Extracción de Funciones de un Cargo usando Minería de Texto en Correos Electrónicos

Jorge A. Alvarado y Constanza Cuervo

Pontificia Universidad Javeriana, Carrera 7 #40-62, Ed. José Gabriel Maldonado, Piso 3, Bogotá-Colombia
(e-mail: jorge.alvarado@javeriana.edu.co; constanza.cuervo@javeriana.edu.co)

Recibido Mar. 13, 2013; Aceptado Abr. 29, 2013; Versión final recibida May. 15, 2013

Resumen

Se presenta el desarrollo de una metodología para extraer, de un grupo grande de correos electrónicos pertenecientes a un trabajador con una definida función de cargo, una muestra de correos electrónicos que sea representativa de las funciones del cargo del trabajador. Esto se hace a través de herramientas de procesamiento del lenguaje natural y de minería de datos. Los resultados muestran que la metodología permite documentar el 65% de las funciones en la fase de evaluación. Por lo tanto esta puede ser usada con relativa confiabilidad para la recuperación de información de trabajadores que dejan su cargo de forma intempestiva, y para entrega de cargos por rotación del puesto de trabajo o ascensos, facilitando la comunicación entre las personas que realicen el empalme de cargos

Palabras clave: minería de texto, funciones de cargo, correo electrónico, procesamiento de lenguaje natural

Extraction of Functions of a Job Position using Text Mining in Electronic Mail

Abstract

A methodology to extract, from a group of electronic mails belonging to workers with defined responsibilities, a sample of electronic mails that are representative of the tasks assigned to the workers. This is done through natural language processing tools and data mining. The results showed that the methodology represents a good approach for extracting the functions of the workers, allowing 65% of the work descriptions in the evaluation stage. Thus, the proposed methodology can be used with relative confidence to gathered information about workers that suddenly quit their job positions and also for position changes and promotions.

Keywords: text mining, position responsibilities, electronic mail, natural language processing

INTRODUCCIÓN

Una de las tareas más importantes en las organizaciones es la descripción clara de las funciones de los cargos y las relaciones de dichos cargos dentro del entorno laboral. La función es entendida como la tarea que corresponde realizar a una institución o entidad, o a sus órganos o personas. 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. 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 gestión del conocimiento (Capó-Vicedo et al., 2007); 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 et al., 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.

Para el estudio de las funciones de cargo, la práctica más usual es la elaboración de hojas de perfil. Comúnmente estas hojas de perfil están basadas en encuestas que se realizan al trabajador sobre lo que éste hace, cuándo y cómo (Chase et al., 2000). Las metodologías tradicionales para documentar funciones presentan algunas dificultades. En primer lugar, implican un tiempo considerable en entrevistas y un conocimiento previo del cargo. Por lo tanto es necesario un entrenamiento previo arduo para las personas a las que se les encomienda esta labor, lo cual restringe el número de personas que pueden ser examinadas en un momento dado (Robinson, 2010).

En segundo lugar, toda recolección de información basada en auto-reportes tiende a ser sesgada. Este tipo de instrumento depende de la buena voluntad del empleado, de la capacidad de su memoria y de la importancia que le otorgue a cada asunto. En consecuencia, los datos derivados de estos informes pueden verse comprometidos por: problemas de datos faltantes (Fox-Wasylyshyn y El-Masri, 2005); sesgos cognitivos, incluida la valoración del tiempo (Barrero et al., 2009); y, por último, los participantes podrían tener la intención de tergiversar la información relativa a su cargo. La consecuencia que esto conlleva es que se generan sesgos en los reportes y se tiene información incompleta u oculta.

Las carencias de las metodologías tradicionales de recolección de funciones resultan más problemáticas aun cuando los empleados son trabajadores del conocimiento (TC) (Hammer et al., 2004) porque los métodos de recolección de funciones fueron establecidos originalmente para trabajadores operativos. Describir las tareas de los TC es más difícil, porque sus flujos de trabajo no son necesariamente repetitivos. El conocimiento de los TC 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 y McFadyen, 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. En muchos negocios, el texto sin formato puede llegar a representar hasta el 80% de los datos de la organización (Yu et al., 2005).

El correo electrónico es una de las herramientas más utilizadas por los TC (Hemp, 2009). Por tanto, es una fuente de conocimiento empresarial que puede ser analizada mediante metodologías para la extracción de conocimiento empresarial (Matos, Chalmeta, y Coltell, 2006). 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 y Wortmann, 2012). 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 y Lau, 2005). Como consecuencia, una buena fuente de información para reconocer las actividades y funciones que realiza un trabajador del conocimiento es su correo electrónico, pues es el medio más común en el que escribe.

Dado que el correo electrónico maneja información textual desestructurada, la Minería de Texto (MT), puede ser una herramienta útil para realizar un procesamiento del lenguaje que facilite la tarea de la extracción de las responsabilidades o funciones de un cargo (Feldman y Sanger, 2007). Aunque el análisis de textos se ha aplicado en una amplia gama de campos (Belhadjali et al., 2010), a la fecha y de acuerdo con el estado del arte realizado, no se encontraron aplicaciones de MT en la extracción de funciones en puestos de trabajo. El aporte principal de esta investigación es el uso de herramientas automatizadas para simplificar la tarea de extracción de las funciones de un cargo seleccionando un grupo representativo de correos electrónicos mediante técnicas de MT y procedimientos automáticos de análisis de texto desestructurado.

La utilización de MT 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 MT orientadas a la detección de fraudes y al hallazgo de delitos informáticos han sido exitosas (Chen et al., 2011; Holton, 2009; Thompson, 2005) particularmente para identificar al autor más plausible de un correo electrónico anónimo de un grupo de posibles sospechosos y para la investigación de la ciberdelincuencia y la recopilación de pruebas, como en los trabajos de Iqbal et al. (2008). Estas investigaciones buscan detectar un fraude, pero no extraen las actividades declaradas que realizan los autores de los correos electrónicos.

Por otra parte, las aplicaciones relacionadas con servicio al cliente pretenden apoyar esas áreas de la empresa, aliviando su labor mediante métodos de filtro y predicción, como las investigaciones realizadas por Sakurai y Suyama (2005) o la investigación de Weng y Liu, (2004). En este caso es la detección y clasificación de quejas y reclamos, y no de actividades, el foco principal.

Existen también investigaciones referidas a las interacciones sociales halladas en el correo electrónico. Por ejemplo, el caso de Enron analizado por Kang et al. (2010), es ampliamente conocido y está basado en la detección de relaciones sociales y que incluye adicionalmente análisis de sentimientos y clima organizacional. Tyler et al., (2005) estudiaron temas de autoridad y liderazgo a partir del correo electrónico. Un paso más adelante hacia la elaboración de funciones lo dan el trabajo de Stuit y Wortmann (2012) -donde se construye un método para reconocer interacciones, atributos y roles en un proceso-, y el trabajo de Leuski (2004) donde se identifica el rol que desempeña el trabajador, según el contexto, al separar los correos electrónicos personales de los profesionales. Sin embargo en ninguna de estas investigaciones se llega a una descripción de las funciones del cargo.

La MT ha resultado útil para la detección de relaciones sociales y personales claves a partir del correo electrónico y es así como podría ser utilizada para detectar las principales relaciones y funciones laborales de un cargo dado. Sin embargo no hemos encontrado una investigación dirigida a la extracción de funciones a partir de un análisis de MT del correo electrónico ni tampoco la hemos hallado en el idioma español.

Por estas razones consideramos que el principal aporte de nuestro artículo radica en una primera aproximación para facilitar la tarea de extracción de funciones laborales basada en MT y procesamiento del lenguaje del correo electrónico en español, ayudando a seleccionar correos electrónicos clave para el entendimiento de las funciones de un cargo.

METODOLOGIA

La metodología utilizada se basó en el Modelo CRISP-DM para el desarrollo de un proyecto. El CRISP-DM fue creado en 1996 en el campo de la minería de datos (Chapman et al., 2000). Es importante aclarar que las seis fases del proyecto de minería de datos no son rígidas. Por el contrario, el movimiento hacia adelante y hacia atrás entre fases diferentes es siempre requerido (Franco et al., 2012) 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. Lo anterior ha hecho que la metodología planteada aquí se haya modificado varias veces desde la propuesta inicial y se espera que sea refinada aún más en el inmediato futuro (Dueñas-Reyes, 2009).

La metodología consiste en la extracción de un subconjunto c de la colección de documentos C , en donde cada documento representa un correo electrónico. Este subconjunto está estratificado de acuerdo a las áreas de la empresa contactadas, y a la dirección o sentido del correo (enviado o recibido) y contiene una gran riqueza informativa relativa a las actividades cuya realización se menciona en el correo electrónico. La figura 1 muestra los pasos seguidos para la extracción del subconjunto c . Dichos pasos se describen a continuación.

Preparación de datos

Tras obtener el archivo que contenga la información del correo electrónico del periodo de tiempo que se desea evaluar, la base de datos se enriquece con el área de la organización (interna o externa) a la cual pertenece cada correo electrónico que el cargo remite o recibe, así como un indicador mostrando si el correo es entrante o saliente. Esta información, analizada estadísticamente y en conjunto con la información temporal del correo (mes, día, hora) permite crear como subproducto un análisis de redes sociales del cargo- con sus principales áreas conectadas y sus cargas de trabajo en el tiempo- que también se suele incluir en los perfiles, pero que no hace parte del análisis de datos no estructurados y por lo tanto no se reporta aquí en detalle.

Fig. 1: Metodología propuesta para la extracción de funciones de un cargo

Procesamiento del lenguaje

El procesamiento del lenguaje incluye las siguientes etapas para cada correo electrónico:

- 1-Limpieza de “colas” en los correos, correspondientes a los mensajes reenviados, las firmas y los mensajes de seguridad automáticos.
- 2- Tokenización por palabra, lematización y etiquetado de palabras con la parte de la oración.
- 3- Eliminación de palabras vacías
- 4- Reducción de la base de datos a verbos no auxiliares y sustantivos sin nombres propios. La razón de esta reducción radica en que se consideran las actividades como parejas sustantivo-verbo que representan acciones sobre objetos o conceptos específicos.

Las etapas 1, 3 y 4 se programaron en el software SAS; para la etapa 2 se utilizó el software *Treetagger*, herramienta desarrollada por Schmid(1994).

Reglas de aprendizaje por asociación

La creación de reglas de aprendizaje se hace siguiendo el método propuesto por Agrawal et al.(1993), el cual ha sido ampliamente usado en sistemas de recomendación (Su y Khoshgoftaar, 2009) y en análisis de carritos de mercado(Hoanca y Mock, 2011). Para el desarrollo de esta metodología, el conjunto de transacciones T se corresponde con una colección de documentos específica (en este caso, correos electrónicos). Los ítems de cada transacción I_t corresponden a las palabras extraídas en el punto 3.2 para la colección de transacciones T . Los conjuntos X e Y son subconjuntos de I_t . El conjunto X contiene los sustantivos y el conjunto Y contiene los verbos.

El soporte para cada posible par de palabras (x,y) se define como el número de correos electrónicos que contienen ese grupo de palabras como porcentaje del número de correos electrónicos en la colección de documentos.

$$\text{Soporte } (x,y)=n(x \cap y)/n(T) \quad (1)$$

La confianza mide la coocurrencia de los ítems x e y dado que ha ocurrido el ítem x .

$$\text{Confianza}(x \rightarrow y)= n(x \cap y)/n(x) \quad (2)$$

Finalmente se calcula el impulso, el cual indica si la co-ocurrencia de la pareja de palabras (x,y) en los correos electrónicos es superior, igual o inferior a lo esperado de una repartición de palabras de forma independiente y aleatoria.

$$\text{Impulso}(x \rightarrow y)= \text{Confianza}(x \rightarrow y)*n(T)/n(y) \quad (3)$$

Cuando el impulso es mayor que 1, se puede decir que las palabras se presentan juntas más frecuentemente de lo esperado. En cambio, si es menor que 1, las palabras se presentan juntas en menor frecuencia de lo esperado.

Se crearon 10 subconjuntos de correos de la colección C , dividiéndolos por cada una de las 5 áreas de la organización con las que se tenía mayor frecuencia de intercambio de correos, tanto en enviados como recibidos, llegando a tener 10 subconjuntos C_{ij} , siendo i el área y j un indicador booleano que muestra si el correo es enviado o recibido. El proceso de reglas de asociación se ejecutó independientemente para cada uno de los C_{ij} . De estos grupos de reglas de asociación para cada C_{ij} se escogieron de manera empírica aquellas reglas para parejas (x, y) con *soporte* superior al 1% e *impulso* superior a 1, llevándolas a un conjunto de reglas candidatas R_{ij} para cada uno de los 10 grupos.

Extracción de correos representativos

En esta etapa se obtuvieron, para cada correo electrónico en C_{ij} , el número de reglas candidatas contenidas en R_{ij} que también estaban presentes en el respectivo correo electrónico. Posteriormente, los correos en C_{ij} se ordenaron de mayor a menor por frecuencia de reglas candidatas presentes. Finalmente, se seleccionaron los correos con mayor frecuencia de reglas candidatas para cada uno de los C_{ij} para conformar el subconjunto c final, bajo la suposición de que estos correos contenidos en c resumen de manera densa la actividad que es propia del cargo.

Generación de las funciones

Los correos extraídos se sometieron a evaluación humana para extraer las funciones del cargo, indicando a las personas a qué área pertenecía cada correo, y si se trataba de un correo enviado o recibido para que ellas procedieran a redactar las funciones.

RESULTADOS

La empresa donde se realizó 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 a nivel latinoamericano. Tiene un 30% del mercado en los productos que ofrece. En Colombia tiene presencia en las principales ciudades y cuenta con una amplia infraestructura propia. Las dos personas que prestaron su correo laboral tienen cargo de coordinador. La primera pertenece al área de compras que en el momento de la investigación pertenecía al área de compras editoriales y esta área a su vez pertenecía al área de operaciones, y su correo electrónico durante un año consta de 21.370 mensajes de correo electrónico.

Tras efectuar el proceso descrito en la sección 3, se obtuvieron parejas de palabras candidatas, de las cuales un ejemplo se muestra en la tabla 1, donde frente a cada sustantivo aparece el verbo que resultó asociado al mismo, junto con su soporte e impulso. Se muestran aquellas parejas de palabras que obtuvieron el mayor impulso dentro de todos los correos electrónicos. A manera de ejemplo, a partir de la asociación de palabras presentada, una persona cercana al cargo extrajo como posible función la de confirmar y enviar especificaciones de libros, la cual es efectivamente una de las funciones del cargo. Sin embargo, resultó evidente que se requería que la persona tuviera conocimiento del dominio específico del cargo para deducir sus funciones sólo a partir del análisis de asociaciones.

Para el ejemplo presentado, se requiere saber que las palabras solapa, tinta, papel, e hilo con el verbo acabar se refieren a la elaboración de un libro. En MT resulta claro que en la construcción de las relaciones entre las entidades es importante el uso de un dominio específico de conocimiento (Feldman y Sanger, 2007). Por ello se decidió agregar a la metodología el paso de extracción de correos representativos, pensando que los correos completos facilitarían que personas con menor conocimiento del área del cargo pudieran extraer las funciones. Los resultados para el primer caso fueron prometedores, pero era necesario someterlos a una validación rigurosa con un cargo que fuese desconocido para los generadores de las funciones. Por ello se decidió realizar una validación con un segundo cargo que permitiera revisar los resultados logrados en el primer caso de estudio.

La segunda persona pertenece al área editorial y hace referencia a la coordinadora del proceso editorial que pertenece a la Gerencia de Contenidos y Soluciones Educativas. Sus datos textuales constan de 5.380 correos electrónicos. Tras ejecutar la metodología, se seleccionaron empíricamente 7 correos por área, para un total de 35 correos que deben ser leídos por los generadores de funciones (0.6% del total de documentos de la colección). Adicional a ello, se extrajeron 35 correos electrónicos aleatorios divididos por las cinco áreas para generar una línea de base ingenua. De ese modo, se tuvieron dos grupos de correos a ser evaluados.

Tabla 1: Ejemplo de parejas de palabras candidatas

Sustantivo	Verbo	soporte	impulso
cal	calentar	0,016	31,82
cartulina	calentar	0,016	31,45
solapa	calentar	0,016	30,39
tinta	calentar	0,014	29,51
solapa	acabar	0,031	20,76
cartulina	acabar	0,029	20,27
cal	acabar	0,029	20,26
hilo	acabar	0,010	17,91
formato	calentar	0,017	17,65
tinta	acabar	0,022	17,23
papel	calentar	0,017	16,63

Dos personas externas a la organización y sin conocimiento del cargo hicieron el papel de generadores de funciones. Se les informó el nombre del cargo, y el nombre de la empresa y su objeto social. A continuación se les pidió leer los dos grupos de correos y a partir de ellos elaborar la Misión del cargo, las actividades del cargo y los problemas típicos del cargo, puesto que estas tres solicitudes hacen parte de los formatos de perfiles de cargo que usualmente se elaboran en la empresa. Uno de ellos leyó primero los correos aleatorios y el otro primero los correos seleccionados por la metodología.

Las actividades elaboradas por los generadores de funciones fueron comparadas una a una contra las funciones del perfil del cargo que el área de Recursos Humanos había elaborado, indicando si coincidían completamente (1 punto), parcialmente (medio punto) o no estaban presentes (0 puntos). Los correos generados por la metodología obtuvieron en promedio el 65% de las funciones, mientras que las funciones generadas por correos aleatorios obtuvieron en promedio el 45% de las mismas funciones, de un total de 10 funciones.

Entre las funciones que se identificaron basadas en la metodología y no en los correos aleatorios se encuentran actividades muy precisas como “Participar en el análisis de viabilidad de proyectos de Colombia. Orientar y hacer seguimiento al análisis de viabilidad de los países” o “Presupuestar y calcular la distribución de planta de Contenidos y Soluciones Educativas de Colombia y monitorear sus cambios”. Entre las funciones no identificadas por ninguna de las metodologías se encuentra una que representa más una actitud (Ser flexible a los procesos de cambio) y otra referida a temas de capacitación empresarial (Identificar necesidades de capacitación para los equipos editoriales). Ello muestra la habilidad de la metodología propuesta para identificar funciones muy específicas del cargo, pero también su dificultad para detectar acciones de segundo orden relacionadas con el dinamismo y mejora de un puesto de trabajo de un TC.

A la persona actual que se encuentra en el cargo se le presentaron las funciones generadas por la metodología. A este TC se le realizó una entrevista abierta semiestructurada para conocer su opinión acerca de las funciones generadas. En particular, el trabajador destacó que en las funciones generadas encontró información novedosa, sorpresiva y verdadera acerca de su cargo, y evidenció que las actividades generadas por la metodología resumían gran parte de las funciones más importantes del cargo pero que la descripción de la misión del cargo se quedaba corta en su alcance. Así mismo destacó actividades que no se evidenciaban por realizarse de manera telefónica y personal. Destacó que el proceso de levantamiento de sus funciones por parte de Recursos Humanos tomó más de una semana, mientras que con una hora del generador de funciones y una hora de entrevista con él fue posible tener una idea de su cargo igual o mejor a la obtenida anteriormente.

CONCLUSIONES

La metodología desarrollada durante este trabajo presenta características que la hacen viable y funcional como una primera aproximación para la extracción de funciones de un cargo. Logra además 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 o un proceso de levantamiento de

funciones se reduce a un proceso de síntesis automatizado y un perfilamiento final. En el segundo caso, la extracción automática de información permite darle una mirada diferente a un cargo, disminuyendo los sesgos del auto-reporte y descubriendo actividades que por olvido u ocultamiento no se encuentran en los perfiles tradicionales.

Si bien se puede pensar que la metodología no es mucho mejor que la línea de base con correos aleatorios, resulta claro que los resultados basados en la metodología son menos variables y garantizan el hallazgo de correos densos- en el sentido de que incluyen relaciones de palabras que son significativas para el cargo-.

En su estado actual, la metodología es funcional para extraer funciones específicas o tácticas de un cargo. Sus aplicaciones actuales están en la recuperación de información de trabajadores que dejan su cargo de forma intempestiva, y para entregas de cargos por rotación del puesto de trabajo o ascensos, facilitando la comunicación entre las personas que realicen el empalme de cargos.

REFERENCIAS

Agrawal, R., Imielinski, T., y Swami, A., *Mining association rules between sets of items in large databases*, Proceedings of the 1993 SIGMOD Conference, 22(2), 207-216 (1993).

Barrero, L. H., Katz, J. N., y Dennerlein, J. T., *Validity of self-reported mechanical demands for occupational epidemiologic research of musculoskeletal disorders*, Scandinavian Journal of Work, Environment and Health, 35(4), 245-260 (2009).

Belhadjali, M., Abbasi, S. M., y Whaley, G. L. A., *Text Mining Approach to the Health Care Debate*, Competition Forum, 8(2), 260-264 (2010).

Capó-Vicedo, J., Tomás-Miquel, J. V., y Expósito-Langa, M., *La Gestión del Conocimiento en la Cadena de Suministro: Análisis de la Influencia del Contexto Organizativo*, Información Tecnológica, 18(1), 127-136 (2007).

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., y Rudiger, W., *CRISP-DM 1.0: Step by step data mining guide* (en línea), (2000).
<ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf>. Acceso: 5 de mayo de 2012.

Chase, R. B., Aquilano, N. J., y Jacobs, F. R., *Administración de Producción y Operaciones*, 8ª edición, Mc Graw Hill, México, (2000).

Chen, H., Chau, M., y Li, S. H., *Enterprise risk and security management: Data, text and Web mining*, Decision Support systems, 50(4), 649-650 (2011).

Dueñas-Reyes, M. X. *Searching for true information with spatial data mining*. Ingeniería y Universidad, 13(1), 137-156 (2009).

Feldman, R., y Sanger, J., *The Text Mining Handbook: advanced approaches in analyzing unstructured data*. 1ª edición, Cambridge University Press, Cambridge, USA (2007).

Fox-Wasylyshyn, S. M., y El-Masri, M. M., *Handling missing data in self-report measures*. Research in Nursing y Health, 28(6), 488-495 (2005).

Franco, D. A., Perea, J. L., y Puello, P., *Metodología para la Detección de Vulnerabilidades en Redes de Datos*. Información Tecnológica, 23(3), 113-120 (2012).

Hammer, M., Leonard, D., y Davenport, T. *Why don't we know more about knowledge?*, Mit Sloan Management Review, 45(4), 14-18 (2004).

Hemp, P., *Death by information overload*. Harvard business review, 87(9), 82-88, (2009).

Henard, D. H., y McFadyen, M. A., *Making knowledge workers more creative*, Research-Technology Management, 51(2), 40-46 (2008).

Hoanca, B., y Mock, K. *Using Market Basket Analysis to Estimate Potential Revenue increases for a Small University Bookstore*. Conference for Information Systems Applied Research, 1822-1833, Wilmington, NC, USA, 3 a 6 de Noviembre de (2011).

- Holton, C., *Identifying disgruntled employee systems fraud risk through text mining: A simple solution for a multi-billion dollar problem*, Decision Support Systems, 46(4), 853-864 (2009).
- Iqbal, F., Hadjidj, R., Fung, B. C. M., y Debbabi, M., *A novel approach of mining write-prints for authorship attribution in e-mail forensics*. Digital Investigation, 5, S42-S51 (2008).
- Kang, H., Plaisant, C., Elsayed, T., y Oard, D. W., *Making Sense of Archived E-mail: Exploring the Enron Collection With NetLens*, Journal of the American Society for Information Science and Technology, 61(4), 723-744 (2010).
- Kushmerick, N., y Lau, T. *Automated email activity management: an unsupervised learning approach*. Paper presented at the Proceedings of the 10th international conference on Intelligent user interfaces, San Diego, California, USA. (2005).
- Leuski, A., *Email is a stage: discovering people roles from email archives*. Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, 502-503 Sheffield, Reino Unido, Julio 25 a 29 de (2004).
- Matos, G., Chalmeta, R., y Coltell, O., *Metodología para la extracción del conocimiento empresarial a partir de los Datos*, Información Tecnológica, 17(2), 81-88 (2006).
- Parise, S., Cross, R., y Davenport, T. H., *Strategies for preventing a knowledge-loss crisis*, Mit Sloan Management Review, 47(4), 31-41(2006).
- Robinson, M. A., *Work Sampling: Methodological Advances and New Applications*. Human Factors and Ergonomics in Manufacturing and Service Industries, 20(1), 42-60 (2010).
- Sakurai, S., y Suyama, A., *An e-mail analysis method based on text mining techniques*. Applied Soft Computing, 6(1), 62-71 (2005).
- Schmid, H., *Probabilistic Part-of-Speech Tagging Using Decision Trees*. Proceedings of International Conference on New Methods in Language Processing, 44-49, Manchester, Reino Unido, 14 a 16 de Septiembre de (1994).
- Stuit, M., y Wortmann, H., *Discovery and analysis of e-mail-driven business processes*, Information Systems, 37(2), 142-168 (2012).
- Su, X., y Khoshgoftaar, T. M., *A survey of collaborative filtering techniques*, Advances in Artificial Intelligence, 2009(4), 2-2 (2009).
- Thompson, P., *Text mining, names and security*, Journal of Database Management, 16(1), 54-59 (2005).
- Tyler, J. R., Wilkinson, D. M., y Huberman, B. A., *E-mail as spectroscopy: Automated discovery of community structure within organizations*, Information Society, 21(2), 133-141(2005).
- Weng, S. S., y Liu, C. K., *Using text classification and multiple concepts to answer e-mails*. Expert Systems with applications, 26(4), 529-543 (2004).
- Yu, L., Wang, S., y Lai, K. K., *A rough-set-refined text mining approach for crude oil market tendency forecasting*. International Journal of Knowledge and Systems Sciences, 2(1), 33-46(2005).