Big data text analytics: an enabler of knowledge management

1. Introduction

The role of big data in effective decision-making and improving many business functions from marketing to supply chain has been acknowledged (Chae, 2015; Chen, Chiang, & Storey, 2012; Davenport, 2013; Waller & Fawcett, 2013). As such, big data was acknowledged by Davenport & Patil (2012) as the next big thing in the 21st century. Testament to this, a Businessweek (2011) survey of the state of business analytics found that 97 percent of companies with revenues exceeding $100 million were to use some form of business analytics. However, according to IBM as much as 80% of the data available to an organization is unstructured (George et al., 2014), and so there is a significant opportunity to leverage in the analysis of unstructured data. Unlocking this potential represents the next Big Data challenge for businesses, concerning how to use big data to extract useful information to make more informed decisions and develop a competitive advantage (Rajaraman & Ullman, 2011).

Companies such as Amazon, eBay and Walmart are using big data text analytics to effectively manage vast amount of knowledge, communicate with their customers and enhance their operations (Davenport & Patil, 2012). This has led to a growing academic interest in big data text analytics, yet there is a dearth of research examining the role of big data text analytics as an enabler of knowledge management (Davenport, 2013; Watson & Marjanovic, 2013). Indeed big data has been characterized as powering the next industrial revolution, so it is somewhat surprising that it has not figured more prominently in the field of knowledge management. Big data has been characterised in terms of volume, variety and velocity (Laney, 2001), while knowledge has been defined in terms of tacit, explicit, implicit, complex, simple, as well as tacit codified and encapsulated (Nonaka & Takeuchi, 1995; Zander & Kogut, 1995; Gao et al., 2008; van den Berg, 2013). There is an opportunity in big data to discover hidden knowledge and generate new knowledge which is important to enable and enhance knowledge management using big data text analytics.

Knowledge management (KM) deals with the processes and practices that enable the creation, acquisition, capturing, and sharing of knowledge (Scarborough & Swan, 2001;...
KM systems have been suggested to be the key for improving the efficiency of business processes and key determinants of competitive advantage (Cockrell & Stone, 2010; Vorakulpipat & Rezgui, 2008; Witherspoon, Bergner, Cockrell, & Stone, 2013). However, there is a paucity of research on the role of big data text analytics in KM (Chen et al., 2012; Davenport, 2013). However, it has previously been stated that big data text analytics is an important part of KM (Chen et al., 2012; King, 2009; Wang & Wang, 2008). It can help not only in the sharing of common knowledge of business intelligence, but also helps in extending human knowledge (Wang & Wang, 2008). However, the application and utility of big data text analytics in the generation of knowledge insights as part of KM is not fully explored. Big data text analytics tools could help organizations in the discovery of hidden knowledge and generation of new knowledge from vast amounts of structured and unstructured data.

The aim of this article is to show the utility of big data text analytics as an enabler of knowledge management (George, Haas, & Pentland, 2014; Grant, 1996). We apply text analytics as an example on 196 articles published in two of the leading journals in the domain of knowledge management—*the Journal of Knowledge Management* and *Knowledge Management Research & Practice* during 2013-14 to show how the vast amount of data can be visualized. The paper demonstrates the value of big data text analytics in visualising data and improving knowledge management. By doing so, the article demonstrates the utility of big data text analytics as a method for the discovery of hidden knowledge and generation of new knowledge. The remainder of the paper is structured as follows: The second section deals with conceptual background; the third section discusses big data text analytics as a method; the fourth section presents the findings; and the final section of the paper is the core discussion and conclusions.

### 2. Conceptual background

#### Big Data Text Analytics

Big data is defined as huge amounts of structured and unstructured data comprising billions of data points or observations, which can be accessed in real time and is characterized by its volume, velocity, and variety (Brynjolfsson & McAfee, 2012; Einav & Levin, 2013; Laney, 2001; O’Leary, 2013). Big data has been suggested to be ‘raw’ in nature and is everywhere, but due to its complexity it is difficult to understand and interpret using traditional methods...
(Mackenzie, 2006). Manyika et al. (2011) labelled big data as the next frontier for competition, innovation and productivity growth. Big data text analytics is a process of extracting and generating useful non-trivial information and knowledge from structured and unstructured data (Chen et al., 2012), which through its categorization, visualization and interpretation can enables more effective KM (Chen et al., 2012; Davenport, 2013).

In this context big data text analytics role becomes even more salient in enabling the processes and practices of capturing and sharing of vast amount of data (Chen et al., 2012; Rajaraman & Ullman, 2011). There has been a growing use, and reliance, on big data in a variety of different industries and commercial contexts from finance, healthcare to supply chain domains (Chae, 2015; George et al., 2014). Big data text analytics applications have even been trialled in detecting epidemic diseases in society (Ginsberg et al., 2009), as a means of digital infectious disease surveillance. Arguably the growing role of big data analytics could lead to the reframing of what constitutes knowledge and how we engage with data and information (Boyd & Crawford, 2012).

As noted above, big data can be structured and/or unstructured, and may originate from multiple sources. Consequently big data is often not well understood due to its complexity. Understanding big data demands a combination of analytic tools and high-level skills that are often not widely available (Tambe, 2014). Indeed the challenge of big data is particularly in its interpretation, converting seemingly data into insights that can improve knowledge management (e.g. Cheung et al., 2005; van den Berg, 2013). In this context, big data text analytics plays an important role in generating valuable knowledge which otherwise would be impossible for organizations to source and share. This is in keeping with previous studies which have found out that ICT-enabled data analytics tools support both the acquisitions and sharing of information and knowledge (Jarle Gressgard et al., 2014).

Beyond answering 'know-what' questions with more information, big data text analytics can also be used to address 'know-why' questions which are critical for developing a competitive advantage through KM (Kogut & Zander, 1992; Witherspoon et al., 2013). Due to these characteristics Rae and Singleton (2015:2) regard big data as a ‘fluid, user-centred concept that emerges as a result of a relative imbalance between the data themselves and the constraints on collection, management and then synthesis by the analyst’. These views have led big data text analytics to be seen as a new approach to research, with its application
permitting the exploration of unique patterns and predicting future trends (Aiden & Michel, 2014). Scholars have also noted how the application of big data text analytics has become increasingly important in the discovery and solution of business problems (Mayer-Schönberger & Cukier, 2013). Similarly the use of data analytics, and text mining specifically, have generated significant research interest (George et al., 2014), with applications including predicting stock market (e.g. Chung, 2014).

Due to the volume of data available, big data text analytics play a key role in capturing and sharing key information (Chen et al., 2012). Similarly Lazer et al. (2009:722) suggest that big data text analytics offers ‘the capacity to collect and analyze data with an unprecedented breadth and depth and scale’. Due to these characteristics Boyd and Crawford (2012:6) suggested that big data text analytics ‘reframes key questions about the constitution of knowledge, the processes of research, how we should engage with information, and the nature and the categorization of reality’. Big data text analytics could transform personal knowledge management, with the role of individual knowledge workers becoming increasingly vital as well (Pauleen, 2009). Similarly scholars suggest that many organizations are developing information systems to facilitate the sharing and integration of knowledge (Alavi & Leidner, 1999; Jarle Gressgard et al., 2014). Despite the academic interest in big data, there is still a limited understanding about its opportunities and challenges, and particularly the paucity of research an enabler of KM (Davenport, 2013; LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2013; Watson & Marjanovic, 2013). This could be due to the inherent challenges associated with big data text analytics and the fact it is difficult to capture, store, analyze and visualize vast amount of data; or it could be the result of lack of availability of skilled big data analysts (e.g. Ahrens et al., 2011; Chen & Zhang, 2014; Tambe, 2014). Next we review characteristics of knowledge and knowledge management before exploring the links between knowledge management and big data.

**Knowledge and Knowledge Management**

As with big data, knowledge as a broad concept that has been classified and defined in many different ways in the extant literature (Nonaka & Takeuchi, 1995; Spender, 1996; Gao et al., 2008; van den Berg, 2013; Crane & Bontis, 2014). Knowledge has been defined as set of justified beliefs, which can be managed to enhance the organization's capability for effective action (Alavi & Leidner, 2001; Nonaka, 1994). There are acknowledged to be three major
KM processes, namely the acquisition, conversion, and application of knowledge (Gold, Malhotra, & Segars, 2001; Alavi et al, 2006; Kulkarni et al, 2007; Gasik, 2011). Knowledge acquisition refers to developing new knowledge from data, information, or knowledge (Gold et al., 2001; Magnier-Watanabe & Senoo, 2010). Knowledge conversion refers to making the acquired knowledge useful for the organization (Gold et al., 2001; Orzano et al., 2008) by structuring it or transforming tacit knowledge into explicit knowledge. Knowledge application refers to the use of knowledge to perform tasks (Sabherwal & Sabherwal, 2005). Thus, KM includes the firm's processes of acquiring new knowledge, converting knowledge into a form that is usable and easily accessed, and applying knowledge in the organisational setting (Verkasolo & Lappalainen, 1998; Gasik, 2011). KM processes enable organizations to capture, store and transfer knowledge efficiently (Grant, 1996; Magnier-Watanabe & Senoo, 2010), and within this context big data text analytics is becoming increasingly important (Chen et al., 2012; Davenport, 2013).

The most widely cited types of knowledge are those of explicit and tacit knowledge (Inkpen & Dinur, 1998; Polanyi, 2009; Crane & Bontis, 2014). Explicit knowledge is that which can be documented and (Nonaka, 1991), and can consequently be easily transmitted Kogut and Zander, 1992) and embedded in standardised procedures (Martin & Salomon, 2003; Nelson & Winter, 1982). Tacit knowledge, by contrast, is often implicit and not codified. Such knowledge is difficult to capture in the form of text and is context dependent (Crane & Bontis, 2014), often derived and shared through a process of learning by doing (Nonaka, 1994). Nonaka and Takeuchi (1995) posit that explicit and tacit knowledge are not mutually exclusive, but rather complementary with knowledge converted from one form to the other in some organisations.

The conversion of knowledge from one type to the other is not always an easy task for organizations, as organisations have to make systematic efforts to reap the benefits of tacit knowledge. It is in this context that the role of big data text analytics becomes vital in capturing, acquiring and sharing huge volumes of explicit knowledge which through big data text analytics may be interpreted through tacit insights (Davenport, 2013; Scarbrough & Swan, 2001). The knowledge-based view (KBV) of the firm considers knowledge and the ability to integrate individual knowledge in organizations as an important source of competitive advantage (Grant, 1996; Kogut & Zander, 1992). Indeed effective KM in organizations is defined as getting the most out of knowledge-based resources including
explicit and tacit knowledge (Sabherwal & Becerra-Fernandez, 2003; Dalkir, 2005; Vitari, 2011; Al-Sudairy & Vasista, 2012).

Knowledge management and big data share similar objectives, as both roles is to create competitive advantage for organizations (Chen et al., 2012; George et al., 2014; Grant, 1996), while big data text analytics allow firms to track and catalogue sources of external knowledge to enable effective sharing of knowledge (Chen et al., 2012; Davenport, 2013; Davenport & Prusak, 1998; Gold et al., 2001). Both roles are important as the bases for creating competitive advantage for firms, and neither can be pursued independently of the other (Alavi & Leidner, 1999; Nonaka, 1994). However, there is hardly any research on the mutual relationship between big data and knowledge management (Davenport, 2013; LaValle et al., 2013; Nonaka & Takeuchi, 1995; O'Leary, 2013).

**Big Data and Knowledge Management**

The relationship between big data and knowledge management is rooted in the knowledge-based view of the firm and it can provide an overarching theoretical framework (Davenport, 2013; Grant, 1996; Kogut & Zander, 1993). The KBV sees knowledge as a key source of competitive advantage, and suggests a similar reciprocal relationship between knowledge and knowledge management. On the one hand, knowledge serves as the basis for knowledge management, for example Grant (1996) notes the complementarity between different kinds of knowledge. Big data text analytics, to this end, has the potential to capture and utilise different sources of explicit and tacit knowledge, and produce new depth of knowledge as a basis of more effective decision making (e.g. Grant, 1996; Kitchin, 2013; Laney 2012). Similarly, knowledge management can improve and strengthen the combinations of knowledge resources (Cockrell and Stone, 2010).

Following the underlying assumptions of KBV, the above literature provides considerable basis for expecting big data to inform knowledge management. The use of big data text analytics affects the processes for absorbing the new knowledge coming from different sources (Cohen & Levinthal, 1990; Andreeva & Kianto, 2011), applying knowledge (Grant, 1996), and its conversion from one form to another (Sabherwal & Becerra-Fernandez, 2003). The extant literature provides reason to expect that big data text analytics will enhance KM by enabling enhanced knowledge creation, integration and sharing (Chen et al., 2012; George...
et al., 2014; King, 2009). Moreover, the use of big data text analytics can further improve KM processes and transactive memory systems in leveraging the value of big data (Alavi & Leidner, 1999; Argote, McEvily, & Reagans, 2003; O’Leary, 2013).

3. Method of Big Data Text analytics

The growing amount of data available to decision-makers in organisation is becoming overwhelming. In practice this means big data cannot be processed manually, and consequently KM tools are required to support more informed and effective decision making in a timely manner. It is in this context that big data text analytics tools offers a means to identify patterns and other non-trivial information and knowledge from vast amount of both structured and unstructured data that may otherwise not be visible. This new and emerging research domain endeavours to address potential data overload issues by using techniques of data mining, information retrieval, machine learning and knowledge management (Feldman & Sanger, 2006; Chen et al., 2012; Davenport, 2013).

There has been a rise in the use big data text analytics in scholarly research. The use of big data text analytics in the social sciences (King, 2014; Varian, 2014), and particularly regional and information sciences (e.g. Chen et al., 2012; Chen & Zhang, 2014; Gandomi & Haider, 2015; Rae & Singleton, 2015), as a means of data collection, processing, extraction and analysis (Chen et al., 2012; Chaudhuri et al. 2011, Watson & Wixom 2007). Big data text analytics offer a high potential value and wide applications in diverse areas to develop a competitive advantage (Chen et al., 2012). For instance, it has been used in understanding supply chain related issues, and guests’ experiences and preferences for hotels (Chae, 2015; Xiang, Schwartz, Gerdes, & Uysal, 2015), as well as in mapping digital businesses (Nathan & Rosso, 2015). However, the application of big data text analytics is still lacking in different fields including KM, which is in part due to lack of understanding about the possibilities of big data text analytics (e.g. Davenport, 2013; LaValle et al., 2013), and of the big data capture, processing and analysis techniques available (see Chen & Zhang, 2014).

In this article, we apply the method big data text analytics to articles published in two of the leading journals in the domain of knowledge management as a means to demonstrate the insights that can be generated. Over a 2 year period (2013 and 2014) a total of 196 articles were published on variety of topics from the Journal of Knowledge Management and
Knowledge Management Research & Practice. All of the articles were downloaded and converted them into plain text format to facilitate processing in the subsequent analysis. During the conversion process, we removed graphs, tables, authors’ related information, journal name and references pages as these could create potential repetitions into the analysis. We then applied text analytics techniques with the aim to show the application of big data text analytics techniques in the capturing, acquisition and sharing of knowledge (Davenport, 2013; Grant, 1996; Scarbrough & Swan, 2001). We applied text analytics approach on the entire document instead of picking particular sections of the documents as we wanted to avoid self-selection bias in the analysis.

Applying big data text analytics approach to the entire document is more comprehensive as it can give depth of information about a particular topic. In this way insights can be identified and presented more effectively and intuitively (Simoff et al., 2008). We also applied custom stop words to overcome any potential bias related issues arising from the analysis, for instance ‘journal of knowledge management’ or the words ‘they’ and ‘research & practice’ could potentially show up as the most frequent word in the words list. We added such words to the stop word query in order to have a reliable list of most frequent words. By applying text analytics techniques on the 196 articles we identified the 50 most frequent words used across these articles. This approach further helps in capturing important knowledge, but also the organization and analysis of knowledge. Various tools for big data text analytics are available to organizations in capturing, acquiring and sharing organization-wide knowledge and table 1 list some of these tools.

<table>
<thead>
<tr>
<th>Big data text analytics tools</th>
<th>Brief Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apache Hadoop Distributed File System</td>
<td>Open source Java based software framework that enables the storage, distribution of large data set on multiple servers</td>
</tr>
<tr>
<td>MapReduce</td>
<td>Provides the interface for the processing, generation and distribution of large data sets. It provides clustered scale-out large data processing capabilities and solutions</td>
</tr>
<tr>
<td>Hive</td>
<td>Data warehousing Hadoop infrastructure, provides data summarization, data query and analysis</td>
</tr>
<tr>
<td>Zookeeper</td>
<td>Facilitate a centralized infrastructure, provide synchronization across a cluster of</td>
</tr>
</tbody>
</table>
computers.

<table>
<thead>
<tr>
<th>Database</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBase</td>
<td>Open source, distributed and non-relational database management systems, based on non-SQL approach</td>
</tr>
<tr>
<td>Cassandra</td>
<td>A distributed database system, handles big data distribution across multiple computers</td>
</tr>
<tr>
<td>Apache Pig</td>
<td>Platform for the analysis of large data sets that consists textual language- PigLatin (process both structured and unstructured data)</td>
</tr>
</tbody>
</table>

Source: based on various sources including Raghupathi and Raghupathi (2014).

4. Findings

**Big data text analytics and KM**

The 196 articles published during 2013 and 2014 were analysed using big data text analytics. The findings suggest that big data text analytics can be an enabler of effective KM, as it generates quality of knowledge. Figure 1 is a word cloud of the most frequent words across the 196 articles. For reasons of space and readability, we only show the most frequent words that appear at least 50 times in the entire sample of articles. The larger and darker the word is, the more frequent it appears in 196 articles. This form of visualization helps in the quick extraction of key knowledge from a large volume of data, in this case the 196 published articles.

**Figure 1. Most Frequent Words across 196 articles published in the Journal of Knowledge Management and the Journal of Knowledge Management Research & Practice during 2013-14**
The frequency of words appearing in the 196 articles published in 2013 and 2014 are shown in Table 2, which clearly shows the focus to range from the organizational level to individual level. The most frequent words 'organizational', 'sharing', 'information', 'innovation', 'social', 'learning' and 'transfer'. These are the terms that are important for KM and creating competitive advantage for firms (Alavi & Leidner, 1999; Argote et al., 2003; Grant, 1996). Such visualizations of both unstructured and structured data facilitate KM and improve timely decision making. These findings show the utility of text analytics for capturing, acquiring and constructing important knowledge (Chen et al., 2012; Davenport, 2013; Grant, 1996). Knowledge of the most frequent and popular words is beneficial for understanding the focus of the research and its emerging domains.

One of the central challenges for organization has been how to codify and share knowledge (more) effectively. Big data text analytics increases the capacity to capture, process and analyse data, as well as it speed compared to traditional KM tools. The importance of visualizing knowledge also further aids in the codification of important knowledge, which was another limitation of traditional KM tools. The analysis also indicates that research in the domain of knowledge management have been mostly done at the organisation level, with less focus on micro level issues concerning knowledge management.

Table 2. Most frequent words appeared across 196 articles published during 2013-14 in the Journal of Knowledge Management and the Journal of Knowledge Management Research & Practice
In addition to the combined analysis, we have also conducted separate analysis for each of the journals to see how the pattern changes across the two journals. The Figure 2 is a comparative analysis of the two journals, showing the difference in rank of the most frequently appearing words ranked in the top 50 of both journals. The pattern extrapolated from the data also shows that while the terms are shared, that certain terms tend to be more dominant in each journal. The analysis of big data text analytics serves to show respective specialisms and priorities in each of the journals as is reflected by the 196 articles.
Beyond analysing the frequency of words, the data can also be analysed to cluster the key information and the findings indicate important patterns of clusters. For example, sharing, information, tacit, people, innovation, international, processes, systems, communication, team, practices are part of transfer cluster, whereas individual, social, trust, capital, intellectual, relationships constitute as another cluster called networks cluster as presented in the cluster analysis shown in Figure 3. Using the 196 journal as an example the findings demonstrate that big data text analytics has the potential to improve KM by clustering important knowledge and such clustering helps for the better organization and analysis of key data.
In table 3 we list four dominant clusters and the most frequent terms that appear in each. This approach highlights the nature of clusters and contextual information that can be distilled using big data text analytics. The findings provide important details about the frequency of topics appearing within each of the clusters. These key clusters also point toward the important topics in the wider knowledge management literature such as processes, people, technology, networks, innovation, learning and value creations for the development of organizational competitive advantage (Almeida, Song, & Grant, 2002; Grant, 1996). This analysis also indicates that big data text analytics driven clustering can serve important role for the enhancement of knowledge management in organization by text categorization, identification of semantic relationship, and visualization. Thus analytics based clustering of data can potentially mitigate information overload problem and facilitate new knowledge creation and its dissemination.
Table 3: The Most Frequent Terms appearing in each Cluster

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Most frequent terms within the cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>1- Transfer</td>
<td>sharing, information, tacit, people, innovation, international, processes, systems, communication, team, practices</td>
</tr>
<tr>
<td>2- Networks</td>
<td>individual, social, trust, capital, network, intellectual, relationships</td>
</tr>
<tr>
<td>3- Resources</td>
<td>creation, practices, structure, activities, strategic, technology</td>
</tr>
<tr>
<td>4-Organizational</td>
<td>external, capacity, customer, strategic, creation, networks, relationship, innovation, strategy, internal, managers, processes</td>
</tr>
</tbody>
</table>

5. Discussion and Conclusions

The purpose of this article is to examine the potential of big data text analytics as a method by highlighting the depth of knowledge that can be generated from big data for effective knowledge management through the visualisation, organization, interpretation and analysis of information that would not otherwise be feasible. Despite the increasing interest in big data and business analytics, the topic remains underdeveloped within the KM literature, and its role as an enabler of KM in particular is not well understood (Davenport, 2013; LaValle et al., 2013; Watson & Marjanovic, 2013). Yet the big data text analytics approach presents opportunities for organizations, the majority of whom are ill-informed about how big data text analytics can support making informed and timely decisions through the generation of valuable knowledge and its subsequent interpretation (e.g. Davenport, 2013; LaValle et al., 2013).

Against this backdrop, performing big data text analytics on the 196 articles published in two of the leading journals begins to highlight the value of big data text analytics to capture and visualize knowledge that would otherwise be impossible to process and codify. The finding drawn from the 196 papers as an example of big data text analytics as an example of effective KM, and in contrast to traditional KM systems, highlight the utility of big data text analytics
in KM (Malhotra, 2005). The key findings of this study in the context of KM are threefold. Table 4 presents the summary of the key findings.

**Table 4. Summary of key Findings**

<table>
<thead>
<tr>
<th>Utility of big data text analytics and role in knowledge management</th>
<th>Key findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Capturing and visualizing most frequent words from a vast amount of both structured and unstructured data</td>
<td>✓ big data text analytics plays an important role for the timely transfer, sharing and management of vast amount of data. ✓ generations of key words through text mining of big data can help in the internalization, sharing and effective management of key knowledge assets. ✓ text analytics of big data could potentially overcome information overload problems and facilitate both codification and personalization knowledge management strategies.</td>
</tr>
<tr>
<td>2. Organization of dispersed knowledge into categories and key words</td>
<td>✓ enable the effective codification of important knowledge. ✓ stop words can be applied for real time filtering of knowledge. ✓ making the delivery and optimization of relevant knowledge available to multiple constituents within an organization for effective as well as timely decision-making. ✓ knowledge generated through text analytics of big data is fine-grained and precise and organizations could potentially benefit from such an efficient, and high-quality knowledge thus enabling the effective retention and sharing of valuable knowledge.</td>
</tr>
<tr>
<td>3. Analytics oriented clustering analysis</td>
<td>✓ provide important contextual details through the clustering of important knowledge. ✓ clustering and the visualization of vast amounts of data can provide an interface that is well suited for effective knowledge management and improved &amp; timely decision making. ✓ analytics oriented clustering and categorization of big data help KM systems to improve the relevancy, quality and timeliness of knowledge that is captured, codified, and shared. ✓ quality of knowledge is generated by enhancing categorization of contents and interpretation of big data and enables effective KM.</td>
</tr>
</tbody>
</table>
First, we demonstrate the utility of big data text analytics by capturing and visualizing most frequent words. Figure 1 and Table 2 indicate that the focus of the research in KM have focused on understanding KM at the organisational level issues related to KM such as organizational knowledge creation, sharing, transfer and innovation. However, less attention has been paid to micro level issues impacting knowledge management such as the role of individuals in the creation and transfer and management of knowledge, routines, processes, as well as particular managerial actions, communications- the micro-foundations of KM. As such, big data text analytics plays an important role for the timely transfer, sharing and management of such vast amount of data as this study has demonstrated. The generations of key words through big data text analytics can help in the internalization, sharing and effective management of key knowledge assets (Davenport, 2013; LaValle et al., 2013; Nonaka & Takeuchi, 1995; O'Leary, 2013), which would otherwise be very time consuming through the use of traditional KM methods (e.g. Chen et al., 2012). For instance, Hansen et al. (1999), noted that it was a mistake on the part of the traditional KM systems to apply codification and personalization strategies at the same time, and indicated that companies should only focus on one of them. As the findings of the study indicate, big data text analytics could overcome such problems and facilitate both codification and personalization knowledge management strategies. Put differently, the conversion of unstructured and structured forms of data could generate explicit knowledge that may lead to easier absorption (Cohen & Levinthal, 1990), which is exemplified by the two journals reviewed. Traditional KM systems have faced information overload issues and emphasized on predetermined workflows and rigid ‘information-push’ approaches (Malhotra, 2005). Big data text analytics can overcome information overload with so-called information-push strategies making important knowledge accessible and available.

Second, the findings indicate that big data text analytics facilitates the organization of dispersed knowledge into categories and key words, which enable the effective codification of important knowledge. As this study shows that text mining enabled stop words can be applied for real time filtering of knowledge, thereby making the delivery and optimization of relevant knowledge available to multiple constituents within an organization for effective as well as timely decision-making. The other implications for KM is that big data text analytics helps analyzing all the data rather than choosing a particular set of data. The knowledge generated through text mining of big data is fine-grained and precise, therefore, organizations
could benefit from such an efficient, and high-quality knowledge thus enabling the effective retention and sharing of valuable knowledge (e.g. Kogut & Zander, 1992).

Big data text analytics can be utilized in the capture of both structured and unstructured data to enable the generation of new depth and codification of knowledge as the basis of competitive advantage (e.g. Grant, 1996; Kitchin, 2013; Laney, 2012). Traditional KM approaches have often been criticized for being expensive, time consuming and unable to provide timely knowledge to the right persons, whereas the application of big data text analytics, as demonstrated in this paper, suggests that rich and quality knowledge can be generated through such approach (e.g. Chen et al., 2012; Laney, 2012). The richness and generation of quality knowledge then facilitates the effective sharing of this knowledge across the organization, thus improving workers productivity and 'transactive memory systems' (e.g. Argote et al., 2003; Argote & Ren, 2012; Lewis & Herdon, 2011). The generation of quality knowledge through big data text analytics has wider implications for organization since knowledge has been suggested to be important for developing competitive advantage in the knowledge-based economy (Grant, 1996; Kogut & Zander, 1992; Spender, 1996), and most of the organizations are not well-informed on the opportunities offered by text mining of big data for making timely decisions (e.g. Davenport, 2013; LaValle et al., 2013).

Third, cluster analysis (see figure 3 and table 3) can be performed through the use of big data text analytics techniques to show important contextual details through the clusters. Clustering and the visualization of vast amounts of data can provide an interface that is well suited for effective knowledge management and decision making (e.g. Chen, 2001). Such clustering can generate insights to enhance knowledge management, to understand the relational aspects of themes. All four clusters show common themes as well as important differences, with terms such as ‘processes’, ‘practices’, ‘innovation’, ‘creation’, and ‘networks’ occurring frequently across the four clusters. These findings indicate the utility of big data text analytics based clustering as a key approach of optimizing, synthesizing and capturing important knowledge in the discovery of hidden knowledge and generation of new knowledge that can support KM.

The big data text analytics clustering approach offers important insights for not only the effective deployment of knowledge assets, but also making informed decisions in the
organization. These findings indicate that big data applications and tools are important enablers of KM (George et al., 2014; Grant, 1996). The results of this study further supports the views of King (2009:87) who noted: ‘text analytics have great potential utility for knowledge management’. Through the application of big data text analytics, the paper has shown how the visualization of important topics as well as clustering-based approaches can show the categorization and proximity of topics, which can be utilized by the KM systems to improve the relevancy, quality and timeliness of knowledge that is captured, codified, and shared. Again this has significant scope to improve knowledge management and organizational performance. The visualization of big data text analytics can also enable effective KM, by presenting data that would otherwise be very time-consuming and difficult to process through the use of traditional KM systems.

5.1. Implications for Research and Practice

The study has several implications for research and practice. The study demonstrates the value of big data text analytics in the discovery of hidden knowledge and generation of new knowledge that would otherwise not be possible through traditional KM systems. The goal of the KM system is to ensure that key knowledge assets are effectively codified and shared. In the context of this paper the focus was the 196 papers, but this could equally be applied to a wider variety of documents to provide insights about individual and organizational practices. In this way managers would be well placed to take advantage of novel approaches, such as big data text analytics, that enable big data to be interrogated and used to inform more effective and timely decision making (e.g. Chen et al., 2012).

The usefulness of such techniques is likely to grow in the coming years as organization adopts new technologies that might have unanticipated consequences such as managing offshoring work through IT-enabled technologies or material source tracking and collecting key clients information. The findings of this article demonstrate the usefulness of big data text analytics in capturing, storing and sharing of knowledge in a timely fashion which can enhance KM processes with quality knowledge (e.g. Davenport, 2013). Furthermore, big data text analytics approaches can be applied to a variety of databases such as consumer complaint databases or even Twitter feeds to capture the temporal and spatial trends in opinions, habits, or events as well as online social interactions (Golder & Macy, 2011, 2014), thus enabling effective knowledge codification and sharing.
The findings presented in this article on the use of big data text analytics provides important insights on the usefulness of using big data text analytics for knowledge management. The use of big data text analytics has wider implications, as we have shown in this article; it provides rich contextual detail so organization can focus their attention on where to dedicate their efforts and how to create value by utilizing valuable assets in this case the application of big data text analytics. However, there are still inherent challenges of using big data, as the skills needed may not be available in different context thus hindering effective knowledge management (Tambe, 2014). In the context of the empirical focus of this paper, the findings indicate that research on knowledge management has mostly been conducted at the organizational level topics and less attention paid to how the micro level processes such as routines and individual level actions impacts knowledge management through the creation, transfer and management of knowledge. In this way, the study highlighted the important application and utility of big data text analytics by demonstrating that depth of knowledge can be generated for effective knowledge management. One of the key contributions of this study is demonstrating the quality and nature of knowledge generated through the utilization of big data text analytics methods. In addition, as a by product, we conceptually develop the link between big data and knowledge based view (Grant, 1996), indicating that it is one of the key theories that provides a much needed theoretical lens to investigate the role of big data in various settings (e.g. George et al., 2014).

6. Limitations and Future Research Directions

This is an exploratory study that applied big data text analytics to demonstrate its value for the effective generation and codification of rich and quality knowledge which enables knowledge management and more effective decision making. Keeping aside its contributions, it also has some limitations and these could serve an important avenues for further research. The application of big data text analytics was applied only on articles published in two leading knowledge management journals over a two year period. Future studies could apply big data text analytics over a longer period to examine trends and consider more sources to develop the robustness of the findings.

Indeed different sources of unstructured data, such as twitter feeds and other that is available on the web, has the potential to generate useful insights and in so doing support effective knowledge management of organizations. On this basis future research could build onto these findings and apply big data text analytics in different context and on different research topics.
such as the role of individuals in the age of big data which would support the case for focusing on the individual as well as the organization as a unit of analysis.

Acknowledgments
The authors would like to thank the Guest Editor of this issue, David Pauleen, and two anonymous reviewers for their insightful suggestions and comments that helped to improve the quality of this manuscript.

References


Polanyi, M. 2009. The tacit dimension, University of Chicago Press, Chicago, IL.


**About the Authors**

Zaheer Khan is a Lecturer (Assistant Professor) in Strategy & International Business at the Sheffield University Management School, University of Sheffield, UK. His research focuses on global technology management, with a particular interest in knowledge transfer through FDI to emerging economies. He received his PhD from the University of Birmingham, UK. His work has been published in the *International Business Review*, the *Global Strategy Journal*, The *Journal of International Business Studies*, *Industrial Marketing Management*, *Critical Perspectives on International Business* and *International Marketing Review*, among others.

Tim Vorley is a Professor and Chair in Entrepreneurship at the Sheffield University Management School. He also heads the Centre for Regional Economic and Enterprise Development (CREED). His research focuses on organisational dynamics of entrepreneurialism, innovation, regional development. His work has appeared in *Economic Geography*, *International Small Business Journal*, *Entrepreneurship and Regional Development*, *Education and Training*, *European Journal of Industrial Relations*, *Environment & Planning C*, and *Growth and Change*, among others.