y referencias a trabajos futuros.

II. TENDENCIA EN MEDIOS DE COMUNICACIÓN

En este capítulo se explicarán los conceptos básicos de tendencia en medios de comunicación a partir del principal metaestudio, *Media bias in presidential elections: a meta-analysis*, continuando con el listado de los tipos de sesgo descritos en ese estudio. En seguida se contextualizará la problemática de la tendencia en medios de comunicación chilenos, para finalizar con un estado del arte de técnicas de análisis computacional de texto empleadas para realizar esta tarea en grandes volúmenes de texto.

II.1 Tendencia en medios de comunicación

La preocupación sobre la presencia de tendencia en medios de comunicación ha dado origen a una gran cantidad de estudios. La principal recopilación sistemática ha sido el trabajo de D'Alessio (D'Alessio & Allen 2000) que incluye 59 estudios cuantitativos a la cobertura de prensa estadounidense de las elecciones presidenciales desde 1948 al año 2000.

La tendencia en medios de comunicación se define por D'Alessio como un acto deliberado, influyente y sostenido, que además cuestione las ideas ampliamente respaldadas por la población. Casos como el sesgo pro-capitalista y anticomunista observado en la prensa estadounidense, no se consideran relevantes porque a pesar de ser deliberados, influyentes y sostenidos, mantiene una idea apoyada por la mayoría de la población.

Por otro lado, la tendencia partidista o ideológica apunta a sectores más diversos de la población, de tal forma que un medio que da énfasis a un punto de

vista particular, causará rechazo en la población que tenga un punto de vista contrario. Con esto, se considera que el sesgo o tendencia partidista cuenta con todas las características necesarias para ser considerado relevante.

II.1.1 Tipos de tendencia

El trabajo de D'Alessio encontró las siguientes categorías a partir de los estudios previos recopilados:

a) Tendencia de Selección

El primer tipo de tendencia es el de selección, donde periodistas y editores seleccionan los artículos de noticias que van a ser presentados al público y, al mismo tiempo, deseleccionan artículos que no van a llegar de ninguna forma al público. En ambos casos por motivos ideológicos relacionados con la línea editorial del medio en cuestión. Se considera la más difícil de detectar porque consiste justamente de las historias que no son publicadas.

b) Tendencia de Cobertura

El segundo tipo de tendencia es de cobertura. Si se tiene dos posiciones respecto a un tema, se cuantifica la cantidad de artículos que presenta cada una de las posiciones y se determina si existe una tendencia cuando se encuentra una diferencia significativa en la cobertura de cada una de las posiciones. En el caso de este estudio, se puede considerar como la cantidad de artículos donde cubre cada candidato.

c) Tendencia Estructural

El último tipo de tendencia se denomina estructural y se refiere a la opinión vertida por el medio en los artículos que publica. Generalmente es presentada en análisis clasificando la cobertura como “favorable” o “desfavorable”, o bien “positiva” y “negativa”.

II.1.2 Tendencia en medios de comunicación chilenos

El caso particular de los medios de comunicación chilenos, a diferencia del caso de los medios estadounidenses mencionado anteriormente, el esfuerzo en encontrar sesgos o tendencias apunta a la concentración de medios de comunicación. Tal es el caso del estudio Concentración de la propiedad de los medios de comunicación (Rogel et al. 2010), donde se analiza la concentración de los medios de comunicación en contexto de control de información y poder político.

Ewa Sapiezynska estudió los niveles de restricción percibidos por periodistas mediante encuestas (Sapiezynska et al. 2013), encontrando un nivel alto especialmente en las restricciones dadas por sus superiores. En las categorías de D'Alessio, este caso corresponde al primer tipo de sesgo, filtrando los artículos de prensa antes de que puedan salir al público.

II.1.3 Cobertura de las elecciones presidenciales chilenas

Siguiendo la conclusión de D'Alessio sobre la relevancia de la tendencia partidista en la cobertura de prensa, resulta de especial interés estudiar la cobertura de prensa a las elecciones presidenciales chilenas del año 2013 por la cantidad de candidatos sin precedentes hasta el momento (TRICEL 2013).

Las elecciones presidenciales chilenas de 2013 presentan dos elementos particulares: por un lado, son las primeras elecciones presidenciales desde la implementación de la inscripción automática y voto voluntario; mientras que también presentan la mayor cantidad de candidatos en la historia electoral

chilena, escenario nunca antes visto.

El principal interés en la cobertura de prensa se centra en la prensa internacional. Un estudio de Fundación Imagen País cuantificó las apariciones de estas elecciones en la prensa internacional e identificó los principales temas tratados (Fundación Imagen de Chile 2013).

Un estudio de la empresa de estudio de medios Conecta publicado por el periódico La Tercera cuantificó las menciones de candidatos en la prensa y estudió la relación entre cobertura causada por una vocería del candidato o motivada por el propio medio. Es decir, se comparó la cantidad de apariciones en medios donde se habla del candidato frente a las veces que el mismo candidato habla. Por un lado, se encontró que para todas las apariciones en prensa Michelle Bachelet, 40% corresponde a menciones mientras que un 23% corresponde a vocerías, por otro lado, para el resto de los candidatos, la mayor parte de apariciones corresponde a vocerías (Tercera 2013). Otros estudios sobre la percepción de los candidatos se basan en los comentarios extraídos de redes sociales, no de la prensa, como en el caso del Observatorio Político, el cual se basa en la clasificación de polaridad de menciones en Twitter a candidatos durante el periodo de la campaña (Marín, M., Mendoza, M., Rodríguez, A., Contreras, A., Caro, D., Cabrera L, & Sanhueza, 2013; Mendoza, 2014).

II.2 Análisis automatizado de artículos de noticias

La gran cantidad de artículos publicados durante un periodo acotado de tiempo puede ser procesada solamente de forma automática, tanto en la recuperación de artículos como en el procesamiento.

Las elecciones presidenciales chilenas del año 2013 presentaron una alta cobertura de prensa respecto a los distintos candidatos a través de diversos medios de comunicación como televisión, radio y prensa.

Solamente en prensa digital se generaron miles de artículos sobre los candidatos en el lapso de tiempo entre las inscripciones oficiales de las candidaturas y el día de la votación. Analizar toda esta información manualmente requeriría una cantidad inabordable de tiempo, por lo que resulta de utilidad utilizar un método automatizado de recuperación y análisis de esta información.

Estudios como *Media coverage in times of political crisis: A text mining approach* (Junqué de Fortuny et al. 2012), analizan la tendencia de medios digitales en el periodo de formación del gobierno belga en 2011 mediante técnicas de minería de textos y minería de opinión, encontrando marcadas diferencias entre la coberturas a determinados partidos políticos en distintos medios. Este método se basa en un diccionario de sentimientos que incluye términos previamente calificados según polaridad y clasifica cada artículo de noticias en una categoría positiva o negativa.

Una publicación reciente (Veltri 2012), fuera del área política, analizó la representación de la nanotecnología en la prensa española. A diferencia el estudio anterior, Veltri utilizó una metodología llamada Análisis de Semántica

Latente (LSA). Dicha metodología permite analizar los contenidos a partir de campos semánticos (clúster temáticos, según el estudio), siendo equivalente a usar un método no supervisado al poder obtener resultados sin un entrenamiento previo, como en el caso anteriormente mencionado.

II.2.1 Análisis de Semántica Latente

Este estudio propone un análisis de los contenidos de los medios de comunicación digital a los candidatos mediante la técnica llamada Análisis de Semántica Latente (LSA). Esta técnica basada en los procesos de adquisición de lenguaje, permite analizar un gran volumen de textos y encontrar campos semánticos compuestos por términos relacionados, permitiendo un nivel más profundo de análisis de los contenidos de prensa que métodos de minería de opinión basados en clasificación al no depender con presupuestos del lenguaje utilizado, sino extrayendo el sentido de los mismos textos.

III. ANÁLISIS DE SEMÁNTICA LATENTE

Este capítulo explica el Análisis de Semántica Latente (LSA) comenzando por su origen en la Recuperación de Información y Psicología. En seguida se analizan los principios lingüísticos que lo sustentan y las limitaciones de esta metodología. Finalmente, se detalla la implementación algebraica de LSA mediante el modelo vectorial y descomposición en vector singular.

III.1. Origen de LSA en la teoría de adquisición del lenguaje

LSA fue propuesto como una teoría de adquisición de lenguaje por Landauer y Dumais (Landauer & Dumais 1997). Se presentó originalmente como “La Solución al Problema de Platón” referente a la pobreza de estímulo, es decir, el hecho de que las personas tienen un conocimiento del mundo mayor que la exposición que han tenido de éste mismo, o bien, expresado como la paradoja entre cómo la observación de una cantidad relativamente pequeña de eventos produce conocimiento que resulta ser generalmente correcto.

La respuesta de Platón es que la gente llega al mundo con conocimiento y solamente necesita ciertas pistas para completarlo. De manera similar, Chomsky ha postulado que la adquisición de lenguaje por parte de infantes no sería posible solamente con la exposición al lenguaje de adultos, sino que debe existir una base innata para tal propósito, sin requerir de experiencia (Landauer & Dumais 1997; Landauer 1998).

El trabajo de Landauer y Dumais propone explicar este fenómeno desde de la noción de que varios dominios del conocimiento contienen una gran cantidad de enlaces

débiles que permiten amplificar el proceso de aprendizaje a partir de la inferencia estadística. Este modelo se comparó con el proceso de aprendizaje de niños en edad escolar y demostró resultados similares en pruebas de selección múltiple de sinónimos, lo que puede interpretarse como que la información necesaria para responder pruebas estandarizadas de vocabulario puede inferirse sólo desde las estadísticas de uso. Otra interpretación más ambiciosa, por este par de autores, consiste en formular una teoría para toda la adquisición de conocimiento humano.

Estas dos interpretaciones consideran solamente información obtenida a partir del aprendizaje, sin mediar conocimiento previo como la cultura. Además, considera únicamente información de las palabras presentes en el análisis sin tomar en cuenta el orden, sintaxis o gramática. Para ilustrar esta idea, considérese las siguientes ecuaciones (Venegas V. 2003; Landauer 2002):

$$\text{ecks} + \text{wye} + \text{aye} = \text{foo} \quad (2.1)$$

$$\text{ecks} + \text{wye} + \text{bie} = \text{foo} \quad (2.2)$$

A pesar que no se puede decir nada sobre *ecks* y *wye*, a partir de ambas ecuaciones se puede inferir que *aye* y *bie* tienen el mismo significado. De esta manera, se puede extraer información semántica simplemente desde la distribución de términos en un texto a partir de un contexto (oración, párrafo, documento u otro).

III.2. Origen de LSA en Recuperación de Información

En la disciplina de Recuperación de Información existe la compleja tarea de comparar literalmente los términos de búsqueda con el contenido de los documentos. El mayor problema subyace en que el lenguaje humano tiene diferentes formas de

expresarse, dando lugar a casos donde un mismo concepto se puede representar con distintos términos (sinonimia) o que un término pueda referirse a varios conceptos (polisemia) (Rosario 2000).

En estos casos, la mayor solución apunta a realizar una búsqueda en base a un *concepto* o *sentido* en el documento.

Deerwester postuló la Indexación Semántica Latente en *Indexing by Latent Semantic Analysis* (Deerwester et al. 1998) como una forma de detectar las estructuras semánticas subyacentes en los documentos a partir de una matriz de términos/documentos donde posteriormente se descompone mediante Descomposición de Valor Singular (SVD, por sus siglas en inglés) y luego se puede hacer una recuperación a partir de los vectores singulares obtenidos.

III.3. Principios Lingüísticos de LSA

En el artículo *The distributional hypothesis* (Sahlgren 2008), se analizan los principios lingüísticos de distintas técnicas de análisis semántico de documentos basadas en la distribución de los términos en estos mismos. Todas estas técnicas son entonces, catalogadas como técnicas de *semántica distribucional*.

La base de los dos principales tipos de técnicas, según Sahlgren, yace en las relaciones que pueden tener los términos en las unidades de lenguaje a partir de los principios delineados por Ferdinand de Saussure.

Saussure, quien sentó las bases del estructuralismo, centró su análisis en la *estructura* del lenguaje sobre su *uso*, o sea, en encontrar la lengua (*langue*), que a la vez, al ser usada, origina el habla (*parole*). Su principal forma de ilustrar esta idea es mediante la analogía con el ajedrez: las reglas, el tablero y las piezas son el equivalente a la

lengua, mientras que todas las jugadas posibles hechas por los jugadores corresponden al habla.

Esta analogía se extiende a la forma de identificar las piezas, cada pieza se identifica en cuanto a las diferencias que tiene con las demás. De la misma manera, los signos en el lenguaje se definen a partir de sus diferencias con el resto de los signos.

Aquí surgen dos tipos de relaciones: si dos términos aparecen en la misma unidad (sea oración, frase, párrafo, etc.), se está hablando de una relación *sintagmática*, definida por una relación en presencia; por el otro lado, si se está hablando de términos que pueden ocurrir en vez del otro, nunca al mismo tiempo, en una misma unidad, estamos hablando de relaciones *paradigmáticas*.

Ejemplificando, todas las palabras de la oración “el lobo tiene hambre” presentan relaciones sintagmáticas, por otro lado, en la combinación “el lobo tiene [hambre|sed]”, tanto “hambre” como “sed” ocurren cuando la otra palabra no está presente, teniendo entonces una relación paradigmática.

A partir de esta teoría, Sahlgren describe dos implementaciones principales:

III.3.1. Modelos sintagmáticos

En esta categoría se encuentra LSA como una de las implementaciones más importantes. Las implementaciones de los modelos sintagmáticos se dependen principalmente de la región de texto donde ocurren los términos, siendo éste el parámetro más importante. La región de texto puede ser desde frase, oración, párrafo hasta documento, esta última es preferida en contextos de Recuperación de Información.

La base de estas implementaciones es una matriz que registra la frecuencia de ocurrencias de un término en una región de texto. En el ejemplo a continuación (Tabla III-1: Ocurrencias de palabras en cada documento), se utiliza el documento como región de texto.

Tabla III-1: Ocurrencias de palabras en cada documento

Palabra/Documento	d_1	d_2	d_3	d_4	d_5	d_6	d_7
p_1	1	0	0	3	0	2	0
p_2	0	0	0	2	2	0	1
p_3	0	3	0	2	0	1	0
p_4	3	0	2	0	1	0	0
p_5	0	2	0	1	0	0	3

Finalmente, se calcula la semejanza entre dos palabras a partir de la similitud del coseno sobre su representación vectorial. Se prefiere esta medida de similitud porque considera solamente la dirección de los vectores, sin importar su magnitud. La similitud del coseno ocupa la siguiente fórmula:

$$a \cdot b = \|a\| \cdot \|b\| \cos \theta \quad (2.3)$$

Que despejando, resulta:

$$\cos \theta = \frac{a \cdot b}{\|a\| \cdot \|b\|} \quad (2.4)$$

Se define el módulo del vector como:

$$\|a\| = \sqrt{\sum a_i^2} \quad (2.5)$$

Y el producto punto entre vectores como:

$$a \cdot b = \sum a_i b_i \quad (2.6)$$

De tal forma, si aplicamos esta similitud para las palabras p_1 y p_3 de la Tabla III-1: Ocurrencias de palabras en cada documento, tenemos los vectores $p_1 = (1,0,0,3,0,2,0)$ y $p_3 = (0,3,0,2,0,1,0)$, se reemplaza en 2.4:

$$\cos \theta = \frac{(1,0,0,3,0,2,0) \cdot (0,3,0,2,0,1,0)}{\|(1,0,0,3,0,2,0)\| \cdot \|(0,3,0,2,0,1,0)\|} \quad (2.7)$$

Finalmente, se reemplaza en 2.7 el producto punto por la fórmula 2.6 y los módulos de los vectores p_1 y p_3 por la fórmula 2.5, obteniendo el resultado 0,571, equivalente a un ángulo de 55,15°.

$$\cos \theta = \frac{8}{3,741 \cdot 3,741} = 0,571 \quad (2.8)$$

III.3.2. Modelos paradigmáticos

La base de los modelos paradigmáticos está, como se explicó anteriormente, en encontrar palabras que no ocurran al mismo tiempo pero sí en contextos similares. Entonces, se vuelve necesario para sus implementaciones, el considerar los siguientes aspectos: 1) El tamaño de la región de contexto en donde se obtiene la información, 2) la posición de las palabras dentro de la región de contexto y finalmente, 3) la dirección en la que se extiende la región de texto.

De esta forma, una implementación de modelos paradigmáticos se realiza también con una matriz (Tabla III-2: Ocurrencias de palabras frente a otras palabras), pero esta vez se mapea cada ocurrencia de una palabra frente a otra en una misma ventana previamente definida, sin importar cuál sea esta ventana.

Tabla III-2: Ocurrencias de palabras frente a otras palabras

Palabra	p ₁	p ₂	p ₃	p ₄	p ₅
p ₁	0	1	0	0	0
p ₂	1	0	0	1	0
p ₃	0	1	0	0	0
p ₄	0	0	1	0	0
p ₅	0	0	0	1	0

Esta matriz no es simétrica, los vectores de las filas representan ocurrencias a la derecha de la ventana mientras que los vectores de las columnas representan las ocurrencias a la izquierda de la ventana.

Una implementación completa de un modelo de semántica distribucional bajo el modelo paradigmático, se puede encontrar en *Hyperspace Analogue to Language* (HAL) (Lund & Burgess 1996).

III.4. Consideraciones lingüísticas

Los métodos de semántica distribucional no son mirados con buenos ojos por parte de los lingüistas porque se basan principalmente en el esquema de “bolsa de palabras”, especialmente en el caso de los modelos paradigmáticos. Esto significa que se contabiliza la ocurrencia de las palabras sin considerar su orden de aparición en el texto, lo que deja de lado toda información sintáctica.

Esta misma reticencia a modelos que eliminan gran parte de la información lingüística llevó a pruebas tales como el estudio *Rules for syntax, vectors for semantics* (Wiemer-Hastings y Zipitria, 2001; en Sahlgren, 2006), que utilizó

etiquetado morfosintáctico en una implementación de LSA, dando como resultado un menor desempeño del análisis ocasionado por la mayor cantidad de términos que aparecen debido a este mismo etiquetado.

La base estructuralista del paradigma de semántica distribucional hace que todos los análisis que se puedan obtener a partir de estos modelos sean cuestionables solamente dentro de contextos estructuralistas. Es decir, el sentido que extraen estos modelos es sólo el sentido extraído a partir de la información distribucional de las palabras en los textos, por ende, no es el sentido que está en la mente de las personas o el sentido que se encuentra en la cultura (Sahlgren 2008).

III.5. Modelo vectorial de LSA

LSA, como modelo sintagmático, se implementa a partir de una matriz de término por unidad de texto (como se ilustró en la Tabla III-1: Ocurrencias de palabras en cada documento). A diferencia del ejemplo básico discutido en la sección II.3.1, en un caso real es posible que la cantidad de documentos sea considerablemente mayor. Dada esta situación, resulta difícil calcular la similitud entre los términos a partir de vectores con un alto número de dimensiones. Por lo tanto, para resolver este problema, las implementaciones de LSA utilizan Descomposición de Valor Singular; SVD, por sus siglas en inglés (Figura III-1: Descomposición en valor singular).

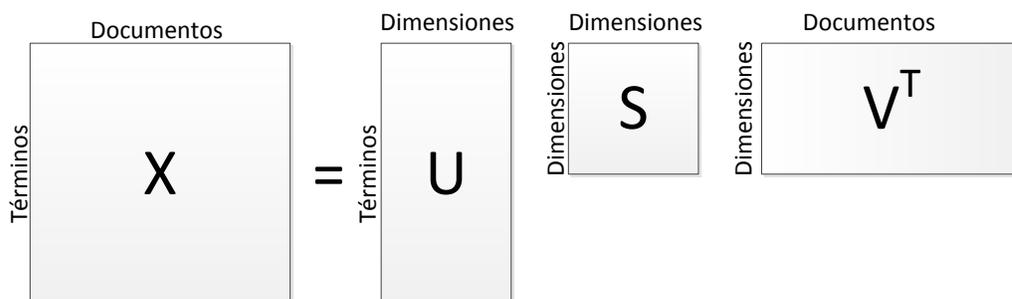


Figura III-1: Descomposición en valor singular

Mediante SVD, es posible tomar la matriz X (Tabla III-3: Matriz X - Términos por documentos) de términos por documentos a un espacio dimensional significativamente reducido, comprendido por tres matrices: Una matriz U (

Tabla III-4: Matriz U - Componentes vectoriales de cada término en dimensiones d_0 y d_1) de términos por dimensiones, una matriz S de dimensiones por dimensiones y finalmente, una matriz V^T de dimensiones por documentos. La idea tras la reducción de dimensiones es conservar sólo aquellas que aportan mayor variación a la colección de documentos.

Tabla III-3: Matriz X - Términos por documentos

	d1	d2	d3	d4	d5	d6
planta	2	1	0	2	0	1
energía	1	0	1	0	0	0
eólica	1	2	0	0	0	0
vegetal	0	0	0	1	2	0
árbol	0	0	0	1	0	2
flor	0	0	0	0	1	0

Tabla III-4: Matriz U - Componentes vectoriales de cada término en dimensiones d_0 y d_1

Término	d_0	d_1
energía	-0.47825	0.46439
eólica	-0.5325	0.4291
planta	-0.42998	-0.06259
vegetal	-0.31282	-0.46384
árbol	-0.4436	-0.5928
flor	-0.09056	-0.17242

Finalmente, a partir de la matriz U, se puede obtener una representación vectorial de los términos a partir de las dimensiones obtenidas. En el ejemplo actual (Figura III-2: Representación vectorial de los términos), se puede observar que por un lado los términos “árbol”, “vegetal” y “flor” tienen un alto grado de similitud semántica, mientras que por el otro, “eólica” y “energía” son también altamente similares.

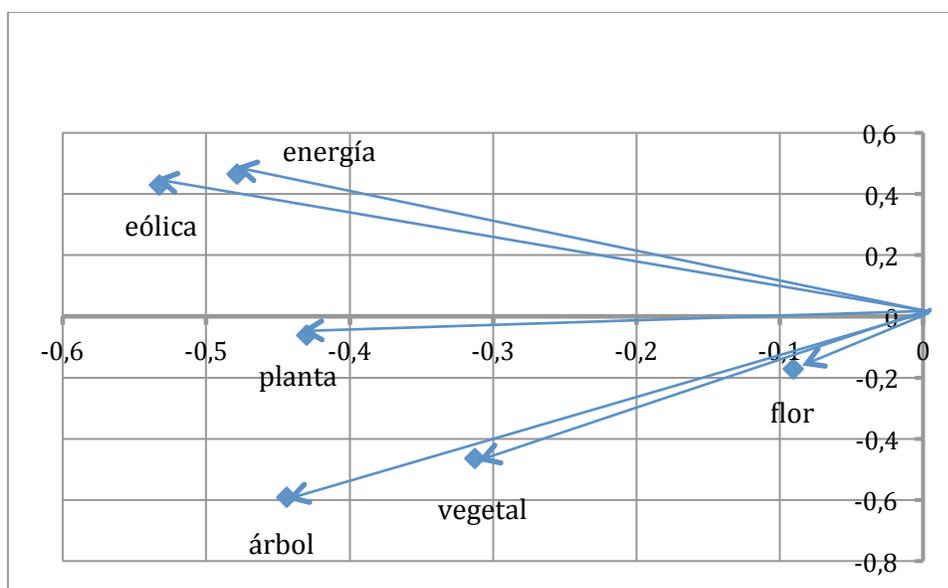


Figura III-2: Representación vectorial de los términos

IV. IMPLEMENTACIÓN

En este capítulo se describe la implementación del análisis, comenzando desde la selección de artículos, recuperación vía scrapping, preprocesamiento, Minería de Datos con LSA, hasta la evaluación y visualización.

La metodología de implementación sigue la estructura de un proceso de Descubrimiento de Conocimiento en Bases de Datos (KDD, por sus siglas en inglés). El proceso se inicia con una etapa de Selección, donde se recuperan los artículos considerados relevantes, luego se preprocesan, se efectúa el proceso de minería de datos con LSA, para luego ser evaluados visualmente lo que da paso finalmente a una interpretación de los resultados (Figura IV-1: Proceso KDD).



Figura IV-1: Proceso KDD

IV.1. Selección

La selección de artículos de prensa consideró medios digitales relevantes que cumplan con las características de ser masivos y al mismo tiempo, contener

información recuperable, considerando información recuperable como aquella presentada como texto (lo que descarta medios que publican su versión en línea como imágenes de sus versiones en papel, por ejemplo LUN) y que permiten ser descargados por medios automáticos. De esta forma, La Tercera, EMOL, La Nación y el Mostrador fueron seleccionados para el presente análisis.

Debido al tiempo requerido para descargar artículos de noticias mediante scraping, se utilizó el servicio ScraperWiki que provee un entorno para ejecutar tales procesos y guardarlos en una base de datos MySQL.

La base de datos guardó los siguientes elementos considerados relevantes para cada artículo descargado:

- Medio: nombre del medio.
- Categoría: nombre de la categoría correspondiente a la taxonomía interna de cada medio.
- Url: Dirección URL al artículo de noticias original.
- Fecha: Fecha de publicación del artículo.
- Título: Título de artículo de noticias
- Contenido: Texto del artículo.

IV.1.1. Parsing

El proceso de parsing se realizó de forma distinta para cada medio analizado utilizando principalmente la librería Lxml de Python, la cual recupera elementos de páginas Html con selectores equivalentes a los de CSS.

De esta forma, el título del artículo y texto se obtuvieron a partir de secciones en la misma página. La fecha del artículo se encontró en la misma Url para todos los casos

excepto para La Tercera, de donde se extrajo como parte del mismo contenido de la página.

IV.1.2. Recuperación

Inicialmente, el proceso de recuperación utilizó un algoritmo a lo ancho (breadth-first). Este fue el caso para El Mostrador y La Tercera, donde cada artículo de prensa contenía referencias a otros artículos relacionados. Sin embargo, EMOL y La Nación carecen de esta característica, por lo que fue necesario recuperar los artículos de noticias a partir de los buscadores que provee cada sitio utilizando palabras clave.

IV.1.3. Búsqueda a lo ancho

Se implementó una versión estándar de búsqueda a lo ancho (Figura IV-2: Algoritmo de Búsqueda a lo ancho) para los medios El Mostrador y La Tercera. Se definió una URL inicial y una lista de URL recorridas. El algoritmo toma la URL inicial, y obtiene el listado de URL que tengan el mismo formato (por ejemplo: “*www.medio.com/categoria/fecha/..*”), garantizando que sólo se recorran enlaces a artículos similares, descartando enlaces externos, publicidad o similares.

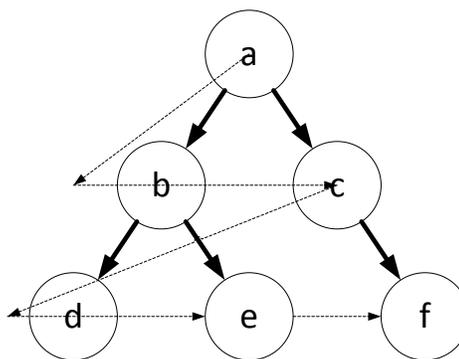


Figura IV-2: Algoritmo de Búsqueda a lo ancho

Para cada URL recorrida, se valida que en su título contenga una palabra clave relacionada con las elecciones presidenciales (nombres de candidatos, “elecciones”, “debate”, etc), de ser el caso, se obtienen los siguientes datos básicos: URL Título de la noticia, fecha y contenido, los cuales se almacenan en la base de datos. Este proceso se repite iterativamente hasta agotar los enlaces nuevos que no aparezcan en el listado de visitados.

IV.2. Preprocesamiento

Los documentos almacenados en la base de datos son descargados en formato JSON, generando un archivo distinto por cada medio, el cual pasa a ambiente Python para ser preprocesado. La fase de preprocesamiento utiliza principalmente la librería NLTK, Natural Language Toolkit (NLTK Project 2015).

El proceso consta de las siguientes fases:

- a) Tokenización: La separación del documento de texto en unidades significativas (tokens), en este caso, de palabras o términos. Esta funcionalidad está implementada como parte de NLTK.
- b) Eliminación de caracteres auxiliares: Los términos obtenidos sólo deben contar de caracteres alfanuméricos, por lo que se elimina todo el resto. Un caso especial corresponde a la secuencia “.-” que se usa en algunos medios para delimitar a la primera palabra en casos cuando ésta corresponde a la ciudad donde ocurrió el suceso reportado (Por ejemplo: “Santiago.- El presidente...”). La frecuencia de estos

términos ocasiona distorsiones el análisis posterior de frecuencias por lo que fueron eliminados.

c) Eliminación de palabras vacías. Las palabras vacías o “stop words” corresponden a aquellas que no entregan información semántica relevante, sean artículos, conjunciones y otras por el estilo. Se confeccionó una lista de stop-words a partir de diversas fuentes en línea. Luego de primeras instancias de análisis se agregaron los nombres de los candidatos a esta lista para eliminar la necesidad de ser tratados como bigramas.

d) Stemming: El Stemming es la reducción de palabras con alguna flexión (conjugación o declinación) a su raíz o *stem* (en inglés). Por ejemplo, para “gato”, “gata” y “gatos”, la raíz correspondiente es “gat”. Esto es necesario para procesar estadísticamente todos los términos con valor semántico similar, de manera de contabilizar distintas flexiones como un mismo término. En este estudio se utilizó para stemming el Algoritmo de Porter, con su implementación Snowball, incluida en NLTK.

IV.3. Minería de Datos

El resultado del preprocesamiento es un listado de documentos compuesto a la vez por un listado de términos, el cual se encuentra listo para entrar la proceso de minería de datos. Este proceso está basado en la librería Gensim para Python, orientada a análisis de texto con modelamiento de tópicos.

El proceso se inicia con la generación de un diccionario. Para Gensim, un diccionario corresponde a una lista con pares de término y número de identificación. Esto permite recuperar un término de vuelta en su representación vectorial.

A continuación, se genera el modelo de bolsa de palabras, que en la terminología de Gensim, se denomina *corpus*. Este corpus se compone de una lista de representaciones de los documentos compuesta por una lista de pares. Estos pares corresponden al número identificador del término correspondiente, y el siguiente a su frecuencia en cada documento.

Para bajar la ponderación en el corpus de las palabras frecuentes en todos los documentos pero que no entregan información, se realiza un proceso de Tf-idf (Text frequency - inverse document frequency, Frecuencia de texto - Frecuencia inversa de documento, en español)

La funcionalidad de cálculo de Tf-idf viene incorporada en Gensim. Toma como entrada el corpus de términos con frecuencias y retorna un corpus con las frecuencias reemplazadas por su valor de Tf-idf.

En seguida se realiza el proceso SVD. Tomando como entrada el corpus Tf-idf, el diccionario de términos y la cantidad de dimensiones deseadas, se obtiene un modelo LSA consistente en las matrices U , S y V^T , equivalente a la descomposición descrita en la sección III.5, Modelo vectorial de LSA

El último paso para el análisis es utilizar la función de Gensim para imprimir temas. Los temas son equivalentes a las dimensiones que se obtuvieron en SVD. De esta forma, si se ejecuta la función para imprimir n términos para m temas, el sistema entrega los primeros n términos que aportan más a cada uno de los m temas,

pudiendo repetirse términos entre los temas. La puntuación que tiene cada término puede ser negativa según la descomposición SVD, sin embargo, se considera su valor absoluto al ordenar su aporte al tema, pero finalmente se retorna el valor con su signo correspondiente.

La salida, entonces, se compone de listas de término por cada tema, acompañados de su valor en la matriz U de términos por temas. Adicionalmente, se agrega una lista de palabras correspondientes a las flexiones de los términos relevantes por tema (raíz), ordenados por frecuencia absoluta de ocurrencias en el total de textos por medio. A continuación, se agrega a modo de ejemplo el tema 0 para el medio La Nación (Tabla IV-1: Detalle de tema 0 para La Nación).

Tabla IV-1: Detalle de tema 0 para La Nación

Tema	Termino	Puntuación	Flexiones
0	primari	-0.10829616460682044	primarias:262 primaria:46 primariasen:2
0	matthei	-0.1030094137045523	matthei:393
0	gobiern	-0.085783828149057967	gobierno:591 gobierne:3 gobiernen:1
0	ex	-0.082202031815041279	ex:857
0	campa	-0.081150014640804222	campaña:356 campañas:10
0	educ	-0.079618159917576259	educacin:232 educativa:4 educativo:2 educadora:1
0	ministr	-0.079482452087252881	ministro:197 ministra:161 ministros:35 ministras:4
0	part	-0.078560678840425224	parte:210 partido:197 partidos:135 partir:29 partes:6 parti:6 partida:4 partir:2 parta:2
0	debat	-0.078529520981929044	debate:190 debates:45 debatir:12 debatido:2
0	com	-0.077648335402385305	comando:318 cómo:37 comado:2 comida:2 comer:1

IV.4. Visualización

Teniendo en cuenta que la salida obtenida por LSA consiste en listados de términos acompañados por su valor en la matriz U luego de la descomposición, la relación semántica entre estos términos fuera del tema no resulta evidente. De esta forma, se desarrolló una visualización basada en la matriz de similitud de términos U y otra por la matriz VT de documentos por temas.

IV.4.1. Similitud de términos

La representación visual de similitud de términos por LSA como grafo ha sido empleada en trabajos como Storylines (Zhu & Chen 2007) para obtener una visualización accesible del espacio de semántica latente y también como resolución de polisemia (Jorge-Botana et al. 2009; Jorge-Botana et al. 2010). Los trabajos como los de Botana, anteriormente mencionados, utilizan una visualización centrada en términos previamente determinados para luego buscar otros términos cercanos a estos mismos, generando sucesivamente un grafo semántico. Sin embargo, al considerar solamente términos previamente seleccionados se pierde la información de otros términos relevantes que aparecen de modo emergente en cada uno de los temas de la descomposición vectorial.

Tomando en cuenta lo anterior, este análisis considera 10 temas seleccionando los 10 términos más relevantes por cada uno, donde ciertos términos aparecen en varios temas. Para su visualización, se construyó un grafo a partir de las relaciones semánticas entre términos, coloreado según la pertenencia a los distintos temas.

El proceso se detalla a continuación:

1. Se construye un listado de términos a partir de las salidas eliminando las duplicaciones por apariciones en distintos temas.
2. El listado se ubica posteriormente en la matriz U para obtener la representación vectorial de cada término.
3. Finalmente, se calcula la similitud de coseno entre cada término. Para un umbral con valores iguales o superiores a 0.5, se considera que la similitud es suficiente y se almacena en un nuevo listado de enlaces.

La salida se construye en un archivo JSON compuesto en una parte por la lista inicial de términos acompañada de su puntuación y tema, y en otra, por el listado de enlaces compuesto por la referencia a cada par de términos y su distancia. El resultado solamente del grafo es el siguiente (Figura IV-3: Grafo Semántico):

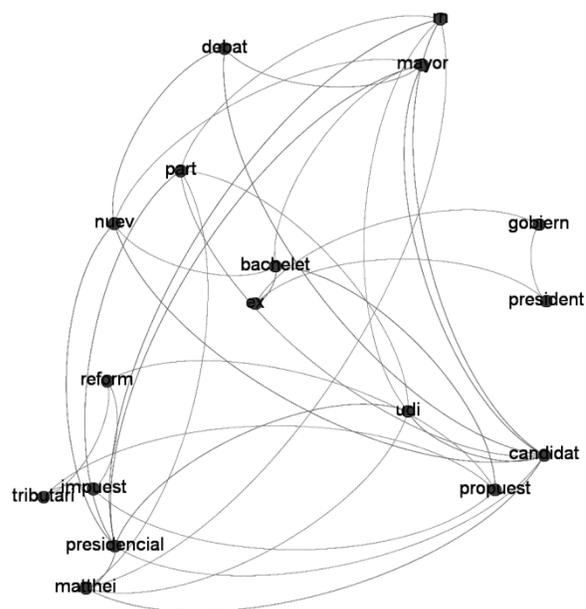


Figura IV-3: Grafo Semántico

El resultado final es una visualización web desarrollada a partir de D3, librería de generación de gráficos vectoriales SVG desde datos. Se compone del grafo semántico en la parte derecha de la pantalla con términos coloreados según sus temas. El listado de temas se muestra en el sector izquierdo de la pantalla, al hacer clic en cada tema se destacan los términos de cada tema y se muestra la lista de términos con la puntuación correspondiente (Figura IV-4: Visualización Web de Grafo Semántico).

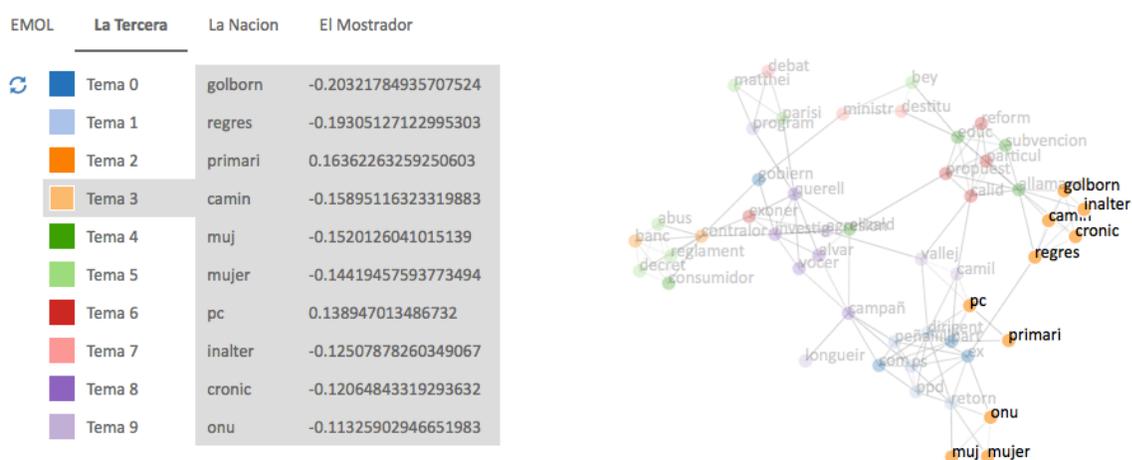


Figura IV-4: Visualización Web de Grafo Semántico

IV.4.2. Dispersión de documentos

En conjunto con la visualización de similitud de términos construida desde la matriz U , se desarrolló una visualización de similitud de documentos desde la matriz V^T . La matriz V^T referencia a todos los documentos del corpus en la matriz original de documentos por términos, la diferencia es que en vez de estar descritos por un vector

con tantas dimensiones como la cantidad de términos totales, la reducción por SVD dejó a cada documento de la matriz V^T descrito con solamente 10 dimensiones. De todas formas, esta matriz resultante no es directamente visualizable ya que visualización bidimensional requiere solamente un par de coordenadas por cada documento para posicionarlo en un plano.

Tal proceso se realizó con la técnica de Escalamiento Multidimensional (MDS, Multi Dimensional Scaling). MDS permite obtener desde un conjunto vectorial, un subconjunto N-dimensional el cual mantiene las distancias entre cada elemento del conjunto vectorial original. En contextos similares, se ha utilizado MDS en la visualización de corpus de texto, como en el trabajo de Fortuna (Fortuna et al. 2005). En este trabajo se ejecutó MDS sobre la matriz V^T para obtener un conjunto de 2 dimensiones, utilizando distancia euclidiana mediante la implementación de MDS de la librería Sklearn del lenguaje Python. La visualización, construida igualmente con la librería D3, resulta en un gráfico de dispersión con cada documento proyectado en un plano, coloreado según su medio de origen (Figura IV-5: Visualización de dispersión de documentos).

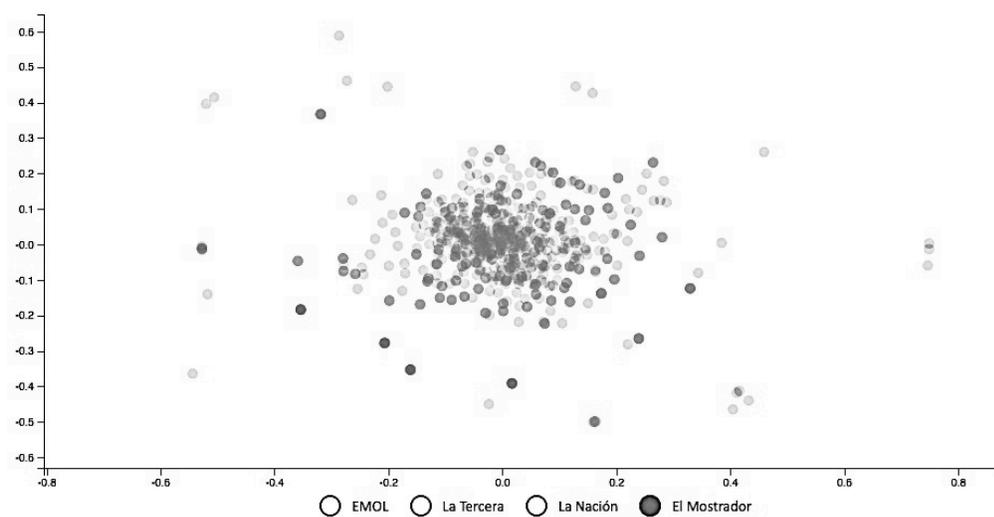


Figura IV-5: Visualización de dispersión de documentos

V. RESULTADOS

En este capítulo se presentan los resultados obtenidos de la recuperación de artículos de noticias y su posterior análisis de semántica latente. Partiendo desde los datos sin procesar, se presenta un análisis básico de cobertura a los candidatos, seguido de una visión global del análisis de semántica latente, destacando las relaciones significativas encontradas.

V.1. Cobertura

Dentro del grupo de artículos de noticia recuperado en el lapso entre el periodo de inscripción de candidaturas presidenciales y la primera elección presidencial, se calculó el porcentaje de menciones del nombre de un candidato en los títulos de los artículos respecto al total de artículos recuperados en el periodo.

Tabla V-1: Menciones de candidatos y votación final obtenida

number	EMOL	La Tercera	El Mostrador	La Nación	Votación
Bachelet	52.70%	49.79%	40.20%	40.63%	46.69%
Matthei	35.53%	19.01%	37.55%	35.70%	25.02%
MEO	12.27%	1.86%	7.96%	14.18%	10.98%
Parisi	10.40%	5.58%	7.76%	19.11%	10.11%
Claude	2.26%	0.21%	1.43%	4.57%	2.80%
Sfeir	0.88%	0.21%	2.45%	2.64%	2.80%
Miranda	0.98%	0.41%	3.88%	3.97%	1.25%
Israel	2.16%	0.21%	6.53%	4.69%	0.57%
Jocelyn-Holt	1.77%	0.21%	1.84%	5.89%	0.19%

En términos generales, la cobertura de prensa a cada candidato corresponde a la votación final obtenida (TRICEL 2013). En específico, los medios EMOL y La Tercera siguen un patrón cercano a la votación final (Figura V-1: Porcentaje de

menciones de candidatos en medios y votación obtenida). Por otro lado, El Mostrador y La Nación dieron una mayor cobertura relativa a los candidatos que resultaron con votación menor a 10%.

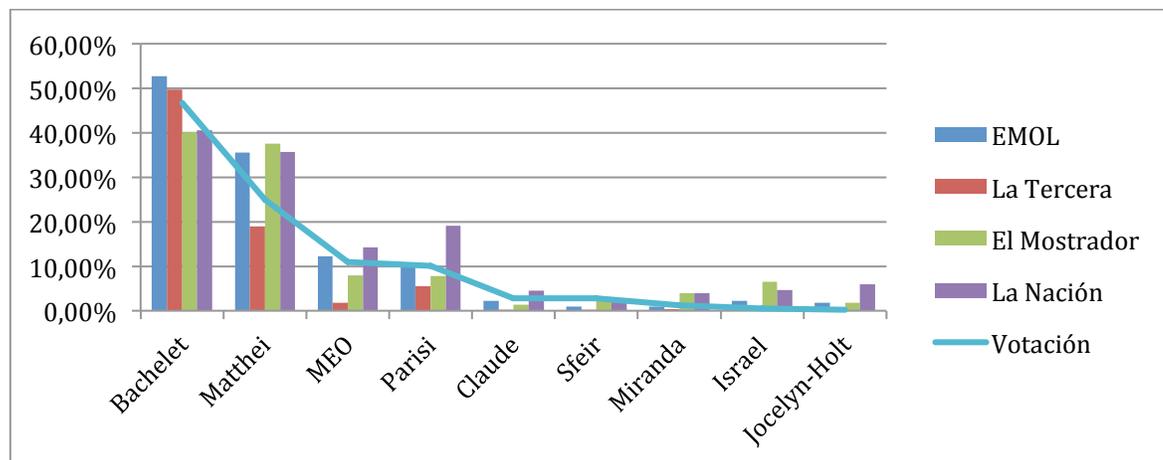


Figura V-1: Porcentaje de menciones de candidatos en medios y votación obtenida

V.2. Dispersión de documentos

A partir de la Matriz de documentos por dimensiones V^T , se realizó una reducción de dimensiones mediante MDS para generar una visualización de dispersión con el fin de comparar la similitud de documentos entre los distintos medios (Figura V-2: Dispersión de documentos por medio. y Figura V-3: Detalle de dispersión por medio (en sentido del reloj, desde arriba a la izquierda: EMOL, La Tercera, El Mostrador y La Nación).).

Para comparar la dispersión en un espacio bidimensional se utilizó la raíz de Distancia Media Cuadrada, DMC ($\hat{\sigma}$) (Ecuación 5.1), que compara la distancia de cada punto con cada elemento del centroide obtenido por los promedios en cada dimensión (\bar{x}, \bar{y}) .

$$\hat{\sigma} = \sqrt{\frac{1}{N} \sum_i ((x_i - \bar{x})^2 + (y_i - \bar{y})^2)} \quad (5.1)$$

El medio con menor similitud de documentos resultó ser El Mostrador con un valor de DMC de 0,29760607267, mientras que La Nación resultó ser el medio con mayor similitud, con un valor de 0,259546006303 (Tabla V-2: Dispersión de documentos en medios analizados.).

Tabla V-2: Dispersión de documentos en medios analizados.

	Todos	EMOL	La Tercera	La Nación	El Mostrador
$\hat{\sigma}$	0,28184	0,27562	0,28499	0,25954	0,2976
Centroide - X	2,47473E-17	-0,02694	-0,04297	0,0477	0,01309
Centroide - Y	3,79345E-18	0,01745	-0,07055	0,02459	-0,00602
Máximo - X	0,5833	0,54628	0,54852	0,5833	0,58104
Mínimo - X	-0,57421	-0,54396	-0,55816	-0,57421	-0,49414
Máximo - Y	0,68917	0,58355	0,68917	0,53278	0,60269
Mínimo - Y	-0,62885	-0,52535	-0,60923	-0,62885	-0,62215

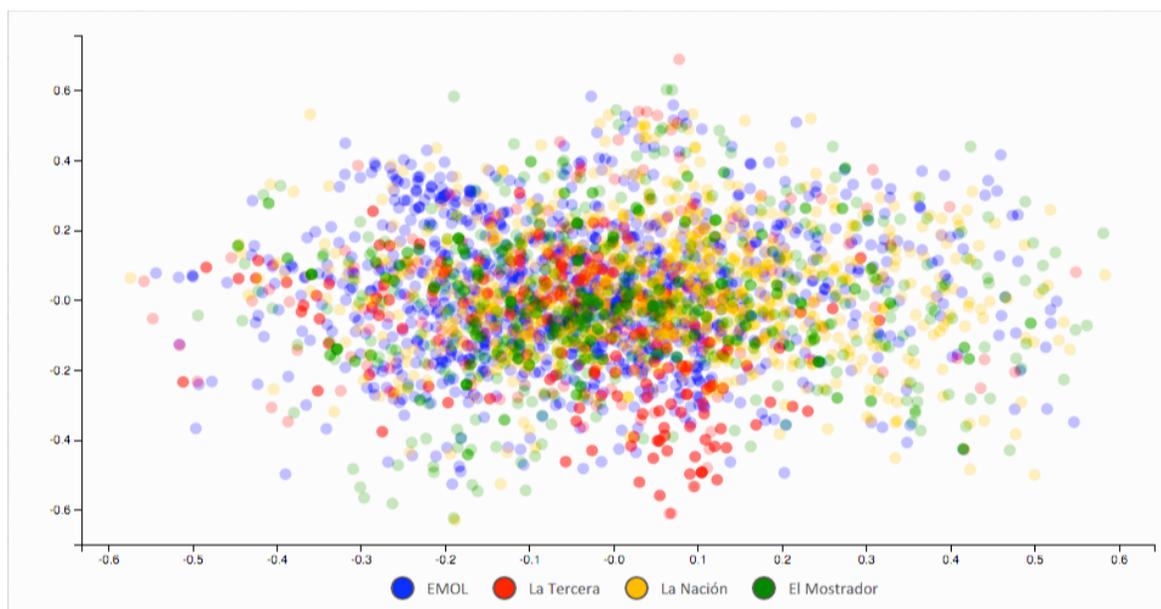


Figura V-2: Dispersión de documentos por medio.

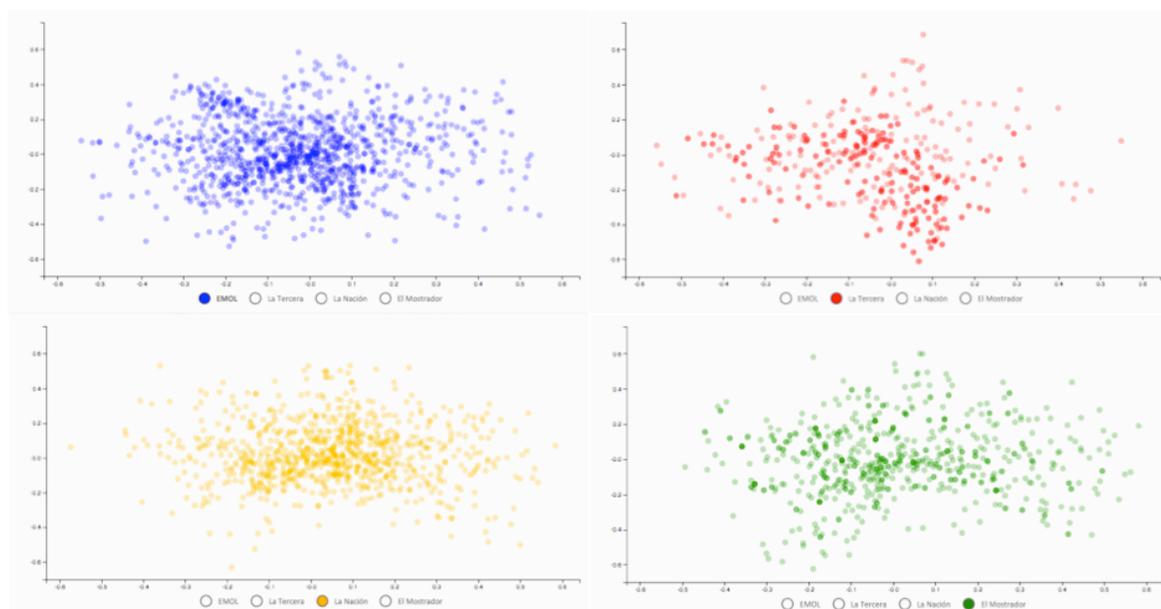


Figura V-3: Detalle de dispersión por medio (en sentido del reloj, desde arriba a la izquierda: EMOL, La Tercera, El Mostrador y La Nación).

V.1. Análisis de Semántica Latente

Para decidir una unidad sintagmática para el Análisis de Semántica Latente (LSA), se estudió la cantidad de caracteres por documento y por párrafo, con el fin de encontrar la unidad con mayor regularidad. Considerando la regularidad de la cantidad de caracteres por párrafo (Tabla V-3: Distribución de caracteres y párrafos.) y la distribución de párrafos por documentos (Figura V-4: Distribución de párrafos por documentos.), se eligió al párrafo como unidad sintagmática de análisis.

Tabla V-3: Distribución de caracteres y párrafos.

	Caracteres por Documentos	Caracteres por Párrafos	Párrafos por Documentos
σ	1.179,220	267,398	4,886
Media	2.087,380	293,674	7,143
Mediana	1.841	256	6
Máximo	16.778	12.608	72
Mínimo	354	4	1

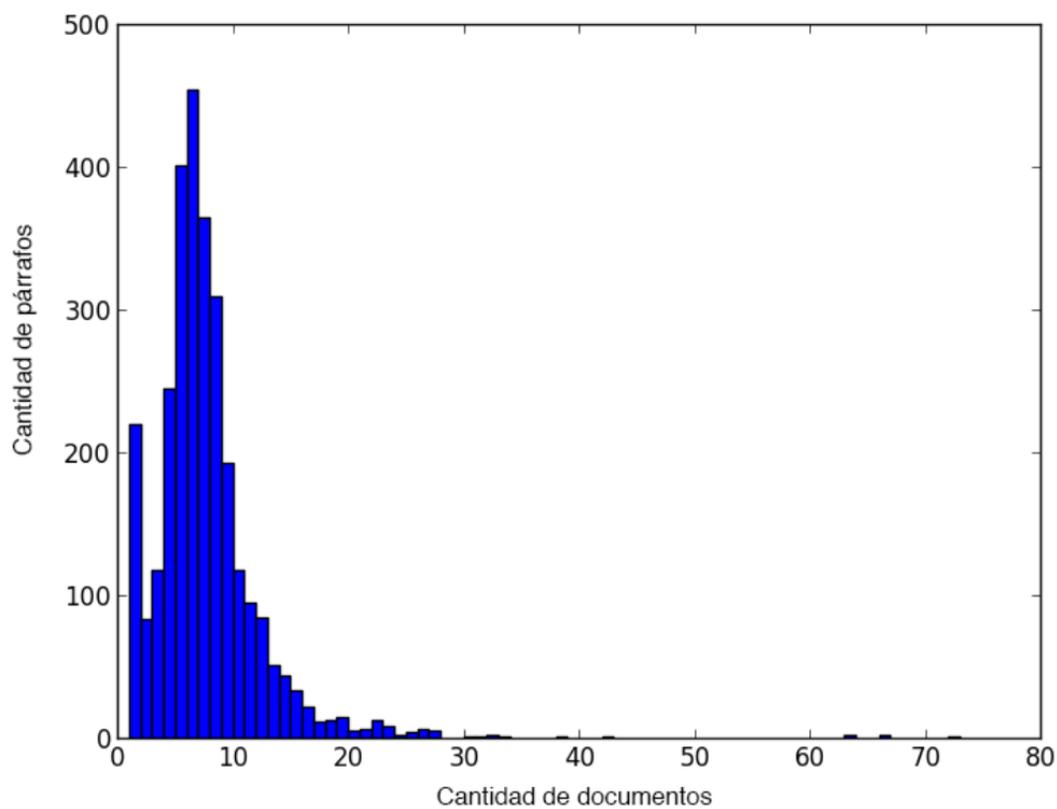


Figura V-4: Distribución de párrafos por documentos.

De esta manera, se ejecutó LSA por cada medio a su correspondiente corpus de noticias para diez dimensiones extrayendo los 10 términos más relevantes por tema. Posteriormente se generó un grafo con las relaciones entre todos los términos obtenidos, coloreando cada término del grafo según el tema al que pertenece cada uno.

V.1.1. EMOL

Los principales tópicos cubiertos por EMOL (Figura V-5: Campos semánticos de EMOL.), se listan a continuación:

- a) Definición de candidatos. Se define a Bachelet como candidata presidencial y ex mandataria perteneciente a la Nueva Mayoría. De manera similar, Evelyn Matthei, como candidata de la Alianza conformada por los partidos UDI y RN. Finalmente, aparecen mencionados también como candidatos Marco Enríquez-Ominami y Franco Parisi, sin detallar su afiliación política, excepto por la independencia de éste último.
- Contenidos programáticos. El principal contenido tratado corresponde a la reforma tributaria en relación con Bachelet, por otro lado, temas relativos a educación se encuentran relacionados principalmente a Enríquez-Ominami y Parisi.
- b) Debate. El debate con mayor relevancia es aquel organizado por ANATEL, la Asociación Nacional de Televisión.
- c) Conflicto. La relación entre Matthei y Parisi denota la cobertura del conflicto entre ambos.
- d) Segunda Vuelta. El último tema relevante es la segunda vuelta de la elección presidencial, centrándose en términos como “votación”, “ganador” y “elección”.

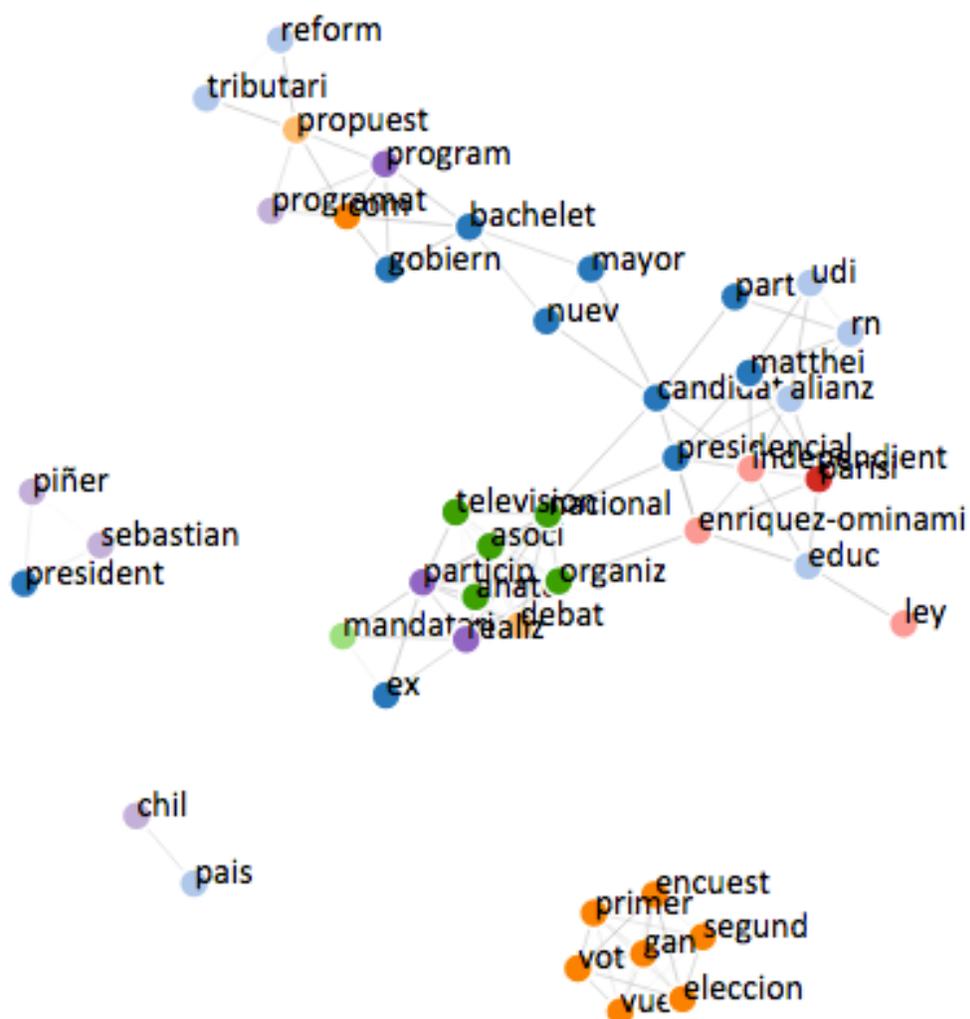


Figura V-5: Campos semánticos de EMOL.

V.1.2. La Tercera

Los principales temas en la cobertura de La Tercera (Figura V-6: Campos semánticos de La Tercera.) son los siguientes:

a) Definición de candidatos. Bachelet es definida nuevamente tanto como ex presidenta y candidata presidencial de la Nueva Mayoría, agregando su calidad como

ex directora de ONU Mujer, su relación con los partidos PPD y PS en contexto de apoyo para la segunda vuelta y el precandidato José Antonio Gómez de su misma coalición. Evelyn Matthei aparece relacionada a la Alianza y también con Andrés Allamand, precandidato por el mismo conglomerado político.

b) Contenidos programáticos. Los principales temas tratados son la reforma constitucional y reforma educacional, esta última en relación con el ministro de educación entonces Harald Beyer y su destitución.

Entre otros temas relevantes, destaca la cobertura a las elecciones primarias y las menciones al presidente durante ese momento, Sebastián Piñera. Sin embargo, de los términos que llevaron a la raíz “president”, la mayoría corresponde a “presidenta” (331 veces), contra “presidente” (250 veces).

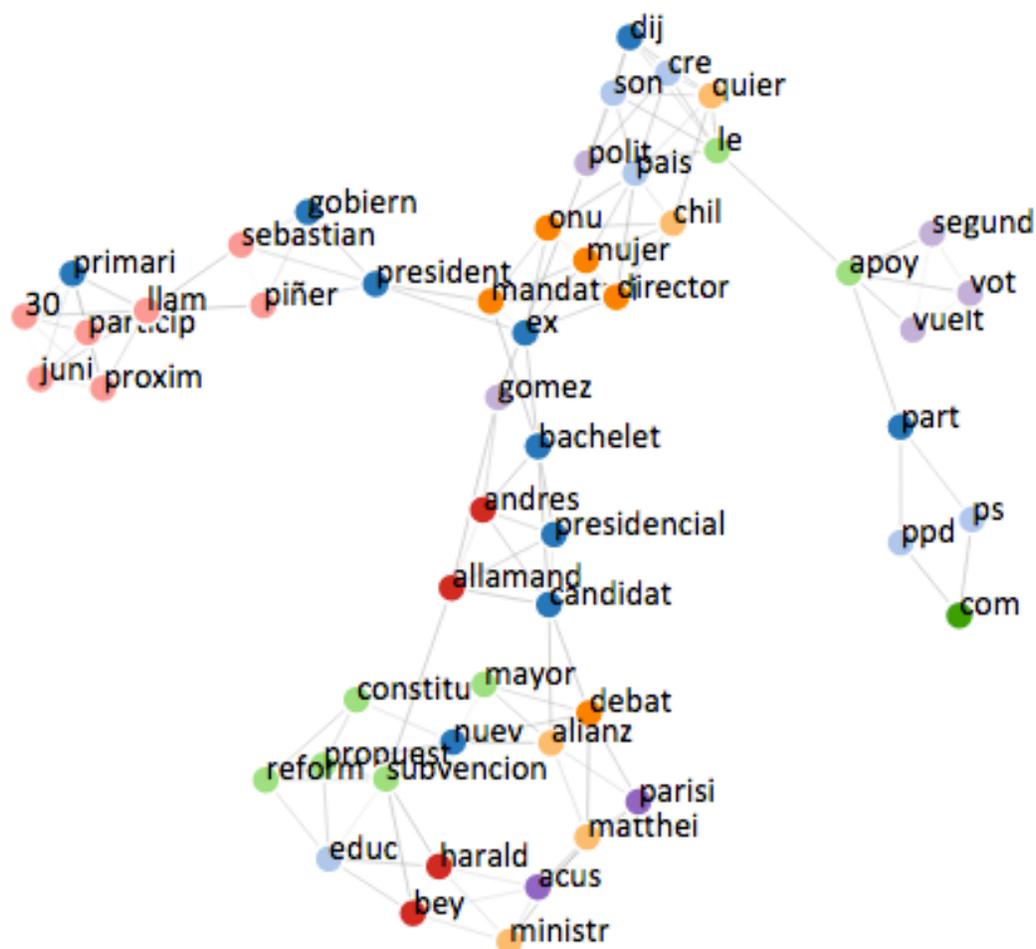


Figura V-6: Campos semánticos de La Tercera.

V.1.3. La Nación

Para el caso de La Nación (Figura V-7: Campos semánticos de La Nación.), los principales tópicos cubiertos son los siguientes:

- a) Definición de candidatos. Bachelet nuevamente es nombrada a partir de ser la candidata o abanderada de la nueva mayoría, igualmente que Evelyn Matthei respecto de la Alianza. Estrechamente ligado aparece Sebastián Piñera, presidente en

ese momento. Ligado a Matthei aparece Parisi debido al conflicto que tuvieron durante la campaña. Además, Parisi aparece caracterizado como independiente al mismo tiempo que guarda relación con Alfredo Sfeir, Ricardo Israel, Roxana Miranda y Marcel Claude, este último perteneciente al Partido Humanista.

b) Contenidos programáticos. El programa de Bachelet aparece relacionado principalmente con educación y economía. De manera similar, Parisi aparece relacionado con economía y colegios. Los temas de DDHH son relevantes pero sin relación a un candidato particular.

En otros temas relacionados aparecen los debates de Anatel y de Archi, también la segunda vuelta presidencial y, al igual que en los medios anteriores Sebastián Piñera.

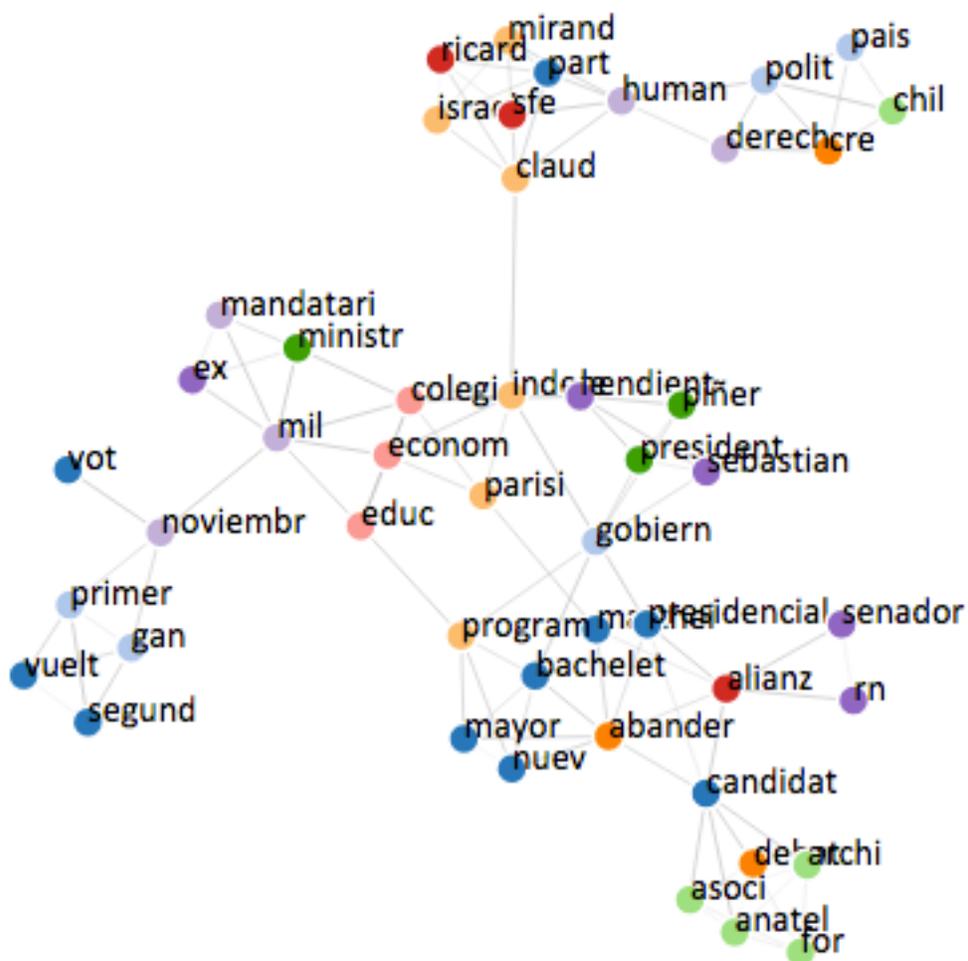


Figura V-7: Campos semánticos de La Nación.

V.1.4. El Mostrador

Finalmente, para El Mostrador (Figura V-8: Campos semánticos de El Mostrador.), los temas de cobertura más relevantes son:

- a) Definición de candidatos. Bachelet y Matthei son definidas en base a sus conglomerados, Nueva Mayoría y Alianza por Chile, respectivamente. Parisi a su

vez, como candidato independiente y muy relacionado a Marco Enríquez-Ominami y Evelyn Matthei.

b) Contenidos programáticos. Respecto a Bachelet, se dio principal cobertura a la reforma laboral y reforma educacional.

Finalmente, otros temas relevantes son relativos al debate Anatel y el paso de candidatos a la segunda vuelta.

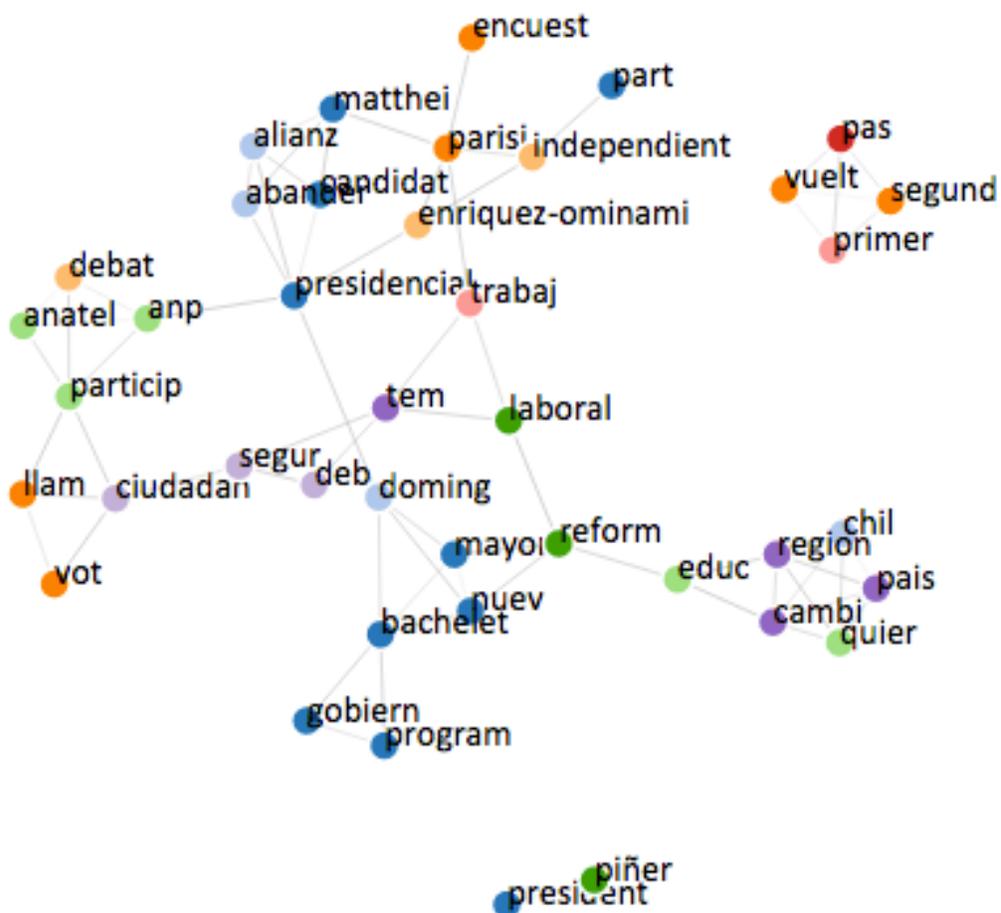


Figura V-8: Campos semánticos de El Mostrador.

VI. CONCLUSIONES

En este capítulo se presentan las conclusiones al trabajo realizado. Consiste en la verificación del cumplimiento del objetivo general referente a la detección de tendencia en los medios de comunicación analizados conforme a las categorías descritas en el Capítulo II y en la verificación de los objetivos específicos descritos durante los procesos de implementación y análisis (Capítulos IV y V).

VI.1. Objetivo General

Respecto a la hipótesis principal sobre la aparición de tendencias en medios de comunicación se puede comentar lo siguiente. En primer lugar, tal como se explicó en el Capítulo II, el primer tipo de tendencia es difícil de comprobar, sin embargo, se pudo observar una tendencia de cobertura en el análisis cuantitativo respecto al espacio que se dio a cada candidato (Tabla V-1: Menciones de candidatos y votación final obtenida) generalmente correspondiente en forma relativa a la votación general obtenida.

En segundo lugar, la tendencia estructural, referente a los contenidos de los artículos de noticias pudo ser observada a lo largo de los grafos semánticos obtenidos. Medios como EMOL y La Tercera mayor cobertura a los contenidos programáticos de Bachelet por sobre todos los otros candidatos, los cuales fueron mencionados en función de ella u otros candidatos, siendo caso representativo el de Matthei que recibió fuerte cobertura sobre su conflicto con Parisi. La Nación y El Mostrador, por otro lado, dieron mayor cobertura a los candidatos que finalmente recibieron menor

votación, incluyendo también contenidos programáticos. Todo esto demuestra las diferencias de contenido en la cobertura de los candidatos a la elección, confirmando la tendencia estructural de todos los medios.

VI.2. Objetivos específicos

a) Grafo semántico

Los grafos semánticos obtenidos entregaron una representación visual que enriquece el análisis de LSA al complementar la lista de términos relevantes por temas con la relación semántica de todos los términos entre sí, observando patrones que quedan fuera de los grupos temáticos, generando una red de términos que destaca tanto las relaciones semánticas y la pertenencia a grupos temáticos. Esto resulta evidente en casos cuando términos relevantes de un tema no presentan relaciones semánticas fuertes entre otros términos del mismo tema resultando incluso a veces en grupos separados visualmente.

b) Similitud de documentos

Los cálculos de DMC arrojaron que los documentos analizados por cada medio presentan distintos grados de similitud, siendo menor en El Mostrador y mayor en La Nación. Si bien, la primera impresión de la visualización de documentos no permite encontrar una clara separación en categorías deseables (por ejemplo temáticas), es de utilidad para detectar artículos que a pesar de ser recuperados y compartir las palabras clave tienen contenidos ajenos a la temática principal (elecciones presidenciales en este caso), un ejemplo son las noticias referentes a los generales Matthei y Bachelet que aparecieron en EMOL.

c) Cuantificación de cobertura

El análisis cuantitativo de la cobertura arrojó, como se describió en el subcapítulo anterior, que comprueba que existe tendencia de cobertura en todos los medios de comunicación al notar que los candidatos con mayor votación obtenida fueros los que coincidentemente obtuvieron mayor cobertura. Casos a destacar son los de Enríquez-Ominami, el cual tuvo una cantidad de menciones considerablemente menor en La Tercera que otros medios y también Parisi con una cobertura mayor en La Nación que en los otros medios.

VI.3. Trabajos Futuros

Complementando el análisis de similitud de documentos, el proyecto de título de María Paz Espinoza utilizará la herramienta ViTA-SSD (Visual Text Analytics tool for Semi-Structured Documents) desarrollada por la Universidad de Dalhousie, con el fin de encontrar tendencia en medios de comunicación mediante el análisis visual de similitud de documentos que provee esta herramienta sobre el mismo set de datos confeccionado para este trabajo.

BIBLIOGRAFIA

- D'Alessio, D. & Allen, M., 2000. Media bias in presidential elections: a meta-analysis. *Journal of Communication*, 50(4), pp.133–156. Recuperado desde: <http://doi.wiley.com/10.1111/j.1460-2466.2000.tb02866.x>.
- Deerwester, S. et al., 1998. Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science and Technology*.
- Fortuna, B., Grobelnik, M. & Mladenić, D., 2005. Visualization of Text Document Corpus. *Informatica*, 29, pp.497–502.
- Fundación Imagen de Chile, 2013. Las elecciones de Chile en los medios internacionales. Recuperado desde: <http://www.imagendechile.cl/wp-content/uploads/2013/11/Elecciones-en-Chile-20112013-VF.pdf>.
- Jorge-Botana, G. et al., 2010. Visualizing polysemy using LSA and the predication algorithm. *Journal of the American Society for Information Science and Technology*, 2010, p.n/a–n/a. Recuperado desde: <http://doi.wiley.com/10.1002/asi.21355>.
- Jorge-Botana, G., León, J.A. & Olmos, R., 2009. Different Senses Visualization With Latent Semantic Analysis. En *XXVII Spanish Applied Linguistics Association (AESLA) International Conference. Ciudad Real 2009*.
- Junqué de Fortuny, E. et al., 2012. Media coverage in times of political crisis: A text mining approach. *Expert Systems with Applications*, 39(14), pp.11616–11622. Recuperado desde: <http://linkinghub.elsevier.com/retrieve/pii/S0957417412006100> [Accedido octubre 2, 2013].
- Landauer, T.K., 1998. Learning and representing verbal meaning: The latent semantic analysis theory B. A. Spellman & D. T. Willingham, eds. *Current Directions in Psychological Science*, 7(5), pp.161–164. Recuperado desde: <http://www.blackwell-synergy.com/doi/abs/10.1111/1467-8721.ep10836862?journalCode=cdir>.
- Landauer, T.K., 2002. On the computational basis of learning and cognition: Arguments from LSA. In N. Ross (Ed.), *The psychology of learning and motivation*, 41, 2002, (1), pp.43–84. Recuperado desde: <papers3://publication/uuid/78891F9D-A9BC-4E14->

A2E8-825114DE4CCB.

- Landauer, T.K. & Dutnais, S.T., 1997. A Solution to Plato ' s Problem : The Latent Semantic Analysis Theory of Acquisition , Induction , and Representation of Knowledge. , 1(2), pp.211–240.
- Lund, K. & Burgess, C., 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, 28(2), pp.203–208. Recuperado desde:
<http://www.springerlink.com/index/10.3758/BF03204766>.
- Marín, M., Mendoza, M., Rodríguez, A., Contreras, A., Caro, D., Cabrera L, and Sanhueza, C., 2013. Observatorio Político: herramienta de análisis de las primarias en tiempo real. Recuperado desde: <http://www.dgt.usach.cl/noticias/observatorio-politico-herramienta-de-analisis-de-las-primarias-en-tiempo-real>.
- Mendoza, M., 2014. Minería de Opiniones en Elecciones Presidenciales. *Bits de Ciencia, Revista del Departamento de Ciencias de la Computación de la Universidad de Chile*, pp.46–47.
- NLTK Project, 2015. NLTK 3.0 Documentation. Recuperado desde: <http://www.nltk.org/>.
- Rogel, A.M., del Valle, C. & Nitrihual, L., 2010. Concentración de la propiedad de los medios de comunicación en Chile: La compleja relación entre oligopolio y democracia. *Anagramas*, 9(17), pp.131–148. Recuperado desde:
http://www.scielo.org.co/scielo.php?pid=S1692-25222010000200011&script=sci_abstract.
- Rosario, B., 2000. Latent Semantic Indexing : An overview. *INFOSYS 240*, pp.1–16.
- Sahlgren, M., 2008. The distributional hypothesis. *Rivista di Linguistica (Italian Journal of Linguistics)*, 20(1), pp.33–54.
- Sapiezyńska, E., Lagos, C. & Cabalín, C., 2013. Press Freedom under Pressure: Restriction Levels Perceived by Chilean Journalists and Influential Factors. *Cuadernos de Información*, 32, pp.11–26. Recuperado desde:
<http://cuadernos.uc.cl/uc/index.php/CDI/article/view/468>.

- Tercera, L., 2013. Las menciones versus sus vocerías de los candidatos presidenciales en los medios. Recuperado desde:
<http://www.latercera.com/noticia/politica/2013/06/674-528855-9-las-menciones-versus-sus-vocerias-de-los-candidatos-presidenciales-en-los-medios.shtml>.
- TRICEL, 2013. Escrutinio y calificación de la elección presidenciales en segunda vuelta 2013. Recuperado desde: [http://www.tricel.cl/informacioncausas/Tramitacion de Causas/156-2013 02.- Sentencia 02-12-2013.pdf](http://www.tricel.cl/informacioncausas/Tramitacion%20de%20Causas/156-2013%2002.-%20Sentencia%2002-12-2013.pdf).
- Veltri, G. a., 2012. Viva la Nano-Revolucion! A Semantic Analysis of the Spanish National Press. *Science Communication*, 35(2), pp.143–167. Recuperado desde: <http://scx.sagepub.com/cgi/doi/10.1177/10755470124440353> [Accedido septiembre 2, 2013].
- Venegas V., R., 2003. Análisis Semántico Latente: una panorámica de su desarrollo. *Revista signos*, 36(53), pp.121–138. Recuperado desde: http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-09342003005300008&lng=en&nrm=iso&tlng=en [Accedido septiembre 3, 2013].
- Zhu, W. & Chen, C., 2007. Storylines: Visual exploration and analysis in latent semantic spaces. *Computers & Graphics*, 31(3), pp.338–349. Recuperado desde: <http://linkinghub.elsevier.com/retrieve/pii/S0097849307000568>.

ANEXOS

ANEXO A : TABLAS LSA EMOL

Tema	Termino	Puntuación	Frecuencias
0	candidat	0.18095630616624858	candidata:1166 candidato:647 candidatos:453 candidatas:44 candidataen:2
0	bachelet	0.15867173109918253	bachelet:1751 bacheleten:5 bacheletista:1 bacheletantes:1
0	nuev	0.15094899709502327	nueva:855 nuevo:178 nuevos:68 nueve:58 nuevamente:51 nuevas:42 nueves:1
0	presidencial	0.14343939246415283	presidencial:1079 presidenciales:181 presidencialen:3 presidencialismo:1 presidencialantes:1
0	matthei	0.14234584474558276	matthei:1317 mattheien:2 mattheia:1
0	mayor	0.13115778631013827	mayoría:741 mayor:193 mayores:60 mayorías:6 mayormente:1
0	ex	0.12938052249364754	ex:1210
0	part	0.1291310278440983	partido:471 parte:438 partidos:199 partir:55 partes:12 partida:7 partió:7 partiendo:5 parta:4 partidas:4 partirá:3 partimos:3 parten:2 parto:1 partieron:1 partamos:1
0	gobiern	0.12502347430336899	gobierno:1056 gobierne:5 gobierna:5 gobiernan:2 gobiernoen:1 gobiernen:1
0	president	0.12071188541099755	presidente:553 presidenta:339 presidencia:85 presidentes:49 presidentaen:1 presidentas:1
1	reform	-0.25651143899154888	reforma:252 reformas:94 reformar:12 reforman:1 reformará:1
1	presidencial	0.2210426186366829	presidencial:1079 presidenciales:181 presidencialen:3 presidencialismo:1 presidencialantes:1
1	tributari	-0.17778291951670189	tributaria:172 tributarias:15 tributario:11 tributarios:4
1	educ	-0.177100680921631	educación:346 educativo:7 educativa:5 educadoras:3 educador:2 educativos:2 educan:1 educadores:1 educado:1 educadora:1 educados:1 educarse:1
1	matthei	0.16715581042943611	matthei:1317 mattheien:2 mattheia:1
1	candidat	0.16028758790488351	candidata:1166 candidato:647 candidatos:453 candidatas:44

			candidataen:2
1	rn	0.15149190125907097	rn:419
1	alianz	0.1499026025682266	alianza:618 alianzas:6 alianzaen:1
1	udi	0.14773950304446015	udi:401 udien:2
1	pais	-0.14236462985496498	país:681 países:45 paísen:2
2	vuel	0.38379569793500606	vuelta:431 vuelta:6 vueltas:5 vueltaen:2
2	vot	0.37312475675377837	votar:220 votos:132 voto:96 votaron:25 votó:20 voten:17 vote:12 votan:12 votado:10 vota:9 votaría:9 votará:5 votando:5 votv©:4 votamos:3 voter:3 votada:2 votarán:2 votarv©:2 votaban:1 votaba:1 votara:1 votarían:1 votaran:1 votaremos:1 votosen:1
2	segund	0.29320012276102869	segunda:370 segundo:88 segundos:18 segundaen:1
2	gan	0.20343441934572284	ganar:144 gana:27 gane:21 ganan:19 ganado:17 ganas:14 ganó:14 ganamos:11 ganada:10 ganando:9 ganarle:6 ganará:6 gano:3 ganen:3 ganadas:2 ganaba:2 ganarlas:1 ganarse:1 ganaban:1 ganara:1 ganaron:1 ganados:1 ganemos:1 ganarán:1 ganaremos:1 ganaran:1
2	eleccion	0.17963864458633105	elecciones:237 elección:221 eleccion:1
2	encuest	0.17591208533584168	encuesta:153 encuestas:86 encuestados:8 encuestadas:3 encuestadoras:1 encuestasen:1
2	primer	0.16702047535258663	primera:268 primeros:53 primeras:17
2	com	-0.12054553586377102	comando:487 comida:9 comer:3 come:1 comidas:1
2	gobiern	-0.12003907388165079	gobierno:1056 gobierne:5 gobierna:5 gobiernan:2 gobiernoen:1 gobiernen:1
2	rn	-0.11248512958015974	rn:419
3	nuev	0.37557804382569054	nueva:855 nuevo:178 nuevos:68 nueve:58 nuevamente:51 nuevas:42 nueves:1
3	mayor	0.324262351359065	mayoría:741 mayor:193 mayores:60 mayorías:6 mayormente:1
3	bachelet	0.23774232088655436	bachelet:1751 bacheleten:5 bacheletista:1 bacheletantes:1

3	reform	0.23283244404397246	reforma:252 reformas:94 reformar:12 reforman:1 reformará:1
3	tributari	0.17361107630856584	tributaria:172 tributarias:15 tributario:11 tributarios:4
3	rn	-0.16343581122474848	rn:419
3	matthei	-0.15198198916213929	matthei:1317 mattheien:2 mattheia:1
3	propuest	0.14541131934660123	propuestas:316 propuesta:233 propuesto:31 propuestos:5
3	udi	-0.14317037533333885	udi:401 udien:2
3	debat	0.13234125724527285	debate:338 debates:131 debatir:59 debatamos:4 debatiendo:3 debatidas:1 debatía:1 debatido:1 debatirlas:1 debativ©ramos:1 debatió:1 debatan:1 debateen:1 debatimos:1 debata:1
4	debat	0.60130372281228894	debate:338 debates:131 debatir:59 debatamos:4 debatiendo:3 debatidas:1 debatía:1 debatido:1 debatirlas:1 debativ©ramos:1 debatió:1 debatan:1 debateen:1 debatimos:1 debata:1
4	asoci	0.1907759290324138	asociación:89 asociados:5 asociaciones:3
4	anatel	0.17623048567220992	anatel:84
4	nacional	0.14660830403890335	nacional:321 nacionales:19 nacionalen:1 nacionalista:1
4	television	0.14191878579587106	televisión:81
4	ex	-0.13437486328331591	ex:1210
4	organiz	0.13093873700619899	organizado:61 organizaciones:22 organización:22 organizada:7 organizando:6 organizar:4 organizadores:4 organizados:3 organiza:3 organizó:3 organizará:2 organizarnos:1 organizamos:1 organizarlo:1 organizadas:1 organizan:1
4	president	-0.1240084826106585	presidente:553 presidenta:339 presidencia:85 presidentes:49 presidentaen:1 presidentas:1
4	vot	-0.12016623008897329	votar:220 votos:132 voto:96 votaron:25 votó:20 voten:17 vote:12 votan:12 votado:10 vota:9 votaría:9 votará:5 votando:5 votv©:4 votamos:3 voter:3 votada:2 votarán:2 votarv©:2 votaban:1 votaba:1 votara:1 votarían:1 votaran:1 votaremos:1 votosen:1

4	gobiern	-0.1106917693512421	gobierno:1056 gobierne:5 gobierna:5 gobiernan:2 gobiernoen:1 gobiernen:1
5	reform	-0.35384241213553902	reforma:252 reformas:94 reformar:12 reforman:1 reformará:1
5	tributari	-0.27522111849927722	tributaria:172 tributarias:15 tributario:11 tributarios:4
5	rn	-0.19366854928728647	rn:419
5	udi	-0.1885955562954067	udi:401 udien:2
5	educ	-0.17834848364992811	educación:346 educativo:7 educativa:5 educadoras:3 educador:2 educativos:2 educan:1 educadores:1 educado:1 educadora:1 educados:1 educarse:1
5	ex	0.17434245932917997	ex:1210
5	mandatari	0.16181604530579694	mandataria:347 mandatario:56 mandatarios:4
5	matthei	-0.15955048729244545	matthei:1317 mattheien:2 mattheia:1
5	presidencial	-0.15723415066557084	presidencial:1079 presidenciales:181 presidencialen:3 presidencialismo:1 presidencialantes:1
5	gobiern	0.14309843108467421	gobierno:1056 gobierne:5 gobierna:5 gobiernan:2 gobiernoen:1 gobiernen:1
6	nuev	0.22829564338062167	nueva:855 nuevo:178 nuevos:68 nueve:58 nuevamente:51 nuevas:42 nueves:1
6	mayor	0.20768453328636038	mayoría:741 mayor:193 mayores:60 mayorías:6 mayormente:1
6	vuelt	-0.19490701706507557	vuelta:431 vuelto:6 vueltas:5 vueltaen:2
6	part	0.18693891283887459	partido:471 parte:438 partidos:199 partir:55 partes:12 partida:7 partió:7 partiendo:5 parta:4 partidas:4 partirá:3 partimos:3 parten:2 parto:1 partieron:1 partamos:1
6	segund	-0.18671166969124633	segunda:370 segundo:88 segundos:18 segundaen:1
6	gobiern	-0.17867377205369167	gobierno:1056 gobierne:5 gobierna:5 gobiernan:2 gobiernoen:1 gobiernen:1
6	parisi	-0.16820908653115127	parisi:420 parisien:2 parisia:1
6	encuest	-0.15147436367617795	encuesta:153 encuestas:86 encuestados:8 encuestadas:3 encuestadoras:1 encuestasen:1
6	matthei	-0.133610470185684	matthei:1317 mattheien:2 mattheia:1
6	rn	0.13076966888093766	rn:419

7	educ	-0.2953125026923667	educación:346 educativo:7 educativa:5 educadoras:3 educador:2 educativos:2 educan:1 educadores:1 educado:1 educadora:1 educados:1 educarse:1
7	reform	0.29288505474083781	reforma:252 reformas:94 reformar:12 reforman:1 reformará:1
7	debat	0.28847349145985313	debate:338 debates:131 debatir:59 debatamos:4 debatiendo:3 debatidas:1 debatía:1 debatido:1 debatirlas:1 debatir@ramos:1 debatió:1 debatan:1 debateen:1 debatimos:1 debata:1
7	tributari	0.23166015949760377	tributaria:172 tributarias:15 tributario:11 tributarios:4
7	parisi	-0.1960022813672796	parisi:420 parisien:2 parisia:1
7	mayor	-0.14905129446322118	mayoría:741 mayor:193 mayores:60 mayorías:6 mayormente:1
7	independient	-0.14784276665314078	independiente:261 independientes:87 independientemente:1
7	nuev	-0.137399053534076	nueva:855 nuevo:178 nuevos:68 nueve:58 nuevamente:51 nuevas:42 nueves:1
7	enriquez-ominami	-0.1143699925759962	enríquez-ominami:228 enriquez-ominami:3 enriquez-ominamies:1
7	ley	-0.10286904255266764	ley:288 leyes:46 leyó:1 leyen:1
8	ex	0.39462445893790499	ex:1210
8	mandatari	0.27660051895767479	mandataria:347 mandatario:56 mandatarios:4
8	gobiern	-0.24898047695271927	gobierno:1056 gobierne:5 gobierna:5 gobiernan:2 gobiernoen:1 gobiernen:1
8	program	-0.22144327561806534	programa:538 programas:52 programado:9 programada:2 programación:1 programó:1
8	educ	0.1909029594097951	educación:346 educativo:7 educativa:5 educadoras:3 educador:2 educativos:2 educan:1 educadores:1 educado:1 educadora:1 educados:1 educarse:1
8	mayor	-0.1862106695401623	mayoría:741 mayor:193 mayores:60 mayorías:6 mayormente:1
8	nuev	-0.18162796984544044	nueva:855 nuevo:178 nuevos:68 nueve:58 nuevamente:51 nuevas:42 nueves:1

8	vot	0.15077599596738886	votar:220 votos:132 voto:96 votaron:25 votó:20 voten:17 vote:12 votan:12 votado:10 vota:9 votaría:9 votará:5 votando:5 votv©:4 votamos:3 voter:3 votada:2 votarán:2 votarv©:2 votaban:1 votaba:1 votara:1 votarían:1 votaran:1 votaremos:1 votosen:1
8	realiz	0.13538110366122516	realizó:122 realizado:84 realizar:69 realizará:42 realizada:24 realización:24 realizadas:23 realizarán:17 realiza:16 realizaron:15 realizados:9 realizan:8 realizarse:8 realizando:6 realizaremos:3 realizaba:3 realizaría:2 realizamos:2 realizarlo:2 realizables:1 realizaran:1 realizador:1 realizas:1 realizable:1
8	particip	0.11833376692735	participación:153 participar:111 participó:29 participaron:28 participarán:15 participará:14 participado:11 participantes:10 participan:7 participe:7 participativa:7 participa:6 participando:6 participen:5 participativo:5 participamos:2 participaba:2 participara:2 participativos:1 participarían:1 participaron:1 participaríamos:1 participaban:1 participv©:1
9	president	0.34434348346068783	presidente:553 presidenta:339 presidencia:85 presidentes:49 presidentaen:1 presidentas:1
9	piñer	0.27280807700241821	piñera:264 piñeraen:1 piñeraante:1
9	program	-0.24080722911685612	programa:538 programas:52 programado:9 programada:2 programación:1 programó:1
9	sebastian	0.20820598317368114	sebastián:159
9	propuest	-0.19757561898770382	propuestas:316 propuesta:233 propuesto:31 propuestos:5
9	com	-0.18418099703740912	comando:487 comida:9 comer:3 come:1 comidas:1
9	chil	0.18349029368644895	chile:919 chileen:1
9	programat	-0.16530740030107663	programático:87 programáticas:32 programáticos:31 programática:24

			votar:220 votos:132 voto:96 votaron:25 votó:20 voten:17 vote:12 votan:12 votado:10 vota:9 votaría:9 votará:5 votando:5 votv©:4 votamos:3 voter:3 votada:2 votarán:2 votarv©:2 votaban:1 votaba:1 votara:1 votarían:1 votaran:1 votaremos:1 votosen:1
9	vot	-0.15194991226975554	
9	rn	-0.12648181918264878	rn:419

ANEXO B : TABLAS LSA LA TERCERA

Tema	Termino	Puntuación	Frecuencias
0	candidat	0.20493447145112231	candidata:519 candidato:293 candidatos:173 candidatas:11
0	bachelet	0.20091661395096064	bachelet:1323 bacheletismo:4 bacheletista:3 bacheletistas:1
0	ex	0.17547022145158281	ex:857
0	president	0.16350069176181387	presidenta:331 presidente:250 presidencia:48 presidentes:23
0	gobiern	0.15040568719780689	gobierno:591 gobierne:3 gobiernen:1
0	presidencial	0.14372134026917996	presidencial:359 presidenciales:82
0	part	0.1384543297041943	parte:210 partido:197 partidos:135 partir:29 partes:6 partió:6 partida:4 partirá:2 parta:2
0	dij	0.11936119231272238	dijo:504 dije:25 dijimos:1
0	primari	0.11609364227623696	primarias:262 primaria:46 primariasen:2
0	nuev	0.11168022475383954	nueva:234 nuevo:65 nuevas:22 nuevos:22 nuevamente:17 nueve:14
1	ppd	0.19365004514716991	ppd:180
1	presidencial	0.1897459759758704	presidencial:359 presidenciales:82
1	educ	-0.18405217703244531	educación:232 educativa:4 educativo:2 educadora:1
1	ps	0.18383193081982507	ps:188
1	ex	0.16908250045711282	ex:857
1	cre	-0.15796741464565375	creo:172 cree:38 creemos:26 crear:21 creen:16 creer:10 creerle:6 creada:6 creado:6 crea:5 creando:3 crearemos:3 creían:2 creaba:2 creó:2 creía:2 crearía:1 creará:1 crearon:1 crees:1
1	primari	0.14929273446408975	primarias:262 primaria:46 primariasen:2
1	gobiern	-0.14366287725683341	gobierno:591 gobierne:3 gobiernen:1
1	pais	-0.13754330454808453	país:299 países:35 páis:2 pais:1
1	son	-0.12397538157673954	son:277 sonido:1
2	ex	-0.29570855387110884	ex:857
2	candidat	0.29140631378979115	candidata:519 candidato:293 candidatos:173 candidatas:11
2	presidencial	0.2473675756259385	presidencial:359 presidenciales:82
2	primari	0.24653145931792994	primarias:262 primaria:46 primariasen:2
2	president	-0.20983073381205705	presidenta:331 presidente:250 presidencia:48 presidentes:23
2	onu	-0.19459305303461821	onu:121

2	mujer	-0.19439743418330577	mujeres:192
2	debat	0.16079914873693027	debate:190 debates:45 debatir:12 debatido:2
2	mandatari	-0.14824996635362389	mandataria:250 mandatario:46 mandatarios:1
2	director	-0.14745660735049534	directora:62 director:29 directores:2
3	primari	0.3296383214764369	primarias:262 primaria:46 primariasen:2
3	matthei	-0.22701004995308949	matthei:393
3	candidat	-0.2129886232664438	candidata:519 candidato:293 candidatos:173 candidatas:11
3	ministr	-0.20966244953544511	ministro:197 ministra:161 ministros:35 ministras:4
3	pais	0.19137349405838616	país:299 países:35 país:2 pais:1
3	gobiern	-0.15913284672731337	gobierno:591 gobierne:3 gobiernen:1
3	quier	0.15491325789151786	quiero:111 quiere:76 quieren:30 quiera:10 quieran:6 quieres:1
3	presidencial	-0.14728733907766317	presidencial:359 presidenciales:82
3	chil	0.14107623228388275	chile:387
3	alianz	-0.12479359679437182	alianza:148
4	mujer	0.22231937863216342	mujeres:192
4	chil	0.21977850129892942	chile:387
4	presidencial	0.20962121157528504	presidencial:359 presidenciales:82
4	onu	0.20258564490762915	onu:121
4	pais	0.18230574129225177	país:299 países:35 país:2 pais:1
4	com	-0.17976917969231104	comando:318 comado:2 comida:2 comer:1
4	ps	-0.17261536297589267	ps:188
4	candidat	0.16006043008663928	candidata:519 candidato:293 candidatos:173 candidatas:11
4	director	0.14539357065800837	directora:62 director:29 directores:2
4	primari	-0.14203348442964661	primarias:262 primaria:46 primariasen:2
5	propuest	0.31101883055241925	propuestas:131 propuesta:119 propuesto:2
5	educ	0.31071463079877243	educación:232 educativa:4 educativo:2 educadora:1
5	nuev	0.30272562886516258	nueva:234 nuevo:65 nuevas:22 nuevos:22 nuevamente:17 nueve:14
5	reform	0.20883521128129309	reforma:100 reformas:40 reformar:3 reformistas:1 reformó:1
5	mayor	0.16755175195713387	mayoría:182 mayor:61 mayores:25 mayorías:2

5	apoy	-0.16481412482589955	apoyo:83 apoyar:38 apoyaron:9 apoyando:9 apoyado:8 apoyó:6 apoyan:4 apoyará:3 apoyos:3 apoyaba:2 apoyaría:2 apoyamos:2 apoyaran:2 apoye:1 apoyarla:1 apoya:1
5	president	-0.14436665120321854	presidenta:331 presidente:250 presidencia:48 presidentes:23
5	constitu	0.13304617856870482	constitución:43 constituye:7 constituir:4 constituirse:3 constituirá:2 constituyen:2 constituiría:1
5	subvencion	0.1279990839621343	subvencionada:22 subvención:21 subvencionados:5 subvenciones:4 subvencionadas:4 subvenciona:2 subvencionar:1
5	le	-0.12025602518999891	le:299
6	educ	-0.35439338207069315	educación:232 educativa:4 educativo:2 educadora:1
6	nuev	0.27893754934963499	nueva:234 nuevo:65 nuevas:22 nuevos:22 nuevamente:17 nueve:14
6	ministr	-0.242697111142498468	ministro:197 ministra:161 ministros:35 ministras:4
6	primari	-0.22612825544361997	primarias:262 primaria:46 primariasen:2
6	mayor	0.18149576532412789	mayoría:182 mayor:61 mayores:25 mayorías:2
6	bey	-0.16762128088752989	beyer:77
6	allamand	-0.16396755276277669	allamand:107
6	andres	-0.15120517059380489	andrés:128
6	subvencion	-0.13525035613996475	subvencionada:22 subvención:21 subvencionados:5 subvenciones:4 subvencionadas:4 subvenciona:2 subvencionar:1
6	harald	-0.13193276115395222	harald:33
7	gobiern	-0.27543160073521405	gobierno:591 gobierne:3 gobiernen:1
7	primari	-0.25059814137395869	primarias:262 primaria:46 primariasen:2
7	juni	-0.2179315615830229	junio:82
7	piñer	-0.21407585973257293	piñera:179
7	30	-0.21236219133600681	30:78
7	proxim	-0.17690765661458979	próximo:100 próxima:20 próximas:14
7	sebastian	-0.16222306171122361	sebastián:95
7	president	-0.15705475028215579	presidenta:331 presidente:250 presidencia:48 presidentes:23

7	particip	-0.14094174210195218	participación:79 participar:70 participara:7 participó:7 participaron:7 participa:7 participado:6 participativo:4 participará:4 participen:3 participemos:2 participativos:2 participarán:2 participe:2 participando:1 participaran:1 participantes:1 participan:1
7	llam	-0.13848404103564071	llamado:86 llamó:14 llamados:14 llamar:6 llamo:4 llamadas:3 llama:3 llaman:2 llamaron:2 llamará:2 llamando:2 llamada:1
8	matthei	-0.25952910653875555	matthei:393
8	gobiern	0.24594893293594089	gobierno:591 gobierne:3 gobiernen:1
8	ministr	-0.21877238880627287	ministro:197 ministra:161 ministros:35 ministras:4
8	acus	-0.20282044786851122	acusación:69 acusaciones:25 acusó:17 acusando:6 acusar:3 acusadores:2 acusarme:2 acusara:2 acusarlas:2 acusan:2 acusa:1
8	propuest	0.17620742396480951	propuestas:131 propuesta:119 propuesto:2
8	bey	-0.14373813968055368	beyer:77
8	parisi	-0.13595716685338463	parisi:116
8	debat	-0.11765384535226257	debate:190 debates:45 debatir:12 debatido:2
8	harald	-0.11361309260238474	harald:33
8	allamand	0.11296988718603691	allamand:107
9	segund	0.30692201568815786	segunda:81 segundo:64 segundos:8
9	vuelte	0.30512511779798468	vuelta:86 vuelto:4
9	vot	0.24633132173169611	votos:54 voto:51 votar:49 votó:12 votaron:10 votan:6 votando:5 votado:5 vota:3 voté:2 votemos:2 votaran:1 vote:1 voten:1 votará:1
9	polit	-0.18567375610280681	política:196 político:113 políticas:52 políticos:47 políticamente:10 politicas:4 politica:3
9	apoy	0.17089642623339149	apoyo:83 apoyar:38 apoyaron:9 apoyando:9 apoyado:8 apoyó:6 apoyan:4 apoyará:3 apoyos:3 apoyaba:2 apoyaría:2 apoyamos:2 apoyaran:2 apoye:1 apoyarla:1 apoya:1
9	com	-0.15613080541197022	comando:318 comado:2 comida:2 comer:1
9	piñer	0.14657286999525854	piñera:179
9	gomez	-0.14434925903010362	gómez:73 goméz:4
9	president	0.1296491811756951	presidenta:331 presidente:250 presidencia:48 presidentes:23
9	sebastian	0.11824171964643104	sebastián:95

ANEXO C : TABLAS LSA LA NACIÓN

Tema	Termino	Puntuacion	Frecuencias
0	candidat	0.19399687263305912	candidata:722 candidato:514 candidatos:390 candidatas:64
0	bachelet	0.17101910537380841	bachelet:1245 bacheletista:2 bacheletismo:1
0	nuev	0.15955800991873187	nueva:694 nuevo:117 nuevos:72 nuevamente:35 nuevas:32 nueve:19
0	matthei	0.15722669022753041	matthei:1166
0	presidencial	0.15562510344507671	presidencial:766 presidenciales:134 presidencialismo:1
0	mayor	0.14442270573010763	mayoría:572 mayor:129 mayores:50 mayorías:8 mayormente:1
0	vot	0.1381141537608303	votar:218 voto:189 votos:184 votaron:23 votaría:20 votó:19 votado:17 votará:16 vota:14 voten:14 vote:13 votan:13 votv©:7 votamos:3 votando:3 votados:3 votaremos:3 votarán:2 votara:2 votemos:1 votaban:1 votaran:1
0	vuelte	0.13239947029183238	vuelta:506 vuelto:6 vueltas:4
0	segund	0.12468798648837098	segunda:427 segundo:78 segundos:19 segundas:2
0	part	0.12081996949984652	partido:356 parte:264 partidos:97 partir:33 partió:12 partiendo:4 partida:4 partes:4 partamos:2 parten:2 partimos:2 partan:1 parta:1
1	vuelte	0.45709821733323963	vuelta:506 vuelto:6 vueltas:4
1	segund	0.39680627029017368	segunda:427 segundo:78 segundos:19 segundas:2
1	vot	0.18599551089868524	votar:218 voto:189 votos:184 votaron:23 votaría:20 votó:19 votado:17 votará:16 vota:14 voten:14 vote:13 votan:13 votv©:7 votamos:3 votando:3 votados:3 votaremos:3 votarán:2 votara:2 votemos:1 votaban:1 votaran:1
1	primer	0.14093737968886222	primera:227 primeros:30 primeras:16
1	polit	-0.13582237034239553	política:316 político:171 políticos:99 políticas:69 políticamente:7 politica:1 politicas:1
1	presidencial	0.12905389766130623	presidencial:766 presidenciales:134 presidencialismo:1

1	gobiern	-0.11642103457320065	gobierno:617 gobierne:4 gobierna:3 gobiernen:1 gobiernan:1
1	pais	-0.11482506515210499	país:437 países:41
1	gan	0.11120377132010097	ganar:85 gana:20 gane:20 ganó:19 ganamos:12 ganas:10 ganan:6 ganarle:6 ganando:6 ganado:6 ganaron:6 ganará:5 ganara:4 ganaba:4 ganaremos:3 ganemos:2 ganarse:2 ganadas:1 ganándose:1 ganen:1 ganaban:1 ganaría:1 ganada:1
1	bachelet	0.10174469273633273	bachelet:1245 bacheletista:2 bacheletismo:1
2	vot	-0.29086815292025509	votar:218 voto:189 votos:184 votaron:23 votaría:20 votó:19 votado:17 votará:16 vota:14 voten:14 vote:13 votan:13 votv@:7 votamos:3 votando:3 votados:3 votaremos:3 votarán:2 votara:2 votemos:1 votaban:1 votaran:1
2	vuelte	-0.27805732018361384	vuelta:506 vuelto:6 vueltas:4
2	presidencial	0.24440074860219124	presidencial:766 presidenciales:134 presidencialismo:1
2	segund	-0.23399687892459259	segunda:427 segundo:78 segundos:19 segundas:2
2	candidat	0.22011891306956308	candidata:722 candidato:514 candidatos:390 candidatas:64
2	debat	0.16417578846719594	debate:339 debates:102 debatir:35 debatiendo:1 debatido:1 debata:1 debatían:1 debatieran:1
2	nuev	0.14098670687870274	nueva:694 nuevo:117 nuevos:72 nuevamente:35 nuevas:32 nueve:19
2	mayor	0.13875435114709325	mayoría:572 mayor:129 mayores:50 mayorías:8 mayormente:1
2	abander	0.13818804438758545	abanderada:322 abanderado:92 abanderados:11 abanderadas:6 abandera:2 abanderad:1
2	cre	-0.13801206815447781	creo:253 creemos:60 cree:55 crear:42 creen:18 creer:15 crea:9 creará:8 creó:8 creado:5 creamos:4 creando:4 creía:3 crearon:2 creada:2 crees:2 crearemos:2 creados:1 crean:1 creándole:1
3	nuev	-0.39374419169345787	nueva:694 nuevo:117 nuevos:72 nuevamente:35 nuevas:32 nueve:19
3	mayor	-0.37376588989097931	mayoría:572 mayor:129 mayores:50 mayorías:8 mayormente:1

3	bachelet	-0.26339015787090414	bachelet:1245 bacheletista:2 bacheletismo:1
3	parisi	0.24490603934510588	parisi:658
3	independient	0.18106858968026393	independiente:246 independientes:56 independientemente:10
3	program	-0.16602211089883445	programa:397 programas:36 programado:12 programó:2 programada:2 programados:1
3	part	0.15936181882077455	partido:356 parte:264 partidos:97 partir:33 partió:12 partiendo:4 partida:4 partes:4 partamos:2 parten:2 partimos:2 partan:1 parta:1
3	israel	0.14358088282341383	israel:160
3	claud	0.14058829129798914	claud:171
3	mirand	0.13522629653431725	miranda:142
4	vot	0.27860801161719695	votar:218 voto:189 votos:184 votaron:23 votaría:20 votó:19 votado:17 votará:16 vota:14 voten:14 vote:13 votan:13 votv©:7 votamos:3 votando:3 votados:3 votaremos:3 votarán:2 votara:2 votemos:1 votaban:1 votaran:1
4	matthei	-0.2260205120848236	matthei:1166
4	parisi	-0.19952406313227139	parisi:658
4	nuev	0.17872088270484124	nueva:694 nuevo:117 nuevos:72 nuevamente:35 nuevas:32 nueve:19
4	president	-0.17296037629081454	presidente:378 presidenta:200 presidencia:86 presidentes:17 presidencias:1 presidentas:1
4	mayor	0.1728028877990157	mayoría:572 mayor:129 mayores:50 mayorías:8 mayormente:1
4	claud	0.15903224049164691	claud:171
4	mirand	0.14662732752190233	miranda:142
4	piñer	-0.14350673769803168	piñera:215
4	ministr	-0.14288181062932628	ministra:166 ministro:104 ministros:27 ministras:4
5	debat	0.56996991098185057	debate:339 debates:102 debatir:35 debatiendo:1 debatido:1 debata:1 debatían:1 debatieran:1
5	vot	-0.24641689948740661	votar:218 voto:189 votos:184 votaron:23 votaría:20 votó:19 votado:17 votará:16 vota:14 voten:14 vote:13 votan:13 votv©:7 votamos:3 votando:3 votados:3 votaremos:3 votarán:2 votara:2 votemos:1 votaban:1 votaran:1

5	anatel	0.15898193303635003	anatel:92
5	asoci	0.15350974565677994	asociación:73 asociados:5
5	segund	0.14763100516771183	segunda:427 segundo:78 segundos:19 segundas:2
5	president	-0.13136393433962593	presidente:378 presidenta:200 presidencia:86 presidentes:17 presidencias:1 presidentas:1
5	piñer	-0.12372543715692244	piñera:215
5	chil	0.12177925218866113	chile:664 chilean:1
5	for	0.12137728261250548	foro:90 foros:8
5	archi	0.12032256871855178	archi:66
6	vot	-0.51782792456567983	votar:218 voto:189 votos:184 votaron:23 votaría:20 votó:19 votado:17 votará:16 vota:14 voten:14 vote:13 votan:13 votV©:7 votamos:3 votando:3 votados:3 votaremos:3 votarán:2 votara:2 votemos:1 votaban:1 votaran:1
6	segund	0.28444880758032431	segunda:427 segundo:78 segundos:19 segundas:2
6	vuelt	0.27136134243465432	vuelta:506 vuelto:6 vueltas:4
6	debat	-0.24117223227690407	debate:339 debates:102 debatir:35 debatiendo:1 debatido:1 debata:1 debatían:1 debatieran:1
6	part	0.18573518733496824	partido:356 parte:264 partidos:97 partir:33 partió:12 partiendo:4 partida:4 partes:4 partamos:2 parten:2 partimos:2 partan:1 parta:1
6	israel	0.15288853361246799	israel:160
6	ricard	0.14784808090361085	ricardo:144
6	alianz	-0.13380388627073878	alianza:459 alianzas:11
6	mirand	0.12386896522063112	miranda:142
6	sfe	0.1238386853928499	sfeir:118
7	parisi	0.36941231069829172	parisi:658
7	educ	0.20158136426243312	educación:270 educativo:5 educativa:3 educar:3 educan:1 educador:1
7	president	-0.17303655947821642	presidente:378 presidenta:200 presidencia:86 presidentes:17 presidencias:1 presidentas:1
7	independient	0.16305105437048695	independiente:246 independientes:56 independientemente:10
7	nuev	0.16213849463442445	nueva:694 nuevo:117 nuevos:72 nuevamente:35 nuevas:32 nueve:19

7	econom	0.15775812270760126	economista:174 economía:56 económico:44 económicos:22 económica:20 económicas:10 economistas:4 económicamente:1
7	matthei	0.13655701291390857	matthei:1166
7	mayor	0.13613094266091824	mayoría:572 mayor:129 mayores:50 mayorías:8 mayormente:1
7	colegi	0.13560509007948732	colegios:81 colegio:41 colegiados:1
7	debat	-0.11821243316323521	debate:339 debates:102 debatir:35 debatiendo:1 debatido:1 debata:1 debatían:1 debatieran:1
8	piñer	0.36568095174920368	piñera:215
8	sebastian	0.30600832748009144	sebastián:150
8	president	0.29378159719589275	presidente:378 presidenta:200 presidencia:86 presidentes:17 presidencias:1 presidentas:1
8	gobiern	0.28093266329463534	gobierno:617 gobierne:4 gobierna:3 gobiernen:1 gobiernan:1
8	bachelet	0.13289467259240656	bachelet:1245 bacheletista:2 bacheletismo:1
8	le	0.13047854533444228	le:687
8	senador	-0.12968037311200292	senador:161 senadora:45 senadores:38 senadoras:2
8	ex	-0.12391982315725773	ex:823
8	polit	-0.11899528433366285	política:316 político:171 políticos:99 políticas:69 políticamente:7 politica:1 politicas:1
8	rn	-0.11566295521559952	rn:152
9	ex	-0.24861671451388481	ex:823
9	derech	0.22600576478727763	derecha:185 derechos:135 derecho:96 derechamente:4 derechas:3
9	alianz	0.20098503375367013	alianza:459 alianzas:11
9	matthei	0.19847672068133773	matthei:1166
9	polit	0.19505768536151258	política:316 político:171 políticos:99 políticas:69 políticamente:7 politica:1 politicas:1
9	human	0.14403360679962368	humanos:82 humanista:61 humano:21 humanidad:7 humanas:7 humana:6 humanidades:2 humanamente:1
9	mandatari	-0.13750176338830317	mandataria:179 mandatario:43 mandatarios:2
9	primer	-0.13444130206465113	primera:227 primeros:30 primeras:16
9	mil	-0.13340436073380602	mil:172 miles:33
9	noviembr	-0.13195887997289826	noviembre:156

ANEXO D : TABLAS LSA EL MOSTRADOR

Tema	Termino	Puntuacion	Frecuencias
0	candidat	0.20539460797100462	candidata:690 candidato:362 candidatos:279 candidatas:26
0	nuev	0.17172038172308451	nueva:643 nuevo:86 nuevos:52 nueve:46 nuevamente:28 nuevas:26 nueves:1
0	presidencial	0.16962214799290648	presidencial:745 presidenciales:92 presidencialismo:1
0	bachelet	0.16121707070690927	bachelet:898 bacheletismo:4 bacheletista:2 bachelete:1
0	matthei	0.15406127227723315	matthei:808
0	mayor	0.15149157117565953	mayoría:521 mayor:119 mayores:37 mayorías:4
0	gobiern	0.14632215310460797	gobierno:717 gobiernen:8 gobierna:6 gobierne:2 gobiernan:1
0	part	0.14196609653215775	partido:354 parte:319 partidos:85 partir:32 partes:6 partiV≥:4 partiendo:2 partirv°n:1 parta:1
0	program	0.13206286910383544	programa:496 programas:43 programado:4 programadas:1 programV≥:1 programada:1
0	president	0.12034954606116394	presidente:356 presidenta:138 presidencia:60 presidentes:32 presidentas:2 presidencias:1
1	presidencial	0.26797609548683948	presidencial:745 presidenciales:92 presidencialismo:1
1	alianz	0.23525246062475849	alianza:354 alianzas:4
1	candidat	0.20280351055524623	candidata:690 candidato:362 candidatos:279 candidatas:26
1	matthei	0.17109883856452027	matthei:808
1	mayor	0.16894985667560222	mayoría:521 mayor:119 mayores:37 mayorías:4
1	abander	0.1661765883963684	abanderada:353 abanderado:99 abanderados:7 abandera:2
1	nuev	0.14959455436203711	nueva:643 nuevo:86 nuevos:52 nueve:46 nuevamente:28 nuevas:26 nueves:1
1	bachelet	0.14192762587659072	bachelet:898 bacheletismo:4 bacheletista:2 bachelete:1
1	doming	0.13526949084445605	domingo:132
1	chil	-0.12903187651133213	chile:671

2	vot	0.38380473581422359	votar:158 voto:120 votos:85 voten:15 votaría:14 votaron:12 votan:10 vota:10 vote:8 votarían:5 votado:5 votarv°:3 votarv°n:3 votamos:2 votv≥:2 votara:1 votarlo:1 votv©:1
2	vuelt	0.29765278431203557	vuelta:231 vuelto:4
2	segund	0.26983403132704947	segunda:200 segundo:72 segundos:1
2	program	-0.24265595785160313	programa:496 programas:43 programado:4 programadas:1 programv≥:1 programada:1
2	encuest	0.21708277245109348	encuesta:116 encuestas:89 encuestados:10 encuestadas:2
2	gobiern	-0.1986765809855254	gobierno:717 gobiernen:8 gobierna:6 gobierne:2 gobiernan:1
2	nuev	-0.18091472759392518	nueva:643 nuevo:86 nuevos:52 nueve:46 nuevamente:28 nuevas:26 nueves:1
2	mayor	-0.15434749981879775	mayoría:521 mayor:119 mayores:37 mayorías:4
2	parisi	0.12892775023874006	parisi:302
2	llam	0.12342930808347033	llamado:117 llamv≥:33 llama:21 llamados:13 llamar:5 llamada:5 llamando:5 llamaron:4 llaman:2 llamarle:1 llamo:1 llamadas:1 llame:1
3	debat	0.26907069985195603	debate:244 debates:69 debatir:35 debatamos:2 debativ©ramos:1 debatido:1 debato:1 debativ≥:1 debatieron:1
3	part	0.20973182594816459	partido:354 parte:319 partidos:85 partir:32 partes:6 partiv≥:4 partiendo:2 partirv°n:1 parta:1
3	vuelt	-0.20500589431063987	vuelta:231 vuelto:4
3	gobiern	-0.18579555417508015	gobierno:717 gobiernen:8 gobierna:6 gobierne:2 gobiernan:1
3	enriquez- ominami	0.17537830457613829	enríquez-ominami:254 enriquez-ominami:5
3	presidencial	0.17525290295019641	presidencial:745 presidenciales:92 presidencialismo:1
3	chil	0.14917976259422011	chile:671
3	bachelet	-0.14811396508894814	bachelet:898 bacheletismo:4 bacheletista:2 bachelete:1
3	independent	0.14810017183064927	independiente:143 independientes:47 independientemente:7
3	segund	-0.14407952132677102	segunda:200 segundo:72 segundos:1
4	reform	-0.21759742228769482	reforma:151 reformas:54 reformar:9

4	nuev	-0.21208988636624762	nueva:643 nuevo:86 nuevos:52 nueve:46 nuevamente:28 nuevas:26 nueves:1
4	president	0.21157110022717249	presidente:356 presidenta:138 presidencia:60 presidentes:32 presidentas:2 presidencias:1
4	laboral	-0.19372356652288653	laboral:163 laborales:40
4	vuelt	-0.18114082849520566	vuelta:231 vuelta:4
4	matthei	0.17468709168215105	matthei:808
4	mayor	-0.16875955626553971	mayoría:521 mayor:119 mayores:37 mayorías:4
4	bachelet	-0.15982271735631334	bachelet:898 bacheletismo:4 bacheletista:2 bachelete:1
4	segund	-0.15777438565715307	segunda:200 segundo:72 segundos:1
4	piV±er	0.13909488142054627	piV±era:163 piV±eristas:1
5	debat	0.66395807836012466	debate:244 debates:69 debatir:35 debatamos:2 debativ@ramos:1 debatido:1 debato:1 debativ≥:1 debatieron:1
5	educ	-0.13952841186977183	educaciV≥n:264 educativa:9 educativo:8 educativos:1 educar:1 educadores:1 educan:1 educarse:1 educadoras:1
5	anatel	0.12664907262610989	anatel:48
5	program	0.11438628797683142	programa:496 programas:43 programado:4 programadas:1 programV≥:1 programada:1
5	laboral	0.10252893514868255	laboral:163 laborales:40
5	vot	0.096839520578304319	votar:158 voto:120 votos:85 voten:15 votaría:14 votaron:12 votan:10 vota:10 vote:8 votarían:5 votado:5 votarV°:3 votarV°n:3 votamos:2 votV≥:2 votara:1 votarlo:1 votV@:1
5	quier	0.09371735884176001	quiero:121 quiere:74 quieren:59 quieres:55 quieran:12 quiera:12
5	anp	0.089092461615773305	anp:31
5	particip	0.087372084457422519	participaciV≥n:79 participar:55 participaron:25 participV≥:20 participado:11 participen:11 participantes:6 participan:6 participativo:5 participa:4 participando:4 participarV°n:3 participativa:3 participarV°:3 participe:3 partícipes:2 participamos:1 participaba:1 participaremos:1 participaban:1 participativos:1
5	nuev	-0.083270686938311467	nueva:643 nuevo:86 nuevos:52 nueve:46 nuevamente:28 nuevas:26 nueves:1

6	vot	0.45155398124265461	votar:158 voto:120 votos:85 voten:15 votaría:14 votaron:12 votan:10 vota:10 vote:8 votarían:5 votado:5 votarv°:3 votarv°n:3 votamos:2 votv≥:2 votara:1 votarlo:1 votv©:1
6	llam	0.32506677129613953	llamado:117 llamv≥:33 llama:21 llamados:13 llamar:5 llamada:5 llamando:5 llamaron:4 llaman:2 llamarle:1 llamo:1 llamadas:1 llame:1
6	segund	-0.27531210379506599	segunda:200 segundo:72 segundos:1
6	vuel	-0.22655545691144177	vuelta:231 vuelta:4
6	parisi	-0.1595898338985946	parisi:302
6	nuev	0.15795933800257431	nueva:643 nuevo:86 nuevos:52 nueve:46 nuevamente:28 nuevas:26 nueves:1
6	matthei	-0.15354368102902718	matthei:808
6	laboral	-0.14529520984925937	laboral:163 laborales:40
6	mayor	0.12896034814419163	mayoría:521 mayor:119 mayores:37 mayorías:4
6	pas	-0.12317241953900256	pasado:100 pasar:64 paso:62 pasa:22 pasv≥:13 pasada:12 pasando:11 pasadas:8 pase:5 pasara:4 pasan:3 pasen:3 pasarv°:3 pasaría:3 pasamos:2 pasarv°n:2 pasos:2 pasarían:1 pasearme:1 pasaron:1 pasear:1 pasados:1
7	vot	-0.22902907807071748	votar:158 voto:120 votos:85 voten:15 votaría:14 votaron:12 votan:10 vota:10 vote:8 votarían:5 votado:5 votarv°:3 votarv°n:3 votamos:2 votv≥:2 votara:1 votarlo:1 votv©:1
7	president	0.2150330952222119	presidente:356 presidenta:138 presidencia:60 presidentes:32 presidentas:2 presidencias:1
7	vuel	0.21054985523029948	vuelta:231 vuelta:4
7	matthei	-0.19610602382227652	matthei:808
7	alianz	-0.19039651178980557	alianza:354 alianzas:4
7	debat	0.18447693234753268	debate:244 debates:69 debatir:35 debatamos:2 debativ©ramos:1 debatido:1 debato:1 debativ≥:1 debatieron:1
7	segund	0.17725944675612124	segunda:200 segundo:72 segundos:1
7	part	0.17593021054023159	partido:354 parte:319 partidos:85 partir:32 partes:6 partiv≥:4 partiendo:2 partiv°n:1 parta:1

7	trabaj	-0.13221272530110634	trabajadores:185 trabajando:44 trabajado:18 trabajv≥:5 trabajadoras:5 trabajaron:5 trabajador:4 trabajarv°:3 trabajos:3 trabajen:1 trabaje:1 trabajarv°n:1 trabajos:1 trabajemos:1 trabajaremos:1
7	primer	0.12335748349676239	primera:141 primeros:27 primeras:4
8	laboral	0.2443673529306456	laboral:163 laborales:40
8	part	0.24425554145366091	partido:354 parte:319 partidos:85 partir:32 partes:6 partiv≥:4 partiendo:2 partiv°n:1 parta:1
8	chil	-0.24150956367147919	chile:671
8	tem	0.17123455168563376	tema:224 temas:182 teme:6 temen:3 temo:2 temidas:1 tememos:1
8	pais	-0.16364587683479559	país:399 países:46
8	region	-0.14364324218382737	regiv≥n:137 regiones:107
8	trabaj	0.14027630897310736	trabajadores:185 trabajando:44 trabajado:18 trabajv≥:5 trabajadoras:5 trabajaron:5 trabajador:4 trabajarv°:3 trabajos:3 trabajen:1 trabaje:1 trabajarv°n:1 trabajos:1 trabajemos:1 trabajaremos:1
8	quier	-0.13572654101759823	quiero:121 quiere:74 quieren:59 quieres:55 quieran:12 quiera:12
8	educ	-0.12524533812336747	educaciv≥n:264 educativa:9 educativo:8 educativos:1 educar:1 educadores:1 educan:1 educarse:1 educadoras:1
8	cambi	-0.12377258424130033	cambio:68 cambios:64 cambia:61 cambiar:48 cambiado:8 cambiando:4 cambiarv°:4 cambie:3 cambiarv°n:2 cambian:2 cambiarlo:2 cambiaría:1 cambiarse:1 cambiemos:1 cambiarlas:1 cambien:1 cambiaremos:1
9	program	-0.33198423128490506	programa:496 programas:43 programado:4 programadas:1 programv≥:1 programada:1
9	gobiern	-0.22126290526453959	gobierno:717 gobiernen:8 gobierna:6 gobierne:2 gobiernan:1
9	encuest	-0.16631891060916562	encuesta:116 encuestas:89 encuestados:10 encuestadas:2
9	ciudadan	0.15284831277160854	ciudadanía:83 ciudadana:63 ciudadanos:49 ciudadano:24 ciudadanas:11

9	tem	0.15215439905143641	tema:224 temas:182 teme:6 temen:3 temo:2 temidas:1 tememos:1
9	reform	0.14458923799965701	reforma:151 reformas:54 reformar:9
9	deb	0.13908256405045616	debe:118 deben:63 debemos:46 debido:20 debería:19 debiera:17 debiv≥:16 deber:13 deberv°n:10 debieron:8 deberían:7 deberv°:7 debieran:6 debiese:5 debían:4 deberíamos:3 debida:3 deban:2 debiv©ramos:2 debiendo:1 debía:1 deberse:1 deba:1
9	debat	0.13460013404127755	debate:244 debates:69 debatir:35 debatamos:2 debativ©ramos:1 debatido:1 debato:1 debativ≥:1 debatieron:1
9	segur	0.12853444596454286	seguridad:68 seguro:36 seguros:24 segura:10 seguramente:6 seguras:1 seguridades:1
9	chil	-0.12839149956454821	chile:671