

# Understanding Natural Disasters as Risks in Supply Chain Management through Web Data Analysis

Jimmy Boon Som Ong<sup>1</sup>, Zhaoxia Wang<sup>1\*</sup>, Rick Siow Mong Goh<sup>1</sup>, Xiao Feng Yin<sup>1</sup>, Xin Xin<sup>1</sup>, Xiuju Fu<sup>1</sup>

<sup>1</sup> Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A\*STAR), Fusionopolis, 1 Fusionopolis Way, #16-16, Connexis, Singapore 138632.

\* Corresponding author. Tel.: +65 6419 1520; email: wangz@ihpc.a-star.edu.sg

Manuscript submitted July 10, 2014; accepted October 5, 2014.

doi: 10.17706/ijcce.2015.v4.391

---

**Abstract:** With the increasing trend of global outsourcing, companies are now facing ever more complex supply chains. When a company operates over a large geographical area, the likelihood of disruptions is potentially increased due to such unforeseen events as natural disasters, union strikes or social unrest. In this paper, we consider natural disasters as a form of risks in supply chains and propose to aid its management by analyzing Web data collected in real-time. Using Twitter "tweets" as our primary source of Web data, a real-time data crawler is developed to collect and analyze tweets that are identified as relevant to natural disasters. In addition, a visualization platform customizable to users' requirements is implemented as part of the decision making dashboard for supply chain risk management. The applicability of such a system and its effectiveness for making informed decisions in risk mitigations are then discussed via a case study.

**Key words:** Natural disaster, earthquake, supply chain risk management, web data, real-time analysis, visualization.

---

## 1. Introduction

The increasing demand for mass customization in many industries has made today's supply chain network more complex in a non-linear interconnected and interdependent global setting [1]. When a company operates over a large geographical area, it exposes itself to the risk of business disruptions due to unforeseen events such as natural disasters [2]. The Japanese earthquake of March 2011 derailed the electronics manufacturing sector and automotive industries due to shortage of electronic supplies. Another example is the Thailand flood. In July 2011 to January 2012, Thailand encountered the worst flooding in five decades, which hit at least five major industrial parks, forcing such companies as Toshiba, Honda Motor Co., Toyota and many others to shut down hundreds of factories. This affected companies operating in other parts of the world belonging to the same supply chain network. Clearly, the links of a supply chain can easily be broken in various ways due to these unforeseen natural incidents. In July, 2013, South Africa earthquake shakes Barberton and Nelspruit, and in 2014 the recent flooding in Cape Town has affected more than 8,000 people [3]. Understanding the effects of natural disasters as risks and employing proper risk mitigation techniques is vital for companies in the management of their global supply chains [4].

Social media today also plays an increasingly important role for broadcasting real-time natural disasters information in the community [5]. Analysis of such media has been leveraged to capture the public's attention of the risks associated with various hazards. Twitter tweets, as a typical social media data, have

been used to detect the epicenter and trajectory of the earthquake locations and the results indicated that the constructed detection model could identify 96% of earthquake events at a scale of 3 and above [6]. With the growing number of social networking websites, social media data such as tweets or audios/videos in YouTube on a particular issue of interest are constantly being accumulated. Many organizations also leverage on social networking platforms to publish event information in a timely manner. Social media analysis is a hot topic and it has been applied in many other areas such as the management of emergent events, marketing strategy planning, and customer service [7]. This has led to the growth of increasingly broad applications of social media data, motivating researchers to explore their applications further in supply chain risk management. This paper detailed how social media data is collected and analyzed to get intrinsic information of natural disasters that are considered as risks posed to supply chain networks.

The applicability of the proposed method is discussed using a case study which includes earthquake monitoring and natural disaster mapping for supply chain management via a platform named "Map Your Risk". In this case study, tweets are collected and analyzed to identify and locate an impending earthquake. The platform's monitoring system will then alert the supply chain managers of the potential risks posed to different supply chain facilities. In addition, earthquake affected areas are further analyzed and displayed on the platform to visually reflect the detailed risks information that are overlaid onto their supply chain network topologies. Furthermore, the network topologies can be transformed to map the exact locations of each supply chain facility on the world map. This paper demonstrates that understanding natural disasters as risks through Web data analysis is not only viable, but also highly valuable and useful to augment the effectiveness of a supply chain risk management policy.

## **2. Web Data Analysis**

Web data takes on many different forms including texts (e.g. Internet forums, weblogs, social and micro blogs, wikis, etc.), images (e.g. scanned documents, graphs, charts, photographs, pictures, etc.), and audio/video (e.g. podcasts) [8]. Social network aggregation can integrate many of the platforms apart from the traditional forms of media data. Web data is the result of information explosion on the Internet, which includes the information disseminated by organizations and Web user generated content [9]. Text however, is the main form of Web data that reflects users' feedbacks, attitudes and opinions on some issues such as discussions, vents, services and policies. Text data analysis thus is becoming a fast and effective way of evaluating public opinion or inherent information value discovery in the social studies.

Text data analysis can be considered as a main objective of text mining and classification and can be generally divided into two categories based on either machine learning or using a lexicon. Machine learning classification methods such as naïve Bayes [10], [11], maximum entropy [12] and support vector machines [13], [14] have been widely used in classifying text data. However, their heavy reliance on training datasets makes this approach prohibitively difficult as datasets are not always readily available. On the other hand, the lexicon-based analysis approach is easily applicable to different datasets and the methods and algorithms are available in many commercial (e.g. IBM LanguageWare, SPSS, Insight, SAS, etc.) and open source software (e.g. Carrot2, GATE, OpenNLP, WEKA, etc.).

## **3. System Architecture**

Besides the text data analysis techniques described in Section II, visualization techniques provide powerful means of making sense of data and there are various implementations to cater to the needs of different real life applications [15]. In this work, we applied text data analysis methods in addition to several visualization techniques to analyze and visualize the Web data. Having the analytical methods and visualizations incorporated into the supply chain management platform, risk managers can make informed decisions and

select suitable strategies based on analytical reports and observations provided. This in turn, is likely to help them further explore or investigate which of the disasters or events are likely to constitute risks and cause potential disruptions to their supply chains in specific areas.

As illustrated in Fig. 1, the browser-server paradigm is selected to design and implement the system. This mode allows easy system maintenance since the client-side user interface on the Web browsers is realized through the server-side application logic [15]. The system architecture comprises of three core modules: Data collection module, Data analysis module and Web service.

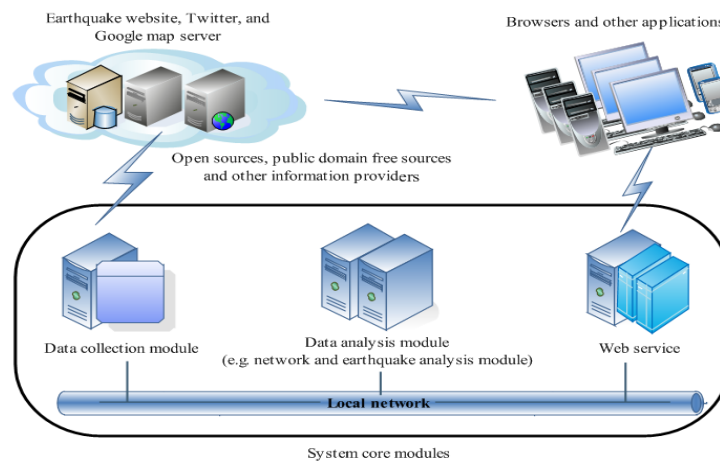


Fig. 1. System architecture overview.

### 3.1. Data collection module

Web data can be downloaded directly from the Web or collected either using various application programming interfaces (APIs) or via third party Web crawlers. We make use of two data sources collected from: open source/public domain directly; 2) information providers using an API.

The first data source in this paper is an earthquake-related website [16] from which the earthquake data is downloaded. The second source is Twitter. Using the provided API, tweets can be filtered and collected in real-time, either individually or collectively, based on a specified language and/or geographical region by using keywords (e.g. "earthquake").

To prepare the Web data collected for analysis, this module includes several pre-processes to clean the data so that irrelevant information is excluded and reformat the data appropriately for the Data analysis module.

### 3.2. Data analysis module

This module performs analysis on the data pre-processed by the Data collection module. Such analysis results as the impact of earthquakes on airports and transportations will be then pushed to the browsers or other applications via a Web service.

### 3.3. Web service

A Web service interfaces between the backend system and the Web browsers or other applications. It allows two software systems to exchange data over the internet. The software system that requests data is called a service requester, whereas the software system that would process the request is called a service provider. For example, the risks status on specific locations can be enquired through a service requester and then the server providers will process the request to visualize the information using Google Maps Web services. This will be illustrated as a case study in the section IV.

The system follows a traditional model-view-controller pattern and the implementation stack consists of

several such technologies as Spark [17] for the main Web framework, Hibernate [18] for object-relational mapping between the application logic and persistence storage, and ExtJS [19] for the user interface on Web browsers. Fig. 2 outlines a modular hierarchical system architecture that is designed to allow scalability, flexibility and easy maintenance [15]. Each main module includes multiple sub-modules. This design is important for the implementation and development of the system as it allows the separation of concern between modules. A module is "pluggable" and can be attached or detached without affecting other modules. For example, a new module can be added to the Data collection module to crawl new sources of data, or the Data analysis module can be enhanced with a more powerful sub-data analysis engine.

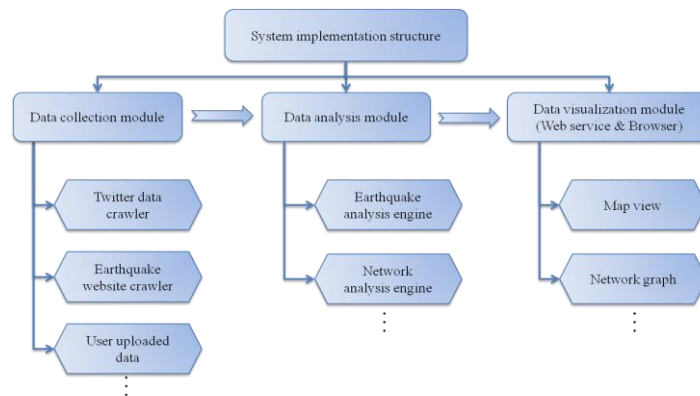


Fig. 2. Hierarchical system architecture modules.

## 4. Case Study

In this case study, earthquakes will be used as an example of natural disasters in the management of risks in supply chain networks.

### 4.1. Visualizing and Monitoring Real-time Earthquake Information

Tweets related to earthquakes are collected in real-time and monitored through the platform. As shown in Fig. 3, earthquakes are visualized using their magnitude and geodata with other relevant information. The yellow translucent circle is an indication of the number of tweets collected around the earthquake region and the size is proportionally related to the damage of the earthquake to some extent [2]. The next subsection will detail the way such information can be further mapped to a supply chain network.



Fig. 3. Visualizing and monitoring earthquakes information in real-time.

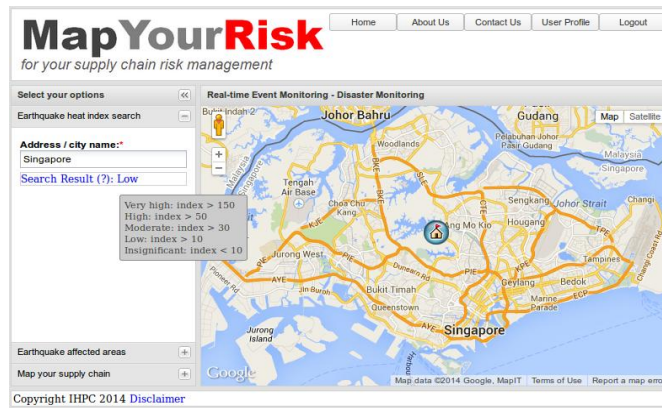


Fig. 4. Map Your Risk: a visualization and analysis platform.

## 4.2. Geo-mapping earthquakes into supply chain networks

Earthquakes are further analyzed through a visualization and analysis platform named “Map Your Risk” (MYR). MYR uses a border layout to arrange the site content. As pictured in Fig. 4 below, the site’s navigation is on the west region, while the main content is in the center. The menu system follows a categorical system with each category having a function or submenu. At this stage, MYR has 3 main functions:

- 1) Earthquake heat index search
- 2) Earthquake affected areas display
- 3) Supply chain networks mapping

Function item 1, as the name suggests, allows the user to search an address or city for its heat index. This heat index score gives an idea of the potential risks involved with that address or city.

Function item 2 (Earthquake affected area) captures earthquake occurrences in real-time using the corresponding geodata, the locations can be demarcated on the map. Fig. 5 illustrates an earthquake location near England and the affected corresponding facilities (in this example we are showing airports) around this earthquake. Potential risks of such facilities are accessed by calculating impacts from the magnitude of the earthquake. Alerts of potential risks, if any, are then raised and sent to facility owners.



Fig. 5. Geo-mapping earthquake locations.

The third function item (Map your supply chain network) has two submenu functions: 1) Supply chain network view; 2) Supply chain map view. Submenu function 1 allows the user to upload his/her supply chain using a pre-defined format. Once the network is uploaded, the network can be visualized in terms of vertices and edges using a force-directed layout as shown in Fig. 6. With the geodata associated with each vertex, facilities in the network can also be viewed alternatively on the map as depicted in Fig. 7. The visualized content in the platform may act as the “digital traces” of risk information to help supply chain managers to visualize and

understand risk levels associated with each supply chain facility.



Fig. 6. Visualizing supply chain network topology.

## 5. Conclusion

Using earthquakes as an example, this work investigates natural disasters through Web data analysis for supply chain risk management. Our approach will be valuable for supply chain risk management researchers in enabling them to monitor events and track the Web data in various settings. Understanding natural disasters as risks in supply chain management is a good utilization of Web data. Our research offers new ideas for organizations and enterprises who wish to probe deeper to leverage Web data in supply chain management as well as issues in other domains. Fully utilizing the available information and adopting appropriate technologies timely to address pressing problems is a priority in this data deluge era, as sufficient and accurate information is necessary for managers to make effective decisions on behalf of their organizations. We look forward to continuing investigation in this area, with the hope that improved intelligent system approaches for Web data collection, analysis, visualization and monitoring will support smarter decisions and enable better risk management in supply chains.

## Acknowledgment

The authors would like to thank Dr. Loganathan PONNAMBALAM, Dr. Haiyan XU and Dr. Jesus Felix Bayta VALENZUELA from IHPC for their valuable discussions, comments and invaluable help in conducting this study. Many thanks to Mr. Sifei LU for the data collection and database support.

## References

- [1] Tse, Y. K., & Tan, K. H. (2012). Managing product quality risk and visibility in multi-layer supply chain. *International Journal of Production Economics*, 139(1), 49-57.
- [2] Goh, R. S. M., et al. (2013). RiskVis: Supply chain visualization with risk management and real-time monitoring. *The 9th annual IEEE International Conference on Automation Science and Engineering* (pp. 207-212).
- [3] Recent natural disaster. Retrieved from: <http://www.disaster-report.com/>.
- [4] Wang, Z., Goh, R. S. M., Yin, X., Loganathan, P., Fu, X., & Lu, S. (2013). Understanding the effects of natural disasters as risks in supply chain management: A data analytics and visualization approach. *2nd Annual Workshop on Analytics for Business, Consumer and Social Insights, Invited presentation*.
- [5] Dabner, N. (2012, January). Breaking Ground' in the use of social media: A case study of a university earthquake response to inform educational design with Facebook. *The Internet and Higher Education*, 15(1), 69-7.
- [6] Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection

- by social sensors. *Proceedings of the 19th international conference on World Wide Web, ACM* (pp. 851-860).
- [7] Pentland, A. S. (2012). Social efficiency: rules for designing social networks and social media [Social Sciences]. *IEEE Signal Processing Magazine*, 9(2), 146-147.
- [8] X. Fu *et al.* (Dec. 2013). Social Media for Supply Chain Risk Management. *Proceedings of in International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 10-13). Bangkok, Thailand.
- [9] Ghiassi, M., Skinner, J., & Zimbra, D. (2013, May). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications*, 40(16), 6266-6282.
- [10] Ortigosa-Hernández, J., Rodríguez, J. D., Alzate, L., Lucania, M., Inza, I., & Lozano, J. A. (2012, Sep.). Approaching sentiment analysis by using semi-supervised learning of multi-dimensional classifiers. *Neurocomputing*, 92,98-115.
- [11] Glorot, X., Bordes, A., & Bengio, Y. (2011). Domain adaptation for large-scale sentiment classification: A deep learning approach. *Proceedings of the 28th International Conference on Machine Learning* (pp. 513-520).
- [12] Deng, H. Ji, H., & Han, J. (2012, Sep.). Uncertainty reduction for knowledge discovery and information extraction on the World Wide Web, *Proceedings of the IEEE*, 100(9), 2658-2674.
- [13] Gokaraju, B., Durbha, S. S., King, R. L., Member, S., & Younan, N. H. (2011). A machine learning based spatio-temporal data mining approach for detection of harmful algal blooms in the Gulf of Mexico. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(3), 710-720.
- [14] Wilson, T., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35(3).
- [15] Wang, Z., Chong, C. S., Goh, R. S. M., Zhou, W., Peng, D., & Chin, H. C. (2012). Visualization for anomaly detection and data management by leveraging network, sensor and GIS techniques. *IEEE 18th International Conference on Parallel and Distributed Systems* (pp. 907-912).
- [16] Earthquake data. Retrieved Dec., 2012, from: <http://earthquakestoday.info/>.
- [17] Spark. Retrieved Sep., 2013, from: <http://www.sparkjava.com/>.
- [18] Hibernate. Retrieved Sep., 2013, from: <http://hibernate.org/>.
- [19] ExtJS. Retrieved Sep., 2013, from: <http://www.sencha.com/products/extjs/>.

**Jimmy Boon Som Ong** is a research engineer in the computing science (CS) Department at the Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A\*STAR), Singapore. His research interests include agent-based epidemiological simulation and modeling, cloud/distributed computing and big data application architectures.

**Zhaoxia Wang** is a scientist and programme manager in the Social & Cognitive Computing (SCC) department, at the IHPC, A\*STAR, Singapore. Her research interests include social data analytics, intelligence systems, pattern recognition, risk/fault prediction and analysis, data mining and applications, and urban systems. She was a research staff of NUS from 2008 to 2011. She received her Ph.D. from Nankai university and she has led and successfully delivered more than 10 projects. She is currently leading technical efforts in big data analytics. She belongs to IEEE.

**Rick Siow Mong Goh** is the Director of the Computing Science (CS) department at the A\*STAR IHPC. At IHPC, he leads a team of more than 50 scientists in performing world-leading scientific research, developing

technologies to commercialization, and engaging and collaborating with industry. The research focus areas include high performance computing (HPC), distributed computing, big data analytics, intuitive interaction technologies, and complex systems. His expertise is in discrete event simulation, parallel and distributed computing, and performance optimization and tuning of applications on large-scale computing platforms. Dr. Goh received his Ph.D. in Electrical and Computer Engineering from the National University of Singapore.

**Xiao Feng Yin** received both his Ph.D. in 2011 and his Master's degrees in 2000 in intelligent manufacturing from Nanyang Technological University (NTU), Singapore. His research areas include logistics and supply chain optimization, planning and scheduling, knowledge discovery and artificial intelligence. He is currently working in IHPC Singapore as a research scientist in supply chain risk management and logistics optimization. He was a research staff of NTU from 2010 to 2012, working on projects of knowledge discovery and anomaly detection.

**Xin Xin** is a Research Engineer in the Social & Cognitive Computing (SCC) department, at the IHPC, A\*STAR, Singapore. His research involves social data analytics, real-time data processing and intelligent systems.

**Xiuju Fu** is a Senior Scientist in the Computing Science (CS) department, IHPC, A\*STAR, Singapore. She graduated with a Ph.D degree in Information Systems from Nanyang Technological University (2003) and B.Eng in Signal Processing, from Beijing Institute of Technology (1995), China. Her research interests include supply chain, data analytics, modeling and simulation and risk management.