International Journal of Disaster Risk Reduction 108 (2024) 104566

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

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

On mining mobile emergency communication applications in Nordic countries

Fuzel Ahamed Shaik, Mourad Oussalah *

University of Oulu, Centre for Machine Vision and Signal Processing, Faculty of Information Technology and Electrical Engineering, PO Box 4500, 90014 Oulu, Finland

ARTICLE INFO

Keywords:
Social media
Emergency communication
Aspect Based Sentiment Analysis
BERT
Ontological vocabulary
Empath categorization

ABSTRACT

Nowadays, the use of mobile devices has become ubiquitous, allowing users to access and share information in almost real-time through various social media platforms. This has provided an edge in emergency communication and disaster handling. Mobile emergency apps have emerged as key technologies in emergency communication. Mining the content of users' reviews of mobile emergency apps has the potential to learn about users' behavior and uncover unforeseen events in the emergency management process. This paper focused on emergency apps present in Nordic countries (Finland, Sweden, and Norway). User feedback was collected for every app from the Google/Apple store, and appropriate text mining techniques were employed to mine the discussion content for a given emergency communication ontology. Next, we investigated the contexts that generate either positive or negative sentiment, highlighting the main factors that impact user behavior most by leveraging the Empath Categorization technique. Finally, we constructed a word association by considering different ontological vocabularies related to mobile applications and emergency response and management systems. The study's findings can help develop early warning systems that trigger alarms whenever a critical event requiring special attention is identified. It also paves the way for developing a more tailored communication strategy that considers the identified community behavior concerning emergency apps.

1. Introduction

Information and Communication Technology, also known as ICT, plays a vital role in the development of our current society. Humans are leveraging ICT tools to solve various problems across different domains, such as business transactions, industrial operations, security and privacy, aid assistance, and all the major aspects of our daily lives [1]. The rapid growth in the use of computers, the Internet, and mobile applications, among other technologies, is the result of rapid development in ICT. Social media and Web 2.0 are two major aspects of ICT that have had a huge impact on our society. Users are leveraging these technologies to gather information on various activities, such as entertainment, recent market trends, climate forecasts, and disaster awareness. These technological advancements also contributed to emergency communications by broadcasting real-time information to a mass audience through social media channels. In essence, the primary factors in emergency communication are information sharing and location gathering, which directly provide an effective and a timely response to both citizens and relevant authorities [2]. These two factors are well accommodated in current mobile emergency applications, establishing a direct link between individuals and

* Corresponding author.
E-mail addresses: Fuzel.Shaik@oulu.fi (F.A. Shaik), Mourad.Oussalah@oulu.fi (M. Oussalah).

https://doi.org/10.1016/j.ijdrr.2024.104566
Received 28 October 2023; Received in revised form 15 May 2024; Accepted 16 May 2024
Available online 20 May 2024

Fig. 1. Mobile emergency applications are considered for our study. (a) 112 Suomi (b) Emergency plus (c) SOS alarm (d) Krisinformation.se (e) Hjelp 113.

emergency service providers [3]. This builds a bridge between individuals and governmental agencies by creating a bidirectional flow of information that acts as a public warning notification system that tracks emergency events.

Communication between users and emergency service providers during an emergency has always been a challenge. Indeed, information interoperability is a noted problem in emergency communication literature [4]. During an emergency, events quickly arise and evolve, requiring emergency communication networks to function at a high level of performance, flexibility, security, and interoperability,among other capabilities [5]. A significant challenge in emergency communication is to provide accurate and relevant information regarding the identification of emergency events and their locations. In some cases, the victims might not be in a state to speak to the emergency service providers to notify them of the incident, which can result in mishandling sensitive information (e.g., the user's disability or unrecognized locations). For this purpose, mobile emergency applications can provide useful insights to fill in this gap. Unlike SOS, calls where the user needs to make a call and communicate all the required details about incident intensity, location, time, etc., to emergency service providers, these mobile emergency apps reduce the effort of the user by automatically transferring most of this information through the mobile app itself through a dedicated mobile device. At the same time, users can also get updates about disaster evolution, emergency service activities, and other relevant announcements and notifications. Alert messages are automatically broadcast at high priority and with a wider dissemination rate. For instance, in the 2007 Southern California Wildfires, more than 500,000 acres of land suffered fires, 1500 homes were damaged, and 85 individuals were injured [6]. People in the impacted area were unable to get important information from the media and turned to mobile and social media to stay updated. They used various social media applications on their phones to contact family and friends , using available information portals and websites to learn about road closures and the condition of fire lines [7].

The European Emergency Number Association (EENA) promoted the use of mobile devices as a means of communication during emergencies, where an estimated 70% of emergency calls to emergency number 112 (in Europe) came from mobile devices [8]. Besides, EENA stressed the need for regulated mobile applications that are accessible across the European Union [3].

The main aim of this study is to devise an approach that uncovers emergency-associated factors, often referred to as *aspect terms* by mining user reviews from emergency mobile applications in Finland, Sweden, and Norway, extracted from the Play Store and the App Store in a way that enhances our understanding of potential contributions of these apps to emergency management and barriers towards the widespread adoption of these mobile emergency apps. Specifically, the objectives of this study are itemized into three parts:

1. To identify aspect terms in the user reviews that provide insights to comprehend users' opinions on the usage of these mobile emergency applications;
2. To mine the reasons and context behind negative user behavior as well as positive behavior; and
3. To devise an ontological vocabulary for emergency apps that best accommodates the content of users' reviews, contributing to future emergency systems' development.

To achieve these goals, our study is divided into four different stages:

1. Collect review documents associated with Finnish, Norwegian, and Swedish emergency mobile applications from two different mobile application stores (Google store and Apple store) to generate data frames, which are then pre-processed to be used in subsequent tasks.
2. Extract aspect terms and their polarity for each review document using a state-of-the-art deep learning model.
3. Negative polarity documents, which reflect users' concerns, as well as positive polarity documents, are further explored to uncover potential insights that cause users' worries and their interests using empath categorization.
4. Construct a vocabulary for emergency mobile apps where the relations between words are highlighted using word association trees and capitalizing from existing emergency ontologies.

The results of this analysis are compiled to act as a manual to comprehend the content of the underlined emergency mobile apps by understanding users' perspectives and needs. In total, we have considered 5 different emergency mobile apps across 3 different countries (112Suomi and Emergency Plus in Finland, SOS Alarm and Krisinformation.se in Sweden, and Hjelp113 in Norway) to perform this study; see Fig. 1.

The structure of this paper is organized as follows. Section 2 provides relevant background, highlighting the importance of mobile emergency applications during emergencies and relevant natural language processing methodological frameworks for mining review documents. The section also elaborates on the importance of ontology design to capture different aspects of emergency communication. Section 3 deals with the data collection step and summarizes the methodology employed to achieve this study's objectives. The results and discussion are reported in Section 4. Finally, we concluded our study by summarizing the key findings and describing the future scope and limitations under Section 5.

2. Background

2.1. Emergency communications

Several studies on formulating optimized communication systems for emergency handling have been conducted in the last two decades. A bidirectional mode of communication using mobile emergency applications has become popular in many European countries [9]. The Nordic countries are no exception to this trend. EENA highlights three significant means of access to emergency services: voice calls, emergency SMS, and mobile emergency applications [9]. Repanovici and Nedelcu [10] studied the current state, barriers, and future potential of mobile emergency notification apps. Their study statistically showed the promise of mobile apps for emergency communications. The authors used a multi-criteria analysis-based approach [11], a decision-making strategy that uses some performance index to rank various communication modes during emergency scenarios. Their study concluded that mobile emergency apps are the preferred means of access over voice calls and emergency SMS. Similarly, Tan et al. [12] provided a comprehensive review of mobile emergency apps by analyzing 49 crisis informatics articles focusing on mobile apps during emergencies [12]. Their study highlighted the need for efficient and friendly use of mobile emergency apps to accommodate the needs of various stakeholder groups, where various categorizations of apps and users have been distinguished. According to Tan et al. [12], presentation layout and visualization are two significant aspects that should be taken into account in the development of any mobile emergency communication app.

The current state of mobile emergency communication apps showcases advancements in real-time information dissemination due to the wide adoption of smartphone technology [3]. Indeed, such apps can directly benefit from the already-equipped mobile location technology using GNSS chipsets that support several satellite navigation systems (GPS, Galileo, Glonass, and more), as well as WiFi-based location, cell ID, and possibly other hybrid location technology. In addition, it enables the sharing of personal user's information, including disability, next-of-kin, and spoken language, together with other features such as multimedia sharing and video calling, thus saving valuable time for rescue operations. This is especially important to deaf people who cannot make a direct 112 call. According to the European Emergency Number Association (EENA), mobile emergency apps have great abilities to quickly identify and characterize emergency events through mobile alert signals, providing the rescue teams with answers to the three basic questions: what? where? and when? [13]. Up to 2022, as part of the EENA-PEMEA project investigation, there are 28 mobile emergency apps registered at EENA developed by the 13 EU states; many of them were ultimately linked to the EU-backed 112 emergency services. Nevertheless, several barriers still prevent enhancing the popularity and adoption of emergency communication apps. First, EENA itself stressed the need for the European mobile emergency apps to be compliant with a pan-European standard and to work across all EU countries in a standardized way. However, many EU countries developed their own mobile emergency apps at the national level only, which reduced their usability across the EU countries, as revealed in the PEMEA project review [13]. So far, only a limited number of countries, including Finland, have adhered to *the Pemea* standard, and the European Commission reported in 2023 that "the lack of a harmonized approach to establishing criteria for location accuracy and reliability hampers Member States' efforts to develop adequate solutions that ensure that emergency services benefit from caller location that is useful to effectively intervene in case of an emergency".[1] Second, the literacy issue can be another barrier, as it requires the user's registration, an active profile, and being able to access and manipulate the app from their phone device. Third, the infrastructure limitations in remote areas often hinder the accessibility of such apps, according to [14]. Fourth, while the "digital divide" exacerbates disparities [15], trust deficits between authorities and the public undermine the credibility of such applications [16]. Fifth, potential usability issues can harm users enthusiasm for the adoption of such apps. Finally, the rich panorama of the social media landscape renders the adoption of new apps and channels a difficult task from the user's perspective [17] due to the already increasing set of services provided to users by various providers.

Addressing these barriers is crucial for enhancing the efficacy of emergency communication apps by ensuring widespread access and understanding during emergencies. The field of emergency communication has experienced significant changes due to various upgrades in communication technologies, e.g., a transition from 3G to 4G networks and from 4G to 5G; an ongoing transition from 911 in the US (resp. 112 in Europe) to enhanced EN911 (resp. Next Generation NG112 in Europe). This ultimately impacted the way different countries tackled this transition phase, which created further difficulties in adopting a common standard. On the other hand, social media platforms have gained momentum in users' daily lives. Therefore, the ability to communicate with emergency centers through social media is of paramount importance, an issue that is yet to be supported by the future NG112 system. At present, the development of NG112 is still in its preliminary stage. The EU Commission is currently gathering individual countries' plans for Public Safety Answering Points (PSAPs) to be able to receive, answer, and process emergency communications through packet-switched technology. Therefore, future developments in mobile emergency communication technology should address the aforementioned issues to ensure effective citizen participation, mutual trust, and successful coordinated actions.

2.2. Mining review documents

The need for analysis of public behavior and motivation towards the use of mobile emergency applications has already been highlighted in several studies. One way to understand public behavior and motivation consists of mining users' reviews of these apps. Natural Language Processing (NLP) and text mining techniques, which include topic modeling and sentiment analysis, are

[1] https://joinup.ec.europa.eu/collection/rolling-plan-ict-standardisation/

well-known methodological frameworks to handle such tasks by identifying key discussion topics, variations in sentiment polarity, and other relevant language patterns that provide insights to comprehend the user's behavior and context. In this regard, topic modeling is an unsupervised learning approach that unfolds a list of topics from the text, where a set of keywords represents each topic. Latent Dirichlet allocation (LDA) [18] and its variants are commonly employed for this task. Topic modeling and sentiment analysis were jointly integrated in Venugopalan et al.'s work [19], where an extensive comparison analysis with recent supervised and unsupervised baseline models was carried out. Using restaurant datasets from SemEval 2014, 2015, and 2016 competitions, the model reported an F-score of 0.81, 0.74, and 0.75, respectively [19].

2.3. Aspect-based sentiment analysis tasks

Typically, sentiment analysis enables us to only classify a review document into positive, negative, or neutral classes, without any insights regarding the content of the review itself. Aspect-Based Sentiment Analysis (ABSA) [20] further refines the polarity analysis by linking the sentiment score to each aspect term (yet to be identified), which corresponds to some relevant attribute (s) in the content of the document (s).

Aspect term extraction (ATE) is one of the key ABSA tasks that aims to extract the aspect terms from the original text using some unsupervised learning algorithms or high-level ontology matching. Hu and Liu [21] employed a frequency-based pattern-mining approach to extract aspect terms from the text by identifying the most common nouns in the sentence. To leverage the complex relationships among various aspect terms and the inherent complexity of language inferences, deep learning and machine learning methods emerged as efficient tools to extract aspect terms instead of traditional frequency-based approaches. Gunathilaka et al. [22] adopted an ABSA approach using the CNN model to use user reviews in the requirement-gathering process, helping developers better understand user needs, leading to improved app development and user satisfaction. Authors in [22] proposed a CNN-based approach that distinguishes Aspect Category Classification and Aspect Sentiment Classification tasks. Their method achieved F1 scores of 0.62, 0.42, and 0.62 for aspect category classification and an accuracy score of 0.80, 0.70, and 0.86 for aspect sentiment classification in the productivity, game, and social networking domains, respectively [22].

Kirange et al. [23] identified aspect terms and the corresponding categories while performing fine-grained sentiment analysis of restaurant reviews. The authors used the Support Vector Machine (SVM) model and compared the results to those of traditional KNN classifier. It was found that the SVM model outperformed the KNN classifier in both the Aspect Term Polarity and Aspect Category Detection tasks, with an average accuracy score of 92.02% and 85.47%, respectively. Liu [24] & Zhang and Liu [25] promoted ontology-based reasoning. For instance, the aspect term (or category) can be (a) a part or a component of an entity (e.g., for the entity "laptop", the battery can be one of the aspect terms), (b) an attribute of the entity (e.g., for entity "laptop", the price can also be another aspect term), or (c) an attribute of a part or a component of the entity (e.g., for entity "laptop", the battery life can be another aspect term) [26]. ABSA has been researched in several SemEval conference competitions. For instance, SemEval2015 Task 12 focused on identifying the opinions expressed about specific entities and their aspects on three different datasets: restaurant, laptop, and hotel reviews [27]. It takes the whole review as input and extracts entity/aspect pairs and their polarities as output. In SemEval2016 Task 5, the study was further expanded into sentence-level ABSA.

Poria et al. [28] proposed a new rule-based framework to identify various aspects of the reviews using common sense and sentence dependency trees to differentiate direct and indirect aspects. A semantic aspect classification approach was proposed by Mukherjee et al. [29] to group various aspect terms into a set of aspect categories. Poria et al. [28] presented a seven-layered deep CNN-based opinion-mining approach to extract aspect terms from a given sentence by initially pinning each word as either an aspect or non-aspect. He et al. [30] suggested a word embedding method to extract the co-occurrence distribution of tokens and utilized an attention mechanism to imitate the importance of the irrelevant tokens, which further highlighted important aspect terms of the sentence. Wang et al. [31] built an end-to-end solution for aspect term extraction using a deep neural network model. Gao et al. [32], emphasis on constructing a short-text aspect-based sentiment analysis method based on a convolutional neural network (CNN) and a bidirectional gating recurrent unit (BiGRU). It predicts the emotional polarity using feature words and corpus sentences as the vector input [32]. The findings of the experiment demonstrate that the enhanced CNN + BiGRU model outperformed the Convolutional Neural Network (CNN), Long-Short Term Memory (LSTM), and Convolutional Neural Network (CLSTM) models in terms of classification accuracy by 12.12%, 8.37%, and 4.46%, respectively [32].

Hoang et al. [33] employed a pre-trained Bidirectional Encoder Representation of a Transformer (BERT) language model, incorporating a fine-tuning approach for an out-domain Aspect-Based Sentiment Analysis (ABSA) task. The authors specifically fine-tuned BERT for a sentence pair classification task, aiming to achieve sentiment polarity classification between the topic and aspect following data from SemEval2014 Task 4. Yang et al. [34] proposed a model that integrates BERT with a multi-head self-attention (MHSA) and local context focus (LCF) mechanism. These studies inspired us to implement a framework leveraging the state-of-the-art BERT model to extract aspects from user reviews of mobile emergency applications. For our study, the ABSA framework developed by Yang et al. [34] was adopted to extract aspect terms from user reviews. The results of this BERT variant model are later analyzed to identify the factors that are to be considered while developing a mobile emergency app or planning an emergency management system using mobile technology.

2.4. Emergency ontologies

Ontologies serve as tools for providing semantically unified representations of concepts and relationships that can be processed by machines in a way that facilitates decision-support during crises where the overwhelming flow of information requires directing

Table 1
Previous ontologies related to mobile applications.

Ontology	Use case	Role of ontology	Common objects/terms
MIDEP	User interface design patterns and ontology models for adaptive mobile applications [44].	Facilitate the choice of design patterns based on user characteristics and the surrounding context, alongside dynamically adapting user interfaces during runtime.	font size, position, interface, and map.
COCCC	Context ontology in mobile applications [45].	Designed for Android mobile applications, to formalize the contextual knowledge embedded within them.	GPS, location, battery, interface, and application.
mlO!	A context ontology for mobile environments [46].	Enables processing and utilization of the context for configuring, discovering, executing, and enhancing various services that may be of interest to the user.	interface, location, network, and coordinates.
CONON	Ontology based context modeling and reasoning using OWL [47].	Offers an overarching contextual ontology, encompassing fundamental concepts related to basic context, while allowing the flexibility to incorporate domain-specific ontologies in a hierarchical structure.	location, application, network, latitude, and longitude.
beAWARE	Ontology-based Representation of Crisis Management Procedures for Climate Events [36].	A more "all-around" compact ontology for managing climate crises, which dramatically simplifies decision-making and combines several essential factors.	location and natural disaster
CROnto	Ontology-Based Representation of Crisis Response Situations [2].	An ontology-based crisis representation that gives stakeholders in crisis response access to a single, shared knowledge base.	disaster, exposure, asset, resource, and impact

each piece of information to a relevant stakeholder group. Crisis management, with its various domains, e.g., healthcare, climate, and fire, has led to the development of several noticeable ontologies; see [35] for an overview. These ontologies differ according to the key concepts employed and the representation language used to depict the relationships among these concepts. Examples of concepts (or classes) include individuals, organizations, resources, damage, infrastructure, topography, meteorology, etc. Similarly, examples of relationships include causation, part-of, and disjoint-from. Representation languages include XML, RDF and OWL (Web-Ontology Language). This study extends the above research by seeking ontologies that may involve mobile emergency apps as well. In this regard, several other relevant ontologies have been identified. Kontopoulos et al. [36] defined beAWARE as an ontology for climate crisis management aimed at supporting decision-makers during climatic disasters. beAWARE uses the concept of spatiotemporal primitive relations between observed real-world objects to improve reusability and applies it to disaster forecasting and management. A fire-related ontology termed 'EmergencyFire' was proposed by Bitencourt et al. [37]. It displays fire emergency response protocols for evacuation purposes. In emergency medicine, Domain Ontology for Mass Gatherings (DO4MG) [38] aims to fill the gap in communication between medical emergency personnel. A linked data-based ontology named Management Of Crisis Vocabulary (MOAC) acts as a vocabulary manual for practitioners and experts to link crisis management activities [39]. Bannour et al. [2] developed a Crisis Response Ontology, called CROnto, by integrating crisis features, crisis effects, and crisis response to provide a complete and sharable knowledge base for the crisis response stakeholders.

Babitski et al. [40] proposed a semantic technological support system called SoKNOS, which aims to highlight the use of semantic technologies in developing disaster management software. Wu et al. [41] leveraged GPS data generated by mobile devices to propose an evacuation decision-making model during wildfires. Baytiyeh [42] showed the importance of mobile apps like WhatsApp during emergencies by highlighting the importance of location, and event characterization, sharing information through images and videos, and extending moral support to family and friends. MIDEP, COCCC, mlO!, and CONON are other examples of context-based ontologies that discuss the design and creation of mobile applications. Othman et al. [43] analyzed 30 metamodels to identify hidden events during disaster management, and later created a unified metamodel to address the events involved in disaster management. The response phase concepts highlighted by Othman et al. [43] can help us understand the factors involved in crisis management, especially the role of various stakeholder groups and coordination action.

A summary of these ontologies and the corresponding use cases is provided in Table 1, highlighting the underlined use-cases, key concepts employed, and the role of the ontology. All in all, this information about existing ontologies motivates us to design a tree vocabulary structure for mobile emergency communication. In this course, the mentioned ontologies provide a high-level description of some concepts related to "mobile applications", "emergency communication", and "disaster management". This inspired us to formulate a semantic and contextual-based vocabulary for mobile emergency applications to understand the influential factors in facilitating user interactions. The aspect terms extracted from the previous analysis are then used to construct a word association tree by highlighting the classes and objects from the corpus data. The classes and the objects are identified with the help of the previous ontologies related to emergency communications.

Table 2
Mobile emergency application's user reviews statistics before filtering.
Max. Tokens - Number of words in the longest review from the dataset
Min. Tokens - Number of words in the shortest review from the dataset.

Country	Mobile emergency application	Total number of users who either rated or commented	Total number of actual comments	Max. tokens	Min. tokens	Total number of comments after filtering
Finland	112 Suomi	4.7k	1470	201	2	414
	Emergency Plus	2.3k	973	442	1	972
Sweden	SOS Alarm	3.7k	591	697	2	451
	Krisinformation.se	825	229	132	1	228
Norway	Hjelp 113	1.6k	536	128	2	164

3. Data description and methodology

3.1. Data collection and preparation

The data used in this study are user reviews of emergency applications in the Nordic countries. We considered 5 different emergency applications across 3 countries ("112 Suomi" and "Emergency Plus" in Finland, "Hejlp 113" in Norway, and "SOS alarm" and "Krisinformation.se" in Sweden). Each country has developed its own emergency application and runs it under the regulations of its governmental body. Specifically, users' reviews located in both the Play Store and Apple Store are retrieved using scraping libraries in Python such as Google_play_scraper and app_store_scraper. The object class of the Google Play scraper library takes the googleId, googleLanguage, and googleCountry of the respective emergency app as inputs. On the other hand, the object class of the App Store scraper library takes appStoreName, appleAppId, and appleCountry as input. Both methods return complete data regarding the user and the user's review. In this study, we are interested in Review ID, User Name, Review, Rating, Date of Review, and Review Created Version.

As of February 27, 2023, we can see data regarding the number of total reviews and the total number of user comments on these applications in Table 2. This data is stored in a Common Separated Value format (CSV) for easy access and transformation. In parallel, the reviews are translated into English using the GoogleTranslator method from the deep_translator python library. The comment data is cleaned by reducing the redundancy and deleting reviews with empty or irregular comments. Also, comments whose token size is less than 5 are filtered out since these comments likely do not contain aspect-based information. This is motivated by the desire to ease the manual scrutiny of the review content whenever needed. The major drawback of this data collection process is the limitation of the Python library's ability to scrape reviews from both the Play Store and App Store, which requires careful manual checking and adaptation of the scrapper to ensure all reviews are reached. This study also reveals that although there are a relatively important number of reviews, only a fraction of them are exploitable. This is because users tend to either leave empty comments or use a small number of tokens, which renders the search of aspect terms of grounded users' comments rather unfeasible for these reviews.

3.2. High-level description of the method

To achieve the study's objectives set in the introduction part of this paper, a high-level description of the approach is outlined in Fig. 2. In particular, we distinguish the steps of data collection, aspect term extraction identification, aspect term sentiment classification, and ontology construction.

3.2.1. Aspect term extraction polarity classification (ATEPC) task

ATEPC aims to extract the aspect terms and their respective polarities from emergency app review documents. Specifically, the ATEPC model takes a sentence (from the user's review document) as input and yields a list of unique aspect terms along with their sentiment polarity. We have used the BERT language model [48] to identify the context according to its pre-trained model. To fine-tune this model for the ATEPC task, we needed to annotate the data for training purposes. The detailed explanation of this annotation and the fine-tuned BERT model are discussed in the next subsection.

3.2.2. Data annotation for ATEPC task

In this study, we follow the spirit of PyABSA, a modular framework to perform ABSA tasks [49]. Specifically, for the data annotation task, we employed DPT [34], a web-based annotation tool for aspect-based sentiment analysis (ABSA) tasks, and converted data into an ATE format. The data annotation for this study is carried out by two research assistants at the University of Oulu under the supervision of a Senior Research Fellow from the University of Oulu and four other collaborators from the DIGeMERGE project. Terms that impact the semantic meaning and are related to the mobile emergency application are considered aspect terms. Their respective sentiments in the sentence are defined as polarity. Fig. 3 shows a sample of DPT-annotated data.

After annotating users' reviews manually using the DPT tool, three files were generated: a CSV file containing the training set for the basic sentiment analysis process, a TXT file containing the training set for the ABSA task, and a JSON file that saves all unfinished

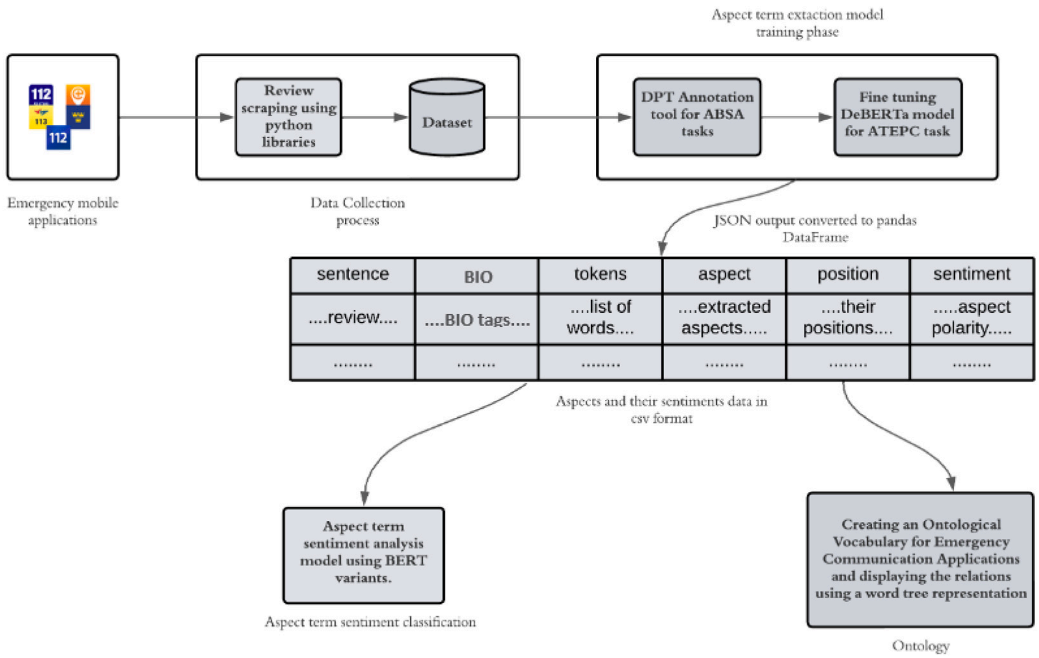

Fig. 2. High-level flowchart diagram of the study. BIO - Beginning of the aspect, Inside the aspect, and Outside the aspect. DPT - a web-based annotation tool.

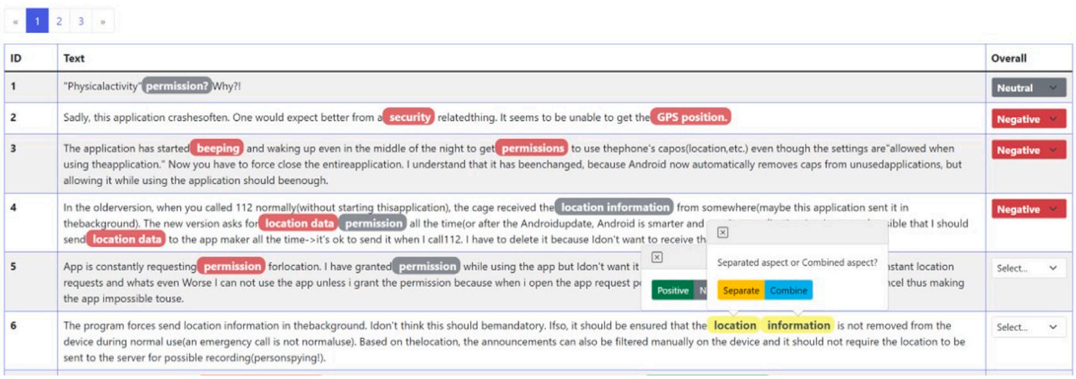

Fig. 3. Web-based interface of the data annotation tool DPT.

annotations. The data in the final output file follows the BIO tagging scheme. This scheme is commonly employed to perform Aspect Term Extraction (ATE) tasks, which are Beginning, Inside, and Outside of the aspect terms. For example, the sentence 'Fine with dark mode but where do you switch back to light mode?' will be divided into tokens of words where each word (w_i, $(1 <= i <= n)$, where n is the total number of tokens in the review sentence, here n=14), is separated with a comma in a list. It is easy to recognize from this sentence the importance of the wordings "dark mode" and "light mode", which carry the attribute "mode" and its intensity ("light" and "dark"). To highlight this observation, according to the BIO scheme, this sentence will be labeled as $\{O, O, B_{asp}, I_{asp}, O, O, O, O, O, O, O, B_{asp}, I_{asp}, O\}$. In the latter, B stands for the beginning of the aspect, I stands for its Inside, O for its Outside, and 'asp' denotes the sentiment of the aspect (negative, positive, or neutral). Fig. 4 illustrates this tagging operation.

3.2.3. Fine-tuning BERT for ATEPC task

The BERT model uses the concept of the Masked Language Modeling (MLM) technique. MLM masks the tokens in a given sentence by replacing them with a [MASK] value. Later, the model is trained to predict the [MASK] value based on the context of previous and forthcoming values in an input vector. This enables the model to learn the context of the words in the sentence. There are other variants of the base BERT model. This includes DeBERTa [50], where the attention weights of words are calculated using the disentangled matrices. Disentangled matrices contain the contents and relative positions of a word in a sentence encoded into two vectors to represent each word. This mechanism is known as the disentangled attention mechanism [50]. This makes DeBERTa v3 a state-of-the-art model to consider in our study.

```
Fine O -999
with O -999
dark B-ASP Neutral
mode I-ASP Neutral
but O -999
where O -999
do O -999
you O -999
switch O -999
back O -999
to O -999
light B-ASP -999
mode I-ASP -999
? O -999
```

Fig. 4. Example of BIO tagging output of a review sentence.

Inspired by the LCF-ATEPC framework proposed by Yang et al. [34], we employed a similar architecture to extract the aspect terms from the user reviews. Instead of the standard BASE-BERT model, we have integrated the LCF-ATEPC framework using the more efficient DeBERTa v3 model to increase performance. The annotated data (as mentioned in Section 3.2.2), which contain both ATE and APC labels, are fed as input to the model. This task is considered to be an unsupervised learning task since we do not know the target labels. After training the model on the training data, the model is tasked with predicting or extracting the aspect terms from the testing data. Later, the results are structured and saved into a CSV file, which we use in a further study.

3.2.4. Aspect terms sentiment classification

Inspired by Hoang et al. [33], a sentence pair classification architecture is adopted to classify sentiments of aspect terms associated with user reviews. For comparison purposes, different variants of BERT models are fine-tuned to perform this task. The architecture employed is illustrated in Fig. 5. The input array is created using a pre-trained BERT tokenizer by joining the context and aspect terms with a [SEP] token. For example, 'The location access is good' and 'location' are context and aspect, respectively, and the input token vector is structured as [CLS]+'The'+'location'+[MASK]+'is'+'good'+[SEP]+location+[SEP]. The BERT tokenizer provides three different encoder pairs; namely, token_ids, attention_masks, and token_type_ids. These encoder pairs are given as input to the model, which aims to predict the probability of each sentiment class (Positive, Negative, and Neutral). This output is achieved by stacking a hidden layer with a dropout rate of 0.1 along with a ReLU activation function on the [CLS] token.

3.2.5. Empath categorization

Identifying the *empath* associated with a user's opinion is very relevant to comprehending the user's attitude. Indeed, this reveals the underlying themes and motives behind users' thoughts by breaking down evaluations into more nuanced emotional aspects. This enables researchers to identify areas of interest and implement focused treatments or changes accordingly. It also adds value to the analysis by supplying thorough insights into users' views and preferences, which are essential for improving satisfaction. Therefore, empath classification has been devised in conjunction with the sentiment analysis task.

More specifically, inspired by Arhab et al.'s study [51] on Twitter data to understand users' car parking behavior, a data processing pipeline has been devised. Initially, the dataset is divided into two data frames according to the aspect term sentiment values: the Positive Aspect Sentiment Data Frame (PASDF) and the Negative Aspect Sentiment Data Frame (NASDF). Next, we perform an Empath categorization-based methodology on these two different data frames to comprehend the content of each data frame.

Fig. 6 illustrates the flow chart for Empath Categorization. We leveraged the Empath open-source library built by Fast et al. [52] to perform this task. The Empath open-source library has nearly 200 default categories and allows the creation of new categories whenever needed. Hence, we created custom empath categories relevant to our study to categorize the reviews accordingly. In a later stage, PASDF and NASDF data are combined to undergo traditional preprocessing, such as the removal of stopwords and redundancy. From this combined data frame, the top 20 common aspect terms from both PASDF and NASDF are considered to investigate the Empath related to each common aspect term. The results of this analysis can be observed in Section 4.3.

3.2.6. Ontology construction

The aspect terms obtained as a result of the Aspect Term Extraction task are categorized into classes and objects to define a relationship between these terms. A class can have multiple objects associated with it. The classes are considered by referring to some existing ontologies mentioned in Table 1. These ontologies stated in Table 1 were selected because of their relevance to Emergency Communication, as highlighted in the literature review section. Therefore, we initially considered that any aspect terms generated by our model could potentially be matched to one of these class terms. To evaluate this matching, we quantified the cosine similarity between the embedding vector [27] of the aspect term and each of the class terms, so that whenever the cosine score is beyond some threshold γ, the associated link is triggered (the aspect term will inherit the functionalities of the corresponding Class term). See Algorithm 1 for an algorithmic description. The value of threshold γ can be set using empirical statistical tests

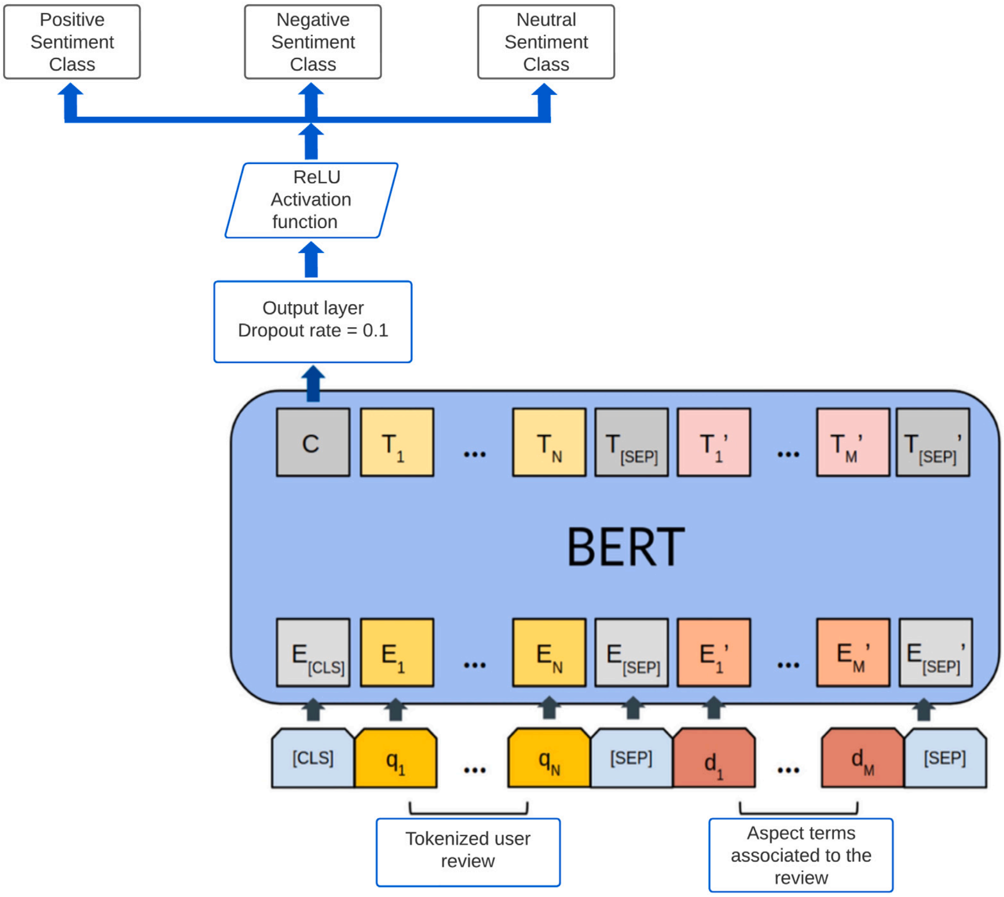

Fig. 5. BERT-based architecture for sentence pair classification.

to ensure trivial assertions in the context of the given classes are fulfilled. Apart from structuring the cosine similarities, we have performed some manual assignment of the object terms to certain classes by considering the context of the terms mentioned in the user reviews. An example of this manual assignment is discussed in Section 4.4.

Algorithm 1 Ontological Vocabulary Representation

```
procedure WordTreeConstruction('class', 'object')
    Load cosine similarity
    Initialize cosine similarity
    Initialize an empty dictionary
    Initialize an empty list for each class
    for each class in the classes do:
        for each object in the objects do:
            If cosine_similarity('class', 'object') ≥ γ:
                Append the 'object' into respective 'class' list
        END for
    END for
    for each class in the classes do:
        dictionary['class'] = respective class list
    END for
    Output the dictionary
```

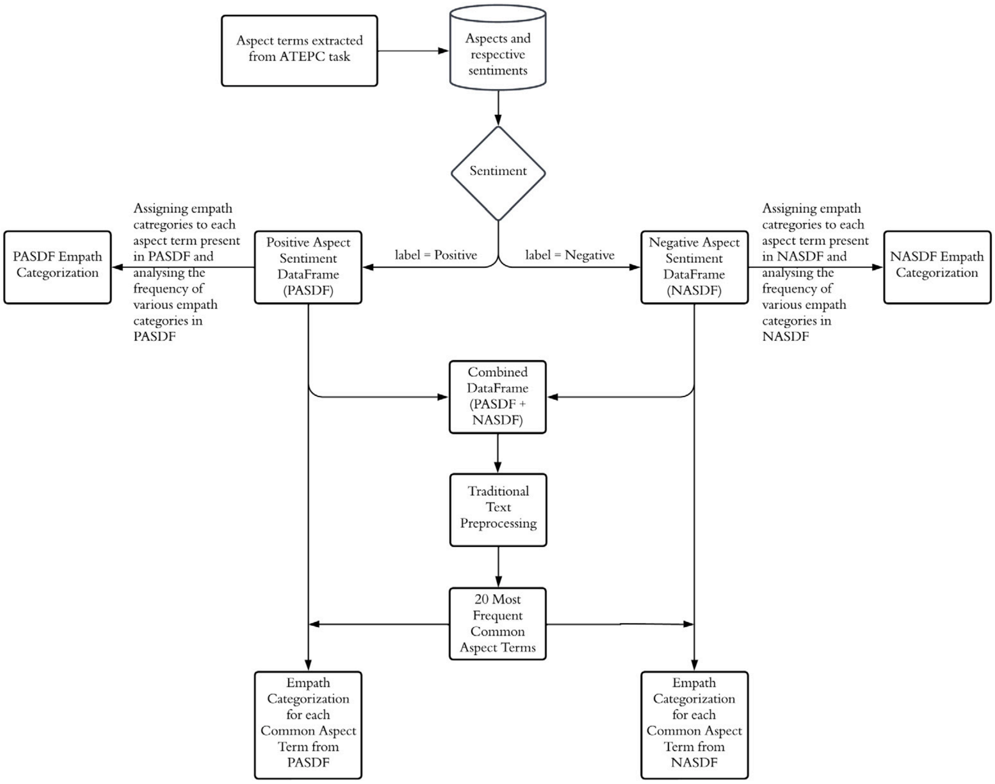

Fig. 6. A flowchart representation for empath categorization to identify the categories in both positive and negative aspect sentiment datasets.

4. Results and discussions

The results of the analysis are highlighted for each subtask, and then the overall results are discussed concerning the mobile emergency application.

4.1. Aspect term extraction task

We used the DeBERTa v3 model to perform the aspect term extraction (ATE) task. The output contains reviews, BIO tagging of each review, sentence tokens, aspect terms, the position of the aspect terms in the review, and their respective sentiments. The distribution of sentiments in the reviews is shown in Fig. 7.

We can observe that the dataset is more directed towards negative polarity, which is somehow widely expected as users are more motivated to use the emergency apps in case of unpleasant event occurrence, but also to comment on the suitability and difficulty of use of the app itself. Later, the sentiment rate (SR_a) of each aspect is calculated to identify the top 30 aspect terms from the whole corpus. The sentiment rate (SR_a) can be explained as the normalized weighted sum of the aspects in the dataset. The equation of SR_a is shown in Eq. (1).

$$SR_a = \frac{q}{N} \sum_{i=1}^{q} w_i \tag{1}$$

Where N is the total number of reviews in the dataset, q is the total number of reviews for the 'a' aspect and $w \in -1, 0, 1$ -1 for negative, 1 for positive, and 0 for neutral sentiment.

After extracting the aspects from the reviews, we highlighted the top 30 frequently repeated aspect terms in the corpus, which are displayed in Fig. 8. The aspect terms are app, location, battery, update, idea, emergency, GPS, address, notifications, position, coordinates, concept, map, open, fire, information, application, use, latest update, network, GPS coordinates, current events, phone number, events, positioning, emergencies, icon, call, security, alarm, and notification. These aspects terms can be considered while building an emergency application and also validate the performance of the mobile emergency applications according to each aspect.

The line plot drawn after calculating the sentiment rate of an aspect SR_a is shown in Fig. 9. 'App' aspect has created a good impression among the user reviews with a high positive SR_a. Whereas the 'location', 'battery', and 'update' aspects generate a negative sentiment, where 'battery' aspect exhibits the most negative score SR_a. Finally, the remaining aspects have mostly shown

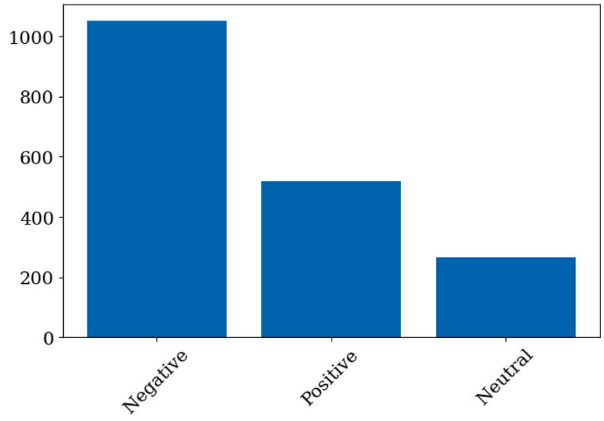

Fig. 7. Frequency of each sentiment in the aspect term extraction task output file which contains aspect term, BIO, aspect sentiment, review, and position of the aspect term in the review.

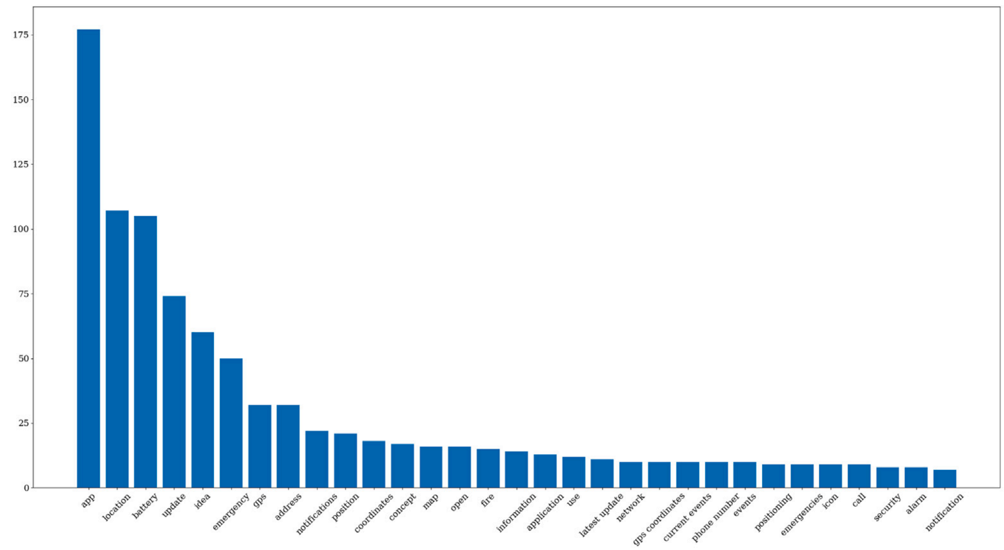

Fig. 8. Top 30 frequently occurred aspects from the aspect term extraction task's output data frame and their respective frequencies sorted in descending order.

neutral polarity as a key insight from this graphical representation. To further understand the sentiment of the particular aspects, we performed a separate analysis on positive and negative sentiment datasets, respectively.

4.2. Aspect sentiment classification task

This section describes the results obtained from the aspect-term sentiment analysis task using the different BERT variants for comparison purposes, and later designs a pipeline to identify the main categories/topics using the Empath category technique by estimating the weights of each of the predefined categories. The sentiment analysis model is a sentence pair classifier model, as discussed in Section 3.2.4. The model takes context and aspect pairs as input and predicts the sentiment of the aspect term in the given context. For comparison purposes, we have considered four different BERT models, which are BERT base, DistilBERT, RoBERTa, and ALBERT. Then, we stacked a hidden layer on top of each model with a dropout rate of 0.1. At the end, we have added a ReLU layer to predict the probability of labels (positive, negative, and neutral). The positive label is encoded as 1, neutral as 0, and negative as 2. Table 3 exhibits the results from each BERT variant model.

The results indicate that the RoBERTa and ALBERT models gave better results compared to the two other variants (Base BERT and DistilBERT). Since our dataset is small, large-batch training will make the learning process faster, which might lead to overfitting scenarios. Roberta may not perform to its best on small datasets since there is limited data available for fine-tuning. ALBERT, on the other hand, can leverage the limited data during the fine-tuning phase through its parameter-sharing technique while requiring only limited computational resources. Hence, we consider the ALBERT model for our final aspect term sentiment classification task. The classification results on the test dataset are shown in Table 4.

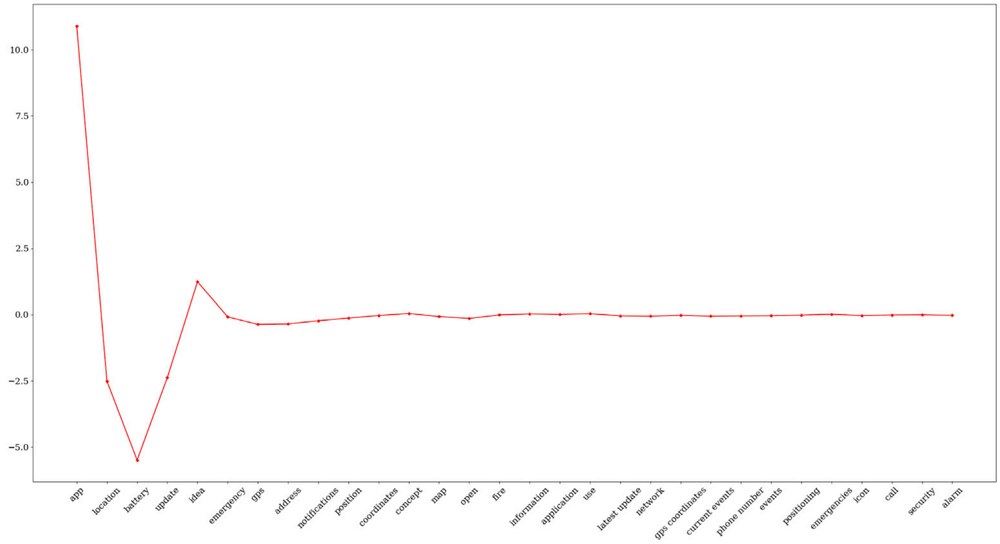

Fig. 9. A line plot of the top 30 frequently occurred aspects from the aspect term extraction task's output data frame and their respective sentiment rate SR_a scores.

Table 3
Results of different BERT variants for each label class Neutral, Positive, and Negative along with total accuracy.

Model	Precision	Recall	f1-score	Accuracy
	1.00	0.11	0.20	
Base-BERT	0.76	0.86	0.81	0.78
	0.78	0.97	0.86	
	0.33	0.12	0.17	
DistilBert	0.52	0.87	0.65	0.66
	0.85	0.69	0.76	
	0.79	0.68	0.73	
RoBERTa	0.76	0.95	0.85	0.84
	0.92	0.83	0.87	
	0.71	0.68	0.70	
ALBERT	0.80	0.85	0.82	0.83
	0.88	0.86	0.87	

Table 4
Classification report on test data from ALBERT model. Neutral class is represented as 0, Positive as 1, and Negative as 2.

	Precision	Recall	f1-score	Support
0	0.71	0.68	0.70	59
1	0.80	0.85	0.82	101
2	0.88	0.86	0.87	175
accuracy			0.83	335
macro avg	0.80	0.80	0.80	335
weighted avg	0.83	0.83	0.83	335

4.3. Empath categorization

In this study, motivated by the scope and variety of our aspect terms, we defined 11 new categories: application, coordinates, device_battery, emergency_services, exposure, fire_emergency, medical_emergency, miscellaneous, natural_site, network, and technology. These categories, together with the associated terms, are shown in Table 5. These categories are considered with the help of the ontological reasoning given in Section 4.3.

Further, we considered the commonly employed feature of word frequency in both positive and negative data frames as an indicator of posts' contents. In this course, the frequency charts of this analysis are illustrated in Figs. 10 and 11. The top 20 common terms from both data frames are noted together with their frequency in two different datasets, as illustrated in Table 6. Next, reviews associated with each common word are combined to undergo an empath categorization-based analysis. Finally, the top empath categories associated with each common word, along with their respective empath scores, are summarized in Table 7. The former displays the reviews from PASDF, and the latter shows the reviews from NASDF.

Table 5
User define empath categories based on the aspect terms association.

Empath category	Sample words associated with each category
application	functionality, interface, new_software, format, update
coordinates	latitude, GPS, address, location, position
device_battery	power, battery_drain, power_consumption, batteries, energy_consumption
emergency_services	evacuation_plans, service_centers, fire_alarm, fire_engines, distress_call, phone_lines
exposure	notification, relevant_information, alerting, inaccurate_information, electronic_messages
fire_emergency	forest_fires, nuclear_explosion, outbreaks, amazon_rain_forest, wildfires, yellowstone_national_park
medical_emergency	cardiac_arrest, ambulance, nearby_hospital, heart_attack, CPR, liver_failure
miscellaneous	permission, necessary, recommending, allowed, save_lives, guidelines
natural_site	hiking, mountains, cross_country_skiing, waterfalls, lake_michigan, valleys
network	signals, transmitter, coordinates, satellites, radio_signals, phone_signals, fiber_optic_cables
technology	initiative, safeguards, security, idea, urgent_need, need, flexibility

Table 6
Top 20 common words from both the PASDF and NASDF.

Common words	Number of occurrences in negative reviews	Number of occurrences in positive reviews
locate	190	67
work	174	43
battery	146	30
update	141	22
emergency	124	78
phone	119	41
good	98	109
use	93	46
notify	88	15
time	79	20
address	75	26
number	75	26
open	74	15
would	73	56
position	65	15
need	64	62
call	60	45
application	59	18
great	58	114
even	58	21

From Table 6, we can notice that common words such as ‘locate’, ‘work’, ‘battery’, ‘update’, ‘emergency’, and ‘phone’ are more frequently mentioned in the negative reviews data frame. Words, such as ‘good’ and ‘great’, are highly used in the positive reviews, indicating that a decent number of users are satisfied with the services provided by the mobile emergency applications.

4.4. Ontology on emergency communication apps

The ultimate aim of mobile emergency application ontology is to provide knowledge on the factors that are important for each mobile application from the user’s perspective and expand these factors to identify relevant variables of each factor. These factors are termed classes, whereas variables are called objects. The classes and objects are considered with the help of other ontologies and the aspect terms extracted from the ATEPC task. The word association tree is created after checking the similarity between the classes and objects, as mentioned in Section 3.2.5. The selection of relevant classes and objects from the extracted aspect terms is based on the various concepts mentioned in the related existing ontologies in the literature. For instance, MIDEP ontology highlights the objects that are supposed to be in mobile applications, such as “Font size”, “Position”, and “Interface”. These are also considered to be objects in our study, which helps in creating a word-tree association. Further, the semantic relations between classes and objects regarding mobile emergency applications are computed.

The ontology word representation, similarity scores for each relation, and relations between terms are shown in Fig. 12. We have considered nine different classes: app, coordinates, initiative, concept, network, accidents, fire, emergency, and function. In word representation, each class contains one or more objects to expand the class characteristics. Some objects share different classes, and some objects act as a class for a small cluster. The relation is called by the ‘subClass of’ notation. For example, in Fig. 12, ‘Idea’ is an object that is a subclass of ‘Initiative’ and ‘Concept’. That means the term ‘Idea’ can be either used as a ‘Concept’ or ‘Initiative’. Similarly, the ‘Medical Emergency’ object acts as a class for a small group of objects such as ‘Heart attack’, ‘Serious illness’, ‘Ambulance’, and ‘First aid’.

Apart from this, we also represented some terms as action–reaction relationships through manual labeling tasks. Typically, an action–reaction relationship indicates a mutual causal effect when changed. For example, in Fig. 12, the user experience of reading

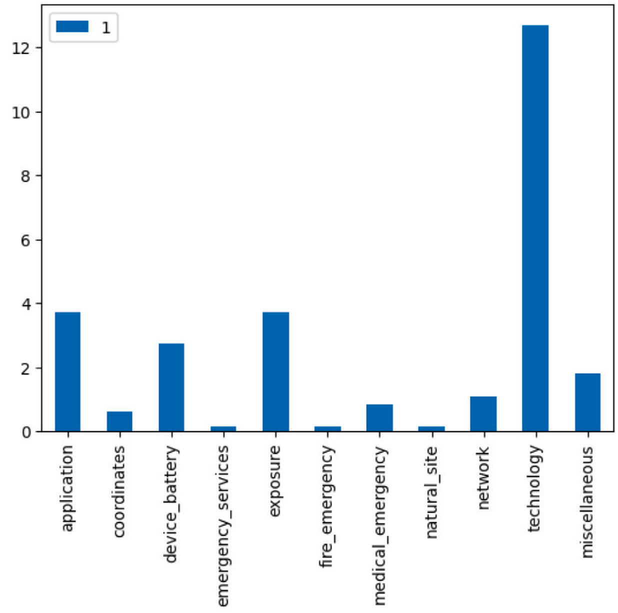

Fig. 10. Frequency of empath categories in positive aspect sentiment data frame (PASDF).

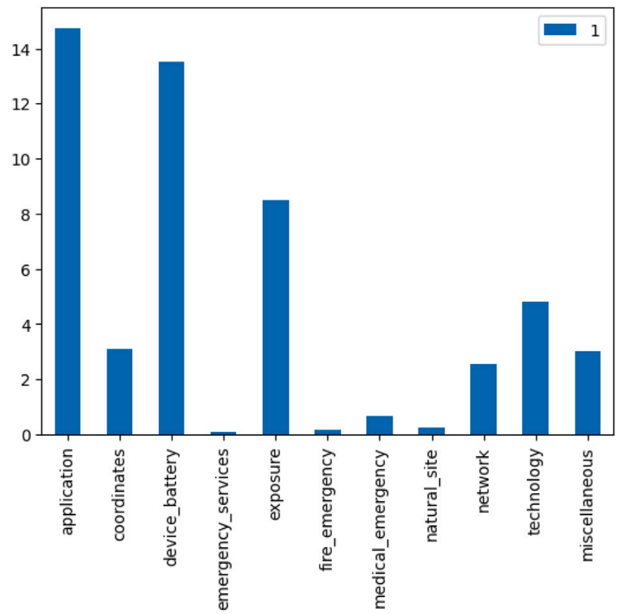

Fig. 11. Frequency of empath categories in negative aspect sentiment data frame (NASDF).

text on the application can be altered by font size, where some users prefer a larger font size, and some like a smaller font size. This implies that the change in font size affects the reading experience for a user. Hence, 'Font size' and 'Read' objects are related through an action–reaction relationship. Other relations, such as 'synonymy or similar context are highlighted using different colors in the word representation shown in Fig. 12.

4.5. Discussions

4.5.1. Key findings

The results illustrated in the above sections helped us achieve the objectives of our study by identifying the aspects impacting user behavior and motivation to use mobile emergency applications. For instance, the aspect Term Extraction task helped us identify the

Table 7
Top 20 common words along with their respective top 3 empath categories and empath category score from PASDF and NASDF.

Word	Empath categories in NASDF	Empath categories in PASDF	Word	Empath categories in NASDF	Empath categories in PASDF
locate	device_battery (1.97), application (1.48), miscellaneous (1.42)	technology (0.87), application (0.51), exposure (0.49)	even	device_battery (0.85), application (0.78), exposure (0.66)	application (0.38), exposure (0.37), technology (0.20)
work	application (3.59), exposure (0.88), device_battery (0.83)	application (0.79), technology (0.36), exposure (0.24)	great	application (0.95), technology (0.66), exposure (0.39)	technology (4.60), application (1.30), exposure (0.80)
battery	device_battery (12.80), application (1.63), technology (0.95)	device_battery (2.38), technology (0.57), coordinates (0.15)	application	device_battery (1.67), exposure (0.70), miscellaneous (0.69)	device_battery (0.46), exposure (0.11), medical_emergency (0.10)
update	application (11.59), device_battery (1.72), exposure (0.60)	application (1.50), exposure (0.31), miscellaneous (0.17)	call	device_battery (0.57), application (0.47), miscellaneous (0.35)	technology (0.93), application (0.45), network (0.42)
emergency	application (1.05), exposure (0.76), technology (0.68)	technology (1.73), miscellaneous (0.63), application (0.51)	need	technology (1.71), application (1.37), exposure (0.30)	technology (3.23), exposure (0.53), application (0.41)
phone	device_battery (1.93), application (0.74), technology (0.67)	technology (0.55), device_battery (0.34), application (0.30)	posit	coordinates (2.73), device_battery (1.49), technology (0.58)	coordinates (0.46), technology (0.36), exposure (0.25)
good	device_battery (3.24), technology (2.08), exposure (1.20)	technology (2.56), device_battery (2.55), exposure (1.32)	would	exposure (0.70), technology (0.54), device_battery (0.51)	technology (0.74), exposure (0.60), application (0.49)
use	device_battery (3.04), application (1.02), miscellaneous (0.66)	exposure (0.65), device_battery (0.56), technology (0.46)	open	application (2.31), exposure (0.49), technology (0.19)	technology (0.28), application (0.20), exposure (0.09)
notify	exposure (5.84), technology (0.46), device_battery (0.35)	exposure (1.36), technology (0.44), fire_emergency (0.11)	number	application (0.32), exposure (0.28), miscellaneous (0.28)	technology (0.38), exposure (0.21), application (0.18)
time	device_battery (1.18), exposure (0.76), coordinates (0.37)	technology (0.55), application (0.24), exposure (0.12)	address	network (0.65), application (0.49), technology (0.36)	technology (0.87), application (0.51), network (0.19)

key factors from the reviews. More specifically, the following can be reported: First, from Fig. 7, we can observe that the corpus of all review documents has more negative aspect sentiments than positive and neutral. This is in agreement with negativity dominance and negativity bias in social psychology [53], which posit that users are more tempted to record negative experiences as a way to manifest their negative emotion or non-satisfaction.

Second, using the sentiment rate SR_a metric, we can notice that mobile emergency applications are negatively impacting battery usage, users' perceptions of location accuracy, and the status of application updates. On a positive note, the applications are found to perform well in informing about traffic announcements, current events, emergency calls, and security updates. Third, the Aspect Sentiment Classification task helped us conduct a joint operation of aspect identification and their associated sentiment polarity. This approach is becoming popular in consumer and advertisement research. For instance, Zhang et al. [54] leveraged the Aspect Based Sentiment Analysis method to capture consumers' preferences, sentiments, and willingness to repurchase. In [54], a multi-aspect stacking model using a weakly supervised BERT model was proposed to enhance accuracy and generalization capabilities. Fourth, regarding the use of Empath Categorization, which aims to understand the empathic reasoning behind emotions in a text, our study revealed several findings. The 'application' empath category yields the highest score in negative sentiment datasets, followed by the 'Device Battery' and 'Exposure' empath categories, with scores of 14.73, 13.50, and 8.49, respectively. This is also supported by the term-frequency results exhibited in Table 6. This indicates that users are more concerned by the status of their mobile phone device and its good functioning state, which includes appropriate execution of the applications, as exemplified in the three above aspects (application, battery, and exposure). This shows the importance of ensuring mobile phone battery availability and reasonable life expectations for the emergency app. This finding is also in line with LG's research, referred to as "low battery anxiety" [55]. This calls for the importance of accounting for users' mobile phone behavior, which includes "low battery anxiety" during the process of emergency app design. In the positive polarity data frame, the 'Technology' empath category yields the highest empath score of 12.69, which testifies to the user's readiness to embarrass new technology if found useful. The common words identified from both the positive and negative sentiment corpus can provide an overview of the functionalities that a mobile emergency application should stress.

Fifth, to comprehend the relationship between these aspect terms, we constructed a word association representation by referring to some of the existing ontologies and the aspect terms extracted. For example, the negative sentiment towards location and battery is due to the location 24/7 app the user needs to run continuously in the background, which results in a high increase in power consumption. This partly explains the negative polarity regarding these aspects. Similarly, the action–reaction relationship provides

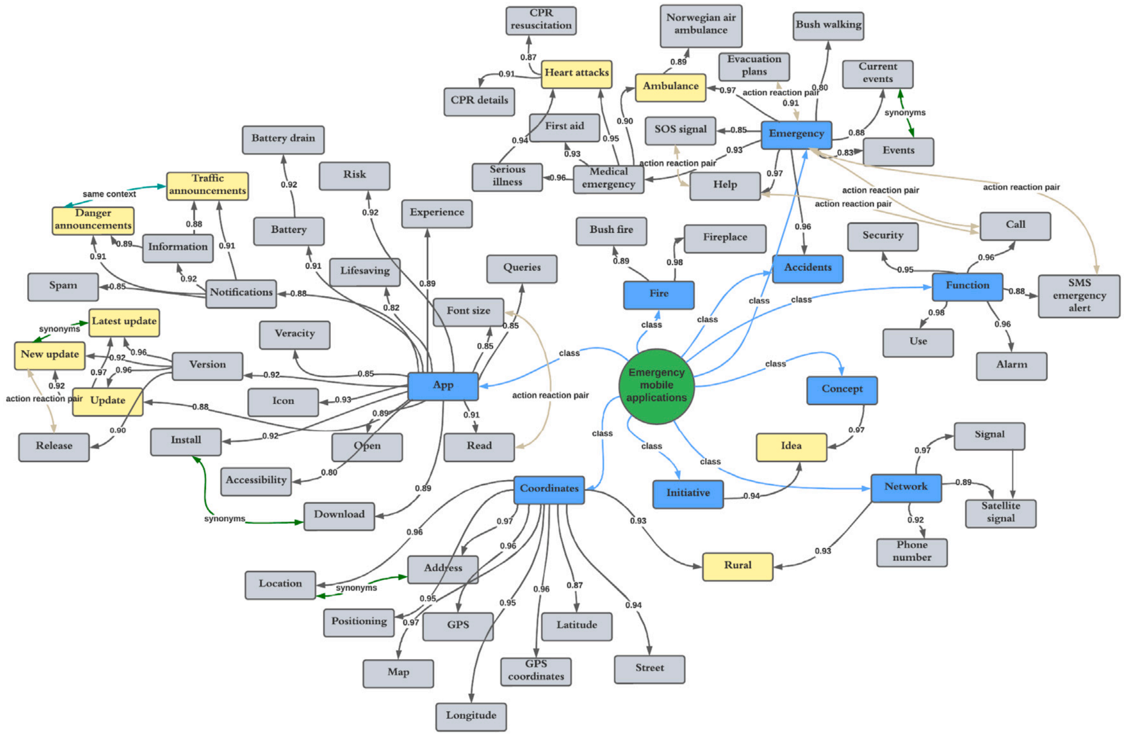

Fig. 12. Word association representation.

insights to understand a user's behavior from a causal perspective. For example, the font size of the text in the app can affect the readability of the user; hence, font size and read objects are considered action–reaction pairs. The reader can refer to Fig. 12 to see more examples of such relationships.

Sixth, mobile emergency applications' usability has significantly risen since the pandemic. In a review article, Islam et al. [56] discussed the functionalities of emergency applications specifically designed during COVID-19. The review's findings emphasize the necessity of developing new applications and improving those that already exist, taking into account not only the functionalities and the objectives that have been made public but also the design features that could encourage user engagement and motivation, such as ease of use, performance, dependability, security, privacy, flexibility, responsiveness, and cultural sensitivity. Similar findings are highlighted in [57] for the development of E!App, a smartphone application for emergency preparedness and response in Upstate New York. Previously, Tan et al. [58] identified several usability concerns specific to disaster apps, which include content relevance based on the app's purpose and the immediacy of information regarding hazards. Users expect efficient functionality without draining critical phone resources. Audio interfaces can enhance usability by improving alerting capabilities. From this perspective, we shall acknowledge that the emergency mobile apps collected in this study lack several of the aforementioned usability features, which would probably have boosted their popularity and the number of reviews collected.

4.5.2. *Theoretical contributions.*

Several theoretical contributions can be pointed out in this study. First, as far as our literature is concerned, this is the first analysis of mobile emergency apps in Nordic countries and Europe, utilizing the content of Google/Apple repositories.

Second, from a methodological approach, our study extends the aspect-based sentiment analysis [20] by providing a user-centric approach where the user can contribute through the manual annotation tool DPT. This distinguishes it from the fully supervised approach advocated in the literature on aspect sentiment analysis. This user-centric approach is particularly useful in disciplines where terminology is case-sensitive, as in the emergency communication field, which requires extensive collaboration among various stakeholders. Next, we have integrated the LCF-ATEPC framework into the DeBERTa-v3 model, contributing to the highly researched area of BERT models. For the classification task (sentiment of identified aspect terms), our work extends the initial BERT architecture proposed by Hoang et al. [33] by integrating the LCF-ATEPC framework and adding an extra layer to perform the classification task.

Third, the developed framework extends our original conceptual model [51] for distinguishing positive and negative reviews in a way to yields balanced coverage of the user's opinions.

Fourth, the approach employed for the emergency ontology construction by capitalizing on identified aspect terms and invoking existing relevant ontologies and term matching using word-embedding similarity to borrow the entity-relationship presents a novel construction that can be extrapolated to several other disciplines. This significantly differs from the popular protégé-like construction [59].

4.5.3. Practical implications

This study revealed several aspects with straightforward practical implications. First, in line with EENA's findings and recommendations, the compliance of the examined mobile emergency apps with the pan-European standard is missing to incomplete and should be given a high priority in the future development of the field. This is highlighted by examining the language of the submitted reviews, where only a limited number used alternative languages outside the national language and English. However, the 112-Suomi app fully complied with the pan-European standard as it was one of the partners in the PEMEA project. Besides, it was found that Denmark does not promote any mobile emergency apps, which makes it an exception in the Nordic and European emergency app landscape. Although we highly anticipate that changes will occur with the progress in the development of NG112, which will likely enforce more coordinated actions.

Second, usability factors, including battery efficiency, layout, and cultural norms, play a key role in sustaining users' motivation towards the use of these emergency apps, and their wide adoption by citizens. Third, as a two-sided communication channel, building mutual trust is of paramount importance to ensure citizens' involvement and confidence in manipulating mobile emergency apps.

Fourth, the framework put forward in this study can be ultimately employed to revise existing emergency apps and motivate the development of more user-friendly apps, by utilizing existing review reports. The approach can be further extended to capitalize on potential users' reviews gathered from alternative sources, e.g., social media platforms.

Fifth, the study revealed that relying on Google/Apple store for feedback gathering is not always sufficient. For instance, 112-Suomi, despite recording close to two million users who downloaded the app, only a few thousand submitted reviews to Google/Apple store (with a smaller number is exploitable). Therefore, identifying other means to collect users' feedback and compensate for the lack of statistically significant samples is equally important. This includes online questionnaires and social media platforms, which convey a substantially high volume of users' generated content.

4.5.4. Limitations

Several limitations are identified throughout this study. First, the total number of exploitable reviews turns out to be insufficient to yield impactful results. This is due to limited submissions to the Google/Apple store. Besides, as pointed out in social psychology literature, the reviews are mainly biased by negativity, where positive reviews or fully satisfied users are less likely to post reviews in Google/Apple stores. This stresses the need to find alternative sources where balanced discussions can be found.

Second, the reviews' distribution highlights the dominance of 112-Suomi as compared to other emergency apps in Norway and Sweden. This can be partly explained by the fact that 112-Suomi complies with Pan-European standards and, therefore, is available across all of Europe in principle. However, the discrepancy in the infrastructure quality and load of national emergency centers may also undermine their availability. Besides, Denmark has not supported the mobile emergency app since 2021 and promotes instead the AML service (Advanced Mobile Location) already embedded in many smartphones. Therefore, the study does not fully capture the existing diversity of emergency apps in Nordic countries, but it only provides some contributions to understanding this landscape.

Third, the data processing pipeline, despite its novelty and use of state-of-the-art methods, still relies on the user's input to conduct the initial labeling task, as in web-based DPT annotation in the aspect-sentiment task. Similarly, ontology mapping requires an initial identification of existing ontologies. Therefore, some subjectivity can be claimed if experts would like to question the quality of annotation and choice of related ontologies.

5. Conclusion

This paper explores the usage of mobile emergency communication in three Nordic countries (Finland, Sweden, and Norway), by mapping the content of users' reviews available in Google/Apple store. For this purpose, a three-step strategy has been devised and implemented. First, it uses a new BERT-based aspect sentiment approach to identify factors (aspect terms) associated with emergency apps from the review documents. Second, using the empath-categorization-based approach, it mines the reasons/context behind negative and positive user appreciation. Third, it uses existing ontological vocabularies related to mobile applications and emergency response and management systems to propose a new ontology-based emergency app that best accommodates the context of users' reviews. Overall, this provides relevant insights into understanding users' behavior and motivation regarding the use of emergency apps for communication during emergencies in Nordic countries. The output of this study can be employed to devise a complementary app that comprehends the content of mobile emergency apps and feeds regulators, practitioners, and researchers with alarm messages to act accordingly. It also provides a basis for the subsequent enhancement of existing mobile emergency apps. Furthermore, a similar study can be carried out on other social media platforms such as X (former Twitter), Reddit, Threads, etc. Since the user activity on these social media platforms is high compared to those in the Play Store and App Store, we can intuitively extract much more information from the huge pile of historical data.

The limited sample sizesample was the main drawback of this study. Although a relatively large number of users have uploaded reviews, only a fraction of those are found relevant because of either empty reviews or very short textual descriptions (less than five tokens). This analysis can be further extended in several ways. First, as there is often a gap between online reviews and genuine user opinion, a questionnaire could be devised to gather users' answers and opinions on the use of these mobile emergency applications, which will then be followed by appropriate statistical analysis to identify relevant arguments that support the widespread of these apps as well as their potential barriers. Second, the questionnaire-like study can also provide insights to enhance the data reasoning pipeline employed in this study by revisiting the selection of aspect terms and utilizing some manual input to fine-tune the deep learning models employed in this study. Third, social media platforms can be leveraged to search for extra user feedback regarding the use of mobile emergency apps.

CRediT authorship contribution statement

Fuzel Ahamed Shaik: Writing – original draft, Validation, Software, Investigation. **Mourad Oussalah:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare no conflict of interest of any kinds.

Data availability

Data will be made available on request.

Acknowledgment

This work is funded by EU Chanse program funding for project DiGEMERGE, which is gratefully acknowledged.

References

[1] N.D. Oye, N. A. Iahad, N. Ab. Rahim, The history of UTAUT model and its impact on ICT acceptance and usage by academicians, Educ. Inf. Technol. 19 (2014) 251–270.
[2] W. Bannour, A. Maalel, B.H. Ghezala, Ontology-based representation of crisis response situations, in: Computational Collective Intelligence: 11th International Conference, ICCCI 2019, Hendaye, France, September 4–6, 2019, Proceedings, Part II 11, Springer, 2019, pp. 417–427.
[3] R. Repanovici, S. Nedelcu, L. Tarbă, S. Busuioceanu, Improvement of emergency situation management through an integrated system using mobile alerts, Sustainability 14 (24) (2022) 16424.
[4] A. Kuehn, M. Kaschewsky, A. Kappeler, A. Spichiger, R. Riedl, Interoperability and information brokers in public safety: an approach toward seamless emergency communications, J. Theor. Appl. Electron. Commer. Res. 6 (2011) 43–60.
[5] C.A. Carreras, J. Navarro, C. Sans, A. Zaballos, Communication technologies in emergency situations, Electronics 11 (2022) 1155.
[6] Jon E. Keeley, Hugh Safford, C.J. Fotheringham, Janet Franklin, Max Moritz, The 2007 southern california wildfires: lessons in complexity, J. Forestry 107 (6) (2009) 287–296.
[7] M. Manso, B. Manso, The role of social media in crisis: A European holistic approach to the adoption of online and mobile communications in crisis response and search and rescue efforts, in: Strategic Intelligence Management, Elsevier, 2013, pp. 93–107.
[8] D. Halliwell, C. Lumbreras, Mobile identity platform for the emergency services, 2018, EENA Operations Document. Mobile Identity Report.
[9] R. Repanovici, A. Nedelcu, Mobile emergency notification apps: Current state, barriers and future potential, in: IOP Conference Series: Materials Science and Engineering, vol. 1009, IOP Publishing, 2021, 012049.
[10] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. AL-Smadi, G. Eryiğit, Semeval-2016 task 5: Aspect based sentiment analysis, in: ProWorkshop on Semantic Evaluation, SemEval-2016, Association for Computational Linguistics, 2016, pp. 19–30.
[11] J.S. Dodgson, M. Spackman, A. Pearman, L.D. Phillips, Multi-Criteria Analysis: A Manual, Department for Transport, Local Government and the Regions, HMSO, London, 2009.
[12] M.L. Tan, R. Prasanna, E. Stock, G. Leonard, D. Johnston, Mobile applications in crisis informatics literature: A systematic review, Int. J. Disaster Risk Reduct. 24 (2017) 297–311.
[13] EENA, Emergency appa PEMEA-final PEMEA report, 2020, EENA Operations Document.
[14] Malcolm J. Moseley, Accessibility: The Rural Challenge, Taylor & Francis, 2023.
[15] Minh Hao Nguyen, Eszter Hargittai, Will Marler, Digital inequality in communication during a time of physical distancing: The case of COVID-19, Comput. Hum. Behav. 120 (2021) 106717.
[16] Frantz Rowe, Ojelanki Ngwenyama, Jean-Loup Richet, Contact-tracing apps and alienation in the age of COVID-19, Eur. J. Inf. Syst. 29 (5) (2020) 545–562.
[17] Relebohile Kobeli, Social Media Landscape Report 2023, Oresti Patricios, Ornico, South Africa, 2023.
[18] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent Dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
[19] M. Venugopalan, D. Gupta, An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis, Knowl.-Based Syst. 246 (2022) 108668.
[20] I. Titov, R. McDonald, Modeling online reviews with multi-grain topic models, in: Proceedings of the 17th International Conference on World Wide Web, Association for Computing Machinery, New York, NY, USA, 2008.
[21] M. Hu, B. Liu, Mining and summarizing customer reviews, in: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004, pp. 168–177.
[22] Sadeep Gunathilaka, Nisansa De Silva, Aspect-based sentiment analysis on mobile application reviews, in: 2022 22nd International Conference on Advances in ICT for Emerging Regions, ICTer, IEEE, 2022, pp. 183–188.
[23] D.K. Kirange, R.R. Deshmukh, M.D.K. Kirange, Aspect based sentiment analysis semeval-2014 task 4, Asian J. Comput. Sci. Inf. Technol. (AJCSIT) 4 (2014).
[24] B. Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, vol. 1, Springer, Berlin, 2011.
[25] L. Zhang, B. Liu, Aspect and entity extraction for opinion mining, in: Data Mining and Knowledge Discovery for Big Data: Methodologies, Challenge and Opportunities, Springer, 2014, pp. 1–40.
[26] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, I. An-droutsopoulos, Semeval-2015 task 12: Aspect based sentiment analysis, in: Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval 2015, 2015, pp. 486–495.
[27] S. Polamuri, Five most popular similarity measures implementation in python, 2015, Internet: http://dataaspirant.com/2015/04/11/five-most-popular-similarity-measuresimplementation-in-python.
[28] S. Poria, E. Cambria, L.W. Ku, C. Gui, A. Gelbukh, A rule-based approach to aspect extraction from product reviews, in: Proceedings of the Second Workshop on Natural Language Processing for Social Media, SocialNLP, 2014, pp. 28–37.
[29] A. Mukherjee, B. Liu, Aspect extraction through semi-supervised modeling, in: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2012, pp. 339–348.
[30] R. He, W.S. Lee, H.T. Ng, D. Dahlmeier, An unsupervised neural attention model for aspect extraction, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2017, pp. 388–397.

[31] B. Viver, Emergency communications and the EU legislative framework, 2019, EENA Operations Document.
[32] Ziwen Gao, Zhiyi Li, Jiaying Luo, Xiaolin Li, Short text aspect-based sentiment analysis based on CNN+ BiGRU, Appl. Sci. 12 (5) (2022) 2707.
[33] M. Hoang, A.O. Bihorac, J. Rouces, Aspect-based sentiment analysis using bert, in: Proceedings of the 22nd Nordic Conference on Computational Linguistics, 2019, pp. 187–196.
[34] H. Yang, B. Zeng, J. Yang, Y. Song, R. Xu, A multi-task learning model for chinese-oriented aspect polarity classification and aspect term extraction, Neurocomputing 419 (2021) 344–356.
[35] S. Liu, C. Brewster, D. Shaw, Ontologies for crisis management: a review of state of the art in ontology design and usability, in: Proceedings of the 10th International ISCRAM Conference, Baden, Germany, 2013, pp. 349–359.
[36] E. Kontopoulos, P. Mitzias, J. Moßgraber, P. Hertweck, H. van der Schaaf, D. Hilbring, I. Kompatsiaris, Ontology-based representation of crisis management procedures for climate events, in: ISCRAM, 2018.
[37] K. Bitencourt, F.A. Durão, M. Mendonça, L.L.B.D.S. Santana, An ontological model for fire emergency situations, IEICE Trans. Inf. Syst. E101.D (2018) 108–115.
[38] P.D. Haghighi, F. Burstein, A. Zaslavsky, P. Arbon, Development and evaluation of ontology for intelligent decision support in medical emergency management for mass gatherings, Decis. Support Syst. 54 (2013) 1192–1204.
[39] M. Limbu, T. Wang, J. Ortmann, Management of a crisis (MOAC) vocabulary specification, 2018, Link: http://observedchange.com/moac/ns/.
[40] G. Babitski, S. Bergweiler, O. Grebner, D. Oberle, H. Paulheim, F. Probst, SoKNOS–using semantic technologies in disaster management software, in: Extended Semantic Web Conference, Springer, 2011, pp. 183–197.
[41] A. Wu, X. Yan, E. Kuligowski, R. Lovreglio, D. Nilsson, T.J. Cova, X. Zhao, Wildfire evacuation decision modeling using GPS data, Int. J. Disaster Risk Reduct. 83 (2022) 103373.
[42] H. Baytiyeh, The uses of mobile technologies in the aftermath of terrorist attacks among low socioeconomic populations, Int. J. Disaster Risk Reduct. 28 (2018) 739–747.
[43] S.H. Othman, G. Beydoun, V. Sugumaran, Development and validation of a disaster management metamodel (DMM), Inf. Process. Manage. 50 (2014) 235–271.
[44] A. Braham, F. Buendía, M. Khemaja, F. Gargouri, User interface design patterns and ontology models for adaptive mobile applications, Pers. Ubiquitous Comput. 26 (2021) 1–17.
[45] F.A. Norki, R. Mohamad, N. Ibrahim, Context ontology in mobile applications, J. Inf. Commun. Technol. 19 (2020) 21–44.
[46] M. Poveda Villalon, M.C. Suárez-Figueroa, R. García-Castro, A. Gómez-Pérez, A context ontology for mobile environments, in: ProceedingsofWorkshopon-Context,Informationand Ontologies — CIAO 2010 Co-Located with EKAW 2010, vol. 626, CEUR-WS, Germany, 2010.
[47] X.H. Wang, D.Q. Zhang, T. Gu, H.K. Pung, Ontology based context modeling and reasoning using OWL, in: IEEE Annual Conference on Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second, Ieee, 2004, pp. 18–22.
[48] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1, 2019, pp. 4171–4186.
[49] H. Yang, K. Li, PyABSA: open framework for aspect-based sentiment analysis, 2022, arXiv preprint arXiv:2208.01368.
[50] P. He, X. Liu, J. Gao, W. Chen, Deberta: Decoding-enhanced bert with disentangled attention, 2020, arXiv preprint arXiv:2006.03654.
[51] N. Arhab, M. Oussalah, Y. Bounab, Analysis of user's car parking behaviour through twitter hashtags, Expert Syst. 40 (2023) e13218.
[52] E. Fast, B. Chen, M.S. Bernstein, Empath: Understanding topic signals in large-scale text, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2016, pp. 4647–4657.
[53] P. Rozin, E.B. Royzman, Negativity bias, negativity dominance, and contagion, Pers. Soc. Psychol. Rev. 5 (4) (2001) 296–320.
[54] Yin Zhang, Jinyang Du, Xiao Ma, Haoyu Wen, Giancarlo Fortino, Aspect-based sentiment analysis for user reviews, Cogn. Comput. 13 (5) (2021) 1114–1127.
[55] "Low Battery Anxiety" Grips 9 Out of Ten People, Technical Report, 2016.
[56] Muhammad Nazrul Islam, Iyolita Islam, Kazi Md Munim, AKM Najmul Islam, A review on the mobile applications developed for COVID-19: an exploratory analysis, IEEE Access 8 (2020) 145601–145610.
[57] Aryamala Prasad, Mila Gasco-Hernandez, J. Ramon Gil-Garcia, Qianli Yuan, Human-centric design in applications for emergency preparedness and response in rural communities: The case of the e!app, in: Proceedings of the 24th Annual International Conference on Digital Government Research, Association for Computing Machinery, New York, NY, USA, 2023, pp. 380–387.
[58] Marion Lara Tan, Raj Prasanna, Kristin Stock, Emma E.H. Doyle, Graham Leonard, David Johnston, Modified usability framework for disaster apps: a qualitative thematic analysis of user reviews, Int. J. Disaster Risk Sci. 11 (2020) 615–629.
[59] M.A. Musen, The protégé project: A look back and a look forward. AI matters, Assoc. Comput. Mach. Specif. Interest Group Artif. Intell. 1 (4) (2015).