

# Text Mining Analysis of Wind Turbine Accidents:

## An Ontology-Based Framework

Gurdal Ertek

College of Business  
Abu Dhabi University  
Abu Dhabi, U.A.E.  
gurdal.ertek@adu.ac.ae

Allan N. Zhang

Singapore Institute of Manufacturing Technology  
(SIMTech), Agency for Science,  
Technology and Research (A\*STAR)  
Singapore  
nzhang@simtech.a-star.edu.sg

Xu Chi

Singapore Institute of Manufacturing Technology  
(SIMTech), Agency for Science,  
Technology and Research (A\*STAR)  
Singapore  
cxu@simtech.a-star.edu.sg

Sobhan Asian

College of Business  
RMIT University  
Melbourne, Australia  
sobhan.asian@rmit.edu.au

**Abstract**— As the global energy demand is increasing, the share of renewable energy and specifically wind energy in the supply is growing. While vast literature exists on the design and operation of wind turbines, there exists a gap in the literature with regards to the investigation and analysis of wind turbine accidents. This paper describes the application of text mining and machine learning techniques for discovering actionable insights and knowledge from news articles on wind turbine accidents. The applied analysis methods are text processing, clustering, and multidimensional scaling (MDS). These methods have been combined under a single analysis framework, and new insights have been discovered for the domain. The results of our research can be used by wind turbine manufacturers, engineering companies, insurance companies, and government institutions to address problem areas and enhance systems and processes throughout the wind energy value chain.

**Keywords**— text mining; ontology; accident analysis; wind turbine accidents; wind energy.

### I. INTRODUCTION

Energy demand in the world is growing at a fast pace. According to International Energy Agency, world electricity demand will increase by more than two-thirds over the period 2011-2035. In this growth scenario, the share of renewable energy sources in total power generation will rise from 20% in 2011 to 31% in 2035, and renewables will eventually overtake gas and eventually coal as the energy source [1]. As the world demand for energy and renewable energy is growing, so is the popularity of wind energy. According to 2016 statistics released by The Global Wind Energy Council (GWEC), the cumulative global wind energy capacity reached a total of 318,137 MW by the end of 2016 and will more than double by 2021 to 800GW by 2021 [2].

Wind turbines (Figure 1) are mechatronic devices that convert wind energy into electrical energy via mechanical

energy. Figure 1 demonstrates the basic components of a wind turbine. A successful wind turbine design depends on the proper design of the turbine's components: Blades are designed with aerodynamic calculations to maximize the torque; rotor is the rotational device that turns the shaft and starts the energy production cycle; the generator is where the energy is finally produced. There are both low-speed shafts and high-speed shafts in a typical wind turbine design; gears in the gear box make the shafts turn at the correct speed; brakes work for emergency stopping or slowing down of the turbines. Anemometer calculates the wind speed. Some turbines control and adjust the shaft speed based on the wind speed read from the anemometer. The yaw control system is another mechanical element that increases the wind turbine's efficiency by minimizing non-symmetrical loads and hence increasing power output [3].

While wind energy industry and the installation of wind turbines are growing, there is comparably little discussion of the possible shortcomings of this energy source, and much less discussion of wind turbine accidents.

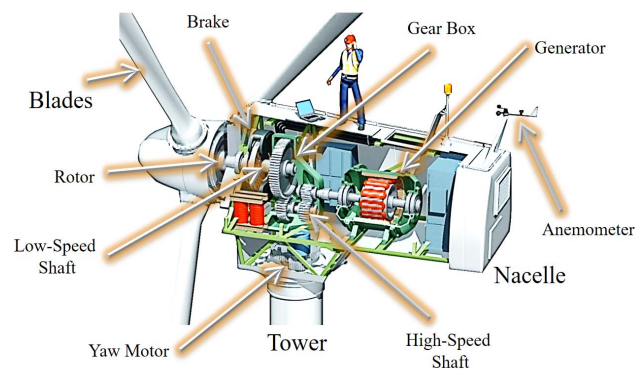


Fig. 1. A wind turbine and its basic mechanical components.

Wind turbine accidents have significance due to several reasons: Firstly, some accidents result in human casualties and injuries, directly and immediately affecting human life. Secondly, wind turbine accidents, especially when they result in human losses, negatively affect the public perception of wind energy and the wind turbine industry. Thirdly, accidents result in various types of costs (loss of revenue, cost of maintenance, loss of reputation, and other costs) and have implications for investors, operators, manufacturers, contractors, and insurance companies, as well as any other party involved in the supply chain. Finally, from a supply chain risk management perspective [4], such accidents may cause disruptions in the energy supply, possibly causing shortages and blackouts at facility, industry, city, or region level.

As of now, the most extensive data on the Internet about wind turbine accidents is available through Caithness Windfarm Information Forum [5], which contains a list of more than 1,900 wind turbine accidents. However, most web links in this list to the news sources do not work properly, and thus the reliability of the list is open to questioning. Other Internet sources contain significantly less data. To our best knowledge, there are no academic studies on mining data from wind turbine accidents, except [6]. The mentioned paper analyses tabularized data derived from wind turbine accident news, and does not apply any text mining techniques. Therefore, there is a significant gap of knowledge and insights throughout the world with regards to the understanding of wind turbine accidents from a broad unbiased perspective, especially through analytics.

Given the growth of the wind turbine industry, and the lack of academic research as well as industry research, we have decided to contribute to the literature in this area. To this end, we have performed a thorough investigation of wind turbine accidents in business news databases and on the Internet, and have created a database of 218 accident news, with full text available, and with confirmed references to the news sources. Then, we set forward to answering the following critical research questions:

- 1) How can the accident news text collection be analyzed to come up with insights into wind turbine accidents?
- 2) What are the terms that appear in the accident news and how are the terms related with each other?

We answer the first research question by introducing an analysis framework based on text mining, for analyzing unstructured wind turbine accident news data. The second research question is answered by applying the developed analysis framework to the text collection, and by discovering new insights into wind turbine accidents and the way they are reported on the media and the Internet.

The remainder of the paper is organized as follows: Section 2 provides a brief review of some relevant literature as the background. Section 3 discusses the methodologies used in the data analysis, including data mining, clustering, multidimensional scaling (MDS), and text mining. Section 4 describes the developed framework. Section 5 communicates the analysis and results, containing the description of the data

and the presentation of discovered insights. Finally, Section 6 presents some conclusive remarks.

## II. LITERATURE

Text mining applications are ubiquitous, spanning a multitude of industries, including construction [7], automotive [8], and process [9] industries. Text mining has also widely been applied to renewable energy, including wind energy, and it has also been utilized for accident data analysis. However, our extensive review shows that none of existing studies has been conducted for accidents related to the wind turbine.

### A. Text Mining for Renewable Energy Industry

Many of the existing works applying text mining to the renewable energy industry focus on identification of technology trends [10][11][12]. In [10], a large taxonomy for renewable energy is developed for comprehensive trend discovery, based on the scientific articles. The study in [11] conducts patent analysis using text mining techniques to explore patterns of innovation and evolution of energy technologies and compares their innovation characteristics. Text mining is applied in [12] to 12,000 patent documents, providing a visual map of the emphasis areas for R&D in renewable energy field.

### B. Text Mining for Wind Energy and Wind Turbine

Another group of studies focus on wind energy and wind turbines, and apply text mining [13]-[15]. For example, [13] proposes a new approach in text-mining to identify promising patents for technology transfer and [14] introduces a semi-automatic approach to build ontology for wind energy domain using Wikipedia articles. However, none of these studies apply text mining to the analysis of wind turbine accidents.

### C. Text Mining of Accident Data

Electronic accident reports usually comprise of large amount of unstructured text and are primary data sources for knowledge discovery of accident information using text mining techniques [14]-[19]. The majority of the existing works for text mining on accident data aims to identify the risk factors, which may be used to improve safety measures and safety training, identify new areas of invention and support new engineering strategies for safety. There does not exist any study in the literature on the text mining of wind turbine accidents.

## III. METHODOLOGY

### A. Data Mining

*Data mining* is the field of computer science that deals with a fundamental question: "How can we analyze data using computational methods, so that we can extract useful information and discover valuable knowledge from it?" [20]. The field of data mining is sometimes also referred to as data analytics (despite subtle differences between the two concepts), and it is the cornerstone of popular concepts of business analytics [21] and big data [22]. Data mining encompasses a multitude of methods that are used to analyze different types of data and for different types of goals [23]-[27]. The ultimate

goal of all these methods is to provide organizations and people in organizations with value, obtained from—typically vast volume, various, and high-velocity— data.

### B. Clustering

*Clustering* is an unsupervised machine learning technique, which aims at grouping data objects (observations, examples) into groups or *clusters*, such that observations within a cluster have high similarity, whereas they are dissimilar to the observations in other clusters (Han et al. [20], page 443). Clustering results in groups of observations represented by centroids, that is, central representative points. In our research, we applied clustering to group the text documents and describe the terms in the text documents.

### C. Multidimensional scaling (MDS)

*Multidimensional scaling (MDS)* is a method for reducing the dimensionality of a dataset by mapping it on a lower-dimensional space [28]. The mapping is performed based on the distance between the data objects, such that strain (a loss function) is minimized. For example, in this study, MDS is used to map text documents and terms, respectively, onto two-dimensional planes. The distance norm in MDS can be one of many different norms, including Euclidean, Pearson correlation, and Hamming. In this study, Pearson correlation was the selected distance norm. MDS was selected because it preserves the distance between data points. Alternative dimensionality reduction technique of Principal Component Analysis (PCA) can also be used if the objective is to preserve covariance of data. The focus of this paper is not the specific technique used for dimensionality reduction and instead it is the text mining process, hence we did not apply any alternative methods at this step.

### D. Text Mining

*Text mining* is the branch of data mining aimed at discovering interesting and nontrivial knowledge from text documents [29]. The common practice in text mining is the extraction of critical attribute information from unstructured text documents through text processing, and then analyzing this structured constructed data with well-known data mining algorithms [30]. Attribute selection allows the removal of the irrelevant and noisy information present in text documents and focuses to only on relevant and informative data for use in text mining.

In the text mining study presented in this paper, we followed the described common approach of processing text documents, extracting term frequencies as features, and using this feature set to perform clustering and multidimensional scaling (MDS).

### E. Ontology Development

*Ontology* is a branch of philosophy. *Descriptive ontology* is concerned with the collection of information, whereas *formal ontology* distills, filters, codifies and organizes the results of descriptive ontology. While ontology is a scientific discipline, an *ontology* is a classification of categories; a formal, explicit specification of shared conceptualization

([31], page 4; [32], page 3). The five basic principles for the design of ontologies are clarity, minimal encoding bias, extendibility, coherence, minimal ontological commitments ([31], pages 8-9).

In this study, initially a set of 474 terms were obtained through filtering by a human domain expert from among the frequent terms that were obtained through text processing. Then an ontology was constructed for terms, by further filtering among these filtered terms and categorizing them. The ontology consists of four categories of *Month*, *Turbine Component*, *Country*, and *Outcome*. In the analysis and results stage of the study, the ontology allowed focusing on the interactions between the terms in these main categories. The 40 terms included in the ontology for the wind turbine accidents are hereafter referred to as *ontology terms*.

## IV. FRAMEWORK

The analysis framework is illustrated in Figure 2. The framework takes the text collection as the primary input. Two term sets are used in analyzing the text collection: Term Set 1 contains terms filtered by a human expert that are relevant to the domain. Term Set 2 contains only the terms within Term Set 1 which are part of a particular ontology.

Firstly, text processing is applied to identify all the frequent terms in the text collection and compute their term frequencies.

For text processing and attribute extraction, the process given in [30] was followed, using the RapidMiner software (<http://rapidminer.com>). Text processing begins with the reading of data from the text collection and continues with the manipulation of this data using text processing algorithms. There are two important settings that are fundamental to the text processing: In the study, the vector creation method was selected as *term frequency*. This method results in the computation of the relative frequencies of each of the terms in each of the documents in the data set. For example, if a term appears 4 times within a document that consist of 100 words, then the relative frequency of that term is  $4/100 = 0.04$ . This value of 0.04 appears in the constructed dataset, under the column for that term, at the row for that document. Another critical setting is the pruning method, which was selected as *percentual pruning* in this study. Value for the *prune below percent* parameter was set as 0.05, meaning that we selected and constructed feature vector for words that appear in at least 5% of the documents in the text collection.

There are seven operators nested within the text processing process, that are serially linked: 1) *Tokenize Nonletters* and 2) *Tokenize Linguistic* operators are both created by selecting the *Tokenize* operator, but with different parameter selections. The former operator tokenizes based on non-letters whereas the latter operator tokenizes based on the linguistic sentences within the English language. 3) *Filter Stopwords* (English) operator removes the stop words in the English language from the text data set. 4) *Filter Tokens (by Length)* operator removes all the words composed of less than min chars characters and more than max chars characters. In our study, words that have less than 2 characters or more than 25 characters were removed from the data set. 5) *Stem (Porter)* operator performs stemming, identifying the stem/root of terms; 6) *Transform*

Cases operator transforms all the characters in the text into lower case. 7) *Generate nGrams* generates terms of length up to a maximum length  $n$ . The only parameter for this operator is *max length*, which was set equal to 1 in our example.

Following text processing, a human-involved activity is conducted, where a domain expert filters terms related with the domain, resulting in Term Set 1. This is followed by another human-involved activity, where the domain expert creates an ontology of terms and filters out only the terms in the ontology, resulting in Term Set 2. In the last step of this process, unsupervised machine learning methods are applied based on Term Sets 1 and 2.

A critical step in the developed and applied data mining framework is the construction of ontology for the domain of wind turbine accidents. While we could have constructed such ontology through mining of Wikipedia articles before the text processing itself, we opted for first observing the text processing outputs and then constructing the ontology based on the observed terms.

The ontology constructed in our study is illustrated in Figure 3. The ontology has four main categories, which contain the ontology terms as word stems (shown in bold and blue color), as their immediate children nodes. Some of the ontology terms contain terms branching under them. These branched terms are merged into the ontology term, with their term frequencies being summed up to compute the term frequency of the ontology term.

The unsupervised machine learning process (Figure 4), which was executed in Orange software (<https://orange.biolab.si/>), begins with the reading of source data, and verification of the data by inspecting it in a data table. Next, the attributes are selected for machine learning. Unsupervised learning is conducted for documents and terms, and is based on the distances among the data objects. The data objects are documents and terms, respectively, in the two analyses. In each analysis, distance map is visualized, multidimensional scaling (MDS) is carried out and the resulting visualization is analyzed, and hierarchical clustering is carried out and the resulting dendrogram visualization is analyzed.

#### A. Data

The most extensive data on the Internet about wind turbines accidents is made available by Caithness Windfarm Information Forum [5], which contains a list of more than 1,900 wind turbine accidents (1,400 at the time of data collection) in the UK and the world. However, most web links in this list to the news sources do not work properly, the original news article cannot be reached, and the data does not enable detailed analysis. Especially the inaccessibility to original news articles raised concerns from an academic point of view in the first year of our project. Similar websites, where lists of accidents are compiled, have the same problem, while containing much less data. After the first year of the project, we decided to focus on compiling our own dataset, also collecting the full text of accident news. This would result in a smaller, but verifiable dataset.

## V. ANALYSIS AND RESULTS

The accident news dataset in this study was collected over a 12-month period, scanning the Ebscohost and Lexis Nexis databases, and searching over the Internet through Google. More than 5,000 search results were scanned, more than 2000 were clicked, skimmed, and/or read, while only a small portion were found highly related. All the 1400+ accidents in the Caithness Farm dataset have been searched for over the Internet through the news titles. Eventually, 218 news, whose sources were verified, were found to directly report wind turbine accidents.

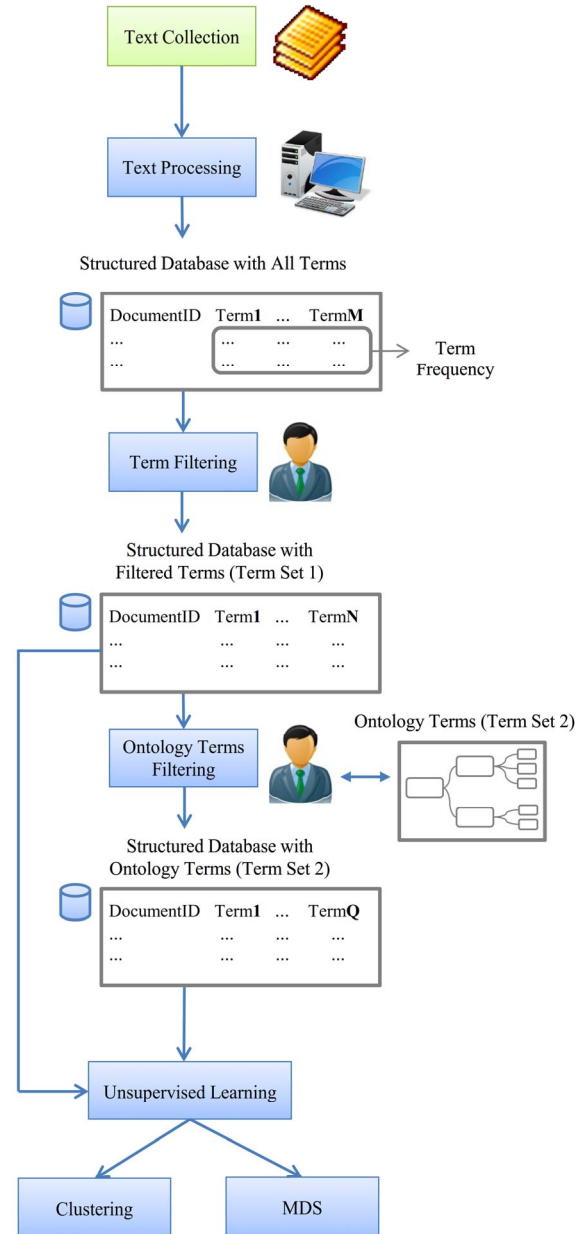


Fig. 2. The framework for the analysis of text collection of wind turbine accident news.

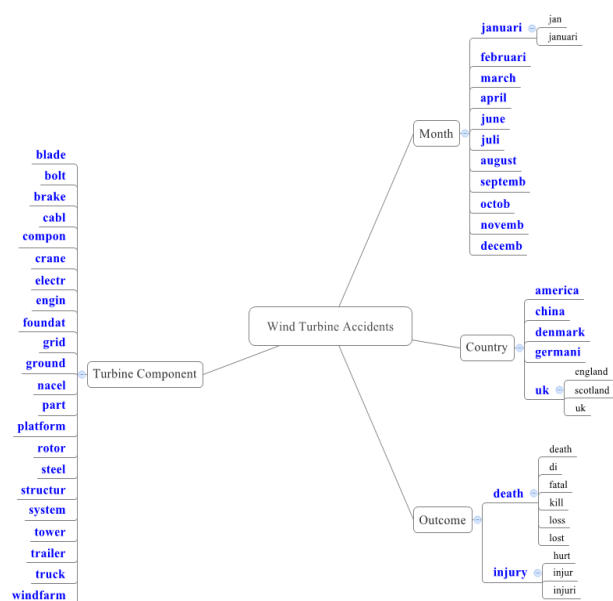


Fig. 3. The ontology developed based on the terms returned from text processing.

### A. Results

The first line of analysis in unsupervised learning is the analysis of the documents in the text collection. While unsupervised learning has been conducted using both databases with Term Sets 1 and 2, only the results for Term Set 2 are presented in this paper.

Figure 5 illustrates the result of MDS, where the documents are mapped based on the Spearman correlation between them. The correlation between two documents is computed based on the frequencies of terms in the documents, i.e., based on two vectors of length 40. Two documents are similar if they are firstly linked and then positioned close to each other on the MDS visualization. This analysis is particularly useful for benchmarking studies, where an organization is interested in finding accident news similar to its own case. One can observe two natural clusters of the documents, and the dense cluster contains many similar documents.

The second line of analysis in unsupervised learning is the analysis of the terms that appear in the text collection. We share in this paper only the results for the *ontology terms*, both due to space limitations, and the fact that more interesting and clear insights can be obtained in this case (compared to Term Set 1).

Figure 6 illustrates the result of MDS, where the ontology terms are mapped based on the correlations among them. The correlation between two documents is computed based on the frequencies, i.e., based on two vectors of length 218. Two terms are associated if they are linked and positioned close to each other on the MDS visualization. This analysis is especially useful for linking the terms under different categories of the constructed ontology.

For example, on the left side of visualization, the words *truck* and *trailer* (under the category *Turbine Component*) are tightly associated, as expected. An unexpected pattern, however, is the fact that the months July, September and (*juli*, *septemb*, *octob* under the category *Month*) are also associated with either one or two of these terms. This pattern suggests that transportation-related accidents may be taking place especially in these three months. The term *death* on the lower left corner of the visualization is especially important, and seems associated with the *foundation* and *crane* components, as well as the months of August and September. Deaths are also closely associated with China, indicating that the accidents regarding China report death more frequently than regarding any other country.

The lower right corner of the MDS visualization contains the terms *nacel*, *compon*, *bolt*, *march*, *april*, indicating the association between terms. This suggests that failure of components in the nacel (especially the bolt), have a tendency to fail in March and April.

The upper right corner of the visualization reveals another pattern: The accidents in Denmark frequently involve *blade* and *brake*, and blade accidents happen frequently in February and brake accidents happen frequently in December. Finally, the upper left corner of the visualization suggests that frequent accidents are observed in the UK in the month of November.

Figure 7 gives the dendrogram obtained through the hierarchical clustering of the ontology words. Terms under the branches with the same color are associated with each other. One cluster contains the terms *august*, *death*, *september*, and *crane*.

This cluster basically reiterates one of the insights that we obtained earlier yet, there are many other patterns and insights that can be observed, some of which are listed below:

- In Germany, foundation of the turbine fails frequently in December.
- In June and July, problems are frequently observed in the engine, cables, structure and tower.
- Accidents in China are associated with the electricity grid.
- Blade failures are most frequently observed in January and February.

The insights obtained above highlight not only the possible risks for each country, but also the timing in the year.

This type of information can be used by manufacturers in designing better components that can endure the conditions in those periods and countries. Engineering companies can use these insights in planning their preventive maintenance activities and developing better safety procedures. Insurance companies can plan their policies in the light of these discovered risk patterns, especially involving death, injury, as well as complete system failure. Governments also can use these insights, in establishing new and better regulations.



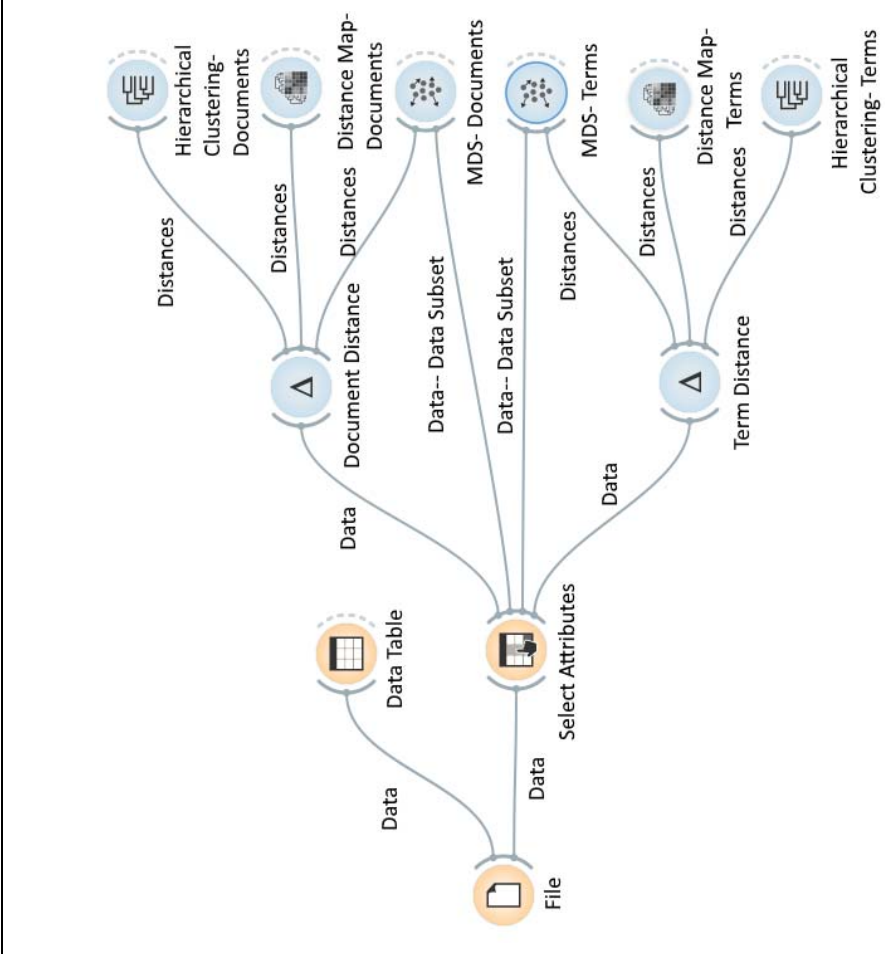


Fig. 4. The unsupervised learning process, consisting of the clustering and multidimensional scaling (MDS), applied on the structured database obtained through text processing.

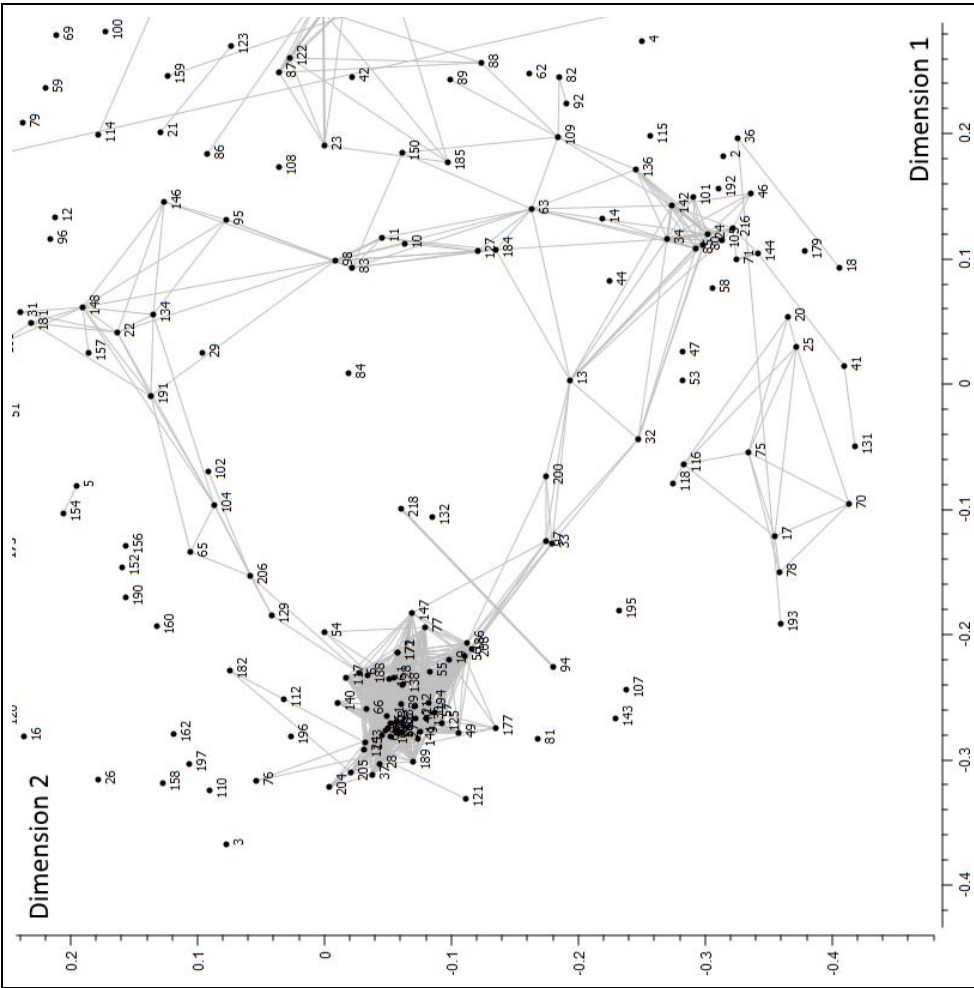


Fig. 5. Multidimensional scaling (MDS) visualization of documents in the text collection..



accident reports can help in the generation of significant new insights. Other research, from a methodological perspective, includes the automatic identification of documents that report particular outcomes, such as death and injuries by using supervised machine learning techniques such as ranking and classification. Furthermore, for collecting the data, techniques from information retrieval can be extensively used and adopted, possibly in interaction with the data mining techniques adopted.

#### ACKNOWLEDGMENT

The authors thank Sena Pakter, Soner Ulun, Dilara Naibi, Servet Büyükkuşoğlu, İnanc Arın, Kamil Çöllü, Mete Sevinç, and Byung-Geun Choi for collecting and/or cleaning the data for the study. The authors thank Çağrı Haksöz for his valuable comments that improved the paper. This research was partially funded by internal grant from Abu Dhabi University's Center of Excellence for Sustainable Business Processes.

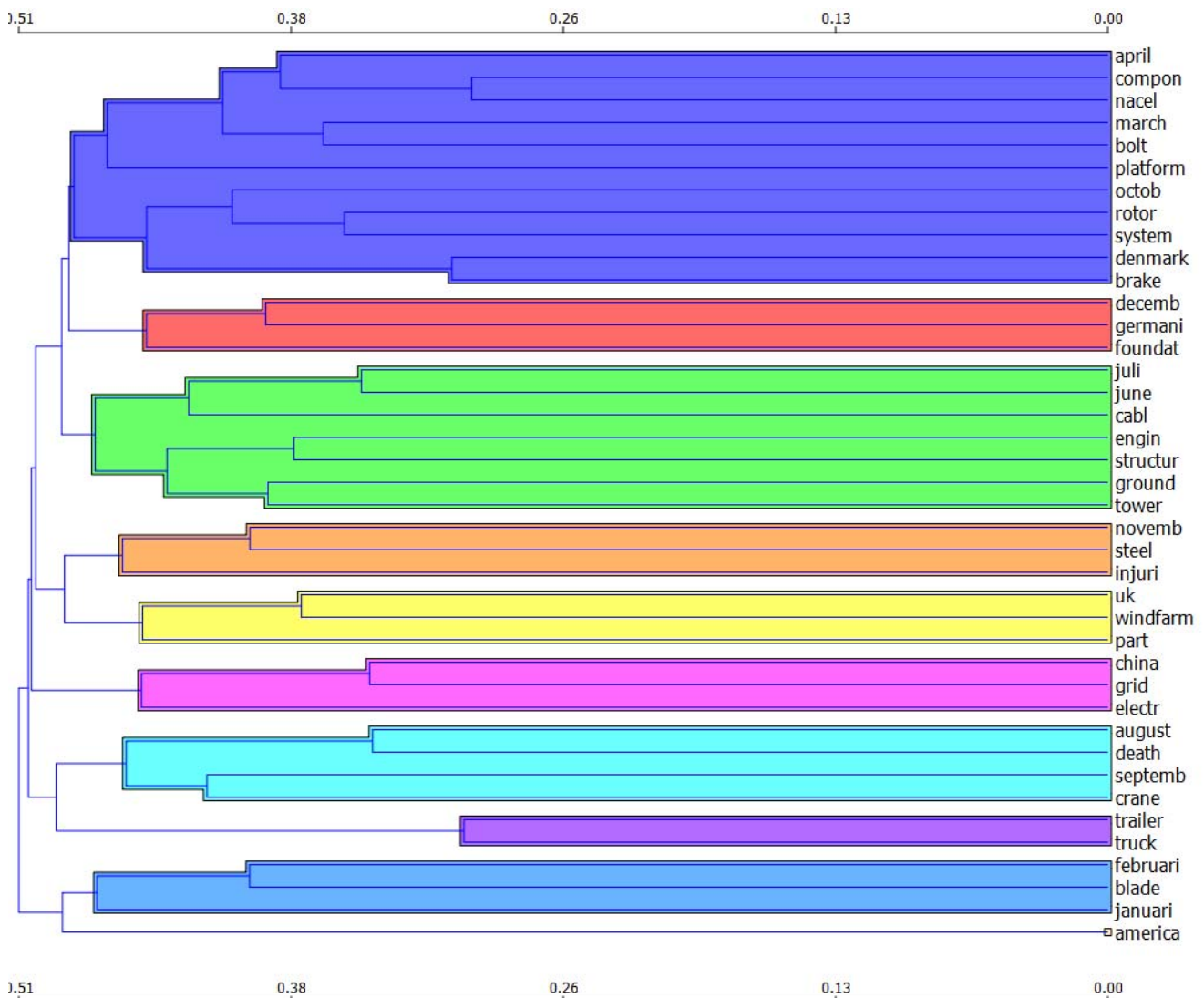


Fig. 7. Dendrogram of terms in the ontology terms set (Term Set 2), showing their hierarchical clustering.



## REFERENCES

- [1] World Energy Outlook 2013. International Energy Agency, 2014 [Online]. Available: <https://goo.gl/Ho8gj4>
- [2] GWEC. Global Wind Report: Annual Market Update 2016. Global Wind Energy Council, 2017 [Online]. Available: <http://gvec.net/global-figures/graphs/>
- [3] Department of Energy. Office of Energy Efficiency & Renewable Energy. Wind Energy Technologies Office. "The Inside of a Wind Turbine" [Online]. Available: <https://goo.gl/5s73kS>
- [4] S. Asian, X. Nie, "Coordination in supply chains with uncertain demand and disruption risks: Existence, analysis, and insights," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, issue. 9, pp.1139-1154, 2014.
- [5] Caithness Windfarm Information Forum. [Online]. Available: <http://www.caithnesswindfarms.co.uk/>
- [6] S. Asian, G. Ertek, C. Haksoz, S. Pakter, and S. Ulun. "Wind turbine accidents: A data mining study," *IEEE Syst. J.*, vol. 11, no. 3, pp. 1567-1578, Sept. 2017.
- [7] A.K. Choudhardy, P.I. Oluikpe, J.A. Harding, P.M. Carillo, "The needs and benefits of text mining applications on post-project reviews," *Computers in Industry*, vol. 60, pp. 728-740, 2009.
- [8] D.G. Rajpathak, "An ontology based text mining system for knowledge discovery from the diagnosis data in the automotive domain," *Computers in Industry*, vol. 64, pp. 565-580, 2013.
- [9] W.T. Liew, A. Adhitya, R. Srinivasan, "Sustainability trends in the process industries: A text mining-based analysis," *Computers in Industry*, vol. 65, no. 3, pp. 393-400, April 2014. <http://dx.doi.org/10.1016/j.compind.2014.01.004>.
- [10] G. Dawelbait, T. Mezher, W.L. Woon, and A. Henschel, "Taxonomy based trend discovery of renewable energy technologies in desalination and power generation", in *PICMET: Proceedings of the Technology Management for Global Economic Growth*, Phuket, pp. 1-8, 2006.
- [11] K. Lee and S. Lee, "Patterns of technological innovation and evolution in the energy sector: a patent-based approach," *Energy Policy*, vol. 59, pp. 415-432, Aug. 2013.
- [12] S. E. Cullen, *Alternative energy powers up*, World IP Today Research Report, Thomson Reuters. [Online] Available: <https://goo.gl/LCaVKj>
- [13] H. Park, J.J. Ree, and K. Kim, "Identification of promising patents for technology transfers using TRIZ evolution trends," *Expert Syst. with Appl.*, vol. 40, pp. 736-743, Feb. 2013.
- [14] D. Kucuk, Y. Arslan, "Semi-automatic construction of a domain ontology for wind energy using Wikipedia articles," *Renewable Energy*, vol. 62, pp. 484-489, 2014.
- [15] A. Cristina, B. Garcia, I.N. Ferraz, and F. Pinto, "The role of domain ontology in text mining applications: the ADDMiner project", in *ICDMW'06: Proceedings of the Sixth IEEE International Conf. on Data Mining-Workshops*, Hong Kong, pp. 34-38, 2006.
- [16] W. Jin, R.K. Srihari, and H.H. Ho, "Improving knowledge discovery in document collections through combining text retrieval and link analysis techniques," in *Proceedings of the 7th IEEE International Conference on Data Mining*, pp. 193-202, 2007.
- [17] T.L. Bunn, S. Slavova, and L. Hall, "Narrative text analysis of Kentucky tractor fatality reports," *Accidents Anal. Prev.*, vol. 40, no. 2, pp. 419-425, Mar. 2008.
- [18] W.-S. Tseng, H. Hguyen, J. Liebowitz, and W. Agresti, "Distractions and motor vehicle accidents," *Ind. Manag. Data Syst.*, vol. 105, no. 9, pp. 1188-1205, Dec. 2005.
- [19] A. Hibino, Y. Niwa, "Graphical representation of nuclear incidents/accidents by associating network in nuclear technical communication," *J. Nucl. Sci. Technol.*, vol. 45, no. 5, pp. 369-377, May. 2008.
- [20] J. Han, M. Kamber, and J. Pei, *Data Mining: concepts and techniques*, 3rd ed., Morgan Kaufmann, 2011.
- [21] T.H. Davenport and J.G. Harris, *Competing on analytics: the new science of winning*, Boston, Mass.: Harvard Business School Press, 2007, ISBN 978-1-4221-0332-6.
- [22] V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Eamon Dolan/Mariner Books, 2014.
- [23] G. Ertek, X. Chi, A.N. Zhang, "A framework for mining RFID data from schedule-based systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, issue 11, pp. 2967-2984, 2017.
- [24] G. Ertek, G. Tokdemir, M. Sevinç, M.M. Tunç, "New knowledge in strategic management through visually mining semantic networks," *Information Systems Frontiers*, vol. 19, issue 1, pp.165-185, 2017.
- [25] E.N. Çinicioglu, G. Ertek, D. Demirer, H.E. Yörük, "A framework for automated association mining over multiple databases," in 2011 *International Symposium on Innovations in Intelligent Systems and Applications (INISTA)*, pp. 79-85, IEEE, 2011.
- [26] G. Ertek, A. Demiriz, F. Cakmak, "Linking behavioral patterns to personal attributes through data re-mining," in *Behavior Computing*, pp. 197-214. Springer, London, 2012.
- [27] G. Ertek, M.M. Tunc, "Re-mining association mining results through visualization, data envelopment analysis, and decision trees," *Computational Intelligence Systems in Industrial Engineering*, pp.601-622, Springer, London, 2012.
- [28] I. Borg, P. Groenen. *Modern Multidimensional Scaling: theory and applications* (2nd ed.). New York: Springer-Verlag. ISBN 0-387-94845-7 (2005).
- [29] G. Miner, J. Elder, T. Hill, R. Nisbet, D. Delen. *Fast, Practical text mining and statistical analysis for non-structured text data applications*, Academic Press, 2012.
- [30] G. Ertek, D. Tapucu, and D. I. Arın, "Text mining with RapidMiner," In: Markus Hofmann, Ralf Klinkenberg (Eds.), *RapidMiner: Data Mining Use Cases and Business Analytics Applications*, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, Chapman and Hall/CRC, 2013.
- [31] O. Corcho, M. Fernández-López, A. Gómez-Pérez, "Ontological Engineering: Principles, Methods, Tools and Languages," In: Calero, C., Ruiz, F., Piattini, M., 2010. *Ontologies for Software Engineering and Software Technology*. Springer, 2010.
- [32] R. Poli, L. Obrst, "The Interplay Between Ontology as Categorical Analysis and Ontology as Technology," In: Poli, R., Healy, M., Kameas, A., *Theory and Applications of Ontology: Computer Applications*, 2010.