

Desafíos y oportunidades de la minería de datos/texto en la selección de personal

Challenges and opportunities for data/text mining in personnel selection

PÉREZ-RAVE, Jorge I.¹

JARAMILLO-ÁLVAREZ, Gloria P.²

CORREA-MORALES, Juan C.³

Resumen

Se provee una revisión de literatura sobre la minería de datos/texto (M-D/T) en la selección de personal (SP). La localización y selección de los estudios consta de dos etapas. La primera es de tipo sistemático y la segunda es una búsqueda tradicional. Este estudio retrata y argumenta tres desafíos para la M-D/T en la SP. También propone integrar la M-D/T con modelos de ecuaciones estructurales y análisis multicriterio para abordar los retos identificados.

Palabras clave: selección de Personal, Minería de Datos, Minería de Texto, Modelos de Ecuaciones Estructurales, Análisis de Decisión Multicriterio.

Abstract

This study provides a literature review on data/text mining (D/T-M) in personnel selection (PS). The location and selection of the studies consider two stages. The first is a systematic search, and the second is a traditional search. This study portrays and argues three challenges for D/T-M in PS. It also proposes integrating D/T-M with structural equation models and multicriteria decision analysis to address the identified challenges.

key words: personnel Selection, Data Mining, Text Mining, Structural Equation Modelling, Multicriteria Decision Analysis

1. Introducción

La gestión del talento humano comprende variedad de funciones, entre ellas: la planificación de necesidades de personal, su captación, desarrollo continuo, compensación y evaluación del desempeño (Strohmeier, & Piazza, 2013). Por lo mismo, es reconocida como un factor estratégico para las organizaciones, que ayuda a generar ventajas sostenidas (Chien, & Chen, 2008; Koutra et al., 2017). Uno de los dominios más importantes de la gestión del capital humano es la selección de personal (SP), también catalogado como un factor crítico de éxito, dado

¹ Director grupo de investigación IDINNOV. IDINNOV S.A.S. investigacion@idinnov.com

² Docente. Departamento de Ciencias de la Computación y la Decisión. Facultad de Minas. Universidad Nacional de Colombia, Medellín. gpjarami@unal.edu.co

³ Docente. Escuela de Estadística. Universidad Nacional de Colombia, Medellín. jccorrea@unal.edu.co

que determina los niveles de calidad inicial del talento humano contratado y fija referentes para su posterior desarrollo (Chavira et al., 2017; Shehu, & Saeed, 2016; Strohmeier, & Piazza, 2013). Dicha selección puede entenderse como el proceso de elegir los mejores individuos, tomando en consideración las cualificaciones duras y blandas requeridas para llevar a cabo el trabajo de forma efectiva (Koutra et al., 2017; Zhang, & Liu, 2011). Así, por medio de un proceso de selección efectivo, la organización aumenta las probabilidades de elegir los talentos más acordes con sus necesidades, retos y expectativas, lo cual puede traducirse en individuos que permanezcan en la organización, ejecuten sus tareas con efectividad y responsabilidad, y adopten comportamientos innovadores (Werner, 2000).

La SP, además de ser un proceso relevante por sus implicaciones para la dirección de organizaciones, también es reconocida como un tema de investigación complejo, al que se suman cada vez más áreas académicas, entre ellas, la minería de datos (MD) y la minería de texto (MT). Estas dos áreas se diferencian principalmente en el tipo de datos abordados (MD: estructurados, MT: textos) y en algunos métodos de preparación y análisis. No obstante, ambas pueden entenderse como un proceso no trivial que utiliza métodos de exploración y análisis automático (o semi-automático) para extraer patrones potencialmente útiles desde los datos (Fayyad, Piatetsky-Shapiro, & Smyth, 1996; Ledolter, 2013). Para esto, hacen uso de varios campos, entre ellos: estadística, inteligencia artificial y visualización de datos.

A pesar de que la minería de datos/texto (M-D/T) en la SP es un campo de investigación con potencial, aún se reconoce incipiente (Chien, & Chen, 2008; Shehu, & Saeed, 2016; Strohmeier, & Piazza, 2013). Apenas en el año 2013 se publicó la primera revisión sistemática acerca de la MD en la gestión del talento humano (Strohmeier, & Piazza, 2013) y esta expuso a la SP como uno de los temas de mayor auge. Las pocas aplicaciones de M-D/T en la SP han ayudado a descubrir patrones exploratorios, automatizar tareas y proveer insumos para la toma de decisiones. Sin embargo, el enfoque de las soluciones de M-D/T en la gestión del talento humano en general, se ha centrado más en el uso o combinación de métodos (*method-driven data mining*), que en la realidad del problema (Strohmeier, & Piazza, 2013). Por consiguiente, la mayor parte de las contribuciones dejan serias dudas sobre la relevancia y utilidad para las aplicaciones reales, lo cual ha motivado un enfoque más emergente, conducido por el dominio del problema (*domain-driven data mining*) (Adejuwon, Mosavi, 2010; Koutra et al., 2017; Strohmeier, & Piazza, 2013). A pesar de este llamado en la gestión del talento humano, la evidencia empírica al respecto es escasa y, más aún, en el dominio de la SP, el cual ofrece complejidades adicionales.

En Strohmeier, & Piazza (2013), revisión sistemática de 100 estudios sobre MD (y también se refirió a datos de texto), en el contexto de la gestión del talento humano, se argumenta la poca evidencia sobre la superioridad de las soluciones de M-D/T, en comparación con métodos validados en tal dominio. En el caso de la SP, vale mencionar: pruebas cognitivas, entrevistas y test psicométricos (Chien, & Chen, 2008; Evers, Anderson, & Smit-Voskuil, 2005). Esta falta de evidencia para legitimar el éxito de la M-D/T en la SP, también es declarada en limitaciones/desafíos de estudios como Han, & Lee (2016).

La poca evidencia para legitimar el éxito funcional de la M-D/T en la SP (Strohmeier, & Piazza, 2013) tiene repercusiones para la investigación científica, la docencia y la dirección de organizaciones. Concerniente a la investigación sobre M-D/T en la SP, porque se desvía de su connotación aplicada y podría estar subestimando (o sobreestimando) las potencialidades de sus efectos en tal dominio, lo cual deja serios “huecos de conocimiento” para la intención de “verdad” que promulga la ciencia. Respecto a la docencia, porque llevaría a aumentar la brecha errónea entre la teoría y la práctica. Con relación a la dirección de organizaciones, porque podrían frenarse los esfuerzos de desarrollo de capacidades analíticas diferenciadas, complejas y difíciles de imitar, que conduzcan a ventajas competitivas a partir de tal dominio (Su, & Linderman, 2016; Teece, Pisano, & Shuen, 1999). En otras palabras, se derivan altos costos monetarios y sociales para individuos y organizaciones, ante la incertidumbre de cómo proveer soluciones integrales de M-D/T que permitan abarcar el proceso completo de la

SP, el cual amerita considerar desde la planificación y preparación del dato, hasta la identificación de patrones relevantes, fiables y válidos, la evaluación y ordenación de candidatos, y la elección final.

Con miras a comprender y aportar soluciones a este desafío global, resumido en la necesidad de materializar el enfoque de M-D/T guiado por el dominio de la SP, este artículo de revisión aborda el siguiente objetivo: identificar desafíos específicos y proponer alternativas de integradoras alrededor de la M-D/T conducida por el dominio de la SP.

Este artículo se estructura en cinco secciones. Esta primera aportó la introducción. La segunda expone un breve marco conceptual de apoyo. La tercera resume la estrategia de recuperación y selección de documentos. La cuarta provee los desafíos específicos inducidos desde la lectura, la interpretación y el análisis de documentos, y, además, propone dos oportunidades para superar dichos desafíos. La quinta sección plasma las conclusiones generales.

2. Conceptos sobre SP

Este apartado no pretende ser exhaustivo en el tema de la SP. Sin embargo, provee conceptos de base que facilitan la comprensión del resto del artículo y que mínimamente debería reconocer un profesional o investigador que busque contribuir, desde la M-D/T, al dominio de la SP, el cual tiene amplio desarrollo a nivel teórico y empírico. En otras palabras, para lograr una solución de M-D/T guiada por el dominio de la SP, no basta con el uso de conceptos y métodos clásicos de la M-D/T, sino que estos deben articularse razonablemente con elementos psicológicos y psicométricos para la SP; lo cual de por sí representa una tarea compleja.

2.1. Análisis de diferencias individuales para la SP

La esencia de la SP es el análisis de las diferencias individuales, con la intención de retratar razonablemente el comportamiento futuro del individuo en el trabajo. Tales diferencias pueden asumirse desde cuatro aspectos: habilidades, personalidad, intereses y autovaloraciones del individuo (Murphy, 2012). Las habilidades describen algo que el individuo hace (no solamente limitada a un atributo que se tiene) para cumplir tareas y resolver problemas (Murphy, 2012). Estas se dividen en habilidades cognitivas y físicas. Las cognitivas tienen que ver con las operaciones mentales para el procesamiento de información. A diferencia de las habilidades físicas, las cognitivas presentan alta relación entre sí (ej: comprensión lectora vs razonamiento lógico) (Murphy, 2012). Al desarrollo de las habilidades se le asocia una componente heredable y otra ambiental (Dickens, & Flynn, 2001; Plomin, & Rende, 1991). La personalidad hace referencia a patrones de comportamiento (Murphy, 2012), lo que se traduce en “formas consistentes de pensamientos, sentimientos y conductas que permiten caracterizar a los seres humanos” (Alonso, Moscoso, & Cuadrado, 2015, p. 81). Los intereses se entienden como respuestas de gusto a personas, cosas o eventos en particular (Strong, 1943). También pueden verse como una respuesta afectiva aprendida a un determinado objeto o actividad, donde a mayor interés, mayor respuesta positiva afectiva y, ante desinterés, más indiferencia o aversión (Murphy, 2012). Las autovaloraciones (o autoevaluaciones) son creencias del individuo sobre sí mismo, las cuales facilitan o inhiben la expresión de habilidades, personalidad e intereses en un amplio rango de situaciones (Murphy, 2012). También son entendidas como juicios amplios que la gente hace sobre su propio valor y su habilidad para influenciar eventos (Judge, 2009).

La medición de las habilidades (cognitivas), la personalidad, los intereses y las autovaloraciones parten de reconocer que estos elementos son constructos psicológicos. Es decir, variables latentes que no pueden ser medidas de forma directa sino a través de variables observables, también llamadas indicadores, ítems o reactivos. Así, un constructo (Ej: liderazgo, extroversión, responsabilidad, compromiso) se entiende como el factor causal de una serie de manifestaciones observables (mirada reflectiva del constructo). El modelo clásico

de evaluación de diferencias individuales para la SP asume una base predictiva entre estas diferencias y el desempeño laboral; es así como la evaluación de las diferencias individuales ayuda a entender e inferir gran parte de las diferencias en el desempeño laboral de los individuos (Binning, & Barrett, 1989; Ployhart, & Schneider, 2012; Murphy, 2012).

2.2. Propiedades fundamentales para la medición de diferencias individuales en la SP

En vista de que el análisis de diferencias individuales se realiza fundamentalmente a través de variables latentes (constructos psicológicos), es determinante el estudio de varias propiedades mínimas que deben superar los instrumentos (test) empleados para ello. Una de ellas es la validez, entendida como el máximo de los principios y hace referencia al grado en el cual el uso que se pretende hacer de las puntuaciones del test está justificado, tanto a nivel teórico como empírico (Martínez, Hernández, & Hernández, 2006). Hoy se reconoce esta propiedad como un concepto unitario, representado por medio de distintas evidencias (Sacket, Putka, & McCloy, 2012; Martínez, Hernández, & Hernández, 2006), las cuales dan lugar a diversas etiquetas, siendo las dos más relevantes, las siguientes: Validez de contenido (tipo de evidencia cualitativa, que representa el grado de exhaustividad de un test o medida para cubrir todo el dominio del contenido de lo que se intenta medir; Yaghmale, 2003). Validez de constructo (grado en que la medida de un determinado constructo evalúa lo que se supone que mide; Ployhart, & Schneider, 2012). Respecto a esta última, se espera que las medidas entre sí reflejen convergencia hacia el constructo subyacente, pero también que ayuden a discriminar respecto a las medidas de otros constructos (Sacket, Putka, & McCloy, 2012), lo cual da lugar a dos subtipos de evidencias más: validez convergente (Ej: alta varianza explicada por los ítems de un mismo constructo) y validez discriminante (alta correlación entre ítems de un mismo constructo, en comparación con las correlaciones de estos con los de otros constructos).

Además de la validez, otras dos propiedades son fundamentales a la hora de evaluar diferencias individuales para la SP. Una de ellas es la fiabilidad y la otra la equidad. Respecto a la fiabilidad de los instrumentos, esta es la tendencia a la consistencia de las puntuaciones (Martínez, Hernández, & Hernández, 2006); lo que da una idea de la precisión del test y también resulta relevante para efectos de generalización o repetición de las puntuaciones. La equidad es imperante en la medida en que los test deben evaluar las diferencias individuales de forma imparcial, independiente de características personales como género, edad, etnia, etc (AERA, APA, & NCME, 1999). Su atención es fundamental, entre otros aspectos, para evitar eventuales demandas a causa de rasgos de discriminación en el proceso de selección (Batisda, & Moscoso, 2015). Desde el punto de vista cuantitativo, la equidad se infiere a partir de que los efectos de características personales sean irrelevantes en la medición del constructo determinado (Martínez, Hernández, & Hernández, 2006). Es de anotar que el término equidad no presenta una definición consensuada, pues debe interpretarse considerando diferentes circunstancias sociales y políticas en el dominio de interés (AERA, APA, & NCME, 1999).

Nótese que una solución de M-D/T que realmente pretenda aportar valor a los procesos de SP, debe considerar, como mínimo, los elementos expuestos, para lo cual deberá hacer uso de campos diferentes a las Ciencias de la Computación y la Inteligencia Artificial, e incluso apoyarse en personal experto en psicología o psicometría.

3. Procedimiento

La localización y selección de documentos se llevó a cabo en dos etapas. La primera fue de tipo sistemático para recorrer los documentos centrados, en específico, en soluciones de M-D/T para la SP. En esta etapa se usaron varios criterios de búsqueda, los cuales fueron ejecutados en las dos bases de datos de resúmenes y citas más posicionadas a nivel mundial: Scopus y Web of Science (WoS). En la Tabla 1 se detallan los criterios empleados en esta etapa.

Tabla 1

Criterios de búsqueda y cantidad de documentos recuperados; etapa 1. Actualizado al 20/10/2019

Tema	Fuente	Criterios de búsqueda	Docs
1: SP	Scopus	TITLE("personnel selection" OR "staff selection" OR "recruitment" OR "human resource selection" OR "human resource allocation") AND *	27792
	WoS	TI=("personnel selection" OR "staff selection" OR "recruitment" OR "human resource selection" OR "human resource allocation")**	23590
2: MD	Scopus	TITLE("data mining" OR "knowledge discovery" OR "computer scoring") AND *	22435
	WoS	TI=("data mining" OR "knowledge discovery" OR "computer scoring") **	7167
3: MT	Scopus	TITLE ("text mining" OR "information retrieval" OR "information extraction" OR "sentiment analysis" OR "natural processing language") AND *	19536
	WoS	TI=("text mining" OR "information retrieval" OR "information extraction" OR "sentiment analysis" OR "natural processing language") **	4732

* (LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cr") OR LIMIT-TO (DOCTYPE, "re") OR LIMIT-TO (DOCTYPE, "ip")).

** Índices=SCI-EXPANDED, SSCI, A&HCI, ESCI Período de tiempo=Todos los años.

SP: Selección de personal; MD: Minería de Datos; MT: Minería de Texto.

Obtenidos los documentos de M-D/T en procesos de SP, estos fueron revisados de forma manual, en el siguiente orden: título, resumen, resto del contenido. Luego, se descartaron aquellos que: (1) no corresponden con M-D/T en procesos laborales de SP (Ej: selección de pacientes para una investigación en salud), o (2) son soluciones computacionales para solamente recuperar conjuntos de candidatos o para diferenciar entre perfiles en anuncios de empleo. Entre los estudios no considerados como parte del conjunto de trabajos empíricos de M-D/T en SP seleccionados en esta etapa, están, por ejemplo: (1) Sawleshwarkar et al. (2018), los cuales proveen una memoria de conferencia en la que citan el valor de la MT para apoyar los procesos de selección; en dicha divulgación mencionan que tienen la intención futura de desarrollar una solución de reclutamiento MT que tome algunos datos de test psicométricos y de la hoja de vida de los aspirantes, con el fin de hacer el proceso más eficiente. Para explorar la pertinencia de esta idea, llevan a cabo una encuesta sobre un grupo de buscadores de empleo (no se declara tamaño de muestra, ni descripción de los participantes). (2) Wu, Shi, & Yang (2017), quienes analizan, mediante MT, anuncios de ofrecimiento de empleo para detectar diferencias y similitudes en conocimientos y habilidades manifestadas en dos tipos de ofertas laborales: big data e inteligencia de negocios.

Los documentos finalmente elegidos, como resultado de la presente etapa, fueron siete y se presentan en la Tabla 2.

Tabla 2

Estudios empíricos centrados en procesos de SP usando M-D/T, etapa 1 (sistemática)

Estudio	País*	Fuente	Área	Otros temas relacionados	Tipos de datos
Tai, & Hsu (2006)	Taiwan	Proceedings of the 9th Joint Conference on Information Sciences	MD	Fuzzy data mining; human resource management	Perfil del candidato vs comportamientos en el trabajo
Chien, & Chen (2008)	Taiwan	Expert Systems with Applications	MD	Human capital; semiconductor industry	Perfil del candidato vs comportamientos en el trabajo
Jayaraj, & Mahalakshmi (2015)	India	Indian Journal of Science and Technology	MD	Business intelligence; resume mining	Perfil del candidato vs requerimientos del cargo

Estudio	País*	Fuente	Área	Otros temas relacionados	Tipos de datos
Campion et al., (2016)	Estados Unidos	Journal of Applied Psychology	MT	Adverse impact; big data; statistics; test scoring	Respuestas de candidatos a preguntas abiertas (Ej: describa cómo ud ha demostrado liderazgo) vs valoraciones del humano
Shehu, & Saeed (2016)	Malasia	Journal of Theoretical and Applied Information Technology	MT	Domain-driven data mining; human resource management; method-driven data mining	Perfil del candidato vs si fue o no seleccionado
Han, & Lee (2016)	Korea del Sur	33rd ISARC International Symposium on Automation and Robotics in Construction	MD	Construction engineer; KNIME; Résumé	Perfil del candidato vs requerimientos del cargo
Awan et al. (2019)	Pakistan	Applied Sciences	MT	Semantic web; information retrieval; information extraction; e-recruitment	Perfil del candidato vs requerimientos del cargo

La segunda etapa fue una búsqueda tradicional, usando las mismas bases de datos de la etapa 1 y también Google Académico. Esta permitió recuperar 28 documentos adicionales (véase Tabla 3) que, desde la perspectiva de los autores, se relacionan con al menos uno de temas de interés (M-D/T o SP) y sirven para complementar el análisis de los documentos centrales.

Tabla 3
Lista resumen de los 28 estudios complementarios, etapa 2 (tradicional)

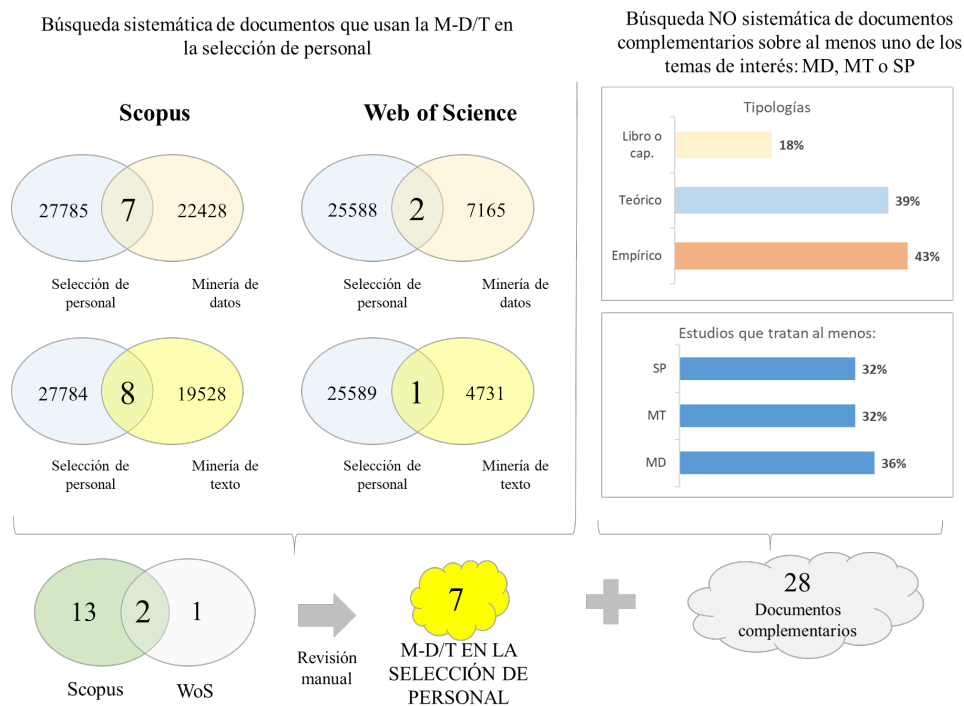
No	Estudios	Como mínimo el área:	Tipologías
1	Abbe et al. (2016)	MT	Teórico
2	Adejuwon, & Mosavi (2010)	MD	Teórico
3	Aghdaie, Zolfani, & Zavadskas (2013)	MD	Empírico
4	Athauda, Tissera, & Fernando (2009)	MD	Teórico
5	Balan, & Mathew (2015)	MT	Teórico
6	Delen, & Crossland (2008)	MT	Empírico
7	Evers, Anderson, & Smit-Voskuijl (2005)	SP	Libro
8	Fayyad, Piatetsky-Shapiro, & Smyth (1996)	MD	Teórico
9	Galinium (2017)	MD	Empírico
10	Herpertz et al. (2017)	SP	Empírico
11	Jereb, Rajkovic, & Rajkovic (2005)	SP	Empírico
12	Jockers (2014)	MT	Libro

No	Estudios	Como mínimo el área:	Tipologías
13	Kasemsap (2015)	MD	Cap. Libro
14	Kelemenis, & Askounis (2010)	SP	Empírico
15	Kjellström, & Golino (2018)	MT	Empírico
16	Koutra et al. (2017)	SP	Empírico
17	Kumari (2012)	MD	Teórico
18	Ledolter (2013)	MD	Libro
19	Lievens, Van Dam, & Anderson (2002)	SP	Teórico
20	Ngai, & Lee (2016)	MT	Teórico
21	Pandey, & Pandey (2017)	MT	Empírico
22	Peng et al. (2011)	MD	Empírico
23	Silge, & Robinson (2017)	MT	Libro
24	Smith, Hanges, & Dickson (2001)	SP	Empírico
25	Strohmeier, & Piazza (2013)	MD	Teórico
26	Werner (2000)	SP	Teórico
27	Zhang, & Liu (2011)	SP	Empírico
28	Zhang, Chen, & Liu (2015)	MT	Teórico

MD: Minería de datos; MT: Minería de texto; SP: Selección de personal

En la Figura 1 se resumen los documentos considerados en ambas etapas.

Figura 1
Resumen de cantidad de documentos centrales (izquierda) y complementarios (derecha). Fuente: elaboración propia



Nótese, a partir de la Figura 1, que esta revisión de literatura se basó fundamentalmente en 35 documentos. Siete fueron documentos centrales, propios de aplicaciones concretas de M-D/T, localizados y depurados de forma sistemática, y los 28 restantes fueron complementarios a los primeros, derivados de una búsqueda/selección tradicional.

4. Resultados

A continuación, se exponen y argumentan tres desafíos identificados para el ejercicio de la M-D/T en la SP, y, seguido, se exponen dos propuestas integradoras para superar dichos desafíos.

4.1. M-D/T aprovechando el valor de los datos narrativos para la SP

El primer desafío, en la M-D/T en la SP, se refiere a la naturaleza de los datos. Según Strohmeier, & Piazza (2013), de 100 estudios que revisaron en el dominio general de la gestión humana, más del 90% de estos se orientó hacia datos estructurados, clásicos de los sistemas de información de recursos humanos. No obstante, los datos de texto, que representan lenguaje narrativo natural, resultaron incipientes, a pesar de las amplias oportunidades que en ellos subyacen. Más aún, cuando la mayor parte de los datos en las organizaciones son textos (Jayaraj, & Mahalakshmi, 2015) sobre documentos de proyectos, entrevistas, hojas de vida, actas, memorandos, informes técnicos y procedimientos, etc.

Adicionalmente, emergen complejidades específicas para datos de texto, a causa de diferentes formatos, estilos, contextos y conocimientos previos del desarrollador del texto (Abbe et al., 2016; Balan, & Mathew, 2015; Jayaraj, & Mahalakshmi, 2015; Zhang, Chen, & Liu, 2015) y de un mayor número de dimensiones, entre ellas: valencia, subjetividad, legibilidad, lingüística, riqueza de información, exactitud, léxica, tamaño y frecuencia de palabras (Balan, & Mathew, 2015). También sobresalen problemas de polisemia, sinonimia y ambigüedad (Jayaraj, & Mahalakshmi, 2015), así como alta subjetividad de las etiquetas asignadas a términos refinados o grupos de estos (Abbe et al., 2016).

En la SP, las complejidades/limitantes descritas sobre este primer desafío (datos textuales) abren un sinnúmero de oportunidades de estudio, cuando dicha selección se realiza con base en la presentación de documentos narrativos por parte de los candidatos (submissions). Es decir, la postulación de propuestas, planes, cartas de intención, etc., para evaluación/juicio de otros. Sin embargo, con relación a este tema, la limitada evidencia de soluciones de M-D/T para la SP, se enfocó en su mayoría en emparejar computacionalmente listas de palabras clave de la hoja de vida del candidato o de algún tipo de preguntas abiertas, con requerimientos del cargo (Campion et al., 2016; Han, & Lee, 2016; Jayaraj, & Mahalakshmi, 2015). A excepción, por ejemplo, de Campion et al. (2016), fueron escasas las soluciones que tomaron en consideración ideas, pensamientos, planes, intenciones, enfoques u orientaciones, etc., que el candidato presente con respecto al cargo y la organización oferente, o su estilo, personalidad y emociones, o que exploren propiedades psicométricas de los hallazgos. Es decir, ha pasado desatendida una riqueza de aspectos del candidato que puede inducirse a partir de la palabra escrita.

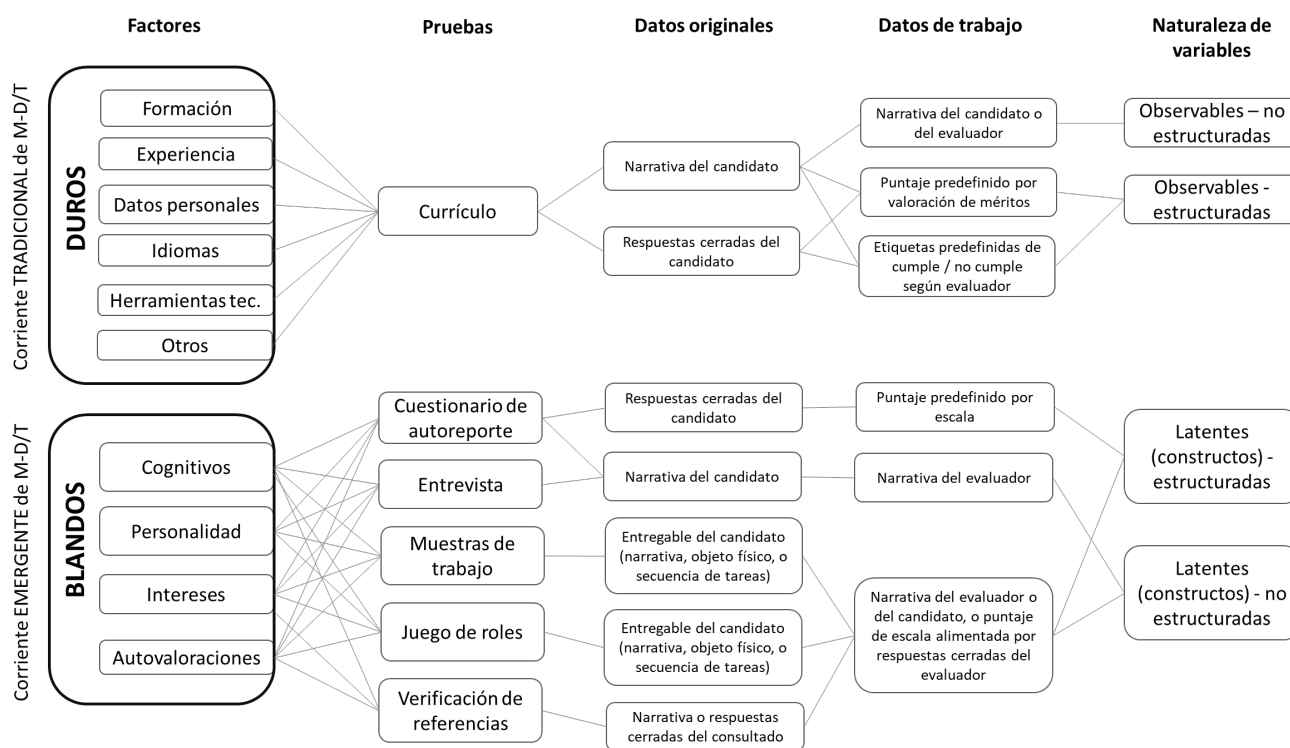
El estudio de Campion et al. (2016) provee bases significativas para futuros trabajos de M-D/T en la SP. Además, deja la puerta abierta a nuevas propuestas desde diversas áreas, que no fueron consideradas en tal estudio. Estas pueden resumirse, inicialmente, en que el marco propuesto por Campion et al., (2016) no fue intensivo en características que favorezcan la flexibilidad, el escrutinio y la reproducibilidad desde el punto de vista de las Ciencias de la Computación. En su lugar, aparte de reflexiones, revisiones, conceptualizaciones, estrategias y pasos valiosos, la sistemática que aporta describe la interacción entre el humano y una herramienta comercial de análisis de texto, entre otras tareas para apoyar la SP. Ello genera costos que podrían reducirse usando software libre, restringe la incorporación de nuevas estrategias, pasos, métodos y herramientas, bien sea de la

M-D/T o de otras áreas, y ocasiona que el marco de trabajo sea intensivo en la interacción humano – computador. Es decir, continuamente el investigador debe intervenir, luego alimentar el programa, después el programa le arroja resultados para que este los analice, interprete y modifique; una vez más los nuevos insumos deben alimentar el programa, y así sucesivamente. Por lo tanto, varias de las tareas ejecutadas desde el humano y desde la máquina no son totalmente reproducibles y no están abiertas al escrutinio. Este caso refleja lo que señalan Delen, & Crossland (2008) y Zhang, Chen, & Liu (2015) al referirse a los sistemas de MT en general, enfatizando en que suelen considerarse una “caja negra”.

Otra de las oportunidades que emergen del estudio de Campion et al. (2016), es que exige abundante cantidad de documentos para desplegarse en su totalidad (mínimo 500 ensayos narrativos de 200 palabras; uno por candidato), lo cual deja por fuera gran parte de los procesos de selección reales, como ellos mismos lo declaran. Sin embargo, si la intensión no es predictiva, con menor cantidad de documentos pueden ejecutarse varias etapas.

En la Figura 2 se resumen dos corrientes de abordaje de datos mediante la M-D/T para la SP, identificadas a lo largo del proceso de revisión, interpretación y síntesis.

Figura 2
Corrientes identificadas para el abordaje de datos por medio de M-D/T en la SP



En la Figura 2 pueden verse dos corrientes; una de ellas (parte superior), resume la corriente tradicional bajo la cual la M-D/T en la SP, ha tendido a concebir y abordar el análisis de diferencias individuales. Nótese que dicha corriente está basada principalmente en datos del currículum del candidato y asume un enfoque de pensamiento “duro”, con variables observables (explícitas en los documentos o puntuaciones). No obstante, la otra corriente (parte inferior de la figura) es la denominada “emergente”, en el sentido de que amerita trascender la práctica tradicional de la M-D/T, con el fin de profundizar en factores blandos del candidato, a través de una riqueza de pruebas que ofrece la SP. Mediante estas, no solo es posible obtener datos estructurados, sino también no estructurados (ej. textos), que en su conjunto posibilitan explotar el potencial de análisis automáticos y descubrimiento de patrones desde la M-D/T.

4.2. M-D/T desarrollando la interpretación de los hallazgos y la toma de decisiones en la SP

El segundo desafío tiene que ver con que, por definición, la M-D/T busca apoyar la toma de decisiones informada, a través de hallazgos no obvios que aporten valor a la organización. Esta promueve que mediante su uso las organizaciones podrán ser más competitivas, gracias al aprendizaje adquirido y a mejores cursos de acción (Kasemsap, 2015; Turban et al., 2008). Sin embargo, el estado actual de la práctica de M-D/T en la SP no ha venido desarrollando de forma fiable, válida y reproducible las etapas de interpretación de los hallazgos arrojados ni la toma de decisiones. A modo de evidencia, considere los siguientes estudios centrados en M-D/T en la SP:

Tai, & Hsu (2006) plantean dos etapas metodológicas: base de datos de transacción (1. Agregación de datos, 2. Transformación) y base de datos de regla. Sin embargo, estas no desarrollan la interpretación de patrones ni la validación de estos con base en propiedades de fiabilidad y validez. Chien, & Chen (2008) desarrollan seis etapas que van desde definir el problema y estructurar el objetivo hasta la interpretación, la extracción de conocimiento y el uso del conocimiento. No obstante, la interpretación de patrones se delega en juicio de “expertos”: “los resultados fueron presentados a un grupo de expertos en recursos humanos para interpretación y discusión de los usos potenciales de las reglas extraídas” (p. 287). A su vez, la toma de decisiones se basa en estrategias generadas bajo métodos no estructurados e independientes del marco desplegado: “basados en los resultados y las interpretaciones a través de MD y discusiones, desarrollamos estrategias de reclutamiento específicas con el fin de cumplir con el ajuste correcto de la mejor política” (p. 288). Jayaraj, & Mahalakshmi (2015) trabajan ocho etapas que van desde base de datos de currículos hasta la ordenación de candidatos. El marco facilita la interpretación y puede reproducirse gracias a un modelo de puntuación de “etiquetas” (Ej: 1. bachiller, 2. Magíster, 3. Doctorado, ...). Sin embargo, no está delimitado al proceso completo de SP, sino al filtro preliminar de hojas de vida, por lo que su utilidad se restringe a la preselección de personal. Champion et al. (2016) desarrollan un marco basado en tres partes: MT, regresión con datos de entrenamiento y predicción con nuevos datos. La interpretación y uso de los patrones no es posible reproducirse completamente debido a operaciones internas de un software propietario basado en botones y por alta dependencia de tareas humanas subjetivas. Sí trabaja con constructos y aporta evidencia de fiabilidad y validez, usando análisis de correlaciones. Shehu, & Saeed (2016) plantean cuatro etapas que van desde estudio preliminar de hojas de vida hasta la implementación de un modelo de SP para reclutar. El alcance de la evaluación de reglas en el dominio de la SP también se delega al juicio de “expertos”. A su vez, el alcance tampoco está delimitado al proceso completo de SP, sino al filtro preliminar de candidatos a partir de datos del currículo. Han, & Lee (2016) trabajan tres etapas: recolección de currículos y requerimientos del trabajo, diseño conceptual de ponderación por distancia entre “etiquetas” del currículo y requerimientos, ejecución y validación de resultados de MT. De igual forma, no está delimitado al proceso completo de SP, sino al filtro preliminar a partir de datos del currículo y bajo un proceso de evaluación de reglas delegado al juicio de “expertos”.

La M-D/T, al formar parte de áreas de las Ciencias de la Computación, se rige por el paradigma de la investigación reproducible (Krause, 2016; Peng, 2011). No obstante, al considerar marcos de M-D/T en la SP, la interpretación de patrones extraídos y el proceso de decisión se siguen descargando en el juicio de “expertos”. En otras palabras, las soluciones de M-D/T, en su alcance actual, no ha logrado desarrollar el proceso completo desde la planificación del dato, hasta la interpretación y el uso del conocimiento (toma de decisiones y la generación de valor en tal dominio) (Strohmeier, & Piazza, 2013). Esto mismo puede verse en el alcance de libros guía, como: Jockers (2014); Ledolter (2013); Silge, & Robinson (2017). De ahí el llamado constante, en el caso de la M-D/T en la gestión humana en general (Strohmeier, & Piazza, 2013), a integrar otros pasos para ajustarse de mejor manera al enfoque conducido por el dominio y finiquitar el proceso completo de SP con criterios de fiabilidad, reproducibilidad, equidad y validez.

4.3. M-D/T complementada con procedimientos confirmatorios que soporten la fiabilidad y validez de constructos psicológicos

Para aspirar a lograr una solución de M-D/T conducida por el dominio de la SP, es preciso reconocer que dicho dominio es intensivo en constructos y demanda emplear criterios, métodos e instrumentos fiables y válidos, soportados por áreas como la Psicometría. Es así que una solución de M-D/T en la SP, también debe considerar el enfoque guiado por constructos, el cual busca comprender qué métodos/instrumentos funcionan, por qué funcionan y qué es lo que realmente miden (Lievens, Van Dam, & Anderson, 2002).

Así, el tercer desafío se debe a la inviabilidad de los marcos actuales de M-D/T para fijar posiciones confirmatorias que den mayor solidez al proceso racional de toma de decisiones desde lo que demanda el enfoque conducido por constructos. Esto se sustenta en que el principal desarrollo que ha tenido la M-D/T es principalmente exploratorio, guiada por los datos (observacionales) (Abbe et al., 2016; Athauda, Tissera, & Fernando, 2009; Chien, & Chen, 2008; Kumari, 2012). Es decir, los hallazgos obtenidos con el uso tradicional de M-D/T, que pueden resumirse en clustering, clasificación, asociación y predicción (Fayyad, Piatetsky-Shapiro, & Smyth, 1996; Ledolter, 2013) (en el caso de datos estructurados) o en extracción/recuperación de información, clasificación de documentos, análisis de sentimientos y de tendencias (en cuanto a textos; Jayaraj, & Mahalakshmi, 2015; Zhang, Chen, & Liu, 2015), no son de naturaleza confirmatoria.

En las pocas aplicaciones identificadas de la M-D/T en la SP, la etapa confirmatoria (empíricamente) que se demanda para cerrar el proceso de generación de conocimiento (Swanson, & Holton, 2005), se suele dejar a juicio y/o iniciativas posteriores de los decisores. Por consiguiente, las actuales aplicaciones de M-D/T en la SP no necesariamente están asegurando soluciones fiables (consistencia de las puntuaciones), válidas (evidencia de lo que realmente miden), equitativas (frente al dilema validez-diversidad) y reproducibles respecto a la justificación de las relaciones subyacentes en (o entre) los constructos. Más aún, de acuerdo con Abbe et al. (2016), la fiabilidad de los hallazgos es raramente discutida.

Nótese que, ante las complejidades y particularidades de la SP, en el marco del desarrollo psicométrico y el enfoque conducido por constructos, no basta incluso con soluciones de M-D/T que meramente busquen confirmar los hallazgos a través de estrategias efectivas de análisis predictivo – computacional, como la partición de la muestra en dos conjuntos: observaciones de entrenamiento del modelo y observaciones de validación de su capacidad predictiva. Esto, debido a que los instrumentos de medida empleados deben primero superar propiedades psicométricas fundamentales ya descritas. De hecho, de acuerdo con Jereb, Rajkovic, & Rajkovic (2005), si no se confirman la fiabilidad y la validez de las medidas de los constructos, ninguna posterior relación puede ser demostrada (Jereb, Rajkovic, & Rajkovic, 2005).

Las limitantes de las aplicaciones de M-D/T para la SP pueden subsanarse al considerar campos como la Psicometría, el cual tradicionalmente ha soportado con éxito varios métodos de selección basados en constructos (Herpertz et al., 2016; Smith, Hanges, & Dickson, 2001). En este campo, fiabilidad y validez están mejor definidas y existen métricas concretas para juzgarlas (ej: Alfa de Cronbach, Fiabilidad Compuesta, Validez Convergente, Validez Discriminante) (Lévy, & Varela, 2006).

El estudio de Champion et al. (2016) toma en cuenta el enfoque conducido por constructos en el ámbito de la M-D/T aplicada un proceso de SP, bajo un enfoque articulador entre Ciencias de la Computación y Psicología. Dicho trabajo considera ensayos de más de 40000 candidatos para explotar la M-D/T y, además, compara los resultados con las calificaciones dadas por tres expertos y con varios test psicométricos. Dicho aporte muestra un caso de

éxito de las potencialidades de la M-D/T en la selección masiva de personal, entre ellas: estudiar la validez y la fiabilidad con métodos psicométricos, proveer enfoque de descubrimiento de constructos, incorporar modelamiento predictivo, comparar puntuaciones computacionales vs. humanas y favorecer la eficiencia del proceso. A pesar de estas bondades y enfatizando en el enfoque conducido por constructos, los procedimientos de análisis de la fiabilidad y la validez de los patrones inducidos de texto, usados en Champion et al. (2016), pueden seguirse desarrollando, a través de métodos/técnicas confirmatorias que vayan más allá del análisis de correlaciones simples.

4.4. Oportunidades integradoras para la M-D/T en la SP

Este apartado propone dos oportunidades para superar los desafíos expuestos, bajo una perspectiva que busca nutrir el potencial de la M-D/T mediante su integración con dos áreas de estudio: modelos de ecuaciones estructurales y análisis de decisión multicriterio.

Modelos de ecuaciones estructurales para confirmar los hallazgos de M-D/T en el dominio de la SP

Para confirmar empíricamente la fiabilidad y la validez de las soluciones de M-D/T en la SP, bajo el enfoque conducido por constructos, resulta útil recurrir a “reglas de oro” de dicho dominio, provenientes de la Psicometría. Esta es catalogada como una de las áreas científicas más exitosas para la medición de constructos psicológicos (Martínez, Hernández, & Hernández, 2006; Nunnally, & Bernstein, 1995). A partir de ella, se cuenta con recursos para comenzar a analizar y confirmar la fiabilidad, la validez y el ajuste de los patrones inducidos desde el análisis automático de datos/textos. Una de las herramientas psicométricas de mayor auge son los modelos de ecuaciones estructurales (SEM, en inglés; se emplea en adelante esta sigla, dada su popularidad en el medio) (Lévy, & Varela, 2006). Este tipo de modelos consta de dos componentes, uno referido a contrastar el modelo de medida (fiabilidad, validez) y otro el modelo estructural (relaciones entre constructos). El modelo de medida está conformado por cada constructo (ej: liderazgo) y sus indicadores. Su contraste se realiza por medio de la técnica Análisis Factorial Confirmatorio, que consta de: especificación del modelo, identificación, estimación, ajuste, interpretación y/o reespecificación (Rial et al., 2006). Además, arroja varias métricas útiles para este contexto, como: validez de cada indicador, validez convergente (según la proporción de varianza común entre los indicadores de un mismo constructo), validez discriminante (diferenciación entre los constructos) y fiabilidad compuesta (consistencia entre los indicadores de un mismo constructo). Todo esto puede profundizarse en Lévy, & Varela (2006); Rial et al., (2006), entre otros.

Ante las limitantes causales de la investigación no experimental, SEM es un recurso estadístico en auge y valioso para validar constructos y confirmar relaciones entre ellos, desde luego, bajo algún fundamento teórico que dé sentido a las relaciones por contrastar. En sí, para la SP asistida por M-D/T, resulta razonable considerar que SEM pueden ayudar en la generación de soluciones cada vez más centradas en el dominio, lo cual abre nuevos abanicos de estudio. Entre los pocos usos de aprovechamiento de datos de texto (MT) que tomen en cuenta aspectos psicométricos de SEM, vale señalar a Pandey, & Pandey (2017) en el estudio del constructo “cultura organizacional” (300 reportes narrativos de empresas) y Kjellström, & Golino (2018) en responsabilidad de salud (233 documentos de entrevistas semiestructuradas). A pesar de que dichos referentes no se enfocaron en el dominio de la SP, sí proveen evidencia de la utilidad y viabilidad de articular elementos de M-D/T con SEM bajo un enfoque conducido por constructos.

M-D/T con análisis de decisión multicriterio para favorecer la toma de decisiones en el dominio de la SP

El análisis de decisión multicriterio (ADM) se sustenta en que permite reconocer y entender la SP como un proceso complejo, donde el decisor debe elegir la alternativa “óptima” (o más satisfactoria) entre un determinado número finito de candidatos, los cuales son evaluados de forma simultánea en varios criterios (atributos) con diferentes preferencias (Dağdeviren, 2010; Kelemenis, & Askounis, 2010). La complejidad radica

en la existencia de varias alternativas, múltiples criterios “duros” y “blandos”, restricciones individuales y colectivas, objetivos en conflicto y preferencias subjetivas (Chavira et al., 2017; Dağdeviren, 2010; Marttunen, Lienert, & Belton, 2017; Zhang, & Liu, 2011). Estas características han hecho de la SP bajo un enfoque multicriterio, uno de los principales tópicos de aplicación de las Ciencias de la Decisión a la gestión del capital humano. La incorporación del ADM cobra sentido en las fases de interpretación y uso de los patrones extraídos por M-D/T, ya que el ADM divide el problema en subproblemas pequeños y menos complejos (Jereb, Rajkovic, & Rajkovic, 2005) y ofrece posibilidades de evaluar, ordenar candidatos y elegir el “mejor” de ellos (Chavira et al., 2017; Dağdeviren, 2010; Kelemenis, & Askounis, 2010). Así, la M-D/T prepara los datos y descubre patrones de forma automática (o semi-automática) y aporta insumos al ADM para evaluar candidatos, decidir de forma sistemática y lograr un mayor número de tareas reproducibles.

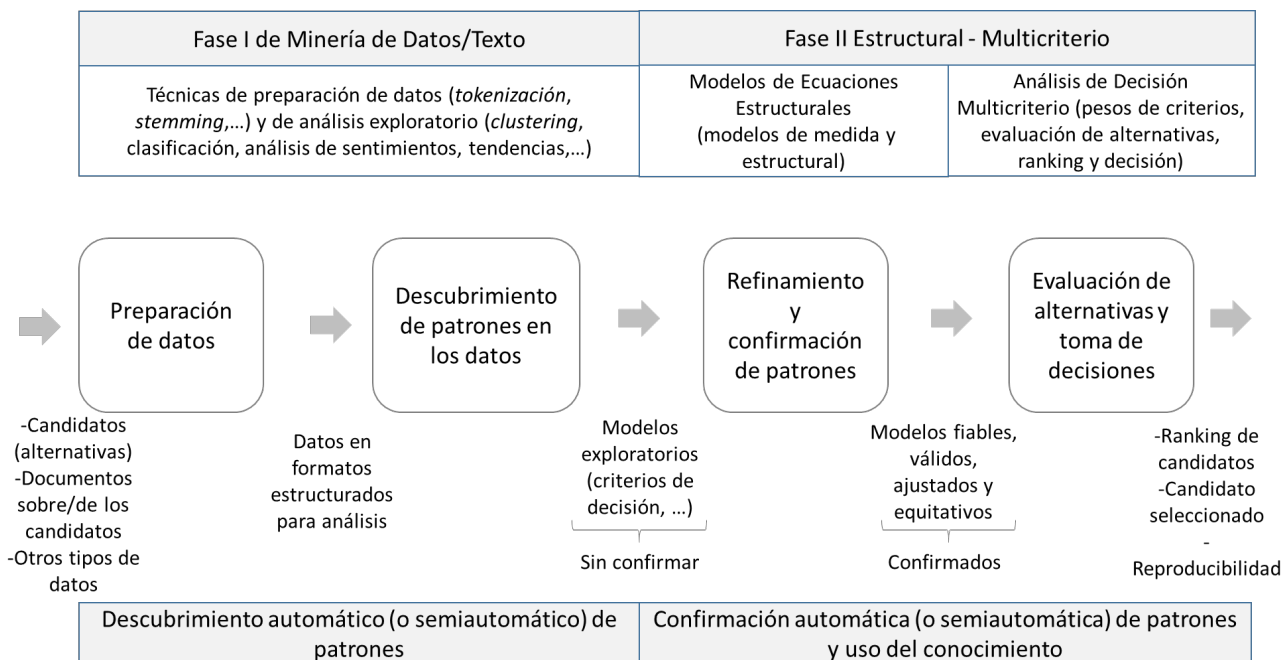
En resumen, el ADM ayuda a complementar las soluciones lideradas por M-D/T, de modo que abarquen más etapas de la SP, por medio de procedimientos de toma de decisiones que favorecen transparencia, comprensión, explicación y justificación de la selección del candidato. Pero, además, la M-D/T puede complementar las soluciones lideradas por el ADM, ya que esta última no logra desarrollar la etapa de estructuración del problema, a pesar de ser crítica para el éxito de su despliegue (del ADM) (Banville et al., 1998; Marttunen, Lienert, & Belton, 2017). Según Marttunen, Lienert, & Belton (2017), el ADM asume que ya se cuenta con un problema debidamente estructurado, lo cual implica que se tiene precisión y fiabilidad acerca de las alternativas, los criterios y las relaciones entre componentes, entre otros aspectos. Esto amerita el desglose de una serie de actividades para la comprensión del contexto, la identificación de grupos de interés, la indagación, el análisis de percepciones y el tratamiento de información adicional sobre el problema, los objetivos y las posibilidades de solución. Todo ello, hasta lograr una posición común que derive en la estructuración del problema. De hecho, Marttunen, Lienert, & Belton (2017) señala que varios métodos de estructuración de problemas han venido complementando el ADM, entre ellos: análisis de stakeholders, mapas cognitivos, DOFA (Debilidades, Oportunidades, Fortalezas y Amenazas). Este tipo de métodos son propios del pensamiento “blando”, el cual se sustenta en el paradigma interpretativista, que es intensivo en significados, percepciones y métodos cualitativos (Swanson, & Holton, 2005). A través de los métodos de estructuración de problemas se busca hacer frente, de manera sistemática, a variedad de situaciones subjetivas y de síntesis de información. Según Rosenhead, & Mingers (2001), estos deben tomarse como “medios” para la comprensión unificada y no para resolver el problema.

Precisamente, un área poco considerada, que puede apoyar la estructuración de problemas de ADM y no se nombra en la extensa revisión de Marttunen, Lienert, & Belton (2017), es la M-D/T. A través del descubrimiento de patrones por medio de esta, es posible identificar alternativas y criterios de decisión. No obstante, la literatura arrojó escasas aplicaciones de la M-D/T asistiendo el ADM. Entre los pocos estudios al respecto, sobresalen dominios de: riesgos de incidentes (Peng et al., 2011), bienes raíces (Galinium, 2017) y segmentación de mercados (Aghdaie, Zolfani, & Zavadskas, 2013); todos estos, usando datos estructurados. La combinación de M-D/T con ADM es un campo en infancia, cuyo potencial radica en que la estructuración del problema no se basa en métodos blandos, como tradicionalmente ocurre (Rosenhead, & Mingers, 2001). De hecho, Marttunen, Lienert, & Belton (2017) concluyen que aún hay mucho espacio para innovaciones que surjan de la combinación del ADM con otras áreas, para favorecer la estructuración de problemas, y también invitan a considerar nuevos dominios. Al proponer la integración de M-D/T con ADM, se espera favorecer la reproducibilidad y la objetividad de las tareas de estructuración del problema de selección y de toma de decisiones, además de aumentar la eficiencia del proceso, fruto de la automatización de tareas.

Síntesis de las dos alternativas integradoras bajo un enfoque de procesos

La Figura 3 resume la integración de la M-D/T con SEM y ADM, bajo una estructura de relaciones cliente – proveedor.

Figura 3
Integración de M-D/T con SEM y ADM bajo un enfoque de procesos:
Minería de Datos/Texto – Estructural - Multicriterio



Nótese, en la Figura 3, que la integración de M-D/T con SEM y ADM ha sido denominada Minería de Datos/Texto – Estructural – Multicriterio y consta de dos macroprocesos, denotados como Fase I y Fase 2. El primer macroproceso tiene como misión explotar el potencial de la M-D/T para la preparación de datos y el descubrimiento de patrones (constructos psicológicos), a partir de la riqueza de datos (estructurados y no estructurados) que provee la SP. El segundo macroproceso despliega el uso del conocimiento e información proveídos por el primer macroproceso (Fase I), con el fin de validar psicométricamente las propiedades de estos, así como llevar a cabo el proceso racional de toma de decisiones (evaluación de candidatos y elección final). Pero es de anotar que ambos macroprocesos explotan también el potencial de la M-D/T en lo que respecta al procesamiento y análisis automático (o semi-automático) para el abordaje, sobre todo de datos masivos. A su vez, la Figura 3 muestra que, dentro de los macroprocesos, es posible concebir cuatro procesos, los cuales presentan entradas y salidas concretas, interpretadas en el contexto de la SP.

5. Conclusiones

Esta revisión de literatura sobre M-D/T en la SP retrata escasa disponibilidad de estudios previos sobre el tema, así como, en las pocas aplicaciones reales, una prevalencia de enfoques en su mayoría fragmentados para hacerle frente a particularidades y necesidades integrales de la SP. Por lo mismo, este trabajo resalta la necesidad de lograr, en una misma solución de M-D/T para la SP, la preparación de datos, el descubrimiento de patrones (modelos, criterios, ...), la confirmación de estos (métricas de fiabilidad, validez y equidad), la evaluación y ordenación de candidatos, y la elección final. Todo ello, bajo el paradigma de la investigación reproducible.

Este trabajo invita a considerar dos áreas (SEM y ADM) para complementar las bondades y debilidades actuales de la M-D/T en la SP, con miras a lograr soluciones conducidas por dicho dominio. Con la integración de tales áreas se busca favorecer el ajuste a la evidencia, la fiabilidad, la validez, la equidad y la reproducibilidad de los hallazgos de M-D/T, así como mejorar la eficiencia de los procesos. Esto último, dadas las posibilidades de

automatizar, lo mayor posible, algunas tareas inductivas (paradigma cualitativo), así como tareas de deducción de hipótesis y contraste de estas (propias de la investigación cuantitativa). Bajo esta óptica, se busca propiciar, mediante un enfoque más integral, soluciones de M-D/T mejor guiadas por el dominio de la SP.

Articular M-D/T con SEM y ADM exige recurrir a la investigación mixta (cuali-cuanti). Por un lado, el enfoque cualitativo para definir objetos conceptuales y proponer modelos de partida para explorar en textos narrativos y, por el otro, el enfoque cuantitativo para la deducción de hipótesis y su posterior contraste empírico; todo ello, tomando como columna vertebral la M-D/T. La pertinencia de articular las áreas descritas, en el dominio de la SP, puede entenderse desde la teoría de recursos (Barney, 1991) y el enfoque de capacidades dinámicas (Teece, Pisano, & Shuen, 1999), a partir de los cuales se reconoce que los recursos, *per se*, no generan ventajas competitivas. Sin embargo, el arte y la ciencia de estructurarlos, relacionarlos, combinarlos, automatizarlos y utilizarlos es todo un desafío, que puede conducir al desarrollo de capacidades analíticas que deriven en ventajas competitivas (Su, & Linderman, 2016; Teece, Pisano, & Shuen, 1999).

Se espera que este estudio motive agendas de investigación alrededor de una nueva rama denominada “Minería de Datos/Texto – Estructural – Multicriterio”, con el fin de estimular soluciones cada vez más acordes con las realidades y retos de la SP, así como de áreas en las que esta se indexa (gestión humana, psicología organizacional, dirección de organizaciones, ...).

Referencias bibliográficas

- Abbe, A., Grouin, C., Zweigenbaum, P., & Falissard, B. (2016). Text mining applications in psychiatry: a systematic literature review. *International journal of methods in psychiatric research*, 25(2), 86-100.
- Adejuwon, A., & Mosavi, A. (2010). Domain driven data mining—Application to business. *IJCSI International Journal of Computer Science Issues*, 7(4), 41-44.
- Aghdaie, M., Zolfani, S., & Zavadskas, E. (2013). A hybrid approach for market segmentation and market segment evaluation and selection: an integration of data mining and MADM. *Transformations in Business and Economics*, 12(2B), 431-459.
- Athauda, R., Tissera, M., & Fernando, C. (2009). Data mining applications: Promise and challenges. *Data Mining and Knowledge Discovery in Real Life Applications*. InTech.
- Awan, A., Nabeel, M., Khan, S., Latif, K., & Khattak, A. (2019). A New Approach to Information Extraction in User-Centric E-Recruitment Systems. *Applied Sciences*, 9(14), 2852.
- Balan, U. & Mathew, S. (2015). Online word of mouth using text mining: A review of literature and future directions. *Computational Intelligence: Theories, Applications and Future Directions (WCI)*, 2015 IEEE Workshop on (1-6). IEEE.
- Banville, C., Landry, M., Martel, J. M., & Boulaire, C. (1998). A stakeholder approach to MCDA. *Systems Research and Behavioral Science*, 15(1), 15-32.
- Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17, 99–120.
- Campion, M., Campion, M., Campion, E., & Reider, M. (2016). Initial investigation into computer scoring of candidate essays for personnel selection. *Journal of Applied Psychology*, 101(7), 958.
- Chavira, D., Lopez, J., Noriega, J., & Retamales, J. (2017). A multicriteria outranking modeling approach for personnel selection. In *Fuzzy Systems (FUZZ-IEEE)*, July 2017 IEEE International Conference on (1-6). IEEE.

- Chien, C., & Chen, L. (2008). Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. *Expert Systems with applications*, 34(1), 280-290.
- Dağdeviren, M. (2010). A hybrid multi-criteria decision-making model for personnel selection in manufacturing systems. *Journal of Intelligent manufacturing*, 21(4), 451-460.
- Delen, D., & Crossland, M. (2008). Seeding the survey and analysis of research literature with text mining. *Expert Systems with Applications*, 34(3), 1707-1720.
- Evers, A., Anderson, N., & Smit-Voskuijl, O. (Eds.) (2005). *The Blackwell handbook of personnel selection* (1). John Wiley & Sons.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 36–53.
- Galinium, M. (2017). Integrating data mining technique and AHP in market analysis to propose new product development in real estate. In *IOP Conference Series: Materials Science and Engineering* (166(1), 012030). IOP Publishing.
- Han, S., & Lee, G. (2016). A Preliminary Study on Text Mining-Based Human Resource Allocation in a Construction Project. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction* (33, 1). Vilnius Gediminas Technical University, Department of Construction Economics & Property.
- Herpertz, S., Nizielski, S., Hock, M., & Schütz, A. (2016). The relevance of emotional intelligence in personnel selection for high emotional labor jobs. *PloS one*, 11(4), e0154432.
- Jayaraj, V., & Mahalakshmi, V. (2015). Augmenting efficiency of recruitment process using IRCF text mining algorithm. *Indian Journal of Science and Technology*, 8(16), 1-8.
- Jereb, E., Rajkovic, U., & Rajkovic, V. (2005). A hierarchical multi-attribute system approach to personnel selection. *International Journal of Selection and Assessment*, 13(3), 198-205.
- Jockers, M., 2014. *Text analysis with R for students of literature*. New York: Springer.
- Kasemsap, K. (2015). The role of data mining for business intelligence in knowledge management. *Integration of data mining in business intelligence systems*, 12-33.
- Kelemenis, A., & Askounis, D. (2010). A new TOPSIS-based multi-criteria approach to personnel selection. *Expert systems with applications*, 37(7), 4999-5008.
- Kjellström, S., & Golino, H. (2018). Mining concepts of health responsibility using text mining and exploratory graph analysis. *Scandinavian journal of occupational therapy*, 1-16.
- Koutra, G., Barbounaki, S., Kardaras, D., & Stalidis, G. (2017). A Multicriteria Model for Personnel Selection in Maritime Industry in Greece. *Business Informatics (CBI), July 2017 IEEE 19th Conference on* (1, 287-294). IEEE.
- Krause, A. (2016). Reproducible Research in Real Estate: A Review and an Example. *Journal of Real Estate Practice and Education*, 19(1), 69-85.
- Kumari, M. (2012). Data Driven Data Mining to Domain Driven Data Mining. *Global Journal of Computer Science and Technology*, 1 (23), art. 422.

- Ledolter, J. (2013). *Data mining and business analytics with R*. John Wiley & Sons, New Jersey.
- Lévy, J., & Varela, J. (2006). *Modelización con estructuras de covarianzas en ciencias sociales. Temas esenciales, avanzados y aportaciones especiales*. A Coruña: Netbiblo.
- Lievens, F., Van Dam, K., & Anderson, N. (2002). Recent trends and challenges in personnel selection. *Personnel Review*, 31(5), 580-601.
- Martínez, M., Hernández M. & Hernández, M. (2006). *Psicometría*. Madrid: Alianza Editorial.
- Marttunen, M., Lienert, J., & Belton, V. (2017). Structuring problems for Multi-Criteria Decision Analysis in practice: A literature review of method combinations. *European Journal of Operational Research*, 263(1), 1-17.
- Ngai, E., & Lee, P. (2016). A review of the literature on applications of Text Mining in Policy Making. *PACIS*, June (343).
- Nunnally, J. & Bernstein, I. (1995). *Teoría psicométrica (3ª ed)*. México, D.F.: Editorial McGrawHill Latinoamericana.
- Pandey, S., & Pandey, S. K. (2017). Applying natural language processing capabilities in computerized textual analysis to measure organizational culture. *Organizational Research Methods*, 1-33.
- Peng, R. (2011). Reproducible research in computational science. *Science*, 334(6060), 1226-1227.
- Peng, Y., Zhang, Y., Tang, Y., & Li, S. (2011). An incident information management framework based on data integration, data mining, and multi-criteria decision making. *Decision Support Systems*, 51(2), 316-327.
- Rial, A., Valera, J., Abalo, J., & Lévy, J. (2006). El análisis factorial confirmatorio. En: Lévy, J., & Varela, J. (Dir. y Comp.). *Modelización con estructuras de covarianzas en ciencias sociales (119–143)*. La Coruña, España: Editorial Netbiblo S.L. (Editora: Seco, C).
- Rosenhead, J., & Mingers, J. (2001). *Rational analysis for a problematic world revisited (2nd)*. John Wiley and Sons.
- Shehu, M., & Saeed, F. (2016). An adaptive personnel selection model for recruitment using domain-driven data mining. *Journal of Theoretical & Applied Information Technology*, 91(1), 117-129.
- Silge, J., & Robinson, D. (2017). *Text Mining with R: A tidy approach*. O'Reilly Media, Inc.
- Smith, D., Hanges, P., & Dickson, M. (2001). Personnel selection and the five-factor model: Reexamining the effects of applicant's frame of reference. *Journal of Applied Psychology*, 86(2), 304.
- Strohmeier, S., & Piazza, F. (2013). Domain driven data mining in human resource management: A review of current research. *Expert Systems with Applications*, 40(7), 2410-2420.
- Su, H., & Linderman, K. (2016). An Empirical Investigation in Sustaining High-Quality Performance. *Decision Sciences*, 47 (5), 787-819.
- Sawleshwarkar, S., Rangnani, N., Mariwalla, V., & Halbe, A. (2018). Simplified Recruitment Model Using Text-Mining on Psychometric and Aptitude Tests. In *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 586-589). IEEE.

Swanson, R., & Holton, E. (2005). Research in organizations: Foundations and methods in inquiry. Berrett-Koehler Publishers.

Tai, W. & Hsu, C. (2006). A realistic personnel selection tool based on fuzzy data mining method. In 9th Joint International Conference on Information Sciences (October, JCIS-06). Atlantis Press.

Teece, D., Pisano, G., & Shuen, A. (1999). Dynamic capabilities and strategic management. Knowledge and strategy, 77-115.

Turban, E., Sharda, R., Aronson, J. E., & King, D. (2008). Business intelligence: A managerial approach (58-59). Upper Saddle River, NJ: Pearson Prentice Hall.

Werner, J. (2000). Implications of OCB and contextual performance for human resource management. Human resource management review, 10(1), 3-24.

Wu, J., Shi, H., & Yang, J. (2017). Are big data talents different from business intelligence expertise? Evidence from text mining using job recruitment advertisements. Service Systems and Service Management (ICSSSM), 2017 International Conference on (1-6). IEEE.

Zhang, S., & Liu, S. (2011). A GRA-based intuitionistic fuzzy multi-criteria group decision making method for personnel selection. Expert Systems with Applications, 38(9), 11401-11405.

Zhang, Y., Chen, M., & Liu, L. (2015). A review on text mining. In Software Engineering and Service Science (ICSESS), September 2015 6th IEEE International Conference on (681-685).

Anexos

Anexo 1

Evidencias de ejecución de algoritmos de búsqueda en Scopus y la escasa cantidad de estudios recuperados (actualizado al 20/10/2019)

Search history	Combine queries...
11 (TITLE ("personnel selection" OR "staff selection" OR "recruitment" OR "human resource selection" OR "human resource allocation")) AND (TITLE ("text mining" OR "information Retrieval" OR "information extraction" OR "sentiment analysis" OR "natural processing language")) AND (LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cr") OR LIMIT-TO (DOCTYPE, "re") OR LIMIT-TO (DOCTYPE, "ip")) ... View More ✓	8 document results
10 (TITLE ("personnel selection" OR "staff selection" OR "recruitment" OR "human resource selection" OR "human resource allocation")) AND (TITLE ("data mining" OR "knowledge Discovery" OR "computer scoring")) AND (LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cr") OR LIMIT-TO (DOCTYPE, "re") OR LIMIT-TO (DOCTYPE, "ip"))	7 document results
9 TITLE ("text mining" OR "information Retrieval" OR "information extraction" OR "sentiment analysis" OR "natural processing language") AND (LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cr") OR LIMIT-TO (DOCTYPE, "re") OR LIMIT-TO (DOCTYPE, "ip"))	19,536 document results
8 TITLE ("data mining" OR "knowledge Discovery" OR "computer scoring") AND (LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cr") OR LIMIT-TO (DOCTYPE, "re") OR LIMIT-TO (DOCTYPE, "ip"))	22,435 document results
1 TITLE ("personnel selection" OR "staff selection" OR "recruitment" OR "human resource selection" OR "human resource allocation") AND (LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cr") OR LIMIT-TO (DOCTYPE, "re") OR LIMIT-TO (DOCTYPE, "ip"))	27,792 document results

Anexo 2

Evidencias de ejecución de algoritmos de búsqueda en WoS y la escasa cantidad de estudios recuperados (actualizado al 20/10/2019)

Conjunto	Resultados	<input type="button" value="Guardar historial/Crear alerta"/> <input type="button" value="Abrir historial guardado"/>	<input type="button" value="Editar conjuntos"/>
# 5	2	#2 AND #1 <i>Índices=SCI-EXPANDED, SSCI, A&HCI, ESCI Período de tiempo=Todos los años</i>	Editar
# 4	1	#3 AND #1 <i>Índices=SCI-EXPANDED, SSCI, A&HCI, ESCI Período de tiempo=Todos los años</i>	Editar
# 3	4.732	TI=("text mining" OR "information Retrieval" OR "information extraction" OR "sentiment analysis" OR "natural processing language") <i>Índices=SCI-EXPANDED, SSCI, A&HCI, ESCI Período de tiempo=Todos los años</i>	Editar
# 2	7.167	TI=("data mining" OR " knowledge Discovery" OR "computer scoring") <i>Índices=SCI-EXPANDED, SSCI, A&HCI, ESCI Período de tiempo=Todos los años</i>	Editar
# 1	23.590	TI=("personnel selection" OR "staff selection" OR "recruitment" OR "human resource selection" OR "human resource allocation") <i>Índices=SCI-EXPANDED, SSCI, A&HCI, ESCI Período de tiempo=Todos los años</i>	Editar