

# **Diseño de una metodología para la extracción de funciones y mapas relacionales a partir de herramientas de minería de texto**

**Autor: Constanza Cuervo Vargas**

**Director de Trabajo de Grado: Jorge Andrés Alvarado Valencia (en coautoría)**

***Maestría en Ingeniería Industrial, Pontificia Universidad Javeriana, Bogotá, Colombia***

## Tabla de Contenido

1. Resumen	3
2. Introducción	3
3. Planteamiento del problema	4
4. Marco Teórico	6
4.1. Elaboración de perfiles	6
4.2. Minería de Texto en correo electrónico	6
5. Objetivos	8
6. Metodología	9
7. Desarrollo del trabajo	11
7.1. Comprensión del problema	11
7.2. Comprensión y preparación de los datos	12
7.2.1. Extracción de variables	12
7.2.2. Selección de las variables extraídas y generadas	12
7.2.3. Análisis de las variables no textuales (Caso 1)	14
7.2.3.1. Correos Enviados y Recibidos	14
7.2.3.2. División por el origen de la persona: país, externo/interno (en relación a la organización) y área a la que pertenece	15
7.2.3.3. División por el tiempo (hora, día, mes) o momento en que el correo es enviado o recibido	17
7.2.3.4. División referente sobre si contiene en el correo adjunto	18
7.2.3.6. Mapa Relacional cargo 1	19
7.2.4. Análisis de las variables relacionadas con el texto (variable contenidos)	21
7.3. Modelado	21
7.4. Validación con un segundo correo Cargo 2	25
7.4.1. Análisis de las variables que no textuales	26
7.4.1.1. Correos Enviados y Recibidos	26
7.4.1.2. División por el origen de la persona: país, externo/interno (en relación a la organización) y área a la que pertenece	26
7.4.1.3. División por el tiempo (hora, día, mes) o momento en que el correo es enviado o recibido	28
7.4.1.4. División referente sobre si contiene en el correo archivos adjuntos	30
7.4.1.5. División por el tamaño del correo	30
7.4.1.6. Mapa Relacional cargo 2.	31
7.4.2. Análisis para las variables textuales	32
7.5. Evaluación y resultados de la investigación	38
7.6. Desarrollo	39
8. Conclusiones	41
9. Recomendaciones futuras	42
10. Referencias Bibliográficas	43
11. Anexos	46

## 1. Resumen

En las organizaciones, uno de los problemas más frecuentes es la ausencia de un procedimiento para documentar, sistematizar y recopilar la información relativa a las funciones de cargos específicos. Esta situación se da en especial, cuando se trata de trabajadores del conocimiento. Estos son entendidos como aquellos cuya labor dentro de una organización está relacionada a sus habilidades para crear, transformar y comunicar conocimiento. Frente a lo anterior, la Minería de Texto se convierte en una herramienta para extraer una fracción de tales actividades y relaciones, para encontrar dentro en el correo electrónico la respuesta a este conocimiento desaprovechado.

## 2. Introducción

Para el estudio de perfiles de cargo, la práctica más usual es la elaboración de diagramas (Chase, 2000) o la realización de procedimientos tradicionales. Comúnmente se realizan encuestas al trabajador, preguntando, por ejemplo, ¿qué hace? ¿Cuándo y cómo realiza una tarea? (Chase, 2000). Este tipo de metodologías implican un tiempo considerable en entrevistas o un conocimiento previo del cargo y por lo tanto es necesario un entrenamiento previo arduo a las personas a las que se les encomienda esta labor. En el caso que la realice el mismo trabajador trae consigo problemas de autorreporte. El detalle de este problema se hablará en la siguiente sección.

Esta investigación plantea una alternativa adicional a la metodología de la extracción de las funciones tradicional, pues se centra en trabajadores del conocimiento, se busca identificar funciones básicas y funciones escondidas que no sean evidentes hasta para el mismo trabajador. Este trabajo se dividirá en un marco teórico, una metodología planteada en varias actividades discutidas en detalle durante el desarrollo del trabajo, la presentación de unos resultados en un producto final (funciones y un mapa relacional) y finalmente, las conclusiones que buscan abrir el interrogante sobre una mejor y más fácil metodología en ese campo.

### 3. Planteamiento del problema

Una de las tareas más importantes en las organizaciones es la descripción clara de las funciones de los cargos y las relaciones de estos dentro del entorno laboral. “Es probable que el diseño del cargo sea tan antiguo como el trabajo humano” (Chiavenato, 2000. Pag 291). Tal descripción es útil porque ayuda en todos los procesos de desarrollo del trabajador al interior de una organización, lo que a su vez facilita sus procesos de selección, capacitación, promoción, la distribución de su trabajo y su evaluación. (Niebel, 1996). Por otra parte, la Norma de calidad ISO 9001 en su cláusula 5.5.1, requiere que las responsabilidades y la autoridad en los cargos sean definidas y comunicadas. Adicionalmente, la documentación de los conocimientos aprendidos por el trabajador hace parte importante de la administración del conocimiento (knowledge management); ésta permite que la organización mantenga su conocimiento aún en el caso en el que los empleados se marchen de la empresa, llevándose consigo contactos, relaciones y el conocimiento acumulado en su puesto de trabajo (Parise, 2006). En consecuencia, la documentación de las responsabilidades y relaciones de un cargo es una ayuda importante en la administración del talento humano, el aseguramiento de la calidad y la administración del conocimiento empresarial.

Las metodologías tradicionales para documentar funciones y relaciones presentan algunas dificultades porque son sesgadas. Por ejemplo, las medidas de auto-reporte son ampliamente utilizadas en la investigación, pero los datos derivados de estos informes pueden verse comprometidos por un problema de datos faltantes (Fox, 2005). Es claro que en estos métodos de auto-reporte los participantes podrían tener la intención de tergiversar la información relacionada con el uso del tiempo dedicado a cada función y el manejo de las mismas. (Robinson, 2010). Los autorreportes dependen de la buena voluntad del empleado, de la capacidad de su memoria y de la importancia que le otorgue a cada asunto. La consecuencia que conlleva es que se generan sesgos en los reportes e información incompleta. Ahora bien, si se utilizara el método tradicional de observación, se restringiría el número de personas que pueden ser examinadas en un momento dado (Robinson, 2010). Entonces, para el logro de dicha documentación la restricción sería de recursos y tiempo.

Esta carencia resulta más problemática aun cuando los empleados son trabajadores del conocimiento (en inglés se conoce este término como *Knowledge Worker* (Hammer & Leonard, 2004)), en adelante nos referiremos a ellos como KW) porque los métodos tradicionales para crear perfiles de cargos fueron establecidos originalmente para trabajadores operativos. Describir escribir las tareas de los KW es más difícil, porque sus flujos de trabajo no son necesariamente repetitivos (Hodson, 1996). Los KW se enmarcan dentro de la Administración del Conocimiento, área que estudia los procesos de adquirir, conservar, compartir y usar el conocimiento y la experiencia (Lloria, 2008)

El conocimiento de los KW se basa en una cantidad cada vez mayor de información transmitida en diferentes medios y por diversas experiencias a la cual un individuo está expuesto (Henard, 2008). Entre las diferentes formas de transmisión de la información utilizadas están las salas de conversaciones (chats), la creación de documentos en programas de procesamiento de texto y de datos y la comunicación vía correo electrónico. Sobre esta última en particular, se quiere anotar que el correo es una de las herramientas más utilizadas por este tipo de trabajadores (Hemp, 2009). Por tanto, el correo electrónico es una valiosa fuente que puede permitir un acercamiento a la obtención de las funciones y las relaciones de los puestos de los KW.

Dado que el correo electrónico no maneja información estructurada, la Minería de Texto (en inglés se conoce este término como *Text Mining*, en adelante nos referiremos a ella como TM), puede ser la herramienta adecuada para encontrar las estructuras de los textos y, en consecuencia, las funciones y relaciones de los KW. La TM promueve “el proceso de descubrimiento de conocimiento y su obtención mediante técnicas de clasificación para la indexación de texto” (Capuano, 2008). Al mismo tiempo pretende profundizar en “el conocimiento de alto nivel y patrones útiles a partir de datos de texto”. Igualmente, trata de “analizar y aprender el significado de la información implícita estructurada de forma automática” (Dorre, Gerstl, & Seiffert, 1999). Aunque el análisis de textos se han aplicado en una amplia gama de campos (Belhadjali, Abbasi & Whaley 2010), a la fecha, de acuerdo con el estado del arte realizado, no se

encontraron aplicaciones de TM en la extracción de funciones, el establecimiento de relaciones y de perfiles de trabajo.

A través de esta investigación se busca crear una fracción de los perfiles de cargo, particularmente sobre las responsabilidades y las relaciones clave, a partir de la aplicación de TM en los correos electrónicos de los KW. Con ello se podrán complementar y comparar, o eventualmente remplazar, los resultados obtenidos a través de las técnicas actuales de levantamiento de esta información. A su vez que servirá de soporte en el caso de que las técnicas tradicionales no hayan podido ser utilizadas por ausencia del empleado en el cargo.

Por tanto nos preguntamos:

¿Es posible crear partes de los perfiles a partir de la aplicación de minería de texto a los correos electrónicos de los Trabajadores del Conocimiento (Knowledge Workers) que sea comparable a las técnicas actuales de levantamiento de esta información?

## 4. Marco Teórico

### 4.1. Elaboración de perfiles

Los perfiles de un cargo pueden ser representados de diversas formas que van desde planteamientos gráficos muy simples, tales como listas de selección, histogramas, y listas desplegables hasta cuadros de radio; de ese modo, apoyan la construcción de consulta y la extracción de los patrones interesantes. Es fundamental en la construcción de las relaciones entre las entidades el uso de un dominio específico de conocimiento (Ronen, 2007).

Para la concepción del perfil del cargo es importante tener en cuenta dos aspectos: El primero consiste en la definición de funciones. La función es entendida como la tarea que corresponde realizar a una institución o entidad, o a sus órganos o personas (RAE). El segundo, la utilización de mapas relacionales. Estos tienen como objeto proporcionar al usuario una herramienta para observar como el trabajador interactúa con otros. En dichos mapas se concentran las relaciones generales entre clientes internos, y entre procesos y funciones (Cuesta, 2010).

### 4.2. Minería de Texto en correo electrónico

En cualquier negocio, “la información está disponible en múltiples formatos. Uno de los más utilizados es el texto sin formato. Algunas investigaciones han calculado que cerca del 80% de los datos está en documentos de texto” (Yu, Wang & Lai, 2005). Por ello, una buena fuente de información para reconocer las actividades que realiza un trabajador del conocimiento es su correo electrónico, pues es el medio más común en el que escribe. El correo electrónico es tan importante para los trabajadores que llega a ser el hábitat central desde el cual se recibe, se administra y se delega en las organizaciones. (Stuit & Wortmann, 2011). Los empleados usan gran parte de su tiempo en él. Por eso, la interacción generada en el correo electrónico es vital para los procesos en las empresas (Kushmerick & Lau, 2005).

La herramienta para extraer la información del correo electrónico es la TM, entendida como un proceso intensivo de conocimiento en el cual el usuario interactúa con un documento usando una herramienta de análisis. La TM identifica y extrae las características representativas de un lenguaje natural, mediante el uso de técnicas y metodologías en el área de la información (Ronen & Feldman, 2007). La tecnología de Minería de Texto “proporciona la flexibilidad para explotar la información a partir de múltiples formatos de datos diferentes como bases de datos relacionales, almacenamiento de datos y bases de datos transaccionales” (Rahman & Harding, 2012).

Los archivos analizados pueden estar en cualquier clase de archivo, desde correos electrónicos, hasta archivos HTML, PDF, procesadores de texto etc. (Ronen & Feldman, 2007). Para esta investigación con el correo electrónico, estaremos centrados en documentos semi-estructurados, por la forma de su semántica, por el uso de los signos de puntuación, las mayúsculas, las palabras, la estructura de las oraciones y la presencia de cualquier tipo de elemento tipográfico.

La TM facilita el trabajo puesto que se realiza automáticamente. Para extraer datos eficientemente en TM, “se debe utilizar un software con las funcionalidades suficientes. En la actualidad, hay muchos programas diferentes, comerciales o gratuitos, disponibles en el mercado” (Zhang y Segall, 2010). En este campo son varios los avances en TM. Amir (2005), por ejemplo, describe una nueva herramienta llamada “máxima asociación” que permite descubrir las asociaciones de interés. Por su parte, Spasic (2005) describe un software de TM para extraer automáticamente la información y los hechos. De esta forma se descubren asociaciones ocultas. Sin embargo, no existe un software estándar para todas las tareas y el análisis de TM usualmente requiere algún nivel de programación de acuerdo con la tarea elegida o debe soportarse en un software de múltiples fuentes.

La utilización de TM en correo electrónico empresarial se ha dirigido en los últimos años a 3 campos principales: detección de fraudes, mejoramiento de servicio al cliente y entendimiento de la red social a partir del correo electrónico.

Las aplicaciones de TM orientadas a la detección de fraudes y al hallazgo de delitos informáticos han sido exitosas (Holton, 2009 & Chen, 2011 & Thomson, 2005) particularmente para identificar al autor más plausible de un correo electrónico anónimo de un grupo de posibles sospechosos (Iqbal & Hadjidj, 2008). La extracción de conocimiento y la información de un texto de correo electrónico se ha convertido en un importante paso para la investigación de la ciberdelincuencia y la recopilación de pruebas. (Iqbal & Hadjidj 2010).

Son diversas las investigaciones realizadas a partir de correos electrónicos relacionadas con servicio al cliente. Por ejemplo de detección de abuso en internet (Chou, 2008) o la investigación realizada por Sakurai (2004) donde propone un método que emplea técnicas de TM para analizar los correos electrónicos recogidos en un centro de servicio; el método extrae los conceptos clave de correos electrónicos y presenta su información estadística. Sakurai al continuar con su investigación, descubre las frecuencias de secuencias de modelos de ítems de información de texto. Los modelos analizan la tendencia de los datos de texto y pueden predecir eventos futuros utilizando dichos modelos. (Sakuari, 2008). Weng & Liu, (2004) usan técnicas de clasificación de texto aplicados a las sugerencias por correo electrónico mediante el uso de una plantilla de respuesta, con el fin de reducir la carga del personal de servicio al cliente para responder a mensajes de correo electrónico. El objetivo primordial de estas investigaciones es apoyar las áreas de servicio al cliente, aliviando su labor mediante métodos de filtro y predicción. De donde se deduce, que la TM podría usarse para reducir la carga de análisis de información por parte de personas.

Existen también investigaciones referidas a las interacciones halladas en el correo electrónico. Por ejemplo, Kang (2010) descubrió relaciones entre los individuos que participan en este medio, al llevar su análisis hasta las redes sociales existentes. Otras investigaciones encuentran las conexiones explícitas e implícitas entre las personas. La Minería se desarrolla en las asociaciones semánticas deducidas de sus comunicaciones de correo electrónico (Mc Athur & Bruza, 2003). También está la que estudian temas de autoridad y liderazgo a partir del correo electrónico (Tylor & Wilkinson, 2005). También encontramos investigadores que trabajan métodos para descubrir el manejo de los procesos de negocio, como es el caso de Stuit & Wortmann (2011). En esta investigación se construye un método para reconocer interacciones, atributos y roles. Podemos deducir que la TM ha resultado útil para la detección de relaciones sociales y personales claves a partir del correo electrónico, es así como podría ser utilizada para detectar las principales relaciones laborales de un cargo dado.

El análisis de Enron (2010) y el acercamiento de Leuski (2004) son dos ejemplos importantes de señalar. En el análisis del correo electrónico de Enron se describen con detalle las redes sociales de las personas de la compañía, el contenido y la visualización de los datos encontrados con la ayuda de un sistema diseñado para apoyar la exploración de la red de contenido y de los actores que participan en las grandes colecciones de correos (Kang 2010). Sin embargo, los análisis de Enron han estado altamente enfocados a la detección de sentimientos previos a la quiebra de la compañía y a rastrear el clima organizacional y la intensidad de las interacciones, de acuerdo con la cronología de la quiebra económica de la compañía. El segundo ejemplo, es el acercamiento al análisis del correo electrónico que realiza Leuski (2004). Al revisar algunos roles, a través del análisis del correo electrónico, identificó el rol que desempeña el trabajador, según el contexto, al separar los correos electrónicos personales de los profesionales. Cabe señalar que este estudio es un primer acercamiento al análisis del correo electrónico para revisar la clase de correos electrónicos que se manejan a nivel laboral.

Sin embargo, no hemos encontrado una investigación dirigida a la extracción de funciones a partir de un análisis de TM del correo electrónico. Tampoco la hemos hallado en el idioma español. Por estas razones consideramos que el principal aporte de nuestro artículo radica en extraer funciones laborales en el análisis del TM del correo electrónico en español.

## 5. Objetivos

### Objetivo General

Diseñar una metodología para la extracción de información para los perfiles de cargo de trabajadores de conocimiento utilizando herramientas de minería de texto en su correo electrónico.

### Objetivos específicos

1. Diseñar una metodología para describir funciones de un trabajador del conocimiento extrayendo por medio de minería de texto la información de su correo electrónico.
2. Diseñar una metodología para crear el mapa relacional de un trabajador del conocimiento extrayendo por medio de minería de texto la información de su correo electrónico.



## 6. Metodología

La metodología utilizada es el Modelo CRISP-DM. El CRISP-DM fue creado en 1996 en el campo de Minería de datos (Chapman & Clinton, 1996). Este modelo contiene fases de un proyecto, sus tareas respectivas, y las relaciones entre estas tareas. Si bien, el ciclo de vida del proyecto de Minería consiste en seis fases, estas no son rígidas. Por el contrario, el movimiento hacia adelante y hacia atrás entre fases diferentes es siempre requerido. Lo anterior hace que la naturaleza de esta metodología sea cíclica, ya que en minería no se termina una vez la solución sea desplegada; siempre habrá informaciones ocultas durante los procesos, las cuales pueden provocar nuevas preguntas. Aunque la actividad de validación no está incluida en la metodología original, la incluimos por considerarla importante. Ver Gráfica 1.



Gráfica 1. Metodología CRISP-DM

Dentro de la extracción de la información, hay dos pasos fundamentales que ayudan a la realización de la definición de las funciones (Ronen, 2007): 1. La definición de entidades (personas), atributos (tipo de cargo y área de la organización a la que pertenece) y factores (relaciones entre las personas) 2. La identificación de grupos nominales y verbales. Estos elementos se utilizan como bloques de construcción para armar las relaciones.

Paralelamente se creó un pequeño glosario básico de terminología de TM para lograr una mejor comprensión de la metodología. A continuación se presentan brevemente los métodos y pasos intermedios que se utilizaron para el desarrollo del modelo:

- *Stop Words*<sup>1</sup>: son palabras comunes que se ignoran, para este trabajo se definieron que las stop words serán los adjetivos, adverbios, conjunciones, y artículos, porque no se consideraron.
- Las partes de la oración (*Part of speech*): en la gramática tradicional se clasifica las palabras basándose en ocho partes de la oración: el verbo, el sustantivo, el pronombre, el adjetivo, el adverbio, la preposición, la conjunción y la interjección.
- Lematización: Proceso lingüístico que consiste en obtener la forma original de la forma flexionada de una palabra, es decir, en plural, en femenino, conjugada. Es la palabra que nos encontraríamos como entrada en un diccionario tradicional: singular para sustantivos, masculino singular para adjetivos, infinitivo para verbos.

<sup>1</sup> Se incluye este término en inglés porque se considera que su traducción en español no contiene toda la definición necesaria.

- *Collaborative filtering*<sup>2</sup>, Existen hoy sistemas de recomendación que aumentan el proceso social natural de ayudar a filtrar cosas interesantes como libros disponibles, artículos, páginas web, películas, música, restaurantes, chistes, productos comestibles, y hasta encontrar la información más interesante y valiosa. Los sistemas de recomendación se les conoce como Collaborative Filtering (Su & Khoshgoftar).
- *MBA (Market Basket Analysis)*<sup>3</sup>: Herramienta creada para analizar el proceso de compra de un consumidor revisando sus hábitos de consumo y encontrando asociaciones entre los diferentes ítems que compra el consumidor, pero en nuestro caso se utilizó para revisar la asociación entre los verbos y los sustantivos, y hallar las relaciones existentes. (HoanCan & Mock, 2011)

Para comprender el segundo paso mencionado en la página anterior, vale la pena mencionar que una palabra puede ser clasificada dentro del TM como parte de una oración (“part of speech”). La extracción de ellos han sido objeto de muchos estudios en los últimos años (Petrovic, 2010). Las partes básicas de una oración son el sustantivo, el pronombre, el verbo, el adjetivo, el adverbio, la preposición, la conjunción y la interjección (Nebrija, 2012). Para este estudio se tomaron dos partes fundamentales de la oración: primero, el verbo, indispensable para identificar una acción. Un verbo básicamente responde a la pregunta: ¿qué está haciendo el sujeto? En segundo lugar, los sustantivos, porque con ellos se identifica sobre quién o qué recae la acción, y de esa forma complementan el entendimiento de la acción.

Para empezar, la automatización del proceso de clasificación de los datos de entrada estos deben ser representados en un formato adecuado. Esto incluye la aplicación de diferentes técnicas de minería de datos. Uno de los pasos fundamentales es la eliminación de *stop words*. (Rahman & Harding, 2012).

---

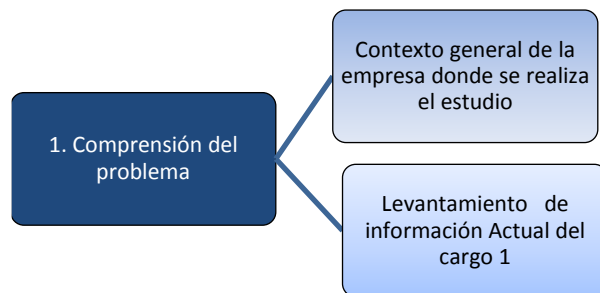
<sup>2</sup> Se incluye este término en inglés porque se considera que su traducción en español no contiene toda la definición necesaria.

<sup>3</sup> Se incluye este término en inglés porque se considera que su traducción en español no contiene toda la definición necesaria.

## 7. Desarrollo del trabajo

### 7.1. Comprensión del problema:

En esta etapa se realizaron 2 actividades. Ver Gráfica 2.



Gráfica 2. Actividades principales de la comprensión del problema

La empresa donde se realiza este estudio es de tipo productivo, líder dentro del sector de la Educación donde es un facilitador del proceso de aprendizaje, la adquisición del conocimiento y del entretenimiento de las personas. Cuenta con un equipo de trabajo conformado por 2.667 colaboradores. De ellos, el 40% está ubicado en Colombia, le siguen México con el 24%; Brasil con 12%; Ecuador, Perú, Venezuela y Argentina, cada uno con el 4%, y el 8% restante lo conforman Puerto Rico, Chile, Guatemala, El Salvador, Costa Rica y Panamá. Tiene un 30% del mercado en los productos que ofrece. En Colombia tiene presencia en las principales ciudades y cuenta con una gran infraestructura propia.

Las dos personas que prestaron su correo laboral tienen cargo de coordinadoras, la primera pertenece al área de compras y la segunda pertenece al área editorial. A continuación se da más detalles:

- El primer caso con el que se realizó el modelo hace referencia a la coordinadora de compras de impresión de libros de literatura que en el momento de la investigación pertenecía al área de compras editoriales y esta área a su vez pertenecía al área de operaciones. El investigador tiene conocimiento cercano al cargo precisamente para realizar un modelo cercano a la realidad en la fase de entrenamiento.
- El segundo caso con el que se validó el modelo hace referencia a la coordinadora del proceso editorial que pertenece a la Gerencia de Contenidos y Soluciones Educativas. De este segundo caso, el investigador solo realizará el análisis de las variables no textuales y la validación del modelo de las variables textuales la hacen dos personas externas de la organización que no tienen conocimiento alguno del cargo, para garantizar la validación.

Se recopiló la información del proceso actual que se utiliza para el levantamiento de la descripción de funciones dentro de la empresa donde revisaremos los dos casos de estudio. Se recopiló la terminología necesaria para la descripción de funciones y relaciones. Entre los hallazgos encontramos que solo existe un formato para la descripción del cargo. Este formato no aplica para todos los cargos: Solo cubre el 50% de las personas que está laborando directamente en empresa, de acuerdo con la información suministrada por el área de Recursos Humanos.

Se revisó el formato de descripción de cargo. El formato diligenciado no se tomó en cuenta durante el estudio para evitar sesgos. Ver Anexo 1.

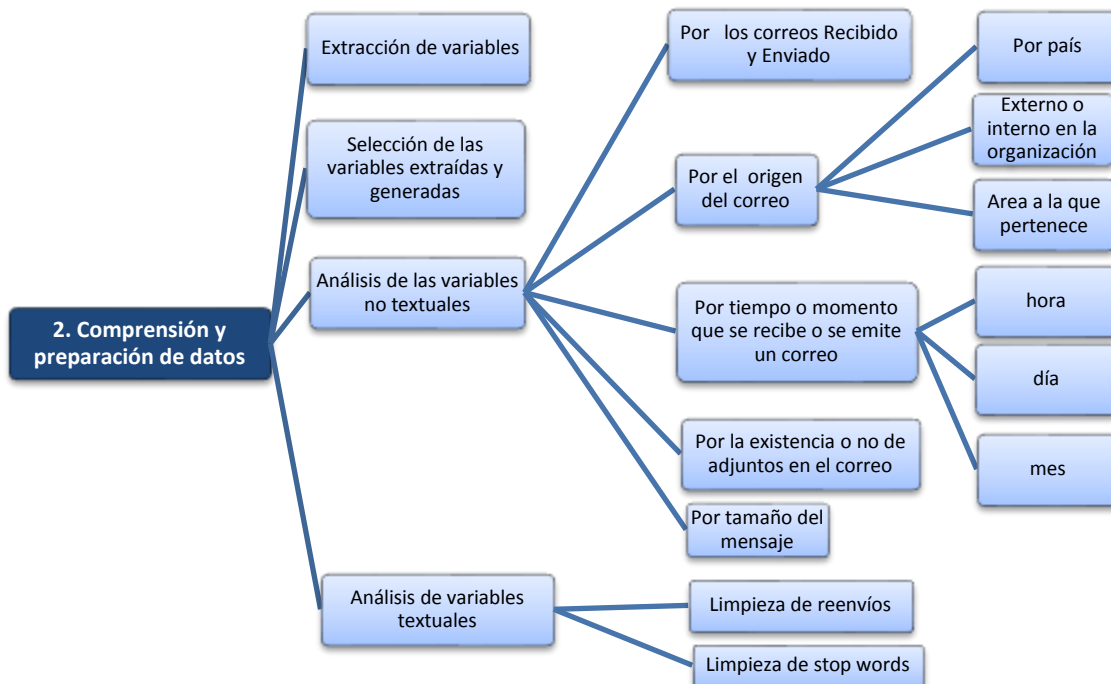
En esta investigación se pretende dar apoyo únicamente a las siguientes partes de la descripción de cargo, las cuáles están incluidas dentro del formato anteriormente descrito:

- La Misión del cargo: Qué hace el trabajador del conocimiento
- Actividades del cargo: Cuáles son sus principales funciones
- Problemas típicos del cargo
- Relaciones con otras áreas: Con qué áreas se relaciona y con qué frecuencia

Los demás aspectos contenidos en el perfil no se contemplan porque a partir de la información contenida en el correo no se pueden determinar aspectos como la experiencia, la responsabilidad en Oracle, o los salarios, por ejemplo.

## 7.2. Comprensión y preparación de los datos:

La siguiente gráfica muestra las actividades más importantes en este paso. Ver Gráfica 3 :



Gráfica 3. Principales actividades del paso de Comprensión y preparación de datos

### 7.2.1. Extracción de variables

Para realizar el muestreo y el modelo del primer caso de estudio, se obtuvo un archivo de correo electrónico (.PST) durante el lapso de un año. Inicialmente se utilizaron programas de Outlook, Excel, Access, con los cuales se realizaron los procesos de importación y exportación de los datos al software y al SPSS para el análisis de datos. Ver Tabla 1.

Variable	Ubicación	Etiqueta	Nivel de medida	Tipo
<b>Id</b>	1	Consecutivo de control	Escala	Numérico
<b>EnviadoRecibido</b>	2	Determina si el correo enviado es Enviado o Recibido	Nominal	Cadena
<b>Importancia</b>	3	Este calificativo lo utiliza la persona dueña del correo para calificar su correo como importante	Escala	Numérico
<b>Icono</b>	4	Información propia del correo	Nominal	Cadena
<b>Prioridad</b>	5	Este calificativo lo utiliza la persona dueña del correo para darle prioridad a su correo	Escala	Numérico
<b>Asunto</b>	6	Describe en forma general el título del correo	Nominal	Cadena
<b>De</b>	7	Es el nombre de la persona que envía el correo	Nominal	Cadena
<b>Mensajeparamí</b>	8	Mensaje para mí (control del sistema)	Escala	Numérico
<b>MensajeCCparamí</b>	9	Mensaje CC para mí (control del sistema)	Escala	Numérico
<b>Nombre del remitente</b>	10	Nombre del remitente (es el mismo De)	Nominal	Cadena
<b>CC</b>	11	Persona a la que le enviamos el correo	Nominal	Cadena
<b>Recibido</b>	12	Fecha en que se recibe el mensaje	Escala	Fecha
<b>Tamaño del mensaje</b>	13	Tamaño del mensaje	Escala	Numérico
<b>Contenidos</b>	14	Cuerpo del contenido del correo	Nominal	Cadena
<b>Creado</b>	15	Fecha de creación del correo	Escala	Fecha
<b>Modificado</b>	16	Fecha de modificación del correo	Escala	Fecha
<b>Prefijo de asunto</b>	17	Prefijo de asunto (puede identificar cuando es reenviado o cuando es aceptado)	Nominal	Cadena
<b>Contiene datos adjuntos</b>	18	Contiene datos adjuntos (es una variable binaria 0=no contiene, 1=si contiene)	Escala	Numérico
<b>Asunto normalizado</b>	19	Asunto normalizado: es el mismo título pero sin la identificación si fue reenviado o no	Nominal	Cadena
<b>Tipo de objeto</b>	20	Tipo de objeto	Escala	Numérico
<b>Contenido no leído</b>	21	Contenido no leído	Escala	Numérico
<b>Índice 1</b>	22	Identificación si es la primera o segunda o tercera persona a la que se le envió el correo	Nominal	Cadena
<b>Nombre Enviado</b>	23	Nombre de la persona a la que se le envió el correo	Nominal	Cadena
<b>País</b>	24	País de la persona a la se le envía el correo. En el caso de los recibidos es el país de la persona de la que se recibe el correo	Nominal	Cadena
<b>Interno Externo</b>	25	Interno-Externo: Interno refiriéndose a dentro de la empresa. Externo refiriéndose a la persona fuera de la empresa	Nominal	Cadena
<b>Area</b>	26	Área de la persona que envía o recibe el correo	Nominal	Cadena
<b>MDAY1.31</b>	27	Día del mes (1-31) en que el correo fue enviado o recibido	Escala	Numérico
<b>MONTHY1.12</b>	28	Mes del año (1-12) en que el correo fue enviado o recibido	Escala	Numérico
<b>WKDAY1.7</b>	29	Día del año (1-7) en que el correo fue enviado o recibido, entendiéndose como 1 el Domingo y 7 el día Sábado)	Escala	Numérico
<b>HOUR1.23</b>	30	Hora del día (1-23) en que el correo fue enviado o recibido	Escala	Numérico
<b>AÑO</b>	31	Año en que el correo fue enviado o recibido	Escala	Numérico

Tabla 1. Variables del archivo de trabajo (extraídos del SPSS)

### **7.2.2. Selección de las variables extraídas y generadas**

El correo electrónico nos proporcionó bastantes datos. No obstante, solo se utilizaron las siguientes variables, tanto para los correos enviados como para los recibidos. Estas variables se extrajeron directamente del correo electrónico:

- Id
- De
- Para o CC
- Recibido o enviado
- Contenidos
- Contienedatosadjuntos
- Tamañodelmensaje

Éstas variables son extraídas directamente del correo electrónico.

Se crearon las siguientes variables a partir de la persona que envía o recibe:

- País
- Externo /Interno
- Área

Esta información se obtuvo a través de Gestión Humana, quién posee un directorio de la empresa donde se acumula la identificación de cargo de cada persona que labora; si la persona no está contenida en este directorio se preguntó al directamente con el área de Gestión Humana o una persona del área de estudio.

Igualmente se transformaron variables a partir de la fecha contenida en la variable Recibido:

- Día del mes enviado o recibido (MONTHY1.12)
- Día de la semana enviado o recibido (MDAY1.31)
- Mes del año enviado o recibido (AÑO)
- Hora del día enviado o recibido (HOUR1.23)

Esta transformación se realizó de acuerdo con las fechas de envío o recepción de documentos. Lo anterior permitió la creación de variables para identificar día, mes u hora, con el fin de observar el comportamiento o las frecuencias de los mensajes. Estas variables se crearon en SPSS mediante funciones y desde Excel.

Para una mejor comprensión de los datos obtenidos, hubo otra división en dos tipos: variables: no textuales y variables textuales. A continuación se presenta el análisis de esta información.

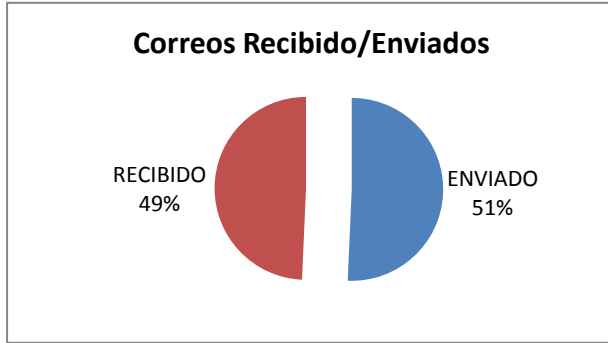
Es importante cuando se realice el análisis de las áreas, tener en cuenta que aparecerán correos personales, estos correos se identificarán porque no están en el directorio de Gestión Humana o en un directorio de proveedores. Para el estudio de las funciones no se utilizará esta información ya que éticamente se considera incorrecto aunque el correo sea laboral. Igualmente se debe tener un consentimiento informado de las personas dueñas de los archivos de correo. (Anexo 5).

### **7.2.3. Análisis de las variables no textuales (Caso 1):**

Se trabajaron 5 dimensiones: por los correos recibidos y enviados, por su origen, por el momento (tiempo) en que fue enviado y recibido, por su identificación si tiene archivos adjuntos y por el tamaño del correo, a continuación mostramos el detalle de cada uno.

#### **7.2.3.1. Correos Enviados y Recibidos**

Se analizaron los correos enviados y recibido por aparte para lograr un mejor entendimiento del correo. Hay una distribución equilibrada entre recibo y enviado. Ver Gráfica 4.



Gráfica 4. Frecuencia de correos enviados y recibidos Cargo 1

### 7.2.3.2. División por el origen de la persona: país, externo/interno (en relación a la organización) y área a la que pertenece

Se agruparon las personas por asunto, área y país hallando las relaciones más fuertes. Se revisaron las áreas claves que con más frecuencia se relaciona. Igualmente se encontraron frecuencias entre los correos internos (que pertenecen a la empresa) y correos externos, se identificó en el cargo un alto grado de relación con proveedores. Se identificaron claramente las áreas con las que se relaciona e intercambia información y la frecuencia con los países donde es notable la muy relevante la comunicación con Colombia. Ver Tabla 2.

**Frecuencia por país**

PAIS	Datos Recibido		Datos Enviado	
	Frecuencia.	%	Frecuencia.	%
COLOMBIA	10258	97,37%	10458	96,52%
VENEZUELA	80	0,76%	104	0,96%
PERU	63	0,60%	144	1,33%
PUERTO RICO	45	0,43%	65	0,60%
GUATEMALA	20	0,19%	10	0,09%
ARGENTINA	16	0,15%	15	0,14%
MEXICO	16	0,15%	14	0,13%
ECUADOR	15	0,14%	12	0,11%
CHILE	8	0,08%	10	0,09%
PANAMA	6	0,06%	0	0,00%
COSTA RICA	5	0,05%	0	0,00%
ALEMANIA	0	0,00%	3	0,03%
NO IDENTIFICADO	3	0,03%	0	0,00%
<b>Total general</b>	<b>10535</b>	<b>100,00%</b>	<b>10835</b>	<b>100,00%</b>

Tabla 2. Frecuencia de correos por país

### Distribución de correos externos e internos de recibidos y enviados

Hay una alta relación con externos por su relación con proveedores. Ver Tabla 3.

	Datos Recibidos		Datos Recibidos	
	Frecuencia.	%	Frecuencia.	%
INTERNO	8010	76,03%	7086	65,40%
EXTERNO	2525	23,97%	3749	34,60%
Total general	10535	100,00%	10835	100,00%

Tabla 3. Frecuencia de correos recibidos y enviados interno/externo

#### Distribución de correos recibidos y enviados por área

Entre las áreas de compras, proveedor, mercadeo, comunicación interna, CEDI (Centro Logístico), editorial y ventas está el 92% de los correos recibidos y el 90% de los correos enviados. Ver Tabla 4.

AREA	Recibidos		Enviados	
	Frecuencia	%	Frecuencia	%
COMPRAS	3960	38%	2875	26,53%
PROVEEDOR	2288	22%	3625	33,46%
MERCADEO	1149	11%	1085	10,01%
INFORMACIÓN INTERNA	730	7%	0	0,00%
CEDI	618	6%	1201	11,08%
EDITORIAL	599	6%	488	4,50%
VENTAS	237	2%	504	4,65%
PLANEACIÓN DE LA DEMANDA	226	2%	249	2,30%
PERSONAL	225	2,14%	199	1,84%
OPERACIONES	100	0,95%	85	0,78%
COSTOS	88	0,84%	62	0,57%
DISEÑO	78	0,74%	137	1,26%
EXPORTACIONES	77	0,73%	94	0,87%
PROMOCION	31	0,29%	86	0,79%
CREDITO	25	0,24%	0	0,00%
GESTION HUMANA	23	0,22%	8	0,07%
REGALIAS	14	0,13%	18	0,17%
ADMINISTRATIVO	14	0,13%	27	0,25%
FINANCIERO	14	0,13%	38	0,35%
PRODUCCION	14	0,13%	0	0,00%
TECNOLOGIA	11	0,10%	3	0,03%
CLIENTE	10	0,09%	3	0,03%
PRESIDENTE	4	0,04%	3	0,03%
CREDITO	0	0,00%	31	0,29%
VARIOS	0	0,00%	14	0,12%
Total general	10535	100,00%	10835	100,00%

Tabla 4. Frecuencia de correos recibidos y enviados por área

#### 7.2.3.3. División por el tiempo (hora, día, mes) o momento en que el correo es enviado o recibido

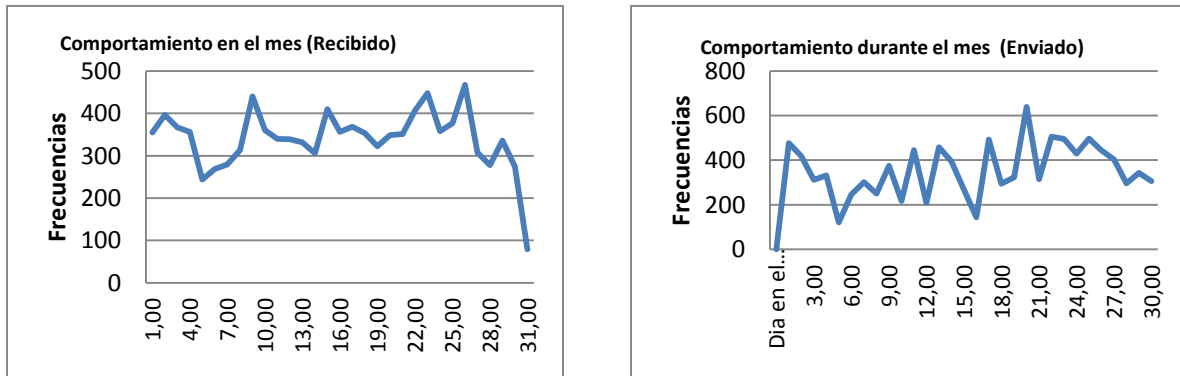
Se analizó la relación de las frecuencias Vs el tiempo. En este aspecto se crearon más variables de tiempo, se identificaron picos durante el día, la semana, el mes y el año. Dentro de este aspecto de la investigación, se



determinaron frecuencias y promedios, se revisaron los correos electrónicos que fueron enviados o recibidos en tiempos no usuales de horario de trabajo y se buscaron los destinatarios de estos correos electrónicos. Igualmente se identificaron la frecuencia diaria de correos electrónicos y los horarios más congestionados para recibir y enviar correos electrónicos (tráfico de e-mails). Por último, pudimos establecer una tendencia por mes y por año.

### Comportamiento durante el mes

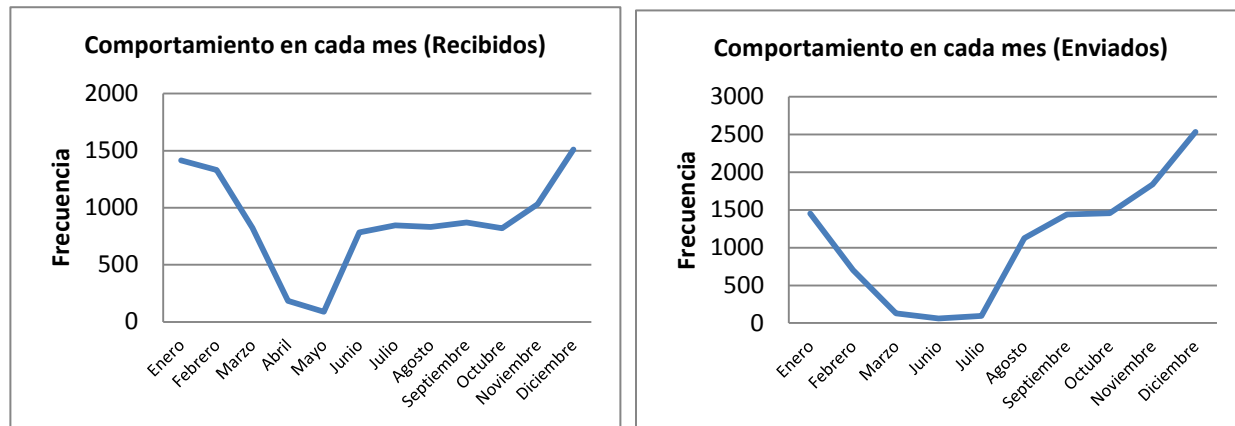
Existencia de picos hacia el día 10 y 21 del mes. Ver Gráfica 5.



Gráfica 5. Frecuencia de correos recibidos y enviados en el mes

### Comportamiento en cada mes

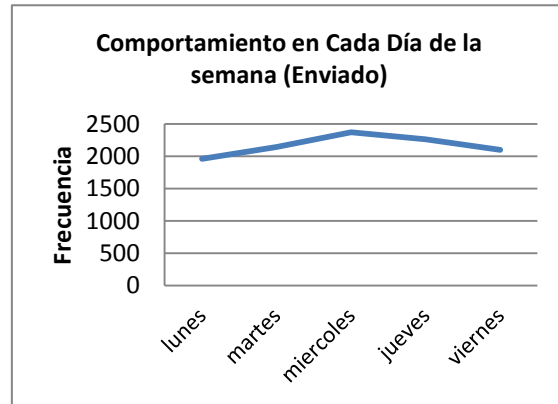
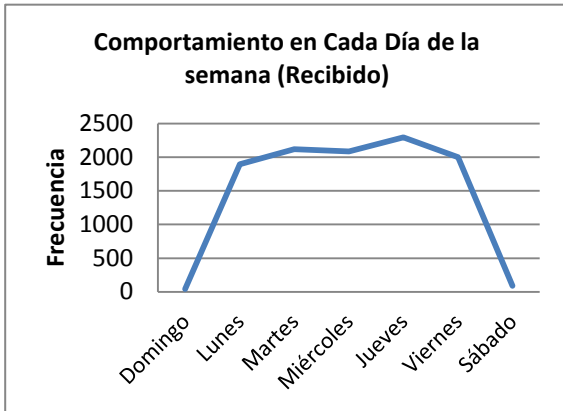
Las gráficas muestran un tráfico alto en los meses de enero a febrero y de Octubre a Diciembre y un valle en los meses de abril a mayo. Precisamente correspondientes a la temporalidad de los productos que compra. Ver Gráfica 6.



Gráfica 6. Frecuencia de correos recibidos y enviados por cada mes

### Comportamiento en cada día de la semana

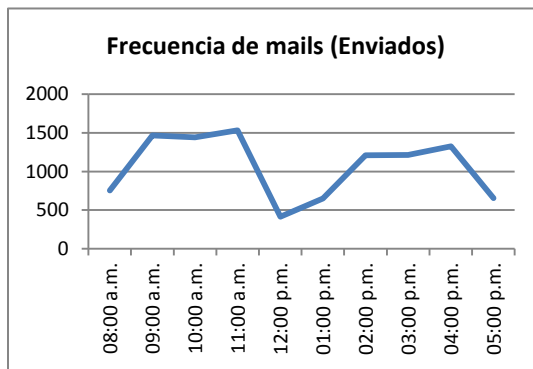
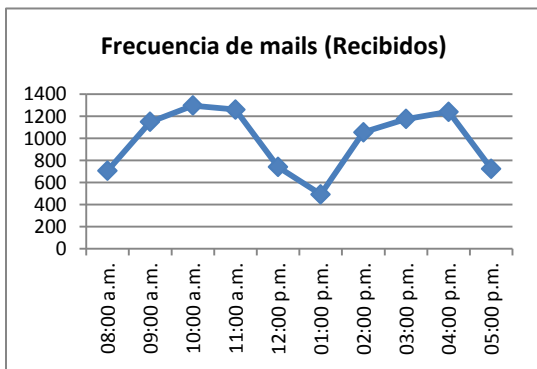
Las gráficas muestran una frecuencia alta de lunes a viernes durante la jornada laboral. Ver Gráfica 7.



Gráfica 7. Frecuencia de correos recibidos y enviados por cada día de la semana

#### Comportamiento por hora

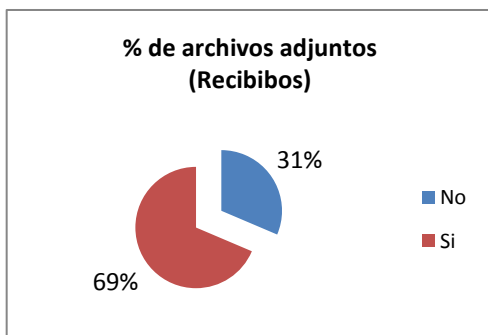
Las gráficas muestran una frecuencia alta durante las 10 y 11 am y nuevamente de 2 a 4 pm. Ver Gráfica 8.



Gráfica 8. Frecuencia de correos recibidos y enviados por hora de la semana

#### 7. 2.3.4. División referente sobre si contiene en el correo adjunto

En este momento se analizaron cuántos correos tienen un archivo adjunto. Esto nos permitió identificar si además de la información escrita en el correo se requerían archivos adicionales para comprender los datos. La mayoría de veces contiene correos adjuntos para que el destinatario analice. El envío de archivos gráficos y de archivos en Excel a los proveedores hace que el 95% de los correos enviados contengan adjuntos. Ver Gráfica 9.



Gráfica 9. Frecuencia de correos recibidos y enviados por hora de la semana

#### 7.2.3.5. División por el tamaño del correo

La tabla que se muestra a continuación presenta el tamaño del correo en recibidos y enviados. Ver Tabla 5.

Tamaño archivo (MB)	Tamaño de los correos Recibidos		Tamaño de los correos Enviados	
	Frecuencia	%	Frecuencia	%
0-0,95	9729	92,35%	10271	94,79%
0,95-1,9	265	2,52%	172	1,59%
1,90-2,86	139	1,32%	103	0,95%
2,86-3,81	74	0,70%	51	0,47%
3,81-4,76	112	1,06%	117	1,08%
4,76-5,72	95	0,90%	53	0,49%
5,72-6,67	11	0,10%	5	0,05%
6,67-7,62	26	0,25%	7	0,06%
7,62-8,58	21	0,20%	20	0,18%
8,58-9,53	26	0,25%	12	0,11%
9,53-10,4	10	0,09%	0	0,00%
10,4-11,4	4	0,04%	7	0,06%
11,4-12,3	5	0,05%	0	0,00%
12,3-13,3	2	0,02%	4	0,04%
13,3-14,3	2	0,02%	0	0,00%
14,3-15,2	5	0,05%	0	0,00%
15,2-16,2	4	0,04%	0	0,00%
16,2-17,1	1	0,01%	0	0,00%
17,1-18,1	1	0,01%	0	0,00%
19,0-20	1	0,01%	0	0,00%
24,7-25,7	2	0,02%	0	0,00%
25,7-26,7	0	0,00%	13	0,12%
Total general	10535	100,00%	10835	100,00%

Tabla 5. **Tamaño de los correos Recibidos y enviados**

La gráfica 14 nos muestra que en un 95% aproximadamente los correos enviados o recibidos son de 1MB. La mayoría de los adjuntos son correos en Excel.

### 7.2.3.6. Mapa Relacional cargo 1

Basados en esas cinco dimensiones o divisiones, se efectuaron análisis univariados y bivariados que permitieran identificar patrones y crear el mapa relacional

A partir del análisis de las variables no textuales establecidas para los correos electrónicos, para el cargo de estudio podemos concluir lo siguiente:

**NOMBRE: A<sup>4</sup>**

**CARGO : Coordinadora de Compras de impresión**

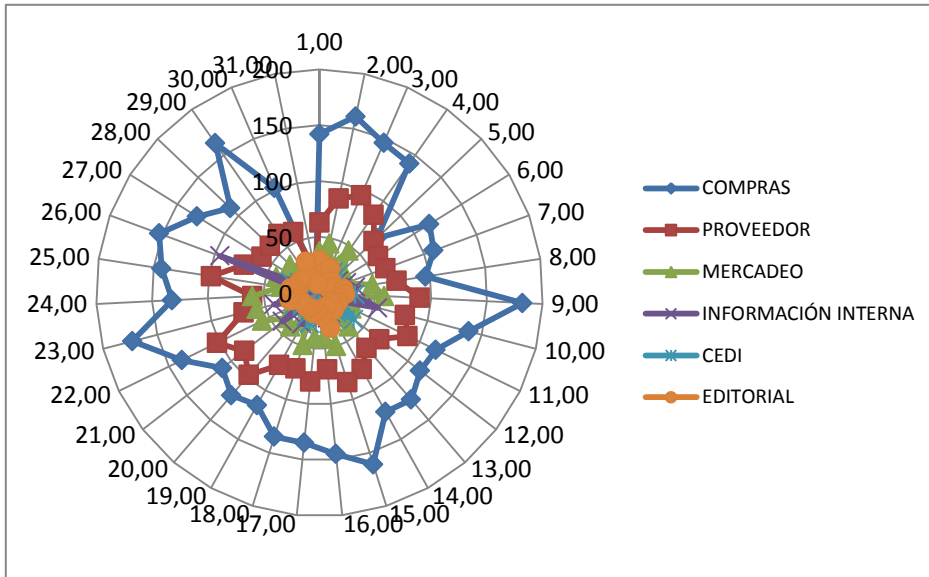
Según sus correos podemos concluir que:

1. Las personas con las que más se relaciona son: las dos Auxiliares de Compras y las otras Coordinadora de Compras (31% de las comunicaciones).
2. Recibió 10535 correos (38 mails diarios) y envió 10835 correos (40 correos diarios)
2. El 97% de las comunicaciones las recibe de Colombia
3. Las áreas de las que más recibe información es: Compras (38%), Proveedores (22%), Mercadeo (11%), Información Interna (7%), Centro Logístico (6%), Editorial (6%), Ventas (2%) y Planeación de la Demanda

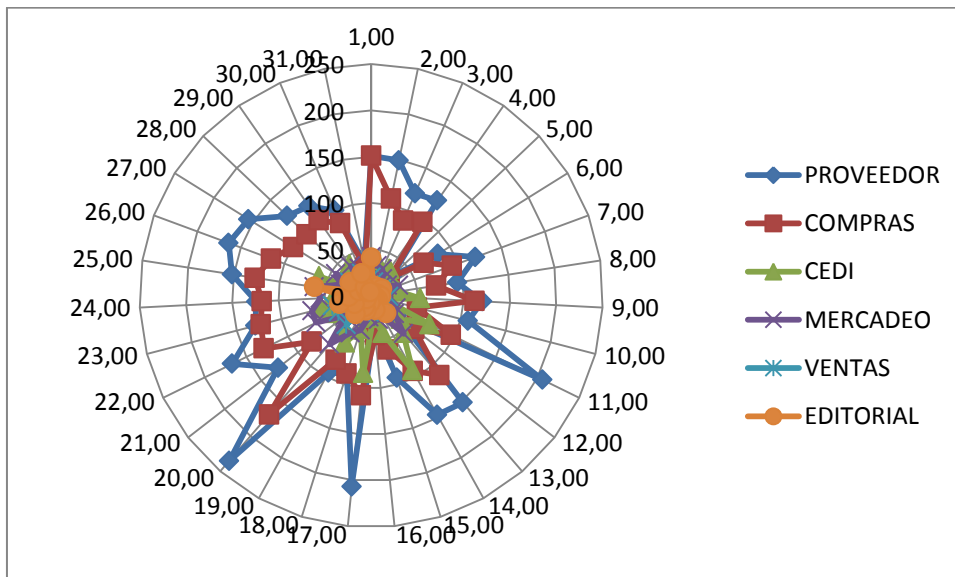
<sup>4</sup> No se coloca el nombre propio de la persona por motivos de confidencialidad

- (2%). En donde el 70% de los correos serán internos (de personas de la empresa)
4. Recibirá la mayoría de sus correos entre las 10 a 11 am y entre las 3 a 5 pm
  5. Debe ser cuidadoso en la lectura de los mail porque de 3 mails al menos 2 tendrán adjuntos
  6. En el 95% los correos enviados deben tener adjuntos, razón por la cual debe tener un correo de buena capacidad
  7. Los miércoles y los jueves son los días con más alto volumen de correos

**Mapa Relacional** durante el mes .Ver Gráfica 10 y 11 de los correos enviados y recibidos.



Gráfica 10. Correos recibidos del cargo con otras áreas durante un mes (en una frecuencia diaria).



Gráfica 11. Correos enviados del cargo con otras áreas durante un mes (en una frecuencia diaria).

#### 7.2.4. Análisis de las variables relacionadas con el texto (variable contenidos):

En cuanto al contenido del correo se realizaron las siguientes actividades:

- Limpieza de reenvíos. En este aspecto se dejó solo el último correo. Los avisos de seguridad que aparecen como mensajes finales se eliminaron por considerar que no aportan información relevante en la investigación.
- Eliminación de stop words: Se excluyeron adjetivos, conjunciones, artículos, adverbios y preposiciones del contenido del correo,, ya que únicamente se analizaron las parejas sustantivo-verbo.

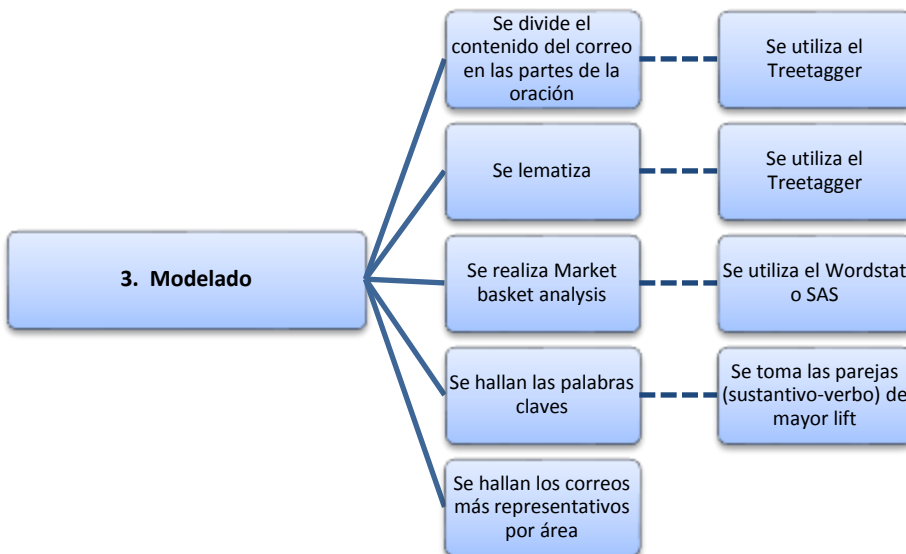
El objetivo de esta etapa es determinar las relaciones sustantivo y verbo en el contenido del correo. A partir de ello, pudimos como base para empezar a delimitar las funciones básicas de un trabajador del conocimiento KW en la empresa estudiada.

### 7.3. Modelado

En esta segunda etapa se enfoca el TM.

Debemos mencionar que en la investigación se ensayaron varios software de Minería de textos, como Gate y Weka, y Statword y SAS. Sin embargo, el uso de Gate y Weka requería de un conocimiento avanzado en la programación de la herramienta. Mientras que el software Statword y SAS cuenta con herramientas internas ya programadas. Este era un software del que ya existía un conocimiento de la herramienta.

A continuación se describe en un diagrama de flujo cada una de las actividades que se programó. Ver Gráfica 12.



Gráfica 12. **Flujograma de análisis del contenido de los correos**

En el Treetagger, herramienta creada por Schmid, Helmut (1994 y 1995), se exportan e importan los correos, para realizar la división de las partes de una oración (Part of the speech). Como ya lo hemos enunciado antes, en la gramática tradicional se clasifica las palabras basándose en ocho partes de la oración: el verbo, el sustantivo, el pronombre, el adjetivo, el adverbio, la preposición, la conjunción y la interjección. El Treetagger etiqueta cada palabra del texto con la parte del lenguaje a la que pertenece.

En la siguiente tabla se muestra la parte de la oración que identifica el Tree Tagger. Para nuestro estudio solo utilizamos lo correspondiente a: NC (Nombre común), NP (Nombre Propio), VCLinf (Verbo en infinitivo), Vldj y Vlfm (otros verbos), Vlger (Verbo en gerundio); y se obviaron los verbos ser, estar, deber, poder o haber porque son genéricos y nosotros estábamos interesados en observar el nivel de detalle de la acción. Ver Tabla 6.

Parte de la oración	Observaciones
ACRNM	acronyms, initialism & abbreviations dictionary.
ADJ	adjetivos
ADV	adverbios
ART	artículos
CARD	números
CC	conjunciones (copulativa)
CCAD	conjunciones adversativas o de contraposición
CCNEG	conjunciones (negativas)
CODE	número
CQUE	pronombre relativo
CSUBF	adverbio de modo
CSUBI	conjunción
CSUBX	conjunción
DM	adjetivo determinado
INT	interrogación
NC	nombre propio, varios no identificados, varios letra con número
NEG	negativo
NMEA	magnitud
NMON	nombre relativos a tiempo
NP	nombres propios
ORD	ordinales
PAL	sufijo
PDEL	preposición
PE	en otro idioma
PPC	artículo
PPO	posesivo
PPX	pronombre
PREP	preposición
QU	pronombre determinado
REL	pregunta
VCLinf	verbos en infinitivo
VEfm	ser o estar
VHfm	haber
VHger	haber
VHinf	haber
VLadj	otros verbos
VLfm	otros verbos
VLger	verbos en gerundio
VMadj	otros verbos
VMfm	deber o poder
VMinf	deber o poder
VSadj	ser
VSfm	ser o estar
VSger	ser en gerundio

Tabla 6. Partes de la Oración identificadas en el Tree Tagger

Es importante aclarar que Tree Tagger realiza las siguientes acciones: etiqueta el *part of speech* utilizando cadenas de Markov, busca si las palabras son independientes unas de otras (la identidad de una palabra

sólo depende de su etiqueta) y realiza un árbol para detallar y hallar la raíz de la palabra. Schmid, Helmut (1994 y 1995). El Tree Tagger tiene un 96% de exactitud en clasificar correctamente una palabra

Después se realiza la lematización. Esta raíz es la forma que por convenio se acepta como representante de todas las formas flexionadas de una misma palabra, Durante la importación se mantiene el número del caso o id, que corresponde al número del correo de tal forma que podamos identificar el área a la que pertenece.

En caso de que la empresa utilice dos idiomas, herramientas como Tree Tagger o Wordstat cuentan con diccionarios que incluyen uno o varios idiomas. Lo importante es incluirlos antes de hacer la separación de la oración y de la lematización.

Una vez realizado el filtro de Tree Tagger, utilizamos *Collaborative Filtering*. Una de las herramientas del *Collaborative Filtering*, es el *Market Basket Analysis*. Se obtuvieron las frecuencias por cada palabra en cada correo y se determinó una asociación entre las palabras utilizando *Market Basket Analysis*. Cabe anotar que el proceso se realizó para pares (verbo y sustantivo). Para ello primero se determinaron las frecuencias totales en que aparecen las palabras sobre los casos. Segundo se hizo un *support*<sup>5</sup> que establece la medida de la frecuencia de la colección de elementos de una asociación cuando ocurren juntos como un porcentaje de todos los casos o correos electrónicos.

Entre los pasos para hallar la información que requerimos estaba elaborar una fórmula que nos permitiera acceder a la información. La fórmula usada fue:

$$\text{Support } (X \rightarrow Y) = P(XUY) = n(XUY)/N$$

El support identifica si las palabras se presentan conjuntamente en un buen porcentaje de los correos electrónicos.

El siguiente paso fue hallar el *confident*<sup>6</sup>: la regla "X dado Y" es una medida de cuánto mayor es la probabilidad de que B se produzca cuando se ha producido A. Si el valor es de 100%, Lo anterior significa que B siempre se produce si A ha ocurrido. Su fórmula es:

$$\text{Confidence } (X \rightarrow Y) = P(X|Y) = n(XUY)/n(Y)$$

Además, se halla una medida adicional, el *Lift*<sup>7</sup>. Este revisa si una regla es mucho mejor que lo esperado, es decir, para este caso si la pareja de palabras se presenta más de lo esperado en diferentes correos frente a un par de palabras cualesquiera. El indicador que arroja nos da la relación de los registros que soportan toda la regla para el número que se esperaba, asumiendo que no hay relación entre las palabras X. En términos de fórmula, el Lift se expresa:

$$\text{Lift } (X \rightarrow Y) = \text{Confidence}/P(Y)$$

Cuando el Lift es mayor que 1, se puede decir que las palabras se presentan juntas más frecuentemente de lo esperado. I. En cambio, si es menor que 1, las palabras se presentan juntas en menor frecuencia de lo esperado y la regla tiende a ser un instrumento negativo que se basa en adivinar los resultados. Es decir, no logra ser precisa.

---

<sup>5</sup> Se incluye este término en inglés porque se considera que su traducción en español no contiene toda la definición necesaria.

<sup>6</sup> Se incluye este término en inglés porque se considera que su traducción en español no contiene toda la definición necesaria.

<sup>7</sup> Se incluye este término en inglés porque se considera que su traducción en español no contiene toda la definición necesaria.

Tomamos un *lift* mayor que 1 y organizamos la información de los *support* de mayor a menor para garantizar mejores resultados.

Al determinar los verbos y las palabras claves, obtuvimos un grupo de palabras importantes para determinar a partir de allí las funciones del trabajador del conocimiento. Este análisis lo realizamos por área y desde los correos enviados y recibidos. Encontramos que el 80% de las funciones es obtenido del área a la que pertenece el trabajador en este caso en particular.

A continuación presentamos algunos ejemplos de la información obtenida sobre las palabras claves obtenidas y la función. Ver Tablas 7 y 8.

ROOT	root2	support	confidence	expected	lift
cal	calentar	0,016336	0,552941	0,017379	31,81624
calentar	cal	0,016336	0,94	0,029545	31,81624
calentar	cartulina	0,016336	0,94	0,029892	31,44628
cartulina	calentar	0,016336	0,546512	0,017379	31,44628
calentar	solapa	0,016336	0,94	0,030935	30,38629
solapa	calentar	0,016336	0,52809	0,017379	30,38629
calentar	tinta	0,013903	0,8	0,027112	29,50769
tinta	calentar	0,013903	0,512821	0,017379	29,50769
acabar	solapa	0,030587	0,642336	0,030935	20,76405
solapa	acabar	0,030587	0,988764	0,047619	20,76405
acabar	cartulina	0,028849	0,605839	0,029892	20,26744
cartulina	acabar	0,028849	0,965116	0,047619	20,26744
acabar	cal	0,028502	0,59854	0,029545	20,25882
cal	acabar	0,028502	0,964706	0,047619	20,25882
acabar	hilo	0,01008	0,211679	0,011818	17,91177
hilo	acabar	0,01008	0,852941	0,047619	17,91177
calentar	formato	0,017379	1	0,056656	17,65031
formato	calentar	0,017379	0,306748	0,017379	17,65031
acabar	tinta	0,022245	0,467153	0,027112	17,23077
tinta	acabar	0,022245	0,820513	0,047619	17,23077
calentar	papel	0,017379	1	0,060132	16,63006
papel	calentar	0,017379	0,289017	0,017379	16,63006

**Función obtenida:** Confirmar y enviar especificaciones de libros

Tabla 7. **Palabras relacionadas y función generada**

ROOT	root2	support	confidence	lift
adjudicar	especificación	0,017379	0,625	11,60081

Función: **Adjudicar** libros con ciertas **especificaciones**

Tabla 8. **Palabras relacionadas y función generada**



Al analizar las parejas de palabras, se encontró que había dos razones para no utilizar el MBA como producto final para deducir las funciones:

- Primero la existencia de verbos como agradecer, tener, ir y, saludar dificultaba el análisis porque no proporcionan por sí solas información detallada de la acción.
- Segundo, se halló que se requería que la persona tuviera conocimiento del cargo para deducir sus funciones a partir del MBA, esto no es posible muchas veces por tiempo o porque la persona ya no se encuentra desempeñando ese cargo. Esto es visible en la gráfica 19, donde se requiere saber que las palabras solapa, tinta, papel, e hilo con el verbo acabar se refieren a la elaboración de un libro.

Se realizó entonces un conteo de las parejas que obtuvieron mejor lift en los correos. Finalmente, se decidió hacer un algoritmo en SAS que contara cuántas parejas, de las extraídas por el MBA y que tuvieran un alto lift, tenía cada correo electrónico. De ese modo se obtuvieron los correos más representativos o densos (aquellos que contienen más parejas de palabras que aparecen juntas en un correo). Después se realizó una lectura total de dichos correos en casos específicos para inferir a partir de un número pequeño de correos por área, las funciones del trabajador del conocimiento, en la validación la persona que esté leyendo la información de los títulos y de los contenidos de los correos no conoce de forma cercana el cargo del estudio.

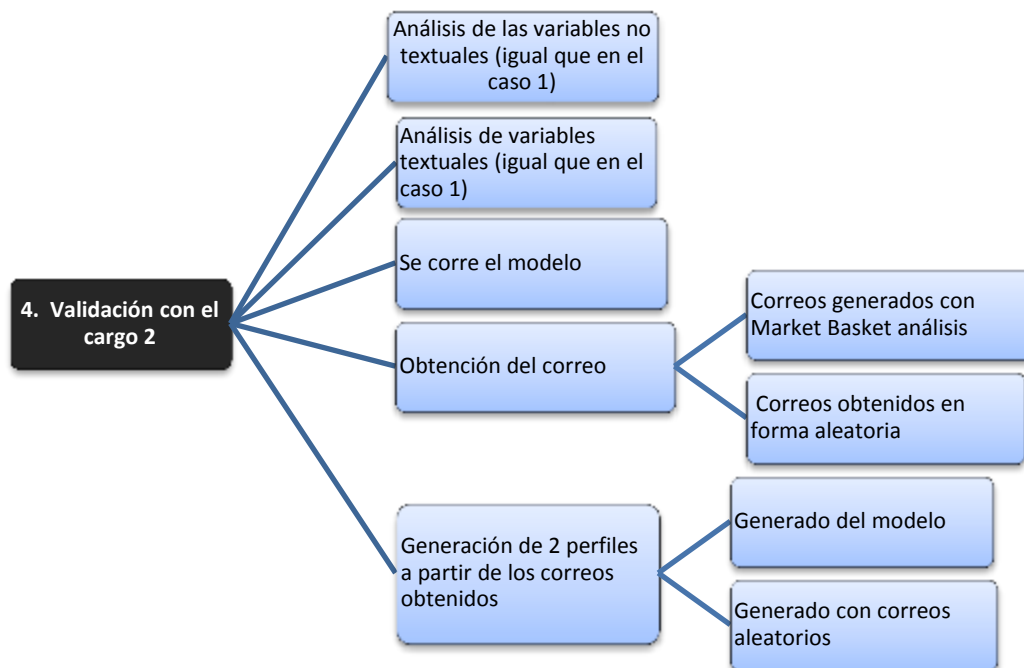
Tomando los correos se obtuvieron mejores y más claros resultados. A continuación damos el resultado por área con sus funciones correspondientes desarrolladas por el investigador:

- Envío de especificaciones a los impresores
- Cotización de libros.
- Negociación con los proveedores (Plazo, precio, entrega, flete, estándar de calidad, apoyo logístico de almacenamiento y cantidad de producción de acuerdo a la necesidad) con los proveedores seleccionados.
- Envía órdenes de compra
- Coordina con el proveedor, entrega de la mercancía y valida las especificaciones definitivas de impresión editorial de los libros
- Priorizar la programación de producción de impresiones editoriales

Igualmente después de esta revisión se considera necesario obtener no solo el contenido del correo electrónico sino relacionar Título o el asunto del correo electrónico para tener mejores resultados.

#### **7.4. Validación con un segundo correo Cargo 2**

Durante la evaluación se tomó un segundo correo con las mismas bajo los mismos parámetros y pasos descritos anteriormente que nos sirven de validación, precisamente para asegurarnos que la metodología funciona . Ver Gráfica 13.



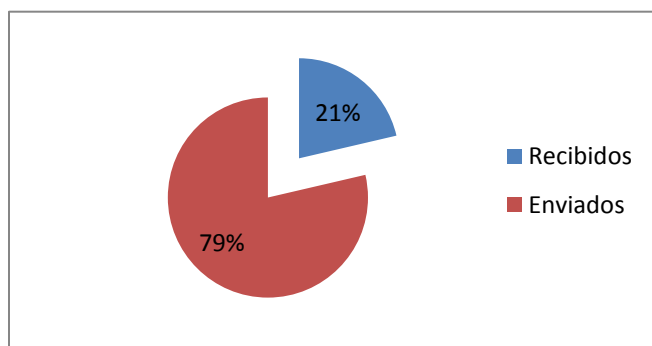
Gráfica 13. Pasos de la validación utilizando el Cargo 2.

#### 7.4.1. Análisis de las variables que no textuales:

Se trabajaron las mismas 5 dimensiones:

##### 7.4.1.1. Correos Enviados y Recibidos

Los correos enviados y recibidos no tienen la misma distribución, son casi cuatro veces más los correos que se envían que los que se recibe. Debe ser un área que conecta con otras áreas en forma neurálgica. Ver Gráfica 14.



Gráfica 14. Frecuencia de los correos enviado y recibidos correo Cargo 2

##### 7.4.1.2 División por el origen de la persona: país, externo/interno (en relación a la organización) y área a la que pertenece

### Frecuencia por país

La mayor relación la tiene en Colombia pero entre los países distintos son México y Argentina con los que mantiene una mayor comunicación. Ver tabla 9.

PAIS	Datos de enviado		Datos Recibido	
	Frecuencia	%	Frecuencia	%
COLOMBIA	1056	91,91%	3714	87,80%
MEXICO	30	2,61%	129	3,00%
ARGENTINA	26	2,26%	136	3,20%
PUERTO RICO	12	1,04%	16	0,40%
PERU	11	0,96%	162	3,80%
ECUADOR	10	0,87%	15	0,40%
CHILE	3	0,0026	44	1,00%
PANAMA	1	0,0009	0	0,00%
GUATEMALA	0	0,00%	10	0,20%
CHINA	0	0,00%	3	0,10%
VENEZUELA	0	0,00%	2	0,00%
<b>Total general</b>	<b>1149</b>	<b>100,00%</b>	<b>4231</b>	<b>100,00%</b>

Tabla 9. Frecuencia de correos recibidos y enviados por país Cargo 2

### Distribución de correos externos e internos de recibidos y enviados

La mayoría de los correos que maneja son internos. Tiene muy poca relación con proveedores o clientes (externos). Ver Tabla 10.

INTERNO/EXTERNO	Datos de enviado		Datos Recibido	
	Frecuencia	%	Frecuencia	%
INTERNO	1113	96,87%	3780	89,30%
EXTERNO	36	3,13%	451	10,70%
<b>Total general</b>	<b>1149</b>	<b>100,00%</b>	<b>4231</b>	<b>100,00%</b>

Tabla 10. Frecuencia de correos recibidos y enviados interno/externo  
Distribución de correos recibidos y enviados por área

Entre un 70% a un 80% está la relación con Editorial, Costos y Diseño. Ver Tabla 11.

AREA	Datos enviados		Datos Recibidos	
	Frecuencia	%	Frecuencia	%
EDITORIAL	729	63,45%	2511	59,30%
COSTOS	114	9,92%	150	3,50%
DISEÑO	95	8,27%	280	6,60%
COMPRAS	41	3,57%	103	2,40%
MERCADEO	38	3,31%	212	5,00%
FINANCIERO	36	3,13%	76	1,80%

PERSONAL	35	3,05%	457	10,80%
VENTAS	18	1,57%	119	2,80%
GESTION HUMANA	12	1,04%	61	1,40%
REGALIAS*	11	0,96%	94	2,20%
PROVEEDOR	5	0,44%	48	1,10%
CEDI	4	0,35%	12	0,30%
VARIOS	3	0,26%	0	0,00%
INNOVACIÓN	2	0,17%	16	0,40%
OPERACIONES	2	0,17%	3	0,10%
PLANEACIÓN DE LA DEMANDA	2	0,17%	7	0,20%
ADMINISTRATIVO	1	0,09%	16	0,40%
PROMOCIÓN	1	0,09%	23	0,50%
PRODUCCION	0	0,00%	19	0,40%
VARIOS	0	0,00%	1	0,00%
OPTIMIZACION	0	0,00%	15	0,40%
CALIDAD	0	0,00%	8	0,20%
<b>Total general</b>	<b>1149</b>	<b>100,00%</b>	<b>4231</b>	<b>100,00%</b>

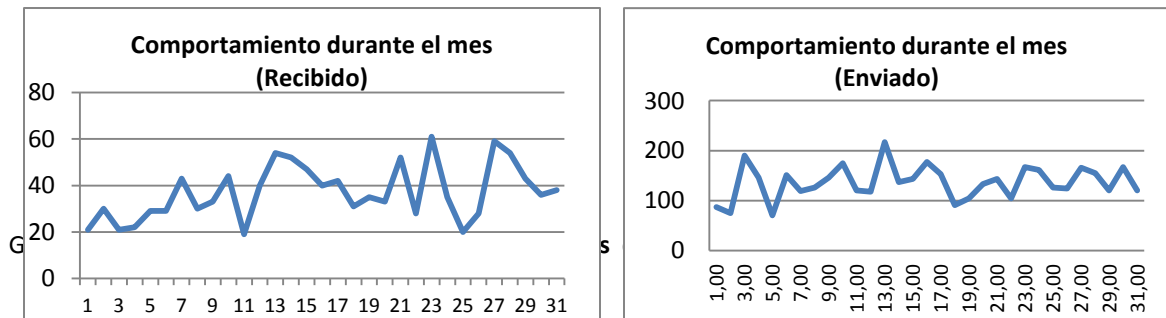
Tabla 11. Frecuencia de correos recibidos por área

\*(Aunque Regalías está contenida dentro de costos, se separó porque regalías es un área muy específica)

#### 7.4.1.3. División por el tiempo (hora, día, mes) o momento en que el correo es enviado o recibido

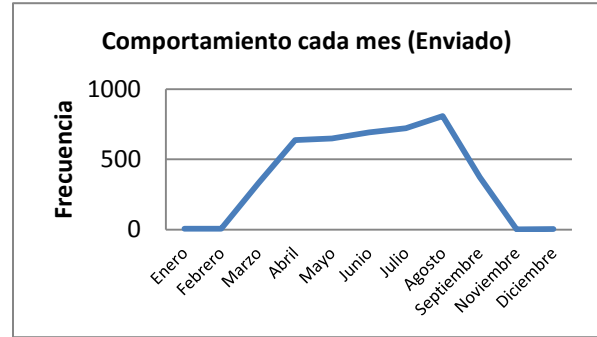
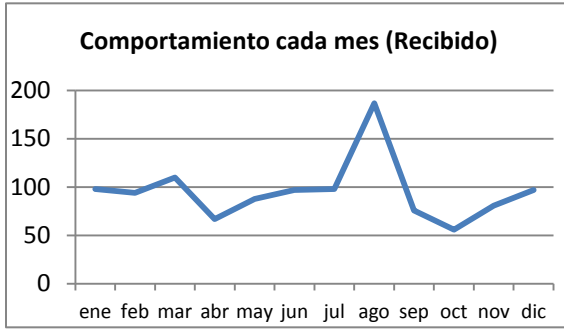
##### Comportamiento durante el mes

No se ve un comportamiento marcado durante el mes, unos picos de correos recibidos a finales del mes y un pico no tan sobresaliente hacia el 13 y el 15 del mes. Ver Gráfica 15.



##### Comportamiento en cada mes

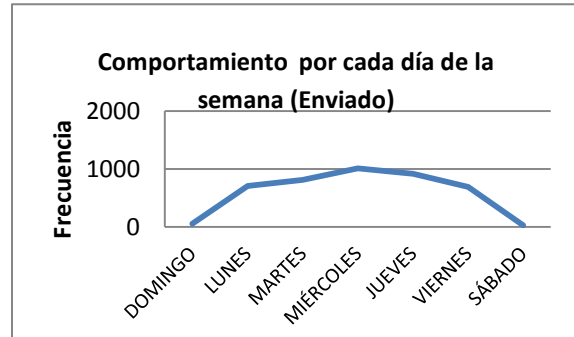
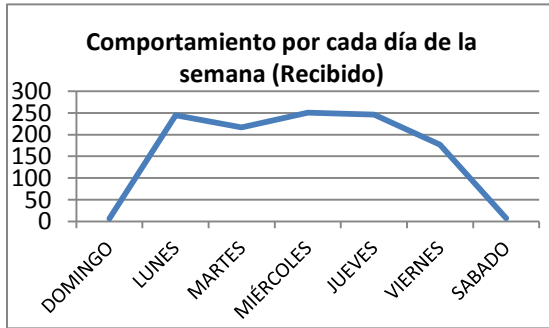
Hay un pico alto de correos recibidos entre agosto y septiembre por finalización de proyectos y en los enviados hay un valle en diciembre enero y febrero, época en que editorial hasta ahora comienza nuevamente sus proyectos. Ver Gráfica 16.



Gráfica 16. Frecuencia de correos recibidos y enviados por cada mes

### Comportamiento en cada día de la semana

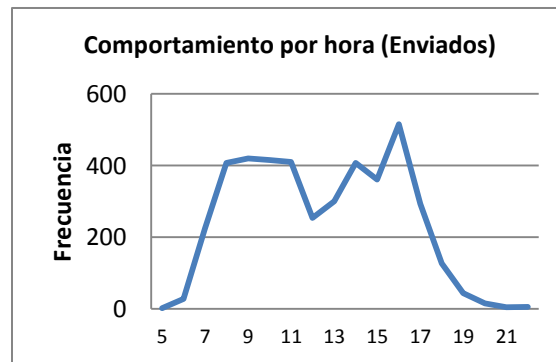
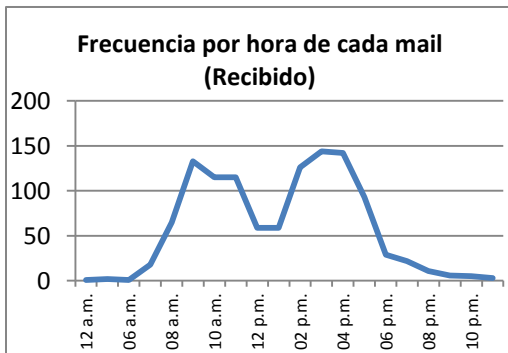
Trabaja de Lunes a Viernes. Ver Gráfica 17.



Gráfica 17. Frecuencia de correos recibidos y enviados por cada día de la semana

### Comportamiento por hora

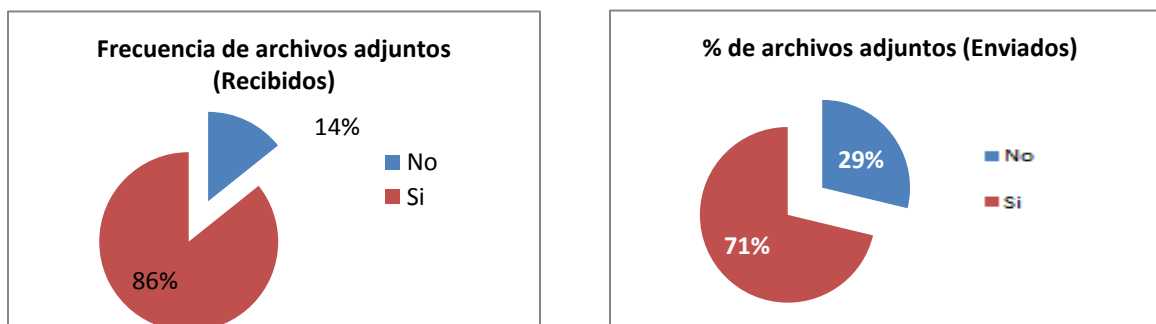
Hay picos de 8 a.m. a 10 a.m. y de 2 p.m. a 4 p.m. en los correos recibidos y en los enviados de 7 a.m. a 11 a.m., con un gran pico a la 4 p.m. en los enviados. Ver Gráfica 18.



Gráfica 18. Frecuencia de correos recibidos y enviados por hora de la semana

#### 7.4.1.4. División referente sobre si contiene en el correo archivos adjuntos

Existe un gran porcentaje de archivos adjuntos en los correos enviados y recibidos. El KW debe tener cuidado al leerlos para no perder esta información, para este estudio no se analizan los archivos adjuntos. Ver Gráfica 19.



Gráfica 19. Frecuencia de correos recibidos y enviados por hora de la semana

#### 7.4.1.5. División por el tamaño del correo

A continuación las tablas muestran el tamaño del correo en recibidos y enviados. Ver Tabla 12.

Tamaño del mensaje	Recibidos		Enviados	
	Frecuencia Recibido	%	Frecuencia.	%
0,00-0,96	1027	89,38%	3939	93,08%
0,96-1,91	51	4,44%	104	2,46%
1,91-2,87	14	1,22%	51	1,21%
2,87-3,82	5	0,44%	67	1,58%
3,82-4,77	6	0,52%	10	0,24%
4,77-5,73	15	1,31%	3	0,07%
5,73-6,68	3	0,26%	9	0,21%
6,68-7,63	11	0,96%	7	0,17%
7,63-8,59	7	0,61%	5	0,12%
8,59-9,54	2	0,17%	5	0,12%
10,49-11,45	5	0,44%	10	0,24%
11,45-12,40	1	0,09%	1	0,02%
12,40-13,36	0	0,00%	13	0,31%
13,36-14,31	0	0,00%	2	0,05%
14,31-15,26	0	0,00%	2	0,05%
15,26-16,22	2	0,17%	1	0,02%
16,22-17,17	0	0,00%	1	0,02%
24,80-25,75	0	0,00%	2	0,05%
<b>Total general</b>	<b>1149</b>	<b>100,00%</b>	<b>4232</b>	<b>100,00%</b>

Tabla 12. Tamaño de los correos Recibidos y enviados

La tabla nos muestra que en un 96% aproximadamente los correos enviados o recibidos son de 1MB. La mayoría de los adjuntos son archivos en Excel.

#### 7.4.1.5. Mapa Relacional cargo 2

Basados en esas cinco dimensiones o divisiones, se efectuaron análisis univariados y bivariados que permitieran identificar patrones y crear el mapa relacional.

A partir del análisis de las variables no textuales establecidas para los correos electrónicos, para el cargo 2 de estudio pudimos concluir lo siguiente:

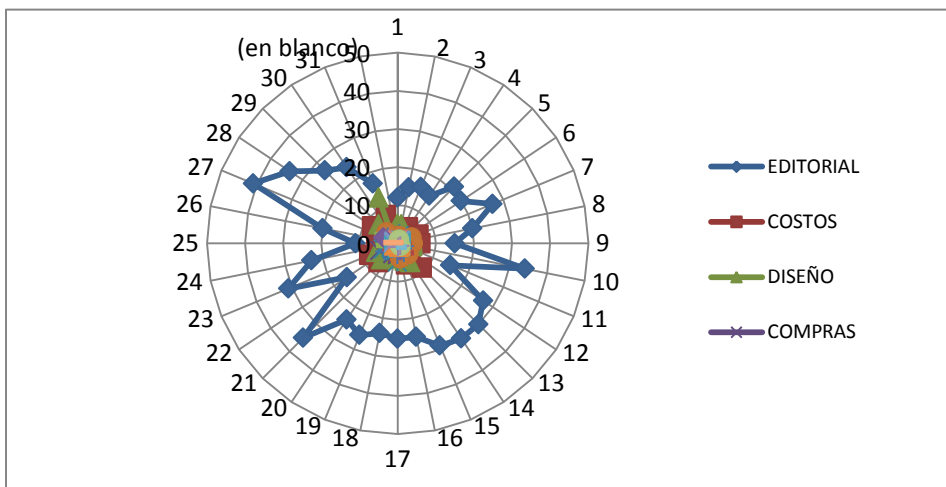
**NOMBRE:** B<sup>8</sup>

**CARGO :** Coordinadora del Proceso Editorial

Según sus correos podemos concluir que:

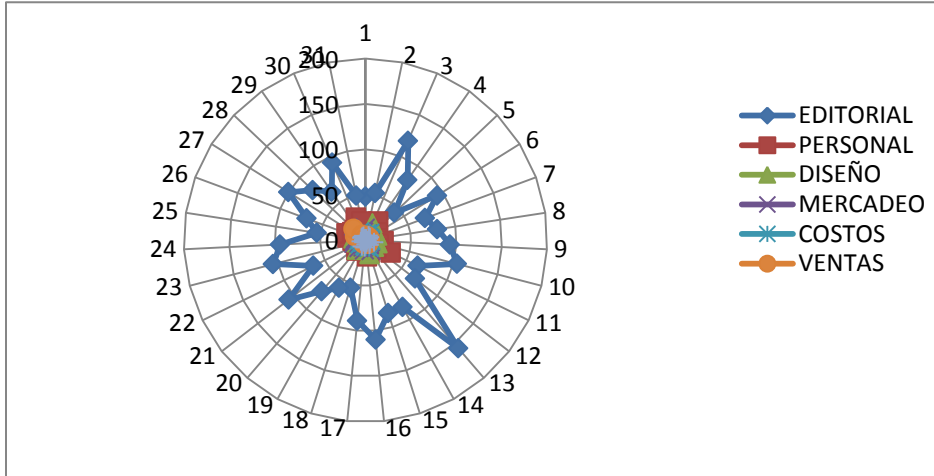
1. Recibió 1149 correos (4 mails diarios) y envió 4231 correos (15 correos diarios)
2. El 92% de las comunicaciones las recibe de Colombia. Y el 88% son correos enviados a personas de Colombia
3. Las áreas de las que más recibe información es: Editorial (63%), Costos (10%), Diseño (8%). Y envía a Editorial (59%), Personal (11%); Diseño (7%), Mercadeo (5%) y Costos (3,5%)
4. Recibirá la mayoría de sus correos entre las 8 a 11 am y entre las 3 a 5 pm
5. Debe ser cuidadoso en la lectura de los mail porque de 4 mails al menos 3 tendrán adjuntos
6. En el 71% los correos enviados deben tener adjuntos, razón por la cual debe tener un correo de buena capacidad
7. Los martes, miércoles y jueves son los días con más alto volumen de correos

**Mapa Relacional** durante el mes. Ver Gráficas 20 y 21.



Gráfica 20. Correos recibidos del cargo con otras áreas durante un mes (en una frecuencia diaria).

<sup>8</sup> No se coloca el nombre propio de la persona por motivos de confidencialidad



Gráfica 21. Correos enviados del cargo con otras áreas durante un mes (en una frecuencia diaria).

#### 7.4.2. Análisis para las variables textuales

Revisamos primero una muestra del MBA obtenido y encontramos algunas relaciones verbo/sustantivo interesantes . A continuación se muestra una sección de las relaciones obtenidas. Ver Tabla 13.

RAIZ 1	RAIZ 2	support	confidence	lift
comité	presentar	0.0139033716	0.7547169811	33.926.886.792
actualizar	proyección	0.0118178658	0.5573770492	32.725.995.316
editorial	presentar	0.0104275287	0.4411764706	19.832.261.029
comprar	editorial	0.0222453945	0.4383561644	18.546.333.602
costo	cerrar	0.010775113	0.1901840491	17.098.734.663
comprar	orden	0.0128606187	0.2534246575	16.570.516.812
deber	proyección	0.0100799444	0.2660550459	15.621.231.979
espacio	definir	0.0118178658	0.3695652174	15.189.130.435
actualizar	archivo	0.0135557873	0.6393442623	13.830.025.884
esperar	entrega	0.0128606187	0.4868421053	12.733.133.971
cronograma	esperar	0.0104275287	0.3296703297	12.479.757.085
pago	realizar	0.015641293	12.162.162.162	11.397.570.209
cierre	realizar	0.0212026416	11.730.769.231	10.993.297.419
cargar	hoja	0.0100799444	0.3052631579	10.710.269.576
desarrollar	proyecto	0.0111226973	0.6808510638	10.531.228.552
costo	presupuestar	0.0111226973	0.1963190184	10.459.441.036
ejercicio	realizar	0.0173792145	10.869.565.217	10.186.234.244
fecha	desarrollar	0.0111226973	0.1658031088	10.149.266.894
ejecución	encontrar	0.010775113	0.3780487805	89.888.127.394
hoja	realizar	0.0257212374	0.9024390244	84.570.588.703
realizar	desarrollo	0.0142509559	0.1335504886	83.527.120.804
proyecto	esperar	0.0142509559	0.2204301075	83.444.397.284
publicar	fecha	0.010775113	0.5535714286	82.519.430.052
cargar	código	0.0118178658	0.3578947368	79.818.849.449



unidad	enviar	0.0100799444	0.3186813187	79.725.752.508
trabajar	reunión	0.010775113	0.4305555556	79.404.380.342
esperar	libro	0.013208203	0.5	78.606.557.377
realizar	revisión	0.0177267987	0.1661237785	77.086.792.056
realizar	viabilidad	0.0118178658	0.1107491857	74.098.931.899
encontrar	matriz	0.0139033716	0.3305785124	69.931.939.718
actualizar	información	0.0166840459	0.7868852459	6.737.704.918
tener	autoría	0.0104275287	0.063559322	6.530.720.339
tener	contenido	0.0118178658	0.0720338983	64.762.976.695
proyecto	revisar	0.0163364616	0.252688172	64.334.855.838
revisar	libro	0.0159888773	0.407079646	63.998.259.103
código	realizar	0.0302398332	0.6744186047	63.202.030.149
información	presentar	0.0163364616	0.1398809524	62.880.859.375
proyecto	cargar	0.013208203	0.2043010753	61.870.967.742
esperar	indicador	0.0100799444	0.3815789474	59.021.646.859
indicador	encontrar	0.0159888773	0.247311828	5.880.298.587
enviar	proyecto	0.0149461244	0.3739130435	57.835.904.628
ayuda	realizar	0.0135557873	0.609375	57.106.575.733
matriz	hacer	0.013208203	0.2794117647	55.059.427.881
realizar	ejecución	0.0166840459	0.1563517915	54.856.598.077
código	encontrar	0.0100799444	0.2248062016	53.451.854.699
proyecto	realizar	0.0361487661	0.5591397849	52.398.865.189
incluir	información	0.015641293	0.6	51.375
encontrar	costo	0.0118178658	0.2809917355	4.959.590.326
pago	tener	0.0104275287	0.8108108108	4.942.166.743
viabilidad	tener	0.0118178658	0.7906976744	48.195.703.587
indicador	enviar	0.0100799444	0.1559139785	39.005.610.098
gestión	tener	0.0111226973	0.4266666667	26.006.779.661
verificar	proceso	0.0100799444	1	19.759.615.385

Tabla 13. **Tabla de Sustantivos/Verbos relacionados correo 2.**

Tal como se notó en el modelamiento, se pueden obtener las funciones pero es indispensable que sea el dueño del cargo el que analice estas relaciones, o al menos alguien con conocimiento del negocio.

Como siguiente paso, se tomaron 7 correos por área teniendo en cuenta las áreas de mayor frecuencia y se le dan a un evaluador para que redacte las funciones a partir de dichos correos. Para verificar si el uso de esta metodología es buena, se decide: a) comparar contra el perfil generado por RR.HH las funciones obtenidas y b) comparar contra un método ingenuo que consiste en dividir los correos por área y escoger aleatoriamente, este ejercicio lo realizamos 3 veces, es decir se obtiene 3 grupos de correos aleatorios, donde no se tuvo en cuenta la metodología de MBA. El evaluador desconoce la naturaleza del cargo y es una persona externa a la organización dónde se desarrolló la investigación, solo conoce el nombre de la empresa y el cargo pero su única fuente son los correos electrónicos obtenidos.

El evaluador realiza una descripción de los siguientes puntos a partir de su lectura de los correos enviados:

- La Misión del cargo
- Actividades del cargo
- Problemas típicos del cargo

- Relaciones con otras áreas

Se generaron dos formatos, se marcaron internamente el Perfil 1 (funciones y datos obtenidos a partir del Market Basket Analysis y el Perfil 2, 3 y 4 como el correo ingenuo. En el Anexo 2 se incluyó el perfil real del trabajador de conocimiento y en el Anexo 3 pueden verse los cuatro perfiles generados.

El evaluador final es el dueño del cargo. Esta persona no tendrá ninguna forma de diferenciarlos ni tendrán numeración para evitar que se sesguen los resultados o conclusiones a las que llegue. A partir de estos perfiles se llevaron a cabo dos análisis:

- Primero, se compararon las funciones del perfil elaborado en la compañía con los perfiles elaborados 1, 2, 3 y 4, si cumplían con la función tomaba el valor de 1, sino el valor de cero, si se acercaba a la función 0,5, luego se ponderó y se obtuvo un porcentaje que representa el nivel en el cual el perfil elaborado corresponde con el perfil elaborado en la compañía. Estos fueron los resultados. Ver tabla 14.
- Segundo, el dueño del cargo realizó una evaluación subjetiva de cada uno de los cuatro perfiles generados, los datos finales están en el anexo 3.

COMPARACIÓN CON LAS FUNCIONES DEL PERFIL	%	PUNTAJE UTILIZANDO MBA	PUNTAJE UTILIZANDO UN ALEATORIO 1	PUNTAJE UTILIZANDO UN ALEATORIO 2	PUNTAJE UTILIZANDO UN ALEATORIO 3
• Participar en la creación, implementación y seguimiento de procesos editoriales estandarizados para Contenidos y Soluciones Educativas.	9%	50%	100%	50%	100%
• Colaborar en la definición de los modelos de reutilización y aprovechamiento de contenidos.	9%	0%	0%	0%	0%
• Generar herramientas de medición de los indicadores de gestión necesarios para garantizar mejora continua del proceso editorial.	9%	100%	100%	100%	100%
• Orientar, medir y publicar indicadores de gestión.	9%	100%	100%	100%	100%
• Participar en el análisis de viabilidad de proyectos de Colombia. Orientar y hacer seguimiento al análisis de viabilidad de los países,	9%	100%	0%	0%	0%
• Presupuestar y calcular la distribución de planta de Contenidos y Soluciones Educativas de Colombia (textos, LIJ, Contenidos Digitales y Greenwich) y monitorear sus cambios.	9%	100%	0%	0%	0%
• Hacer seguimiento al castigo de los costos editoriales generados en proyectos o productos aplazados o cancelados.	9%	100%	100%	0%	100%
• Hacer seguimiento y control a los cronogramas y costos editoriales de los proyectos. Generar y analizar reportes de compras editoriales.	9%	100%	100%	100%	100%
	9%	0%	50%	100%	100%

• Identificar necesidades de capacitación para los equipos editoriales relacionadas con procesos editoriales, indicadores de gestión, manejo presupuestal, gestión documental y propiedad intelectual.	9%	0%	0%	0%	0%
• Ser flexible a los procesos de cambio.	9%	0%	0%	0%	0%
RESULTADO		<b>59%</b>	<b>50%</b>	<b>41%</b>	<b>55%</b>

<b>COMPARACIÓN CON LAS FUNCIONES DEL PERFIL</b>
Participar en la generación de políticas que permitan la rentabilización de los productos y la eficiencia de los procesos de Contenidos y Soluciones Educativas a través del seguimiento, control y la mejora continua de los procesos editoriales.

<b>50%</b>	<b>50%</b>	<b>30%</b>	<b>50%</b>
------------	------------	------------	------------

Tabla 14. Ponderación de los perfiles obtenidos con la extracción de un tercero y la calificación del investigador

Como puede observarse, el promedio del perfil obtenido con el método propuesto con la metodología del MBA supera el promedio del perfil ingenuo obtenido con correos aleatorios. Cabe anotar algunas de las funciones que faltaron en los perfiles aleatorios :

- Participar en el análisis de **viabilidad** de proyectos de Colombia. Orientar y hacer seguimiento al análisis de viabilidad de los países,
- **Presupuestar** y calcular la distribución de planta de Contenidos y Soluciones Educativas de Colombia (textos, LJJ, Contenidos Digitales y Greenwich) y monitorear sus cambios.

Se observa que dentro del MBA estas palabras fueron identificadas. Ver Tabla 15

RAIZ 1	RAIZ 2	support	confidence	lift
proyecto	cubrir	0.0100799444	0.1559139785	13.193.074.004
proyecto	desarrollar	0.0111226973	0.1720430108	10.531.228.552
proyecto	esperar	0.0142509559	0.2204301075	83.444.397.284
viabilidad	realizar	0.0118178658	0.7906976744	74.098.931.899
proyecto	revisar	0.0163364616	0.252688172	64.334.855.838
proyecto	cargar	0.013208203	0.2043010753	61.870.967.742
proyecto	enviar	0.0149461244	0.2311827957	57.835.904.628
proyecto	encontrar	0.015641293	0.2419354839	57.524.660.091
proyecto	realizar	0.0361487661	0.5591397849	52.398.865.189
viabilidad	tener	0.0118178658	0.7906976744	48.195.703.587
proyecto	hacer	0.0152937087	0.2365591398	46.615.112.682
proyecto	presupuestar	0.0128606187	0.1989247312	44.711.441.532
proyecto	agradecer	0.0298922489	0.4623655914	44.340.860.215
proyecto	tener	0.0441432047	0.6827956989	41.618.712.411
proyecto	comprar	0.0135557873	0.2096774194	41.317.940.787
proyecto	ir	0.0114702815	0.1774193548	36.459.677.419
seguimiento	saludo	0.0100799444	0.90625	24.074.619.114
seguimiento	avenida	0.010775113	0.96875	19.142.127.404
seguimiento	proceso	0.010775113	0.96875	19.142.127.404
seguimiento	educación	0.010775113	0.96875	19.128.989.362
proyecto	reflejar	0.0163364616	0.252688172	11.184.367.246
proyecto	tener	0.013208203	0.5671641791	13.712.028.095
proyecto	tener	0.013208203	0.5671641791	13.712.028.095

Tabla 15. **MBA de las palabras claves de las funciones no encontradas en otros perfiles**

## 7.5. Evaluación y resultados de la investigación

El trabajador del conocimiento, dueño del cargo modelado evaluó los cuatro formatos (Ver Anexo 3):

1. 1(MBA)
2. Aleatorio1
3. Aleatorio2
4. Aleatorio3

En estos formatos se describe el perfil del cargo, que hasta ese momento desconocemos, solamente se le agregó la frecuencia por áreas y el mapa relacional de correos recibidos y enviados

El trabajador del conocimiento respondió las siguientes preguntas y se le realizó una entrevista abierta semiestructurada para obtener más información. Luego, se creó un formato (Ver Anexo 4) para tener una guía, durante la conversación. Sin embargo, lo más importante de esta entrevista era la obtención de la opinión que generaba la presentación de los perfiles al Trabajador del conocimiento (dueño del cargo evaluado).

Después de aproximadamente una hora de entrevista, el dueño del cargo concluyó que:

- Las áreas identificadas en el perfil (Tabla 16), estaban dentro de lo identificado en los correos electrónicos (Tabla 17). Adicionalmente dentro del perfil no aparece explícito relaciones importantes como con costos y mercadeo.

Su áreas del perfil	Áreas generales Fueron identificadas en la metodología	
	Áreas generales	Fueron identificadas en la metodología
Administración global de proceso editorial	Editorial	si
Analista Junior de Gestión documental	Editorial	si
Directores Editoriales, Jefes editoriales, Coordinadores editoriales y Jefe de Contenidos Digitales.	Editorial	si
Editores Jefes de Área, editores, Jefe de diseño y diseñadores	Editorial	si
Administrador de Contenidos	Editorial	si
Asistente Administrativa	Administrativa	si
Secretaria Editorial Global	Editorial	si
Compras	Compras	si
Gestión Humana	Gestión Humana	si

Tabla 16. **Áreas del perfil.**

AREA
EDITORIAL
COSTOS
DISEÑO
COMPRAS
MERCADEO
FINANCIERO
PERSONAL
VENTAS
GESTION HUMANA
REGALIAS*

PROVEEDOR
CEDI
VARIOS
INNOVACIÓN
OPERACIONES
PLANEACIÓN DE LA DEMANDA
ADMINISTRATIVO
PROMOCIÓN
PRODUCCION
VARIOS
OPTIMIZACION
CALIDAD

Tabla 17. **Áreas identificadas con la metodología.**

- Las gráficas y las tablas de frecuencias muestran claramente las tendencias reales del trabajo desarrollado, las interacciones con las diferentes áreas y las actividades críticas del cargo. Las gráficas muestran las áreas con las que se tiene interacción, los porcentajes para el caso de Guatemala, Perú y Argentina sorprendieron, porque con ellos, el KW tiene la percepción de tener una mayor interacción gracias a las reuniones y teleconferencias que se programan. Cabe anotar nuevamente que éstas no son objeto del presente estudio.
- Si se tienen en cuenta las funciones y el análisis de las relaciones generadas a partir de la metodología, el perfil es mejor para el KW que el generado por Gestión Humana. Porque es más claro en sus relaciones, toma en cuenta áreas que en el perfil original no están y contiene información del tráfico de los correos en el tiempo.
- Si solo se evalúa el tema de funciones sin las relaciones el perfil es peor el resultado, pero da un acercamiento muy aproximado a las funciones descritas. No obstante, las actividades relacionadas son muy específicas y no permiten llegar a la misión real del cargo, donde el alcance es mayor. Muy posiblemente porque además de la gestión por correo realiza alguna gestión personal y telefónica que no tienen objeto en este estudio. Además también cuenta un factor de interpretación ya que el evaluador que redacta las funciones no tiene conocimiento del entorno en que se desarrolla este KW, para evitar sesgos.
- La rapidez con que se obtuvo la evaluación de los perfiles y lo cercano de las funciones le sorprendió al KW. Generalmente este proceso puede llevar más de una semana y en solo una entrevista de una hora, puede tenerse una idea clara de las funciones y así estructurar el cargo.

## 7.6. Desarrollo

A continuación se enumera cada paso de la metodología realizada:

### 1. Obtención de la información

- Realizar un contexto general de la empresa donde se realiza el estudio
- Seleccionar un cargo al que se le desea realizar la metodología.
- Obtener el archivo que contenga la información del correo electrónico de un periodo de tiempo de información del cargo que se desea evaluar. En el caso del Outlook sería un PST.

### 2. Comprensión y preparación de datos

- Recopilar todos correos en una sola carpeta para aplicarles el mismo análisis

- Extracción de variables Selección de las variables extraídas y generadas (enviado/recibido, Asunto, De (remitente, Para (enviado), CC (Copia), Recibido (fecha), Tamaño mensaje, Contenido, Contiene adjuntos)
  - Obtener un directorio con Gestión Humana para crear las variables faltantes (área, interno/externo, mes, día y hora) y cruzarla con la base de datos
  - Análisis de las variables no textuales (Frecuencias y participación)
    - Analizar por los correos Recibido y Enviado
    - Analizar por el origen del correo
      - Por país
      - Externo o interno en la organización
      - Área a la que pertenece
    - Analizar por tiempo o momento que se recibe o se emite un correo
      - Por hora
      - Por día
      - Por mes
    - Analizar por la existencia o no de adjuntos en el correo
    - Analizar por tamaño del mensaje
    - Hacer el mapa relacional
  - Analizar las variables textuales
    - Realizar limpieza de reenvíos y de mensajes de seguridad para evitar que la información se repita y salgan palabras no representativas
    - Realizar limpieza de stop words
- 3. Obtener las funciones a partir de los correos representativos**
- Dividir el contenido del correo en las partes de la oración (part of speech) a través del Treetagger
  - Lematizar a través del Treetagger
  - Realizar Market basket analysis utilizando el Wordstat o SAS
  - Revisar las parejas (sustantivo-verbo) de mayor lift para encontrar las parejas claves
  - Hallar los correos más representativos por área (aquellos que incluyen mayor cantidad de parejas sustantivo-verbo, con mayor lift y support). Aproximadamente 7 por área.
  - Redactar las funciones a partir de los correos electrónicos representativos.
- 4. Armar o realizar el perfil con las funciones obtenidas y el mapa relacional con las tablas de frecuencias**

El desarrollo o implementación de la metodología podría generar un producto de conocimiento con potencial de ser trasladado de manera práctica a la industria (spin-off). Aunque algunas partes se automatizaron, se requeriría una aplicación exclusiva para llevar su implementación en forma masiva.



## 8. Conclusiones

1. La metodología desarrollada durante este trabajo presenta características que la hacen viable y funcional para la extracción de funciones y mapas relacionales de un cargo. Algunas de sus bondades son:
  - Es práctica. Después que se automatiza la obtención de los correos representativos, su desarrollo es muy rápido y sencillo, lo que permite extraer más fácilmente las funciones y las relaciones del cargo. Se debe realizar una entrevista corta para perfilar las funciones obtenidas pero el tiempo es mínimo comparando el levantamiento normal de un cargo de un KW
  - Permite extraer las funciones de una persona que deja su cargo, esto es, permite tener una visión más amplia si la persona no se encuentra. En este caso un colaborador cercano a la persona que haya abandonado el cargo sería la persona ideal que perfilaría las funciones finales. En el caso de estudio, posibilitó identificar funciones en un 59%. Las relaciones reales se determinan en un 100% porque los correos incluían las áreas del perfil de Gestión Humana y además otras áreas que no estaban identificadas y con las que guardaba mucha relación.
  - Permite medir la importancia de la relación relativa del cargo con otras áreas de forma más objetiva, evitando los sesgos del autorreporte. Hay una focalización en las áreas claves de la organización y permite determinar en qué momento durante el mes el tráfico de comunicación se hace más frecuente. Permite, de forma visual, identificar las áreas claves con las cuales se mantiene una mayor relación.
  - Permite determinar relaciones que no son evidentes ni para el mismo trabajador a quién se le está realizando el estudio, disminuyendo de este modo también los sesgos del autorreporte.
  - Permite agilizar el proceso de extracción de funciones. El KW puede focalizarse en lo realmente importante porque prioriza las funciones, que pueden ayudar en su desarrollo. Con un análisis mayor permitiría una identificación de los problemas más frecuentes o críticos y esto puede ser fuente de información o de entrada para el desarrollo de las competencias de una persona porque puede determinar falencias.
  - En caso de movimientos internos en la empresa por ascensos o estructuración o fusión de cargos, permitiría identificar más rápidamente el perfil.
2. Se obtienen correos representativos que sintetizan las funciones básicas de un KW. Aunque cabe anotar que la diferencia en los resultados no es lejana entre la utilización del MBA y un método aleatorio, entre más cercano sean las funciones obtenidas menor es el tiempo que se debe utilizar para perfilar las funciones finales. Si se realiza la lectura de correos aleatorios sin la metodología, se pueden encontrar con correos interesantes pero también con correos que no corresponden directamente con sus funciones o con tareas menores. Este riesgo puede ser muy alto. La metodología garantiza que los correos escogidos son densos, en el sentido de que incluyen relaciones de palabras que son significativas para el cargo.
3. Se requiere analizar las variables textuales y no textuales porque una no excluye a la otra. Antes bien, la información de las dos se complementan. Las relaciones van amarradas a las funciones del KW, las dos forman una mejor visión del KW. En específico el mapa relacional resume en forma visual las áreas con mayores frecuencias durante el mes. Esto permite un vistazo muy general del cargo e identificar los puntos donde se pueden tener “cuellos de botella” por la cantidad de correos que van muy relacionados a la cantidad o comunicaciones de trabajo.
4. La metodología propuesta logra dos objetivos clásicos de la minería de datos: síntesis de información y descubrimiento de nuevo conocimiento. En el primer caso, la tarea de revisar

completamente un correo electrónico para entender un cargo se reduce a la tarea de ver unos mapas relacionales y leer unas cuantas decenas de correos a lo sumo para darse una idea general del cargo. En el segundo caso, la eliminación de los sesgos humanos en la ponderación de la información permite darle una mirada nueva, diferente, y más objetiva desde un punto de vista cuantitativo a la dinámica de intercambio de información de un cargo. Por ello, se considera lograda una aplicación correcta de la minería de datos, en este caso, minería de texto.

5. La metodología presenta aún las siguientes dificultades y desafíos que deberá enfrentar para un mayor desarrollo:
  - Mayor automatización e independencia del factor humano. A futuro, se deberá intentar reducir la etapa que requiere de una persona para la elaboración de funciones.
  - Encuentro de relaciones triples y cuádruples, y elaboración automática de oraciones a partir del MBA. Ello facilitaría la automatización previa
  - Mejor escogencia de los puntos de corte en el lift, y en general, un refinamiento de la metodología. El objetivo del trabajo no fue optimizar la metodología, sino generar una primera aproximación a la misma para encontrar sus bondades.
  - Utilizar no solo el sustantivo y el verbo dentro del MBA sino utilizar palabras como adjetivos y adverbios que nos pueden mostrar como se siente el KW.
  - Revisión del número de los correos óptimos a revisar para la extracción de las funciones

## 9. Recomendaciones futuras

1. Automatizar cada uno de los pasos para lograr tiempos cortos en la implementación de la metodología.
2. No excluir palabras como los adjetivos donde se pueden realizar estudios sobre cómo se siente el trabajador dentro de la organización, identificando por ejemplo clima laboral, relaciones con sus compañeros o con su jefe.

## 10. Referencias Bibliográficas

Chase, Richard B., Aquilano, Nicholas J & Jacobs, F, Robert. (2000). Administración de producción y operaciones. Cap. 11; Pags: 416-418.

Niebel, Benjamin W (1996). Ingeniería Industrial. Métodos, Tiempos y Movimientos. Cap. 11. Análisis y Evaluación de puestos o cargos. Pags: 316-318.

Hammer, Michael; Leonard, Dorothy & Davenport, Thomas .(2004). Why Don't We Know More About Knowledge? MIT Sloan Management Review. Cambridge: Summer 2004, Vol. 45, Iss. 4. Pags: 14.

Henard DH & McFadyen MA. (2008). Making knowledge workers more creative. Research-technology management, Vol. 51, Iss. 2. Pags: 40-46. Mar-Apr2008.

Parise, Salvatore; Cross, Rob & Davenport, Thomas H.(2006). Fuga de Conocimiento. Gestión\MIT Sloan Management Review. Massachusetts Institute of Technology. Dic 2006 - Ene2007.

Lloria, M Begoña. (2008). A review of the main approaches to knowledge management. Knowledge Management Research & Practice. Houndmills: Mar 2008. Vol. 6, Iss. 1; Pags: 77, 13.

Hemp, Paul. (2009). Death by information overload. Harvard Business Review. Sep 2009. Pags: 83-89.

Belhadjali, M.; Abbasi, S., & Whaley, G.(2010). A Text Mining Approach to the Health Care Debate. Competition Forum, 8(2); Pags: 260-264. Retrieved February 12, 2011, from ABI/INFORM Global. (Document ID: 2174555631).

Capuano, Nicola; Gaeta, Matteo, Ritrovato; Pierluigi, & Salerno Saverio.(2008). How to integrate technology-enhanced learning with business process management. Journal of Knowledge Management, 12(6), Pags: 56-71. Retrieved February 12, 2011, from ABI/INFORM Global. (Document ID: 1572050241).

Dorre, Jochen; Gerstl, Peter & Seiffert Ronald (1999). Text Mining: Finding Nuggets in Mountains of Textual Data. In KDD-99 Proceedings, San Diego, CA: ACM. Pags. 398-401.

Feldman, Ronen, & Sanger, James. (2007). The Text Mining Handbook.

Sakurai, Shigeaku (2008). Rule Discovery From Textual Data. Chapter VI 2008. Emerging Technologies of Text Mining. Pags: 1-18.

Segall, R.S. & Zhang, Q. (2006). Data visualization and data mining of continuous numerical and discrete nominal-valued microarray databases for biotechnology, Kybernetes: International Journal of Systems and Cybernetics. Vol. 35 No. 9. Pags: 1538-66.

Segall, R.S. & Zhang, Q. (2010). Review of data, text and web mining software. Vol. 39 No. 4. Pags: 625-655.

Amir, A.; Aumann, Y.; Feldman R. & Fresko, M. (2005). Maximal association rules: a tool for mining associations in text. Journal of Intelligent Information Systems, Vol. 25 No. 3. Pags: 333-345.

Spasic, I.; Ananiadou, S.; McNaught, J. & Kumar, A. (2005). Text mining and ontologies in biomedicine: making sense of raw text", Briefings in Bioinformatics, Vol. 6 No. 3; Pags: 239-51.

Sakurai, S. & Suyama A. (2004). E-mail analysis method based on text mining techniques. Applied Soft Computing 6; Pags: 62-71.

Weng, Sung-Shun & Liu, Chih-Kai. (2004). Using text classification and multiple concepts to answer e-mails. *Expert Systems with Applications* 26, Pags: 529–543.

Chiavenato, Idalberto. (2000). *Administración de Recursos Humanos*. Editorial Mc Hill. Pags: 291-353.

Cuesta, Armando (2010). *Gestión del Talento Humano y del Conocimiento*. Ecoe Ediciones. Pag 271-272.

Fox, Wasylyshyn (2005). Handling missing data in self-report measures. *Research in nursing & health*, Vol. 28, Iss 6; Pags: 488-495. Dec 2005.

Robinson, Mark A. (2010). *Work Sampling: Methodological Advances and New Applications*. Human factors and ergonomics in manufacturing & service industries, Vol. 20, Iss. 1; Pags: 42-60. Jan-feb 2010.

Stuit, Marco & Wortmann, Hans (2012). Discovery and Analysis of email-driven business processes *Information Systems* 37. *Journal Information Systems Volume 37 Issue 2, April, 2012* Pags 142-168 .

Petrovic, Sasa; Snajder, Jan & Basic, Bojana Dalbelo . (2010). Extending lexical association measures for collocation extraction. *Computer speech and language*. Vol. 24, Iss. 2. Pags: 383-294. Apr 2010.

Kang, Hyunmo; Plaisant, Catherine; Elsayed, Tamer & Oard, Douglas W. (2010). Making Sense of Archieve E-mail : Exploring the Enron Collection with NetLents. *Journal of the American Society for Information Sciencie and Technology*. Vol. 61, Iss. 4. Pags: 723-744.

Leuski, Anton. (2004). E-mail is a Stage:Discovering People Roles from e-mail Archives. July 25–29. 2004. In *Proceedings of the 27th Annual International ACM SIGIR. Conference on Research and Development in Information Retrieval*. Pags: 502–503. NewYork: ACM Press.04/0007.

Mc Arthur, Robert & Bruza, Peter. (2003). Discovery of implicit and explicit connectionsbetween people using e-mail utterance. In *Proceedings of the Eighth European Conference ofComputer - Supported Cooperative Work* . Pags: 21-40.

Tyler, Joshua; Wilkinson, Dennis & Huberman, Bernardo. (2005). E-mail as spectroscopy: Automated discovery of community structure withing organizations. *The Information Sociaty*. Cap. 21. Pags: 133-141.

Chen, Hsinchun & Chau, Michael. (2011). Enterprise risk and security management: data, text and web mining. *Decision Support systems*. Vol. 50,Iss. 4. Pags: 649-650. Marzo 2011.

Holton, Carolyn. (2009) Identifying disgruntled employee systems fraud risk through text mining: A simple solution for a multi-billion dollar problem. *Decision Support System*. Vol. 46, Iss.4. Pags: 853-854.

Chou, Chen-Huei; Sinha, Atish & Zhao, Huimin. (2008) .A text mining approach to internet abuse detection. *Information systems and e-business management*. Vol. 6, Iss. 4. Pags: 419-439.

Thomson, P. (2005). Text mining, names and security . *Journal of database management*. Vol. 16, Iss. 1; Pags: 54-59. Jan-Mar 2005.

Chapman, Pete; Clinton, Julian; Kerber, Randy; Khabaza, Thomas; Reinartz, Shearer & Colin, Thomas, Wirth. (1996). *Guía paso a paso de Minería de Datos*. CRISP-DM 1.0.

Stuit, Marco & Wortmann, Hans. (2011), Discovery and analysis of e-mail-driven business processes. *Journal of Biomedical Informatics*. Vol.4. Iss.6. Pags: 142-168. Sep. (2011).

Kushmerick, N. & Lau, T.. 2005. Automated e-mail activity management: an unsupervised learning approach, in: Proceedings of the 10th International Conference on Intelligent User Interfaces, ACM Press, San Diego, California, USA, 2005. Pags. 67–74.

Yu, L. & Wang, S., & Lai, K. K. (2005). A rough-set-refined text mining approach for crude oil market tendency forecasting. *International Journal of System Sciences*, Vol. 2. Iss. 1. Pags. 33–46.

Ur-Rahman, N & Harding, J.A. (2012). Textual data mining for industrial knowledge management and text classification: A business oriented approach. *Expert Systems with Applications* Vol. 39. Iss. 5. Pags 4729-4739. April, (2012).

Holton, Carolyn. (2009). Identifying disgruntled employee systems fraud risk through text mining: A simple solution for a multi-billion dollar problem. *Decision Support Systems* Vol. 46. Pags: 853–864.

Iqbal\*, Farkhund; Hadjidj, Rachid & Fung, Benjamin C.M., & Debbabi, Mourad. (2008). A novel approach of mining write-prints for authorship attribution in e-mail forensics. *Digital Investigation*. May (2008). Pags: 42–51.

Iqbal\*, Farkhund; Hadjidj, Rachid & Fung, Benjamin C.M., & Debbabi, Mourad. (2010). Mining writeprints from anonymous e-mails for forensic investigation. *Digital Investigation*. Jul (2010). Pags: 56–64.

Nebrija, Antonio. (2012). Gramática de la lengua castellana.

Su, Xiaoyuan & Khoshgoftar, Tagui M. (2009) A Survey of Collaborative Filtering Techniques. Department of Computer Science and Engineering, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA. Received 9 February 2009; Accepted 3 August 2009. Pags: 1-2.

Schmid, Helmut (1995). Improvements in Part-of-Speech Tagging with an Application to German. Proceedings of the ACL SIGDAT-Workshop. Dublin, Ireland. Pags: 1-9.

Schmid, Helmut (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. Proceedings of International Conference on New Methods in Language Processing, Manchester, UK. Pags: 1-9.

RAE, Real Academia Española (2012) . Pag Web: <http://www.rae.es/rae.html>.

Hoanca, Bogdan & Mock, Kenrick (2011). Using Market Basket Analysis to Estimate Potential Revenue increases for a Small University Bookstore. Conference for Information Systems Applied Research. Wilmington North Carolina, USA. 2011 CONISAR Proceedings. v4 n1822.

Hodson, William K. Maynard Manuel del ingeniero Industrial. (1996). Tomo 1. Sección 5 Aplicación y Control de la medición del trabajo. Cap. 1 Procedimientos de Administración y Control. Pags: 5.3, 5.4 y 5.5.

Hodson, William K. Maynard Manuel del ingeniero Industrial. (1996). Tomo 1. Sección 5 Aplicación y Control de la medición del trabajo. Cap. 7 Desarrollo de Estándares Administrativos y de Trabajo de oficina. Pags: 5.127, 5.128, 5.129.

# 11. Anexos

## Anexo 1. Formato original utilizado en la empresa del estudio

### DESCRIPCION DE CARGO

### NOMBRE DEL CARGO

No.

FECHA DE REALIZACIÓN \_\_\_\_\_

PAIS: \_\_\_\_\_  
CARGO: \_\_\_\_\_

SECCIÓN: \_\_\_\_\_  
CARGO JEFE INMEDIATO: \_\_\_\_\_

#### MISION DEL CARGO:

--

#### FORMACIÓN ACADÉMICA REQUERIDA

--

#### FORMACIÓN ACADÉMICA DESEABLE

--

#### EXPERIENCIA REQUERIDA

Años de Experiencia:					Trayectoria / Áreas :	Conocimientos Específicos:
Ning.	< 3	3 a 5	5 a 10	> 10		

#### EXPERIENCIA DESEABLE

Años de Experiencia:					Trayectoria / Áreas :	Conocimientos Específicos:
Ning.	< 3	3 a 5	5 a 10	> 10		

#### DIMENSIONES:

	Actuales	Futuras	Cobertura Geográfica
No de personas directas	<input type="text"/>	<input type="text"/>	<input type="text"/>
No de personas indirectas	<input type="text"/>	<input type="text"/>	
Valor nómina personas directas	\$	<input type="text"/>	
Presupuesto Operativo Asignado	\$	<input type="text"/> ( gastos )	
		Ppto de Salarios de Voluntad y Ppto del área	

#### ACTIVIDADES DEL CARGO:

ACTIVIDADES PRINCIPALES			
%	ACTIVIDAD	RESULTADO FINAL ESPERADO	FORMA DE MEDIR EL LOGRO

<b>ACTIVIDADES DE APOYO</b>

<b>AUTORIDAD PARA TOMAR DECISIONES</b>
--

<b>DECISIONES PROPIAS</b>	<b>DECISIONES CONSULTADAS</b>

<b>PROBLEMAS TÍPICOS DEL CARGO</b>
------------------------------------


<b>RELACIONES DE TRABAJO</b>
------------------------------

<b>RELACIONES INTERNAS:</b>			<b>RELACIONES EXTERNAS:</b>		
Frec.:	Área:	Propósito:	Frec.:	Área:	Propósito:

<b>MODELO DE COMPETENCIAS ASOCIADO AL CARGO:</b>
--

COMPETENCIA	NIVEL	CRÍTICA

COMPETENCIA	NIVEL	CRÍTICA

<b>RESPONSABILIDADES EN EL SISTEMA (ORACLE)</b>
---


<b>VALORACION HAY</b>
-----------------------

<b>Know How</b>	<b>Solución Problemas</b>			<b>Responsabilidad</b>			<b>PUNTOS TOTAL</b>	
							<b>PERFIL</b>	

Anexo 2. Perfil de desempeño correo de validación

**DESCRIPCIÓN DE CARGO**

**COORDINADOR PROCESO EDITORIAL DE CONTENIDOS Y SOLUCIONES EDUCATIVAS**

No.

**FECHA DE REALIZACIÓN:** Julio de 2011  
**PAIS:** Colombia **SECCIÓN:** Contenidos y Soluciones Educativas  
**CARGO:** Coordinador Proceso Editorial de Contenidos y Soluciones Educativas **CARGO JEFE INMEDIATO:** Administrador Global Proceso Editorial de Contenidos y Soluciones Educativas

**MISIÓN DEL CARGO**

Participar en la generación de políticas que permitan la rentabilización de los productos y la eficiencia de los procesos de Contenidos y Soluciones Educativas a través del seguimiento, control y la mejora continua de los procesos editoriales.

**FORMACIÓN ACADÉMICA REQUERIDA**

Título universitario en Ingeniería Industrial, Administración de empresas o carreras afines.  
 Manejo de excel avanzado.

**FORMACIÓN ACADÉMICA DESEABLE**

Especialización o maestría en Sistemas de Gestión.  
 Especialización ó maestría en Dirección estratégica de proyecto.  
 Conocimiento en Propiedad Intelectual.

**EXPERIENCIA REQUERIDA**

Años de Experiencia:					Trayectoria / Áreas :	Conocimientos Específicos:
Ning.	< 3	3 a 5	5 a 10	> 10		
	X				Sistema de Gestión de calidad	Control de procesos, manejo de indicadores de gestión, análisis de resultados y mejora continua.
		X			Software	Manejo de excel avanzado, word, visio
	X				Administrativa	Control presupuestal, seguimiento y control de proyectos, administración de procesos y procedimientos.

**EXPERIENCIA DESEABLE**



Años de Experiencia:					Trayectoria / Áreas :	Conocimientos Específicos:
Ning.	< 3	3 a 5	5 a 10	> 10		
	X				Editorial	Conocimiento del proceso editorial de educación ó de literatura infantil y juvenil. Conocimiento del sector.
	X				Propiedad Intelectual	Legislación en propiedad intelectual.

#### DIMENSIONES

No de personas directas	Actuales	Futuras	
	0	0	Corbertura Geográfica
No de personas indirectas	Actuales	Futuras	Contenidos y Soluciones Educativas
	0	0	
Valor nómina personas directas	\$		N.S y mensual
Presupuesto Operativo Asignado	\$		( gastos )

#### ACTIVIDADES DEL CARGO

ACTIVIDADES PRINCIPALES			
%	ACTIVIDAD	RESULTADO FINAL ESPERADO	FORMA DE MEDIR EL LOGRO
40%	<ul style="list-style-type: none"> <li>• Participar en la creación, implementación y seguimiento de un procesos editorial estandarizado para Contenidos y Soluciones Educativas.</li> <li>• Colobarar en la definición de los modelos de reutilización y aprovechamiento de contenidos.</li> <li>• Generar herramientas de medición de los indicadores de gestión necesarios para garantizar mejora continua del proceso editorial.</li> <li>• Orientar, medir y publicar indicadores de gestión.</li> </ul>	Optimización del proceso editorial.	Estructuras eficientes y ahorro en compras editoriales. Indicadores de gestión. Seguimiento a los resultados del proceso editorial global y de sus variantes nacionales. Implementación de acciones preventivas y correctivas.
30%	<ul style="list-style-type: none"> <li>• Participar en el análisis de viabilidad de proyectos de Colombia. Orientar y hacer seguimiento al análisis de viabilidad de los países,</li> <li>• Presupuestar y calcular la distribución de planta de Contenidos y Soluciones Educativas de Colombia (textos, LIJ),Contenidos</li> </ul>	Desarrollo de proyectos rentables, administración de flujo de caja y control de compras editoriales.	Indicadores de gestión. Cumplimiento de margen de productos. Distribución de planta en cero, al cierre del año.

	Digitales y Greenwich) y monitorear sus cambios. • Hacer seguimiento y control a los cronogramas y costos editoriales de los proyectos. Generar y analizar reportes de compras editoriales. • Hacer seguimiento al castigo de los costos editoriales generados en proyectos o productos aplazados o cancelados.		Disminución de compras editoriales castigadas en el PyG
15%	Cumplir los cronogramas de planeación de área.	Cumplimiento de metas del área.	Matriz de alto desempeño.
10%	Identificar necesidades de capacitación para los equipos editoriales relacionadas con procesos editoriales, indicadores de gestión, manejo presupuestal, gestión documental y propiedad intelectual.	Equipos actualizados.	Resultados de auditoría e indicadores de gestión.
5%	Ser flexible a los procesos de cambio.	Equipo alineado con las políticas de la organización.	Evaluación de desarrollo.

#### ACTIVIDADES DE APOYO

Asesorías a los equipos editoriales en análisis financiero y/o estadístico de los proyectos.			

#### AUTORIDAD PARA TOMAR DECISIONES

DECISIONES PROPIAS	DECISIONES CONSULTADAS
Comunicación y solicitud de información de los países.	Viajes.
	Asistencia a eventos
	Ejecuciones de presupuesto
	Distribución de planta de Colombia para Contenidos y Soluciones Educativas.
	Medición, análisis y publicación de indicadores
	Generación de acciones correctivas y preventivas.

#### PROBLEMAS TÍPICOS DEL CARGO

Desarrollo de múltiples actividades.			
Cruce de fechas de temporada entre países.			
Ubicación geográfica dispersa de equipo de equipos editoriales.			
Alteraciones de clima laboral.			

#### RELACIONES DE TRABAJO

RELACIONES INTERNAS:			RELACIONES EXTERNAS:		
Frec.:	Área:	Propósito:	Frec.:	Área:	Propósito:
Alta	Administración global de proceso editorial	<ul style="list-style-type: none"> <li>• Definición de metodologías de trabajo.</li> <li>• Asesoría, seguimiento, control, comunicación, toma de decisiones del proceso.</li> <li>• Gestión documental.</li> <li>• Ciclos de auditorías.</li> </ul>			

		<ul style="list-style-type: none"> <li>• Propiedad intelectual y manejo de contratos.</li> <li>• Capacitación.</li> </ul>			
Alta	Analista Junior de Gestión documental	<ul style="list-style-type: none"> <li>• Asesoría, seguimiento, control, comunicación, toma de decisiones del proceso.</li> <li>• Indicadores de gestión.</li> <li>• Ciclos de auditorías.</li> <li>• Capacitación.</li> </ul>			
Alta	Directores Editoriales, Jefes editoriales, Coordinadores editoriales y Jefe de Contenidos Digitales.	<ul style="list-style-type: none"> <li>• Administración de presupuestos.</li> <li>• Viabilidad de proyectos.</li> <li>• Evaluación desarrollo proyectos y cierre de productos.</li> <li>• Actualización y optimización de procesos.</li> <li>• Asesorías en la planeación proyectos.</li> <li>• Definición de techos presupuestales de compras editoriales.</li> </ul>			
Media	Editores Jefes de Área, editores, Jefe de diseño y diseñadores	<ul style="list-style-type: none"> <li>• Administración de presupuestos.</li> <li>• Viabilidad de proyectos.</li> <li>• Evaluación desarrollo proyectos y cierre de productos.</li> <li>• Indicadores de gestión.</li> <li>• Compras editoriales.</li> <li>• Actualización y optimización de procesos.</li> <li>• Seguimiento a proyectos editoriales.</li> <li>• Asesoría.</li> </ul>			
Media	Administrador de Contenidos	<ul style="list-style-type: none"> <li>• Definición de metodologías de trabajo.</li> <li>• Asesoría, seguimiento, control.</li> <li>• Indicadores de gestión AGRAL.</li> </ul>			
Alta	Asistente Administrativa	<ul style="list-style-type: none"> <li>• Cumplimiento procedimiento de compras.</li> <li>• Flujo de caja de proyectos.</li> <li>• Seguimiento a compras editoriales.</li> </ul>			
Alta	Secretaria Editorial Global	<ul style="list-style-type: none"> <li>• Solicitud y legalización de anticipos.</li> <li>• Atención telefónica.</li> <li>• Envío y recepción de correspondencia.</li> <li>• Servicios generales.</li> </ul>			
Media	Compras	<ul style="list-style-type: none"> <li>• Seguimiento al proceso de compras.</li> <li>• Definición de parámetros de negociación global de proveedores para compras editoriales.</li> <li>• Contratación de proveedores.</li> <li>• Evaluación de proveedores.</li> </ul>			
Media	Gestión Humana	<ul style="list-style-type: none"> <li>• Despliegue de resultados de evaluaciones de clima y liderazgo.</li> </ul>			

**MODELO DE COMPETENCIAS ASOCIADO AL CARGO:**

**SOPORTE CARVAJAL**

COMPETENCIA	NIVEL	CRÍTICA
Acompañamiento al cliente	3	MC
Conocimiento organizacional	3	MC
Impacto e influencia	2	CRÍTICA
Profundización en la información	3	MC
Compromiso con los resultados	3	MC
Pensamiento innovador	2	CRÍTICA

COMPETENCIA	NIVEL	CRÍTICA
Comunicación oral y escrita	3	MC
Trabajo en equipo	3	MC

**RESPONSABILIDADES EN EL SISTEMA (ORACLE)**

REPORTE DE GASTOS
SIPRO PROCESOS
SIPRO CONSULTA

**VALORACION HAY**

Know How	Solución Problemas				Responsabilidad				PUNTOS TOTAL		
									<b>PERFIL</b>		

Anexo 3. Perfil generado con el MBA

CUESTIONARIO COORDINADORA DEL PROCESO EDITORIAL	OBSERVACIONES DUEÑO DEL CARGO
<p><b>1. La Misión del cargo</b></p> <p>Asegurar que las especificaciones de los proyectos de impresión sean correctamente ingresados en el sistema, adicionalmente realiza seguimiento a la ejecución presupuestal y que el costo de los proyectos esté adecuadamente asignado. Verifica que el área de diseño tenga los guiones e imágenes para desempeñar su función</p>	<p><b>CALIFICACIÓN SIN TENER EN CUENTA EL ANÁLISIS DE RELACIONES=7</b></p> <p>El know how de la organización no es impresión, es desarrollo y producción de contenidos educativos. La misión debe estar definida no en función de la tarea realizada sino en el objetivo que se quiere encontrar con el cargo.</p>
<p><b>2. Actividades del cargo</b></p> <p>Direccionamiento para conocer la viabilidad de un proyecto</p> <p>Coordinar reuniones para avances de proyectos, en las que debe revisar el cronograma de cumplimiento para las diferentes propuestas de diseño.</p> <p>Revisión de indicadores para cumplir con los diferentes proyectos de diseño de libros.</p> <p>Revisar la viabilidad de un proyecto de impresión a través del costo del mismo</p> <p>Coordinar reuniones donde se agilicen temas de costos de proyectos, generando responsabilidades en cada área.</p> <p>Revisión de presupuestos</p> <p>Debe revisar archivos provenientes de diseño para su impresión</p>	<p>Directriz y coordinación, a cambio de direccionamiento</p> <p>Si</p> <p>Si</p> <p>Si</p> <p>Si</p> <p>Si y análisis</p> <p>No</p>
<p><b>3. Problemas típicos del cargo</b></p> <p>Revisar productos que nos se desarrollaron y se deben desmontar de presupuestos.</p> <p>La falta de códigos de libros o textos por otras áreas.</p> <p>Daños en archivos</p>	<p>Si, pero esta enunciada como actividad no como problema</p> <p>Identificado el problema pero no se enunció de manera correcta</p> <p>No es un problema de mi cargo, es de otro cargo</p>
<p><b>4. Relaciones con otras áreas</b></p> <p>Editorial</p> <p>Diseño</p> <p>Mercadeo</p> <p>Costos</p>	

Anexo 3. Perfil generado con los aleatorios 1

CUESTIONARIO COORDINADORA DEL PROCESO EDITORIAL		OBSERVACIONES DUEÑO DEL CARGO
<b>1. La Misión del cargo</b>		<b>CALIFICACIÓN SIN TENER EN CUENTA EL ANÁLISIS DE RELACIONES =7</b>
Coordinar los recursos digitales destinados a la elaboración de proyectos, así como la gestión de costos y las inconsistencias de información en el sistema		Recursos digitales no
<b>2. Actividades del cargo</b>		
Realizar seguimiento a los costos contenidos en las órdenes de compra		Si
Asegurar la entrega de la información correcta de archivos requeridos por los diseñadores para cada proyecto		No
Realizar análisis y seguimiento a los indicadores de gestión de costo de cada uno de los proyectos		Si, pero no solo de costo, de cronograma y monitorear calidad
Solucionar las inconsistencias que se puedan presentar con la creación y utilización de códigos en el sistema		Si
<b>3. Problemas típicos del cargo</b>		
Coordinación efectiva de las reuniones de seguimiento de costos		Si
Archivos de proyectos incorrectos o no actualizados		No
Creación no oportuna o inadecuada de códigos en el sistema		Si
<b>4. Relaciones con otras áreas</b>		
Editorial		
Diseño		
Mercadeo		
Costos		

Anexo 3. Perfil generado con los aleatorios 2

CUESTIONARIO COORDINADORA DEL PROCESO EDITORIAL	OBSERVACIONES DUEÑO DEL CARGO
<p><b>1. La Misión del cargo</b></p> <p>Revisión de los archivos entregado por otras áreas para la impresión de textos o documentos</p>	<p><b>CALIFICACIÓN SIN TENER EN CUENTA EL ANÁLISIS DE RELACIONES = 6</b></p> <p>No es la misión del cargo</p>
<p><b>2. Actividades del cargo</b></p> <p>1. Coordinación de archivos para la impresión de documentos</p> <p>2. Revisión de cronogramas</p> <p>3. Revisión de instructivos</p> <p>4. Coordinar con otras dependencias indicador de cronogramas</p> <p>5. Avanzar temas de licitaciones, reuniendo a las partes involucradas</p> <p>6. Coordinar archivos entregados por diseño para su impresión</p>	<p>No lo realizo</p> <p>Es una de las tareas críticas, pero más que la revisión mi misión es garantizar que se cumplan</p> <p>Revisión cuando los instructivos son de otras áreas pero tienen dentro del alcance los equipos editoriales, y redacción cuando son del proceso editorial directo</p> <p>Si</p> <p>En este caso, debo garantizar que las licitaciones sean rentables, por ello coordino a las partes al igual que en la viabilidad de los otros productos</p> <p>No lo realizo</p>
<p><b>3. Problemas típicos del cargo</b></p> <p>Creación de códigos</p> <p>Revisión de costos</p> <p>Dudas del proceso de cronogramas</p>	<p>No es un problema, es una actividad</p> <p>No es un problema, es una actividad</p> <p>No es un problema, es una actividad</p>
<p><b>4. Relaciones con otras áreas</b></p> <p>Editorial</p> <p>Diseño</p> <p>Mercadeo</p> <p>Costos</p>	<p>Ok</p> <p>Ok</p> <p>Ok</p> <p>Ok</p>

Anexo 3. Perfil generado con los aleatorios 3

CUESTIONARIO COORDINADORA DEL PROCESO EDITORIAL	OBSERVACIONES DUEÑO DEL CARGO
<p><b>1. La Misión del cargo</b>                      Coordinación, planeación de proyectos desde la revisión de costos hasta la revisión de documentación para su impresión</p>	<p><b>CALIFICACIÓN SIN TENER EN CUENTA EL ANÁLISIS DE RELACIONES = 7</b>                      Este es parte del alcance de mi cargo, pero no es la misión del mismo</p>
<p><b>2. Actividades del cargo</b></p> <ol style="list-style-type: none"> <li>1. Revisión de instructivos para entregas de productos libros o textos</li> <li>2. Coordinar reuniones para revisar temas de indicadores de cronogramas</li> <li>3. Revisión de indicadores</li> <li>4. Revisión de costos</li> <li>5. Asistir a Comités para revisar proyectos</li> <li>6. Revisión de procesos</li> <li>7. Revisión de documentación que proviene de diseño</li> <li>8. Solicitar información a costos sobre códigos de los diferentes proyectos</li> </ol>	<p>No es la revisión, es la redacción cuando son del proceso editorial directo</p> <p>No implica solo la coordinación para revisarlos, sino revisarlos para garantizar que se cumplen</p> <p>Análisis y generación de acciones según el resultado de los indicadores</p> <p>Es monitoreo y análisis</p> <p>Si, pero no es la asistencia, es la gestión de los proyectos la que nos hace asistir a los comités</p> <p>Mejora continua e identificación de reprocesos</p> <p>No lo hago, monitoreamos el equipo de diseño por ser un área que cambio en búsqueda de eficiencia en el proceso y de realizar más trabajo interno que externo</p> <p>Si, pero es garantizar que los costos de los proyectos son los reales del mismo</p>
<p><b>3. Problemas típicos del cargo</b></p> <ol style="list-style-type: none"> <li>1. la no corrección de fechas de cronogramas</li> <li>2. No coinciden los indicadores con las estadísticas medidas</li> <li>3. Cuando se va a realizar la relación de costos, al final no coinciden con lo inicialmente propuesto porque los parámetros iniciales los cambiaron.</li> <li>4. Problemas con los documentos</li> </ol>	<p>yo diría que el problema es la variación de las fechas de cronograma sin consultar a las áreas afectadas</p> <p>No entendí este problema, se refiere a que los indicadores estén por debajo de la meta?</p> <p>Esta enunciado como un hallazgo, pero es un riesgo que debemos mitigar desde el área</p> <p>No es un problema si no genera costos</p>
<p><b>4. Relaciones con otras áreas</b></p> <p>Editorial                      Diseño                      Mercadeo                      Costos</p>	



**Anexo 4. Guía para la entrevista hecha al trabajador del conocimiento durante el proceso de validación**

1. Si compara cada uno de los perfiles con el original enviado por recursos humanos el perfil es:

- Es mucho mejor
- Es mejor
- Es más o menos igual
- Es peor
- Es muchísimo peor

En qué aspectos es mejor y cuáles peor: \_\_\_\_\_

2. Encuentra alguna función que no estaba contenida en el perfil original?

- Si
- No

Cuál? \_\_\_\_\_

Observe el gráfico de relaciones:

Usted identifica que esas son las áreas? Son los % de relación correctos? Le sorprende estos porcentajes?

3. Indique sobre la escala cada uno de los formularios entregados . Organice inicialmente de peor a mejor incluyendo el perfil de cargo que le entrega Gestión Humana.

Mala 1 2 3 4 5 6 7 8 9 10 11 Excelente