Text mining of Scottish post-emergency and training exercise debrief reports

Author(s):
Oliver Gunawan and Timothy Aldridge

Report Number:
MSU/2018/04

Text mining of Scottish post-emergency and training exercise debrief reports

Report approved by:	Timothy Aldridge MSc (Data Management and Spatial Analysis Team)
Report authorised for issue to Scotland's National Centre for Resilience by:	Helen Balmforth PhD (Head of Centre for Data Analytics)
Date of Issue:	15/11/2018
Lead Author:	Oliver Gunawan PhD (Data Management and Spatial Analysis Team)
Contributing Author(s):	Timothy Aldridge MSc (Data Management and Spatial Analysis Team)
Customer:	Scotland's National Centre for Resilience
Technical Reviewer(s):	Derek Morgan PhD
Editorial Reviewer:	Derek Morgan PhD
Project number:	PE06449

Disclaimer:

This report and the work it describes were undertaken by the Science Division of the Health and Safety Executive under contract to The National Centre of Resilience. Its contents do not supersede current Health and Safety Executive policy or guidance.

Acknowledgements

This research was funded by the National Centre for Resilience. The authors wish to thank colleagues from the Scottish Resilience Development Service, and the three Regional Resilience Partnerships covering Scotland (North of Scotland, East of Scotland and West of Scotland) for sharing their knowledge of the Scottish Resilience community and their assistance in data collection.

EXECUTIVE SUMMARY

INTRODUCTION

Anecdotal evidence suggests that as much as 80% of an organisation's data are in unstructured formats, such as paragraphs of text, and that this is projected to rise (Taylor, 2018). This means that the majority of data held cannot currently be analysed via traditional methods. Instead, analytical tools such as text mining can be used to discover and extract interesting, non-trivial knowledge from free or unstructured text.

To meet priorities outlined by the Sendai Framework for Disaster Risk Reduction (UNISDR, 2015), countries are encouraged to systematically evaluate, record, share and publically account for disaster losses. This information may be recorded in structured formats such as numerical spreadsheets, but may also be stored as unstructured data in photos, videos, social media and written reports. This is an untapped resource that could hold a wealth of valuable local knowledge about specific civil emergencies and their consequences. Analysis of these data could have benefits for post-disaster compensation schemes, disaster loss reporting, understanding the efficacy of response activities, and the development and validation of risk models.

OBJECTIVES

The Health and Safety Executive (HSE), as part of the Natural Hazards Partnership (NHP), were commissioned by the Scottish National Centre for Resilience (NCR) to explore the value of text mining analysis for extracting impact information from post-emergency and training exercise debrief reports written by Scottish Regional Resilience Partnerships. The aims of the work were explored through four objectives:

1. Extract sentences containing impact information.
2. Analyse words that appear commonly, and their associations with other common words.
3. Extract location information to map the impacts of the hazard.
4. Explore sentiment analysis.

MAIN FINDINGS

Post-event reports were collected from Resilience Direct: an online private network for sharing information to improve preparation, planning and response to civil emergencies. With the assistance of colleagues at the NCR, the Scottish Resilience Development Service, and the three Regional Resilience Partnerships covering Scotland (North of Scotland, East of Scotland and West of Scotland), an online search retrieved 74 valid incident and exercise debrief reports from between 2010 and 2018. Each report was split into two blocks of text to focus analysis. These focussed on background information and lessons learned.

A method was identified for extracting sentences containing impact information from the free text with 85% accuracy. This analysis could be used to compile a list of impacts that have arisen from civil emergencies that could then be added to current stores of disaster loss data to improve our understanding of mitigation, response and forecasting. Location extraction was also demonstrated to enable mapping of impacts, placing the findings in a geographical and visual context which may be useful for further analysis. In both cases, consideration of more sophisticated text mining techniques may improve the accuracy of results. In particular, exploration of Natural Language Processing as a way of acknowledging the meaning and context of words in a sentence may improve the extraction of relevant features.

The most frequently appearing words in the background sections of the body of reports emphasise the role of the police in most emergencies, but also emphasise time (Figure ES1 contains the top 100

most frequently used words). For the lessons learned sections of the text, many of the most common words were the same as in the background text, although there was more emphasis on different partners and their activities. A method of clustering groups of words that commonly appear together was demonstrated. This analysis could be used to classify the reports into different themes.

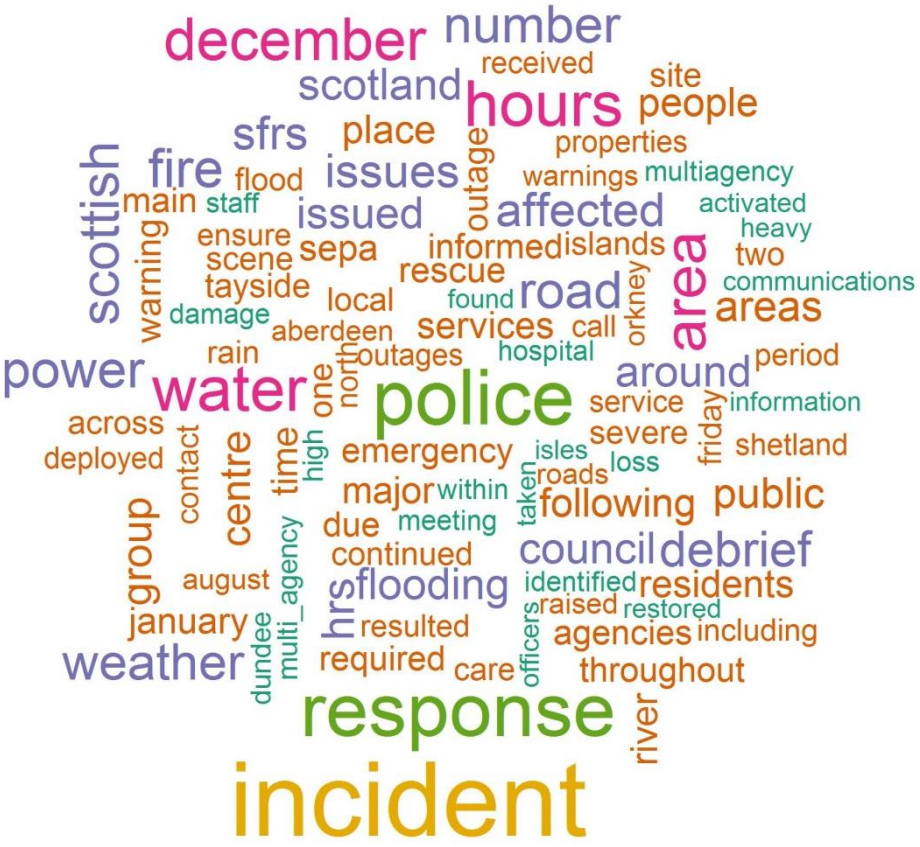


Figure ES1 Word clouds for the 100 most frequently occurring words in the background corpus. Colours and size have been used to clarify word frequency.

Sentiment analysis provides insight into the emotional content of a piece of text. The reports are found to be largely positive in their composition of lessons learned (an average of 78% positive words for incident reports compared to negative words). This report also demonstrates how eight different sentiments (anger, anticipation, disgust, fear, joy, sadness, surprise and trust) contribute. Trust is found to be the most abundant sentiment, followed by anticipation, fear and surprise. There are no discernible patterns over time, but it is possible to characterise reports based on patterns of sentiments. Five clusters of sentiments are identified although further analysis could improve cluster definition.

CONCLUSIONS AND RECOMMENDATIONS

- Text mining techniques, as demonstrated in this report, can be applied to a range of different documents to extract value from unstructured data. Examples of potential use include greater understanding and summarisation of user feedback in relation to policy decisions or public services.

- Extraction of impact-related sentences can directly contribute to global initiatives on disaster risk reduction. Development of the model identified in this report may focus on Natural Language Processing techniques as a method to improve accuracy, although the requirements of the end user must be considered to ensure that the analysis meets the final objectives.

- Integration of text mining techniques may enhance value and enable the characterisation of hazards and responses. Integration of other machine learning techniques may also enable predictive modelling.

- Larger sample sizes of documents would be beneficial for future analyses and more robust scientific results.

CONTENTS

1 INTRODUCTION

1.1 INTRODUCTION

Civil emergency risks include natural hazards such as flooding and severe weather, industrial accidents such as chemical and oil spills, and hazards such as domestic fires, airport incidents and major road traffic accidents. According to the Civil Contingencies Act (2004), in Scotland, organisations involved in civil emergency response include the emergency services (Police, Fire Service, Ambulance), health boards, local authorities, the Scottish Environment Protection Agency, the Maritime and Coastguard Agency, Transport Scotland, other asset owners including energy, utilities and communications, and volunteer organisations.

Co-ordination of multi-agency activities in response to civil emergencies is the remit of the Scottish Government. This is based on three Regional Resilience Partnerships (RRPs), which split the country into North, West and East regions. The RRPs are composed of twelve Local Resilience Partnerships (LRPs). The partnerships bring together all relevant organisations to develop effective approaches for emergency management (Ready Scotland, 2018). Partnerships are also responsible for evaluating multi-agency emergency responses for recent incidents and training exercises using debrief reports. Reports provide a record of the incidents and events, with a particular focus on the lessons that can be learned for future response and business continuity. An individual report provides valuable insight into an incident or event, but summarising over multiple reports, for example to identify underlying trends and patterns or categorisations, can be time-consuming if the report archive is large. Analytical text mining approaches offer a means of extracting and cataloguing information held in reports, which is scalable to large collections of unstructured material.

Text mining is the discovery and extraction of interesting, non-trivial knowledge from free or unstructured text (Kao and Poteet, 2007). This analytical technique is gaining in popularity and is used widely across academia, business and government. By programming a machine to automatically 'read' and analyse whole repositories of texts, text mining has been used for a range of applications including monitoring online article content for national security (Zanasi, 2009), identifying previously hidden linkages in scientific developments (Percha *et al*., 2011), and evaluating emotions in customer feedback to improve public services (Heron, 2016). Text mining presents an opportunity for Scottish Government and their stakeholders, who may own or have access to large volumes of reports, articles, online reviews and feedback, to gain fresh insights from the information that they hold. This report describes text mining, its applications and how it can be used in the analysis of Scottish data sources. The specific focus of this analysis will be post-emergency multi-agency response debrief reports composed by RRPs.

1.2 BACKGROUND

Collecting and managing disaster loss data allows stakeholders to evaluate disaster policies, identify driving factors in loss trends and enable the generation of early warning systems (Bouwer *et al*. 2007). The importance of this activity has been raised at the global level, where the Sendai Framework for Disaster Risk Reduction: Understanding Disaster Risk (UNISDR, 2015), encourages countries to systematically evaluate, record, share and publically account for disaster losses. In a research report for the European Commission, De Groeve *et al.* (2014) cite four general areas, where improving collection of observed impacts can provide benefits to different stakeholders:

1. Businesses and insurance companies are interested in calculating disaster loss compensation. This is a major contributor in assessing the capacity for people, businesses and communities to recover.
2. At the national scale, governments are interested in disaster loss accounting, particularly from large disasters. This allows for a better understanding of how societal loss trends change over time, between hazard types and across nations.
3. Regional governments and their stakeholders are interested in the disaster forensics after a specific emergency to improve future response. This involves a more detailed analysis of specific actions and impacts.
4. Academics and forecasters are interested in improving Early Warning Systems to improve the type of information that responders receive when confronted with an impending emergency. Observed impact data are a key input in calibrating and evaluating the performance of such systems.

Disaster loss data collection methods are diverse. Common approaches include household surveys, site visits, social media, photographs, videos, and assorted print and online reports. These datasets must be intelligently processed and interpreted to account for incomplete coverage, biased reporting and differing levels of detail. The outputs of these data collection sources can be categorised as structured or unstructured. Structured data are well-organised (for example, a spreadsheet of numbers ordered into rows and columns). They can be easily ordered, processed and analysed using computer models. Unstructured data don't conform to an identifiable configuration. They are often built on the human language (for example, a paragraph of text). To analyse unstructured data with computer models, the data must first be codified using specialised techniques. In particular, there has been a recent upsurge in interest in text mining (Figure 1) to extract information from unstructured text. Anecdotal evidence suggests that as much as 80% of an organisation's data are in unstructured formats and this percentage is projected to rise (Spire Technologies, 2016; Taylor, 2018).

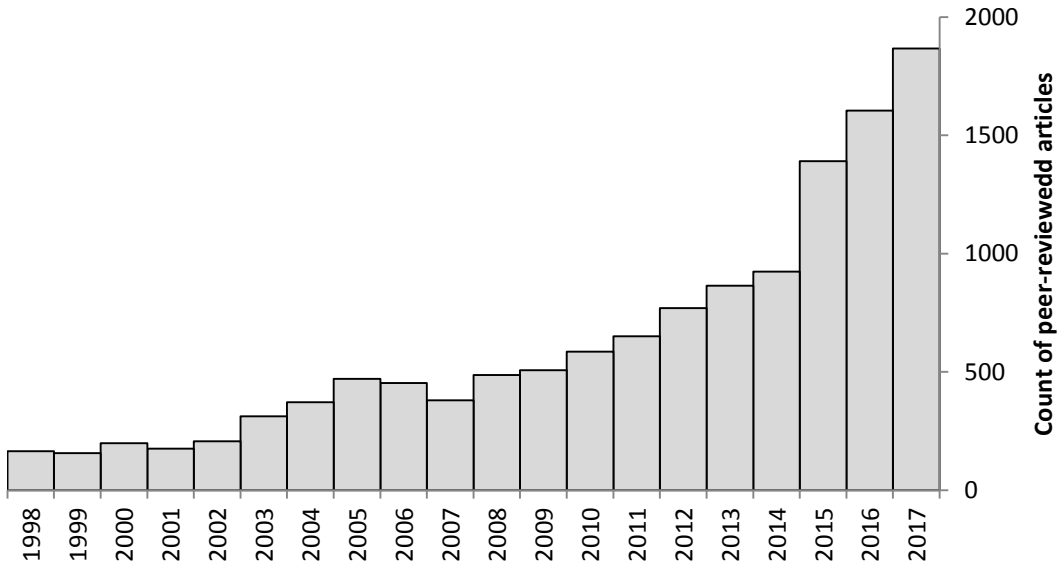


Figure 1 The number of peer-reviewed scientific journal articles featuring the terms 'Text Mining' or 'Natural Language Processing' in the Web of Science website, 1998-2017.

The process of text mining usually involves five commonly agreed steps (Figure 2), described as:

1. **Gather data**. This may be from websites, emails, document files etc. This exercise may be automated if the sources have been organised in a specific format.
2. **Pre-process**. The pre-processing step cleans the raw data. This includes removing unwanted content from text such as punctuation or white spaces. It also includes tokenisation. Tokenisation is the process of codifying the text such that individual words or groups of words become data objects that can then be identified and read by a machine.
3. **Index**. Indexing the text allows data to be recalled more efficiently by machines. A typical example of this is the creation of document-term matrices, where the documents in a collection (or 'corpus') are arranged by the words that are contained in them.
4. **Mining**. Data Mining involves a range of text mining techniques for extracting different types of information from the text.
5. **Analysis**. Finally, the raw results mined from the text need to be evaluated and visualised so that they can be properly interpreted to satisfy the research question.

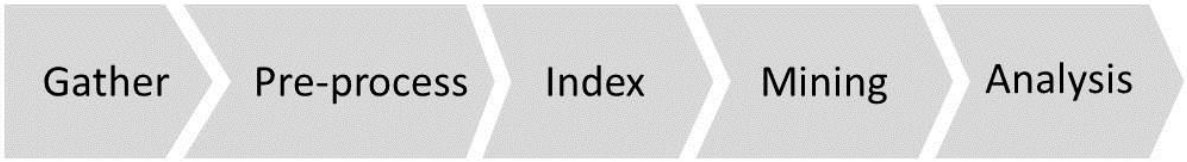


Figure 2 Common steps in text mining. Adapted from Chang *et al.* (2018).

Text mining techniques include information retrieval and extraction, text summarisation, social media mining, document classification, and Natural Language Processing (NLP) (Allahyari *et al.* 2017). Some relevant references for text mining are listed: Ngai and Lee (2016) conducted a review of text mining in relation to policy making, covering a range of governance areas, Sakaki *et al.* (2012) mined geo-located tweets to detect potential earthquakes, using a dictionary of key terms, Chau *et al.* (2002) extracted entities (names, addresses etc.) from narrative police reports for more efficient crime investigation, and Torii *et al.* (2011) classified online articles using machine learning techniques to automate the identification of new disease outbreak information.

Beyond a simple analysis of word patterns, NLP helps a machine 'read' text by summarising, tagging, extracting, recognising and clustering terms within the text. NLP requires a consistent knowledge base to work from such as a thesaurus, ontology or set of rules. NLP may use techniques such as linguistic analysis, speech tagging, artificial intelligence and sentiment analysis. In government, this is most commonly used to extract and interpret public opinion on government decisions (Stylios *et al.* 2010), and to summarise those opinions to produce a 'the voice of the people' (Evangelopoulos and Visinescu, 2012). This approach has helped decision makers evaluate policies in relation to their initial objectives.

Figure 1 highlights the growing popularity of text mining. Current research is developing more sophisticated techniques. For example, Tobback *et al.* (2018) explored the potential for analysing sentiments in news articles to predict uncertainty surrounding economic policy, while Rezaeian *et al.* (2017) reviewed the potential for using text mining with lifecycle analysis and clustering analysis with published papers to predict scientific knowledge gaps for scientific foresight. These examples demonstrate the power of techniques such as text-mining, particularly as conventional government

services are increasingly moving towards online E-government services, which have the capacity to reduce the cost of interaction with the government.

1.3 AIMS AND OBJECTIVES

The primary aim of this report is to identify methods for extracting information on impacts in the report archive. The secondary aim is to demonstrate the wider value that text mining analysis can add to the evaluation of a collection (or corpus) of unstructured texts. The analysis will consider a range of approaches and will evaluate where they have been most effective, and how they might be used for further applications. The objectives designed to achieve the aim are listed below, based on the four methodological approaches taken:

5. **Extract sentences containing impact information.** This is the primary objective using feature extraction. The results of this analysis may contribute directly to international agreements on disaster loss recording. It may also provide a valuable new source of observed impact data for disaster loss compensation and accounting, disaster forensics and risk modelling (Gunawan and Aldridge, 2018).
6. **Analyse words that appear commonly, and their associations with other words.** This objective uses analysis of word frequencies and associations to provide information about the report content, and may identify common response partners and activities.
7. **Extract location information to map the impacts of the hazard.** This objective uses feature extraction to provide additional data for mapping hazard and impact extents and response activities.
8. **Explore sentiment analysis**. This may provide insight into how different hazards and responses are considered via the emotional content of the text. This can include differing reactions over time, emotional responses to different impacts and how future responses are thought of.

1.4 DATA GATHERING

The data used in this research were a series of post-emergency and training exercise debrief reports. These reports are written by RRPs after an emergency that has required a multi-agency response. This includes natural and non-natural hazards and training exercises (both field and desktop). The primary purpose of these reports is to review the circumstances and consequences of the emergency and to determine lessons and recommendations for future learning.

Identification of source material for text mining analysis proved challenging as they are created by different organisations, in different formats, across Scotland. Consequently, data collection was only possible with the co-operation of colleagues at the National Centre of Resilience (NCR), the Scottish Resilience Development Service, and the three RRPs covering Scotland (North of Scotland, East of Scotland and West of Scotland). With their assistance, an online search retrieved 93 reports (between 2010 and 2018) from file stores owned by LRPs and RRPS in Resilience Direct: an online private network for sharing information to improve preparation, planning and response to civil emergencies. Of the 93 reports collected, 19 reports were not considered appropriate due to content or format. 74 reports were retained for analysis (Table 1). The documents were in either Word Document (.docx) or PDF (.pdf) formats.

Table 1 Distribution of incident and exercise report content collected for analysis.

Region	Natural hazard incidents	Natural hazard exercises	Other hazard incidents	Other hazard exercises	**Total incidents**	**Total exercises**	**Grand Total**
East of Scotland	1	0	9	10	**10**	**10**	**20**
North of Scotland	15	0	32	4	**47**	**4**	**51**
West of Scotland	1	1	0	1	**1**	**2**	**3**
Total	**17**	**1**	**41**	**15**	**58**	**16**	**74**

1.5 PRE-PROCESSING

1.5.1 Document segmentation

Initial scoping of the source material suggested that each document should be segmented into component text files to provide a focus for the text mining. The reports did not conform to a common technical structure. This meant that manual separation was required to split raw data into separate text files (.txt). The text files were segmented and categorised into 2 broad themes:

1. Background: These files included the background, overview and summary sections of the debrief reports. These sections contain information on the incident or exercise including the description of the event and information on consequences. A separate background file was produced for each report.

2. Lessons learned: All debrief reports included some information on lessons to be learned. In some reports, this information was grouped into a single paragraph. In other reports, this information was separated into themed sections. Where possible, separate text files were created from themed sections for each report to reflect different types of lessons learned (Table 2). Where named sections did not exactly match the themes in Table 2, the text was assigned the closest theme based on manual evaluation. Where the theme was not obvious, the text was added to the 'Other' theme.

Table 2 Themed 'lessons learned' with code for file naming.

Theme	**Code**
Activation	AC
Command, control and coordination	CC
Communications	CO
Multi-agency working	MA
Plans	PL
Resources	RE
Logistics	LO
People	PE
Environment	EN
Response	RS
Other	OT

Each component text file was named such that reference to the parent document could be made and components could be merged as necessary during analysis (Table 3). Where lessons were not categorised, no *Content 2* code was added.

Table 3 File naming convention: (16 characters)

Unique Report ID	**Exercise or Incident**	**Date of debrief**	**Content1 (Background or Lesson)**	**Content 2 (Lesson theme)**
SGXXX	E/I	DDMMYY	BG or LL	See Table 2

1.5.2 Text pre-processing

The texts were compiled into corpuses by content: background and lessons learned (all lesson themes combined into a single file per report) using the statistical software package, R (R Core Team, 2017). An additional set of corpuses was created for each lesson theme (based on categories in Table 2). The text in each corpus was formatted to improve interpretation. This is a common procedure in text mining, although the formatting differs between analysis types dependant on required outcomes.

The process for each type of analysis conducted in this report is given in Table 4. The analyses are described in more detail in the following section.

Table 4 Corpus pre-processing steps for each analysis conducted in the report.

	Word frequencies and associations	Impact extraction / Location extraction	Sentiment analysis
Removal of punctuation	✓		
Removal of numbers	✓		
Removal of special characters	✓	✓	✓
Removal of empty lines	✓	✓	✓
Removal of new lines (i.e. concatenating the entire text into a single string of words separated by white space	✓	✓	✓
Change all to lower case	✓		
Removal of stop words (common words of little analytical value (the, is, and etc.).	✓		✓
Removal of unnecessary white spaces	✓	✓	✓
Combining words that should stay together and not be analysed separately: • Fire rescue service • Scottish ambulance service • Emergency services • Action plan • Flood alert • Local emergency liaison • Local resilience partnership • Multi agency • Police incident officer	✓		

2 METHODS

2.1 IMPACT EXTRACTION

The primary objective of this report is to understand the feasibility of extracting impact information from debrief reports. Consequently, this analysis extracts sentences that contain key words related to the impacts or consequences of an emergency. To conduct this analysis, punctuation was retained so that sentences could be identified as tokens (a measureable unit in the data). Sentences were then compared against a list of key terms related to impact or loss information:

calls|affected|death|injur|casualt|fatalit|road|block|disrupt|damage|closed|evacuat|rescue|cost

Sentences containing the terms were extracted and placed in a separate list. To avoid capturing sentences that contain information about warnings or likelihoods, the extracted sentences were screened for the following terms which infer probable or forecast outcomes rather than actual losses:

may|possib|likel|probab|forecast|potential

The output was a list of sentences containing impact information by report. These were collated into a single spreadsheet alongside a reference that identifies the relevant report.

To evaluate the accuracy of the impact extraction, all sentences passed through the model were manually assessed and flagged if they contained impact information. This was compared against the results of the text mining model to identify:

- True positives (sentences correctly identified as containing impact information)
- True negatives (sentences correctly identified as not containing impact information)
- False positives (sentences that were identified as containing impacts, but actually not)
- False negatives (sentences that contained impact information, but that were missed by the model).

2.2 LOCATION EXTRACTION

Mapping regional and local impacts provides a useful picture of spatial distributions. Local impact hotspots may be easier to identify with a visual aid. Data for location extraction was prepared in the same way as for impact extraction except that single words and n-grams were used as tokens. N-grams combine individual words with the next *n* number of consecutive words to use as tokens. In this analysis, bi-grams (two-word tokens), tri-grams (three-word tokens) and four-word tokens were analysed. The tokenised text was compared against a list of location names created from the Ordnance Survey (OS) Open Names dataset (Ordnance Survey, 2018). This includes *City, Town, Village, Named Places* and *Named Roads* in Scotland. Both English and Gaelic translations were included. Unitary Authority names were added to capture larger areas. Finally, A-roads and Motorways were added. Tokens that matched the list of locations were extracted. These were manually mapped using ArcGIS software and OS datasets for a single example.

This analysis was validated by referring each mapped location to the raw text. Where an extracted location was an impact site, the map was shaded in red. Where the extracted location was included in the text for a different purpose (e.g. the location of response organisations / duplicate references etc.), the map was shaded blue.

2.3 WORD FREQUENCIES AND ASSOCIATIONS

This analysis assesses each text file as a 'bag of words'. Each word (or group of words) becomes a token in the body of the text in question. A document-term matrix (DTM) was created from the corpuses. This is a common starting point as a method of codifying tokens. A DTM contains word frequencies for each document in the corpus. Rows represent documents while columns represent tokens. It is then possible to identify the most frequently used words as well as trends and associations between pairings of words across the corpus. In this analysis, bi-grams (two-word tokens), tri-grams (three-word tokens) and five-word tokens were analysed. DTMs were created for both background and lessons learned (combined) texts.

Associations between words were assessed in a number of ways. The simplest method correlated the pattern of occurrence of a given word against the patterns of all other words in the DTM. Higher correlations (based on Pearson's *r* correlation scores) indicate that the paired words appear in similar sets of texts within the corpus.

The more sophisticated approach of hierarchical clustering was also demonstrated. Hierarchical clustering groups similar terms together. The algorithm starts with a single cluster containing all terms. This is then split according to the characteristics of the data (in this case, where words appear in the corpus). Stronger, more easily identifiable splits are made first, before moving to more subtle differences. This creates a hierarchical tree or 'dendrogram' of progressively smaller clusters. A line is then drawn across the tree at a selected height to identify the desired number of final clusters.

Word association analysis works better when uncommon terms are removed from the DTM because these terms create a lot of 0s in the matrix. Uncommon terms were removed from the DTM to produce a 'sparse DTM'. After experimentation, all words that appeared in fewer than 50% of the lessons learned DTM, and in fewer than 75% of the texts for the background DTM, were removed. This ensures that the least commonly used terms do not adversely affect analysis. For the purposes of this research, Ward's method of hierarchical clustering (Everitt *et al.* 2001) was applied to both the background and combined lessons learned corpuses as default settings in R.

2.4 SENTIMENT ANALYSIS

Sentiment analysis applies Natural Language Processing via the use of established, predefined dictionaries (Liu, 2010). The dictionaries include information relating to different types of sentiment, allowing classification of other texts. At the most basic, this relates to *positive* and *negative* sentiment. For this research, the combined lessons learned texts offer a useful context for sentiment analysis as they may provide insight into the general feeling of how the response was conducted.

The Syuzhet R package (Jockers, 2017) was used and specifically, the National Research Council Canada (NRC) emotional lexicon[1], which relates words to eight sentiments (*anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise* and *trust*), as well as general *positive*/*negative* classifications. The Syuzhet package was primarily designed for analysis of fictional material and therefore is not optimised for non-fiction text. However, sentiment analysis has been used to assess or predict consumer sentiments from online feedback (Stylios *et al.* 2010; Bai, 2011; Evangelopoulos and Visinescu, 2012) and the concept of the eight sentiments listed above is well established in the literature (Plutchik, 1980).

[1]NRC Word-Emotion Association Lexicon , http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

Each text in the corpus was analysed using the NRC lexicon. The individual lessons learned corpuses were not pursued due to smaller individual text lengths. Results by report were compiled, with the debrief date and mean average proportions of each sentiment recorded. This data was then used to assess trends over time for the entire dataset, and for natural hazards only.

3 RESULTS

3.1 IMPACT EXTRACTION

The results of the impact extraction are a list of sentences for each report. An example output is provided in Figure 3 below. The example demonstrates how a relatively primitive model of feature extraction can still perform well, although there are still areas for improvement. For example, *flood* was not used as a term because it describes the hazard rather than the consequences. However, each of these sentences contains the word. Additionally, sentence 3 indicates that *no* properties were damaged and sentence 4 states that *no* action was taken in fourteen out of over 20 emergency service calls. Further analysis could join other attribute data associated with the case study and compile data across the corpus into a single spreadsheet.

1 “Numerous **road**s including the A90 and A92 arterial routes were **affected** by floodwater as was the Dundee to Aberdeen railway line."

2 "The flooding was widespread and **affected** domestic and commercial premises in Glamis Trinity Charleston Village Rosemill Strathmartine Montrose Monifieth and Brechin."

3 "One of the areas most severely **affected** by floodwater was River Street in Brechin where although the River South Esk breached its banks no properties were **damage**d."

4 "Tayside Fire and Rescue received in excess of twenty **calls** for assistance with flood related incidents however no action was taken in relation to fourteen of these **calls** as no property was threatened."

5 "Tayside Police communications centre recorded almost fifty incidents mainly due to floodwater affecting **road**s."

Figure 3 Example outputs of the impact extraction analysis. Key words used in the extraction are highlighted in **red**.

The overall accuracy of the impact extraction analysis for the entire corpus (Table 13) is 0.85 (85% of sentences were correctly identified as either containing or not containing impact information). These numbers could be improved by developing additional rules around sentence identification. The diversity of hazard and impact types also presented challenges for simplistic extraction.

Table 5 Contingency table for evaluation of impact extraction analysis accuracy.

		Manual Assessment		
		Impact	Not impact	Total
Model Assessment	Impact	**86**	32	118
	Not impact	47	**353**	400
	Total	133	385	518
		65%	92%	Overall Accuracy =85%

When considering only natural hazards (Table 6), the overall accuracy is 84%. The underlying accuracies for sentences containing impacts are similar to those in Table 5.

Table 6 Contingency table for evaluation of natural hazard impact extraction analysis accuracy.

		Manual Assessment		
		Impact	Not impact	Total
Model Assessment	Impact	**45**	12	57
	Not impact	27	**161**	188
	Total	72	173	245
		63%	93%	Overall Accuracy =84%

Taking this analysis one step further, reports could be ranked by severity based on the number of impact sentences extracted. In this case, the top two reports in the corpus relate to flooding with 8 and 7 impact sentences identified in these reports. Impacts listed are diverse and include road closures, rescue and evacuation. The third most severe report is a fire (6 impact sentences) containing evacuations, casualties and damage to property.

3.2 LOCATION EXTRACTION

Figure 4 presents location mapping for a flooding incident based on information extracted from a single report. Regions and lines shaded in red represent locations that have been correctly identified as impact sites. Regions and lines shaded in blue represent locations that have been mentioned in the text for erroneous reasons (e.g. location of response organisation), or that could be misrepresented in the mapping process (e.g. ambiguous location name or extent).

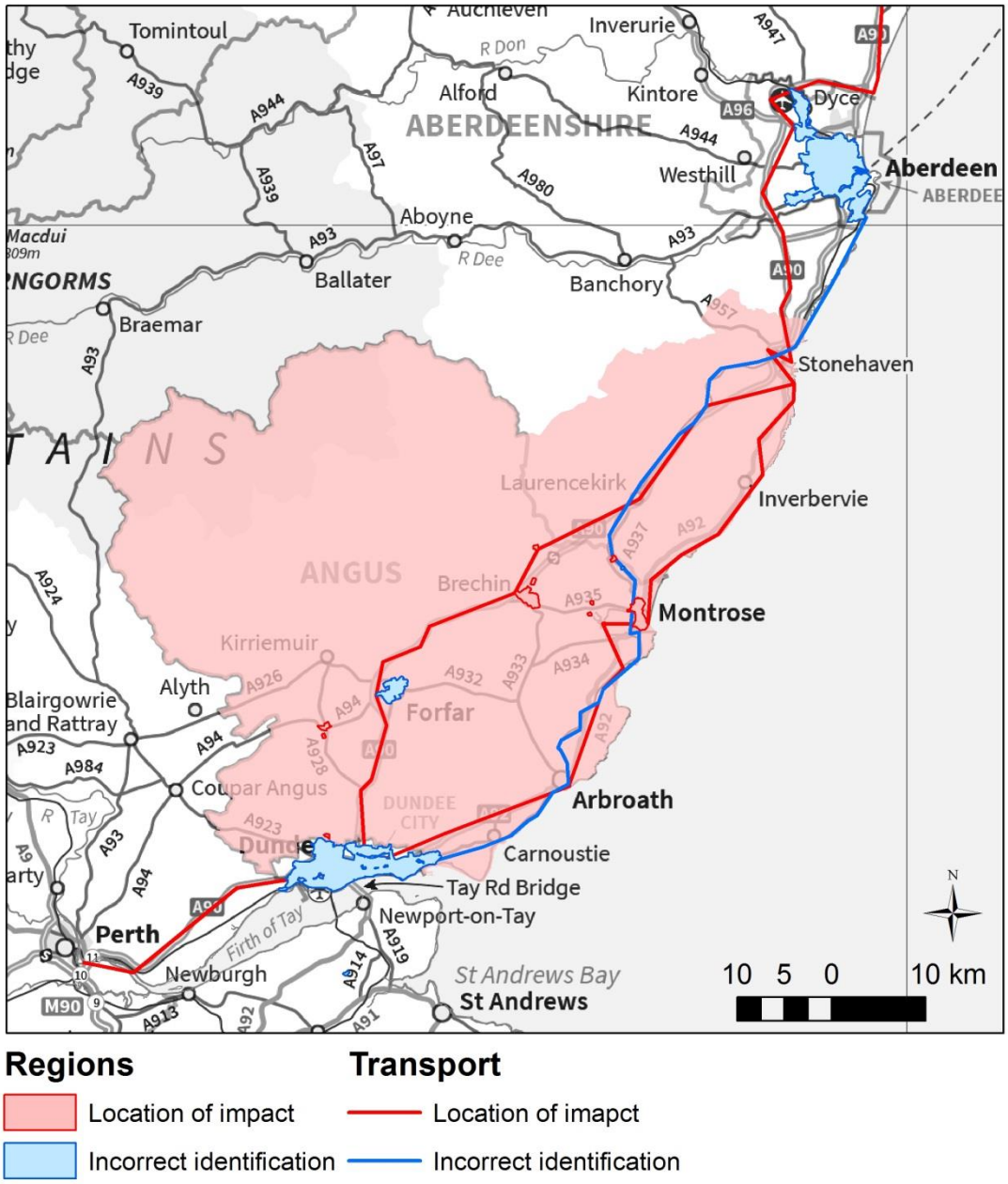


Figure 4 Example location mapping of impacts based on text mining location information extraction. Red regions and lines represent correct impact mapping. Blue regions and lines represent mapping based on incorrect extraction. ©Crown copyright and database rights (2018) Ordnance Survey 100021025.

Figure 4 highlights a number of issues that could be readily addressed by introducing additional rules into the extraction model. For example, Dundee and Aberdeen are both highlighted, but in the text they are mentioned in the context of the railway that joins the two together (the blue line). Forfar is highlighted as the location of the debrief meeting and the village of Logie is highlighted (on the railway line, by the A937) despite multiple sites named Logie in Scotland. Finally, there are known issues with the representation of impact in these maps. For example, Angus is highlighted as the region that was affected, but there is no indication of the distribution or severity of impact within the region. Additionally, the A90 and A92 are both successfully extracted, but both are long roads and without further analysis, it is impossible to identify where along those routes the impact

occurred. Finally, returning to the issue of Dundee, Aberdeen and the disrupted railway in between, it could be argued that closure of the railway does render the cities as impacted although they may not be locations where emergency response action was taken. Analysis of further reports would help to identify the performance of this approach for events with more of a focus in urban areas.

3.3 WORD FREQUENCIES AND ASSOCIATIONS

Tables 7, 8, 9 and 10 present the top ten most frequent words using single word tokens, bi-grams, tri-grams and 5-word n-grams. The tables demonstrate the power of joining words together, but it also highlights challenges when facing a series of reports that are designed to be written in the same way. For example, a number of the phrases in table 10 demonstrate repetition, and are likely to derive from similar sentences, which do not contribute a great deal to the analysis. These could potentially be addressed by filtering out key words from the analysis. The most frequent n-grams in tables 7 and 8 focus more on words relating to the hazard, and identification of response organisations. In particular, the word *water* appears many times, even though less than 30% of incidents were from natural hazards (Table 1). Tables 9 and 10 point towards some of the issues raised.

Tables 7, 8, 9, 10 Top ten most frequently used words in the background corpus for single word tokens (Table 7), bi-grams (Table 8), tri-grams (Table 9) and 5-word groups (Table 10).

Table 7

Word	Freq.
incident	96
police	68
response	67
hours	54
water	53
area	50
december	49
fire	44
scottish	44
power	42

Table 8

Bi-gram	Freq.
severe weather	19
police scotland	17
major incident	14
coordinating group	13
scottish water	13
western isles	12
care people	11
debrief participants	10
issues raised	10
overall debrief	10

Table 9

Tri-gram	Freq.
debrief participants felt	8
main issues raised	7
issues raised debrief	6
overall debrief participants	6
perth kinross council	6
tactical coordinating group	6
felt main issues	5
participants felt main	5
raised debrief around	5
sepa issued flood_alert	5

Table 10

N-gram (5 words)	Freq.
debrief participants felt main issues	5
felt main issues raised debrief	5
main issues raised debrief around	5
participants felt main issues raised	5
debrief highlighted number issues addressed	4
highlighted number issues addressed included	4
incident debrief highlighted number issues	4
issues addressed included action_plan appendix	4
number issues addressed included action_plan	4
overall debrief participants felt main	4

Tables 11, 12, 13 and 14 present the top 10 most frequently occurring words in the combined lessons learned text. Three of the words in Table 7 appear in Table 11 (incident, response, police). The other words are related to resourcing of response. Tables 12 -14 focus on response members and highlight actions taken and general feelings of how the response went. Phrases such as *worked well* (used 38 times)*, lack clarity around* (used 10 times), and *issues raised around* (used 8 time) give an indication of the general feeling around how successful the campaign was. Elsewhere, themes around raising issues and business continuity appear strongly in this corpus.

Tables 11, 12, 13, 14 Top ten most frequently used words in the combined lessons learned corpus for single word tokens (Table 11), bi-grams (Table 12), tri-grams (Table 13) and 5-word groups (Table 14).

Table 11

Word	Freq.
incident	508
response	304
staff	210
agencies	200
information	197
police	190
local	170
exercise	159
around	150
need	150

Table 12

Bi-gram	Freq.
police Scotland	79
incident scene	46
debrief participants	38
worked well	38
within ecc	36
dm crombie	34
public communications	32
major incident	29
issues raised	28
business continuity	26

Table 13

Tri-gram	Freq.
debrief participants felt	16
health social care	11
lack clarity around	10
shared situational awareness	10
business continuity plans	9
forward control point	9
area control room	8
declaration major incident	8
issues raised around	8
single point contact	8

Table 14

N-gram (5 words)	Freq.
escalate issue senior management team	4
incident provided excellent opportunity test	4
key areas consideration highlighted follows	4
sas liaison officer within ecc	4
via education scotland service desk	4
multiagency incident response guide mairg	3
note post debrief comment mca	3
police scotland area control room	3
within area control room acr	3
scottish fire_rescue_service support via	2

The information in Tables 7-14 can be presented in ways that may be more accessible to wider audiences. For example, Figures 5 and 6 present the top 100 most frequently used words in the background and combined lessons learned corpuses as a word cloud. These words have been classified using colour and size by word frequency, such that the largest words are the most frequent and the smallest are the least frequent. Immediately, the focus on the words *response* and *police* highlights the relative importance they may have. Word clouds for individual lesson types can be found in appendix 1.

Figure 5 Word clouds for the 100 most frequently occurring words in the background corpus. Colours and size have been used to clarify word frequency.

Figure 6 Word clouds for the 100 most frequently occurring words in the combined lessons learned corpus. Colours and size have been used to clarify word frequency.

Figures 7 and 8 present dendrograms of hierarchical clustering for the background corpus (Figure 7) and the combined lessons learned corpus (Figure 8). Figure 7 identifies six clusters of words that appear in similar distributions throughout the background corpus. Of these clusters, Cluster 1 appears to describe a commonly occurring hazard type, Cluster 2 describes police response, Cluster 3 seems to infer engagement with the public, while Cluster 5 is characterised by the word *fire*, which appears commonly in the corpus as a specific hazard type. In terms of specific words, the most common words were analysed for word associations, where Pearson's r correlation coefficient was greater than 0.7. *Power* was highly correlated associated with *speed*, *energy* and *outages*; *area* was highly correlated with *badly*. *Scottish* and *water* were highly correlated with a lot of words.

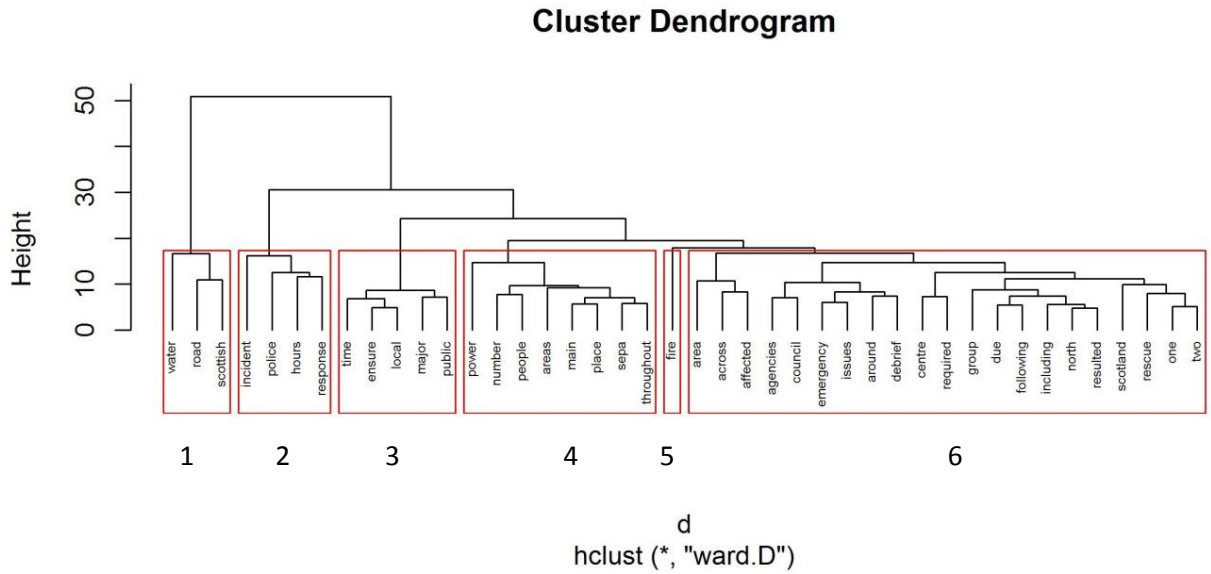


Figure 7 Dendrogram of the background document-term matrix using a sparsity threshold of 0.75. Common terms are listed along the x-axis. Red squares identify six clusters (numbered).

Figure 8 identifies six clusters of words that appear in similar distributions throughout the combined lessons learned corpus. The word list across the x-axis is much more focused on response considerations than the background dendrogram. Additionally, the clusters do not seem to be as well-defined. For example, Cluster 1 is represented only by the term: staff, while Cluster 2 appears to represent local concerns, but Clusters 3, 4 and 6 are very similar in character. In terms of specific words: *need*, *around*, *exercise*, *local*, *police*, *water*, *staff*, *response*, and *incident* are highly correlated ($r > 0.7$) with a lot of words. This indicates that in general, the lessons learned sections of the reports are generally more similar to each other than the background sections. This is to be expected as the hazards typically cover a wide range of different circumstances, while the lessons learned sections are more structured.

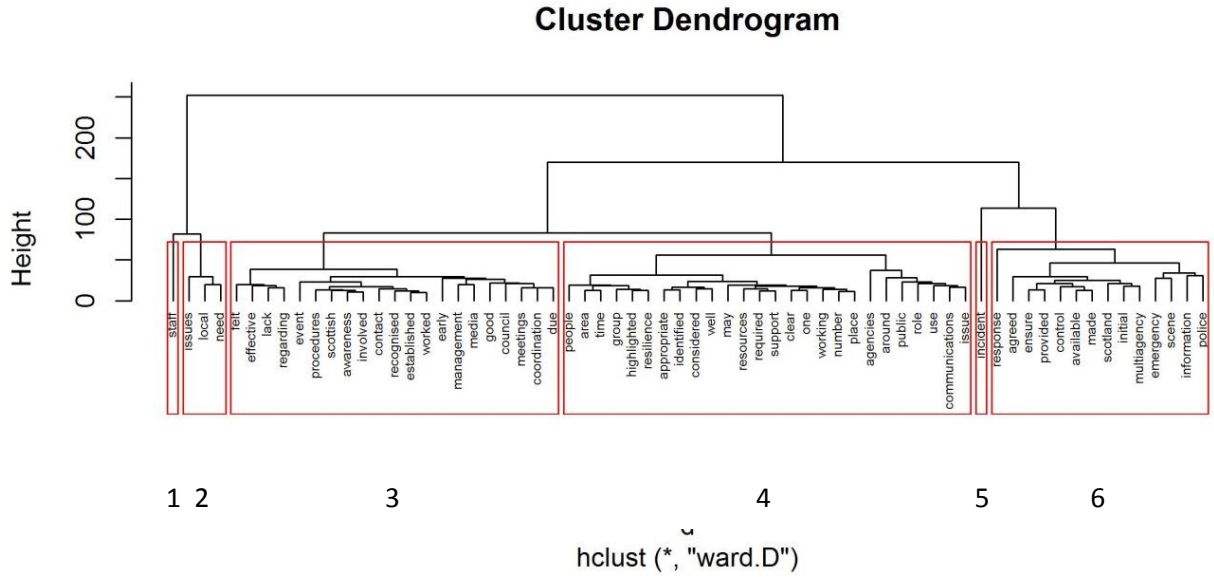


Figure 8 Dendrogram of the combined lessons learned document-term matrix using a sparsity threshold of 0.5. Common terms are listed along the x-axis. Red squares identify six clusters (numbered).

3.4 SENTIMENT ANALYSIS

Tables 15 and 16 present a comparison of mean proportions for each of the sentiments in the NRC function of the Syuzhet R package. Comparisons were made between incident and exercise reports, and natural and non-natural hazard incidents. The bold rows indicate significant differences in the comparisons, based on a t-test ($p<0.05$). Both tables show that the level of *positive* sentiment was generally high – between 75 and 79%. Results of the t-tests indicate that none of these compared differences are significant.

Table 15 Mean proportion of sentiments. t-test comparison of incident and exercise means. Bold rows indicate significant differences ($p<0.05$).

Sentiment	incident mean	exercise mean	t statistic	degrees of freedom	Significance (p)
Positive	0.776	0.751	-0.498	15.54	0.626
Anger	0.039	0.044	0.403	18.298	0.692
Anticipation	0.177	0.171	-0.296	18.265	0.771
Disgust	0.02	0.015	-0.998	24.549	0.328
Fear	0.134	0.153	1.206	21.568	0.241
Joy	**0.108**	**0.072**	**-3.753**	**25.751**	**0.001**
Sadness	0.073	0.081	0.571	20.066	0.574
Surprise	0.13	0.126	-0.25	21.553	0.805
Trust	0.319	0.339	0.876	19.199	0.392

Table 16 Mean proportion of sentiments. t-test comparison of natural hazard and non-natural hazard means. Bold rows indicate significant differences ($p<0.05$).

Sentiment	Natural mean	Non-natural mean	t statistic	degrees of freedom	Significance (p)
Positive	0.785	0.762	-0.96	36.919	0.343
Anger	**0.033**	**0.05**	**2.036**	**35.43**	**0.049**
Anticipation	0.181	0.171	-0.665	43.489	0.509
Disgust	0.015	0.015	0	42	1
Fear	0.127	0.146	1.2	30.524	0.239
Joy	0.115	0.098	-1.752	52.715	0.086
Sadness	0.067	0.082	1.105	34.813	0.277
Surprise	**0.147**	**0.102**	**-3.213**	**39.136**	**0.003**
Trust	0.308	0.337	1.478	34.311	0.149

When sentiments are broken down, *trust* presents the highest proportion by far. The next most prevalent sentiments are *anticipation*, *surprise* and *fear*. As an overall observation, this appears to be a reasonable expectation of the content of the reports. In the incident vs. exercise comparison, the only significant difference is that *joy* is higher in actual incidents than in the simulated exercises. This may reflect the general outcome of real-life incidents as opposed to training exercises. In Table

16, *surprise* is significantly higher in natural hazards than non-natural hazards. This is interesting as there is a more robust system of warning for natural hazards than non-natural hazards. *Anger* also exhibits significant differences, but the difference is relatively small.

Figure 9 presents a temporal view of *positive* sentiment comparing natural and non-natural incidents. *Positive* sentiment is generally high, but there is little other pattern in the data. Of the five case studies presenting the lowest proportion of *positive* sentiment, three case studies describe power outage events on the Scottish Islands, one describes a flood and one describes a chemical exposure emergency. Of the case studies with the highest proportion of *positive* sentiment, two represent missing person cases, the remaining three cases are related to flooding.

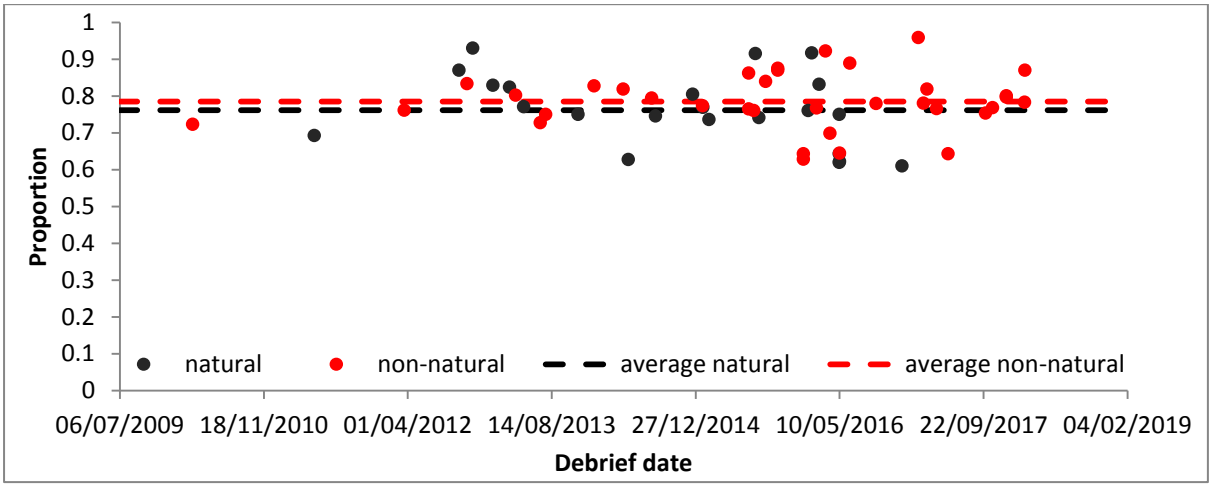


Figure 9 The proportion of *positive* words over time (debrief date) for natural and non-natural hazards. Each data point represents a single report.

Figures 10 and 11 present the composition of sentiment within the corpus over time. In both figures, the highest-proportion sentiment (*trust*) is located at the bottom of the graph, while the least representative is at the top (*disgust*). In both cases (natural hazards and non-natural hazards), *trust* and *anticipation* are the highest-proportion sentiments (30-68% for natural hazards, 36-63% for non-natural hazards). For natural hazards, the highest level of *trust* appears in case studies related to surface water flooding and extreme cold. For non-natural hazards, the highest levels of *trust* appeared in case studies relating to property fire and personal exposure to dangerous chemicals.

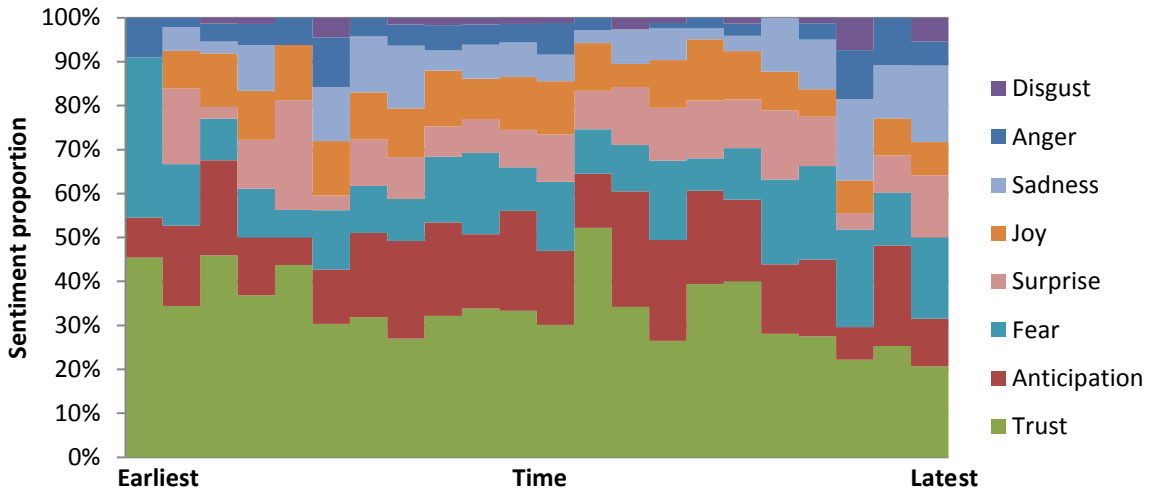


Figure 10 Composition of NRC sentiment analysis over time for natural hazards (left = earliest report, right = latest report).

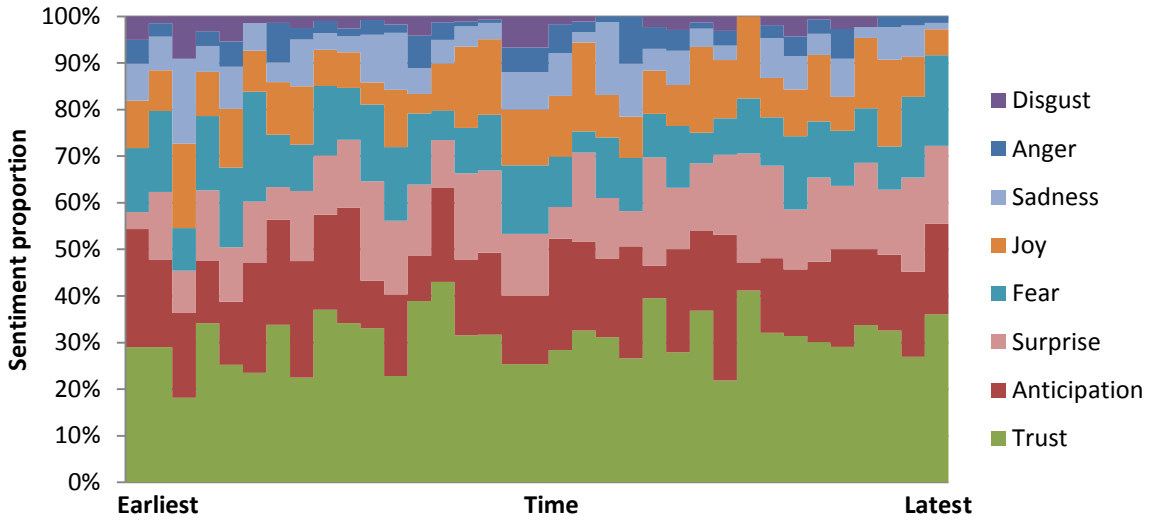


Figure 11 Composition of NRC sentiment analysis over time for non-natural hazards (left = earliest report, right = latest report).

Figure 12 uses the information identified in Figure 10 to cluster the reports by sentiment profile. Five clusters have been identified. Sentiment has been shaded to match Figures 10 and 11. Interpretation should be taken with care because document sample sizes are small and differences may not be statistically significant. Consequently, key differences in the proportion of sentiments have been drawn out. The proportion of *disgust* and *anger* are small and are therefore not considered.

Cluster 1 is the most numerous and has a high proportion of *fear* compared to the other clusters. Of the 21 reports in this cluster, 10 are natural hazard-related, while four reports are related to fires and 3 are related to vehicle accidents. Extracted impacts for these reports are predominantly related to service and access disruption and evacuation. A lot of the impact statements include the mitigation and response activities of the emergency services.

Cluster 2 has the highest proportion of *anticipation*. It is not characterised by any specific hazards. Most of the extracted impacts relate to disruption of roads and services, although this cluster also includes the H1N1 outbreak, where there were significant fatalities.

Cluster 3 has the highest proportion of *trust*. This cluster is based on water-related hazards, four of which are natural and one of which is a burst water main. The extracted impacts are all related to flooding of roads and property and include evacuation and sheltering of residents. The high level of *trust* identified in this cluster may be related to the high level of experience the responders have with severe weather. There are a number of well-used warning systems in place and the hazards are relatively well-understood.

Cluster 4 has the highest proportion of *joy* and *surprise*. This cluster contains five chemical incidents, three missing persons and a severe weather. The extracted impacts include a number of injuries, the unsuccessful search for a missing person and some damage to property. The sensitive nature of these reports is at odds with the outcomes of the sentiment analysis, although much of the impacts appear to be lighter than initially expected and in the very worst case, the phrase 'despite best efforts' may have caused misclassification.

Cluster 5 contains a high proportion of *fear* and the lowest proportion of *trust*. This cluster contains two natural hazards, three chemical incidents and a fire. The extracted impacts for these reports include decontamination activities, evacuation, the use of over 40 firefighters and power outage.

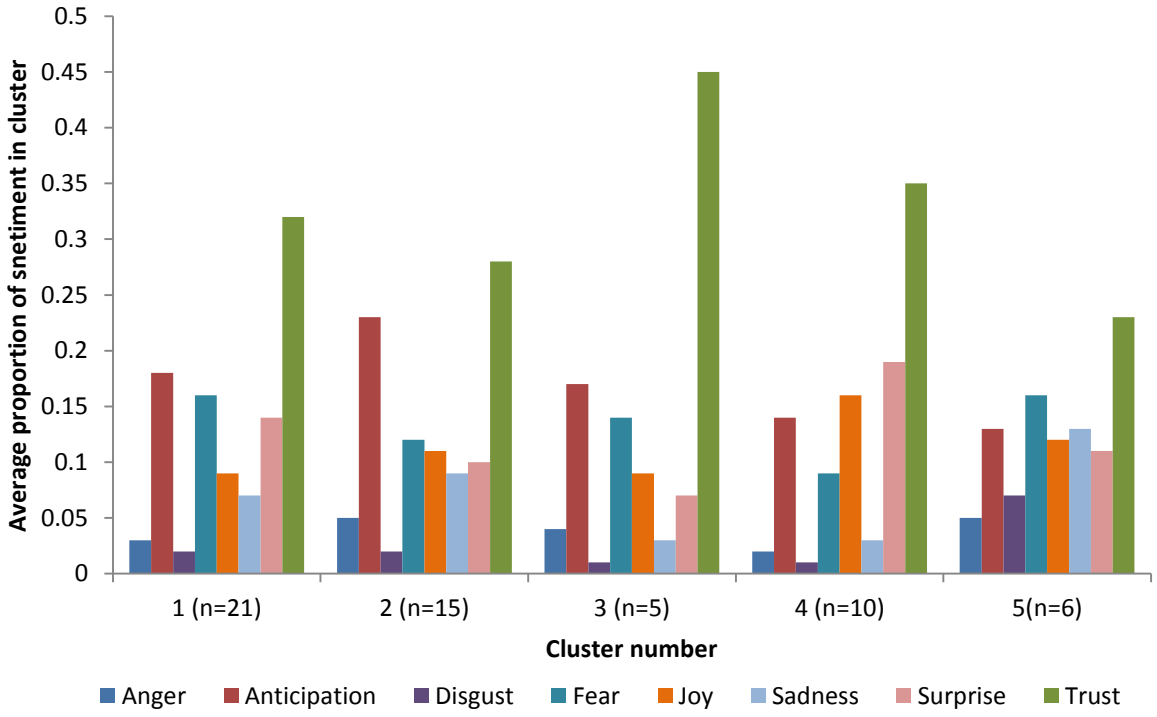


Figure 12 Hierarchical clustering of sentiment analysis (all incident reports). The numbers in brackets indicate the number of reports belonging to each cluster.

4 DISCUSSION

This exploratory analysis has presented a range of text mining techniques. Splitting the raw debrief documents by theme has enabled the analysis to focus on specific dimensions of the data. The results of each analysis have been presented in turn to highlight key words and phrases including an indication of common hazard-impact associations and key response organisations, a method for extracting sentences and locations of interest and identifying the general underlying sentiment of the debrief report. The following section discusses underlying themes across the analysis, suggests future developments and future potential for analysis of other unstructured text resources.

4.1 IMPACT EXTRACTION

The impact extraction task extracted sentences containing impacts, with an overall accuracy of 85%. The results of this analysis fulfil the initial objective of the report, which was to identify an approach to improve the collection of post-event impact data that can be measured. This objective directly relates to national and international targets (UNISDR, 2015). This is the first attempt at extracting impact information from these debrief reports and it could offer a new source of information for a range of applications include hazard loss accountancy, improved understanding and mitigation of the consequences of hazards, informing response decisions and validating and improving early warning system models.

Limitations to the impact extraction approach may be remedied by consideration of word context and by the addition or refinement of extraction rules. For example, the method used does not extract sentences based on specific hazard terms such as ‘flood’ or ‘fire’. This decision was made based on an assumption that sentences are hazard-focused, and may not include impact information. Further, terms such as ‘fire’ are problematic due to descriptions of the activities of the fire service. NLP may be able to improve extraction accuracy, and may assist in weighting or evaluating the contribution of words that modify a sentence (for example, words that enhance other words – *very* or *better*, or words that invert the meaning of a sentence – *not* or *despite*). With or without NLP, any ranking applied to the reports would depend on the requirements of the end user, and would likely require a measure of subjective judgement regarding how certain impacts compare against others and how different types of severity measurement (extent, intensity, type) can be weighted. Additionally, further analysis of extracted sentences could classify sentences into impact types, or impact severities through more sophisticated analyses of numbers and words describing scale within the sentences.

The techniques used for impact extraction have potential further applications in this dataset. For example, a record of warnings and alerts could be produced for each report, which could contribute to the characterisation and validation of the actual severity and extent of the hazard. Alternatively, mitigation and response actions could be identified for each record. This could enable a temporal comparison and correlation of the forecasted hazard, the actual consequences and the implemented responses. Additionally, key term frequencies could be used to characterise the documents. For example, counts of impact terms could be used to classify documents by impact severity and type.

4.2 LOCATION EXTRACTION

The process of mapping locations is relatively simple once the locations have been extracted. This process could be further improved by integrating location and impact extraction information together. This could potentially produce a spatial visualisation of the impact distribution including

severity, extent and type, providing an alternative interpretation of impact to the written document, which may be easier to interpret.

The limitation of vague and ambiguous locations are a known issue in preparation of validation data (Aldridge and Gunawan, 2016), but these could be addressed with local knowledge and by improving the location list used for extraction. The extraction of location not related to the hazard or impacts could be addressed by applying extraction rules evaluating the context of the location within the sentence. Conversely, this information may be important for other reasons. For example, extracting the locations of active response organisations could be useful for route planning or resource allocation. Alternatively, the location of sites indirectly impacted may be useful for understanding wider impacts and economic consequences. Additions to underlying code could apply Ordnance Survey references to locations for easier mapping.

The mapping of impacts into a GIS also allows for a number of further applications. A series of historical impact maps derived from text mining analysis could be combined within a GIS to highlight changing spatial patterns of impact or hazard extent over time for a given hazard type, provide insight during current emergencies or act as supplementary information for emergency and land use planners. These maps would represent a useful source of validation data for spatial hazard impact models such as those currently being created by the Natural Hazards Partnership.

4.3 SENTIMENT ANALYSIS

The sentiment analysis provided some insight into emotional themes running through the corpuses. It is encouraging to see that the sentiment of *trust* is the most representative in the texts. This may be a reflection of the fact that the documents are written in collaboration and much of the content discusses how different responders co-operated during an event or can improve their how they work together. The second most representative sentiments were *fear* and *surprise*, which may reflect the collective reaction to the hazard and the acknowledgment of realised impacts.

The diverse nature and the relatively small size of the corpuses used in this research mean that the full value of sentiment analysis has not been realised. However, the report demonstrates some of the analysis and some of the outputs that can be achieved. Future analysis could integrate with analytical extraction techniques to better understand the reasons for different sentiments. For example, evaluation of impacts against the relative sentiment analysis could provide some context for the composition of emotions present. Additionally, a more sophisticated temporal analysis could focus on how previous debriefs affected future reports. This may be improved by grouping the data by individual author organisations.

Figure 11 presents a demonstration of how sentiment information can be applied to classify reports by sentiment profile. Sample sizes are small, so interpretation cannot be made with great confidence. However, there are differences in cluster profiles which suggest differences in the nature of the hazards and the nature of the extracted impacts. The sentiment analysis presented in this report is limited by the fact that the source R package used for analysis was primarily designed for analysing story arcs within fictional novels. Consequently, some of the contextual assumptions may not fit with non-fiction reports. However, the underlying principles used in sentiment analysis are relevant to non-fiction and have been widely used to classify and summarise user feedback to improve commercial products and public services (Stylios *et al.* 2010; Bai, 2011).

5 CONCLUSIONS

This research has demonstrated the power of text mining using a case study of post-emergency debrief reports. Some of the findings may seem obvious when compared to human interpretation, but the gains are realised in the ability to analyse large bodies of text automatically and systematically, using accessible techniques that are demonstrated to be robust. The value of text mining has been revealed through the achievement of the primary objective. A catalogue of impact statement data can be retrieved from this body of texts. The efficiency and accuracy can be improved by implementing more sophisticated rules, but the key point is that an automated model that works with 70 articles can also work with 10,000.

Validation of results remains challenging as it relies on good sources of validation data. Further, the act of validation in this context is subjective to differing degrees. This is a particular concern for sentiment analysis, where the pre-defined libraries may define which words are related to '*joy'* or '*anticipation'* and key words used frequently in government reports may not be included. A potential solution to this is to create a different set of dictionaries based on word frequencies, or expert-identified words. This may then translate to a more relevant range of sentiments or themes across the texts.

The exploratory scope of this report has meant that some compromises have had to be made. In particular, the overall sample size is relatively small due to the nature of the documents (thankfully, there are relatively few civil emergencies that have required multi-agency response). Further, these reports cover a vast range of very different hazards, requiring different teams of responders and different actions and lessons learned. Consequently, this report should be treated more as a demonstration of the techniques that can be used to unlock the value stored within the documents rather than a scientific record of report contents and trends.

While the majority of the research has focused on text mining techniques in isolation to demonstrate application, there is a great deal of potential in combining text mining techniques with other machine learning processes. For example, predictive analytics could be applied to build models that may be capable of making predictions on impact severities, or response activities in the event of an incident. This could be used to benchmark actual outcomes and indicators of success. Such insights can then be used to form the basis of subsequent decision-making in the prioritisation and distribution of resources.

The techniques and analyses demonstrated here could be easily translated to other stores of documents for purposes ranging from characterising actionable information in tweets during civil emergencies (Verma *et al.,* 2011), to evaluating workers compensation claims data through the mining and classification of occupational injuries (Brooks, 2008), and classifying user feedback and social media to improve public services (Evangelopoulos and Visinescu, 2012). The key starting points in any such projects should be the identification or acquisition of a suitable and accessible body of unstructured text, and a clear idea of the aims and goals to be achieved.

6 REFERENCES

Aldridge, T. and Gunawan O. (2016) Surface Water Flooding Hazard Impact Model: Model Validation, http://www.naturalhazardspartnership.org.uk/science/hims/surface-water-flooding/ [accessed 12 September 2018]. HSL report MSU/2016/17.

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E.D., Gutierrez. and Kochut, K. (2017). *A brief survey of text mining: Classification, Clustering and Extraction Techniques.* KDD Bigdas, August 2017, Halifax, Canada.

Bai, X. (2011). Predicting consumer sentiments from online text, *Decision Support Systems (50),* 4, 732-742.

Brooks, B. (2008). Shifting the focus of strategic occupational injury prevention Mining free-text, workers compensation claims data. *Safety Science 46,* 1-21.

Bouwer, L. M., Crompton, R. P., Faust, E., Höppe, P., & Pielke Jr, R. A. (2007). Confronting disaster losses. *Science-New York then Washington*, *318*(5851), 753.

Chang, J., O'Reilly, C., Pontika, N., Haug, K., Owen, G. & Oudenhoven, M. (2018). *Text Mining 101.* https://www.fosteropenscience.eu/content/text-mining-101 [accessed 26 September 2018].

Chau, M., Xu, J.J. & Chen, H. (2002). *Extracting meaningful entities from police narrative reports.* Proceedings of the 2002 Annual National Conference on Digital Government Research, Los Angeles, USA, May 19-22 2002.

The Civil Contingencies Act (2004) (Contingency Planning) regulations 2005.

De Groeve, T., Poljansek, K., Ehrlich, D. and Corbane, C. (2014). *Current status and best practices for Disaster Loss Data recording in EU Member States: A comprehensive overview of current practice in the EU Member States.* Report by Joint Research Centre of the European Commission, JRC95505, doi: 10.2788/18330.

Evangelopoulos, N. and Visinescu, L. (2012). Text-mining the voice of the people. *Communications of the ACM 55*(2), 62-29.

Everitt, B. S., Landau, S. and Leese, M. (2001), *Cluster Analysis*, 4th Edition, Oxford University Press, Inc., New York; Arnold, London.

Gunawan, O. and Aldridge, T. (2018) *Disaster Loss Data Management in Scotland,* HSE research report DMS/2018/01.

Heron, D. (2016). Data in Government: *Understanding more from user feedback,* https://dataingovernment.blog.gov.uk/2016/11/09/understanding-more-from-user-feedback/ [accessed 26 September 2018].

Jockers, M. (2017). *Package 'syuzhet'* ,R Package.

Kao, A. and Poteet, S.R. (eds) (2007). *Natural Language Processing and Text Mining,* Springer, London.

Liu, B. (2010). Sentiment Analysis and Subjectivity. In Indurkhya, N. & Damerau, F.J. (Eds) *Handbook of Natural Language Processing.* Taylor and Francis.

Ngai, E.W.T. & Lee, P.T.Y. (2016). A review of the literature on applications of text mining in policy making, *Pacific Asia Conference on Information Systems 2016 Proceedings.*

Ordnance Survey (2018). *OS Open Names.* https://www.ordnancesurvey.co.uk/business-and-government/products/os-open-names.html [accessed 27 September 2018].

Percha, B., Garten, Y & Altman, R.B. (2012). Discovery and Exaplanation of drug-drug interactions via text mining. *Pac Symp Biocomput*., 410-421.

Plutchik, R. (1980). *A general psychoevolutionary theory of emotion.* In R. Plutchik & H. Kellerman (Eds.), Emotion: Theory, research, and experience: Vol. 1. Theories of emotion (pp. 3-33). New York: Academic.

R Core Team (2017). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/

Ready Scotland (2018). *Ready in your area,* https://www.readyscotland.org/my-community/ready-in-your-area/ [accessed 26 September 2018].

Rezaeian, M., Montazeri, H. & Loonon, R.C.G.M. (2017) Science foresight using life-cycle analysis, text mining and clustering: A case study on natural ventilation. *Technological Forecasting and Social Change 118,* 270-280.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2013). Tweet analysis for real-time event detection and earthquake reporting system development. Knowledge and Data Engineering, *IEEE Transactions on Knowledge and Data Engineering, 25*(4), 919-931.

Spire Technologies, (2016). *Why organisations should explore their unstructured data*, http://spiretechnologies.com/organizations-explore-unstructured-data/ [accessed 26 September 2018].

Stylios, G., Christodoulakis, D., BEsharat, J., Vonitsanou, M., Kotrotsos, I., Koumpouri, A., & Stamou, S. (2010). Public opinion mining for governmental decisions. *Electrical Journal of e-Government 8*(2), 203-214.

Taylor, C. (2018). Structured vs. Unstructured Data, https://www.datamation.com/big-data/structured-vs-unstructured-data.html [accessed 26 September 2018].

Tobback, E., Naudts, H., Daelemans, W., de Fortuny, E.J., and Martens, D. (2018). Belgian economic policy uncertainty index: Improvement through text mining, *International Journal of Forecasting (34)* 2, 355-365.

Torii, M., Yin, L., Nguyen, T., Mazumdar, C.T., Liu, H., Hartley, D.M., & nelson, N.P. (2011). An exploratory study of a text classification framework for internet-based surveillance of emerging epidemics. *International journal of medical informatics 80*(1), 56-66.

UNISDR (United Nations International Strategy for Disaster Reduction) (2015). *Sendai framework for disaster risk reduction 2015–2030*. http://www.preventionweb.net/files/43291_sendaiframeworkfordrren.pdf [accessed 13 September 2018].

Verma, S., Vieweg, S., Corvey, W.J., Palen, L., Martin, J.H., Palmer, M., Schram, A. and Anderson, K.M., (2011) Natural Language Processing to the Rescue? Extracting" Situational Awareness" Tweets During Mass Emergency. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 385-392).

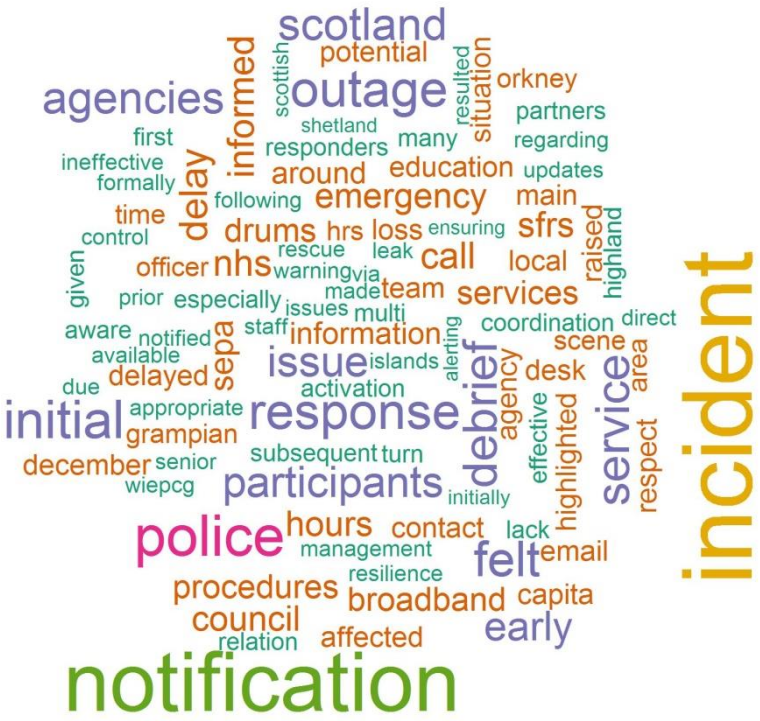
Zanasi A. (2009) *Virtual Weapons for Real Wars: Text Mining for National Security*. In: Corchado E., Zunino R., Gastaldo P., Herrero Á. (eds) Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems CISIS'08. Advances in Soft Computing, vol 53. Springer, Berlin, Heidelberg.

7 APPENDIX

INDIVIDUAL LESSONS LEARNED WORD CLOUDS

The top 100 most frequently used words in individual lessons learned corpuses.

Activation



Command, control and co-ordination

Communications

Environment

Logistics

Multi-agency working

Other

People

Plans

Resources

Response


meadows
site
responders
considered
coordinated
communications
control
decision
resources
provided
worked
scene
good
initial
emergency
rescue
early
officers
fire
centre
quick
public
incident
local
support
coordination
lack
police
national
effective
holiday
home
felt
well
nair
water
sepa
call
given
care
area
park
sfrs
deploy
grampian
services
potential
agencies
procedures
working
ambulance
scottish
service
response

The Health and Safety Executive (HSE) is Britain's statutory regulator of occupational health and safety. Our work covers a varied range of activities: shaping and reviewing regulations, producing research and statistics, investigating incidents and enforcing the law. HSE in its current form was created by the Health and Safety at Work etc. Act 1974.

HSE's Science Division is one of the world's leading providers of workplace health and safety research, training and consultancy, employing around 460 staff across a wide range of disciplines. We have been developing health and safety solutions for over 100 years, so we know what goes wrong in the workplace and why. Our relationship with the rest of HSE gives us a unique insight into the regulatory context of the work we do for our clients.

HSE's Buxton operations are certified to:
ISO 9001 OHSAS 18001

Health and Safety Executive
Science Division

Harpur Hill
Buxton
Derbyshire
SK17 9JN
UK

1.2 Redgrave Court
Merton Road
Bootle
L20 7HS

www.hsl.gov.uk
www.hse.gov.uk/research

T: +44 (0)20 3028 2000
E: hslinfo@hsl.gsi.gov.uk