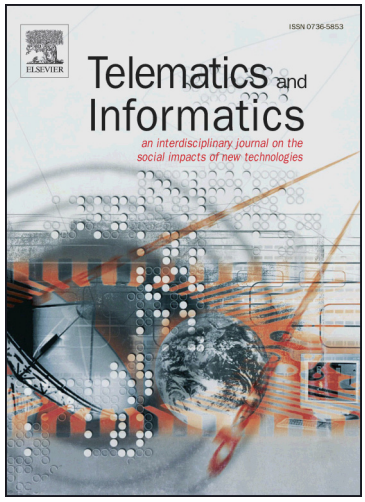

Unstructured big data analytics for retrieving e-commerce logistics knowledge

Pei-Ju Wu, Kun-Chen Lin

PII: S0736-5853(17)30539-7
DOI: https://doi.org/10.1016/j.tele.2017.11.004
Reference: TELE 1030

To appear in: *Telematics and Informatics*

Received Date: 18 August 2017
Revised Date: 25 October 2017
Accepted Date: 6 November 2017

Please cite this article as: Wu, P-J., Lin, K-C., Unstructured big data analytics for retrieving e-commerce logistics knowledge, *Telematics and Informatics* (2017), doi: https://doi.org/10.1016/j.tele.2017.11.004

Unstructured big data analytics for retrieving e-commerce logistics knowledge

First author (Corresponding author)
Pei-Ju Wu, Ph.D.
Assistant Professor, Department of Transportation and Logistics
Chief Innovation & Development Officer, Innovation Center for Intelligent Transportation and Logistics
Feng Chia University
Address: No.100, Wenhwa Road, Seatwen, Taichung 40724, Taiwan, ROC
Tel：+886-4-24517250 Ext. 4668
E-mail: wupj@fcu.edu.tw

Kun-Chen Lin, M.S.
Department of Transportation and Logistics
Feng Chia University
Email: zxz2308525@yahoo.com.tw

Abstract

The divergent evolution of e-commerce has complicated its correspondingly logistics management. However, few studies have explored e-commerce logistics business models via big data analytics. Hence, this investigation explores e-commerce logistics business models from unstructured big data. Specifically, this work develops a hybrid content analytical model to scrutinize essential knowledge of e-commerce logistics. The empirical results of the proposed model incorporate theories of resource dependence theory (RDT) and innovation diffusion theory (IDT) to generate logistical strategies. Ten critical themes of e-commerce logistics from topic mining are “Southeast Asia’s e-commerce logistics payments”, “E-commerce order management”, “E-commerce logistics cloud services”, “E-commerce logistics package management”, “Europe e-commerce trends”, “India’s e-commerce logistics”, “E-commerce distribution

management", "Tax policies", "E-commerce logistics platforms", and "E-commerce logistics networks". Moreover, the fundamental rule of "cross-border e-commerce logistics" is uncovered by the association rules model. The proposed hybrid content analytics framework provides a research foundation for e-commerce logistics management. Furthermore, e-commerce logistics can be implemented by vital strategies: "Establish inter-organizational and technical collaboration to create positive operations performance" and "Comprehend law, policy, and cultural differences to customize appropriate technologies of e-commerce logistics".

Keywords: E-commerce; logistics management; big data analytics; content analytics; unstructured big data

1. Introduction

The success of an e-commerce company correlates strongly with its logistical performance (Cullen and Taylor, 2009; Hsiao et al., 2017) and distribution systems (Lu and Liu, 2015). Superior e-commerce logistics management can raise a company's performance (Bhattacharjya et al., 2016; Cho et al., 2008). The characteristics of various worldwide environments, such as the logistical capacities of different countries, are more likely to induce different logistical operations. Nevertheless, few studies have comprehensively scrutinized e-commerce logistical business models from a global perspective

E-commerce management requires superior delivery services (Iyer et al., 2004), perfect logistics information systems (Auramo et al., 2005), secure e-procurement processes (Vaidyanathan et al., 2012), inter-organizational collaboration (Gimenez and Ventura, 2005), convenient mobile banking (Shaikh and Karjaluoto, 2015), efficient mobile channel (Park and Lee, 2017), and suitable e-collaboration tools (Chong et al., 2009), etc. Improved e-commerce logistics management can also enhance a company's financial performance (Johnson et al., 2007). E-commerce investments need to consider sources for many different possible configurations such as insourcing and outsourcing (Sharp, 2007). Moreover, e-commerce applications influence the supply chain network design over time (Piramuthu, 2005). Additionally, firms need to consider the related logistics activities to obtain the benefit from investing in e-commerce (Delfmann et al., 2002; Mangiaracina et al., 2015; Yu et al., 2015). Hence, how to effectively manage e-commerce logistics is a topic warranting

immediate attention

E-commerce logistics can be analysis in various ways, including a case study approach (Hinson, 2010), multiple discriminant analysis (Saprikis, 2013), interpretive structural modeling (Toktas-Palut et al., 2014), an investment model (Nguyen, 2013), etc. Nonetheless, in market surveys it is usually hard to obtain a large sample size to reflect overall consumer expectations (Singh et al., 2017). Furthermore, big data analytics in the supply chain can be employed to provide up-to-date and meaningful information (Kache and Seuring, 2017) and create value-added operations (Matthias et al., 2017; Tan et al., 2015; Wu et al., 2017). However, few studies have exploited unstructured big data related to e-commerce logistics.

Based on the above motivations, the purpose of this study is to develop unstructured data analytics for retrieving valuable business models of e-commerce logistics. Moreover, theories of resource dependence theory (RDT) and innovation diffusion theory (IDT) are integrated with the empirical results of big data analytics to generate beneficial strategies for e-commerce logistics management. Specifically, this study establishes a hybrid content analytics including web crawler, syntactic analysis, data preprocessing, topic mining, and association rules to extract essential patterns of e-commerce logistics from massive unstructured documents.

The main contributions of this study can be summarized as follows: Few studies have used unstructured big data analytics to analyze e-commerce logistics business models. This pioneering empirical study offers comprehensive

information about how to turn unstructured big data into constructive insights. RDT and IDT theories are applied to the empirical results of the proposed hybrid content model to derive worthwhile e-commerce logistics strategies and to furnish useful knowledge of e-commerce logistics. Moreover, based on the empirical results of the unstructured big data analytics, e-commerce managers can appropriately allocate limited resources to manage their e-commerce logistics business effectively.

The rest of this paper is organized as follows. Section 2 explains the data analytics with a hybrid content analytics. Section 3 shows the empirical results of unstructured big data analytics. Section 4 provides e-commerce logistics strategies with RDT and IDT theories. Finally, Section 5 draws conclusions.

2. Hybrid content analytics

This study devises big data analytics with a hybrid content analytics (Fig. 1) involving web crawler, syntactic analysis, data preprocessing, topic mining, and association rules to tackle critical issues of e-commerce logistics. The proposed unstructured big data analytics for e-commerce logistics is briefly described as follows.

- Web crawler

  The first step of the proposed hybrid content analytics is to choose representative websites related to ecommerce logistics based on domain knowledge. Moreover, keyword search in representative websites can be used to obtain unstructured data of interest. Additionally, the useful information

regarding how to crawl data can be found in Hu et al. (2015).

- Syntactic analysis

  To see the big picture of the collected data, syntactic analysis can be conducted to identify what kind of terms are involved in the collected data. Specifically, natural language processing (NLP) with text parsing methods seeks to analyze text data and to dismantle sentences into constituent terms (Casagrande et al., 2014; Zhai and Massung, 2016). Then, the properties of each term can be obtained such as the frequency of the terms (Casagrande et al., 2014; Zhai and Massung, 2016).

- Data preprocessing

  After syntactic analysis, the large collection of parsed terms must be further processed to ensure that the chosen data is accurate and intact. A blacklist should be compiled for incomplete terms, data sources, parts of speech, etc. Moreover, several parts of speech namely verbs (v.), adverbs (adv.), and articles (art.) could be removed (Farseev, 2014; Gunjan, 2013) to establish a concise parsed terms. Additionally, a synonym list could be established to improve the accuracy of topic mining results (e.g., e-commerce and ecommerce are synonymous). The blacklist and synonym list need to be continuously modified until reasonable results are found. In addition, the valuable information concerning how to implement data preprocessing can be found in Hu et al. (2014).

- Topic Mining

  The singular value decomposition (SVD) can be utilized to decide which

parsed term is appropriate to a topic (Durmusoglu, 2016; SAS Institute Inc., 2016). Equation (1) represents the singular value decomposition. Herein, μ is a term-by-document frequency matrix comprising parsed terms related to e-commerce logistics; γ is a diagonal matrix with singular values; α and β are matrices with orthonormal columns (Durmusoglu, 2016; SAS Institute Inc., 2016). Then, different θ topics of e-commerce logistics are generated based on the calculation of the SVD (Durmusoglu, 2016; SAS Institute Inc., 2016).

$$\mu = \alpha \times \gamma \times \beta^t \quad (1)$$

- Association Rule

The association rules model of the Apriori algorithm (Ting et al., 2014; Wang and Yue, 2017) can be performed to create implicit association rules. In particular, parsed terms grouped into distinct topics are further explored to retrieve the latent relationships of parsed terms related to e-commerce logistics. Criteria for choosing appropriate association rules include Support, Confidence, and Lift. The support measure represents how frequently the e-commerce logistics items occur together. The confidence measure delineates the conditional probability of both e-commerce logistics items A and B occurring given the occurrence of e-commerce logistics item A. The lift measure describes e-commerce logistics items A and B occurring more/less frequently than expected.

Please insert Figure 1 about here

---

3. The empirical results

This investigation accumulated unstructured big data of e-commerce logistics from professional websites involving E27, E-commerce news (Africa), E-commerce times, E-commerce news (Europe), Inside Retail, Practicale-commerce, Supply chain 24/7, Supply Chain Management, The Wall Street Journal, Web of Science, Bigcommerce and DNA covering 2,370 documents in the period from 2010 to 2016.

Web crawling by python programming and software is initially utilized to extract information related to e-commerce logistics. Nonetheless, the progress of web crawling was hindered by the following circumstances. First, lots of websites block an Internet bot from directly acquiring data or only limited information can be retrieved by an Internet bot. Second, an Internet bot is more likely to extract data that is not relevant to research interests. Accordingly, several research assistants are hired to manually crawl data related to e-commerce logistics. When research assistants extract data from e-commerce logistics websites, numerous duplicate documents are found in various websites. Restated, the representative websites collect comprehensive information related to e-commerce logistics,

which implies that representative websites guarantee the completeness of the gathered dataset. A careful pruning of documents (elimination of unrelated and duplicate documents) resulted in a dataset with 2,370 documents.

In syntactic analysis, the parts of speech of terms were inspected to discover the content structure of each document. The analytical results of the top three part of speech of terms were Noun (282,826), Noun group (5,944), and Proper noun (35,369).

In data preprocessing, unnecessary parsed terms are eliminated such as prepositions, participles, adverbs, verbs, adjectives, conjunctions, prefixes, numbers, and punctuation. To avoid potential noise, a blacklist and synonym lists were established, mainly based on the suggestions of SAS® Text Miner 14.2: Reference Help (SAS Institute Inc., 2016). After initial data preprocessing, incomplete terms without meaning appear, primarily due to the process of syntactic analysis that just directly transforms massive text data into smaller parsed terms. All of the terms that do not represent meaningful vocabulary should be blacklisted and eliminated. Moreover, even though some parsed terms are meaningful, they still should be deleted because they are merely data sources (e.g., syntactic analysis may retrieve website names or journal names as parsed terms). Eventually, the remaining parsed terms were Nouns (149,773), Group Nouns (4,904) and Proper Nouns (17,694). Thereafter, the function of topic mining in SAS Text Miner 14.1 was implemented to extract essential topics related to e-commerce logistics. It should be noted that the above process should be executed repeatedly by trial and error to obtain an ideal result.

3.1. The empirical results of topic mining

The proposed unstructured big data analytics of e-commerce logistics was accomplished through SAS Text Miner 14.1 using its default values. Initially, numerous issues mentioned in data preprocessing arose, such as incomplete terms and unnecessary terms. Hence, the generation process of topic mining should be suspended and cycled back to data preprocessing. The cycle between data preprocessing and topic mining should be performed over and over until all issues mentioned in data preprocessing are tackled. Ultimately, this investigation explored ten critical topics related to e-commerce logistics. Table 1 represents the weights of the different terms for the ten generated topics. Unlike clustering, the function of topic mining in SAS Text Miner allows documents and terms be allocated to more than one topic based on their association with a topic (SAS Institute Inc., 2016). Further details can be found in SAS® Text Miner 14.2: Reference Help (SAS Institute Inc., 2016). Moreover, the assigned terms in the ten topics can be viewed as clues to retrieving crucial e-commerce logistics cases from original unstructured documents. Accordingly, ten themes related to e-commerce logistics were generated with practical cases (Fig. 2).

Based on the analytical results of the proposed hybrid content analytical model, the characteristics of the e-commerce logistical trend from 2010 to 2016 involved the following themes: “Southeast Asia’s e-commerce logistics payments”, “E-commerce order management”, “E-commerce logistics cloud services”, “E-commerce logistics package management”, “Europe’s e-commerce

trends", "India's e-commerce logistics", "E-commerce distribution management", "Tax policies", "E-commerce logistics platforms", and "E-commerce logistics networks".

- Topic 1 can be denoted as "Southeast Asia's e-commerce logistics payments". One of the original unstructured data collections represents that Onigi companies integrated with ESL Express, a logistic company, to provide Collect on Deliver (COD) payments so that all their hosted stores could pay via COD Indonesia (Lukman, 2013). In addition, Onigi companies let small- and medium-sized businesses (SMEs) set up their e-commerce stores on Facebook and Twitter, and provide online payment services for SMEs (Lukman, 2013).
- Topic 2 can be represented as "E-commerce order management". One of the original unstructured data collections mentions UPS Order Watch, which is a vendor booking system software application (Burnson, 2012). UPS Order Watch monitors vendor bookings, order details and ship windows to warn e-commerce providers of order volume anomalies, and also provides an online system for approvals so that vendors and customers can communicate in real time (Burnson, 2012).
- Topic 3 can be indicated as "E-commerce logistics cloud services". One of the original unstructured data collections shows that providing open source Swift in the cloud can let app developers use the same language to establish mobile apps (Campisi, 2016).
- Topic 4 can be given as "E-commerce logistics package management". One

of the original unstructured data collections mentions the website SolvingMaze.com, which can check item size to ensure that a shipment uses an appropriate box size (Salvatori, 2014).

- Topic 5 can be denoted as "Europe e-commerce trends." The original unstructured data collections reveals that Finns prefer directly using bank transfers to pay, while Danes prefer using credit card to pay (Ecommerce News, 2014a) as well as online shoppers in the Netherlands, Spain and Germany consider free shipping and delivery within 3-5 days to be important requirements (Ecommerce News, 2015).
- Topic 6 can be represented as "India's e-commerce logistics." One of the original unstructured data collections mentions the BlackBuck app, which customers use to select the transport attributes such as material type, truck type and number of trucks (Rao, 2016). The truck owners then inform customers to confirm that the order is correct (Rao, 2016). Moreover, BlackBuck's app provides an interface in six languages for the truck owners, and sets the GPS on every truck, so that customers can safely release payment on the platform (Rao, 2016).
- Topic 7 can be indicated as "E-commerce distribution management". One of the original unstructured data collections discusses Amazon Prime and ShopRunner providing free two-day shipping, which establishes barriers for small and medium-sized online retailers to complete with their speedy distribution. (Roggio, 2013).
- Topic 8 can be given as "Tax policies". One of the original unstructured data

collections notes that items with value less than US$320 are shipped duty-free and have simple customs and tax procedures in Singapore (Wong, 2014). In Malaysia, items with value below US$160 are shipped duty-free, but customs and tax procedures are complex, as the tax depends on the product type (Wong, 2014).

- Topic 9 can be denoted as "E-commerce logistics platforms". One of the original unstructured data collections mentions the use of multiple channels such as INTTRA, GT Nexus and CargoSmart to provide shippers with an e-commerce platform including sailing schedules, booking, shipping instructions, loading bills, track & trace and reports (Burnson, 2016a).
- Topic 10 can be represented as "E-commerce logistics networks". One of the original unstructured big data collections indicates that a sizeable Chinese electronics retailer, Suning, with its vast logistics network has integrated with Alibaba's Cainiao logistics platform, resulting in 2-hour deliveries (Millward, 2015).

Please insert Figure 2 about here

3.2. Association Rules

An essential association rule ("delivery & customer" with support=40.94,

confidence=74.58, lift=1.17) related to e-commerce logistics is explored in this investigation. Then, the association rule of "delivery & customer" is used to extract vital e-commerce logistics cases from original unstructured documents. E-commerce logistics cases from the original unstructured documents of the association rule reveal that cross-border e-commerce logistics should consider transnational agreements such as the ASEAN Economic Community agreement (Burnson, 2016b). Moreover, cultural differences such as laws (Ecommerce News, 2014a) and languages (Ecommerce News, 2014b) should be considered for cross-border e-commerce logistics. Furthermore, a fashion accessory brand Evita Peroni adopts the NetSuite SuiteCommerce platform with multi-currency, local tax compliance, and logistics functions to manage global e-commerce business (PR Newswire, 2012). Accordingly, the association rule of "delivery & customer" can be named as "cross-border e-commerce logistics".

4. E-commerce logistics strategies

RDT and IDT theories are incorporated into the empirical results of a hybrid content analytics to create e-commerce logistics strategies for those who are interested in managing e-commerce logistics.

- *Establish inter-organizational and technical collaboration to create positive operations performance*

  The resource dependence theory created by Pfeffer and Salancik (1978) argues that firms' leverage of external resources can strengthen their competitiveness. On the basis of the resource dependence theory, inter-organizational teams have a positive impact on team effectiveness,

especially in cases of high uncertainty (Stock, 2006); collaboration has a positive impact on the performance of firm suppliers (Cai and Yang, 2008); the survival of cross-border alliance can be enhanced by mutual trade dependence (Xia, 2011); information sharing and logistics performance have positive impact on interdependence quality (Chu and Wang, 2012); closer relationships of supply chain partners induce a better impact of supply chain partner innovativeness on innovation strategy (Oke et al., 2013); firms can utilize resources from collaborators to successfully achieve sustainable supply chain management (Esfahbodi et al., 2016); and resource dependence is a key to preserving a long-term relationship (Wang et al., 2016).

The RDT theory is incorporated into the analytical results of topic mining (Topics 1, 2, 3, 5, 6, 7, 9, and 10) to derive the strategy that e-commerce logistics firms could set up inter-organizational and technical cooperation to engender positive operational performance. For example, one of the original unstructured data reveals that European logistics providers cooperate with Europe's Galileo global satellite navigation system to deliver ads and promotions, and to identify location proximity so that logistics providers can locate customers (Nelson, 2013).

- *Comprehend law, policy, and cultural differences to customize appropriate technologies of e-commerce logistics*

Innovation diffusion theory (IDT) has been utilized in technology adoption to explore why a new technology successfully diffuses (Hsu et al., 2014; Jongchul and Sung-Joon, 2014; Rogers, 1995). The five characteristics of IDT

that influence the diffusion of an innovation technology involve relative advantage, compatibility, complexity, trialability, and observability (Hsu et al., 2007; Kim and Ammeter, 2014; Venkatesh et al., 2003; Wu and Wang, 2005); these traits can be utilized to help customers accept innovative technologies through integrating them into the national culture (Choi et al., 2014) or political dynamics (Fedorowicz and Gogan, 2010).

The IDT is incorporated into the analytical results of topic mining (Topics 1, 5, 7, and 8) and the association rule of cross-border e-commerce logistics. Specifically, in order to motivate customers to adopt innovative e-commerce logistics technologies, e-commerce logistics technologies should be tailored to meet local law, policy and cultural differences. Moreover, e-commerce firms need to consider tax issues and preferred payment methods to maximize after-tax profit. For instance, one of the original unstructured data points out that German consumers prefer paying by credit card, debit card and PayPal; French customers prefer paying by Carte Bancaire or Carte Bleue, Dutch customers prefer paying by iDEAL online wire transfer (Linke, 2011). Hence, comprehending law, policy and cultural differences to customize appropriate technologies of e-commerce logistics could raise the probability of innovative technology acceptance and successful logistics investment.

5. Conclusions

Many firms have little knowledge of effective management of e-commerce

logistics. This investigation gives academics and managers a holistic view of e-commerce logistics through unstructured big data analytics. Furthermore, e-commerce logistics companies can employ the proposed hybrid content analytics to identify e-commerce logistics issues, and thereby develop vital e-commerce logistics strategies.

The analytical results of the unstructured data analytics of e-commerce logistics discover essential implications for practice and research. First, the empirical results of topic mining reveals critical themes of e-commerce logistics including "Southeast Asia's e-commerce logistics payments", "E-commerce order management", "E-commerce logistics cloud services", "E-commerce logistics package management", "Europe' e-commerce trends", "India's e-commerce logistics", "E-commerce distribution management", "Tax policies", "E-commerce logistics platforms," and "E-commerce logistics networks". Moreover, an essential rule is generated from the association rule model, namely "Cross-border e-commerce logistics". Furthermore, practical implications represent the following paramount strategies of e-commerce logistics: "Establish inter-organizational and technical collaboration to create positive operations performance" and "Comprehend law, policy, and cultural differences to customize appropriate technologies of e-commerce logistics." More importantly,

the proposed hybrid content analytical model can be applied and extended for academia and industry to explore e-commerce logistics issues.

This work differs from existing investigations by discussing e-commerce logistics in various ways. First, this study systematically absorbs the valuable insights of e-commerce logistics through unstructured text data. Second, this study establishes a hybrid unstructured data analytics involving web crawler, syntactic analysis, data preprocessing, topic mining, and association rules to convert unstructured data into valuable know-how. Companies related to e-commerce logistics can utilize the proposed hybrid analytics to discover e-commerce logistics trends and devise e-commerce logistics strategies. Finally, this study incorporates the RDT and IDT theories with the analytical results of the proposed analytics to create e-commerce logistics strategies and to augment the literature of the RDT and IDT theories.

The proposed model may stimulate future studies in e-commerce logistics management as well as help to address issues regarding the applications of unstructured big data analytics to e-commerce logistics. Future studies may investigate different types of unstructured big data, such as videos related to e-commerce logistics, to conduct large-scale unstructured data analytics. Furthermore, future studies may cooperate with e-commerce logistics companies

to obtain their structured e-commerce logistics data, and to holistically analyze structured and unstructured data of e-commerce logistics.

Acknowledgments

This work is partially supported by Ministry of Science and Technology, Taiwan, R.O.C under grant MOST 106-2410-H-035-013-MY2.

Reference

Auramo, J., Kauremaa, J., Tanskanen, K., 2005. Benefits of IT in supply chain management: an explorative study of progressive companies. Int. J. Phys. Distrib. Logist. Manag. 35, 82–100.

Bhattacharjya, J., Ellison, A., Tripathi, S., 2016. An exploration of logistics-related customer service provision on Twitter The case of e-retailers. Int. J. Phys. Distrib. Logist. Manag. 46, 659–680.

Burnson, P., 2016a. E-Commerce Reshaping the Logistics Landscape. Supply Chain 247. URL http://www.supplychain247.com/article/ecommerce_reshaping_the_logistics_landscape (accessed 11.15.16).

Burnson, P., 2016b. ASEAN is Refining Supply Chain Networks. Supply Chain

Manag. Rev. URL http://www.scmr.com/article/asean_is_refining_supply_chain_networks (accessed 8.16.17).

Burnson, P., 2012. New Cloud-Based Platform Introduced by UPS for International Shipments. Supply Chain Manag. Rev. URL http://www.scmr.com/article/new_cloud-based_platform_introduced_by_ups_for_international_shipments (accessed 11.15.16).

Cai, S., Yang, Z., 2008. Development of Cooperative Norms in the Buyer-Supplier Relationship: The Chinese Experience. J. Supply Chain Manag. 44, 55–70.

Campisi, N., 2016. IBM Meets VMware and Apple Swift in the Cloud. E-Commer. Times. URL http://www.ecommercetimes.com/story/83147.html (accessed 11.15.16).

Casagrande, E., Woldeamlak, S., Woon, W.L., Zeineldin, H.H., Svetinovic, D., 2014. NLP-KAOS for Systems Goal Elicitation: Smart Metering System Case Study. Ieee Trans. Softw. Eng. 40, 941–956.

Cho, J.J.-K., Ozment, J., Sink, H., 2008. Logistics capability, logistics

outsourcing and firm performance in an e-commerce market. Int. J. Phys. Distrib. Logist. Manag. 38, 336–359.

Choi, J., Lee, H.J., Sajjad, F., Lee, H., 2014. The influence of national culture on the attitude towards mobile recommender systems. Technol. Forecast. Soc. Change 86, 65–79.

Chong, A.Y.-L., Ooi, K.-B., Sohal, A., 2009. The relationship between supply chain factors and adoption of e-Collaboration tools: An empirical examination. Int. J. Prod. Econ. 122, 150–160.

Chu, Z., Wang, Q., 2012. Drivers of Relationship Quality in Logistics Outsourcing in China. J. Supply Chain Manag. 48, 78–96.

Cullen, A.J., Taylor, M., 2009. Critical success factors for B2B e-commerce use within the UK NHS pharmaceutical supply chain. Int. J. Oper. Prod. Manag. 29, 1156–1185.

Delfmann, W., Albers, S., Gehring, M., 2002. The impact of electronic commerce on logistics service providers. Int. J. Phys. Distrib. Logist. Manag. 32, 203–222.

Durmusoglu, A., 2016. A pre-assessment of past research on the topic of

environmental-friendly electronics. J. Clean. Prod. 129, 305–314.

Ecommerce News, 2015. Alternative delivery options increasingly popular in Europe. Ecommerce News. URL https://ecommercenews.eu/alternative-delivery-options-increasingly-popular-in-europe/ (accessed 8.16.17).

Ecommerce News, 2014a. 8 tips for cross-border success in Europe. Ecommerce News. URL https://ecommercenews.eu/8-tips-for-cross-border-success-in-europe/ (accessed 8.16.17).

Ecommerce News, 2014b. IMRG launches guide to support cross-border e-retailers. Ecommerce News. URL https://ecommercenews.eu/imrg-launches-guide-to-support-cross-border-e-retailers/ (accessed 11.15.16).

Esfahbodi, A., Zhang, Y., Watson, G., 2016. Sustainable supply chain management in emerging economies: Trade-offs between environmental and cost performance. Int. J. Prod. Econ. 181, 350–366.

Farseev, A., 2014. Social media computing- Lecture 2: Text Processing. Lect. 2

Text Process. URL http://farseev.azurewebsites.net/slides/Data%20Representation%20-%20Text.pdf (accessed 11.15.16).

Fedorowicz, J., Gogan, J.L., 2010. Reinvention of interorganizational systems: A case analysis of the diffusion of a bio-terror surveillance system. Inf. Syst. Front. 12, 81–95.

Gimenez, C., Ventura, E., 2005. Logistics-production, logistics-marketing and external integration - Their impact on performance. Int. J. Oper. Prod. Manag. 25, 20–38.

Gunjan, 2013. Stop Words for Social Media Analytics. SmartData Collect. URL https://www.smartdatacollective.com/social-media-analytics-stop-words/ (accessed 11.15.16).

Hinson, R., 2010. The value chain and e-business in exporting: Case studies from Ghana's non-traditional export (NTE) sector. Telemat. Inform. 27, 323–340.

Hsiao, Y.-H., Chen, M.-C., Liao, W.-C., 2017. Logistics service design for cross-border E-commerce using Kansei engineering with

text-mining-based online content analysis. Telemat. Inform. 34, 284–302.

Hsu, C.-L., Lu, H.-P., Hsu, H.-H., 2007. Adoption of the mobile Internet: An empirical study of multimedia message service (MMS). Omega-Int. J. Manag. Sci. 35, 715–726.

Hsu, P.-F., Ray, S., Li-Hsieh, Y.-Y., 2014. Examining cloud computing adoption intention, pricing mechanism, and deployment model. Int. J. Inf. Manag. 34, 474–488.

Hu, H., Wen, Y., Chua, T.S., Li, X., 2014. Toward scalable systems for big data analytics: A technology tutorial. IEEE Access 2, 652–687.

Hu, H., Wen, Y., Gao, Y., Chua, T.S., Li, X., 2015. Toward an SDN-enabled big data platform for social TV analytics. IEEE Netw. 29, 43–49.

Iyer, K.N.S., Germain, R., Frankwick, G.L., 2004. Supply chain B2B e□commerce and time□based delivery performance. Int. J. Phys. Distrib. Logist. Manag. 34, 645–661.

Johnson, P.F., Klassen, R.D., Leenders, M.R., Awaysheh, A., 2007. Utilizing e-business technologies in supply chains: The impact of firm characteristics and teams. J. Oper. Manag. 25, 1255–1274.

Jongchul, O., Sung-Joon, Y., 2014. Validation of Haptic Enabling Technology Acceptance Model (HE-TAM): Integration of IDT and TAM. Telemat. Inform. 31, 585–596.

Kache, F., Seuring, S., 2017. Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management. Int. J. Oper. Prod. Manag. 37, 10–36.

Kim, D., Ammeter, T., 2014. Predicting personal information system adoption using an integrated diffusion model. Inf. Manage. 51, 451–464.

Linke, D., 2011. Cultivating E-Commerce Trust Across the Globe. E-Commer. Times. URL http://www.ecommercetimes.com/story/73892.html (accessed 8.16.17).

Lu, Q., Liu, N., 2015. Effects of e-commerce channel entry in a two-echelon supply chain: A comparative analysis of single- and dual-channel distribution systems. Int. J. Prod. Econ. 165, 100–111.

Lukman, E., 2013. Onigi Integrates with ESL Express Ahead of Launching its Own Logistics Platform. Tech Asia. URL https://www.techinasia.com/onigi-esl-express-jasakirim (accessed

11.15.16).

Mangiaracina, R., Marchet, G., Perotti, S., Tumino, A., 2015. A review of the environmental implications of B2C e-commerce: a logistics perspective. Int. J. Phys. Distrib. Logist. Manag. 45, 565–591.

Matthias, O., Fouweather, I., Gregory, I., Vernon, A., 2017. Making sense of Big Data - can it transform operations management? Int. J. Oper. Prod. Manag. 37, 37–55.

Millward, S., 2015. Alibaba tames a growing rival with $4.6B investment in Suning. Tech Asia. URL https://www.techinasia.com/alibaba-suning-deal (accessed 8.16.17).

Nelson, P., 2013. A Satellite's-Eye View on Mobile Payments. E-Commer. Times. URL http://www.ecommercetimes.com/story/78906.html (accessed 8.16.17).

Nguyen, H.-O., 2013. Critical factors in e-business adoption: Evidence from Australian transport and logistics companies. Int. J. Prod. Econ. 146, 300–312.

Oke, A., Prajogo, D.I., Jayaram, J., 2013. Strengthening the Innovation Chain:

The Role of Internal Innovation Climate and Strategic Relationships with Supply Chain Partners. J. Supply Chain Manag. 49, 43–58.

Park, S., Lee, D., 2017. An empirical study on consumer online shopping channel choice behavior in omni-channel environment. Telemat. Inform.

Pfeffer, J., Salancik, G.R., 1978. The External Control of Organizations: A Resource Dependence Perspective. Harper & Row, New York, NY.

Piramuthu, S., 2005. Knowledge-based framework for automated dynamic supply chain configuration - Production, manufacturing and logistics. Eur. J. Oper. Res. 165, 219–230.

PR Newswire, 2012. EVITA PERONI Bands With NetSuite SuiteCommerce To Accelerate International Expansion. DNA. URL http://www.dnaindia.com/press-releases/press-release-evita-peroni-bands-with-netsuite-suitecommerce-to-accelerate-international-expansion-1753260 (accessed 8.16.17).

Rao, M., 2016. How a startup cracked the formula to bring Uber-like efficiency to India's trucks. Tech Asia. URL https://www.techinasia.com/blackbuck-marketplace-for-freight-truck-ow

ners (accessed 11.15.16).

Rogers, E.M., 1995. Diffusion of innovations. Free Press, New York.

Roggio, A., 2013. Amazon Prime, ShopRunner Forcing Faster Free Shipping. Pract. Ecommerce. URL http://www.practicalecommerce.com/Amazon-Prime-ShopRunner-Forcing-Faster-Free-Shipping (accessed 8.16.17).

Salvatori, J., 2014. 7 ways to lose less money on shipping. Pract. Ecommerce. URL http://www.practicalecommerce.com/7-ways-to-lose-less-money-on-shipping (accessed 11.15.16).

Saprikis, V., 2013. Suppliers' behavior on the post-adoption stage of business-to-business e-reverse auctions: An empirical study. Telemat. Inform. 30, 132–143.

SAS Institute Inc., 2016. SAS® Text Miner 14.2: Reference Help. SAS Institute Inc., Cary, NC, USA.

Shaikh, A.A., Karjaluoto, H., 2015. Mobile banking adoption: A literature review. Telemat. Inform. 32, 129–142.

Sharp, J.A., 2007. Evaluating the effect of sourcing decisions on e-commerce process operations risk. Eur. J. Oper. Res. 180, 815–833.

Singh, A., Shukla, N., Mishra, N., 2017. Social media data analytics to improve supply chain management in food industries. Transp. Res. Part E Logist. Transp. Rev.

Stock, R.M., 2006. Interorganizational teams as boundary spanners between supplier and customer companies. J. Acad. Mark. Sci. 34, 588–599.

Tan, K.H., Zhan, Y., Ji, G., Ye, F., Chang, C., 2015. Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph. Int. J. Prod. Econ. 165, 223–233.

Ting, S.L., Tse, Y.K., Ho, G.T.S., Chung, S.H., Pang, G., 2014. Mining logistics data to assure the quality in a sustainable food supply chain: A case in the red wine industry. Int. J. Prod. Econ. 152, 200–209.

Toktas-Palut, P., Baylav, E., Teoman, S., Altunbey, M., 2014. The impact of barriers and benefits of e-procurement on its adoption decision: An empirical analysis. Int. J. Prod. Econ. 158, 77–90.

Vaidyanathan, G., Devaraj, S., D'Arcy, J., 2012. Does Security Impact

E-procurement Performance? Testing a Model of Direct and Moderated Effects. Decis. Sci. 43, 437–458.

Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D., 2003. User acceptance of information technology: Toward a unified view. MIS Q. 27, 425–478.

Wang, J., Yue, H., 2017. Food safety pre-warning system based on data mining for a sustainable food supply chain. Food Control 73, 223–229.

Wang, Y., Wang, N., Jiang, L., Yang, Z., Cui, V., 2016. Managing relationships with power advantage buyers: The role of supplier initiated bonding tactics in long-term buyer-supplier collaborations. J. Bus. Res. 69, 5587–5596.

Wong, C.K., 2014. Cross-border e-commerce in Singapore and Malaysia. Ec Insid. URL http://www.ecinsider.my/2014/02/cross-border-ecommerce-in-singapore-malaysia.html (accessed 11.15.16).

Wu, J.H., Wang, S.C., 2005. What drives mobile commerce? An empirical evaluation of the revised technology acceptance model. Inf. Manage. 42, 719–729.

Wu, P.-J., Chen, M.-C., Tsau, C.-K., 2017. The data-driven analytics for investigating cargo loss in logistics systems. Int. J. Phys. Distrib. Logist. Manag. 47, 68–83.

Xia, J., 2011. Mutual Dependence, Partner Substitutability, and Repeated Partnership: The Survival of Cross-Border Alliances. Strateg. Manag. J. 32, 229–253.

Yu, J., Subramanian, N., Ning, K., Edwards, D., 2015. Product delivery service provider selection and customer satisfaction in the era of internet of things: A Chinese e-retailers' perspective. Int. J. Prod. Econ. 159, 104–116.

Zhai, C., Massung, S., 2016. Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining. ACM.

List of Tables and Figures

Table 1 Ten topics with the top five terms and weights of different terms

Topic ID	Document cutoff	Term cutoff	The top five terms and their weights	Number of terms	Number of documents
1	0.134	0.032	Indonesia (0.404), Singapore(0.405), Lazada(0.511), Thailand(0.487), Payment(0.271)	121	214
2	0.153	0.032	Cloud(0.421), Vendor(0.337), Application(0.334), Software(0.579), Procurement(0.596)	136	181
3	0.125	0.033	Customer(0.129), Purchase(0.262), Marketing(0.272), Payment(0.271), Shipment(0.305)	196	289
4	0.149	0.032	Shipment(0.305), Package(0.361), Fedex(0.487), Carrier(0.390), Fulfillment(0.349)	114	260
5	0.144	0.031	Germany(0.421), Uk(0.403), Ecommerce(0.195), Netherlands(0.482), France(0.458)	109	214
6	0.127	0.032	India(0.288), Flipkart(0.454), Snapdeal(0.469), Bangalore(0.473), Mumbai(0.475)	105	239
7	0.084	0.033	Analyst(0.333), Drone(0.558), Grocery(0.385), Vehicle(0.375), Apps(0.370)	218	272
8	0.096	0.033	Government(0.320), Tax(0.496), Payment(0.271), Policy(0.342), Law(0.419)	176	270
9	0.127	0.033	Supplier(0.415), Channel(0.311), Retailer(0.196), Inventory(0.322), Manufacturer(0.351)	173	281
10	0.113	0.032	Alibaba(0.400), Ecommerce(0.195), Tmall(0.514), Taobao(0.482), Retailer(0.196)	103	182

Note: *Document Cutoff* means the threshold that a document should meet to be allocated to the topic; *term cutoff* means the threshold that a term should meet to be allocated to the topic; *number of terms* means the total number of terms in a topic; *number of documents* means the total number of documents involving the topic.

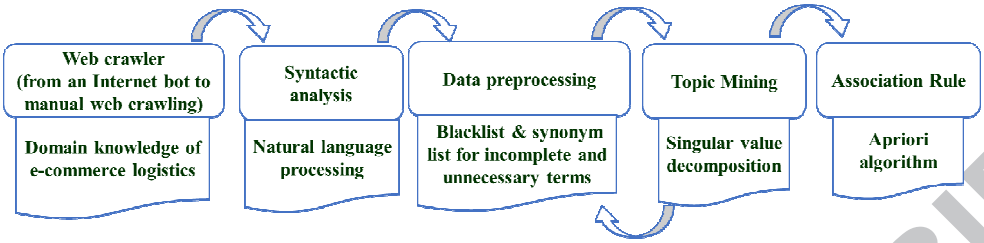

Fig. 1. Proposed hybrid content analytics

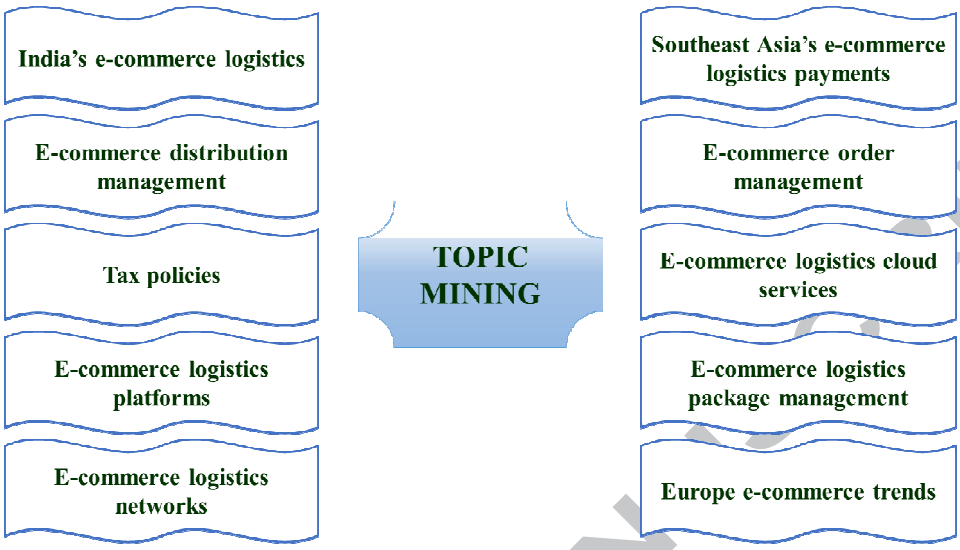

Fig. 2. The empirical results of topic mining

Highlights

- A hybrid unstructured big data analytics is proposed.
- Ten critical themes of e-commerce logistics are generated from topic mining.
- Two vital strategies of e-commerce logistics are created by RDT and IDT theories.