

# REVISTA

Ciencias de la Documentación



Volumen 8 - Número 1  
enero/junio 2022

ISSN 0719-5753

Editorial  
Cuadernos de Sofia

**CUERPO DIRECTIVO**

**Director**

**Eugenio Bustos Ruz**

*Editorial Cuadernos de Sofía, Chile*

**Editora**

**Dra. Antonia Isabel Nogales-Bocio**

*Universidad de Zaragoza, España*

**Editor Científico**

**Dr. Martino Contu**

*Universidad de Sassari, Italia*

*Fondazione Mons. Giovannino Pinna, Italia*

**Cuerpo Asistente**

**Traductora: Inglés**

**Pauline Corthorn Escudero**

*Editorial Cuadernos de Sofía, Chile*

**Portada**

**Graciela Pantigozo De los Santos**

*Editorial Cuadernos de Sofía, Chile*

Revista Ciencias de la Documentación  
Editorial Cuadernos de Sofía

**COMITÉ EDITORIAL**

**Dra. Kátia Bethânia Melo de Souza**

Universidade de Brasília – UNB, Brasil

**Dr. Carlos Blaya Perez**

Universidade Federal de Santa María, Brasil

**Lic. Oscar Christian Escamilla Porras**

Universidad Nacional Autónoma de México,  
México

**Ph. D. France Bouthillier**

MgGill University, Canadá

**Dr. Miguel Delgado Álvarez**

Instituto Griselda Álvarez A. C., México

**Dr. Juan Escobedo Romero**

Universidad Autónoma de San Luis de Potosí,  
México

**Dr. Jorge Espino Sánchez**

Escuela Nacional de Archiveros, Perú

**Dr. José Manuel González Freire**

Universidad de Colima, México

**Dra. Patricia Hernández Salazar**

Universidad Nacional Autónoma de México, México

**Dra. Trudy Huskamp Peterson**

Certified Archivist Washington D. C., Estados  
Unidos

**Dr. Luis Fernando Jaén García**

Universidad de Costa Rica, Costa Rica

**Dra. Elmira Luzia Melo Soares Simeão**

Universidade de Brasília, Brasil

**Lic. Beatriz Montoya Valenzuela**

Pontificia Universidad Católica del Perú, Perú

**Mg. Liliana Patiño**

Archiveros Red Social, Argentina

**Dr. André Porto Ancona Lopez**

Universidade de Brasília, Brasil

**Dra. Glaucia Vieira Ramos Konrad**

Universidad Federal de Santa María, Brasil

**Dra. Perla Olivia Rodríguez Reséndiz**  
Universidad Nacional Autónoma de México, México

**COMITÉ CIENTÍFICO INTERNACIONAL**

**Dr. Héctor Guillermo Alfaro López**  
Universidad Nacional Autónoma de México, México

**Ph. D. Juan R. Coca**  
Universidad de Valladolid, España

**Dr. Martino Contu**  
Università Degli Studi di Sassari, Italia

**Dr. José Ramón Cruz Mundet**  
Universidad Carlos III, España

**Dr. Carlos Tulio Da Silva Medeiros**  
Instituto Federal Sul-rio-grandense, Brasil

**Dr. Andrés Di Masso Tarditti**  
Universidad de Barcelona, España

**Dra. Luciana Duranti**  
University of British Columbia, Canadá

**Dr. Allen Foster**  
University of Aberystwyth, Reino Unido

**Dra. Manuela Garau**  
Universidad de Cagliari, Italia

**Dra. Marcia H. T. de Figueredo Lima**  
Universidad Federal Fluminense, Brasil

**Dra. Rosana López Carreño**  
Universidad de Murcia, España

**Dr. José López Yepes**  
Universidad Complutense de Madrid, España

**Dr. Miguel Angel Márdero Arellano**  
Instituto Brasileiro de Informação em Ciência e  
Tecnologia, Brasil

**Lic. María Auxiliadora Martín Gallardo**  
Fundación Cs. de la Documentación, España

**Dra. María del Carmen Mastropiero**  
Archivos Privados Organizados, Argentina

**Dr. Andrea Mutolo**  
Universidad Autónoma de la Ciudad de  
México, México

**Mg. Luis Oporto Ordoñez**  
Director Biblioteca Nacional y Archivo  
Histórico de la Asamblea Legislativa  
Plurinacional de Bolivia, Bolivia  
Universidad San Andrés, Bolivia

**Dr. Alejandro Parada**  
Universidad de Buenos Aires, Argentina

**Dra. Gloria Ponjuán Dante**  
Universidad de La Habana, Cuba

**Dra. Luz Marina Quiroga**  
University of Hawaii, Estados Unidos

**Dr. Miguel Ángel Rendón Rojas**  
Universidad Nacional Autónoma de México, México

**Dr. Gino Ríos Patio**  
Universidad San Martín de Porres, Perú

**Dra. Fernanda Ribeiro**  
Universidade do Porto, Portugal

**Dr. Carlos Manuel Rodríguez Arrechavaleta**  
Universidad Iberoamericana Ciudad de México, México

**Mg. Arnaldo Rodríguez Espinoza**  
Universidad Estatal a Distancia, Costa Rica

**Dra. Vivian Romeu**  
Universidad Iberoamericana Ciudad de México, México

**Mg. Julio Santillán Aldana**  
Universidade de Brasília, Brasil

**Dra. Anna Szlejcher**  
Universidad Nacional de Córdoba, Argentina

**Dra. Ludmila Tikhnova**  
Russian State Library, Federación Rusa

**Indización, Base de datos y Repositorios Institucionales**

Revista Ciencias de la Documentación, se encuentra en:



GRUPOS DE INVESTIGACIÓN



ResearchGate





**EL PROCESAMIENTO DEL LENGUAJE NATURAL COMO HERRAMIENTA  
EN LA ADMINISTRACIÓN DE RIESGOS DE LAS ENTIDADES PÚBLICAS:  
MODELO APLICADO AL CASO COLOMBIANO**

**NATURAL LANGUAGE PROCESSING AS A TOOL IN RISK MANAGEMENT  
FOR PUBLIC ENTITIES**

**Dr. Ennio Prada Madrid**

Pontificia Universidad Javeriana, Colombia  
ORCID: <https://orcid.org/0000-0002-4998-1837>  
ennioprada@gmail.com

**Dra. Gigiola Prada Madrid**

Pontificia Universidad Javeriana, Colombia  
ORCID: <https://orcid.org/0000-0001-9775-1713>  
gigioprada@gmail.com

**Dra. Yanira Patricia Vargas Cubillos**

Pontificia Universidad Javeriana, Colombia  
ORCID: <https://orcid.org/0000-0002-2019-6114>  
pvc0774@gmail.com

**Dr. Miguelangel Clavijo Perdomo**

Universidad Universidad de la Salle, Colombia  
ORCID: <https://orcid.org/0000-0003-1012-1298>  
miguelangel.clavijo@hotmail.com

**Fecha de Recepción:** 07 de abril de 2021 – **Fecha Revisión:** 10 de junio de 2021

**Fecha de Aceptación:** 02 de noviembre de 2021 – **Fecha de Publicación:** 01 de enero de 2022

**Resumen**

Este artículo pretende demostrar de manera práctica que el Procesamiento de Lenguaje Natural - PLN-, es una herramienta útil en el marco de la gestión de riesgos de las entidades públicas, para tal efecto se presenta un ejemplo aplicado al análisis del mapa de riesgos, este análisis determina que los campos más importantes (nucleares) de este son: riesgos, causas y efectos, se argumenta que con el PLN se puede determinar métricamente la relación semántica entre estos campos nucleares. Se diseñó un Pipeline que determina los pasos del PNL que permitan determinar métricas semánticas, se modeló dicho Pipeline en las herramientas del software Orange 3.26 y se aplicó el corpus de datos que permitió establecer la relación semántica entre los campos nucleares. Finalmente se presentan conclusiones sobre los resultados obtenidos.

**Palabras Claves**

Procesamiento del Lenguaje Natural – Gestión de Riesgos – Análisis Semántico

**Abstract**

This article aims to demonstrate in a practical way that Natural Language Processing -NLP- is a useful tool in the framework of risk management of public entities, for this purpose an example applied to the analysis of the risk map is presented, this analysis determines that the most important (nuclear) fields of this are: risks, causes and effects, it is argued that with the NLP the semantic relationship

between these nuclear fields can be determined metrically. A Pipeline was designed that determines the steps of the NLP that allow determining semantic metrics, said Pipeline was modeled in the Orange 3.26 software tools and the data corpus that allowed establishing the semantic relationship between the nuclear fields was applied. Finally, conclusions on the results obtained are presented.

### Keywords

Natural Language Processing – Risk management – Semantic Analysis

### Para Citar este Artículo:

Prada Madrid, Ennio; Prada Madrid, Gigiola; Vargas Cubillos, Yanira Patricia y Clavijo Perdomo, Miguelangel. El procesamiento del lenguaje natural como herramienta en la administración de riesgos de las entidades públicas: Modelo aplicado al caso Colombiano. Revista Ciencias de la Documentación Vol: 8 num 1 (2022): 01-19.

Licencia Creative Commons  
Licencia Creative Commons Attribution-NonCommercial 3.0 Unported  
(CC BY-NC 3.0  
Licencia Internacional





## Introducción

Las entidades públicas deben desarrollar la administración de riesgos en el marco normativo vigente, el mapa de riesgos es una herramienta definida por la administración de riesgos, esta herramienta, es una representación final de la probabilidad e impacto de uno o más riesgos<sup>1</sup>.

El mapa de riesgos puede adoptar la forma de una tabla resumen con datos relativos al riesgo, sin embargo, se identifican datos nucleares en el mapa de riesgos los cuales son: el riesgo, las causas y las consecuencias, estos campos nucleares definen los aspectos de control del riesgo, acciones preventivas, medición del riesgo, medición de aplicación de controles, etc. Teniendo en cuenta lo descrito, se plantea la necesidad de que los campos nucleares (Riesgo, Causas y Consecuencias) tengan una relación semántica entre sí, pues se postula que en la medida en que dicha relación sea coherente, el mapa de riesgos estará mejor estructurado con relación a todos los datos que lo conforman, los cuales dependen de los campos nucleares.

Al lograr tener un mapa de riesgos en el cual exista una relación semántica entre los conceptos de los campos nucleares, las entidades públicas tendrán una herramienta más efectiva, que le aportará valor a la administración de riesgos y por ende al mejoramiento de la entidad estatal.

Por todo lo anterior se considera viable el desarrollo de un método para calcular una métrica de las relaciones semánticas entre los campos nucleares del mapa de riesgo. A continuación, se comparte la construcción del modelo de PNL que permite identificar la métrica en la relación semántica entre los campos nucleares del mapa de riesgo.

## La administración de riesgos y las entidades públicas

La relación entre riesgo y organizaciones es casi íntima, pues la organización nunca estará libre de su manto, el riesgo siempre estará allí, latente, cercano y expectante de poder materializarse, por su parte la organización debe conocerlo, identificarlo, comprenderlo y controlarlo, es casi un matrimonio eterno.

En el contexto colombiano, la relación entre riesgo y entidades públicas se presenta en el marco de la gestión pública como una actividad dinámica, que continuamente está redefiniéndose y en una continua revisión sistemática, en este marco, la administración de los riesgos juega un papel fundamental incluso llegando a definirse como una política administrativa. La administración de los riesgos en la gestión pública toma mayor relevancia al concebirse como un medio que posibilita un conocimiento más profundo que apunta al mejoramiento de las entidades públicas, pues aporta significativamente en aspectos de productividad y perfeccionamiento de los procesos de negocio.

Se podría afirmar que el cumplimiento de los objetivos retadores de las entidades, así como la razón de ser del estado, toma fortaleza con la administración de los riesgos, pues mediante esta, se logra establecer mecanismos para identificarlos, valorarlos, controlarlos y minimizarlos.

---

<sup>1</sup> Administración de Riesgos Corporativos. Técnicas de Aplicación (Colombia: Pricewaterhouse Coopers, 2005).

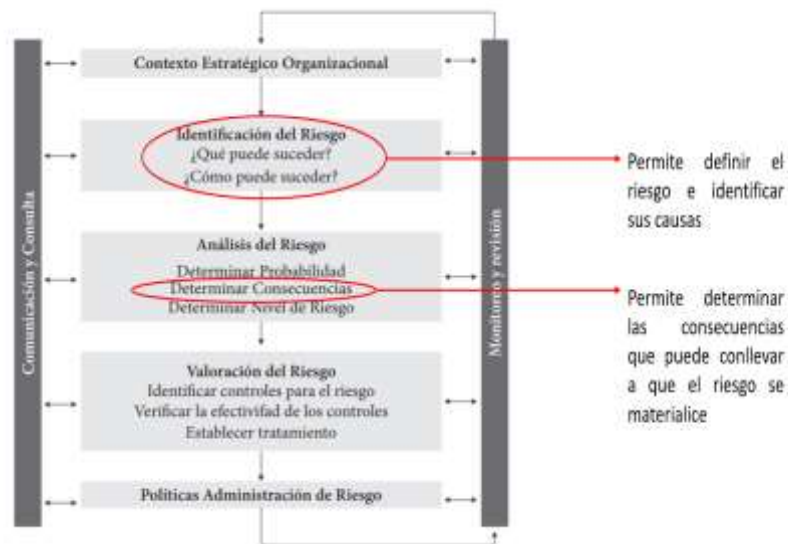


Para lograr una eficiente administración de riesgos, las entidades públicas definen una política de riesgos mediante la cual se propende a identificarlos, controlarlos y medirlos, pero el riesgo siempre existirá, incluso en términos de una referencia muy citada en internet, Peter Drucker indica que tratar de eliminar el riesgo en las empresas es algo inútil. El riesgo es algo inherente al hecho de comprometer recursos actuales en busca de resultados futuros. De hecho, el progreso económico se define como la habilidad de tomar riesgos.

Las entidades públicas fundamentan la administración de riesgos en un marco soportado principalmente por las normas, para el caso colombiano puede verse el marco normativo del Anexo 1.

La administración de riesgos es un componente del subsistema de control estratégico y esta se sirve de la planeación estratégica, de la aplicación de dicha planeación en la estructura de los sistemas integrados de gestión y de los procesos de control interno de las entidades. En la administración de riesgos se tiene en cuenta la agrupación de controles y sus relaciones subyacentes, de esta forma las entidades públicas logran ejercer acciones sobre los eventos que puedan impactar sobre los retos que se ha definido la entidad, la administración de riesgos fomenta una cultura organizacional orientada a la continua autoevaluación y autocontrol de la entidad.

Para la identificación de riesgos en las entidades públicas, debe conformarse un equipo interdisciplinario de personas que estén asignadas a diferentes áreas de la entidad y que posean un conocimiento profundo de los procesos. Desde la óptica de los autores de este documento, la administración de riesgos se fundamenta en la identificación de los riesgos y la determinación de sus causas y sus consecuencias, sin estos aspectos definidos, una política de administración de riesgos sería inoperable. Con relación al proceso de administración de riesgos en Colombia, se indica que el Departamento Administrativo de la Función Pública presenta gráficamente un conjunto de operaciones, continuación se muestra la gráfica:



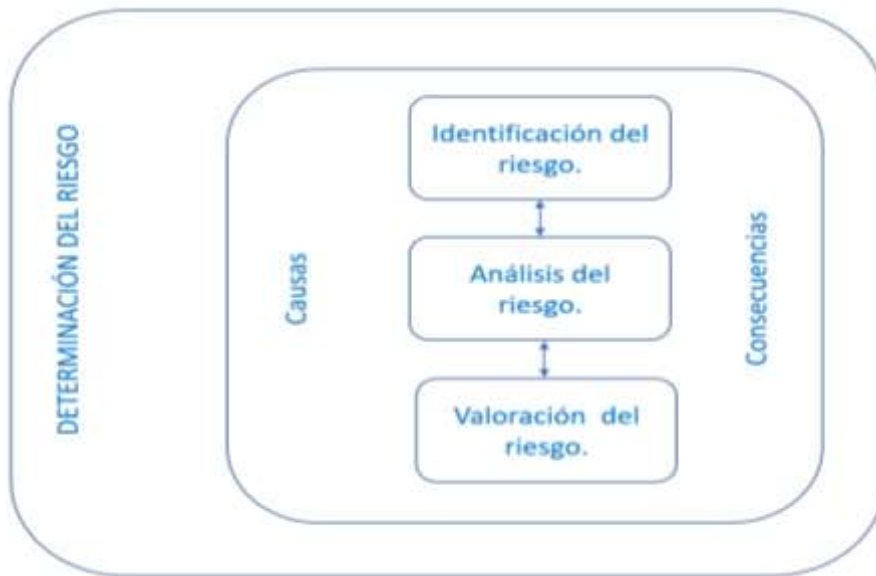
Fuente: Adaptación de los autores del gráfico contenido en la Guía para la Administración del Riesgo del Departamento Administrativo de la Función Pública.

Ilustración 1  
Proceso para la administración del riesgo

La consideración al respecto de que los datos nucleares de la administración de riesgo son los riesgos, las causas y las consecuencias, se fundamenta en que para que exista una administración de riesgos se debe determinar los riesgos que se van a administrar, en esta determinación es necesario establecer lo siguiente:

- Probabilidad ponderada de que el riesgo pueda ocurrir
- Identificar causas de que se presente el riesgo
- Identificar consecuencias de que suceda el riesgo
- Nivel de riesgo (Bajo, medio, alto, extremo)
- Controles que permitirán reducir el riesgo

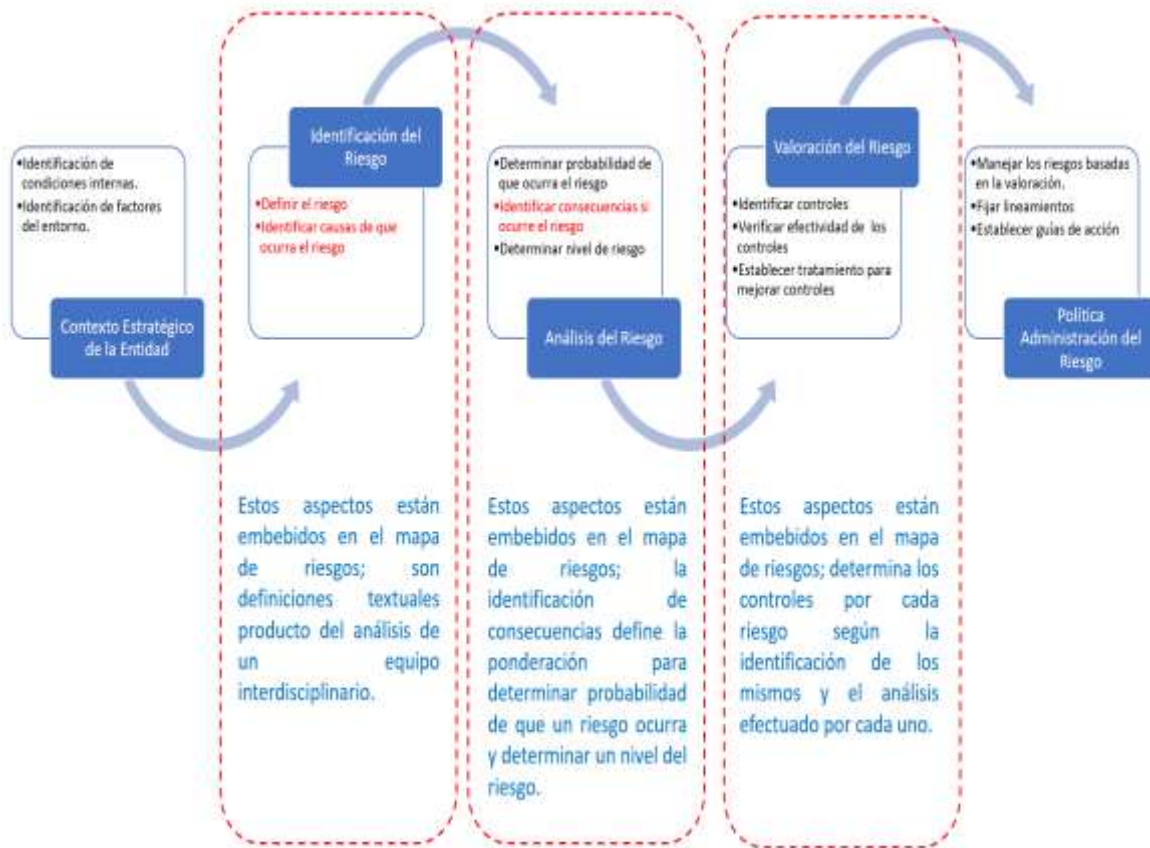
Los anteriores aspectos hacen parte del proceso de administración de riesgos en sus etapas de identificación de riesgo, análisis de riesgo y valoración de riesgo. Sin embargo, lo anterior, los aspectos de identificación del riesgo y sus causas son primordiales para determinar el riesgo, los otros aspectos se consideran son subproductos que dependen de estos, la siguiente gráfica representa el enfoque de las causas y las consecuencias en la determinación del riesgo:



Fuente: Diseño propio de los autores

Ilustración 2  
Modelo de formulación del riesgo

El proceso de administración de riesgos se comprende como una secuencia entre etapas, en las cuales se identifica que las causas del riesgo y las consecuencias de este son elementos estructurales que soportan la confección del mapa de riesgos, el cual es una herramienta de la administración de riesgos que se basa en los distintos sistemas y fuentes de información que intentan reconocer las actividades o procesos susceptibles a riesgos. Mediante el mapa de riesgos se cuantifica la probabilidad de que el riesgo se materialice y se mide la afectación potencial en caso de que suceda. La siguiente gráfica muestra la relevancia estructural de las causas del riesgo y las consecuencias de este en su relación con la información que se registra en el mapa de riesgos.



Fuente: Diseño propio de los autores

Ilustración 3  
Proceso de identificación de riesgos

### La coherencia semántica entre riesgos, causas y consecuencias

Teniendo en cuenta la premisa de que una adecuada administración de riesgos se explicita en el mapa de riesgos de las entidades públicas y que los datos nucleares de este mapa deben tener una coherencia semántica para aportar valor al mejoramiento de la entidad, es necesario establecer que estos datos nucleares deben estar interrelacionados semánticamente entre sí, de esta forma se asegura una congruencia entre los conceptos a nivel del lenguaje natural,

Para determinar la congruencia se plantea la necesidad de establecer relaciones entre los riesgos, las causas y las consecuencias, a continuación una gráfica que representa un ejemplo de riesgos, causas y consecuencias tomado los datos nucleares registrados en el mapa de riesgos del Departamento Administrativo de la Función Pública del año 2020.



Fuente: Diseño propio de los autores

Ilustración 4

Ejemplo de articulación del riesgo, su causa y su consecuencia

En el ejemplo gráfico descrito se establece que el riesgo de **pérdida de la confidencialidad** es causado por la cultura de inseguridad, colaboradores que divulgan información, claves genéricas y divulgación no autorizada de claves. Por su parte las consecuencias de que se materialice el riesgo son: Procesos disciplinarios, deterioro del clima laboral de la Entidad y el IM de determinar el responsable de la divulgación.

Analicemos por un momento la frase del riesgo “**Pérdida de la confidencialidad**”, de otra parte, tengamos en cuenta la frase de la consecuencia “**Deterioro del clima laboral de la Entidad**”, lo pertinente sería elaborar la representación semántica entre estas frases bajo el entendido de que una es consecuencia de la otra, en otras palabras, el deterioro del clima laboral de la entidad es consecuencia de la pérdida de la confidencialidad.

Sin embargo, lo anterior, es complejo construir este tipo de relaciones al momento de realizar la determinación de los riesgos que se registra en los mapas de riesgos que elaboran las entidades del estado, para hilar más fino con la complejidad manifiesta, permitámonos analizar la frase de la consecuencia: “**IM de determinar el responsable de la divulgación**”, ahora revisemos la siguiente relación entre riesgo y consecuencia:

**El IM para determinar el responsable de la divulgación es una consecuencia de la pérdida de confidencialidad**

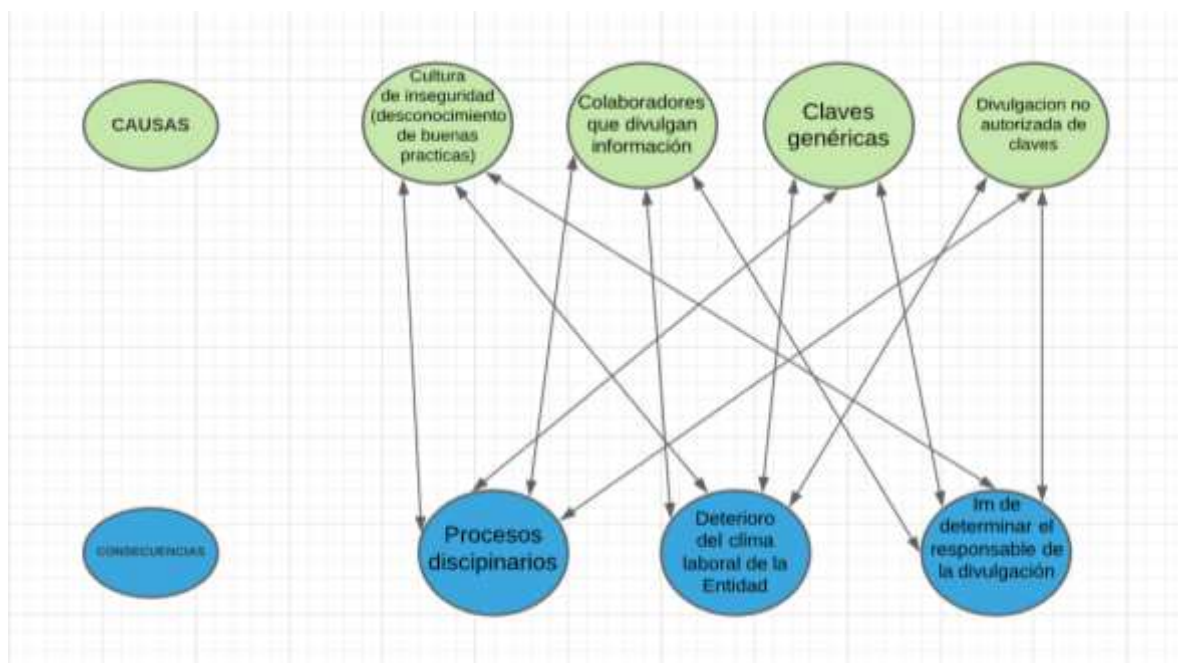
Desde la perspectiva del lenguaje natural comprensible por humanos, la frase anterior no representa significado para el lector, pues no es posible determinar el concepto: “**IM**”, se infiere que es un rol que determina el responsable de la divulgación, pero no es clara la relación semántica entre el riesgo y la consecuencia, la frase **IM de determinar el responsable de la divulgación** genera para el lector que la interpreta una serie de cuestionamientos entre los cuales se presentan los siguientes:

- ¿El concepto IM es un error ortográfico?
- ¿El concepto IM es un término generalizado al interior de la entidad pública?
- ¿El concepto IM es un rol que determina una responsabilidad de la divulgación?

Los procesos mentales del humano que interpreta la relación semántica en el ejemplo anterior generan confusión, por ende, es necesario que la entidad pueda contar con un método que apoye la elaboración de frases que guarden coherencia semántica entre riesgos, causa y consecuencias.

Adicionalmente, se indica que en la determinación de los riesgos participan equipos de personas interdisciplinarias, lo que hace que llegar a un consenso generalizado de conceptos que guarden relación entre sí se hace más difícil.

El relacionamiento semántico ideal no debe estar centralizado desde la definición de un riesgo y las relaciones que este pueda tener con las causas y las consecuencias, el alcance del análisis semántico debe traslaparse también a las relaciones entre las causas y las consecuencias, la siguiente gráfica representa el enmallado de las relaciones semánticas que se pueden establecer con los datos extractados del mapa de riesgo del Departamento Administrativo de la Función Pública del año 2020:

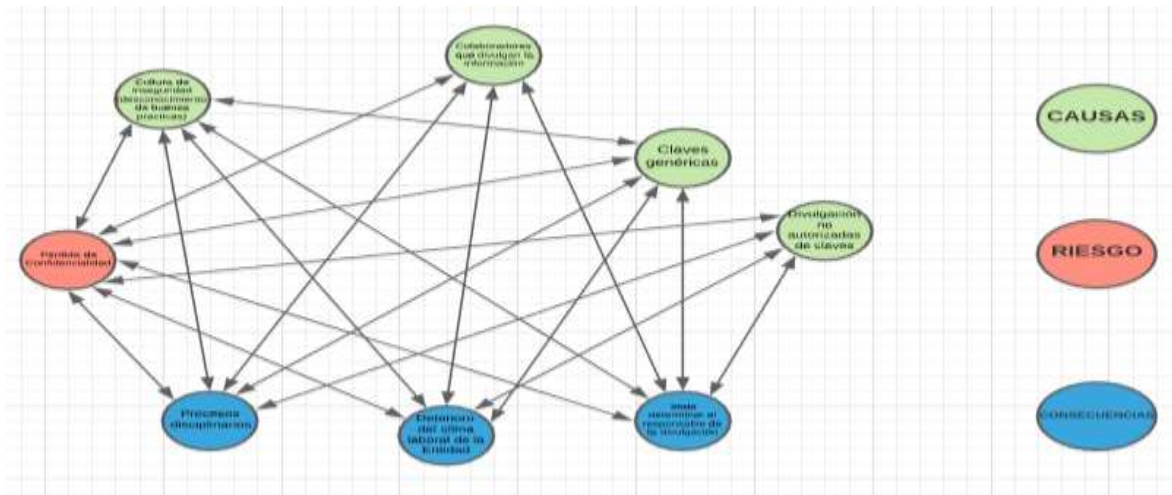


Fuente: Diseño propio de los autores

Ilustración 5  
Posibles relaciones semánticas entre las causas y las consecuencias.

En el marco de las relaciones semánticas de los datos nucleares de los mapas de riesgos, en últimas lo que se pretende es establecer una armonía semántica entre dichos campos nucleares, es decir, las relaciones entre riesgo y causas, riesgo y consecuencias y causas y riesgos, la siguiente gráfica representa las interrelaciones entre todos los datos nucleares del mapa de riesgo:





Fuente: Diseño propio de los autores

Ilustración 6

Posibles relaciones semánticas entre el riesgo, las causas y las consecuencias.

En todas las relaciones semánticas de los datos nucleares, se presenta la complejidad de lograr determinar una coherencia, ante esta dificultad, se esboza como una respuesta eficaz, el echar mano del procesamiento semi-automático del lenguaje natural orientado al análisis de la estructura semántica de los conceptos. En el siguiente numeral se desarrolla la manera de solucionar el análisis de texto con este tipo de procesamiento.

### El procesamiento del lenguaje natural en los datos nucleares de los mapas de riesgos

El procesamiento de Lenguaje Natural -PLN- es una disciplina en el marco de la Inteligencia Artificial asociada a la lingüística computacional, su orientación está dirigida a la investigación del lenguaje natural como aspecto de comunicación entre máquinas y humano<sup>2</sup>. Teniendo en cuenta la problemática que se ha descrito en este documento, se plantea que la solución propuesta es el procesamiento de lenguaje natural para establecer relaciones semánticas entre los datos nucleares de los mapas de riesgos de las entidades públicas. Para comprender la solución, primero hay que introducirnos en el PLN, por ello es necesario indicar que este tiene aplicaciones variadas, entre ellas se pueden mencionar las siguientes:

- Extracción de información
- Generación automática de resúmenes desde textos.
- Identificación de entidades y conceptos
- Extracción de relaciones semánticas entre conceptos

De las mencionadas aplicaciones, se expresa que la que tiene mayor relevancia para la solución del problema descrito es la referida a: **Extracción de relaciones semánticas entre conceptos**, sin embargo, es necesario decir que el principal obstáculo del PLN es que el lenguaje natural tiene gran ambigüedad y debe ser procesado exhaustivamente en sus diferentes niveles, los cuales son:

<sup>2</sup> D. Jurafsky y J. H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition (Prentice Hall Series in Artificial Intelligence). 1<sup>st</sup> ed. Prentice Hall; 2000. Available from: <http://www.worldcat.org/isbn/013095696>

- Léxico
- Estructura
- Pragmático

Como componentes fundamentales de las técnicas de PLN se indican las siguientes:

- Análisis de léxico o morfológico: Radica en analizar las palabras que elaboran sentencias con el fin de sustraer raíces, unidades léxicas, rasgos flexivos, etc.
- Análisis sintáctico. Radica en analizar la estructura de las sentencias de las palabras conforme una gramática establecida en el lenguaje analizado.
- Análisis Semántico: Consiste en la extracción del significado de las sentencias y eliminar ambigüedades.
- Análisis pragmático: Consiste en el análisis del contexto en el cual está embebido el texto analizado y cómo influye este en el significado de este.

El objetivo que nos proponemos consiste en utilizar los métodos del PLN y calcular la relación semántica de los datos nucleares del mapa de riesgos de las entidades públicas para facilitar la identificación de la coherencia con la cual se estructura la determinación de los riesgos en el marco de la administración de riesgos.

Retomando lo registrado en la investigación de Pájaro<sup>3</sup>, se tiene que el estudio de las relaciones entre el significado de los conceptos se define como relaciones semánticas, el estudio de estas relaciones está en el marco de la ciencia computacional y tiene el fin de que las máquinas puedan interpretar el lenguaje de los humanos.

El análisis de las relaciones semánticas de los datos nucleares del mapa de riesgos buscaría entonces la coherencia entre estos conceptos de esta forma las entidades públicas se beneficiarán con una herramienta de gestión de riesgos clara y coherente.

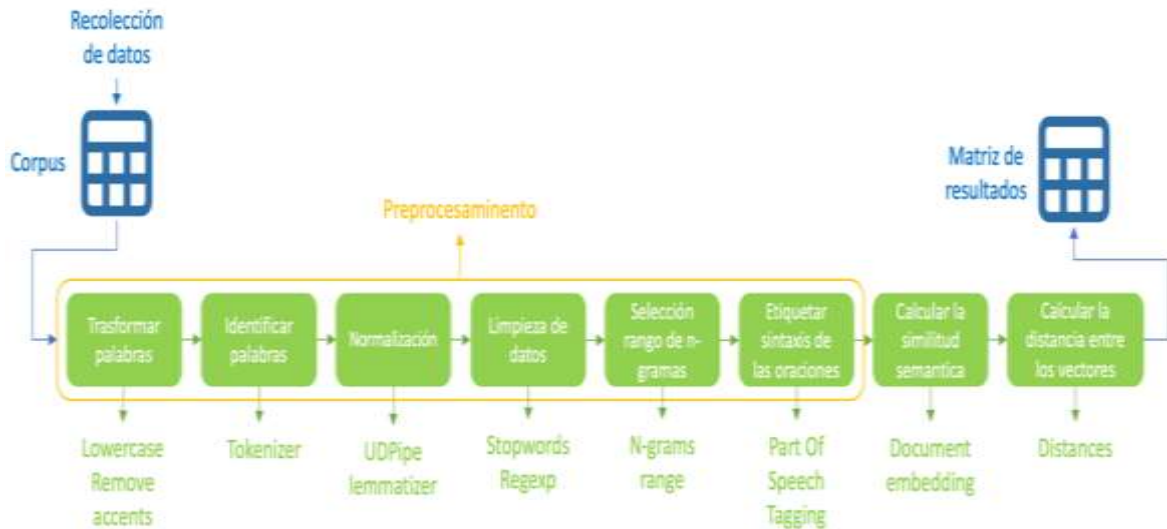
Para el análisis semántico de las oraciones embebidas en los datos nucleares del mapa de riesgo se plantea utilizar la metodología que Pájaro<sup>4</sup> define para un análisis enfocado a las relaciones semánticas entre los conceptos de un árbol del problema en el contexto del modelo de marco lógico, se determina que dicha metodología tiene semejanza a la pretensión de esta investigación, el modelo que Pájaro manifiesta se denomina como Pipeline NPL, a continuación un diagrama extraído de la investigación del autor mencionado, el cual representa el método Pipeline para análisis semántico de oraciones.

---

<sup>3</sup> J. Pajaro Hernández & R. González Rivera, Procesamiento de lenguaje natural para la evaluación de problemas sociales (Bogotá: Pontificia Universidad Javeriana, 2018). Recuperado el 19 de 10 de 2020, <https://repository.javeriana.edu.co/bitstream/handle/10554/40774/Documento.pdf?sequence=4&isAllowed=y>

<sup>4</sup> J. Pajaro Hernández & R. González Rivera, Procesamiento de lenguaje natural





Fuente: Elaboración propia de los autores

Ilustración 7  
Pipeline propuesto

El Pipeline aplicado a esta investigación tendría como entradas las oraciones registradas en los campos nucleares del mapa de riesgo y finaliza con la métrica de las relaciones semánticas entre los conceptos de estas relaciones.

### Metodología aplicada para el procesamiento del lenguaje natural en los datos nucleares de los mapas de riesgos

Para el análisis de la pertinencia semántica entre el riesgo, la causa y el efecto para la evaluación de los riesgos, se partió de la selección de una entidad pública cuya matriz de riesgos se encontrará publicada en su página web, se procedió a realizar la búsqueda en Internet mediante el buscador Google con la siguiente combinación: “mapa de riesgos entidad pública Colombia”.

El primer resultado arrojado fue el mapa de riesgos de la función pública entidad técnica, estratégica y transversal del Gobierno Nacional, cuya función es contribuir al bienestar de los colombianos mediante el mejoramiento continuo de la gestión de los servidores públicos y las instituciones en todo el territorio nacional.

Esta entidad caracteriza la exposición a riesgos que puedan interrumpir el desarrollo de sus funciones mediante el documento “Mapa de Riesgos Institucional y de Corrupción” para efectos del análisis a realizar se seleccionó la versión para el año 2020, dicho documentos reúne los requisitos de identificar los riesgos, causas y efectos objeto del análisis semántico que se pretende desarrollar.

#### 1.1 Recolección de datos

En la gráfica que se presenta a continuación, se observa el Mapa de Riesgos Institucional y de Corrupción de la Función Pública, en la misma se señalan en color verde las tres columnas que contienen el Nombre del riesgo, Causa del Riesgo y Consecuencia

del Riesgo que contienen la información que conforma el corpus a analizar, estas columnas es lo que en este documento se ha denominado como campos nucleares.

El mapa de riesgos institucional muestra una matriz con tres columnas principales: Nombre del riesgo, Causa del riesgo y Consecuencia del riesgo. Las filas representan diferentes tipos de riesgos, como Pérdida de Confidencialidad, y se detallan sus causas y consecuencias. El mapa incluye un encabezado con el título 'Mapa de Riesgos Institucional Versión 1.3 - Julio 7 de 2020' y una barra de navegación superior.

Ilustración  
Mapa de Riesgos Institucional y de Corrupción.

A partir de la matriz se extrajeron los datos nucleares contenidos en las tres columnas, correspondientes al nombre del riesgo, la causa del riesgo y la consecuencia del riesgo. De estos se realizó la extracción de los correspondientes al riesgo seleccionado para el análisis “Pérdida de confidencialidad”, los cuales están contenidos en una tabla de 3 columnas y 4 registros.

Nombre del Riesgo	Causas del Riesgo	Consecuencia del Riesgo
Pérdida de Confidencialidad	Cultura de inseguridad (desconocimiento de buenas prácticas)	Procesos disciplinarios
Pérdida de Confidencialidad	Colaboradores que divulgan la información	Deterioro del clima laboral de la Entidad
Pérdida de Confidencialidad	Claves genéricas	Im de determinar el responsable de la divulgación
Pérdida de Confidencialidad	Divulgaciones no autorizadas de claves	Im de determinar el responsable de la divulgación

Tabla 1  
Datos seleccionados para conformar el corpus.

Una vez extraídos los datos se procedió a estructurar el corpus en dos columnas correspondientes al tipo de oración: Riesgo, Causa y Consecuencia, estos datos se ordenaron de forma tal, que, al presentarse los resultados del análisis no supervisado, se pueda realizar la interpretación de forma controlada, estos datos se exportaron a un archivo CSV, con codificación UTF 8.

Tipo	Oración
Riesgo	Pérdida de confidencialidad
Causa	Cultura de inseguridad (desconocimiento de buenas prácticas)
Causa	Colaboradores que divulgan la información
Causa	Claves genéricas
Causa	Divulgación no autorizadas de claves
Consecuencia	Procesos disciplinarios
Consecuencia	Deterioro del clima laboral de la Entidad
Consecuencia	Imposibilidad de determinar el responsable de la divulgación

Tabla 2  
Estructuración del corpus

Una vez seleccionado el corpus se seleccionan las tareas requeridas para el procesamiento semántico de lenguaje natural, en concordancia con el Pipeline formulado.

## 1.2 Preprocesamiento

Una vez seleccionado el corpus, se procede a extraer una muestra de datos a partir de los cuales se verifica las tareas básicas del PLN definidas en el Pipeline construido, estos son:

**Transformar palabras:** Una vez identificadas las palabras que componen el corpus y de las cuales es posible extraer información, se hace necesario transformarlas de manera que se puedan agrupar palabras con el mismo significado, pero que, por características como acentuaciones, uso de mayúsculas y que al normalizarlas mediante la transformación pueden ser agrupadas por frecuencia o significado. Así mismo de esta actividad es posible identificar palabras como “im” entre otras, que pueden corresponder a abreviaturas no normalizadas por lo que se debe incluir en el modelo a implementar la transformación de dichas palabras, por ejemplo “im” puede corresponder a la palabra “imposibilidad”.



Fuente: Diseño propio de los autores

**Identificación de palabras (tokenización):** La primera actividad a realizar para el procesamiento del lenguaje natural consiste en la separación de los datos contenidos en el corpus en sus unidades con significado, las palabras, luego partiendo de la identificación de las palabras como unidades mínimas con significado que permiten el análisis del lenguaje natural.



Fuente: Diseño propio de los autores

**Limpieza de los datos (stopwords):** Se eliminan los artículos, preposiciones y otras formas verbales que hacen parte de la estructura gramatical de una oración, pero que se considera que carecen de significado, debido a que son de escasa utilidad y de alta frecuencia, por lo que al suprimirlas del corpus se reduce el consumo de recursos computacionales, sin afectar los resultados del procesamiento del lenguaje natural. Lo mismo sucede con los caracteres especiales, como los signos de interrogación, exclamación, paréntesis con los cuales se presenta la misma condición comentada con los artículos y las preposiciones.



Fuente: Diseño propio de los autores

**Etiquetar la sintaxis de las oraciones:** El proceso de etiquetado de la palabra permite identificar la estructura gramatical de la conformación de una oración de tal forma que se pueda calcular la probabilidad de aparición de un tag en un nuevo texto y de igual forma permite clasificar cada palabra según su función en la oración.



Fuente: Diseño propio de los autores

De esta manera se han realizado las actividades básicas del Procesamiento de Lenguaje Natural, brindado a la máquina una fuente a partir de la cual pueda modelarse su entrenamiento, generar un modelo, evaluarlo para determinar su efectividad y posteriormente aplicarlo.

**Calcular similitud semántica:** Mediante este paso básico se busca identificar oraciones compuestas por diferentes palabras, o mediante estructuras gramaticales pero que tienen significados similares, así como la identificación de oraciones con la misma estructura gramatical pero que con el cambio de una sola palabra, pueden cambiar totalmente de sentido o que no permiten la interpretación del mismo concepto.

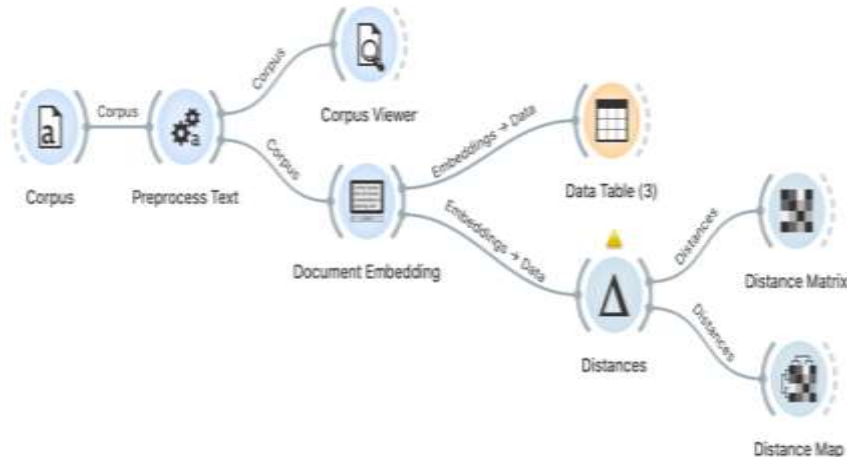
De esta forma se permitirá la agrupación para el entrenamiento de estas oraciones con similitud semántica y descarte de aquellas, cuyo significado sea diferente.

Esta actividad consiste en cargar los  $n$  gramas que componen las oraciones en un plano vectorial de acuerdo con modelos formulados de FastText, estos asignaron un valor vectorial según relaciones semánticas entre los  $n$  gramas, lo que permite asociar diferentes  $n$  gramas dada su relación gramatical.

**Calcular la distancia entre los vectores:** Ya cargados los valores vectoriales correspondientes a los  $n$  gramas que componen cada una de las oraciones del corpus, se realizó el cálculo de distancias entre los mismos, para esto se utilizó la métrica de coseno, a partir del cual se obtendrá un valor entre 0 a 1 en el que 1 será la máxima similitud semántica y 0 la menor.

## Resultados

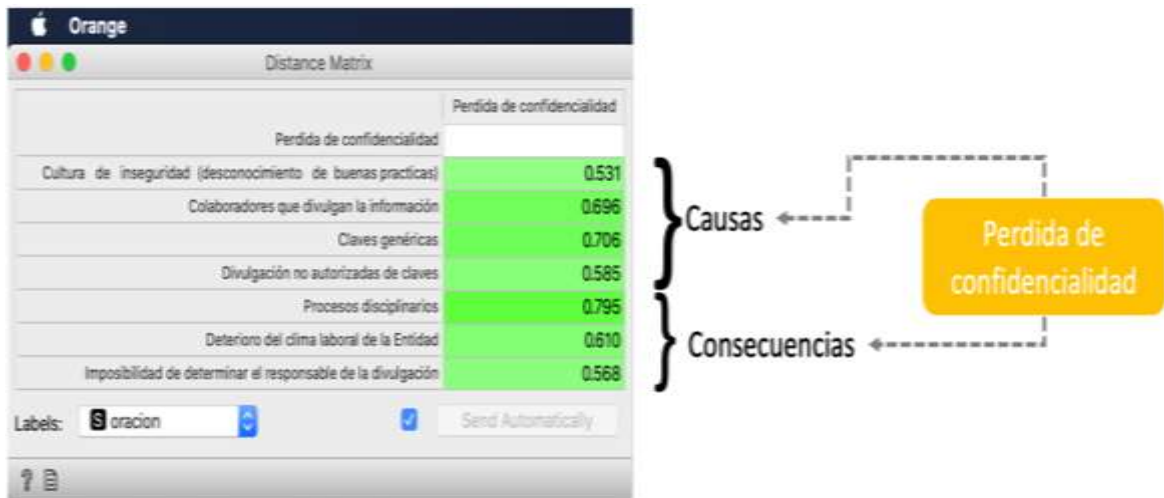
Para realizar el análisis semántico de las diferentes oraciones que componen la matriz de riesgos de la función pública, en específico de aquellas que se refieren al riesgo de “Pérdida de confidencialidad”, se optó por hacer el modelado en la herramienta Orange, llevando el modelo planteado en el Pipeline a la estructura de Widgets de la herramienta seleccionada.



Fuente: Diseño propio de los autores

Ilustración 9  
Modelo estructurado en la herramienta Orange

Terminado el proceso de modelado, se cargó el corpus elaborado en el modelo final, posteriormente a la ejecución de los procesos se desplegó la matriz de distancia (Distance Matrix), el cual arrojó los resultados que se muestran a continuación.

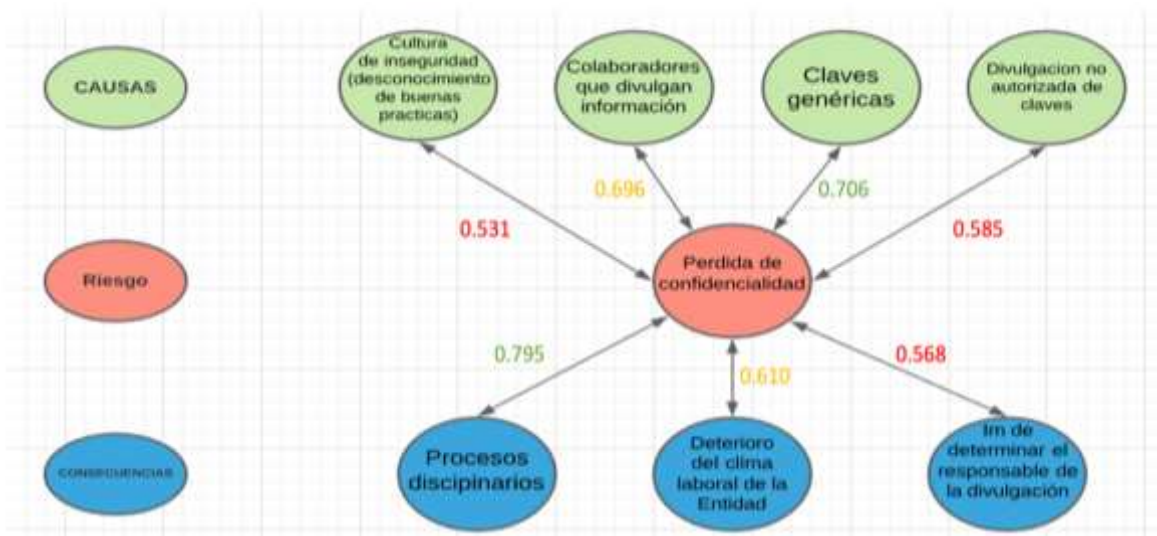


Fuente: Diseño propio de los autores

Ilustración 10  
Resultados matriz de distancia

Los datos arrojados, acorde con lo esperado, mostraron valores entre 0 y 1 de acuerdo con la escala que se planteó para el análisis, lo que permitió abordar la interpretación partiendo del primer dato el cual contiene el riesgo a ser evaluado “pérdida de la confidencialidad”, seguido por las cuatro (4) causas en las filas 2 a 5, y finalmente las tres (3) consecuencias de la 6 a la 8, lo que simplificó la lectura de la matriz, requiriendo solamente la lectura de los datos registrados en la primera columna y no en la totalidad de la matriz.

Esto permitió actualizar el “Ejemplo de articulación del riesgo, su causa y su consecuencia” planteado en la Ilustración 4, con los resultados del cálculo de distancia de los vectores, quedando como se muestra a continuación:



Fuente: Diseño propio de los autores

Ilustración 11  
Resultados calificación del modelo frente a la relación de la matriz de riesgo



Para la pregunta planteada respecto a la medición de la relación semántica en la definición de riesgos, causas y consecuencias, se observa que el riesgo de “pérdida de confidencialidad”, tiene mayor relación semántica con la causa que hace referencia al uso de “Claves genéricas” con un puntaje de 0.706, lo que se entiende como una debilidad de seguridad, lo que expone a la entidad a posibles intrusiones no autorizadas por parte de terceros; así mismo se observa que la consecuencia “Procesos disciplinarios” obtiene la mayor puntuación del análisis general con 0.795, por lo que se entendería que estos son los que presentan la mayor similitud semántica, mientras que “Cultura de inseguridad (desconocimiento de buenas prácticas)” sería la causa con menor similitud semántica con un 0.531 y la “Im de determinar el responsable de la divulgación” la consecuencia con menor similitud con 0.568.

No obstante, es importante resaltar que no se presentaron valores inferiores a 0.5, de los que se podría concluir que no se encuentran oraciones que presenten baja similitud o incluso sin similitud con el riesgo identificado.

Si bien el análisis al lenguaje natural realizado, arrojó resultados en los que encuentra relaciones semánticas, según los modelos semánticos de FastText utilizados por la herramienta orange, llama la atención que la consecuencia con mayor puntaje fuera “Procesos disciplinarios”, esto en el sentido que al analizar el riesgo de “pérdida de confidencialidad” desde la comprensión del equipo de trabajo, no tiene relación para efectos del sujeto que se ve afectado al manifestarse el riesgo, ya que la posible fuga de la información afecta es a la entidad custodia de la información, en este caso la Función Pública, mientras que el proceso disciplinario es aquel que evalúa la responsabilidades de un colaborador por presentar un comportamiento no ético, moral, legal o fiscal, por lo que no son correspondientes.

En el mismo sentido en el caso de las causas, se observan dos causas “Claves Genéricas” el cual como se mencionó anteriormente es el de mayor similitud semántica y “Divulgación no autorizada de claves”, que fue la segunda con menor puntaje según los resultados obtenidos, si bien no pareciera tener relación si concuerdan en ser causas relacionadas con la palabra “clave” y en ese sentido, las dos tienen relación en la posibilidad que se manifieste el riesgo objeto del análisis planteado.

En el sentido de los dos ejemplos enunciados anteriormente, se concluye que si bien los resultados arrojados tienen relación semántica, pueden presentarse resultados en los que el análisis ejecutado por máquina mediante NPL en los cuales no se identifica relación aparente pero que desde el concepto humano si lo tiene, por lo que se entendería que los resultados aún carecen de precisión o exactitud.

## Conclusiones

Es posible determinar que es viable realizar análisis semántico a través del procesamiento de lenguaje natural, se considera en el procesamiento masivo de formulaciones de causas y consecuencias se presenta ante nosotros como una alternativa para validar la pertinencia de las caracterizaciones realizadas, lo que permite retroalimentar a quien caracteriza los riesgos en las entidades públicas para que mejore la composición de estas los campos nucleares.

De los resultados obtenidos se podría entender globalmente que los riesgos y consecuencias caracterizados en la matriz de riesgo de la Función Pública, presentan una



relación semántica que oscila entre el 0.531 y el 0.706, los que permite llegar a la conclusión que no se identifican formulaciones sin relación, pero que el ejercicio realizado puede ser optimizado, redactando oraciones que logren una mayor similitud semántica, lo que resultaría en una mejora en la caracterización de riesgos.

Así mismo los resultados evidenciados, plantean la necesidad de formular y usar modelos lingüísticos pre entrenados, propios del dominio de la evaluación de riesgo, al igual que del lenguaje español y que hagan referencia al contexto del país, que permitan obtener análisis más precisos o exactos según corresponda.

## Anexo 1

### Fundamentados normativo de la Administración de riesgos

- Ley 1474 de 2011, la cual define el Estatuto Anticorrupción. Artículo 73. “Plan Anticorrupción y de Atención al Ciudadano” el cual deben elaborar anualmente todas las entidades, incluyendo **el mapa de riesgos de corrupción**, las medidas concretas para mitigar esos riesgos, las estrategias antitrámites y los mecanismos para mejorar la atención al ciudadano.
- Decreto 4485 de 2009, Por el cual se adopta la actualización de la NTCGP a su versión 2009. Numeral 4.1 Requisitos generales literal g) “establecer **controles sobre los riesgos** identificados y valorados que puedan afectar la satisfacción del cliente y el logro de los objetivos de la entidad; cuando un riesgo se materializa es necesario tomar acciones correctivas para evitar o disminuir la probabilidad de que vuelva a suceder”.
- Decreto 1599 de 2005, por medio del cual se adopta el Modelo Estándar de Control Interno (MECI) para todas las entidades del Estado, en el que la **“Administración del Riesgo”** se define como uno de los componentes del Subsistema de Control Estratégico.
- Decreto 1537 de 2001, mediante el cual se establecen elementos técnicos requeridos para el desarrollo adecuado y fortalecimiento del Sistema de Control Interno de las diferentes entidades y organismos de la Administración Pública, siendo uno de estos la **“Administración del Riesgo”**.
- Ley 87 de 1993, por la cual se establecen normas para el ejercicio del control interno en las entidades y organismos del Estado y se dictan otras disposiciones. (Modificada parcialmente por la Ley 1474 de 2011). Artículo 2 Objetivos del control interno: literal a). Proteger los recursos de la organización, buscando su adecuada administración ante **posibles riesgos** que los afectan. Literal f). Definir y aplicar medidas para **prevenir los riesgos**, detectar y corregir las desviaciones que se presenten en la organización y que puedan afectar el logro de los objetivos.

## Bibliografía

Departamento Administrativo de la Función Pública. (19 de 10 de 2020). Guía para la administración del riesgo. Obtenido de <https://www.funcionpublica.gov.co/documents/418537/506911/1592.pdf/73e5a159-2d8f-41aa-8182-eb99e8c4f3ba>

Departamento Administrativo de la Función Pública. Mapa de Riesgos Institucional versión 12 - Julio 7 de 2020. Obtenido de <https://www.funcionpublica.gov.co/documents/418537/506987/mapa-riesgos-institucionales.xlsx.pdf/ba3b6bd9-a4ed-8a6c-6440-05392bc153bb?t=1594163160657>

García Hanson, J.; Salazar Escobar, P. & Olivares A. E. Métodos de Administración y Evaluación de Riesgos. (F. d. Negocios, Ed.) Primavera (Chile): Universidad de Chile. 2015. Recuperado el 19 de 10 de 2020, de [http://repositorio.uchile.cl/tesis/uchile/2005/garcia\\_j2/sources/garcia\\_j2.pdf](http://repositorio.uchile.cl/tesis/uchile/2005/garcia_j2/sources/garcia_j2.pdf)

Jurafsky, D. & Martin, J. Speech and language processing: an introduction to natural language processing, computational linguistics and speech Recognition. Prentice Hall. 2020. Recuperado el 19 de 10 de 2020, de <http://www.worldcat.org/isbn/013095696>

Morales, A. (06 de 07 de 2016). La administración de Riesgos en las organizaciones. Obtenido de Portafolio. Blogs: <https://blogs.portafolio.co/buenas-practicas-de-auditoria-y-control-interno-en-las-organizaciones/3000-2/>

Pajaro Hernández, J. & González Rivera, R. Procesamiento de lenguaje natural para la evaluación de problemas sociales. Bogotá: Pontificia Universidad Javeriana. 2018. Recuperado el 19 de 10 de 2020, de <https://repository.javeriana.edu.co/bitstream/handle/10554/40774/Documento.pdf?sequence=4&isAllowed=y>



CUADERNOS DE SOFÍA  
EDITORIAL

Las opiniones, análisis y conclusiones del autor son de su responsabilidad y no necesariamente reflejan el pensamiento de la **Revista Ciencias de la Documentación**.

La reproducción parcial y/o total de este artículo debe hacerse con permiso de **Revista Ciencias de la Documentación**.