Computers and Electrical Engineering 68 (2018) 140–155

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Risk assessment for global supplier selection using text mining☆

Chuan-Jun Su*, Yin-An Chen

Department of Industrial Engineering & Management, Yuan Ze University, 135, Far-East Rd., Chung-Li, Taiwan, ROC

ARTICLE INFO

Keywords:
Global supplier selection
Text mining
Social media
Twitter Enabled Supplier Status Assessment (TESSA) tool

ABSTRACT

Adequate global sourcing makes a strategic difference to an organization's ability to reduce cost and improve the quality of its product. However, the global sourcing makes the process of supplier selection riskier and more complex. The data required for confronting the criteria needs to be specific and up-to-date. Now that we have entered the Big Data era with the prominence of social media and social networks such as Twitter, Facebook, LinkedIn; we now have access to global insights and knowledge regarding international suppliers. In this research, we propose a Twitter Enabled Supplier Status Assessment (TESSA) tool that can assist companies in their global supplier selection process. TESSA firstly retrieves a target supplier's related information from the most popular microblog Twitter and then obtains potential risk and uncertainty regarding the supplier through text mining. The discovered risks and uncertainties for companies making better decision on their global supplier selection process.

1. Introduction

Due to global competition modern organizations are paying particular attention to Global Sourcing. This practice has been embraced due to the cost savings it generates, the access to technologies and higher quality products in some cases. Organizations can choose suppliers from anywhere in the world, developing countries are becoming more competitive given their low labor and operating costs. Global supplier selection is riskier than the domestic supplier selection, consequently the decision making process is strongly affected by perceived risks. Suppliers with low price products can be offset by a history of delayed deliveries, or suppliers with state of the art technology can be undermined by excessive tariffs and costs.

Risk can originate from economics and political uncertainty in the supplier's country. Natural disasters are also within the risk factor, any natural disaster can have catastrophic consequences in today's interconnected global supply network. All of these types of disruptions can damage profitability, stock price, and market reputation for the organization with significant long lasting consequences.

In [1] Sawik affirms that taking risk into consideration will allow the buyer to decide whether it should cooperate with a low cost yet risky supplier, over a more expensive but possibly more reliable supplier. There is a crucial need to identify these risk factors and take them into consideration when selecting a supplier. However the supplier selection team struggles to obtain precise, complete and up to date information [2]. Data resources are usually only available at a low frequency of monthly or quarterly levels [3]. The data is sparse through reports, external databases, ERP and MIS systems which are limited and not able to provide sufficient information regarding the risks and uncertainty from the suppliers. The supplier selection process demands more transparency and up to date information.

Given the new Big Data era we are in, gathering data is not a problem anymore and we can utilize global insight and knowledge to

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. J-S Sheu.

* Corresponding author.
E-mail addresses: iecjsu@saturn.yzu.edu.tw (C.-J. Su), s998906@mail.yzu.edu.tw (Y.-A. Chen).

https://doi.org/10.1016/j.compeleceng.2018.03.042
Received 5 September 2017; Received in revised form 26 March 2018; Accepted 26 March 2018

assess the risk and uncertainty for each supplier. In regards of supply chain management some researchers have already used the data in social media to revolutionize their organizations. For instance, social media has played an important role in demand prediction for supply chain management [4,5], according to [6] social media offers insight on preferences and consumer behavior. The information in social media is updated rapidly and spreads virally at an exceptional speed; this provides us with first-hand information. We now have the opportunity to analyze this vast portfolio of information to assist the supplier selection process. In this research, we focus on social media rather than conventional online data due to its ability to be generated and diffused in a quicker manner; we mainly focus on the microblogging tool Twitter. This is because up to date and relevant information is required due to the nature of the risk and uncertainty criteria. Tweets are compact and fast. This is why it has become widely used to spread and share breaking news.

The main objective of this research is to provide a tool called Twitter Enabled Supplier Status Assessment (TESSA) that can assist the procurement team when selecting a supplier, TESSA can provide information on the risk and uncertainty for each supplier helping the decision team to reduce their potential supplier list and make a final decision. As a secondary objective we aim to open a new window on the research field regarding supplier selection and the use of social networks, as a result, companies can be more prone to exploit the extensive data social networks has to offer and utilize it in the decision making process.

The remaining of the paper is organized as follows: Section 2 reviews related literature on global supplier selection, text-classification methodologies, and ontology. Section 3 presents the system architecture and methodology of TESSA. Section 4 describes the implementation with the usage scenarios. Finally, Section 5 provides the contributions and conclusion of the paper.

2. Literature review

2.1. Global supplier selection

Global supplier selection is the process in which the buyer identifies, evaluates, and contracts with providers. Monczka et al. [6] have defined supplier selection as an essential task of procurement, it represents a key role in the company's long term strategy and competitive positioning. Supplier selection has gained great attention in business management literature and practice, mainly due to the growing business environment of global sourcing, strategic buyer-supplier relationship, e-commerce, and so forth.

The method of evaluation will depend on the organization's priorities, capabilities, and strategy. The tool proposed in this research is designed to work as an assistant to provide information on the risk of suppliers regardless of the method the organization is deploying.

2.1.1. Global supplier selection criteria

Selecting a new supplier is not a new problem. A great number of conceptual and empirical research has been published suggesting different techniques and approaches. Relying on a single criterion makes the supplier selection process uncertain. Therefore, a multi-criteria approach is recommended.

Developing countries have become a supplying source for many organizations due to their low labor and operating cost. Price is no longer sufficient and total cost has become an important factor for some organizations [7]. Total Cost of Ownership (TCO) is a philosophy to understand the true cost of a particular product, the TCO criteria attempts to look at a life cycle cost considering all costs associated with acquisition, use, and maintenance [8]. This provides a more effective clarification of supplier performance within the organization, however adopting this approach is very complex, the required account system to capture all the relevant cost of each supplier is a major disadvantage, especially when selecting new suppliers due to the lack of information. Aghai et al. [9] also includes a risk criteria in their model for supplier selection, their risk factors are economic, environmental, and supplier satisfactory ratings. And Vinodh et al. [10] also includes risk criteria within their supplier selection process, they consider the supplier constraint and supplier profile.

2.2. Risk within the supply chain

Over the years the market has become globalized and organizations that once focused on domestic sourcing later sought for suppliers around the world, this makes the supplier selection riskier and more complex. The main objective of global sourcing is to exploit both the supplier's competitive advantages and the comparative location advantages of various countries in global competition. As companies seek for international suppliers they must be aware not only of the opportunities, but of the risk and threats as well. Sreedevi and Saranga have proposed [11] the best way of risk avoidance strategy is to take care of risks when selecting the suppliers.

Organizations need to have all the information they can get in order to make an adequate decision. In this research, we focus on the external risks, which are the ones caused from outside the supply chain, usually related to economic issues, social, political, climate, terrorism, and financial stability. Obtaining data related to external risk has been a challenge, however in our current Big Data era we can now take advantage of the global insight and knowledge.

2.3. Microblogging tool, Twitter

Twitter is a free microblogging service that allows users to communicate with one another using short text based messages, or tweets [12]. Among the different microblogging tools available today, Twitter has become the prevalent platform. It has grown at an unprecedented rate. According to Bennett [13], there are 175 million tweets per day and more than one million new accounts are added every day.

The intensity of Twitter usage varies considerably, the knowledge we can obtain from Twitter is more than just sentiment and opinions. This microblogging platform is becoming the standard domain for event selection and many researchers have already used Twitter as a news source. Sakaki et al. [14] proposed a framework to use Twitter for early detection of earthquakes, they state that Twitter is faster than the detection based on traditional media. Given the speed and coverage of Twitter some researchers have used tweets for event detection.

Within the Supply Chain, Twitter has been mainly used to provide information on transporting goods. O'Leary [15] explains how Twitter has been used in the Supply Chain to provide information about a range of events like the arrival and departure of a shipment, or information about accidents and closed roads; this provides a more easily coordination and mobilization through the supply chain. However, organizations can use Twitter for much more, this microblogging tool can play an important role in the supply chain gathering information from disparate sources and increase available information, and it can also make this information visible in a faster manner.

2.4. Text classification

2.4.1. Natural language toolkit (NLTK)

Natural Language Toolkit (NLTK) is ideal for text classification; it is an easy-to-use interface with a wide variety of lexical resources and text processing libraries for classification, tokenization, stemming, parsing, and semantic reasoning. It has been used for several researchers and has provided excellent results. Yu et al. [16] employs NLTK in a product title classification and mentions the speed is quite remarkable. Kamruzzaman [17] presents an algorithm for text classification using machine learning techniques and NLTK. In this paper, our TESSA system utilizes NLTK to perform text preprocessing of retrieved tweets.

2.4.2. Recurrent neural network (RNN)

A relative new development in text-mining is the use of Recurrent Neural Network (RNN), a branch of Deep Learning to process a sequence of text data while keeping the changes in state for a particular sequence. RNN has been widely applied in the areas of unsegmented, connected handwriting recognition, speech recognition, etc. [18]. Liu et al. presented the RNN recurrent structure that is suitable for processing the variable length text [19].

2.4.3. Word2vec

Word2vec is a group of Deep Learning models developed by Google, which aims at capturing the context of words while provides an effective mechanism for preprocessing raw text data. The model takes input from such large corpus of documents as tweets or news and generates a vector space of high dimension. Each word in the corpus is assigned a unique vector in the vector space [20]. Continuous representations of words as vectors has proven to be an effective technique in many NLP tasks, including sentiment analysis [21].

2.5. Ontology

Ontology formally defines different concepts of a domain and the relationships between these concepts. Ontologies are an important resource to deal with semantic heterogeneity. They are used to achieve the semantic interoperability and retrieval of relevant documents [22]. Ontologies can be used to model a domain with the definition of concepts together with their properties and relations by a shared vocabulary and taxonomy. One of the most challenging tasks in knowledge retrieval from text-based data is the need to examine the data in different contexts and different perspectives. In order to properly classify the text-based tweets within its corresponding risk and uncertainty categories an ontology needs to be developed. Most of traditional approaches lack of taking the semantic relations between words into account. It is out conviction that the ontology-based approach is a better approach.

Harman and Koohang [23] define taxonomy as a way of classifying or categorizing a set of things using a hierarchical structure. Each node, including the root node, is an information entity that represents some object in the real world that is being modeled. Each link between two nodes in a taxonomy represents a “sub classification-of” relation or a “super classification-of” relationship. Ontology defines the terms used to describe and represent an area of knowledge. Ontologies are used by people, databases, and applications that need to share domain knowledge. Ontologies range from simple taxonomies, to metadata schemes, to logical theories [24]. However, with new emerging demands of data rich applications, we have seen many new challenges for machine learning and knowledge discovery. One of these challenges is the need to examine data in different contexts and different perspectives, methods for context-dependent and ontology-aware information extraction from data with user-specified ontologies are needed.

Some researchers have proposed the deployment of ontology-based techniques towards a more fine-grained classification. Iwanaga et al. [25] present a methodology for populating existing earthquake evacuation ontology with instances based on tweets. The proposed approach extracts related information like evacuation center names, products offered at the centers and the timestamp of each tweet. Baldoni et al. [26] designed an application software for associating the predominant emotions to artistic resources of a social tagging platform, they aimed to extract a rich emotional semantics of tagged resources through an ontology driven approach. In these examples, researchers use ontologies to represent characteristics of the categories, so documents are classified in real time not with trained data or learning process.

Traditional Machine Learning algorithms have some drawbacks, one of them is that users must collect a large number of training data and the process is very laborious, the other one is that the traditional approach doesn't consider the semantic relations between

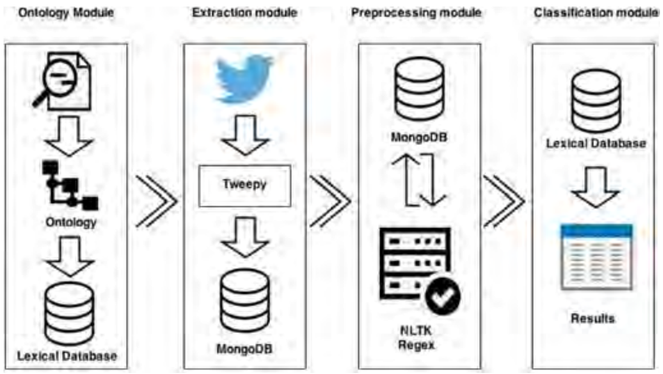


Fig. 1. TESSA Architecture.

words. In order to solve this some researchers have stated that an ontology could be used to organize training sets for some conventional text classification algorithm. Kontopoulos et al. [27] created a domain ontology and then used it to train the classifier for sentiment analysis on Twitter, they obtained positive and accurate results.

3. Methodology

In this research we proposed TESSA (Twitter Enabled Supplier Status Assessment), to work as a tool and assist in the supplier selection process. TESSA's architecture is formed by 4 modules: Ontology Building, Extraction, Preprocessing, and Classifier as shown in Fig. 1.

TESSA's process flow diagram is shown in Fig. 2. TESSA's first module involves building a Risk and Uncertainty Ontology (RUO) which encapsulates risk and uncertainty factors regarding global supplier. A lexical database can then be derived from the ontology, this lexical database encompasses all the aspects and attributes from the ontology and serves as the classifier for the tweets. The Data Extraction process is responsible for retrieving related tweets by using Twitter's streaming Application Programming Interface (API) through a python script crawler. The extraction is done using the supplier name as keyword. These tweets are saved in Mongo Database (Mongo DB). Afterwards in the Preprocess part, a series of processes (Tokenization, Lower case conversion, etc.) are subsequently performed to normalize the retrieved tweets. In the Classification module, once the tweets are preprocessed they are classified using the lexical database.

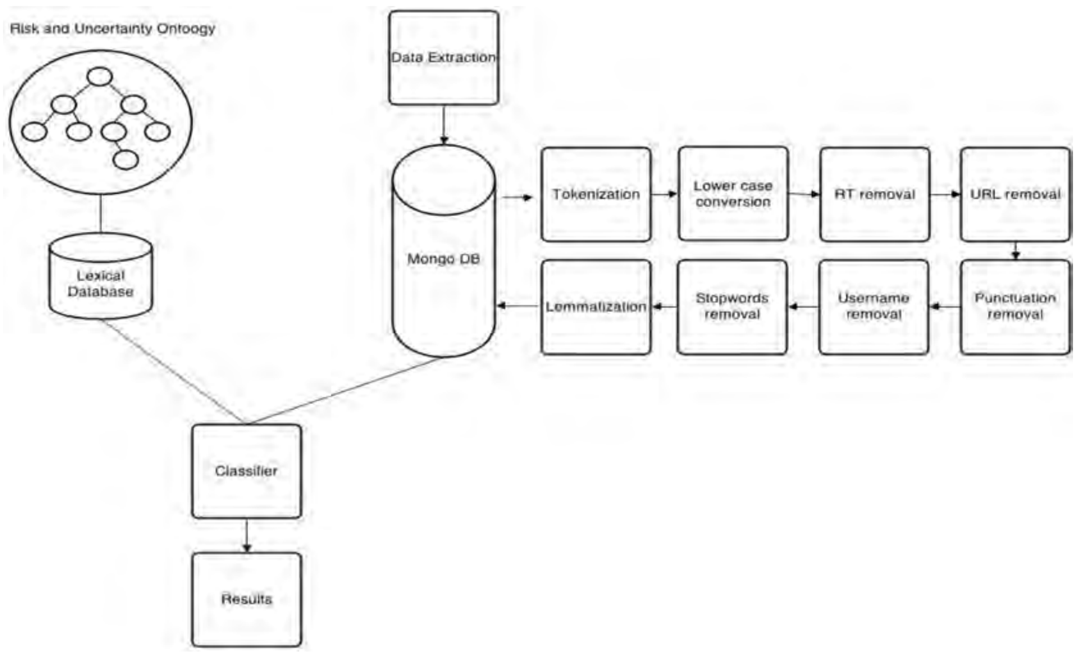


Fig. 2. TESSA's process flow diagram.

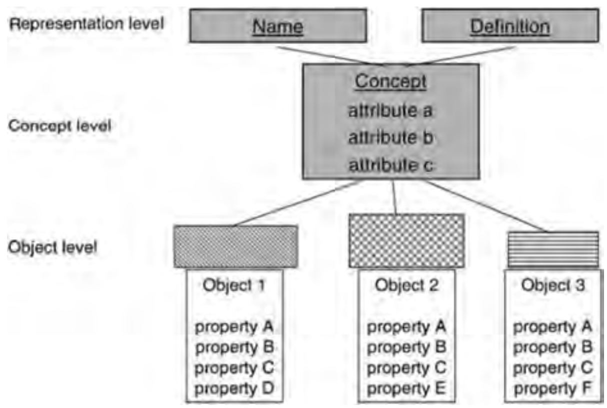


Fig. 3. FCA architecture [28].

3.1. Risk and uncertainty ontology (RUO)

In order to properly classify the tweets within its corresponding risk category an ontology needs to be developed. In this phase RUO is developed based on the Formal Concept Analysis (FCA). FCA explicitly formalizes extension and intension of a concept and their mutual relationship. A strong community of academic, government, and corporate users have utilized tool Protégé when generating an ontology; this tool is used to develop the RUO.

In this research the data set of risk factors that we use is based on various literatures including supplier risk, global risk reports and other ontologies. An example regarding our RUO is shown in Fig. 4. FCA differs from other knowledge representation formalisms (like RDF, description logics, or OWL). It distinguishes three levels: object, concept, and representation levels as depicted in Fig. 3 [28].

3.2. Data extraction

This research relies primarily on data from Twitter, in order to retrieve Tweets we must access Twitter API. We use the Twitter Streaming API in order to retrieve live tweets; this API provides a constant stream of public tweets. Since the information on Twitter is protected we need an Open Authentication (OAuth) to access the data, which involves in setting up Twitter account and registering an application.

A tweet script crawler is built with the programming language Python in order to extract the tweets. In order to connect to the Twitter API we access a Python library called Tweepy, this was chosen due to its quickness and efficiency of response. The tweet script crawler is designed to retrieve the Tweets based on selected keywords, which are the names of the targeted supplier in our case. With the tweet script crawler the tweets can be retrieved and subsequently stored in a NOSQL (Not only SQL- Structured Query Language) database called Mongo DB. In order to allow the communication between Mongo DB and Python, we implement the Python driver called Pymongo into the tweet script crawler.

3.3. Preprocessing

The retrieved tweets have to be cleansed by removing the undesired (noisy) data for further analysis. Two Python modules, the Natural Language Toolkit (NLTK) and the Regular Expression module (Regex) are used for accomplishing the preprocessing. The

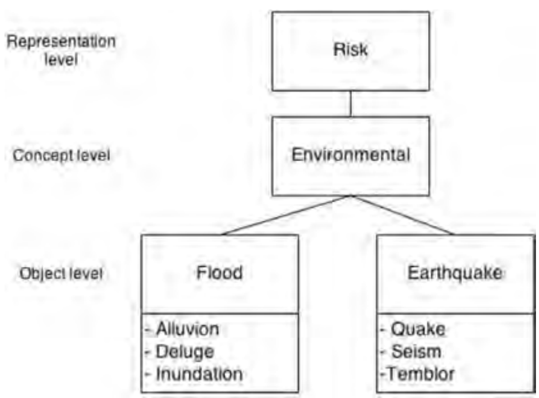


Fig. 4. FCA example.

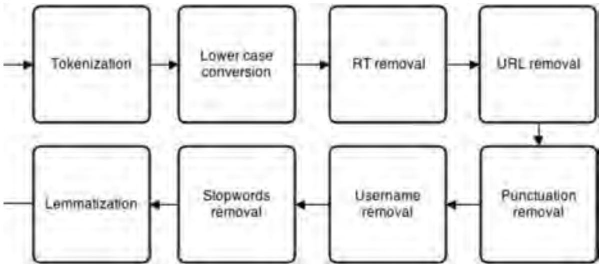


Fig. 5. Preprocess tools.

stages of preprocessing are illustrated in Fig. 5.

3.4. Classification

Once the tweets are preprocessed they go through the lexical database, in the lexical database reside all the aspect and attributes from the ontology. Here all the tweets go through the database and are matched within the aspects and attributes. If there is a match a number 1 is reflected and the tweet is saved in a collection under the name of the aspect, if there is no match a number 0 is shown, the tweets that don't belong to any aspect are classified as trash.

3.5. Evaluation

In Kohavi and Provost [29] explained that:

The exactitude of a classifier can be evaluated by computing the number of correctly recognized class examples, the number of correctly recognized examples that do not belong to that class, and examples that either were incorrectly assigned to the class or that were not recognized as class examples.

These 4 concepts are the entries of the confusion matrix (Table 1). The Confusion matrix is useful to evaluate the performance of a classifier, showing the number per class of well classified and mislabeled instances. It is a simple and understandable way to show the classifier performance. More sophisticated measures of the classifier performance can be calculated from the confusion matrix.

where

tp is the number of correct predictions that an instance is actually similar.
fp is the number of incorrect predictions that an instance is actually similar.
fn is the number of incorrect predictions that an instance is actually dissimilar.
tn is the number of correct predictions that an instance is actually dissimilar.

According to Sokolova and Lapalme [30], the most common evaluation metrics in text classification are accuracy, precision, recall and f-score. Given the confusion matrix we can evaluate TESSA's accuracy, precision, recall and f-score; we will do it by implementing the equations presented in Table 2.

4. Implementation

4.1. Risk and uncertainty ontology creation

The Risk and Uncertainty Ontology (RUO) is developed in order to properly classify the tweets within its corresponding entity, which is a specific risk category. The ontology is built using the FCA methodology, which consists of concept, objects and properties. Following the aspect extraction approach the ontology architecture is designed as follows: the concept is the entity, the objects are the aspects and the properties are the attributes of each aspect. An example is shown in Fig. 6.

In this case our entity is the financial risk, since it's the object we are evaluating. Its aspects are the different financial risk there can be such as bankruptcy and debt, among others. And the attributes of these aspects are insolvency and arrears, among others; in

Table 1
Confusion matrix.

		Predicted	
		Positive	Negative
Actual	Positive	*tp*	*fn*
	Negative	*fp*	*tn*

Table 2
Classification evaluation metrics.

Metric	Formula	Description
Accuracy	$(tp + tn)/(tp + fp + fn + tn)$	Proportion of the total number of predictions that were correct
Precision	$tp/(tp + fp)$	Proportion of the predicted positive cases that were correct
Recall	$tp/(tp + fn)$	Proportion of positive cases that were correctly identified
F-score	$(2 * precision * recall)/(precision + recall)$	Relations between data's positive labels and those given by a classifier

Fig. 6. Ontology design.

this research the attributes are the synonyms of the aspects.

Some researchers have studied different risk and events that create disruptions within the supply chain, other researchers have started to address the risk factors right from the supplier selection process. The literature used is from different industry domains such as automotive, retail and manufacturing. The risk classification varies from researcher to researcher; therefore a review was made to the different literature and some of the categories were merged to build our ontology.

Another source of information for our ontology is the Industry portfolio of risk (as shown in Fig. 7). This portfolio was presented by Elkins in [8]; the portfolio was created from a brainstorming exercise with the aim of assisting manufacturing engineers and supply chain analysts to begin identifying risks that would affect their work.

Based on this literature review we developed the RUO where the entities are: Economic risk, Environmental risk, Financial risk and Sociopolitical risk. A main view of the RUO is shown in Fig. 8.

As shown in Fig. 8 each of the entities (Economic, Sociopolitical, Environmental, and Financial) has a set of aspects, these aspects each contains a list of attributes which are synonyms obtain through a lexical English database called WordNet. Some of the attributes are shown in Fig. 9. The column called [rdf: type] show the aspects and the column called Resource show the attributes of the aspects.

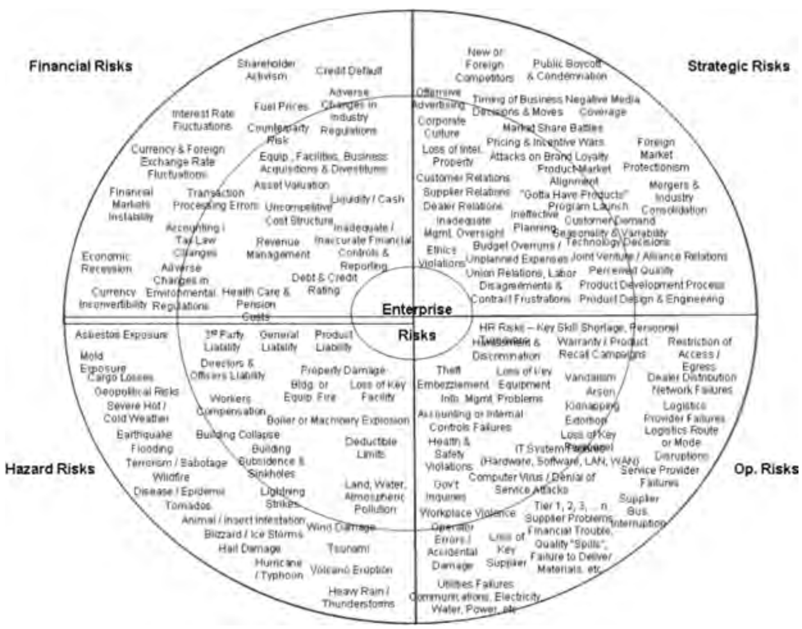


Fig. 7. Elkin's risk portfolio [8].

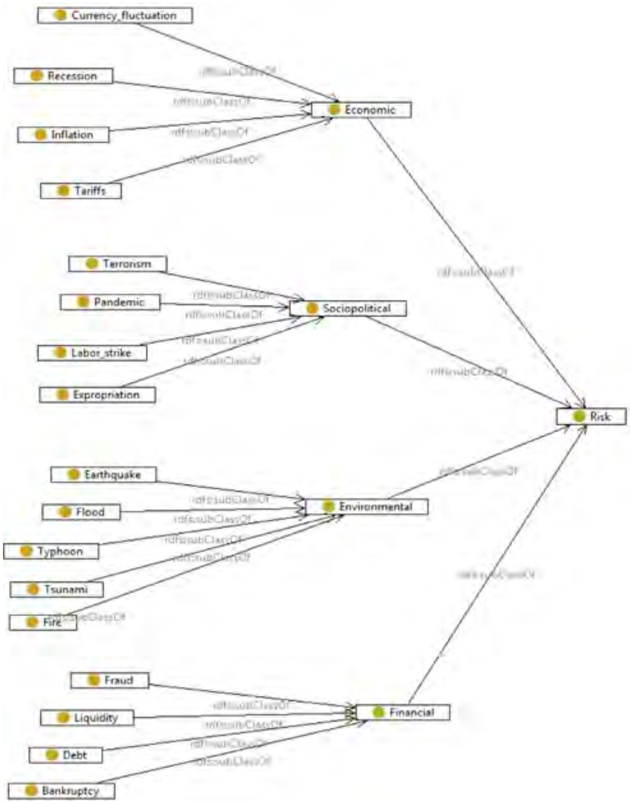


Fig. 8. Risk and uncertainty ontology.

Resource	[rdf:type]
Tax	Tariffs
Arrogation	Expropriation
Epidemic	Pandemic
Seiche	Tsunami
Miasmatic	Pandemic
Indebtedness	Debt
Alluvion	Flood
Impoverish	Bankruptcy
Deluge	Flood
Conflagration	Fire
Protest	Labor_strike
Strife	Labor_strike
Impounding	Expropriation
Outbreak	Pandemic
Belly_up	Bankruptcy
Confiscation	Expropriation
Duty	Tariffs
Inundation	Flood
Quake	Earthquake
Wildfire	Fire
Riot	Labor_strike

Fig. 9. Risk ontology instances.

Table 3
TESSA's dependencies.

Dependency	Version	Description
Python	2.7	Python programing language
Pip	1.5.2	Python package installer
Tweepy	2.3	Python-based package for extraction of twitter text
Regex	2.2.1	Python-based package for Regular Expression
NLTK	2.0.4	Python-based package for Natural Language Processing
Pymongo	2.7.1	Python-based package to interact with MongoDB
MongoDB	2.4.10	Cross-platform document oriented database

4.2. Requirement for TESSA

TESSA is a system built on Python programming language and it works with different packages to perform all the processes. The dependencies in order to create TESSA are all open sources and are shown in Table 3.

4.2.1. Twitter API requirements

Tweepy allows us to access the Twitter API. Twitter allows access to this huge corpus of data through different APIs that facilitate developers building new and creative applications. One of the APIs is the REST service, this service has limits on the search parameters and a number of queries per 15 minutes and therefore has a limited number of results; also limits the search due to the size of the bounding box which does not cover more than a small city. The other Twitter API is the streaming service. Twitter Streaming API is used to retrieve tweets from the public timeline which are the tweets publicly available in real-time only. The streaming API is most suitable for data mining and analytics research, it allows for large quantities of keywords to be specified and tracked and not have a limit on the geographical area it can retrieve tweets from, and we also have access to real-time data.

In this research we utilize the Streaming API from Twitter. In order to have authorized access by Twitter we must create an account on the Twitter developer site and generate the credentials. The steps to create the account are the following:

- Go to https://dev.twitter.com/. Here you can access using your Twitter account or create a new account.
- Create a new application to generate the credentials, as shown in Fig. 10.
- A permission model must be selected. In this case since we only needed to collect information and no modifications we select “Read only”, as shown in Fig. 11.
- The application is created and we now have the authorization tokens needed, as shown in Fig. 12.

Here we are provided to the consumer key, consumer secret, user token, and user secret in order to have authorized access to

Fig. 10. Application details.

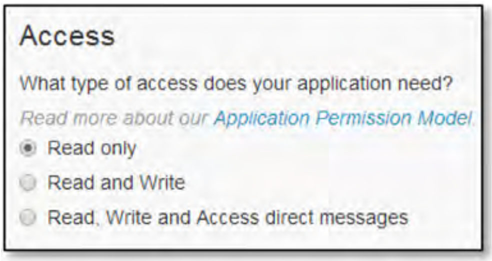


Fig. 11. Access permission model.

Fig. 12. Authorization tokens.

Twitter Streaming API.

4.3. *TESSA implementation*

TESSA is a tool that assists the supplier selection team to evaluate a potential supplier by providing current information regarding the risk the selected supplier holds. In this part we show a scenario where a sample company called "XYZ" is interested in 2 footwear manufacturing companies, company A and company B. Company A is a Taiwanese footwear manufacturing company with headquarters in Hong Kong; it has production facilities in China, Vietnam and Indonesia. Company B is a global sourcing company with headquarters in Hong Kong; it has production facilities in Bangladesh, China, India and Vietnam, among others. TESSA will assist in identifying what risk these 2 suppliers might present.

4.3.1. *Data extraction*

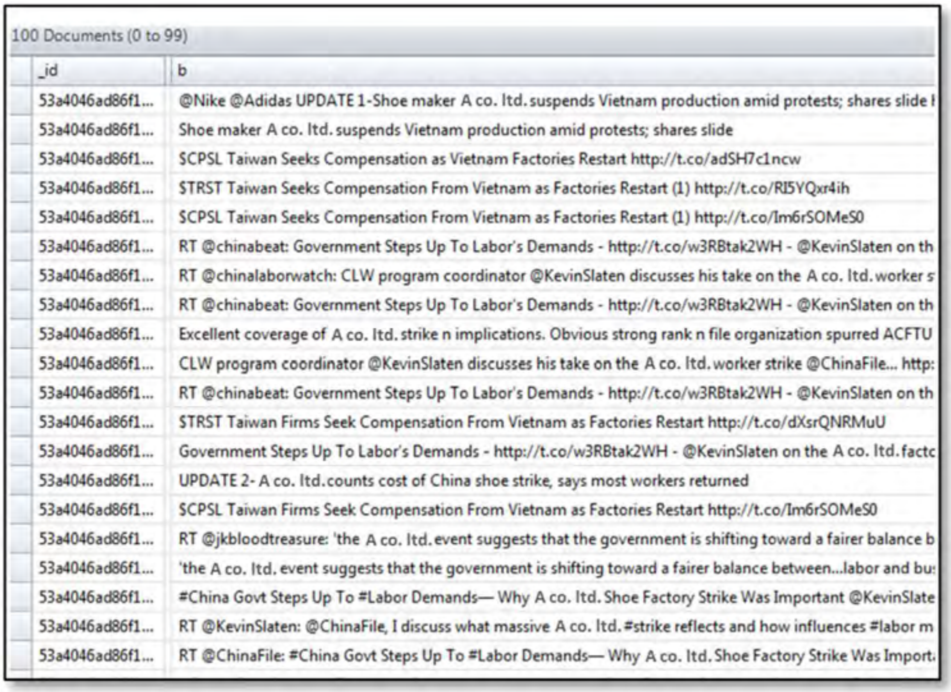
The requirements have already been set up and we have our authorization codes with access to Twitter's streaming API. The next step is to extract information regarding the suppliers company A and B, the search of the tweets is done by using the name of the supplier as keywords of the search, and the search is done by one supplier at the time.

Automatically a database is created in Mongo DB called "test01" and "test02" (one for each supplier), within this database a collection called "Tweets" is created where only the text of the tweet is stored so we can later preprocess it and classify it. The keyword used is the name of the supplier, in this case company A is the first one and then we do another search for company B, we also limit our search for tweets in English only.

We obtained 100 tweets from May 14th to May 17th, for company A the tweets are stored in the database "test01", as shown in Fig. 13. And we obtained 200 tweets regarding for company B, which are stored in the database "test02" as shown in Fig. 14.

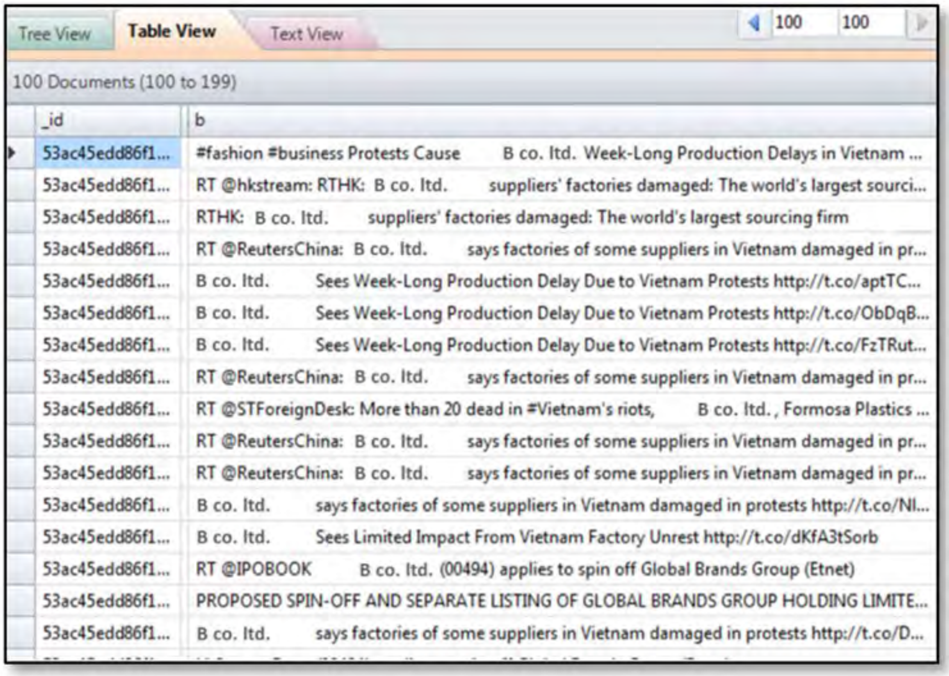
4.3.2. *Data preprocessing*

Now that the tweets have been normalized, they are saved in the database "Test01" for company A tweets and "Test02" for company B tweets; however they are saved into another collection called "cleantweets" . The tweets are now ready to be classified. The normalized tweets are shown in Fig. 15 for company A suppliers and Fig. 16 for company B.


100 Documents (0 to 99)

_id	b
53a4046ad86f1...	@Nike @Adidas UPDATE 1-Shoe maker A co. ltd. suspends Vietnam production amid protests; shares slide I
53a4046ad86f1...	Shoe maker A co. ltd. suspends Vietnam production amid protests; shares slide
53a4046ad86f1...	$CPSL Taiwan Seeks Compensation as Vietnam Factories Restart http://t.co/adSH7c1ncw
53a4046ad86f1...	$TRST Taiwan Seeks Compensation From Vietnam as Factories Restart (1) http://t.co/RI5YQxr4ih
53a4046ad86f1...	$CPSL Taiwan Seeks Compensation From Vietnam as Factories Restart (1) http://t.co/Im6rSOMeS0
53a4046ad86f1...	RT @chinabeat: Government Steps Up To Labor's Demands - http://t.co/w3RBtak2WH - @KevinSlaten on th
53a4046ad86f1...	RT @chinalaborwatch: CLW program coordinator @KevinSlaten discusses his take on the A co. ltd. worker s
53a4046ad86f1...	RT @chinabeat: Government Steps Up To Labor's Demands - http://t.co/w3RBtak2WH - @KevinSlaten on th
53a4046ad86f1...	Excellent coverage of A co. ltd. strike n implications. Obvious strong rank n file organization spurred ACFTU
53a4046ad86f1...	CLW program coordinator @KevinSlaten discusses his take on the A co. ltd. worker strike @ChinaFile... http:
53a4046ad86f1...	RT @chinabeat: Government Steps Up To Labor's Demands - http://t.co/w3RBtak2WH - @KevinSlaten on th
53a4046ad86f1...	$TRST Taiwan Firms Seek Compensation From Vietnam as Factories Restart http://t.co/dXsrQNRMuU
53a4046ad86f1...	Government Steps Up To Labor's Demands - http://t.co/w3RBtak2WH - @KevinSlaten on the A co. ltd. factc
53a4046ad86f1...	UPDATE 2- A co. ltd. counts cost of China shoe strike, says most workers returned
53a4046ad86f1...	$CPSL Taiwan Firms Seek Compensation From Vietnam as Factories Restart http://t.co/Im6rSOMeS0
53a4046ad86f1...	RT @jkbloodtreasure: 'the A co. ltd. event suggests that the government is shifting toward a fairer balance b
53a4046ad86f1...	'the A co. ltd. event suggests that the government is shifting toward a fairer balance between...labor and bu:
53a4046ad86f1...	#China Govt Steps Up To #Labor Demands— Why A co. ltd. Shoe Factory Strike Was Important @KevinSlate
53a4046ad86f1...	RT @KevinSlaten: @ChinaFile, I discuss what massive A co. ltd. #strike reflects and how influences #labor m
53a4046ad86f1...	RT @ChinaFile: #China Govt Steps Up To #Labor Demands— Why A co. ltd. Shoe Factory Strike Was Import

Fig. 13. Tweets of company A.


Tree View Table View Text View 100 100

100 Documents (100 to 199)

_id	b
53ac45edd86f1...	#fashion #business Protests Cause B co. ltd. Week-Long Production Delays in Vietnam ...
53ac45edd86f1...	RT @hkstream: RTHK: B co. ltd. suppliers' factories damaged: The world's largest sourci...
53ac45edd86f1...	RTHK: B co. ltd. suppliers' factories damaged: The world's largest sourcing firm
53ac45edd86f1...	RT @ReutersChina: B co. ltd. says factories of some suppliers in Vietnam damaged in pr...
53ac45edd86f1...	B co. ltd. Sees Week-Long Production Delay Due to Vietnam Protests http://t.co/aptTC...
53ac45edd86f1...	B co. ltd. Sees Week-Long Production Delay Due to Vietnam Protests http://t.co/ObDqB...
53ac45edd86f1...	B co. ltd. Sees Week-Long Production Delay Due to Vietnam Protests http://t.co/FzTRut...
53ac45edd86f1...	RT @ReutersChina: B co. ltd. says factories of some suppliers in Vietnam damaged in pr...
53ac45edd86f1...	RT @STForeignDesk: More than 20 dead in #Vietnam's riots, B co. ltd., Formosa Plastics ...
53ac45edd86f1...	RT @ReutersChina: B co. ltd. says factories of some suppliers in Vietnam damaged in pr...
53ac45edd86f1...	RT @ReutersChina: B co. ltd. says factories of some suppliers in Vietnam damaged in pr...
53ac45edd86f1...	B co. ltd. says factories of some suppliers in Vietnam damaged in protests http://t.co/NI...
53ac45edd86f1...	B co. ltd. Sees Limited Impact From Vietnam Factory Unrest http://t.co/dKfA3tSorb
53ac45edd86f1...	RT @IPOBOOK B co. ltd. (00494) applies to spin off Global Brands Group (Etnet)
53ac45edd86f1...	PROPOSED SPIN-OFF AND SEPARATE LISTING OF GLOBAL BRANDS GROUP HOLDING LIMITE...
53ac45edd86f1...	B co. ltd. says factories of some suppliers in Vietnam damaged in protests http://t.co/D...

Fig. 14. Tweets of company B.

4.3.3. Data classification

With the previously built Risk and Uncertainty Ontology a lexical database is built in order to classify the tweets within the different aspects. The tweets go through the lexical database and are matched within the aspects and attributes. If there is a match a number 1 is reflected and the tweet is saved in a collection under the name of the aspect, if there is no match a number 0 is shown, the tweets that don't belong to any aspect are classified as trash. An example is shown in Fig. 17.

100 Documents (0 to 99)

_id	text
53ac42db89cd...	update 1shoe maker A co. ltd. suspends vietnam production amid protests shares...
53ac42db89cd...	shoe maker A co. ltd.suspends vietnam production amid protests shares slide
53ac42db89cd...	cpsl taiwan seeks compensation vietnam factories restart
53ac42db89cd...	trst taiwan seeks compensation vietnam factories restart 1
53ac42db89cd...	cpsl taiwan seeks compensation vietnam factories restart 1
53ac42db89cd...	government steps laboru2019s demands A co. ltd. factory strikes importance
53ac42db89cd...	clw program coordinator discusses take A co. ltd. worker strike u2026
53ac42db89cd...	government steps laboru2019s demands A co. ltd. factory strikes importance
53ac42db89cd...	excellent coverage A co. ltd. strike n implications obvious strong rank n file organi...
53ac42db89cd...	clw program coordinator discusses take A co. ltd. worker strike
53ac42db89cd...	government steps laboru2019s demands A co. ltd. factory strikes importance
53ac42db89cd...	trst taiwan firms seek compensation vietnam factories restart
53ac42db89cd...	government steps laboru2019s demands A co. ltd. factory strikes importance
53ac42db89cd...	update 2 A co. ltd. counts cost china shoe strike says workers returned
53ac42db89cd...	cpsl taiwan firms seek compensation vietnam factories restart
53ac42db89cd...	A co. ltd. event suggests government shifting toward fairer balance betweenlabor...
53ac42db89cd...	A co. ltd. event suggests government shifting toward fairer balance betweenlabor...
53ac42db89cd...	china govt steps labor demandsu2014 A co. ltd. shoe factory strike important
53ac42db89cd...	discuss massive A co. ltd. strike reflects influences labor movement china u2026
53ac42db89cd...	china govt steps labor demandsu2014 A co. ltd. shoe factory strike important http...
53ac42db89cd...	china govt steps labor demandsu2014 A co. ltd. shoe factory strike important

Fig. 15. Normalized tweets for company A.

Out of the 100 tweets for company A only 1 category was detected which is labor strike, 68 tweets were identified in this category (see Fig. 18) and 32 tweets were classified as trash. The classification took 11 s in total.

Out of the 200 tweets for company B, only 1 category was detected which is labor strike, 87 tweets were identified within this category (see Fig. 19) and 113 were identified as trash. The classification took 22 s in total.

These 2 suppliers present some sociopolitical risk, due to the labor strikes aspects. Now the company "XYZ" is aware of this critical information when making their final decision.

4.3.4. Evaluation

After manually checking the tweets for company A and B and we can build our confusion matrix and evaluate the accuracy, precision, recall and f-score. First we review company A, see Table 4.

TESSA's classification result for company A were 68 tweets for Labor Strike and 32 tweets for Trash, out of 100 tweets; however, after manually revising we found that there are 6 tweets that actually correspond to the Trash category, meaning that TESSA classified 26 tweets as Trash when they actually were supposed to be classified as Labor strike. Using the equation previously stated in chapter 3 we can calculate the accuracy, precision, recall and f-score. Results are shown in Table 5.

In the results we can see that TESSA scored high results in all the metrics and the tool has 0 false positives, showing to be an efficient classifier for this aspect with 100% score in precision. However, it missed 26 tweets from the strike category and labels them as trash.

We also manually revised the 200 tweets for company B, the confusion matrix is shown in Table 6.

TESSA's classification result for company B were 87 tweets for Labor Strike and 113 tweets for Trash, out of 200 tweets; however, after manually revising we found that there are 86 tweets that actually correspond to the Trash category, meaning that TESSA classified 27 tweets as Trash when they actually were supposed to be classified as Labor strike. Using the equation previously stated in chapter 3 we can calculate the accuracy, precision, recall and f-score. Results are shown in Table 7.

In the results for company B TESSA scored high in all the metrics and showed no false positives. The f-score is very close to 100% providing evidence of the good performance the classifier has.

In the case of both of the suppliers TESSA missed several tweets. When manually revising them we noticed the tweets that were labeled as trash, but where supposed to be labeled as Labor strike had attributes that were not included in our ontology, such as

100 Documents (100 to 199)

_id	text
53ac462189cd1...	fashion business protests cause B co. ltd. weeklong production delays vietnam
53ac462189cd1...	B co. ltd. suppliers factories damaged worlds largest sourcing firm B co. ltd. says
53ac462189cd1...	rthk B co. ltd. suppliers factories damaged worlds largest sourcing firm B co. ltd.
53ac462189cd1...	B co. ltd. says factories suppliers vietnam damaged protests
53ac462189cd1...	B co. ltd. sees weeklong production delay due vietnam protests
53ac462189cd1...	B co. ltd. sees weeklong production delay due vietnam protests business fashi
53ac462189cd1...	B co. ltd. sees weeklong production delay due vietnam protests fashion busin
53ac462189cd1...	B co. ltd. says factories suppliers vietnam damaged protests
53ac462189cd1...	20 dead vietnams riots B co. ltd. formosa plastics group voice concern u2026
53ac462189cd1...	B co. ltd. says factories suppliers vietnam damaged protests
53ac462189cd1...	B co. ltd. says factories suppliers vietnam damaged protests
53ac462189cd1...	B co. ltd. says factories suppliers vietnam damaged protests
53ac462189cd1...	B co. ltd. sees limited impact vietnam factory unrest
53ac462189cd1...	B co. ltd. 00494 applies spin global brands group etnet
53ac462189cd1...	proposed spinoff separate listing global brands group holding limited main b
53ac462189cd1...	B co. ltd. says factories suppliers vietnam damaged protests industries via

Fig. 16. Normalized tweets for company B.

```
liquidity crisis Score:
0
expropriation Score:
0
labor strike Score:
1
pandemic Score:
0
terrorism Score:
0
 labor strike
{"text": " protests vietnam cause major footwear
manufacturer halt production details    "}
```

Fig. 17. Lexical Base classifier.

"production delays", "production suspended", "restart production" and "factories damaged" (see Fig. 20).

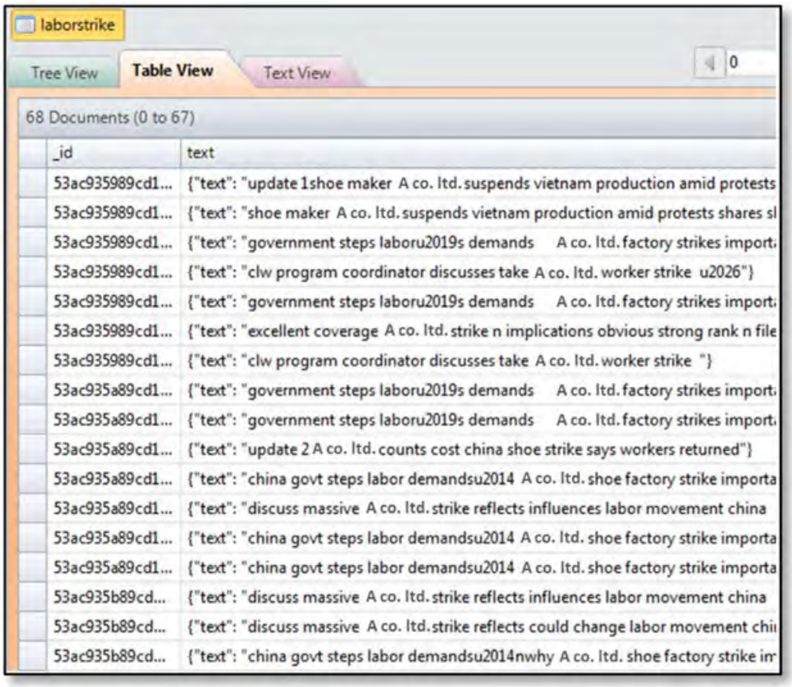
TESSA exposed the risk that company A and B present, which is Labor Strike, this is due to the riots happening in Vietnam in the month of May. Since both companies have production facilities in Vietnam they present risk for the Company XYZ interested in making business with them. Now the XYZ Company can use this knowledge in order to make their final decision.

4.4. Limitation

The proposed TESSA retrieves data by using Twitter Streaming API, meaning that it gathers live tweets without storing historical data because the process generally involves huge data volume. In case that very few data is retrieved with respect to a target supplier, TESSA is still capable of making suggestion on the risk level of selecting the supplier. However, the suggestion could be inadequate due to the lack of data.

5. Conclusion

The development of Twitter Enabled Supplier Status Assessment can be critical for any company in need of an international supplier, the results obtained show high levels of accuracy and performance. This tool allows companies to have a better insight when selecting an international supplier. It can assist the decision team by narrowing down their list of potentials suppliers by exposing



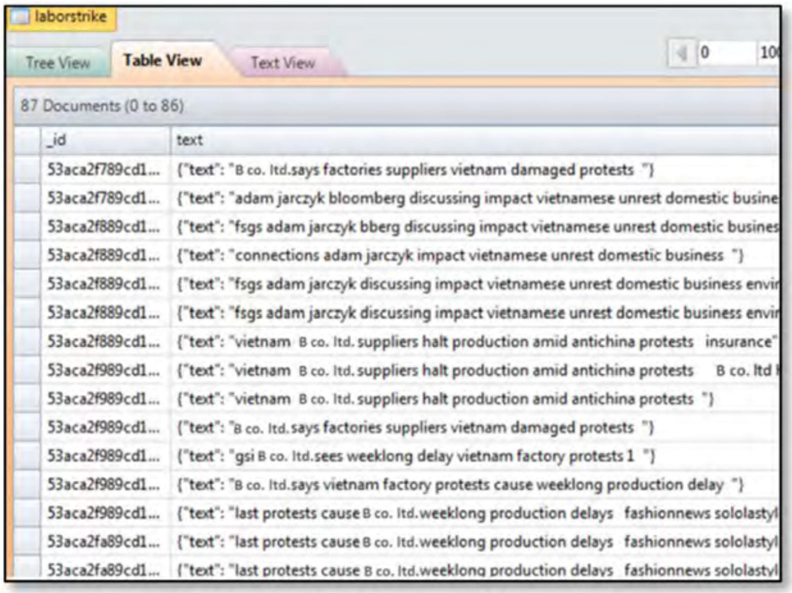
laborstrike

Tree View | Table View | Text View

68 Documents (0 to 67)

_id	text
53ac935989cd1...	{"text": "update 1shoe maker A co. ltd. suspends vietnam production amid protests
53ac935989cd1...	{"text": "shoe maker A co. ltd. suspends vietnam production amid protests shares sl
53ac935989cd1...	{"text": "government steps laboru2019s demands A co. ltd. factory strikes import.
53ac935989cd1...	{"text": "clw program coordinator discusses take A co. ltd. worker strike u2026"}
53ac935989cd1...	{"text": "government steps laboru2019s demands A co. ltd. factory strikes import.
53ac935989cd1...	{"text": "excellent coverage A co. ltd. strike n implications obvious strong rank n file
53ac935989cd1...	{"text": "clw program coordinator discusses take A co. ltd. worker strike "}
53ac935a89cd1...	{"text": "government steps laboru2019s demands A co. ltd. factory strikes import.
53ac935a89cd1...	{"text": "government steps laboru2019s demands A co. ltd. factory strikes import.
53ac935a89cd1...	{"text": "update 2 A co. ltd. counts cost china shoe strike says workers returned"}
53ac935a89cd1...	{"text": "china govt steps labor demandsu2014 A co. ltd. shoe factory strike importa
53ac935a89cd1...	{"text": "discuss massive A co. ltd. strike reflects influences labor movement china
53ac935a89cd1...	{"text": "china govt steps labor demandsu2014 A co. ltd. shoe factory strike importa
53ac935a89cd1...	{"text": "china govt steps labor demandsu2014 A co. ltd. shoe factory strike importa
53ac935b89cd...	{"text": "discuss massive A co. ltd. strike reflects influences labor movement china
53ac935b89cd...	{"text": "discuss massive A co. ltd. strike reflects could change labor movement chi
53ac935b89cd...	{"text": "china govt steps labor demandsu2014nwhy A co. ltd. shoe factory strike im

Fig. 18. Classified tweets of company A.



laborstrike

Tree View | Table View | Text View

87 Documents (0 to 86)

_id	text
53aca2f789cd1...	{"text": "B co. ltd.says factories suppliers vietnam damaged protests "}
53aca2f789cd1...	{"text": "adam jarczyk bloomberg discussing impact vietnamese unrest domestic busine
53aca2f889cd1...	{"text": "fsgs adam jarczyk bberg discussing impact vietnamese unrest domestic busines
53aca2f889cd1...	{"text": "connections adam jarczyk impact vietnamese unrest domestic business "}
53aca2f889cd1...	{"text": "fsgs adam jarczyk discussing impact vietnamese unrest domestic business envi
53aca2f889cd1...	{"text": "fsgs adam jarczyk discussing impact vietnamese unrest domestic business envi
53aca2f889cd1...	{"text": "vietnam B co. ltd. suppliers halt production amid antichina protests insurance"
53aca2f989cd1...	{"text": "vietnam B co. ltd. suppliers halt production amid antichina protests B co. ltd
53aca2f989cd1...	{"text": "vietnam B co. ltd. suppliers halt production amid antichina protests "}
53aca2f989cd1...	{"text": "B co. ltd. says factories suppliers vietnam damaged protests "}
53aca2f989cd1...	{"text": "gsi B co. ltd.sees weeklong delay vietnam factory protests 1 "}
53aca2f989cd1...	{"text": "B co. ltd.says vietnam factory protests cause weeklong production delay "}
53aca2f989cd1...	{"text": "last protests cause B co. ltd.weeklong production delays fashionnews sololastyl
53aca2fa89cd1...	{"text": "last protests cause B co. ltd.weeklong production delays fashionnews sololastyl
53aca2fa89cd1...	{"text": "last protests cause B co. ltd.weeklong production delays fashionnews sololastyl

Fig. 19. Classified tweets of company B.

Table 4
The confusion matrix of company A.

		Predicted	
		Labor strike	Trash
Actual	Labor strike	68	26
	Trash	0	6

Table 5
The metrics results of company A.

Metric	Result
Accuracy	74%
Precision	100%
Recall	72%
F-score	83%

Table 6
The confusion matrix of company B.

		Predicted	
		Labor strike	Trash
Actual	Labor strike	87	27
	Trash	0	86

Table 7
The metrics results of company B.

Metric	Result
Accuracy	86%
Precision	100%
Recall	76%
F-score	86%

Fig. 20. Example of false negative.

their risks and uncertainties; It can give the supplier selection team more knowledge to make the appropriate decision regarding their needs.

In the supplier selection process a big issue is the lack of sufficient information, companies use complex methodologies to overcome the lack of data by using different statistical methods; however, the tool of Twitter Enabled Supplier Status Assessment holds a better solution, to take advantage of an immense pool of information which is Social Media. This tool can be used with other social media sites not just Twitter and allow companies to have access to the data they long for. This data is accurate, up to date and non-biased.

Another contribution from this research is the Risk and Uncertainty Ontology developed, it portraits a general view of the risks in the industry, with different aspects and attributes to help the classification be more fine-grained. One of Twitter Enabled Supplier Status Assessment tool main perks is that is created with open source tools, making it accessible for everyone.

Twitter Enabled Supplier Status Assessment can help companies get more involved in this Big Data era and take advantage of the knowledge. This tool can help companies integrate Social Media as a source of information not just for their future suppliers, but for their current ones as well. They can be up to date regarding any risks their current suppliers might be exposed to, providing them more control over their supply chain.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.compeleceng.2018.03.042.

References

[1] Sawik T. Selection of supply portfolio under disruption risks. Omega 2011;39:194–208.
[2] Amid A, Ghodsypour SH, O'Brien C. A weighted max–min model for fuzzy multi-objective supplier selection in a supply chain. Int J Prod Econ 2011;131:139–45.
[3] Luo X, Zhang J, Duan W. Social media and firm equity value. Inf Syst Res 2013;24:146–63.
[4] Tiwari S, Wee HM, Daryanto Y. Big data analytics in supply chain management between 2010 and 2016: Insights to industries. Comput Ind Eng 2018;115:319–30.
[5] Gunasekaran A, Papadopoulos T, Dubey R, Wamba SF, Childe SJ, Hazen B, Akter S. Big data and predictive analytics for supply chain and organizational performance. J Bus Res 2017;70:308–17.
[6] Monczka RM, Handfield RB, Giunipero LC, Patterson JL. Purchasing & Supply Chain Management. 5th ed. State of Ohio: Thomson; 2011.
[7] Wisner JD, Tan KC, Leong GK. Principles of supply chain management: a balanced approach. 4th ed. Cengage Learning; 2011.
[8] Elkins D, Kulkarni D, Tew J. Identifying and assessing supply chain risk. Supply Chain Risk Manag: Minimizing Disrupt Glob Sourc 2008:51–6.
[9] Aghai S, Mollaverdi N, Sabbagh MS. A fuzzy multi-objective programming model for supplier selection with volume discount and risk criteria. Int J Adv Manuf Technol 2014;71:1483–92.
[10] Vinodh S, Ramiya RA, Gautham S. Application of fuzzy analytic network process for supplier selection in a manufacturing organisation. Expert Syst Appl 2011;38:272–80.
[11] Sreedevi R, Saranga H. Uncertainty and supply chain risk: The moderating role of supply chain flexibility in risk mitigation. Int J Prod Econ 2017;193:332–42.
[12] Bravo-M F, Frank E, Pfahringer B. Building a Twitter opinion lexicon from automatically-annotated tweets. Knowl Based Syst 2016;108:65–78.
[13] Bennett S. Just how big is twitter in 2012. Infographic, News, Portfolio; 2012 Filed Under:Tagged With: infographics, Social Media, twitter. http://www.adweek.com/digital/twitter-statistics-2012/?red=at; [Accessed on 14.03.28].
[14] Sakaki T, Okazaki M, Matsuo Y. Tweet analysis for real-time event detection and earthquake reporting system development. IEEE Trans Knowl Data Eng 2013;25:919–31.
[15] O'Leary DE. The use of social media in the supply chain: survey and extensions. Intell Syst Account Finance Manag 2011;18:121–44.
[16] Yu HF, Ho CH, Arunachalam P, Somaiya M, Lin CJ. Product title classification versus text classification 2012 Technical Report. https://www.csie.ntu.edu.tw/~cjlin/papers/title.pdf; [Accessed on 15.01.05].
[17] Kamruzzaman SM. Text classification using artificial intelligence. Int J Electr Eng 2006:33.
[18] Wikipedia. Recurrent neural network. Ed: February 2018 https://en.wikipedia.org/wiki/Recurrent_neural_network [Accessed on 18.02.23].
[19] Liu P, Qui X, Huang X. Recurrent neural network for text classification with multi-task learning. Proceedings of IJCAI'16 Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, USA. 2016 July 2016. p. 2873–9.
[20] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. Cornell University Library, Computation and Language; 2013 September https://arxiv.org/abs/1301.3781 [Accessed on 18.02.23].
[21] Tang D, Wei F, Yang N, Zhou M, Liu T, Qin B. Learning sentiment-specific word embedding for twitter sentiment classification. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), Maryland, USA, 2014. 2014. p. 1555–65.
[22] Rajeswari V, Varughese DK, Selvan A. A novel method to transform relational data into ontology in the bio-medical domain. Int. J. Eng. Technol 2013;5:0975–4024.
[23] Harman K, Koohang A. Learning objects 2: standards, metadata, repositories, and LCMS. 1st ed. Informing Science; 2006.
[24] Gerber A, Morar N, Meyer T, Eardley C. Ontology-based support for taxonomic functions. Ecol Inf 2017;41:11–23.
[25] Iwanaga ISM, Nguyen TM, Kawamura T, Nakagawa H, Tahara Y, Ohsuga A. Building an earthquake evacuation ontology from twitter. Granular Computing (GrC), 2011 International Conference on. 2011. p. 306–11. Kaohsiung, Taiwan. November.
[26] Baldoni M, Baroglio C, Patti V, Rena P. From tags to emotions: Ontology-driven sentiment analysis in the social semantic web. Intell Artif 2012;6:41–54.
[27] Kontopoulos E, Berberidis C, Dergiades T, Bassiliades N. Ontology-based sentiment analysis of twitter posts. Expert Syst Appl 2013;40:4065–74.
[28] Staab S, Studer R. Handbook on Ontologies. In: Staab S, Studer R, editors. International handbooks on information systems. 2009.
[29] Kohavi R, Provost F. Glossary of terms. Mach Learn 1998;30:271–4.
[30] Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf Process Manag 2009;45:427–37.

Chuan-Jun Su is currently with the Department of Industrial Engineering and Management at the Yuan Ze University, Taiwan, ROC. His interests include Information Technology, IOT, Big Data Analytics, and industrial applications of AI.

Yin-An Chen is currently a Ph.D. student in the Department of Industrial Engineering and Management at the Yuan Ze University, Taiwan, ROC. His interests include Information Technology, Computer-communication Network, Internet of Things, and AI applications in design and manufacturing.