

**Creación de un método de analítica de noticias en medios de
comunicación**

Javier Galvis Montero

**Escuela Colombiana de Ingeniería Julio Garavito
Decanatura de Ingeniería de Sistemas
Maestría Gestión de información
Bogotá D.C; Colombia
2020**

Creación de un método de analítica de noticias en medios de comunicación

Javier Galvis Montero

Trabajo de grado para optar al título de
Magister en Gestión de Información

Directores

Dr. Dante Conti

Dra. Victoria Eugenia Ospina

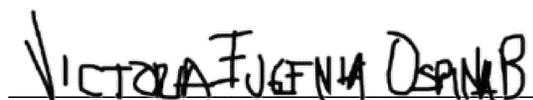
Escuela Colombiana de Ingeniería Julio Garavito
Decanatura de Ingeniería de Sistemas
Maestría Gestión de información
Bogotá D.C; Colombia
2020

© Únicamente se puede usar el contenido de las publicaciones para propósitos de información. No se debe copiar, enviar, recortar, transmitir o redistribuir este material para propósitos comerciales sin la autorización de la Escuela Colombiana de Ingeniería. Cuando se use el material de la Escuela se debe incluir la siguiente nota “Derechos reservados a Escuela Colombiana de Ingeniería” en cualquier copia en un lugar visible. Y el material no se debe notificar sin el permiso de la Escuela.

Publicado en 2020 por la Escuela Colombiana de Ingeniería Julio Garavito. Avenida 13 No 205-59 Bogotá. Colombia. TEL: +57 -1 668 36 00

PÁGINA DE ACEPTACIÓN

El trabajo de grado de maestría titulado “**Creación de un método de analítica de noticias en medios de comunicación**” presentado por JAVIER GALVIS MONTERO cumple con los requisitos establecidos y recibe nota aprobatoria para optar al título de Magíster en Gestión de información.



VICTORIA EUGENIA OSPINA BECERRA

Directora del trabajo de grado



DANTE CONTI

Director 2 del trabajo de grado



OSWALDO CASTILLO NAVETTY

Jurado 1



DANIEL DE LA ROSA

Jurado 2

Bogotá, D.C., 31 de Julio de 2020

Reconocimiento o agradecimientos

A Alba María, Juliana y Manuela mi mayor motivación.

Dante Conti, guía constante y polo a tierra en este proyecto. Excelente maestro y director quien me hizo mi experiencia de aprendizaje la más enriquecedora de todas.

A Vicky Eugenia Ospina, directora del programa quien desde el principio estuvo motivándome a ir más allá de mis propias expectativas.

A la escuela Colombiana de Ingeniería Julio Garavito que me brindó las herramientas necesarias para materializar este proyecto.

Resumen

Uno de los retos más importantes en el análisis de los datos es transformar la data en conocimiento útil y no trivial, para que éste pueda contribuir y soportar los procesos de toma de decisiones en la organización. Resulta fundamental una gestión eficiente en el proceso de análisis y en la misma Gestión de la Información. En el campo de los medios de comunicación y en especial en el análisis de las noticias, esto constituye una razón de alta relevancia, pues la gran cantidad de medios y fuentes generan información en tiempo real, por lo que el procesamiento y análisis de los datos generados debe hacerse de una manera efectiva, ágil y enmarcada en procesos o metodologías guía. Es por ello, que en esta investigación se abordó el reto de proponer una metodología detallada y robusta que a través de un sistema paso a paso pueda gestionar la llegada, procesamiento, transformación y análisis de la data para que posteriormente, ésta pueda agruparse y catalogarse para que junto a herramientas de Analítica Avanzada se generen hallazgos y conocimiento enfocados en enriquecer el input recibido.

El método propuesto en esta investigación se enmarca en la estandarización de la data por medio de atributos únicos que son clasificados y agrupados con características similares mediante la creación de diccionarios personalizados. La selección y profundización de dichos diccionarios de atributos se desarrollaron con diferentes técnicas en la combinación y selección ajustándolos a niveles descriptivos asociados al contexto de los medios de comunicación, ampliando así los niveles de explotación y visualización a la hora de implementar herramientas de analítica avanzada (Minería de Texto, búsqueda inteligente y relacional en las noticias, y su posterior visualización). Por tanto, la creación de los diccionarios con su profundización de atributos básicos de las categorías o tags de clasificación enriquecen los hallazgos logrados y facilitan la aplicación de técnicas de analítica avanzada.

De manera puntual se obtuvieron tres productos en esta investigación: (1) Metodología guía paso a paso para gestionar datos de medios de comunicación, (2) Creación de Diccionarios a varios niveles que enriquecen los tags inputs para mejorar los análisis y (3) Aplicación de técnicas de analítica avanzada estándares y desarrolladas en softwares especializados para la generación de conocimiento que soporte el proceso de toma de decisiones dentro de un caso de estudio empresarial (Siglo Data, Bogotá). Con esto, la investigación cumplió en ofrecer de manera robusta, un conjunto de herramientas de análisis y comprensión de las noticias para actuar de una forma más eficiente y segura ante el desafío creciente que aborda la gestión en tiempo real del mundo noticioso.

Palabras claves: Gestión de la Información, Analítica Avanzada, Big Data, Minería de Texto, Diccionarios.

Abstract

One of the most important challenges in data analysis is to transform information into useful and non-trivial knowledge, so that we can contribute and help decision-making processes in the organization. Efficient management in the analysis process and in Information Management itself is essential. In the field of media and especially in news analysis, this is a highly relevant reason, since the large number of media and sources of information are in real time, the processing and analysis of data duties should be generated in an effective and agile way framed in processes or guide methodologies. That is why, in this research, the challenge of proposing a specific and robust methodology was addressed through a step-by-step system that can manage the arrival, processing, transformation and analysis of data so that later, it can be group and catalog with Advanced Analytical tools, findings and knowledge focused on enriching the input received.

The method proposed in this research is framed in the standardization of the data by means of unique attributes that are classified and grouped with similar characteristics through the creation of personalized dictionaries. The selection and deepening of these attribute dictionaries will be developed with different techniques in the combination and selection, adjusting them to descriptive levels associated with the context of the media. This will allow the exploitation and visualization levels to be disseminated by implementing advanced analytical tools (Text mining, intelligent and relational search in the news, and its subsequent visualization). Therefore, the creation of dictionaries with their deepening of basic attributes of the categories or classification labels will enrich the findings and facilitate the application of advanced analytical techniques.

In a timely manner, three products will be obtained in this research: (1) Step-by-step guide methodology for managing media data, (2) Creation of Dictionaries at various levels that enrich the input labels to improve analysis and (3) Application of advanced analysis techniques developed in specialized software for the generation of knowledge that supports the decision-making process within a business case study (Siglo Data, Bogotá). With this contributions, the research was successful in offering a robust set of news analysis and understanding tools to act more efficiently and safely in the face of the growing challenge of real-time management of the news world.

Key words: Information Management, Advanced Analytics, Big Data, Text Mining, Dictionaries.

Tabla de contenido

Capítulo 1. Introducción	1
1.1 Contexto general.....	1
1.1.2 .Inmediatez de la información	2
1.1.3 Analítica de los medios	2
1.2. Objetivo General	3
1.3. Objetivos específicos.....	3
1.4. Alcances y limitaciones	3
1.5 Contribución de la tesis	4
Capítulo 2. Marco teórico.....	5
2.1. Conceptos básicos para la gestión de la información en medios de comunicación	5
2.1.1 Matriz de macrosegmentación.....	5
2.1.2 Cadena de valor.....	6
2.1.3 Variables de la cadena de valor	6
2.1.4 Flujograma del proceso de monitorización de noticias	7
2.1.5 Matriz de Macrosegmentación y Cadena de Valor en Siglo Data: “La necesidad de un marco metodológico para la gestión de los datos y su explotación con Analítica Avanzada”	7
2.1.6 Tendencias Mundiales.....	8
2.1.7 Tendencias del sector	9
2.2 Análisis de Competitividad. El reto de incorporar un marco guía de desarrollo para la gestión de Información en Siglo Data con Analítica Avanzada.....	10
2.2.1 Fortalezas y Debilidades	10
2.3 Conceptos básicos de Analítica Avanzada bajo el contexto de la investigación	11
2.3.1 Analítica de datos	11
2.4 Analítica Avanzada.....	11
2.4.1 Análítica Descriptiva.....	11
2.4.2 Analítica Predictiva	11
2.4.3 Analítica Prescriptiva	11
2.5 Minería de Texto	12
2.5.1 Procesos en la analítica de texto	12
2.5.2 El etiquetado gramatical.....	12
2.5.3 El clustering o agrupamiento.....	12
2.5.4 Clasificación o categorización.....	12
2.5.5 La extracción de información	12
2.5.6 El análisis de sentimiento	12
2.6 Nube de palabras	13
2.7 Diccionarios específicos	13
2.8 Visualización de resultados.....	14
2.9 KPIs y la importancia en la investigación.....	14
Capítulo 3. Desarrollo y propuesta del Marco metodológico para la Gestión de la Información	16

3.1 La Gestión de los Procesos – Visión Macro	16
3.1.1 Gestión de procesos.....	16
3.1.2 Medios de comunicación.....	16
3.1.3 Servidores de captura	16
3.1.4 Selección de categoría.....	17
3.1.5 Servidores.....	17
3.1.6 Diccionario de datos.....	17
3.1.7 Servidores nube.....	17
3.2 Marco metodológico para la gestión de información y analítica avanzada	17
3.2.1 Importar datos de los tags generados diariamente (volumen de 60 mil registros diarios).....	17
3.2.2 Limpiar los datos importados de acuerdo con las reglas de Negocio	18
3.2.3 Cargar información de medios prensa nacionales en formato .txt.....	20
3.2.4 Limpiar y ajustar los .txt cargados de acuerdo con la estructura de cada periódico y las reglas de negocio establecidas (caracteres especiales y mala redacción)	20
3.2.5 Modelación con los 3 tipos de diccionario de datos	20
3.2.6 Generar información estructurada a partir de los .txt.	20
3.2.7 Construir reglas de negocio para la lectura de la información previamente estructurada para procesarla a través Analytics.....	20
3.3 Creación del diccionario.....	21
3.3.1 Atributo de cada uno de los datos.....	21
3.3.2 Diccionario selección de monitoreo	22
3.3.4 Esquema global de clasificación de noticias	23
3.4 Diccionario selección de monitoreo	24
3.4.1 Categoría	24
3.4.2 Medio	24
3.4.3 Fuente.....	24
3.4.4 Sección.....	24
3.4.5 Página / seg	25
3.4.6 Vpe.....	25
3.4.7 Base audiencia.....	25
3.4.8 Tema.....	25
3.4.9 Fecha	26
3.4.10 Peso	26
3.4.11 Número de noticias.....	26
3.4.12 % de noticias	26
3.5 Diccionario Etiquetado por atributos.....	26
3.5.1 Categoría	26
3.5.2 Tipo	26
3.5.3 Sector primario.....	26
3.5.4 Sector Secundario.....	27
3.6 Resultado esperado.....	27
3.6.1 Técnicas analíticas para trabajar la información	27
3.6.1.1 Tabla.....	27
3.6.1.2 Gráfico de barras	28
3.6.1.3 Gráfico circular	29
3.6.1.4 Tópicos.....	29
3.6.1.5Categorías relacionadas.....	30
Capítulo 4. Validación del modelo por medio de la herramienta de Analítica.....	31

4.1 “Siglo Data” La empresa como caso de uso de la investigación	31
4.1.1 Misión	32
4.1.3 Objetivos Actuales	32
4.2 Clasificación de Etiquetas.....	34
4.2.1 Tipo	34
4.2.2 Sector Primario.....	35
4.2.3 Sector Secundario.....	36
4.2.4 Resultado y aporte	36
4.3 Ejemplos o usos del dashboard para el consumo de la información de forma analítica 37	37
4.3.1 Tipo empresas	37
4.3.2 Sector primario	38
4.3.3 Sector secundario	38
4.3.4 Proceso de documentación.	39
4.3.5 Proceso de selección de categorías.....	39
4.3.6 Visualización en el dashboard	40
4.3.7 Explotación de la Data	40
4.4 Herramienta Analytics	41
4.4.1 Módulo Analytics	41
4.4.2 Informe Descripción de noticias por grupos de selección de interés	42
4.4.3 Nube de Categorías	42
4.4.4 Análisis de categorías asociadas.....	43
4.4.5 Tópicos y Nube de términos de la categoría principal o de interés	44
4.4.6 Búsqueda de varios términos en Categorías o Titulares con la opción de y/o	44
4.5 Módulo Gráfico	45
4.5.1 Informe de Categoría Principal	46
4.5.2 Informe de Categoría Principal y Competencia.....	46
Capítulo 5. Conclusiones y recomendaciones	48
5.1 Aportes	49
5.2 Trabajos o líneas futuras sugeridas o recomendadas	49
6. Referencias	50
7. Anexos	52
7.1 Anexo 1. Listado de etiquetas sector primario	52
7.2 Anexos secundarios	52

Lista de figuras

Figura 1 Cadena de valor	6
Figura 2 Flujograma del proceso de monitorización de noticias	7
Figura 3 Nube de palabras	13
Figura 4 Visualización de resultados	14
Figura 5 Gestión de procesos	16
Figura 6 Diccionario de selección de monitoreo.....	22
Figura 7 Diccionario etiquetado por atributos	23
Figura 8 Esquema global de clasificación de noticias	23
Figura 9 Visualización de atributos clasificados	28
Figura 10 Clasificación por peso (positivo, negativo, neutro).....	29
Figura 11 Nube de terminos y clasificación de tópicos.....	29
Figura 12 Selección y/o entre categorías.....	30
Figura 13 Comparativo de ventas anuales de las compañías de monitoreo en Colombia....	32
Figura 14 Selector MMI sobre el proceso de documentación	39
Figura 15 Selector de televisión MMI sobre el proceso de selección de categoria.....	39
Figura 16 Visualización del Dashboard	40
Figura 17 Dashboard SAS VIYA 3.4	42
Figura 18 Nube de términos SAS VIYA 3.4.....	43
Figura 19 Visualización de atributos clasificados.....	43
Figura 20 Nube de terminos y clasificación de tópicos	44
Figura 21 Selección y/o entre categorías.....	45
Figura 22 Nube de términos y clasificación de tópicos.....	45

Lista de tablas

Tabla 1 Matriz de macrosegmentación.....	5
Tabla 2 Variables de la cadena de valor	6
Tabla 3 Importación de datos por tags	18
Tabla 4 Limpiar los datos importados de acuerdo con las reglas de negocio.....	19
Tabla 5 Diccionario de datos.....	21
Tabla 6 Grupos de categoría.....	27
Tabla 7 Selección de atributos en categorías tipo.....	34
Tabla 8 Selección de atributos en sector primario.....	35
Tabla 9 Selección de atributos en sector secundario.....	36

Tabla 10 Ejemplo empresa.....	37
Tabla 11 Ejemplo selección de atributos en categoría tipo.....	37
Tabla 12 Ejemplo selección de atributos en categoría primaria.....	38
Tabla 13 Ejemplo selección de atributos en categoría secundaria	38
Tabla 14 Data estructurada de selección de categorías y atributos.....	40

Capítulo 1. Introducción

1.1 Contexto general

Hoy en día, los medios de comunicación constituyen una herramienta persuasiva que nos permiten mantenernos en continua comunicación con los distintos sucesos sociales, políticos y económicos tanto a escala nacional como internacional.

“La rapidez y dramatismo con que los MCM (Medios de Comunicación para las Masas) se han ido incorporando en nuestra realidad, no nos dio tiempo para adaptarlos y adaptarnos. Ante ellos nos es difícil tomar una actitud serena: o nos aferramos a un mundo estático, pre-industrial; o nos desarraigamos por completo quedando a merced de las novedades o estímulos superficiales, e ignorando el pasado” (Leonardi, 2018). El Conocimiento que proporcionan la información emitida en los medios de comunicación nos convierte en actores principales de la realidad.

Cada día, en cada momento los medios de comunicación ocupan espacio en nuestras vidas de una manera en muchos casos intrusiva. Medios denominados “calientes” (radio y televisión) han pasado a segundo plano, siendo ahora los digitales los que tienen una mayor importancia en el saber de los acontecimientos diarios.

Actualmente nuestra sociedad está influenciada por el constante bombardeo de información a través de diferentes medios. Esto modifica nuestro modo de vida, nuestras costumbres, el consumo de unos productos, la opinión pública, entre otros. La información que nos llega sobre los distintos sucesos sociales, políticos o económicos puede hacer que las personas cambiemos nuestra forma de pensar respecto a la realidad que nos rodea (Muñoz, 2014). Fake News, o tomar posiciones poco objetivas y de muchas formas deliberadas y sin mayor análisis se convierten en una de las maneras de informarse. Informarse y decidir en tan solo 280 caracteres.

Prensa, radio, televisión e internet entendiéndolos como industrias productoras de cultura y fenómenos discursivos ideológicos, permiten la imposición y mantenimiento masivo del discurso del bloque hegemónico actual. Dicha lógica se genera en el contexto de la modernidad, en que las directrices y formas de ver, estar y sentir en el mundo están normadas por el sistema económico neoliberal, estableciéndose un entrecruzamiento de discursos del consumo y del sometimiento naturalizado (Toro, 2011).

A razón de esto, las instituciones y empresas deben estar en constante seguimiento a todos los volúmenes de información emitida por los diferentes canales mediáticos. Las crisis y también las oportunidades no dan espera (Toro, 2011).

Los medios quieren ser redes sociales, pero no tienen clara cuál es la fórmula que les garantice el éxito. Entre las vías desarrolladas nos encontramos desde la posibilidad de comentar las noticias, algo ya bastante extendido en los medios *online*, hasta plataformas de interconexión e intercambio de contenido multimedia dentro del propio medio. Éste es el caso de USA Today, donde los lectores pueden crear perfiles personales y compartir sus propios documentos multimedia (Fundación Telefónica Telos, 2018).

1.1.2 .Inmediatez de la información

Esta característica de los medios de comunicación facilita la “*explosión de noticias*” y es que hoy día debido a la comunicación directa que tiene el usuario con el medio, las personas del común también se convierten en periodistas. Videos y audios testifican de la veracidad de la información recogida, lo que hace que los medios puedan difundir de forma rápida las noticias de último momento.

Todo esto genera que la gran proliferación de noticias determine una problemática en sus análisis, comprensión y acción ante los eventos.

1.1.3 Analítica de los medios

Cada uno de los medios de comunicación contiene diferentes formatos técnicos en la presentación de sus contenidos. Mientras Televisión y radio es audiovisual, los medios Impresos lo hacen en papel; internet y redes sociales en digital. Todos estos formatos deben ser estandarizados para poderlos trabajar.

“La estructura de datos describe el formato en que los valores van a ser almacenados, cómo van a ser accedidos y modificados, pudiendo así existir una gran cantidad de estructuras de datos” (Sena, 2019).

La estructuración, explotación y visualización de la data estandarizada recogida de los medios de comunicación nos podrá dar como resultado hallazgos para entender de mejor manera la incidencia que tienen las noticias. Proceso importante para la toma de decisiones (Sena, 2019).

La situación actual es determinante en el tiempo que los analistas tardan en comprender y analizar todo el volumen de noticias recolectadas de un determinado tema o situación .La

importancia de este proyecto es conocer de qué manera se puede extraer hallazgos informativos a partir de un banco de datos estructurados conformado por las noticias emitidas en medios de comunicación para ayudar en el análisis de la información.

1.2. Objetivo General

Crear un método para transformar oportunamente volúmenes de noticias emitidas en medios de comunicación en datos y a su vez en información, para así generar hallazgos, tendencias y conocimiento.

1.3. Objetivos específicos

- a) Crear un método para convertir la forma que hoy día se presentan las noticias en televisión, radio, prensa e internet en datos estructurados.
- b) Crear un método de limpieza, explotación y visualización de estos datos.
- c) Generar métodos de conocimiento y tendencias mediáticas con base en los hallazgos.
- d) Realizar investigación de campo para confirmar que esta metodología si da valor a los departamentos de comunicación de empresas, compañías de comunicación estratégica y relaciones públicas en un caso de estudio empresarial (Empresa “Siglo Data”, Bogotá – Colombia)

1.4. Alcances y limitaciones

El alcance de esta investigación consiste en proponer un marco de desarrollo para una herramienta de análisis de data agrupada basado en una gestión de los procesos de la información y el uso de técnicas de analítica avanzada (Minería de texto, Diccionarios relevantes y Visualización de resultados) para así, poder encontrar y resaltar hallazgos en grupos de etiquetas o categorías que contienen características únicas extraídas de las fuentes de datos desde los medios de comunicación.

No corresponde a este trabajo el desarrollo computacional del software de analítica, sino su supervisión (en caso de materializarse) mediante el aporte del conocimiento estructurado con un marco metodológico de desarrollo. Tampoco corresponde a esta investigación la proyección de situaciones futuras, debido a que los asuntos de las noticias dependen de las coyunturas y de los hechos que día a día suceden bajo los diferentes marcos temáticos. Las noticias son un reflejo de ellos.

1.5. Metodología

La propuesta del marco de desarrollo como producto fundamental de esta investigación y que persigue la gestión eficiente de la data obtenida de los diferentes medios y su transformación en conocimiento útil que soporte procesos de toma de decisiones, se sustenta en un enfoque paso a paso resumido a continuación:

- a) Importar datos de los tags generados diariamente (Volumen de 60 mil registros diarios).
- b) Limpiar los datos importados de acuerdo con las reglas de Negocio.
- c) Cargar información de medios nacionales en formato .txt.
- d) Limpiar y ajustar los .txt cargados de acuerdo con la estructura de cada periódico y las reglas de negocio establecidas (caracteres especiales y mala redacción).
- e) Modelación con los 3 tipos de diccionario de datos.
- f) Generar un modelo predictivo para las categorías y tags generados.
- g) Generar información estructurada a partir de los .txt.
- h) Construir reglas de negocio para la lectura de la información previamente estructurada para procesarla a través de Text Analytics.
- i) Creación del diccionario.
- j) Reportes con la información actual, categorías y tags más relevantes, así como búsquedas personalizadas.

1.5 Contribución de la tesis

- a) Aporta hallazgos informativos de una forma eficiente brindando herramientas para mejorar la toma de decisiones en el análisis de grandes cantidades de noticias seleccionadas.
- b) En la exploración y profundización de temas y coyunturas dar lugar a la posibilidad de segmentar la información seleccionada de una forma más precisa.
- c) Realizar análisis comparativos de temas, situaciones y coyunturas desarrollados en distintos lapsos de tiempo.
- d) Comparar ágilmente las diferentes temáticas en grupos sectoriales homogéneos.

Capítulo 2. Marco teórico

2.1. Conceptos básicos para la gestión de la información en medios de comunicación

2.1.1 Matriz de macrosegmentación

La herramienta de macro segmentación contribuye en el entendimiento del entorno competitivo de Siglo Data referenciando su mercado, valores competitivos, comparativos, debilidades y fortalezas.

¿Qué?	¿Cómo?	¿A quién?	¿En Dónde?
Monitoreo, seguimiento, análisis y archivo de noticias emitidas en medios de comunicación (prensa, radio, televisión, internet) y redes sociales	<ul style="list-style-type: none">● Sistema de software especializado de propiedad de MMI el cual permite capturar, digitalizar y publicar información emitida.● Red de servidores en Colombia con respaldo en España.● Personal capacitado para captura e inclusión de información.● Dashboard personalizado de acuerdo a las necesidades de cada cliente.● Servicio de alertas y reportes	Compañías nacionales e internacionales del sector privado, institucional y agremiaciones.	Colombia

*Baquero; Camargo; García; González; Piracón; Galvis 2017
Tabla 1 Matriz de macrosegmentación*

2.1.2 Cadena de valor

La cadena de valor permite describir el desarrollo de las actividades de SigloData y como se va generando valor al producto final.



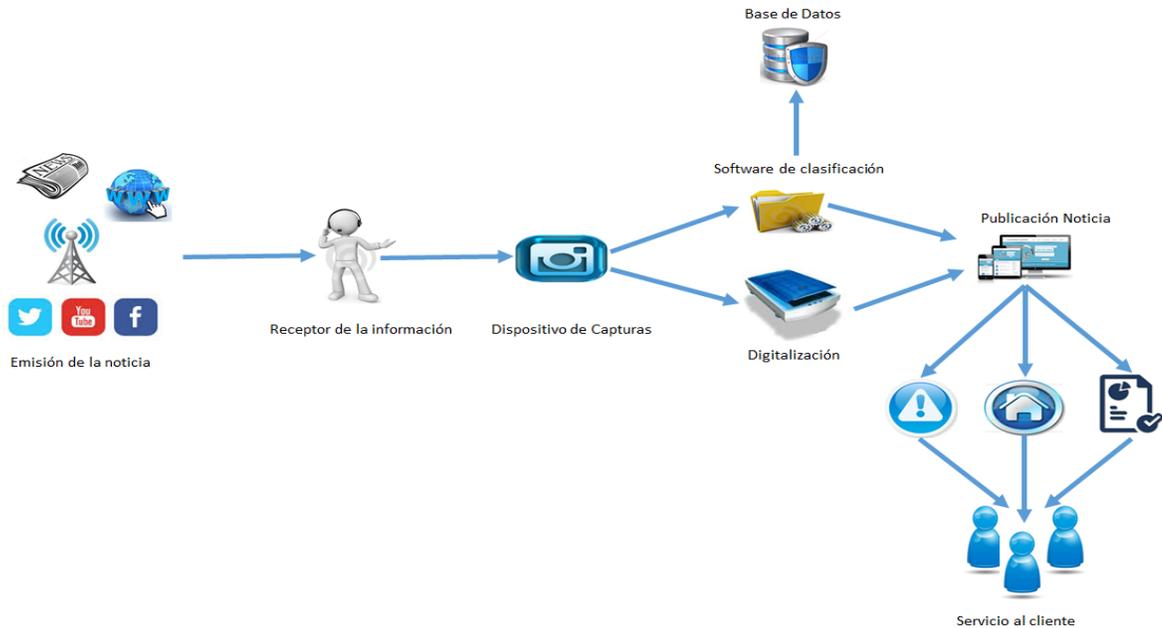
*Baquero; Camargo; García; González; Piracón; Galvis
 Universidad de la Sabana, especialización Gerencia estratégica 2017
 Figura 1. Cadena de valor*

2.1.3 Variables de la cadena de valor

*Baquero; Camargo; García; González; Piracón; Galvis 2017
 Tabla 2. Variables de la cadena de valor*

Variables de la Cadena de Valor	Siglo Data	Punto Cardinal
Emisión de la noticia	Fuerte	Fuerte
Receptor de la información	Fuerte	Fuerte
Dispositivo de capturas	Débil	Fuerte
Digitalización	Fuerte	Fuerte
Software de clasificación	Fuerte	Débil
Base de Datos	Fuerte	Débil
Publicación de la noticia	Fuerte	Débil
Cliente	Fuerte	Fuerte

2.1.4 Flujograma del proceso de monitorización de noticias



*Baquero; Camargo; García; González; Piracón; Galvis 2017.
Figura 2. Proceso de monitorización de noticias*

2.1.5 Matriz de Macrosegmentación y Cadena de Valor en Siglo Data: “La necesidad de un marco metodológico para la gestión de los datos y su explotación con Analítica Avanzada”

Siglo Data se caracteriza por ser una compañía fuerte en el sector en el ámbito de noticias, ya que tiene un importante respaldo tecnológico por MMI en España, lo cual le ha permitido ser más competitivo en la oportunidad, disponibilidad e integridad de la información (Baquero; Camargo; García; González; Piracón; Galvis 2017).

Siglo Data es la única compañía del sector que categoriza y etiqueta toda la información emitida en los medios de comunicación de su cobertura. Sin embargo, esta ventaja resulta siendo punto débil frente a sus competidores cuando se encuentra frente a las dificultades de versatilidad del software en la presentación de la información que sí tienen otras empresas (Baquero; Camargo; García; González; Piracón; Galvis 2017).

2.1.6 Tendencias Mundiales

La evolución de Siglo Data en los últimos 5 años ha tenido una gran relevancia en el mercado de compañías de monitoreo de noticias. La empresa ha tenido crecimiento promedio del 22.5% durante este tiempo. Entre las principales causas de expansión se encuentran:

- Incorporación de servicio de social media en su portafolio de productos.
- Regionalización.
- Incorporación de nuevas herramientas de monitoreo para lograr un mayor alcance y oportunidad.
- Agresiva fuerza de ventas.
- Incorporación de conceptos gráficos en la presentación.
- Crecimiento de la categoría (50 % en estos 5 años).

El sistema de monitoreo de Siglo Data (barrido del medio) es considerado como uno de los más confiables según agremiados a “Fibep”, estamento privado que reúne a las compañías de monitoreo de noticias en el mundo, de la cual es socio MMI. La metodología de clasificar toda la información sin distinguir clientes ni temas obliga al documentalista a prestar atención a todos los protagonistas del medio y no solo a los clientes. El sistema de control de calidad de Siglo Data también verifica la información de todas las categorías y protagonistas y no solo de sus clientes.

El sistema de monitoreo de noticias está cambiando, pasando de un trabajo extensivo en escucha a uno consolidado en las fuentes que influyen en la opinión del grupo primario; para ello los medios de análisis tendrán que estar segmentados por área geográfica, sector económico, gustos, actividades etc. Es así que una compañía de maquillaje tendrá una selección de medios principales, tier. (denominación a los medios de comunicación según su audiencia objetiva, generalmente va de tier 1 a tier 4) muy diferentes a una compañía o institución agraria; es aquí donde las compañías de monitoreo dan valor a sus clientes ubicándose en la escucha específica de su audiencia (Baquero; Camargo; García; González; Piracón; Galvis 2017).

Adicional a ello toda la información debe presentarse de una forma ágil pero impactante en su contenido dejando claro 2 o 3 hallazgos y/o recomendaciones relevantes. La técnica del “*elevator pitch*” está marcando tendencia en la entrega de información (Baquero; Camargo; García; González; Piracón; Galvis 2017).

Data mining juega un factor importante en la elaboración de productos. Esta técnica está basada en análisis de grandes cantidades de información registradas en diversidad de formatos y tiempos (Baquero; Camargo; García; González; Piracón; Galvis 2017).

Todo profesional y empresa dedicada a labores de Comunicación y Relaciones Públicas, sabe que el principal objetivo de su actividad se centra en conquistar y conseguir la simpatía de las audiencias con las que determinado negocio u organización debe de estar vinculado; así como hacer que esa audiencia perciba de forma positiva a la empresa, sus productos y servicios, e incluso a sus líderes. Esta es una labor que se realiza, sobre todo, a través de los

medios de comunicación, que sirven como canales para la difusión de los mensajes que una organización quiere dar a conocer acerca de ella misma y/o de sus productos y servicios (“El Valor del Monitoreo de Noticias en las Estrategias de Comunicación y Relaciones Públicas,” s.f) (Baquero; Camargo; García; González; Piracón; Galvis 2017).

La actividad de monitoreo de noticias es tan antigua como los mismos medios de comunicación, y es que las personas, instituciones y empresas buscan lo que publican los medios de ellos, de su entorno, su competencia etc.

A medida que crecen y se segmentan los medios de comunicación en el mundo se hace más difícil la tarea de rastrear información de forma oportuna y precisa las diferentes fuentes. En Colombia la primera compañía de monitoreo de noticias nace alrededor de los años 60's como empresas de recortes de prensa. En los años 80's, una vez aparecen sistemas de grabación de televisión y radio, son incorporados en el seguimiento de noticias (Baquero; Camargo; García; González; Piracón; Galvis 2017).

A comienzos de los años 2000 cuando se inicia la digitalización de la información, esta actividad se hace más rápida y versátil; ya no era necesario los recortes de prensa, ni la grabación de vídeos y audios en Vhs o cassettes, todas las noticias se canalizan en portales web, y es allí donde se aloja toda la información para tener acceso a ella de manera casi inmediata. La digitalización también conlleva a una expansión de las compañías de monitoreo a nivel internacional, convirtiéndolas en verdaderas fuentes de información (Baquero; Camargo; García; González; Piracón; Galvis 2017).

En el mundo existen cuatro compañías que son hoy día las más fuertes del mercado:

Kantar Media <http://www.kantarmedia.com/es>

Cision <http://www.cision.com/us/>

Acceso <http://monitoreodemedios.acceso.com/>

Ninestar <http://ninestars.in/index.php/media-monitoring/>

Una de las características adicionales que tienen estas empresas es que han incorporado negocios paralelos de investigación de mercados, comunicación estratégica, relaciones públicas y publicidad. Estas actividades les dan músculo financiero que las respaldan y fortalecen.

2.1.7 Tendencias del sector

En los últimos 5 años y dada la relevancia de las redes sociales, se incorpora al negocio otro eje de información relevante, Social Media entra de una forma disruptiva en la comunicación tradicional; ya no son los medios de comunicación entes que manejan la información de manera cerrada y en exclusiva, las redes sociales crean voz a todas las personas, haciendo importante en la información la fuerza de la multitud (Mendes, Augusto, Michelotti & Souza, 2015).

Todo esto genera grandes volúmenes de información que debe ser analizada en tiempo record para agilizar la toma de decisiones. Todo ese Big Data hace importante incorporar herramientas y que ayuden al análisis datos y texto: Data mining y text mining.

2.2 Análisis de Competitividad. El reto de incorporar un marco guía de desarrollo para la gestión de Información en Siglo Data con Analítica Avanzada

2.2.1 Fortalezas y Debilidades

Siglo Data MMI fortalece su presencia nacional con información íntegra y oportuna a nivel nacional y regional (13 corresponsalías), con un servicio al cliente soportado con herramientas de búsqueda de información.

Es la única compañía en Colombia certificada con el ISO 9001 otorgado por el Icontec. El buen servicio es un valor que aprecian los clientes, esto se ve manifestado en las encuestas de percepción. En la actualidad el dashboard no es amigable al usuario y hoy día no permite realizar gráficas ni búsquedas por rango de fechas entre otras funcionalidades. Las otras dos compañías de monitoreo si tienen una herramienta web amigable para el usuario. Siglo Data tiene los precios más altos del sector, lo que hace que el mercado institucional sea muy difícil de captar (Baquero; Camargo; García; González; Piracón; Galvis 2017).

Punto cardinal es una compañía con sede en Medellín con un portal muy moderno y versátil el cual es apreciado por sus clientes. La calidad de la información es impecable. Tiene 2 sucursales en Bogotá y Barranquilla, sin embargo, la oportunidad y la integridad en las noticias regionales no es un punto fuerte de su gestión. No tienen una base completa de información de archivo y el servicio al cliente en momentos no es bueno. Los precios que trabaja esta compañía son intermedios (Baquero; Camargo; García; González; Piracón; Galvis 2017).

GNI es una compañía colombiana respaldada por GNI internacional con sede en Israel. Esta empresa es la tercera en el ranking de facturación; tiene sucursal en Medellín, la calidad de la información y el servicio de apoyo al cliente no es un punto fuerte; sin embargo, tiene un departamento comercial bien establecido y maneja precios bajos (Baquero; Camargo; García; González; Piracón; Galvis 2017).

Bajo la premisa del concepto de monitoreo de noticias SigloData busca incorporar dentro de sus implementos actuales herramientas de analítica avanzada que permita pasar de ser recolector de noticias a ser generador de hallazgos de conocimiento en el campo informativo y noticioso de medios (Baquero; Camargo; García; González; Piracón; Galvis 2017).

2.3 Conceptos básicos de Analítica Avanzada bajo el contexto de la investigación

2.3.1 Analítica de datos

Es la ciencia que examina datos en bruto con el propósito de sacar conclusiones sobre la información. El análisis de datos es usado en varias industrias para permitir que las compañías y las organizaciones tomen mejores decisiones empresariales y también es usado en las ciencias para verificar o reprobador modelos o teorías existentes. El análisis de datos se distingue de la extracción de datos por su alcance, su propósito y su enfoque sobre el análisis. Los extractores de datos clasifican inmensos conjuntos de datos usando software sofisticado para identificar patrones no descubiertos y establecer relaciones escondidas. El análisis de datos se centra en la inferencia, el proceso de derivar una conclusión basándose solamente en lo que conoce el investigador (Lescano & Ávila, 2019).

2.4 Analítica Avanzada

La analítica avanzada puede ayudar a probar ideas, predecir resultados y a crear reportes basados en datos de las grandes redes. Son herramientas que no solo ayudan a resolver la complejidad de los datos, sino que también permiten que los empleados de las empresas puedan usar la información transformada por los ingenieros y generen reportes rápidamente muchas veces con información en tiempo real (Garzón, s.f).

División

2.4.1 Analítica Descriptiva

A través del histórico de datos muestra qué ha pasado con el tema de búsqueda y las razones detrás de este (Ariza, 2019).

2.4.2 Analítica Predictiva

Permite transformar esas métricas descriptivas en un conjunto de previsiones, pronósticos y tendencias precisas; dirá lo que pasará en un futuro (Ariza, 2019).

2.4.3 Analítica Prescriptiva

Identifica las decisiones más óptimas teniendo en cuenta los grandes volúmenes de datos y las infinitas variables y restricciones que se han generado anteriormente (Ariza, 2019)

2.5 Minería de Texto

Define el proceso de identificar patrones significativos en los datos que sean válidos, novedosos, potencialmente útiles y comprensibles para un usuario. Este proceso global consiste en transformar información de bajo nivel en conocimiento de alto nivel (Riquelme, Ruiz & Gilbert, s.f.)

2.5.1 Procesos en la analítica de texto

Los procesos de minería de texto se suelen construir como combinación de una serie de tareas, entre las que se encuentran las siguientes:

2.5.2 El etiquetado gramatical

(En inglés, part-of-speech o PoS tagging) “consiste en identificar la estructura de un texto y en asignar a cada una de las palabras que lo componen su categoría gramatical, en función del contexto en que aparecen”. (Meaning Cloud, s.f)

2.5.3 El clustering o agrupamiento

“Permite descubrir los temas relevantes y las relaciones dentro de una colección de documentos, agrupando estos dentro de un conjunto de documentos similares, internamente homogéneas pero diferentes entre ellas” (Meaning Cloud, s.f).

2.5.4 Clasificación o categorización

“Consiste en asignar un texto a una o varias categorías entre una taxonomía predefinida teniendo en cuenta el contenido global del texto” (Meaning Cloud, s.f).

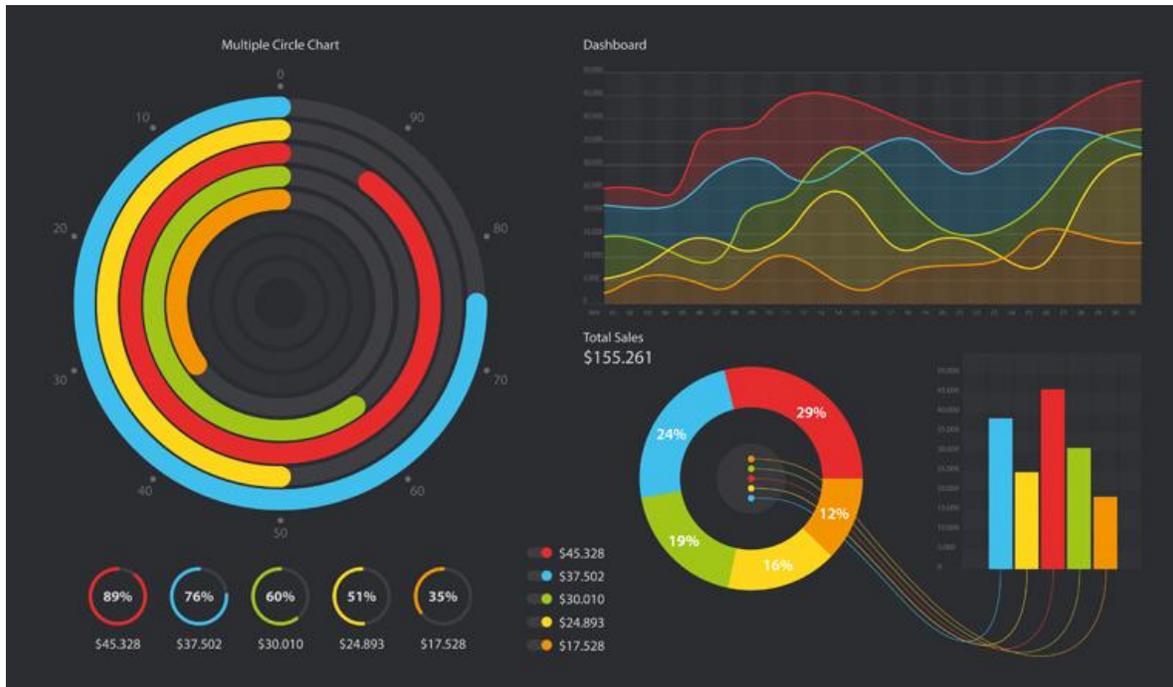
2.5.5 La extracción de información

“Consiste en identificar dentro del texto apariciones de entidades (nombres propios de personas, empresas, marcas, lugares), conceptos abstractos y otros elementos de información específicos: cantidades, relaciones, etc. Se utiliza para detectar menciones e identificar aquellos elementos que llevan el mayor peso informativo de un texto” (Meaning Cloud, s.f).

2.5.6 El análisis de sentimiento

“Consiste en detectar la polaridad (positiva, negativa, neutra o ausencia de polaridad) que lleva un documento. Esta polaridad puede ser debida a una opinión subjetiva o a la expresión de un hecho objetivo de uno u otro signo” (Meaning Cloud, s.f.).

2.8 Visualización de resultados



Sergio Ruiz, 2018

Figura 4. Tableau: visualización de datos en gráficos radiales.

La visualización de datos es la presentación gráfica de información con dos propósitos. Por un lado, la interpretación y construcción de significado a partir de los datos (es decir, el análisis); y por otro lado, la comunicación. La visualización es una herramienta muy potente para descubrir y comprender la lógica que se encuentra detrás de un conjunto de datos, así como para compartir esta interpretación con otras personas desde un punto de vista objetivo. Como se suele decir: “una imagen vale más que mil palabras”, especialmente cuando el significado que se quiere comunicar se representa mejor gráfica que verbalmente, y esta representación gráfica está diseñada de acuerdo a los principios formales que rigen la visualización de datos (Iniciativa aporta; red.es, 2016).

2.9 KPIs y la importancia en la investigación

KPIs (key performance indicator, KPI por sus siglas en inglés), es decir, indicadores clave de rendimiento que, a través de mediciones, nos dan información sobre si una determinada acción está yendo por el camino correcto.

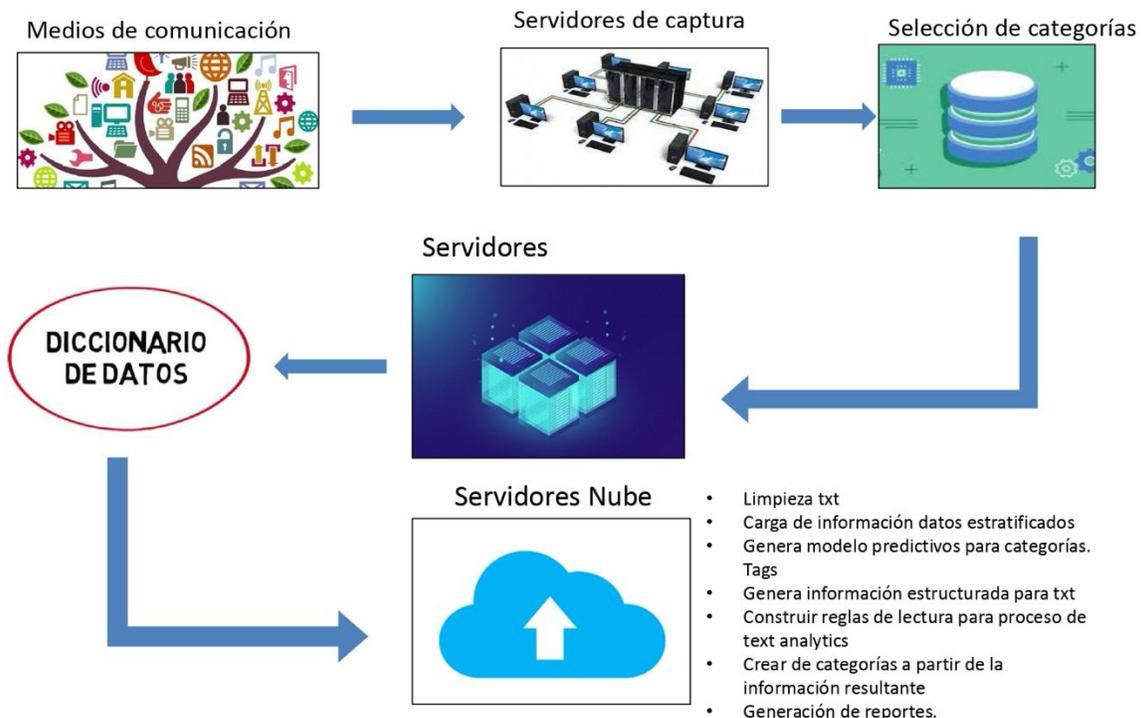
Gracias a los KPIs es posible conocer si se están alcanzando los objetivos o por el contrario, si es necesario cambiar la estrategia. Por lo tanto **obtener datos e informes dinámicos** ayuda a identificar el cumplimiento de metas. Los principales KPIs están clasificados en dos grandes grupos: **KPIs financieros** y **KPIs de marketing y ventas** (Tecon: soluciones informáticas, s.f.)

Cualquier actividad debe ser cuantificable, y por ende medible. Susceptible de establecer KPIs para el monitoreo de su gestión y mejora a través del tiempo: “lo que no se mide no se mejora”.

Capítulo 3. Desarrollo y propuesta del Marco metodológico para la Gestión de la Información

3.1 La Gestión de los Procesos – Visión Macro

3.1.1 Gestión de procesos



Javier Galvis, 2020.

Figura 5. Gestión en procesos

3.1.2 Medios de comunicación

Base de recolección de información: Señales de televisión ,señales de radio, impresos de prensa y revistas. Estas son herramientas de recolección de noticias en portales web.

3.1.3 Servidores de captura

En televisión y radio la captura de señal se realiza por Dvrs para digitalizarla y adaptarla a los formatos requeridos. En prensa la publicación física es digitalizada.

3.1.4 Selección de categoría

Los videos de televisión, los audios de radio ,los pdfs de prensa y las noticias de internet se procesan y etiquetan según los tags base de recolección.

3.1.5 Servidores

Las noticias procesadas ya estandarizadas y clasificadas, junto con los recursos, son pasados a los servidores donde está alojada la plataforma y el dashboard de consulta.

3.1.6 Diccionario de datos

Diariamente la plataforma genera un resumen estratificado en formato CSV del resumen de las noticias categorizadas y etiquetadas. Cada noticia y sus características se repite según el número de categorías que tenga.

3.1.7 Servidores nube

El CSV entra a la herramienta analítica que está en nube .Allí se hace la limpieza del txt, se estructura, se generan los modelo predictivos, se construyen las reglas de lectura para el text analytics, se crean las categorías a partir de la información resultante y se generan los reportes.

3.2 Marco metodológico para la gestión de información y analítica avanzada

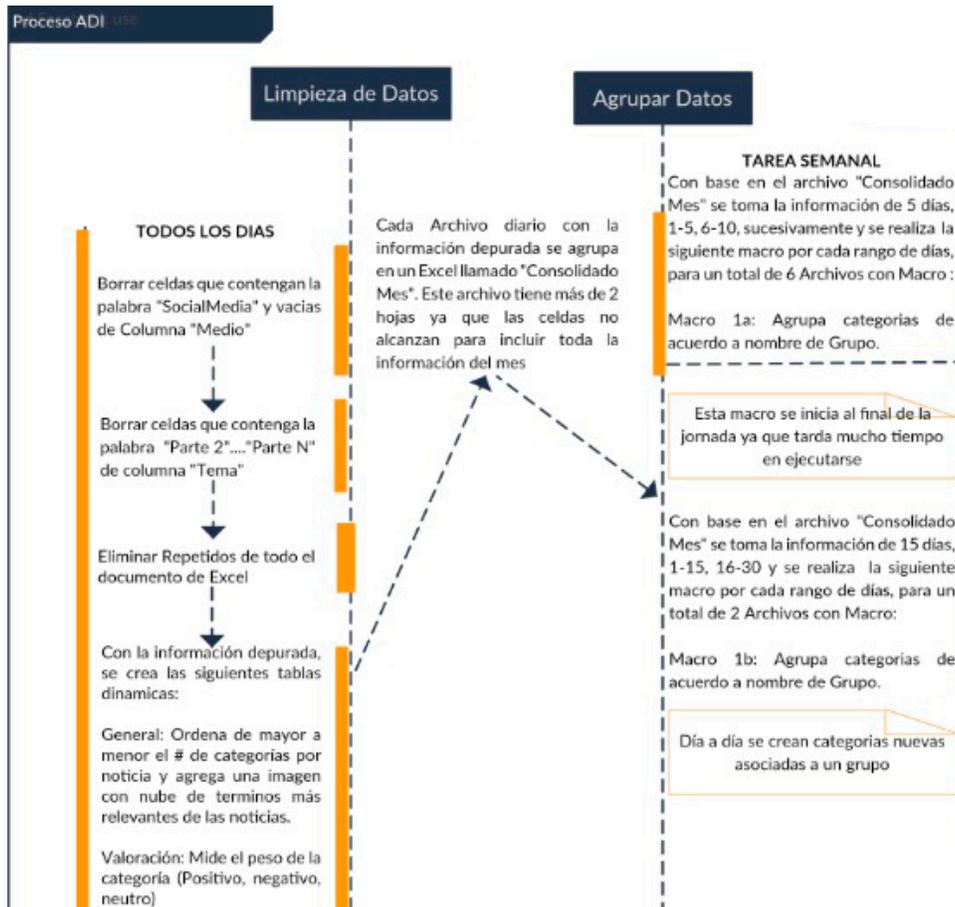
3.2.1 Importar datos de los tags generados diariamente (volumen de 60 mil registros diarios)

La plataforma base envía por correo de forma estructurada y en formato Excel o Csv el resumen de la gestión documental en los márgenes de tiempo que se requieran.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
CATEGORIA	TEMA	SOPORTE	FECHA	MEDIO	SECCION	VPE COSTE	PESO	PAGINA/SEGI	AUDIENCIA	ID_NOTA	URL			
LABORAL	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo			4	250254	9472	http://co.mmi-e.com/accesopdf.php?PRUTA=VOLSI	
ECONOMIA	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo			4	250254	9472	http://co.mmi-e.com/accesopdf.php?PRUTA=VOLSI	
COMERCIO	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo			4	250254	9472	http://co.mmi-e.com/accesopdf.php?PRUTA=VOLSI	
SOCIEDAD	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo			4	250254	9472	http://co.mmi-e.com/accesopdf.php?PRUTA=VOLSI	
TRANSPORTE	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo			4	250254	9472	http://co.mmi-e.com/accesopdf.php?PRUTA=VOLSI	
TECNOLOGIA	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo			4	250254	9472	http://co.mmi-e.com/accesopdf.php?PRUTA=VOLSI	
TURISMO	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	negativo			4	250254	9472	http://co.mmi-e.com/accesopdf.php?PRUTA=VOLSI	
MINISTERIO DE LA PROTEC	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo			4	250254	9472	http://co.mmi-e.com/accesopdf.php?PRUTA=VOLSI	
DESEMPLEO	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	negativo			4	250254	9472	http://co.mmi-e.com/accesopdf.php?PRUTA=VOLSI	
INFLACION	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	negativo			4	250254	9472	http://co.mmi-e.com/accesopdf.php?PRUTA=VOLSI	
CONTROL INTERNO	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	neutro			4	250254	9472	http://co.mmi-e.com/accesopdf.php?PRUTA=VOLSI	
BANCO DE LA REPUBLICA	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo			4	250254	9472	http://co.mmi-e.com/accesopdf.php?PRUTA=VOLSI	
FMI (FONDO MONETARIO	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo			4	250254	9472	http://co.mmi-e.com/accesopdf.php?PRUTA=VOLSI	
DANE (DEPARTAMENTO AD	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo			4	250254	9472	http://co.mmi-e.com/accesopdf.php?PRUTA=VOLSI	
SIPSA SISTEMA DE INFORN	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo			4	250254	9472	http://co.mmi-e.com/accesopdf.php?PRUTA=VOLSI	

Tabla 3.
SigloData, 2020

3.2.2 Limpiar los datos importados de acuerdo con las reglas de Negocio



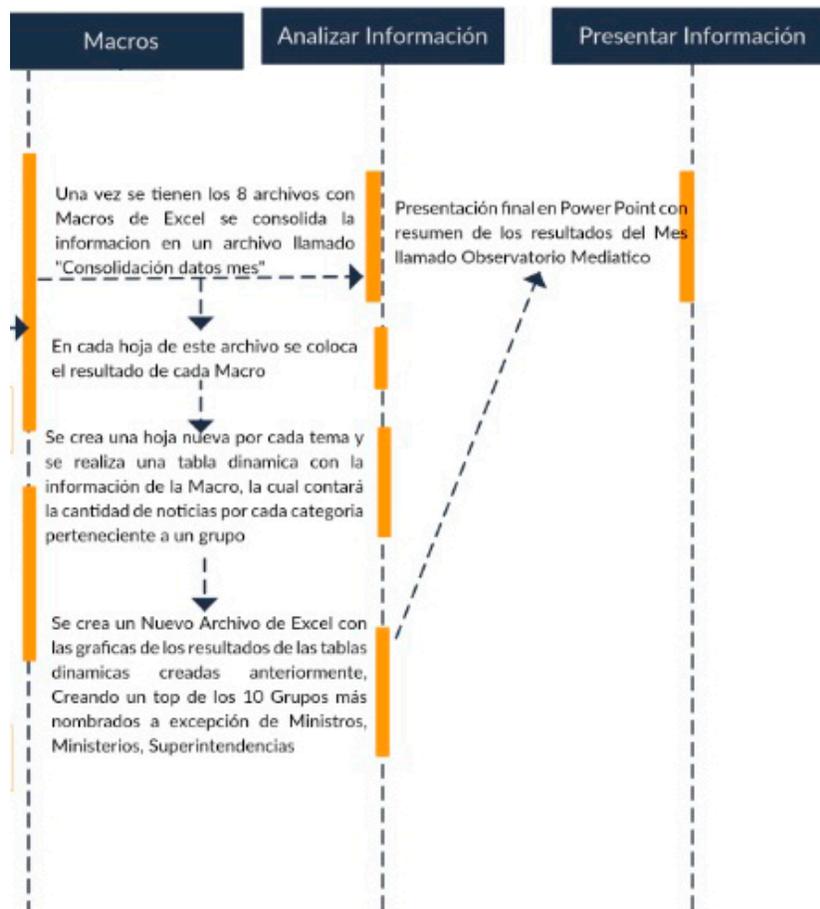


Tabla 4.
Bigband data sas

- En la limpieza de datos se eliminan las filas que contengan la palabra “*parte + #*” se unifican en una sola y se suma el coste de las filas o canal.
- La columna soporte tiene la fuente del medio pero en información de radio y tv aparece la franja horario mas no el nombre de la emisora .Hay entonces que agregar una columna nueva y según su franja horaria aparecerá el nombre de la emisora.
- Con la información de la nueva columna, se crean 2 columnas adicionales: la primera se incluye la ciudad y el medio y la segunda el cubrimiento del medio (nacional-regional).

3.2.3 Cargar información de medios prensa nacionales en formato .txt.

Este proceso es posible mediante el formato OCR (del inglés Optical Character Recognition) cuyo proceso esta dirigido a la digitalización de textos los cuales se identifican automáticamente a partir de las imágenes de periódicos convertidos en formato pdf.

3.2.4 Limpiar y ajustar los .txt cargados de acuerdo con la estructura de cada periódico y las reglas de negocio establecidas (caracteres especiales y mala redacción)

Según la gráfica 3.2.2 los textos son organizados y depurados con técnicas de autocompletar palabras y borrado de texto con % de caracteres ilegibles. Sobre los textos se corre las queries de categorías o etiqueta del diccionario principal. Las palabras o frases que se ajustan a una query son seleccionadas e incluidas en la base de datos.

3.2.5 Modelación con los 3 tipos de diccionario de datos

Existe un diccionario principal compuesto por 11.582 atributos que lo componen empresas, instituciones, voceros y temas.

A Partir de este diccionario se generan 3 diccionarios adicionales (Tipo , sector primario y sector secundario) con cualidades económicas, políticas o sectoriales de cada uno de los 11.582 atributos principales.

3.2.6 Generar información estructurada a partir de los .txt.

Los txt aplicables a las categorías y asociados a los diccionarios son los que utilizan para la generación de reportes.

3.2.7 Construir reglas de negocio para la lectura de la información previamente estructurada para procesarla a través Analytics

La creación del diccionario y la visualización del mismo generando cluster en la agrupación de las etiquetas permite originar reportes:

- Coyunturales.
- Sectoriales.
- Temáticos.
- Personajes.
- Combinar por y /o etiquetas y categorías.

3.3 Creación del diccionario

Según el portal Dblearner (2012) “el diccionario de datos es un repositorio de metadatos que contiene las definiciones de los objetos de datos, descripciones y relaciones entre sí”. El diccionario es muy importante en el desarrollo de aplicaciones, ya que permite a los involucrados en el proceso conocer a detalle las bases de datos.

En esta investigación, la base de selección de información está compuesta por cada una de las noticias que emiten los medios de comunicación ya sea prensa, radio, televisión e internet. Cada noticia se trabaja como una unidad y se incluye en la plataforma (herramienta que se mostrará en su capítulo respectivo) y son el input para la creación de 2 diccionarios relacionados que constituyen el primer aporte de este trabajo.

En el proyecto de la creación del método de analítica de noticias se trabajó con un diccionario base que se componen de dos partes:

- a) Selección de monitoreo: Estandarización de cada una de las partes que componen la selección de las noticias recolectadas en las tablas de datos de la plataforma.
- b) Etiquetado de atributos: Ampliación dentro del diccionario de categorías en 3 atributos adicionales. Esto son: tipo, sector primario y sector secundario. La adaptación de estas 3 nuevas cualidades permitirá una mayor segmentación de los objetos alcanzando la profundidad deseable para su análisis.

3.3.1 Atributo de cada uno de los datos

Símbolo	Significado
=Ó :=	Composición: está compuesto o es equivalente a
+	Inclusión : Y
[]	Selección
{ }	Opción
*Texto *	Comentario en texto
#	Alfanumérico
#	Numérico
\$	Moneda
%	Porcentaje
	Categoría relacionada

Javier Galvis, 2020.
Tabla 5. Diccionario de datos

3.3.2 Diccionario selección de monitoreo

1) Diccionario selección de monitoreo

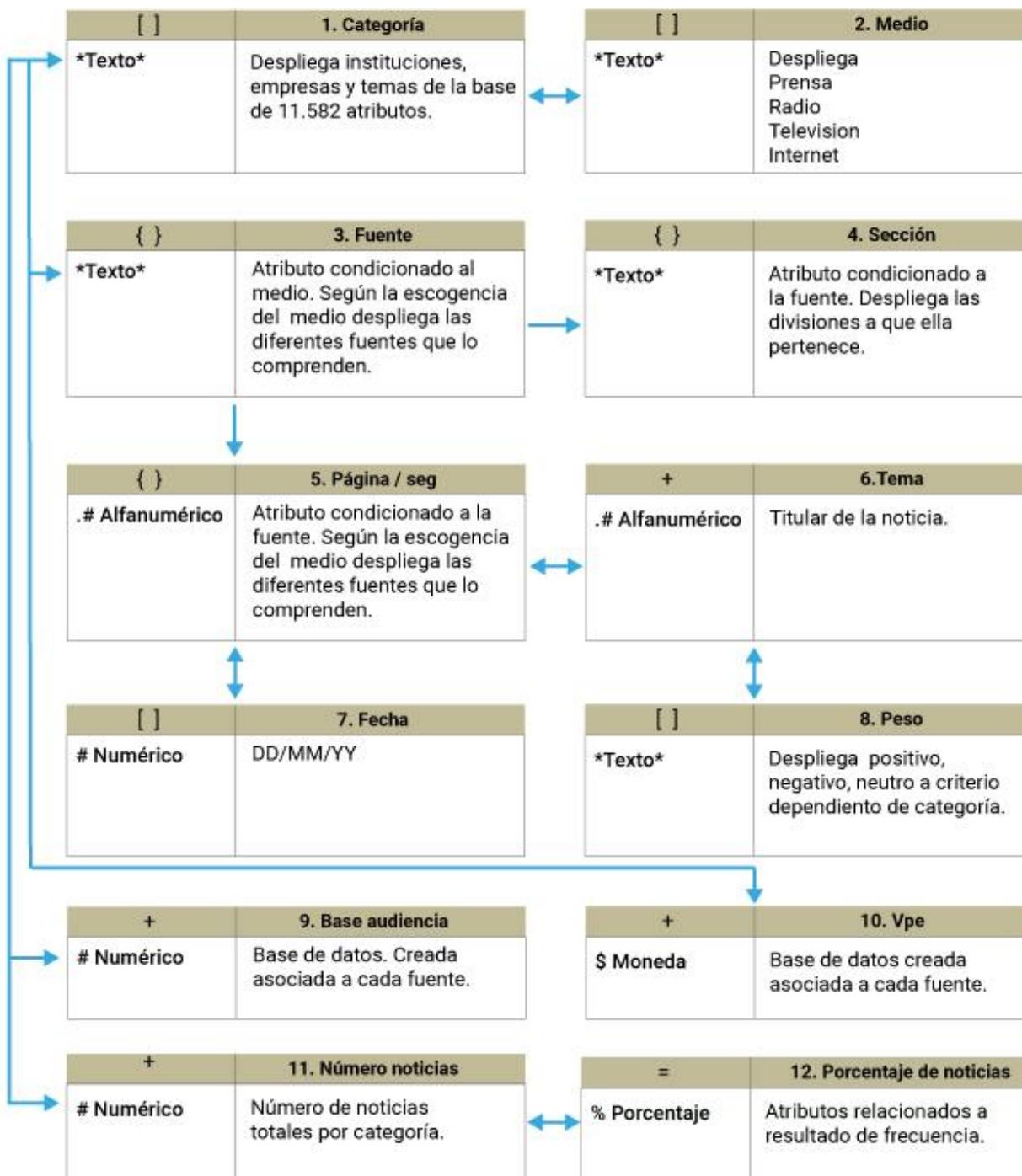


Figura 6. Manuela Galvis y Javier Galvis, 2020. Selección de monitoreo

3.3.3 Diccionario etiquetado por atributos

2) Diccionario etiquetado por atributos

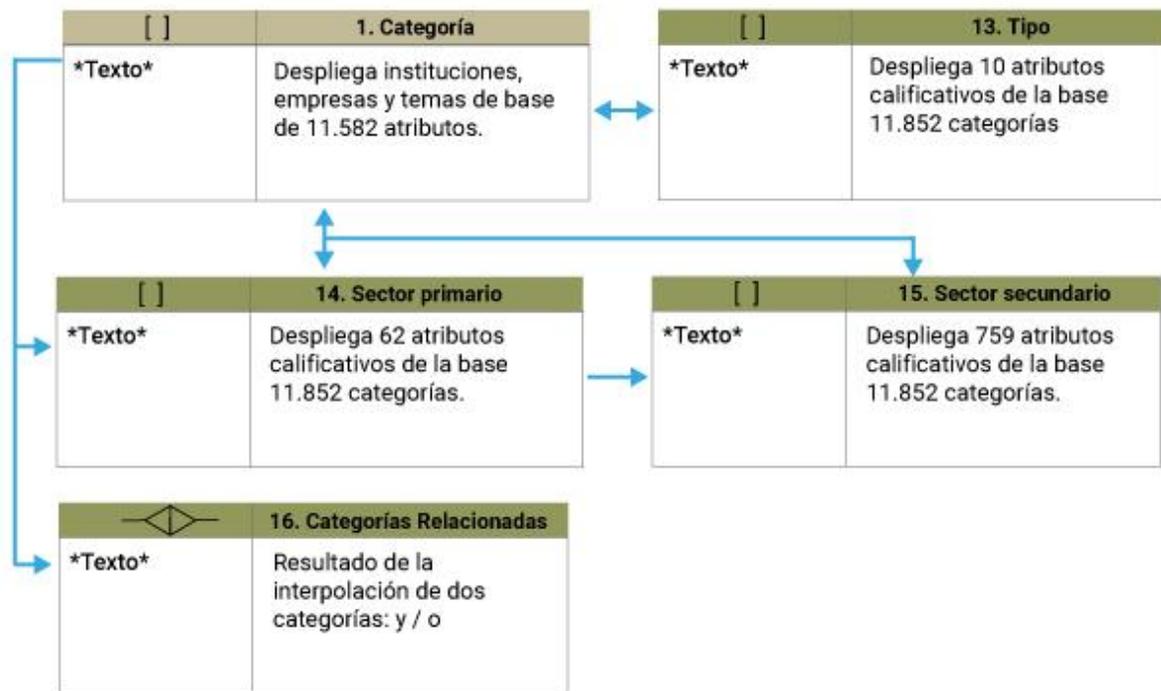


Figura 7. Manuela Galvis y Javier Galvis, 2020. Etiquetas por atributos

3.3.4 Esquema global de clasificación de noticias

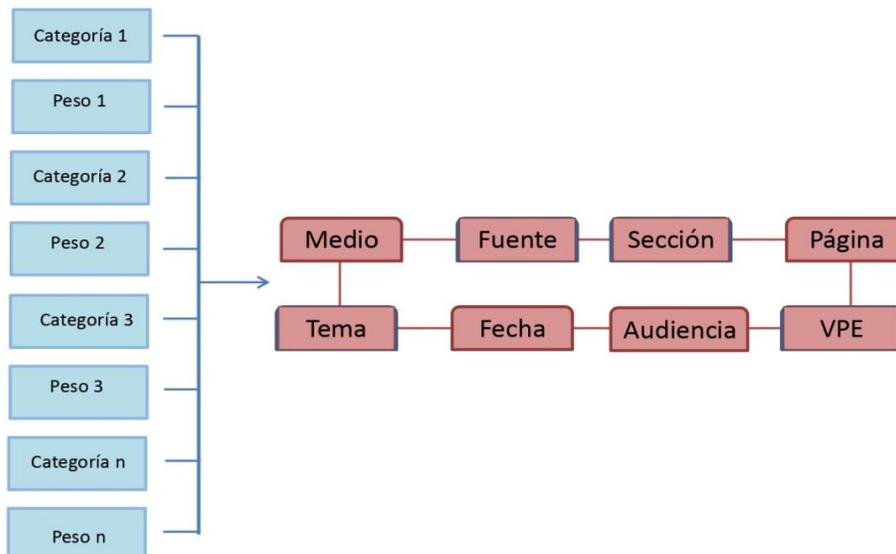


Figura 8. Javier Galvis, 2020 . Clasificación de noticias global.

3.4 Diccionario selección de monitoreo

La base de selección de información es cada una de las noticias que emiten los medios de comunicación ya sea prensa, radio, televisión e internet. Cada noticia se trabaja como una unidad y se incluye en la plataforma dándole las características que a continuación se describen.

3.4.1 Categoría

Este es el aspecto principal en la implementación del sistema. La categoría (o etiquetas) es la selección automática o manual de los atributos que componen una noticia y que la convierten en particularmente única. La selección objetiva de ellas hace que no sea necesario la lectura de toda la noticia (medios impresos, internet) oírla y/o verla (radio y televisión) para tener conocimiento objetivo de su temática. El conjunto sistemático de la categorización de todas las noticias es la base que enriquece la herramienta de analítica pues logra estructurar en datos la información emitida desde diferentes formatos.

3.4.2 Medio

Atribuye a cada categoría la procedencia original del formato, este puede ser: internet, radio, prensa o televisión.

3.4.3 Fuente

Atribuye al medio seleccionado la fuente que la compone. Cada una de ellas tiene característica según su cobertura geográfica (nacional, regional, local) o temática (económica, política, deportiva, social, general, entre otros). El medio prensa la componen los diferentes periódicos y revistas de cobertura nacional, regional, local o sus temáticas. En radio las cadenas de noticias de ámbito nacional y regional o especializado. En televisión son las cadenas nacionales, regionales y especializadas. En internet puede ser nacional, regional y especializada. La fuente al permitir la segmentación de la categoría facilita focalizar el resultado de la información.

3.4.4 Sección

Característica de ubicación dentro de la fuente. La sección logra segmentar aún más la categoría pues le atribuye propiedades de localización dentro de su espacio. Las fuentes

tienen secciones al igual que los medios geográficos o temáticas como pueden ser sección nacional internacional, regional , económicas , deportivas etc.

3.4.5 Página / seg

Atributo de dimensión o espacio de la noticia seleccionada que está enmarcada dentro de la sección, la fuente y el medio. Para los medios impresos de prensa se divide en cms x columnas, donde los centímetros es el alto de la nota y la columnas el ancho .Para revistas las dimensiones son página completa, 1/2 pagina, 1/4 de página. Radio y televisión el atributo son el número de segundos que tiene de duración cada noticia. Para internet no se tiene en cuenta estos atributos.

3.4.6 Vpe

Valor Publicitario Equivalente. Este es un atributo numérico de valor estimado. Los espacios en medios tienen un costo publicitario que es el que el medio vende a los anunciantes. Estos valores están dados por atributos de audiencia (número de personas que ven, oyen o leen una determinada fuente) por el medio, la fuente y la sección .Todas estas particularidades hace que la fuente le dé un valor para los anunciantes. Este costo está determinado también por extensión; o sea la duración o tamaño de la pieza publicitaria. En el análisis de noticias el VPe se utiliza para darle al atributo un carácter cuantitativo. La sumatoria de todos los VPe que alcance una categoría a través del volumen de noticias donde tenga presencia será un atributo de estudio que se tendrá en cuenta.

3.4.7 Base audiencia

Similar que el Vpe. la base de audiencia es el número de personas estimados que ven, oyen o leen una determinada fuente. A diferencia del Vpe este valor no esta determinado por la sección ni por la extensión de la noticia. Es un valor único por cada fuente. Este valor está prediseñado en la base de datos y se adiciona directamente cuando se es seleccionada la fuente. La sumatoria de todas las audiencias que alcance una categoría a través del volumen de noticias donde tenga presencia es también atributo importante de alcance.

3.4.8 Tema

Es un Espacio Alfanumérico para digitar de forma automática o manual el titular de cada una de las noticias.

3.4.9 Fecha

Es un espacio numérico automático de la fecha de la noticia. Por defecto incluye la fecha de inclusión. En caso de requerir incluir información de fechas anteriores la herramienta lo permite.

3.4.10 Peso

Atributo calificativo de la categoría dentro de la noticia. Esta valoración afecta a la categoría y no a la noticia. El peso puede ser automático (según query de frases y términos cercanos a la categoría) o manual a criterio del documentalista. Este atributo es polémico por la misma subjetividad que lo compone, pero es un aspecto importante pues marca tendencias valorativas de la categoría.

3.4.11 Número de noticias

Atributo numérico. Suma las categorías de todos los registros encontrados en un lapso de tiempo sumando uno o más atributos específicos.

3.4.12 % de noticias

Atributos relacionales a resultados de frecuencia. % de registros encontrados vrs. base total de registros.

3.5 Diccionario Etiquetado por atributos

3.5.1 Categoría

La categoría del diccionario de selección de monitoreo es la base del diccionario de etiquetado. 11.582 atributos que lo componen instituciones, empresas y temas seleccionadas de cada una de las noticias monitoreadas y recolectadas.

3.5.2 Tipo

Cada uno de estos 11.582 atributos es a su vez categorizado en una de las 10 propiedades calificativas que lo componen.

3.5.3 Sector primario

De forma secuencial nuevamente cada uno de los 11.582 atributos base son categorizados en una de las 62 propiedades calificativas con mayor alcance que las 10 anteriores propiedades del número: 2 Tipo.

3.5.4 Sector Secundario

De forma secuencial cada uno de los 11.582 atributos base son categorizados por tercera vez en uno de las 759 atributos dando una mayor especificación de agrupación.

3.6 Resultado esperado

La segmentación de los datos utilizados en los 3 tipos de diccionarios aquí descritos conlleva a obtener una serie de hallazgos que podrán ser interpolados y valorados con información adicional que la misma herramienta permite desarrollar.

3.6.1 Técnicas analíticas para trabajar la información

3.6.1.1 Base de datos relacionales

Se parte de la base de datos compuesta de una colección de tablas (relaciones), los datos son almacenados en filas y columnas. Los sistemas de gestión de bases de datos relacionales almacenan la data de forma estructurada.

3.6.1.1 Tabla

GRUPOS DE CATEGORIAS	CATEGORIA	PESO	positivo	neutro	negativo	Total
	ASOBANCARIA		112	114	.	226
	CORONAVIRUS / COVID -19		10	50	20	80
	ECONOMIA		40	13	1	54
	SANTIAGO CASTRO GOMEZ ASOBANCARIA		28	22	.	50
	BANCOS		34	8	1	43
	SUPERFINANCIERA (SUPERINTENDENCIA FINANCIERA		23	13	.	36
	CREDITOS/FINANCIACION		29	5	.	34
	TARJETAS DE CREDITO		13	17	.	30
	TASAS DE INTERES		17	12	.	29
← Otros	BANCO BILBAO VISCAYA ARGENTARIA BBVA		7	17	.	24
	WHATSAPP		21	3	.	24
	BANCA Y FINANZAS		17	5	1	23
	INFRAESTRUCTURA		3	17	.	20
	COMERCIO ELECTRONICO		4	15	.	19
	EMPRESAS		16	3	.	19
	ESTADO		18	.	1	19
	REDES SOCIALES		19	.	.	19
	BANCO DE LA REPUBLICA		12	5	.	17
	SUPER INTENDENCIA FINANCIERA		5	12	.	17
	E-COMMERCE - COMERCIO ELECTRONICO		4	12	.	16

SAS VIYA 3.4, 2020.

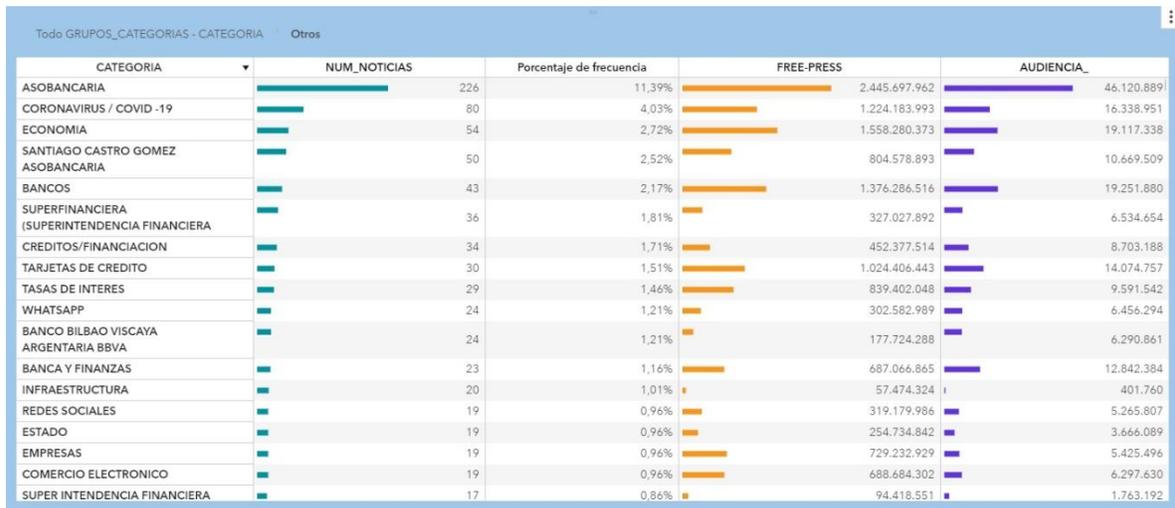
Tabla 6. Grupos de categorías

Visualización del grupo principal del diccionario: categorías con resultados de atributos de los diccionarios escogidos. En el ejemplo se modelo con:

- Categoría

- Fecha
- Peso
- Número de noticias

3.6.1.2 Gráfico de barras

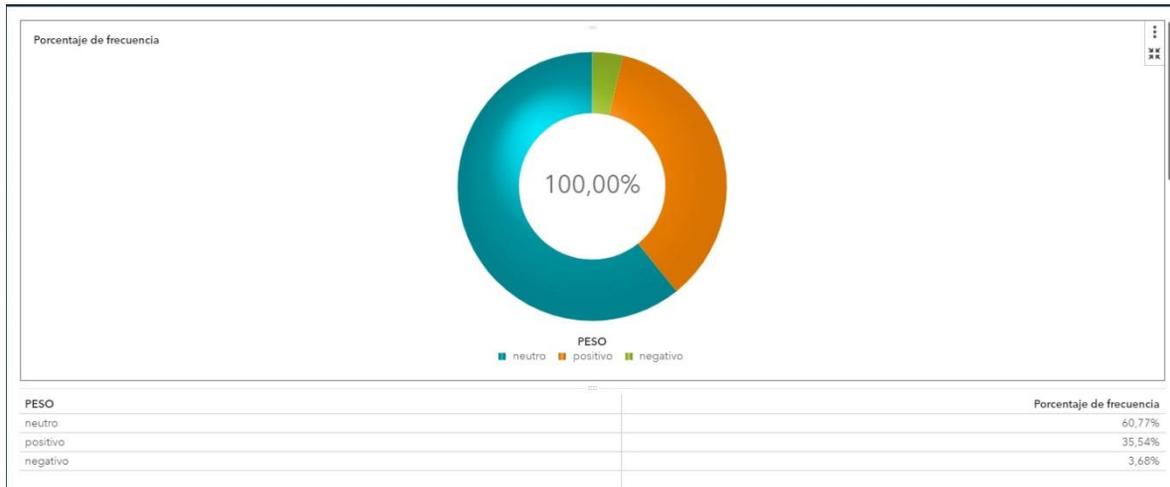


SAS VIYA 3.4, 2020.

Figura 9. Visualización de atributos clasificados

Visualización de atributos clasificados en base al grupo principal del diccionario: Categorías. La categoría se rankea de mayor a menor por número total de noticias que contienen esta etiqueta, porcentaje de frecuencia, Vpe y audiencia.

3.6.1.3 Gráfico circular



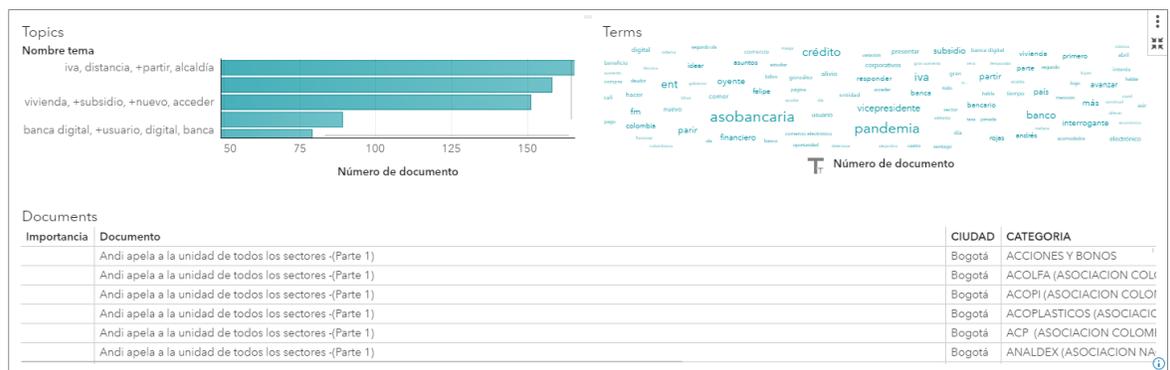
SAS VIYA 3.4, 2020.

Figura 10. Clasificación por peso (positivo, negativo, neutro)

Visualización de los atributos de la base de la categoría relacionado con:

- Fecha
- Número de noticias
- Peso
- Porcentaje de frecuencia

3.6.1.4 Tópicos



SAS VIYA 3.4, 2020.

Figura 11. Nube de términos y clasificación de tópicos

Visualización de resultados de los atributos de la categoría base con los diccionario de selección de monitoreo. La nube de término es generada de los textos de las noticias seleccionadas y también con el ranquin de tópicos de palabras más utilizadas en la selección.

3.6.1.5 Categorías relacionadas

CATEGORIA	NUM_NOTICIAS	FREE-PRESS	AUDIENCIA_
ASOBANCARIA	44	1.376.846.516	19.331.675
BANCOS	43	1.376.286.516	19.251.880
ECONOMIA	23	987.888.938	11.595.450
CREDITOS/FINANCIACION	16	274.356.158	4.163.995
SANTIAGO CASTRO GOMEZ ASOBANCARIA	13	643.542.383	5.903.408

SAS VIYA 3.4, 2020.

Figura 12. Selección y/o entre categorías

Visualización de los resultados de la interpolación de y/o de dos categorías. Las categorías relacionadas permiten unir dos etiquetas y generan un ranquin grupal donde es posible determinar el número categorías diferentes a las seleccionadas que tienen ascendencia sobre estas. También permite conocer en el global de noticias el porcentaje y número de ellas que contienen las dos etiquetas.

Capítulo 4. Validación del modelo por medio de la herramienta de Analítica

Para demostrar la utilidad de dicho modelo, esta investigación utilizará la herramienta de analítica dentro de la empresa Siglo Data. Se buscara como primera medida introducir a la compañía, sus funciones y su potencial para la puesta en práctica del modelo, y, posteriormente, el uso de la plataforma de analítica para su validación.

4.1 “Siglo Data” La empresa como caso de uso de la investigación

Historia de la empresa

Siglo Data nació en el año 1998 apéndice de una central de medios que se llamaba CIGLO Comunicaciones, para venderle a los clientes de publicidad el servicio de monitoreo de noticias de su negocio. Cuando comenzó la crisis del año 1999-2000, en la cual influyó la entrada de canales privados como Caracol y RCN y nuevas centrales de medios, el negocio cambió. A raíz de esto se decidió reforzar la compañía de monitoreo de noticias, y cerrar la compañía de pauta de medios en el año 2000 (Baquero; Camargo; García; González; Piracón; Galvis 2017).

En el comienzo, el envío de la información era físico, es decir recortes de periódicos, VHS y grabaciones de audio en cassettes. En el año 2004 se migró a la página web de la compañía como canal de entrega a los clientes de la información, lo cual fue inmanejable debido a los elevados volúmenes de información, siendo necesaria la búsqueda de nuevas alternativas. A raíz de esto se contactó a la empresa MMI de España, una de las más importantes en monitoreo de medios en Europa, quienes a partir del 2006, brindan el servicio de plataforma de carga, consulta y almacenamiento de la información emitida en medios. Hoy en día, el 50% de la información que se almacena en MMI pertenece a Colombia, más específicamente a Siglo Data (Baquero; Camargo; García; González; Piracón; Galvis 2017).

En el año 2010 se adquirieron los servicios de una compañía llamada Augure para la lectura de noticias en portales de internet, la cual se incorporó a la aplicación de la herramienta de MMI, con el fin de ampliar el monitoreo de medios a noticias que se encontraran en los sitios web. En el año 2013, se contrató el servicio de seguimiento de social media (twitter, Facebook, Instagram, YouTube, entre otros) con la empresa Brand Watch, para lograr una mayor cobertura en los servicios de monitoreo incluyendo el seguimiento de redes sociales (Baquero; Camargo; García; González; Piracón; Galvis 2017).

En agosto 2016, Siglo Data compró la mayoría de las acciones de la empresa número 7 del mercado a nivel nacional: Infomedios, con la idea de abarcar el mercado institucional público, en el cual Siglo Data no compite por precios en el mercado, así como para entrar en el negocio de tracking publicitario que era un nicho de mercado sin profundizar aún (Baquero; Camargo; García; González; Piracón; Galvis 2017).

En la actualidad hay 12 compañías de monitoreo de medios en Colombia, de las cuales Siglo Data ocupa el puesto No. 2 por cifras en ventas, con una participación del 31% del mercado y aproximadamente 180 clientes del sector privado y público, agremiaciones, instituciones y agencias de comunicación estratégica, relaciones públicas y de publicidad (Baquero; Camargo; García; González; Piracon; Galvis 2017).



Figura 13. Comparativo de ventas anuales de las compañías de Monitoreo en Colombia Superintendencia de sociedades-Cámara de comercio .2018

4.1.1 Misión

Somos una compañía colombiana dedicada al monitoreo, clasificación, análisis y archivo, tanto de noticias en medios de comunicación masivos, a nivel regional y nacional, como de análisis de información en redes sociales.

2.1.2 Visión

Ser en el 2022 un Centro de Pensamiento Sectorial Estratégico, basado en la información recolectada en medios de comunicaciones y en redes sociales, para convertirnos en un aliado de nuestros clientes, que soporte la toma de sus decisiones, orientando nuestra labor con nuevos modelos de gestión como la seguridad de la información y la responsabilidad social empresarial.

4.1.3 Objetivos Actuales

A continuación, se detallan los objetivos de SigloData:

- Lograr la satisfacción de nuestros clientes.

- Alcanzar los resultados esperados en el cubrimiento y presentación de la información, cumpliendo los requisitos establecidos para nuestros servicios y productos.
- Asegurar el envío oportuno de la información.
- Desarrollar y mantener la infraestructura física y tecnológica requerida para la operación.
- Consolidar aliados estratégicos que cumplan los requisitos de la organización.
- Consolidar un equipo de trabajo competente con resultados satisfactorios en su desempeño.
- Atender las solicitudes y manifestaciones de satisfacción de los clientes y lograr los niveles de ocurrencia mínimos esperados de quejas y reclamos.
- Lograr la eficacia de las acciones de mejora que se generen en los procesos.
- Alcanzar los resultados de ventas y financieros esperados por la junta directiva para la operación de la compañía.
- Esta investigación aportará a SigloData herramientas claves que ayudarán a soportar su gestión empresarial otorgando ventajas adicionales a sus clientes aportando conocimiento; aspecto clave en el proceso en la toma de decisiones.

Siglo Data cuenta actualmente con un sistema de monitoreo denominado “*barrido por categorías*”. El sistema consiste en la clasificación de cada noticia emitida en los medios de cobertura: Prensa, Radio, Televisión e Internet. Cada noticia es etiquetada. La base de etiquetación la componen 11.582 atributos; a su vez cada uno de ellos se reclasifica en 3 niveles de precisión así:

4.2 Clasificación de Etiquetas

4.2.1 Tipo

NOMBREWEB	ACTIVIDAD	Tipo	Sector primario	Sector
Araujo Ibarra & Asociados S.A.			Servicios profesionales	Abogados
Baker & Mc Kenzie - Abogados			Servicios profesionales	Abogados
Brigard & Urrutia Abogados			Servicios profesionales	Abogados
Cardenas & Cardenas Abogados			Servicios profesionales	Abogados
DLA Piper Martínez Beltrán			Servicios profesionales	Abogados
Gómez - Pinzón Zuleta Abogados			Servicios profesionales	Abogados
Pinilla Gonzalez & Prieto Abogados			Servicios profesionales	Abogados
Posee & Herrera Abogados			Servicios profesionales	Abogados
GAGEL LAW FIRM			Servicios profesionales	Abogados
LEGOPS – INTEGRA LEGAL ABOGADOS			Servicios profesionales	Abogados
Lloreda Camacho & Abogados			Servicios profesionales	Abogados
Zarama & Asociados			Servicios profesionales	Abogados
Abimgra			Agroindustria	Abonos
Abocol (Abonos Colombianos S.A.)	1	Empresa	Agroindustria	Abonos
Aborganicos		Empresa	Agroindustria	Abonos
Cremalleras Rey	1	Empresa	Confeccion vestuario textil	Accesorios
Induboton S.A.	1	Empresa	Confeccion vestuario textil	Accesorios
Indulanas	1	Empresa	Confeccion vestuario textil	Accesorios
Nefromédicas Ltda	1	Empresa	Salud	Accesorios
Sun Microsystems de Colombia S.A.	1	Empresa	Tecnología	Accesorios

Javier Galvis, 2019.

Tabla 7. Selección de atributos en categoría tipo

Cada uno de las 11.582 categorías se clasifica según lo que corresponda con un atributo único inicial. Estos son:

- Empresa.
- Eventos.
- Institución.
- Ong's.
- Personaj

- Política.
- Producto.
- Tema.
- Ubicación geográfica.
- Vocero.

4.2.2 Sector Primario

Al igual de la anterior esa misma etiqueta se subdivide en 62 atributos de mayor extensión.

A	B	C	D	E	F
NOMBREWEB	ACTIVIDAD	Tipo	Sector primario	Sector	
Araujo Ibarra & Asociados S.A.	1	Empresa	Se	Abogados	
Baker & Mc Kenzie - Abogados	1	Empresa	Se	Abogados	
Brigard & Urrutia Abogados	1	Empresa	Se	Abogados	
Cardenas & Cardenas Abogados	1	Empresa	Se	Abogados	
DLA Piper Martínez Beltrán	1	Empresa	Se	Abogados	
Gómez - Pinzón Zuleta Abogados	1	Empresa	Se	Abogados	
Pinilla Gonzalez & Prieto Abogados	1	Empresa	Se	Abogados	
Posee & Herrera Abogados	1	Empresa	Se	Abogados	
GAGEL LAW FIRM	1	Empresa	Se	Abogados	
LEGOPS – INTEGRA LEGAL ABOGADOS	1	Empresa	Se	Abogados	
Lloreda Camacho & Abogados	1	Empresa	Se	Abogados	
Zarama & Asociados	1	Empresa	Se	Abogados	
Abimgra	1	Empresa	Aq	Abonos	
Abocol (Abonos Colombianos S.A.)	1	Empresa	Agroindustria	Abonos	
Aborganicos	1	Empresa	Agroindustria	Abonos	
Cremalleras Rey	1	Empresa	Confeccion vestuario textil	Accesorios	
Induboton S.A.	1	Empresa	Confeccion vestuario textil	Accesorios	
Indulanes	1	Empresa	Confeccion vestuario textil	Accesorios	

Javier Galvis, 2019.

Tabla 8. Selección de atributos en sector primario.

- | | | |
|------------------------|----------------------------|---------------------------|
| 1) Accesorio automóvil | 8) Aseo personal | 15) Cajas de compensación |
| 2) Accesorio hogar | 9) Astronomía geología | 16) Ciencia |
| 3) Agroindustria | 10) Banca y finanzas | 17) Colombia |
| 4) Alimentos | 11) Bebidas alcohólicas | 18) Comercio |
| 5) Arte y cultura | 12) Bebidas ni alcohólicas | 19) Comercio Exterior |
| 6) Artículos oficina | 13) Belleza | 20) Comunidad |
| 7) Aseo hogar | 14) Borrar | |

*Para ver mayor detalle de estas tablas consulte anexo 1

4.2.3 Sector Secundario

El sector secundario nos da una mayor profundidad al etiquetado de las categorías logrando una mayor precisión en la búsqueda de información. (752 atributos).

1	NOMBREWEB	ACTIV	Tipo	Sector primario	Sector Secundario
2	Araujo Ibarra & Asociados S.A.	1	Empresa	Servicios profesionales	Aboga
3	Baker & Mc Kenzie - Abogados	1	Empresa	Servicios profesionales	Aboga
4	Brigard & Urrutia Abogados	1	Empresa	Servicios profesionales	Aboga
5	Cardenas & Cardenas Abogados	1	Empresa	Servicios profesionales	Aboga
6	DLA Piper Martínez Beltrán	1	Empresa	Servicios profesionales	Aboga
7	Gómez - Pinzón Zuleta Abogados	1	Empresa	Servicios profesionales	Aboga
8	Pinilla Gonzalez & Prieto Abogados	1	Empresa	Servicios profesionales	Aboga
9	Posee & Herrera Abogados	1	Empresa	Servicios profesionales	Aboga
10	GAGEL LAW FIRM	1	Empresa	Servicios profesionales	Aboga
11	LEGOPS – INTEGRA LEGAL ABOGADOS	1	Empresa	Servicios profesionales	Aboga
12	Lloreda Camacho & Abogados	1	Empresa	Servicios profesionales	Aboga
13	Zarama & Asociados	1	Empresa	Servicios profesionales	Aboga
14	Abimgra	1	Empresa	Agroindustria	Abonc
15	Abocol (Abonos Colombianos S.A.)	1	Empresa	Agroindustria	Abonos

Javier Galvis, 2019.

Tabla 9. Selección de atributos sector secundario.

A su vez y dependiendo de la clasificación inicial cada una de las 11.582 categorías se subdivide en:

- | | | |
|------------------------------|--|--------------------------|
| 1) Abogados. | 7) Suministros. | 13) Aguardiente ron. |
| 2) Accesorios. | 8) Administradora riesgos profesionales. | 14) ahorro y crédito. |
| 3) Accesorio ferreterías. | 9) Aerolíneas. | 15) Alfombras. |
| 4) Aceite mesa y cocina. | 10) Aeropuertos. | 16) Alimentación animal. |
| 5) Acero aluminio. | 11) Agencia de viajes. | 17) Analítica. |
| 6) Acueducto alcantarillado. | 12) Agronomía | 18) Apicultura. |
| | | 19) App. |
| | | 20) Arquitectura |

**Para ver mayor detalle de estas tablas consulte anexo 2*

4.2.4 Resultado y aporte

La planeación y elaboración de los diccionarios anteriormente expuestos es uno de los aportes al trabajo de grado. Una vez terminados se pusieron a prueba con las herramientas de analítica avanzada y como resultados mostro aspectos fundamentales en la presentación de los hallazgos de la información mediática.

4.3 Ejemplos o usos del dashboard para el consumo de la información de forma analítica

La categoría Avianca

A	B	C	D	E	F
NOMBREWEB	AC.TI.	Tipo	Sector primario	Sector Secundario	
Avianca	1	Empresa	Transporte aereo		Aerolinea

Javier Galvis, 2020.

Tabla 10. Ejemplo empresa

Avianca es una:

Empresa – Clasificada como de transporte aéreo y dentro de transporte aéreo es una aerolínea.

4.3.1 Tipo empresas

A	B	C	D	E	F
NOMBREWEB	AC.TI.	Tipo	Sector primario	Sector Secundario	
Araujo Ibarra & Asociados S.A.	1	Empresa	Servicios profesionales	Abogados	
Baker & Mc Kenzie - Abogados	1	Empresa	Servicios profesionales	Abogados	
Brigard & Urrutia Abogados	1	Empresa	Servicios profesionales	Abogados	
Cardenas & Cardenas Abogados	1	Empresa	Servicios profesionales	Abogados	
DLA Piper Martínez Beltrán	1	Empresa	Servicios profesionales	Abogados	
Gómez - Pinzón Zuleta Abogados	1	Empresa	Servicios profesionales	Abogados	
Pinilla Gonzalez & Prieto Abogados	1	Empresa	Servicios profesionales	Abogados	
Posee & Herrera Abogados	1	Empresa	Servicios profesionales	Abogados	
GAGEL LAW FIRM	1	Empresa	Servicios profesionales	Abogados	
LEGOPS – INTEGRA LEGAL ABOGADOS	1	Empresa	Servicios profesionales	Abogados	
Lloreda Camacho & Abogados	1	Empresa	Servicios profesionales	Abogados	
Zarama & Asociados	1	Empresa	Servicios profesionales	Abogados	
Abimgra	1	Empresa	Agroindustria	Abonos	
Abocol (Abonos Colombianos S.A.)	1	Empresa	Agroindustria	Abonos	
Aborganicos	1	Empresa	Agroindustria	Abonos	
Cremalleras Rey	1	Empresa	Confeccion vestuario textil	Accesorios	
Induboton S.A.	1	Empresa	Confeccion vestuario textil	Accesorios	
Indulanas	1	Empresa	Confeccion vestuario textil	Accesorios	
Nefromédicas Ltda	1	Empresa	Salud	Accesorios	

Javier Galvis, 2020.

Tabla 11. Ejemplo selección de atributos en categoría tipo

Etiquetada como empresas, con el mismo atributo de Avianca se tienen en el directorio 5.090 etiquetas adicionales.

4.3.2 Sector primario

NOMBREWEB	ACTIV	Tipo	Sector primario	Sector Secundario
Aeroespacial	1	Empresa	Transporte aereo	Aeroespacial
GRUPO THALES	1	Empresa	Transporte aereo	Aeroespacial
Aces	1	Empresa	Transporte aereo	Aerolinea
Aero Medellin	1	Empresa	Transporte aereo	Aerolinea
Aero Northwest	1	Empresa	Transporte aereo	Aerolinea
Aero Sucre	1	Empresa	Transporte aereo	Aerolinea
Aero Sur	1	Empresa	Transporte aereo	Aerolinea
Aero Territoriales de Colombia	1	Empresa	Transporte aereo	Aerolinea
Aerocali S.A	1	Empresa	Transporte aereo	Aerolinea
Aerocentro de Colombia S.A.	1	Empresa	Transporte aereo	Aerolinea
AeroContinente	1	Empresa	Transporte aereo	Aerolinea
Aeroenvios S.A.	1	Empresa	Transporte aereo	Aerolinea
Aeroexpreso	1	Empresa	Transporte aereo	Aerolinea
Aerolínea Air Plus Comet	1	Empresa	Transporte aereo	Aerolinea

Javier Galvis, 2020.

Tabla 12. Ejemplo selección de atributos en categoría primaria

Dentro de la categoría transporte aéreo como esta clasificada Avianca existen en el directorio 179 etiquetas adicionales.

4.3.3 Sector secundario

NOMBREWEB	ACTIV	Tipo	Sector primario	Sector Secundario
Aces	1	Empresa	Transporte aereo	Aerolinea
Aero Medellin	1	Empresa	Transporte aereo	Aerolinea
Aero Northwest	1	Empresa	Transporte aereo	Aerolinea
Aero Sucre	1	Empresa	Transporte aereo	Aerolinea
Aero Sur	1	Empresa	Transporte aereo	Aerolinea
Aero Territoriales de Colombia	1	Empresa	Transporte aereo	Aerolinea
Aerocali S.A	1	Empresa	Transporte aereo	Aerolinea
Aerocentro de Colombia S.A.	1	Empresa	Transporte aereo	Aerolinea
AeroContinente	1	Empresa	Transporte aereo	Aerolinea
Aeroenvios S.A.	1	Empresa	Transporte aereo	Aerolinea
Aeroexpreso	1	Empresa	Transporte aereo	Aerolinea
Aerolínea Air Plus Comet	1	Empresa	Transporte aereo	Aerolinea
Aerolínea Copa	1	Empresa	Transporte aereo	Aerolinea
AEROLINEA DE ANTIOQUIA - ADA	1	Empresa	Transporte aereo	Aerolinea
AEROLINEA FAST COLOMBIA	1	Empresa	Transporte aereo	Aerolinea
AEROLINEA INTERJET	1	Empresa	Transporte aereo	Aerolinea
Aerolínea Regional Express	1	Empresa	Transporte aereo	Aerolinea
Aerolínea SAM	1	Empresa	Transporte aereo	Aerolinea
Aerolínea Satena	1	Empresa	Transporte aereo	Aerolinea

Javier Galvis, 2020.

Tala 13. Ejemplo selección de atributos en categoría secundaria

112 etiquetas de nivel secundario están registradas en aerolíneas. Todas ellas pertenecen al transporte aéreo y son empresas como Avianca.

Estos niveles de profundidad dan la facilidad de agrupar las etiquetas para traer la información específica de análisis.

4.3.4 Proceso de documentación.

Selector Web. Aplicativo web donde se realiza la gestión de documentación.

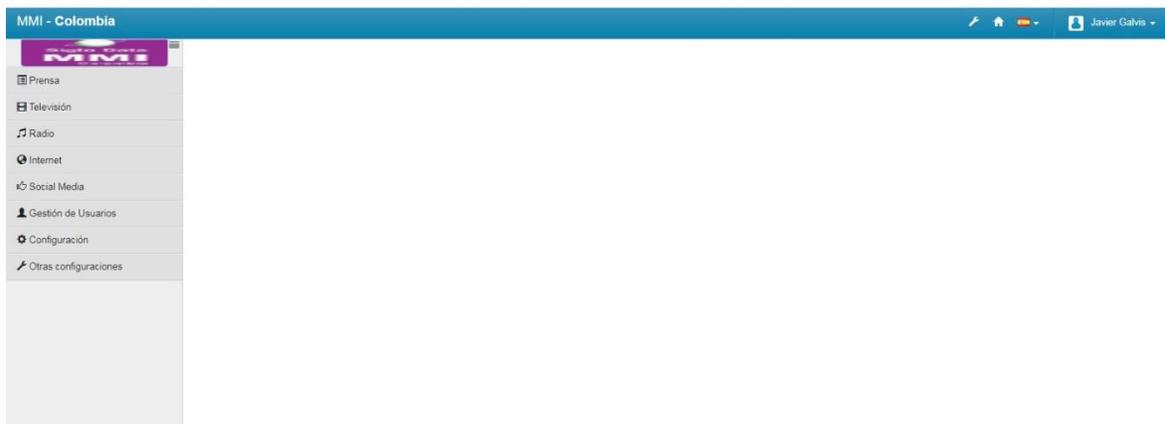


Figura 14. Selector MMI, 2015

Página de entrada donde se presentan alternativas de inclusión de noticias por medios

4.3.5 Proceso de selección de categorías

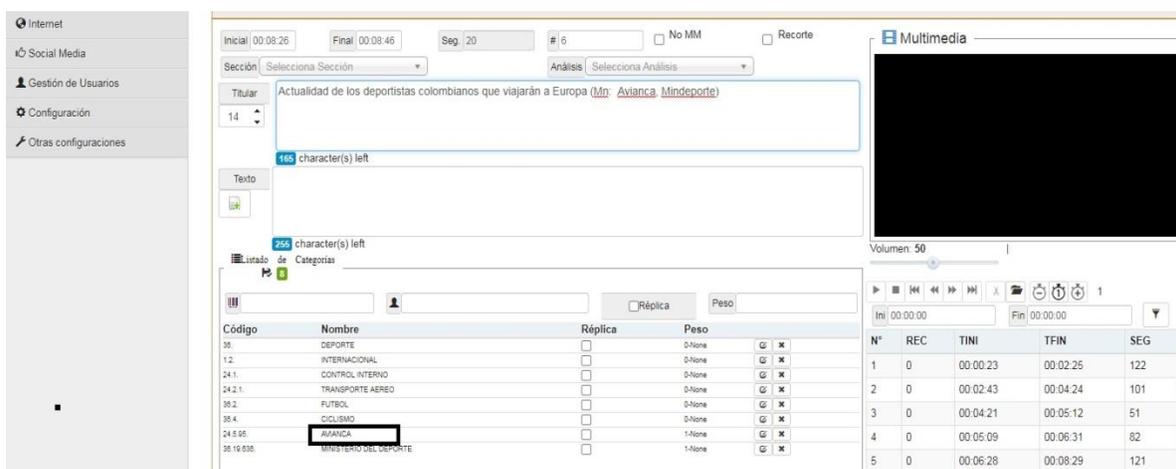


Figura 15. Selector de televisión MMI, 2015

Herramienta web que permite la documentación, selección e inclusión de los videos

4.3.6 Visualización en el dashboard



Figura 16. Dashboard MMI, 2015
Ejemplo de noticia de televisión en el servicio del cliente

4.3.7 Explotación de la Data

Mediante datos estructurados la herramienta se programa para el envío de un correo diario con el resumen de la información estratificada.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CATEGORIA	TEMA	SOPORTE	FECHA	MEDIO	SECCION	VPE COSTE	PESO	PAGINA/SEG:	AUDIENCIA	ID_NOTA	URL			
2	LABORAL	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo	4	250254	9472	http://co.mmi-e.com/accesopdf.php?RUTA=VOLSNI			
3	ECONOMIA	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo	4	250254	9472	http://co.mmi-e.com/accesopdf.php?RUTA=VOLSNI			
4	COMERCIO	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo	4	250254	9472	http://co.mmi-e.com/accesopdf.php?RUTA=VOLSNI			
5	SOCIEDAD	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo	4	250254	9472	http://co.mmi-e.com/accesopdf.php?RUTA=VOLSNI			
6	TRANSPORTE	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo	4	250254	9472	http://co.mmi-e.com/accesopdf.php?RUTA=VOLSNI			
7	TECNOLOGIA	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo	4	250254	9472	http://co.mmi-e.com/accesopdf.php?RUTA=VOLSNI			
8	TURISMO	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	negativo	4	250254	9472	http://co.mmi-e.com/accesopdf.php?RUTA=VOLSNI			
9	MINISTERIO DE LA PROTEC	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo	4	250254	9472	http://co.mmi-e.com/accesopdf.php?RUTA=VOLSNI			
10	DESEMPEÑO	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	negativo	4	250254	9472	http://co.mmi-e.com/accesopdf.php?RUTA=VOLSNI			
11	INFLACION	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	negativo	4	250254	9472	http://co.mmi-e.com/accesopdf.php?RUTA=VOLSNI			
12	CONTROL INTERNO	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	neutro	4	250254	9472	http://co.mmi-e.com/accesopdf.php?RUTA=VOLSNI			
13	BANCO DE LA REPUBLICA	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo	4	250254	9472	http://co.mmi-e.com/accesopdf.php?RUTA=VOLSNI			
14	FMI (FONDO MONETARIO)	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo	4	250254	9472	http://co.mmi-e.com/accesopdf.php?RUTA=VOLSNI			
15	DANE (DEPARTAMENTO AD)	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo	4	250254	9472	http://co.mmi-e.com/accesopdf.php?RUTA=VOLSNI			
16	SIPSA SISTEMA DE INFORM	El valor de la información en momentos d El Espectador		07/04/2020	PRENSA	Economía	16.440.000,00	positivo	4	250254	9472	http://co.mmi-e.com/accesopdf.php?RUTA=VOLSNI			

MMI, 2010
Tabla 14. Data estructurada de selección de categorías y atributos

Dependiendo del número de categorías o etiquetas que tenga cada noticia, esta se multiplica en las siguientes variables:

- 1) Categoría o etiqueta.
- 2) Tema: Titular de la nota.
- 3) Soporte fuente de publicación.

- 4) Fecha.
- 5) Medio.
- 6) Sección de emisión dentro de la fuente y medio.
- 7) VPE.coste. Costo free-press.
- 8) Peso. Calificación dentro de la noticia (positiva , negativa o neutra).
- 9) Página / segundo. Ubicación de la noticia.
- 10) Id de la nota.
- 11) Url al testigo.

4.4 Herramienta Analytics

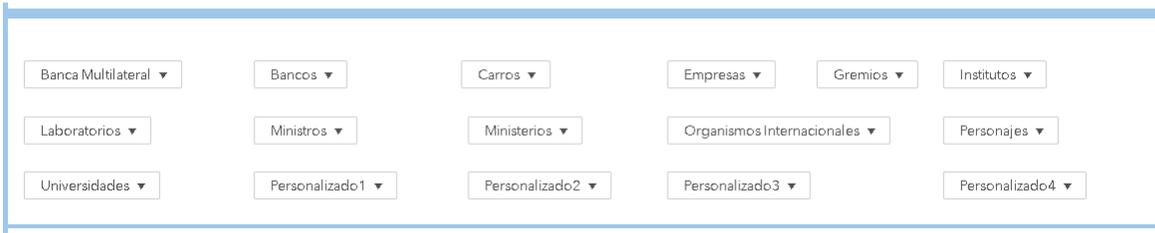
La Herramienta de Analytics está conformada por dos módulos (Módulo Analytics - Módulo Gráfico) la cual podrá consultar e interactuar entre sí.

4.4.1 Módulo Analytics

Este módulo da la opción de obtener datos cuantitativos como: Categorías asociadas, Tópicos y Nube de Términos de la empresa de interés, empresas del sector o competencia y sector donde se encuentra su nicho de mercado.

El módulo Analytics da la facilidad de filtrar diferentes opciones como:

- Año.
- Mes.
- Día.
- Cubrimiento (Nacional, Regional, Internacional).
- Ciudad.
- Medio (Prensa, Radio, Televisión e Internet).
- Peso (Positivo, Negativo, Neutro).
- Fuentes en medios impresos.
- Fuentes de radio.
- Fuentes de televisión.
- Fuentes Online.
- Sección donde fue publicada la noticia en medios impresos.
- Franja horaria donde fue emitida la noticia en medios de radio y televisión.
- Grupos de selección de interés.



*Figura 17.
Dashboard SAS VIYA 3.4 -Big Bang Data, 2019
Visualización de los árboles que se pueden generar con los diccionarios creados*

Al poder interactuar con las diferentes opciones de filtro podrá obtener diferentes informes y hallazgos (Descripción de noticias por grupos de selección de interés; Nube de categorías asociadas al sector donde se encuentra su nicho de mercado; Análisis de Categorías asociadas a la marca seleccionada; Tópicos y Nube de términos de la categoría principal o de interés; Búsqueda de varios términos o titulares con la opción de y/o; Informes predictivos).

4.4.2 Informe Descripción de noticias por grupos de selección de interés

En este informe encontrará información gráfica cuantitativa de los grupos seleccionados encontrando número de noticias, % de frecuencia, Free-press y audiencia.

- Trazabilidad de los grupos de selección de interés.
- Peso (Positivo, negativo, neutro) de los grupos de selección de interés.
- Medio (Prensa, Radio, Televisión e Internet) de los grupos de selección de interés.
- Top 10 de las fuentes donde se generó mayor visibilidad mediática de los grupos de selección de interés.

4.4.3 Nube de Categorías

Encontrará una nube de categorías con los tags incluidos dentro de los grupos de selección de interés. En esta opción también podrá encontrar unos filtros de selección como: Grupo o grupos de interés, Peso (Positivo, negativo, neutro) y Medio (Prensa, radio, televisión e internet).



Big Bang Data- SAS VIYA 3.4. 2020
Figura 18. Nube de términos

4.4.4 Análisis de categorías asociadas

En este informe podrá detectar cuáles fueron las categorías o tags asociadas a la categoría o tag principal. Esta categoría o tags está relacionada al contenido de la noticia. Esta información viene acompañada del total de noticias, % de frecuencia, Free-press y Audiencia.

- Peso (Positivo, negativo, neutro) de la categoría o tags principal y asociadas.
- Medio (Prensa, Radio, Televisión e Internet) de la categoría o tags principal y asociadas.
- Top 10 de las fuentes donde se generó mayor visibilidad mediática de la categoría o tags principal y asociadas.

GRUPOS DE CATEGORÍAS	CATEGORIA	NUM_NOTICIAS	Porcentaje de frecuencia	FREE-PRESS	AUDI
Otros	MINISTERIO DE LA PROTECCION SOCIAL - MINISTERIO DE SALUD	52	7,37%	738.958.310	
	AZUCAR	37	5,24%	521.080.268	
	PRESIDENTE IVAN DUQUE MARQUEZ	32	4,53%	456.254.325	
	ANDI (ASOCIACION NACIONAL DE EMPRESARIOS DE COLOMBIA)	27	3,82%	482.433.508	
	CAMARA DE LA INDUSTRIA DE ALIMENTOS DE LA ANDI	26	3,68%	294.748.397	
	MINISTERIO DE COMERCIO INDUSTRIA Y TURISMO	25	3,54%	125.270.342	
	CAMARA DE INDUSTRIA DE BEBIDAS DE LA ANDI	21	2,97%	288.606.394	

Big Bang Data- SAS VIYA 3.4. 2020.
Figura 19. Visualización de atributos clasificados

BÚSQUEDA DE VARIOS TÉRMINOS EN TITULARES O CATEGORÍAS

Búsqueda en Categorías

Seleccione la primera palabra (en titulares) o categoría que quiera buscar:

Búsqueda en Titulares

Seleccione la segunda palabra (en titulares) o categoría que quiera buscar:

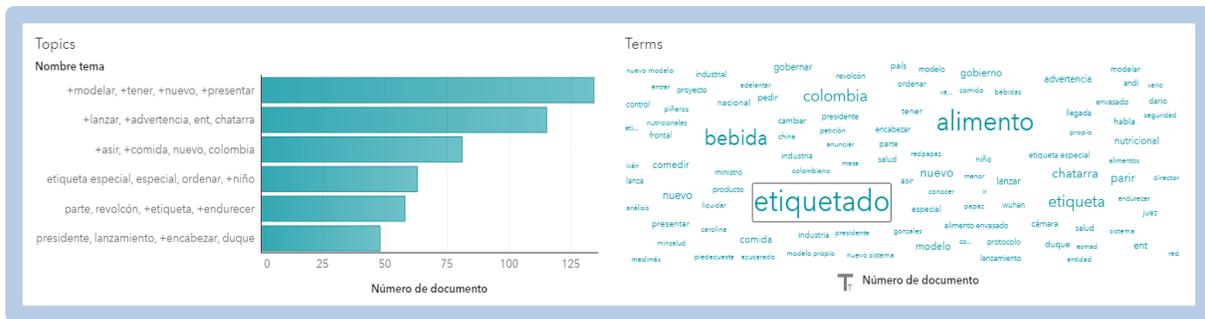
O

Seleccione la tercera palabra (en titulares) o categoría que quiera buscar:

Y

Seleccione la cuarta palabra (en titulares) o categoría que quiera buscar:

*Big Bang Data- SAS VIYA 3.4, 2020.
Figura 21. Selección y/o entre categorías.*



*Big Bang Data- SAS VIYA 3.4, 2020.
Figura 22. Nube de términos y clasificación de tópicos*

4.5 Módulo Gráfico

Este módulo da la opción de obtener datos cuantitativos de la empresa de interés, empresas del sector o competencia y sector donde se encuentra su nicho de mercado. El módulo Gráfico da la facilidad de filtrar diferentes opciones como:

- Año.
- Mes.
- Día.
- Cubrimiento (Nacional, Regional, Internacional).
- Ciudad.
- Medio (Prensa, Radio, Televisión e Internet).
- Peso (Positivo, Negativo, Neutro).
- Fuentes en medios Impresos.

- Fuentes de radio.
- Fuentes de televisión.
- Fuentes Online.

Al poder interactuar con las diferentes opciones de filtro podrá obtener dos tipos de informes (Informe de Categoría Principal – Informe Categoría principal y competencia).

4.5.1 Informe de Categoría Principal

En este informe encontrará información gráfica cuantitativa de la marca como:

- Exposición en medios (Prensa, radio, Televisión, e Internet).
- Trazabilidad mensual.
- Trazabilidad diaria.
- Cubrimiento (Nacional, regional, internacional).
- Ciudad.
- Peso (Positivo, negativo, neutro).
- Detallado por peso.
- Exposición mediática fuentes en medios impresos.
- Exposición mediática fuentes de radio.
- Exposición mediática fuentes de televisión.
- Exposición mediática fuentes online.

4.5.2 Informe de Categoría Principal y Competencia

En este informe encontrará información gráfica cuantitativa de la marca, competencia y sector:

- Exposición mediática comparativa por la empresa de interés, empresas del sector o competencia y sector donde se encuentra su nicho de mercado.
- Detallado de las empresas que conforman el informe (Empresa de interés, empresas del sector o competencia y sector donde se encuentra su nicho de mercado).
- Exposición mediática comparativa por medios (Prensa, radio, televisión e internet) de la empresa de interés, empresas del sector o competencia y sector donde se encuentra su nicho de mercado.
- Exposición mediática comparativa por fuentes en medios impresos de la empresa de interés, empresas del sector o competencia y sector donde se encuentra su nicho de mercado.
- Exposición mediática comparativa por fuentes de radio de la empresa de interés, empresas del sector o competencia y sector donde se encuentra su nicho de mercado.
- Exposición mediática comparativa por fuentes de televisión de la empresa de interés, empresas del sector o competencia y sector donde se encuentra su nicho de mercado.
- Exposición mediática comparativa por fuentes online de la empresa de interés, empresas del sector o competencia y sector donde se encuentra su nicho de mercado.

- Exposición mediática comparativa por Cubrimiento de la empresa de interés, empresas del sector o competencia y sector donde se encuentra su nicho de mercado.
- Exposición mediática comparativa por Ciudad de la empresa de interés, empresas del sector o competencia y sector donde se encuentra su nicho de mercado.
- Exposición mediática comparativa por Peso (Positivo, negativo, neutro) de la empresa de interés, empresas del sector o competencia y sector donde se encuentra su nicho de mercado.

Nota: Cada una de las gráficas muestra el total de noticias, valor free-press y audiencia.

Capítulo 5. Conclusiones y recomendaciones

Terminada la investigación y en base a su alcance, objetivos y resultados obtenidos se listan a continuación las conclusiones y recomendaciones:

- a) Se desarrolló un método con líneas guías específicas para el manejo completo de las fuentes de datos “inputs” de medios de comunicación en términos de extracción, limpieza, transformación y enriquecimiento. El pre y procesamiento garantiza su idoneidad para aplicar técnicas de Minería de Texto y búsqueda inteligente/relacionada para generar información consumible con alto valor agregado soportada en herramientas de visualización.
- b) El paso a paso propuesto permitió estandarizar los datos en datos estructurados y gobernables. Con ello se logró corregir fallas desde la fuente: datos incompletos, incorrectos, inexactos, o no pertinentes generados por la plataforma, garantizando así un monitoreo de control estable que permite corregir o suprimir los datos según sea el caso.
- c) Los datos preprocesados y enriquecidos con los diccionarios por niveles, permitieron aplicar técnicas de analítica avanzada, sustentadas básicamente en Minería de datos (Minería de texto) cuyos outputs son amigables para el usuario y extrapolables a plataformas de visualización que contribuyen a soportar las decisiones mediante dashboards interactivos y resumibles gerenciales.
- d) Los puntos anteriores se pusieron en práctica en un caso de uso empresarial (Empresa Siglo Data, Bogotá). Esta fase puede etiquetarse como proyecto de inspección, producción y despliegue (“deployment” como suele llamarse en el argot de la Analítica y TI). En Siglo Data, se hicieron pruebas y validaciones con el área comercial y clientes. La retroalimentación recibida tuvo buenos niveles de aceptación. La visualización y entrega final de los hallazgos detectados al generar los diferentes clústeres (grupos asociados a los diferentes niveles obtenidos con los diccionarios de las etiquetas y tags) en la clasificación de las categorías, ratificó un mayor enriquecimiento y valor agregado en la información que se generaba previamente. Actualmente, se continúa desarrollando una herramienta computacional completa (bajo patrocinio con SAS y sus motores de Big Data, Analítica y Visualización) que sigue los principios metodológicos descritos en este proyecto de investigación.

Bajo las premisas anteriores, se pueden resumir los siguientes aportes y posibles líneas de investigación futuras:

5.1 Aportes

- Obtención de un método base para gestionar la data y obtener conocimiento de ella.
- Estandarización de diccionarios de atributos alcanzando tres niveles de profundización adaptado y enriquecido para medios de comunicación. La granularidad a tres niveles mejoró los aportes en detalle y descripción de las noticias analizadas.
- Mejoramiento en el soporte a la toma de decisiones en cuanto a la información ofrecida por Siglo Data a sus clientes.

5.2 Trabajos o líneas futuras sugeridas o recomendadas

- a) Ampliar en la herramienta de minería de texto las transcripciones de los contenidos de prensa radio y televisión. A diferencia de los titulares que hoy día se tienen, el contenido total de la noticia generara una mayor riqueza en los hallazgos.
- b) Generar KpIs teniendo como base el resultado de los informes actuales para motivar a los usuarios la realización de planes de acción consolidando así la herramienta como parte activa de gestión.
- c) Generación automática o sistemática de listados: Rankings en los diccionarios de atributos por categorías, por citar un ejemplo.
- d) Envíos automatizados de la información.
- e) Generación de alertas vía correo o WhatsApp en la generación de listados del punto anterior.
- f) Alertas por correo o WhatsApp inteligentes: generar alertas cuando en una serie de tiempo se den por cumplido, en una misma noticia o noticias, ciertos atributos predeterminados.
- g) Redes sociales: Integrar en este desarrollo las conversaciones en redes sociales mediante las herramientas de minería de texto que permitan categorizar la información como inputs adicionales en las fuentes estudiadas hasta ahora.

6. Referencias

Acceso <http://monitoreodemedios.acceso.com/>

Analítica de texto. Recuperado de <https://www.meaningcloud.com/es/soluciones/analitica-de-texto>

Ariza, F. (2019). *¿Qué es la Analítica Avanzada y cuáles son sus beneficios?*. Recuperado del 21 February 2019, de <https://www.itop.es/blog/item/que-es-la-analitica-avanzada-y-cuales-son-sus-beneficios.html>.

Baquero, S; Camargo, Z; García, F; González, A; Piracón, D; Galvis, J. (2017) *Direccionamiento estratégico*. Universidad de la Sabana

Cision <http://www.cision.com/us/>

De Rossi, A. (2012). *El diccionario de datos*. Recuperado del 6 August 2012, de <http://dblearner.com/el-diccionario-de-datos/>

El Valor del Monitoreo de Noticias en las Estrategias de Comunicación y Relaciones Públicas. (s.f). Recuperado de https://www.infosol.com.mx/espacio/Articulos/Desde_la_Trinchera/el_valor_del_monitoreo_de_noticias_en_las_estrategias_de_comunicacion_y_relaciones_publicas.html#.Xwd-fC1t_jB

Fernández, R. (2014). *Los medios de comunicación: ¿cómo influyen en la sociedad?*. Recuperado el 2 de Junio 2014, de <https://blog.uclm.es/ricardofdez/2014/06/02/los-medios-de-comunicacion-como-influyen-en-la-sociedad/>

Garzón, G. *Watson Analytics. una solución cloud para procesos de big data*. Gerson Garzón. Recuperado de <https://repository.unilibre.edu.co/bitstream/handle/10901/11219/Watson%20Analytics%20una%20Solución%20Cloud%20para%20procesos%20de%20BigData.pdf?sequence=1&isAllowed=y>

IBM Knowledge Center, *Configurar el análisis de texto*. Recuperado de https://www.ibm.com/support/knowledgecenter/es/SS8NLW_12.0.0/com.ibm.discovery.es.ta.doc/iypofnv_textanal_cont.html

Iniciativa aporta; red.es (2016). Visualización de datos, definición, tecnologías y herramientas. Recuperado de https://datos.gob.es/sites/default/files/doc/file/informe_herramientas_visualizacion.pdf

Kantar Media <http://www.kantarmedia.com/es>

Leonardi, B. (2018). Influencia de los medios de comunicación de masas. Recuperado de <https://www.monografias.com/trabajos/influmcm/influmcm.shtml>

Lescano, D., & Ávila, I. (2019). *Analítica de datos para el soporte en la toma de decisiones en el área de distribución y ventas de la distribuidora farmacéutica Trujillo*: Universidad Privada Antenor Orrego. Recuperado de http://repositorio.upao.edu.pe/bitstream/upaorep/5900/1/RE_ING.SIST_DANIEL.ACOSTA_SOLEDAD.VASQUEZ_SOPORTE.EN.LA.TOMA.DE.DECISIONES_DATOS.PDF

Mendes, G., Augusto, A., Michelotti, E., & Souza, C. (2015). *Modelo de monitoreo de redes sociales para orientar a tomada de decisiones de organizaciones de gestión de destinos* (5 ed.). Revista Brasileña de investigación turística. Recuperado de <https://www.rbtur.org/rbtur/article/view/835>

Ninestar <http://ninestars.in/index.php/media-monitoring/>

Riquelme, J., Ruiz, R., & Gilbert, K. *Minería de Datos: Conceptos y Tendencias* [Ebook]. Recuperado de <https://idus.us.es/bitstream/handle/11441/43290/Miner%C3%ADa%20de%20datos.pdf?sequence=1&isAllowed=y>

Sena, M. (2019). Estructura de datos. Recuperado el 31 de enero de 2019 de <https://medium.com/techwomenc/estructuras-de-datos-a29062de5483#:~:text=Una%20estructura%20de%20datos%20describe,tipo%20de%20problema%20que%20resuelve.>

Tecon: soluciones informáticas *¿Qué es un KPI? Cómo definir los KPIs básicos para tu negocio*. Recuperado de <https://www.tecon.es/que-es-un-kpi-como-definir-los-kpis-basicos-para-tu-negocio/>

Toro, B. (2015). Medios Masivos de Comunicación: Una construcción de la realidad. Revista Pequén, (1). Retrieved from <http://revistas.ubiobio.cl/index.php/RP/article/view/1824>

5 beneficios de utilizar las nubes de palabras en el aula. (2015). Recuerado el 22 April 2015, de <https://noticias.universia.edu.pe/consejos-profesionales/noticia/2015/04/22/1123727/descubre-5-beneficios-utilizar-nubes-palabras-aula.html>

7. Anexos

7.1 Anexo 1. Listado de etiquetas sector primario

- | | | |
|----------------------------------|-----------------------------------|------------------------------|
| 1) Accesorio automóvil. | 22) Conflicto armado. | 44) Medio ambiente. |
| 2) Accesorios hogar. | 23) Construcción civil | 45) Medios de comunicación. |
| 3) Agroindustria. | 24) Construcción infraestructura. | 46) Minería minas. |
| 4) Alimentos. | 25) Construcción inmobiliario. | 47) Ocio y entretenimiento. |
| 5) Arte y cultura. | 26) Deportes. | 48) Opinión. |
| 6) Artículos oficina. | 27) Dulces y snacks. | 49) Política. |
| 7) Aseo hogar. | 28) Economía. | 50) Religión. |
| 8) Aseo personal | 29) Educación. | 51) Restaurantes y bares. |
| 9) Astronomía geología. | 30) Elecciones y votaciones. | 52) Salud. |
| 10) Banca y finanzas. | 31) Embajadas. | 53) Servicios profesionales. |
| 11) Bebidas alcohólicas. | 32) Energía y combustible. | 54) Servicios públicos. |
| 12) Bebidas no alcohólicas. | 33) Fauna Mascotas. | 55) Tabaco. |
| 13) Belleza. | 34) Fronteras migraciones. | 56) Tecnología. |
| 14) Borrarr. | 35) Fundaciones-Ongs. | 57) Telecomunicaciones. |
| 15) Cajas de compensación. | 36) Geografía. | 58) Trabajo. |
| 16) Ciencia. | 37) Gobierno estado. | 59) Transporte aéreo. |
| 17) Colombia. | 38) Industria. | 60) Transporte férreo |
| 18) Comercio. | 39) Infraestructura hídrica. | 61) Transporte terrestre. |
| 19) Comercio exterior. | 40) Infraestructura vial. | 62) Turismo. |
| 20) Comunidad. | 41) Joyas relojes perfumes. | |
| 21) Confección vestuario textil. | 42) Justicia y seguridad. | |
| | 43) Literatura- textos. | |

7.2 Anexos secundarios

- | | | |
|---------------------|----------------------|----------------------------|
| 1) Accesorios hogar | 4) Arte y cultura | 8) Astronomía geología |
| 2) Agroindustria | 5) Artículos oficina | 9) Banca y finanzas |
| 3) Alimentos | 6) Aseo hogar | 10) Bebidas alcohólicas |
| | 7) Aseo personal | 11) Bebidas no alcohólicas |
| | | 12) Belleza |

- 13) Borrarr**
14) Cajas de compensación
15) Ciencia
16) Colombia
17) Comercio
18) Comercio exterior
19) Comunidad
20) Confección vestuario textil
21) Conflicto armado
22) Construcción civil
23) Construcción infraestructura
24) Construcción inmobiliario
25) Deportes
26) Dulces y snacks
27) Economía
28) Educación
29) Elecciones y votaciones
30) Embajadas
31) Energía y combustible
32) Fauna Mascotas
33) Fronteras migraciones
34) Fundaciones-Ongs
35) Geografía
36) Gobierno estado
37) Industria
38) Infraestructura hídrica
39) Infraestructura vial
40) Joyas relojes perfumes
41) Justicia y seguridad
42) Literatura- textos
43) Medio ambiente
44) Medios de comunicación
45) Minería minas
46) Ocio y entretenimiento
47) Opinión
48) Política
49) Religión
50) Restaurantes y bares
51) Salud
52) Servicios Profes
53) Servicio Público
54) Tabaco
55) Tecnología
56) Telecomunicaciones
57) Trabajo
58) Transporte aéreo
59) Transporte férreo
60) Transporte terrestre
61) Turismo
62) Abogados
63) Abonos
64) Accesorios
65) Accesorios ferretería
66) Accidentes
67) Aceite mesa y cocina
68) Acero aluminio
69) Acueducto alcantarillado
70) Acuicultura
71) Administradora riesgos profesional.
72) Aduanas aranceles
73) Adulto mayor
74) Aeroespacial
75) Aerolínea
76) Aeropuertos
77) África
78) Agencia de viajes
79) Agremiación
80) Agremiación
81) Agro
82) Aguardiente ron
83) Aguas
84) Aguas mares
85) Ahorro y crédito
86) Ajedrez
87) Alcaldías
88) Alfombras
89) Algodón
90) Alimentación
91) Alimentación animal
92) Almacenamiento
93) América
94) América del norte
95) América del sur
96) Analítica
97) Antioquia
98) Apicultura
99) App
100) Armamento
101) Arquería
102) Arquitectura
103) Arroz
104) Arte y cultura
105) Asambleas departamentales
106) Ascensores elevadores
107) Aseo basura
108) Asfaltos
109) Asia
110) Asistencia
111) Asistencia
112) Astronomía
113) Atletismo
114) Audio y video
115) Automóvil
116) Automóvil renta
117) Automovilismo
118) Aves
119) Avícola
120) Ayudas rescates
121) Azúcar
122) Bacteriología
123) Baloncesto
124) Banano
125) Banca multilateral
126) Banca y finanzas
127) Banco central
128) Banda ancha
129) Bares
130) Barranquilla
131) Baterías
132) Bebidas alcohólicas
133) Béisbol
134) Belleza
135) Belleza
136) Bicicletas
137) Bienestar
138) Bilingüismo
139) Biocombustible
140) Biodiversidad
141) Bioinformática

- 142) Biomédica**
143) Biotecnología
144) Bogotá
145) Bolígrafos
146) Bomberos
147) Boxeo
148) Bpo
149) Buscador
150) Buses
151) Cableado
152) Cacao
153) Café
154) Cajas de compensación
155) Cajas fuertes
156) Cajeros automáticos
157) Calcáneos
158) Caldos sopas
159) Calzado accesorios marroquinería
160) Cámaras de comercio
161) Camaras fotograficas
162) Cambios divisas
163) Camiones
164) Canales
165) Candidatos cargos popular
166) Capilar
167) Carbón
168) Cárceles prisiones
169) Cardiología
170) Carga
171) Carnes cárnicos
172) Carrocerías
173) Cartagena
174) Cartón papel
175) Católica
176) Caucho plásticos
177) Cemento concreto
178) Centrales de riesgo
179) Centroamérica
180) Centro convenciones - comercial
181) Cerámica
182) Cereales
183) Cerraduras
184) Cerveza
185) Champiñones
186) Chatarra
187) Chiclets
188) Chocolate de mesa
189) Chocolates
190) Ciclismo
191) Ciencia
192) Cigarrillos
193) Cine
194) Cirugía
195) Cítricos
196) Clima
197) Clinicas hospitales
198) Clubes
199) Clubes
200) Cobranzas
201) Cobre
202) Colchones
203) Colegios
204) Colombia
205) Comercio
206) Comercio
207) Comercio exterior
208) Competitividad
209) Complemento alimenticio
210) Computadores
211) Comunidad
212) Consejos
213) Concesionario
214) Concesionario
215) Condimentos especias
216) Congelados
217) Congregaciones y marchas
218) Conservas
219) Construcción inmobiliario
220) Constructora
221) Consultoría
222) Contact center
223) Contaduría
224) Contaminación ambiental
225) Contrabando
226) Cooperativa
227) Copiadora
228) Corporaciones autónomas
229) Corredores de bolsa
230) Corrupción
231) Cosméticos
232) Courier
233) Courier mensajería
234) Cremas personal
235) Cruceros
236) Cultura
237) Curtiembre
238) Danza
239) Datacenter
240) Deforestación
241) Delitos
242) Departamentos
243) Deportes
244) Derechos humanos
245) Dermatología
246) Desastres
247) Desastres naturales
248) Desodorante
249) Desplazamiento
250) Detergente
251) Devaluación-Inflación
252) Diagnóstica
253) Diálisis
254) Diamantes
255) Discapacidad
256) Diseño
257) Dispositivos
258) Dispositivos antipersona
259) Distribución geográfica
260) Divisas
261) Documento identidad
262) Donaciones Auxilios
263) Donuts
264) Drogadicción
265) Drones
266) Dulces
267) Ecología
268) Economía
269) Edificio
270) Editorial
271) Educación
272) Ejercicio
273) Elecciones
274) Elecciones

- | | | |
|--|-----------------------------|-------------------------------|
| 275) Electrodomésticos | 320) Fundaciones | 365) Ingeniería civil |
| 276) Electrónica | 321) Funerarias | 366) Ingenios azucareros |
| 277) Empaques envases | 322) Fútbol | 367) Inmobiliaria |
| 278) Emprendimiento | 323) Gafas deportivas | 368) Insectos |
| 279) Empresas | 324) Galletas | 369) Insectos |
| 280) Energía eléctrica hidroelectricidad | 325) Ganadería | 370) Institutos |
| 281) Energía eléctrica | 326) Gas | 371) Instrumentación |
| 282) Energía eólica | 327) Gaseosas | 372) Insumos |
| 283) Energía nuclear | 328) Gastronomía | 373) Insumos |
| 284) Energía solar | 329) Gatos | 374) Internacional |
| 285) Energía térmica | 330) Gelatina | 375) Internet |
| 286) Energizantes | 331) Género | 376) Inversiones |
| 287) Eps-Ips | 332) Geofísica | 377) Investigación |
| 288) Equinos | 333) Geografía | 378) Investigación |
| 289) Esmeraldas | 334) Gimnasio | 379) Isla |
| 290) Espacio público | 335) Girardot | 380) Jabones corporales |
| 291) Estadística | 336) Giros pagos | 381) Joyas |
| 292) Estética | 337) Gobernaciones | 382) Judaísmo |
| 293) Europa | 338) Golf | 383) Juegos |
| 294) Evaluación | 339) Gráfica impresión | 384) Juegos de azar |
| 295) Eventos | 340) Grupo empresarial | 385) Jugos |
| 296) Explosivos mineros | 341) Grupos armados | 386) Juguetería |
| 297) Extranjeros | 342) Hamburguesa | 387) Justicia y seguridad |
| 298) Fabricante | 343) Hardware | 388) Laboratorio clínico |
| 299) Factura electrónica firma | 344) Harinas-Pan - cereales | 389) Laboratorio farmaceutico |
| 300) Falsificación | 345) Helados | 390) Lácteos |
| 301) Fauna | 346) Helados | 391) Ladrillos |
| 302) Ferróníquel | 347) Herbicidas plaguicidas | 392) Laminados |
| 303) Fertilización Fertilizantes | 348) Hidráulica | 393) Leasing |
| 304) Fiduciaria | 349) Hidrocarburos | 394) Leches derivados |
| 305) Filosofía | 350) Hidropónicos | 395) Legislación |
| 306) Filtros | 351) Higiene femenina | 396) Lencería |
| 307) Financiación | 352) Historia | 397) Licitaciones |
| 308) Fiscalía General | 353) Horticultura | 398) Literatura |
| 309) Floricultura | 354) Hotelería | 399) Llantas |
| 310) Fondo de empleados | 355) Huevos | 400) Lobby |
| 311) Fondo de pensiones | 356) Idiomas | 401) Logística |
| 312) Fósforos | 357) Iluminación | 402) Lubricantes |
| 313) Fotografía | 358) Impresos | 403) Lucha libre |
| 314) Franquicias | 359) Impuestos | 404) Luminarias |
| 315) Frigorífico | 360) India | 405) Madera |
| 316) Fronteras exterior | 361) Indígenas | 406) Mallas -alambres |
| 317) Fruticultura | 362) Industria | 407) Malta |
| 318) Fuerzas militares | 363) Infantil | 408) Maquinaria |
| 319) Fumigación | 364) Infecciones | 409) Marihuana medicinal |

- 410) Marroquinera
411) Mascotas
412) Mecánica
413) Medellín
414) Medicamento-Ansiolítico
415) Medicamento-Antibiótico
416) Medicamentos
417) Medicina interna
418) Medicina prepagada
419) Medio ambiente
420) Medio oriente
421) Medios
422) Medios de comunicación
423) Mercurio
424) Metal
425) Metalmecánica
426) México
427) Miel
428) Migraciones
429) Minería minas
430) Ministro
431) Moda
432) modelaje
433) monoteísmo
434) Montañismo
435) Morrales maletas
436) Motos
437) Mudanzas
438) Muebles
439) Multinivel
440) Municipios
441) Museo bibliotecas
442) Música
443) Mutilación
444) Narcotráfico
445) Natación
446) Naviera
447) Negritudes
448) Niñez
449) Normas y leyes
450) Notarías
451) Nutrición
452) Obesidad
453) Oceanía
454) Ocio y entretenimiento
455) Ocupacional
456) Odontología
457) Oftalmología
458) Oleoductos
459) Operador
460) Opinión
461) Orinoquia
462) Ornamentación
463) Ornitología
464) Oro
465) Ortopedia
466) Osteología
467) Otorrino
468) País
469) Palma de aceite
470) Panela
471) Pañales
472) Papel higiénico
473) Papelería carton
474) Parqueadero
475) Parques diversiones
476) Parques naturales
477) Partido político
478) Patinaje
479) Patinetas electricas
480) Patologías
481) Patologías-Aborto
482) Patologías-Alergias
483) Patologías-Alopecia
484) Patologías-Alzheimer
485) Patologías-Anestesiología
486) Patologías-ansiedad
487) Patologías-Artritis
488) Patologías-Audición
489) Patologías-cancer
490) Patologías-Cardiovascular
491) Patologías-Celular
492) Patologías-Dermatología
493) Patologías-Diabetes
494) Patologías-Dolor
495) Patologías-Drogadiccion
496) Patologías-Enfermedades infecto contagiosas
497) Patologías-Enfermedades r
498) Patologías-Epidemias
499) Patologías-esclerosis
500) Patologías-Esquizofrenia
501) Patologías-Eutanasia
502) Patologías-Fibrosis
503) Patologías-Gastroenterologia
504) Patologías-Genetica
505) Patologías-Ginecologia
506) Patologías-Hematologia
507) Patologías-Hormonas
508) Patología-Sida
509) Patologías-Infecciones
510) Patologías-Metabolismo
511) Patologías-Muscular
512) Patologías-Nefrologia
513) Patologías-Neumologia
514) Patologías-Neurologia
515) Patologías-Opiáceos
516) Patologías-Osteoporosis
517) Patologías-Parkinson
518) Patologías-Sindrome de down
519) Patologías-Tuberculosis
520) Patología-Suicidio
521) Patologías-Intoxicaciones
522) Patologías-Leucemia
523) Peajes
524) Peces
525) Pediatría
526) Perfumes
527) Pegantes
528) Películas revelado
529) Pensiones y cesantías
530) Perfumes
531) Periodismo
532) Perros
533) Persianas
534) Pesas
535) Pescado
536) Petroquímica
537) Piedras
538) Pinturas
539) Pizzeria
540) Planetas estrellas
541) Plásticos
542) Plátano
543) Población

- 544) Pobreza**
545) Poliuretano
546) Pollerías
547) Pólvora
548) Porcicultura
549) Presidencia
550) Prevención y desarrollo
551) Procesadores
552) Proceso de paz
553) Producción radio y tv
554) Programas
555) Proyectoros
556) Proyectos
557) Psicología
558) Psiquiatría
559) Publicidad y mercadeo
560) Puertos
561) Purificador aire
562) Química
563) Radio
564) Reciclaje
565) Red social
566) Reforestación
567) Reforma
568) Refrigeración
569) Refugiados
570) Regalías
571) Regulación
572) Rehabilitación
**573) Reintegración –
Reincorporación**
574) Relaciones exteriores
575) Religión
576) Relojes
577) Remesas divisas
578) Renal
579) Renta
580) Represas
581) Reservas
582) Residuos
583) Respiratorio
584) Restaurantes
585) Retail
586) Retail
587) Reumatismo
588) Riego
589) Riesgos
590) Rios
591) Robótica
592) Sal
593) Salsas aderezos
594) Salud
595) San jose de guaviare
596) Saneamiento básico
597) Satélite
598) Secretarías de educación
599) Secretarías de hacienda
600) Secretarías de salud
601) Secuestro
602) Seguridad
603) Seguridad
604) Seguridad industrial
605) Seguridad social
606) Seguros
607) Semillas
608) Servicios
609) Servicios temporales
610) Sexualidad
611) Siliconas
612) Sindicatos
613) Sistema metro
614) Sistema nervioso
615) Snacks
616) Sociedad
617) Software
618) Soldadura
619) Soluciones
620) Sonido
621) Sopas
622) Sopas
623) Subastas
624) Suministros
625) Suministros
626) Suramérica
627) Tabaco
628) Talcos
629) Talento humano
630) Tarjeta de crédito
631) Taurino
632) Taxi
633) Te
634) Teatro
635) Tecnología
636) Tejas
637) Tejas
638) Tejo
639) Telecomunicaciones
640) Telefonía
641) Telefonía móvil
642) Telégrafo
643) Televisión
644) Tema
645) Tenis
646) Tequila
647) Terapéutico
648) Territorial
649) Terrorismo
650) Textil
651) Tiendas
652) Tierras
653) Tiquetes cupones
654) Tiro con arco
655) Toros
656) Trabajo
657) Transferencias de capitales
658) Transgénicos
659) Trasplantes
660) Transporte
661) Transporte terrestre
662) Transporte aéreo
663) Transporte férreo
664) Triatlón
665) Tubérculos
666) Tubería tubos
667) Túneles puentes
668) Turismo
669) Tutelas
670) Ungüentos
671) Universidades
672) Urbanismo
673) Urología
674) Utensilios
675) Vacunación
676) Vales -tiquetes
677) Vasectomía

678) Venta al aire libre
679) Vestidos de baño
680) Vestuario
681) Vestuario infantil
682) Veterinaria
683) Vías
684) Video

685) Vidrio
686) Vigilancia
687) Vinos
688) Violencia
689) Violencia
690) Virus bacteria
691) Vivienda

692) Vivienda
693) Vocero
694) Vodka
695) Whisky
696) Zona franca