Development of IoT Mining Machine for Twitter Sentiment Analysis:

Mining in the Cloud and Results on the Mirror

Salha M. Alzahrani
Vice-Dean of Graduate Studies in Taif University
Assoc. Prof. of Computer Science, College of Computers and Information Technology, Taif University
Taif, Saudi Arabia
s.zahrani@tu.edu.sa

***Abstract*— Microblogs sentiment analysis of people's attitudes, appraisals and emotions has become one of the most active research areas for business marketing, decision making, political campaigns, and alike. As people publish short snippets of texts through the social networks expressing their ideas, thoughts and opinions, an instant and reliable mining machine should be utilized. In this paper, we proposed an IoT mining machine for Twitter sentiment analysis. Firstly, we used Twitter's API for harvesting tweets in real time. Then, a mining engine was developed on the Raspberry Pi single-board microcomputer as an IoT platform due to its availability and connectivity. The IoT device was programmed for sentiment analysis and opinion mining using state-of-the-art Naïve Bayes classifier which after training was used to classify the trending tweets into either positive or negative. We used a gold standard dataset from SemEval 2017 for training our classifier which achieved 0.992 of accuracy. We aggregated the sentiments of tweets streamed in daily trend hashtags into visualized graphs. Finally, the visualized results from opinion mining were displayed on two-way smart mirror without any need for application installment. Our experimental results on the IoT mining machine demonstrate its feasibility and effectiveness.**

Keywords—IoT; Sentiment Analysis; Opinion Mining; Twitter; Social Internet of Things

I. Introduction

At glance, one may ask what does the "Internet of Things" even mean, and why should we care about? The Internet of Things, shortly IoT, is all about data. Before going to a formal definition, let us start with a simple example which indicates the existence of IoT and its importance in our daily life. Suppose that a company chain needs to hire hundreds of employees to count its stock manually which costs a lot. The company instead may hire very few (e.g. ten) employees if an IoT solution is implemented for automatically counting the stock and sending notifications when specific items are out of stock. This scenario would positively influence the company to save efforts and money as it will not have to pay for employees where technology can replace them. Another scenario is using IoT for smart home automation where lights can be adjusted automatically according to the outside brightness, air conditioning machines can be adjusted automatically according to the outside temperature, doors can be automatically locked/unlocked according to comers' face recognition, and many more.

From the above scenarios, IoT is the evolution of things as embedded devices that are connected to the Internet and speak to each other. Things can be cars, homes, appliances, refrigerators, air conditioners, shelves, coffee machines, spoons, and even our bodies. In IoT systems, each device (i.e. thing) is assigned a unique identifier (IP address) and therefore has the ability to transfer/receive data over a network without requiring human-to-human or human-to-computer intervention.

Nowadays, using IoT offers many ways to solve our daily life problems and challenges. A study surveyed the use of different types of sensors for soil, smoke, water, temperature, etc. for the development and implementation of IoT applications [1]. Other studies include the utilization of IoT for social data exchange and analytics, and incorporate IoT for social networks applications [2-4].

Thus and more formally, IoT is all about data exchange, processing, and acting accordingly. Data comes in many formats from sensor readings, input instructions, trigger commands, to human microblogs and business information. Consequently, the area of data analysis is too broad. Data mining is a major one which can be defined as the process of extracting valid, previously unknown, and comprehensible information from digitized data in order to improve and optimize decision-making [5, 6]. No matter how powerful the data mining algorithm is, the resulting model will not be valid if the data are not selected and pre-processed correctly [5]. As a major subfield of data mining, dealing with text data collections such as documents, web pages, articles, abstracts, news, tweets is called *text mining* [7, 8]. Text Mining is also related to other research fields, including Machine Learning (ML), Information Retrieval (IR), Natural Language Processing (NLP), Information Extraction (IE), Statistics, Pattern Recognition (PR), and Artificial Intelligence (AI).

There are many data mining algorithms that have been studied for text analytics [9, 10]; for example, the use of

978-1-5386-4817-9/18/$31.00 ©2018 IEEE

sentiment classification in microblogs [11] as will be explained shortly. Approaches for characterization, clustering, classification, association, and prediction have been applied for unstructured textual mining widely. Text pre-processing and cleansing are the first steps which strongly affect the outcome of text mining algorithms. Pre-processing of texts includes sentence and word tokenisation, and dealing with hyphens, diacritics if any, punctuations, and numbers. Other pre-processing methods include stop words removal such as the, a, an., etc., and part of speech tagging, knowledge-based lemmatisation, and domain-based feature extraction. Applications of text mining are numerous in the biomedical domain, business intelligence domain, financial domain and many more [12-14].

Sentiment analysis, also referred to as opinion mining, denotes the computerised techniques that can be used to analyse and predict someone's opinion whether it is positive or negative, agree or disagree, towards or against [15]. Due to the growth of social networks, sentiment analysis has become one of the most active research areas for business marketing decisions, political campaigns, and even people's attitudes, emotions and appraisals. A study, for instance; proposed a method for opinion mining of coffee service quality and customer satisfaction by mining tweets reported by their customers [16]. Typical research problems in sentiment analysis include: (i) feature extraction and identification of textual data elements which contain user opinions [17]; (ii) sentiment classification and prediction of users' attitudes and sentiments (e.g., positive or negative) [16, 18, 19]; (iii) opinion aggregation and visualisation of a condensed set of predicted user opinions; and (iv) automatic analysis of opinions, emotions, and subjectivity associated with a topic in microblog posts [18, 19]. Such techniques developed for microblog sentiment analysis can also be applied to classify social media data in a real-time manner.

Therefore, this paper aims to investigate the use of IoT for data mining problems; specifically, sentiment analysis and opinion mining. Consider the scenario when people every day look at Twitter hashtags raised for different events. People read lots of tweets in each hashtag to grasp its content and know what it is all about. In the opposite side, imagine a device which automatically harvest tweets for each event (i.e. hashtag), download hundreds of tweets, analyse the content, and present you with a summary in a bar graph for example, or even visualise the people's sentiments about that event. Furthermore, imagine if that device not your mobile phone but instead is a smart mirror which displays the events' summary and people's sentiments daily without the need to spend efforts or use programming tools, or otherwise buy apps to do so. That is, in this paper we proposed the use of IoT mining machine namely Raspberry Pi for sentiment analysis and visualization of results on a smart mirror connected to our IoT device.

The rest of the paper is organized as follows. Section II presents an overview of related works in two sides: one side is for sentiment analysis methods for social networks such as Twitter, and the other side is for studies that comprises IoT with social networks. Section III shows the methodology carried out for developing an IoT device with mining capabilities. Section IV demonstrates the experimental works for sentiment analysis on Raspberry Pi, how to stream tweets, and use NLTK built-in classifiers such as Naïve Bayes classifier to predict tweets as positive and negative; and visualise the results on smart mirror which can be used at homes, companies, etc. Section V gives the concluding remarks and future works to accomplish this research and open new directions for social IoT.

II. Related Works

With the appearance of Web 2, people nowadays are authoring the web. Bloggers have appeared to write long textual entries which we called blogs. Another concept of writing microblogs, small snippets of texts, has become dominant such as using Twitter, Facebook statuses, and alike. People have become micro-bloggers as they like to express their opinions and attitudes about daily life events in short phrases. Due to its instantaneous nature, "Twitter and other microblogging platforms have been widely employed for political and marketing campaigns, tracking of emergencies, opinion surveys, live news reporting, etc" [20].

Microblog sentiment detection is however quite challenging for several reasons. Primarily, microblog posts are normally very short, in Twitter it does not exceed 140 characters, and people often use shortcuts, irregular words and emojis which impose the difficulty of understanding the text by machines. Another reason is that the many micro-bloggers aim to establish their unique wiring style and hence it is difficult to come up with a list of frequent word sets to analyse their behaviour and know their opinions. More importantly, microblogs are very dynamic and users establish/break links between them easily which require efficient scalable solutions. Due to these challenges, the role of pre-processing in microblogs sentiment analysis is crucial. A study shows that preserving emotions and linguistic features improve the accuracy of sentiments classification [21].

A. Sentiment Analysis in Social Netowrks

Several research studies have been conducted for sentiment analysis in social networks as massive platforms for people to express their opinions, feedbacks, and reviews. A study [22] has shed light on combining qualitative and quantitative analysis of texts from opinion forums which improve product reviews and help companies to improve their services, products, etc. Different research studies for opinion mining have mainly reported this problem as a classification task; for example, a study worked for reporting controversial events from Twitter using three different classifiers and reported encouraging results [23], a study used binary classifier for sentiment analysis in Twitter [24], another study investigated exploiting data for choosing the best classifier model for sentiment analysis for Spanish words and reported their results [25]. Two classifiers were designed based on Naive-Bayes and

Maximum Entropy classifiers, and their accuracies were compared on different feature sets for sentiment analysis of Twitter feeds in [26]. Another research paper used support vector machines for experimental works in opinion mining [30].

Research works have investigated the use of emotions and visual multimedia information posted in Twitter for sentiment analysis [27, 28]. Agave, a collaborative visual analysis system for exploring events and sentiment over time in large tweet data sets was developed and evaluated [28]. Another study [29] explored unsupervised sentiment analysis in Twitter, MySpace, and Digg social media in contrast to supervised machine learning methods applied normally for these problems. The study showed that though unsupervised learning has been implemented, it showed robust and encouraging results with social texts compared with supervised domain-specific learning methods. Another direction of research [30] has not only employed language features but also incorporated user-level approaches that are induced either from the Twitter follower/followee network or from the network in Twitter formed by users referring to each other using "@" mentions. Their results indicated that incorporating user network data yielded statistically significant sentiment classification enhancement. Applications of mining tweets have included comparison of sentiments of users over time such as the case of the Boston Bombing Tragedy [31], and determining the popularity of city locations based on users tweets [32].

Successful sentiment analysis tasks have been conducted through SemEval (Semantic Evaluation) conference series. In this regard, contemporary computational linguistics and natural language processing techniques have been widely proposed for sentiment analysis with the development of rich data sets from tweets, SMS messages, LiveJournal messages, and a special test set of sarcastic tweets [33]. The task attracted more than 40 group of participants every year who used various computational techniques. The winner teams have developed sentiment analysis techniques which outperformed the baselines by substantial margins.

Nevertheless, sentiment is often implicitly expressed via latent semantic relations, patterns and dependencies among words in tweets. Therefore, advanced statistical and semantic language models have been developed for sentiment analysis. For example, an approach of adding semantics as additional features into the training set reported supreme improvements of F harmonic accuracy score around 6.5% and 4.8% over the baselines of unigrams and part-of-speech features respectively [34]. The experiments were done on three different Twitter data sets and compared with sentiment-bearing topic analysis and found that semantic features produced better results for sentiment classification. Unlike most existing approaches which consider that opinions always expressed by explicit words, another study focused on exploiting semantic patterns for Twitter sentiment analysis [35]. The study used semantic patterns on tweet level and entity level and evaluated the approach on nine datasets and showed an encouraging improvement of accuracy compared with six state-of-the-art baselines. A recent study investigated the use of statistical models built based on n-gram language models for sentiment analysis of Twitter posts [36]. In this study, n-gram models were built for positive and negative tweets and the polarity of a given tweet was defined by the closest perplexity with these models.

B. Social Internet of Things

Although of the huge increment of research body in social networks analysis and the fast advancements in IoT technologies, less attention have been given for designing the social internet of things [2]. This trend will give people more relatedness with the “things”, and on the other hand, IoT things will serve socially better with more creativity, engagement, relatedness, participation with people. Another challenge is achieving autonomicity in IoT [37]. This trend aims to develop IoT autonomic devices able to manage themselves by considering situational-awareness, knowledge, smartness and social behaviour. Things will evolve and act in a more autonomous way, becoming more reliable and smarter. A third trend is to connect places and things to online communities to enable more interactions between not only the people themselves but also between people and their things like preferred cafes [38]. A fourth trend in social IoT is to design networked things to independently communicate with each other. A study by [39] developed a so-called Morse things, a set of ceramic bowls and cups networked together to autonomously interact with each other using IoT. The study claimed that “we reflect on the nature of living with IoT things and discuss insights into the gap between things and humans that led to the idea of a new type of thing in the home that is neither human-centred technology nor non-digital artifacts”. Similarly, another study [4] introduces the concept of "Social Thing" where things have unique personalities and autonomously build network to cooperate with each other as if they are living entities. Social things can autonomously exchange information about each other and take actions accordingly.

Recent yet important trend in social IoT is the utilisation of social “big” data through IoT devices. For instance, a study explored people recommendations using IoT platforms [40]. Their study proposed a unified probabilistic based framework by fusing information across relationships between users (i.e., users' social network) and things (i.e., things correlations) to make more accurate recommendations. The proposed approach not only inherits the advantages of the matrix factorization, but also exploits the merits of social relationships and thing-thing correlations. Another study was directed to the same aspects of empowering the social IoT paradigm which combined and merged people and devices (i.e. things) allowing people and connected devices as well as the devices themselves to interact within a social network framework [3]. The study focused on creating social rules using ontology models for social relationships between the people and the things.

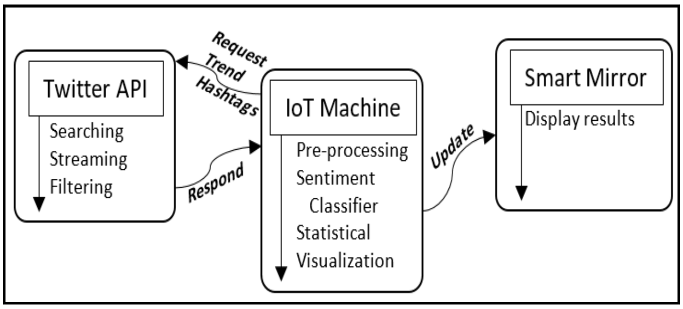


Fig. 1. General methodology framework for IoT mining machine proposed in this research

III. METHODOLOGY

This work sheds light on two different aspects: one is the sheer scale of interaction with things at home [41], and the other aspect is the increase need of social microblogs analysis of sentiments [18]. That is, the methodology of this study incorporates the field of sentiment analysis and the field of IoT. Fig. 1 shows the general methodology framework used in this research.

Below we explained the Raspberry Pi characteristics as an IoT platform, the interaction with online services using IoT platform, and the utilization of Twitter's API for searching hashtags and streaming tweets. Then, we discussed the methods employed for opinion mining of tweets and results visualization by the IoT with a two-way smart mirror which will show events' summary and people's sentiments daily without the need to spend efforts with programming tools, or otherwise buy apps to do so.

A. *Raspberry Pi*

An embedded system can be thought of as a computer hardware system having software embedded in it. An embedded system can be an independent system or it can be a part of a large system, wherein a microcontroller or microprocessor is designed to perform a specific task [42]. Examples of emended systems are those we can see in the washing machines, automatic vending machines, and so on. Once any embedded system is connected to the Internet, we would say that an IoT device is established for data transfer and control via the Internet. A good platform for developing IoT platforms is *Raspberry Pi*; a microcomputer which can provide a cloud storage sever with internet connectivity [43]. Raspberry Pi has many features compared with other platforms such as Arduino, UDOO, Banana Pi, and Pc Duino. These features include fast processing using the Raspbian OS, ability to programming using Python, embedded Wi-Fi for the Internet, feasible data processing, ability to use HDMI port, and many more. In this work, we used Raspberry Pi 3 Model B as a single-board computer with wireless LAN and Bluetooth connectivity, which is shown in Fig. 1.

Fig. 2. Raspberry Pi 3 Model B as a single-board computer with wireless LAN and Bluetooth connectivity

B. *Interaction with Online Services*

Networking libraries are provided on the Raspberry Pi. For example, the *socket* and *SocketServer* are low-level networking libraries in Python whereby the programmer composes the message content and the details of the protocol. Furthermore, there are other protocol-specific libraries whereby the programmer does not need to compose the protocol details as the built-in functions can be used to handle the connection and protocol details such as *http*, *http.client*.

These networking libraries give the capabilities to the Raspberry Pi for interaction programmatically with online web services such as Facebook, Twitter, Google Maps, and YouTube. Such web-based services run on the cloud (i.e. remote server) and accept requests from clients such as the Raspberry Pi mining machine proposed for this research. Interaction can be made via HTTP messages and an Application Programming Interface (API) between the programs. We used API to define the format of the request/response messages and the messages sequences; for example,

```
GET /maps/<web_address> HTTP/1.1
```

C. *Using Twitter API on Rapberry Pi*

There are tools such as *AlchemyAPI* SDK[1] developed by IBM for machine learning, language understanding and can be certainly used for sentiment analysis. Another Twitter's API tools which work on Raspberry Pi include *Tweepy* and *Twython* which allow the Raspberry Pi to send tweets, respond to the tweets, and look for a hashtag. First of all, we registered our app to access Twitter's API and received several keys needed for authentication from Twitter. We decided to use in this work. We used the registered keys to create a Twython object to transmit messages. For example, sending a tweet from the Raspberry Pi can be easily performed as follows:

```
from twython import Twython
twitter   =   Twython   (APP_KEY,   APP_SECRET,
AUTH_TOKEN, OAUTH_TOKEN_SECRET)
twitter.update_status(status='My First Tweet!')
```

D. *Searching and Filtering Tweets on Rapberry Pi*

Our IoT mining machine was programmed to react to trend hashtags. So, when a hashtag appears, scan the Twitter stream

[1] https://www.ibm.com/watson/alchemy-api.html

for the hashtag and download tweets for further analysis. Using the same keys with Twython library, we can search for a hashtag using the following code:

```
results = twitter.cursor(twitter.search,
q='python')
for result in results:
    print result[‘text’]
```

We also streamed tweets to track a specific hashtag which we intend to analyse as follows:

```
class MyStreamer(TwythonStreamer):
    def on_success(self, data):
        if 'text' in data:
            print data['text'].encode('utf-8')
    def on_error(self, status_code, data):
        print status_code

stream = MyStreamer(APP_KEY, APP_SECRET,
        OAUTH_TOKEN, OAUTH_TOKEN_SECRET)
stream.statuses.filter(track='twitter')
```

In this work, we proposed to analyse global trend hashtags every 12 hours based on the location ID for Saudi Arabia[2]. We proposed to analyse the top five hashtags. The following code shows how to retrieve the hashtags appeared in the trend. Then, we streamed the top 2000 tweets appeared in each hashtag using the techniques described earlier in this section.

```
results = twitter.trends.place(_id = 23424938)
for location in results:
      for trend in location["trends"]:
               print " - %s" % trend["name"]
```

E. Opinion Mining of Tweets

As we mentioned earlier, sentiment analysis is the process of automatically and computationally determining whether a microblog post is positive, negative or neutral using machine learning classification techniques. Literature studies have used neural networks [44], SVM (Support Vector Machine) [45], Naive Bayes classification techniques [45] for opinion mining. In this paper, we proposed to use a simplest but effective Naïve Bayes classifier using NLTK (Natural Language Toolkit) [46] following to the research work done by Goel *et al.* (2016) [47].

The architectural methodology used for Twitter sentiment analysis in our IoT mining machine is summarized in Fig. 3. As can be seen in the Figure, the methodology starts by crawling the tweets in the top ten trending hashtags. We used trending hashtags in Arab region and focused on tags in Arabic. We constructed the tweets dataset which go further under preprocessing steps such as text cleaning, sentence splitting, word tokenization and finally feature identification. From the same dataset, we constructed the list of positive and negative words in Arabic. The dataset was then divided into training dataset and test dataset, and Naïve Bayes classifier was trained to classify tweets into either positive and negative as we will explain shortly. After training, the classifier is ready to predict the opinion in tweets appeared in trending hashtags and generate summary of opinions using our IoT mining machine.

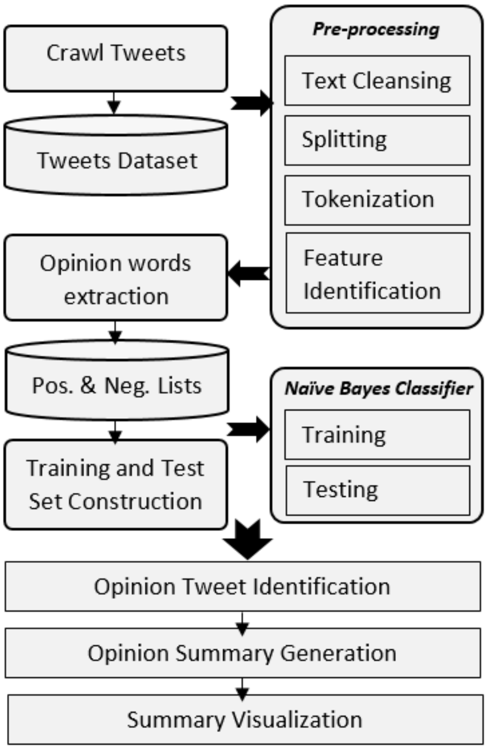


Fig. 3. Architectural methodology used for Twitter sentiment analysis

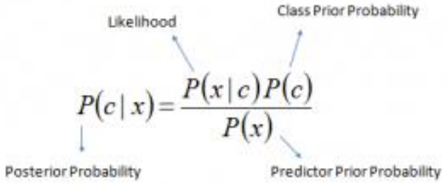
Naïve Bayes classifier is defined as “is a method of classification that does not use rules, a decision tree or any other explicit representation of the classifier. It uses the probability theory to find the most likely classification of an unseen (unclassified) instance” [6]. The Naïve Bayes classifier combines two probabilities for each event (i.e. class): the prior probability and posterior probabilities. The prior probability is calculated using the following formula:

$$\boldsymbol{P}(\boldsymbol{class} = \boldsymbol{c}) = (\boldsymbol{frequency\ of\ c})/(\boldsymbol{total\ number\ of\ instances})$$

The posterior probability is calculated given an additional information:

$$\boldsymbol{P}(\boldsymbol{class}=\boldsymbol{c} \mid \boldsymbol{a}=\boldsymbol{v}) = (\boldsymbol{frequency\ of\ c\ given\ attribute\ a})/(\boldsymbol{total\ number\ of\ instances\ in\ a})$$

However, the posterior probabilities are computed the other way round as follows: $\boldsymbol{P}(\boldsymbol{a}=\boldsymbol{v} \mid \boldsymbol{class}=\boldsymbol{c})$ and combined with the prior probability in a single formula as follows:



$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

[2] https://dev.twitter.com/rest/reference/get/trends/place

In this work, our text classifier is trained on labelled dataset with positive and negative *tweets*. In order to create the classifier, we used the following steps:

First, the list of distinct words and their frequency (i.e. number of times appeared in the dataset tweets) were extracted and arranged ascendingly.

Second, words with the highest frequency were used as the *word_features* list.

Third, for each input tweet we can extract the features based on the *word_features* list using *feature_extractor* method. When a word in the input appeared in the *word_features* list, it takes 1. Otherwise, it takes 0.

Fourth, textual features from all tweets in the training set were applied using NLTK tool as below.

```
training_set    =    nltk.classify.apply_features
(feature_extractor, tweets)
```

Fifth, the classifier was then trained on the training set using NLTK tool as follows:

```
classifier    =    nltk.NaiveBayesClassifier.train
(training_set)
```

Sixth, the accuracy of our classifier was computed using the test dataset as follows:

```
nltk.classify.accuracy(classifier, test_set)
```

Seventh, we can use our classifier to predict the sentiment in any new tweet; thence, in the list of tweets appearing in the hashtags as below:

```
Class = classifier.classify  (feature_extractor
(tweet))
```

Finally, we aggregate the tweets that are classified as positive and those as negative to show the results summary of sentiments in the in the trending hashtags as below.

```
#trending _hastag_x

Tweets = list of crawled tweets appeared in
        #trending _hastag_x
For each tweet in Tweets:
    Class = classifier.classify
    (feature_extractor    (tweet).split(‘ ‘)

    if Class == ‘Positive’:
          positive_tweets.append(tweet)
    else :
          negative_tweets.append(tweet)
Positivity = len (positive_tweets)

Negativity = len (negative_tweets)
```

where `len` denotes the length of the list (i.e. number of items in the list) as in *Python* language.

IV. Experiemntal Results

In this section, we presented the experimental works we have done to accomplish this research. We describe the hashtag and dataset description in Section A, and samples of results from hashtag streaming and filtering is shown in Section B. Section C shows results from opinion mining and Section D shows the visualisation results of opinion's summary as proposed to appear in the IoT smart mirror.

A. Dataset Description

In this paper, we used the dataset for Arabic tweets released recently by SEM-Eval 2017 [48]. The dataset consists of 25540 words in a total of 1656 tweets. The sentiment types distribution in this dataset is either positive or negative as can be seen in Table I. We divided the dataset into two parts: training dataset with 999 tweets, and test dataset with 657 tweets. Table II shows sentiment types distribution in the training and test datasets. The dataset contains 31 different topics; each contains tweets annotated by positive or negative sentiment, as summarised in Table III.

TABLE I. Sentiment types distribution in the dataset

Dataset	**Sentiments Distribution**		
	Positive	***Negative***	***Total***
Arabic Tweets Dataset realsed by SEM-Eval 2017	885	771	1656

TABLE II. Sentiment types distribution in the Training and Test datasets

Dataset	**Sentiments Distribution**		
	Positive	***Negative***	***Total***
Training Dataset	460	539	999
Test Dataset	425	232	657

TABLE III. Tweets distribution by topic and sentiment

Topic	**Sentiments Distribution**		
	Positive	***Negative***	***Total***
ابل	33	18	51
اندرويد	40	4	44
اسرائيل	3	60	63
ايران	4	59	63
اردوغان	12	46	58
الارهاب	0	96	96
الاسلام	69	5	74
السعودية	16	20	36
العراق	8	17	25
امازون	33	2	35
باراك أوباما/اوباما	16	41	57
ايفون	44	3	47
برشلونة	18	7	25
بشار الأسد	5	35	40
بوكيمون	24	10	34
بيونسيه	65	3	68
جستن بيبر	37	11	48
جوجل/ غوغل	65	13	78
حلب	4	20	24
داعش	0	93	93
دونالد ترامب	5	50	55
رمضان	64	0	64
ريال مدريد	24	4	28
سوريا/سوريه	9	23	32
سيسي	6	79	85
غوتشي	53	3	56
فيدرر	78	2	80
ميسي	60	9	69
هاري بوتر	49	4	53
هيلاري كلنتون	17	26	43
ويندوز ١٠	24	8	32

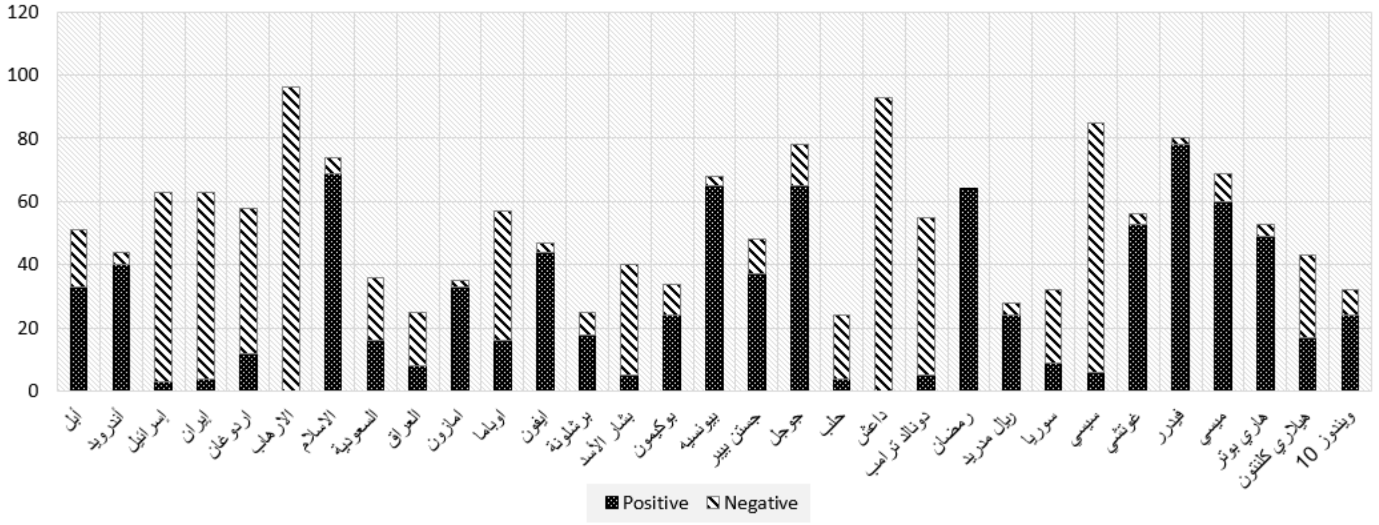


Fig. 4. Topics included in the Arabic Tweets Dataset released by SEM-Eval 2017 and percentages of positive and negative tweets in each topic

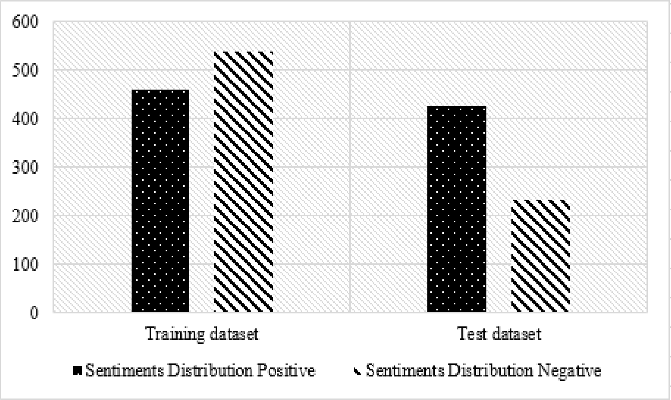


Fig. 5. Distribution of Postive and Negative Tweets in the Training and Test Datasets

Topics included in the Arabic Tweets Dataset released by SEM-Eval 2017 are numerous but basically includes up-to-date topics varies from technological side; for example, apple, iPhone, Windows 10, and religious and political side such as Obama, Trump, terrorist, Islam, and so on. A visualization of each topic and percentages of positive and negative tweets in each topic is demonstrated in Fig. 4. The totals of tweets in the training and test datasets are shown in Fig. 5.

B. Results from Hash Tag Streaming and Filtering

We followed the steps explained in Section III part D to generate the list of hashtag trends, and then crawl tweets in each hashtag. In our IoT mining machine we set a timer to gather trending hashtags every half day. Fig. 6 shows and example of universal hashtags trending on 30 Oct. 2017, and Fig. 7 shows another hashtags trending on 13 Dec. 2017 determined in Saudi Arabia. In this work, we focused on trending hashtags by Saudi Arabia region, and crawled Arabic tweets. In each trending hashtag, we gathered more than one thousand top tweets and entered them into a cleansing process including removing mentions, URLs, numbers, special characters, etc. For the cleaning purpose, we employed *TwitterCleanup* library in Python which obtained clean tweets for further classification and summary generation as we will explain shortly.

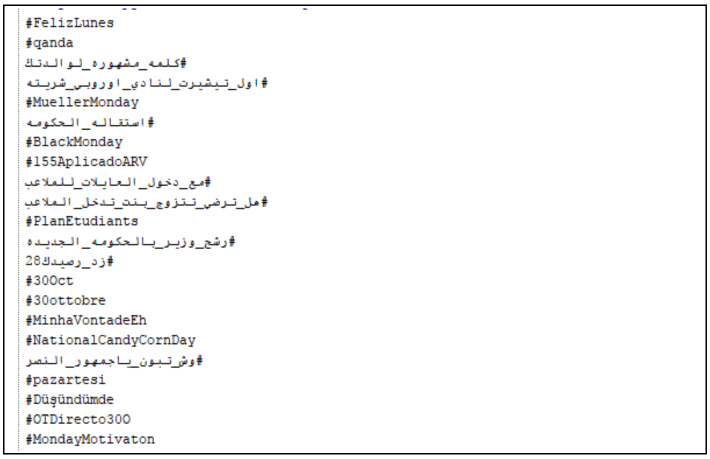

```
#FelizLunes
#qanda
#كلمه_مشهوره_لوالدتك
#اول_تيشيرت_لنادي_اوروبي_شريته
#MuellerMonday
#استقاله_الحكومه
#BlackMonday
#155AplicadoARV
#مع_دخول_العايلات_للملاعب
#هل_ترضي_تتزوج_بنت_تدخل_الملاعب
#PlanEtudiants
#رشح_وزير_بالحكومه_الجديده
#زد_رصيدك28
#30Oct
#30ottobre
#MinhaVontadeEh
#NationalCandyCornDay
#وش_تبون_ياجمهور_النصر
#pazartesi
#Düşündümde
#OTDirecto300
#MondayMotivaton
```

Fig. 6. Sample of universal hashtag list obtained on 30 Oct. 2017

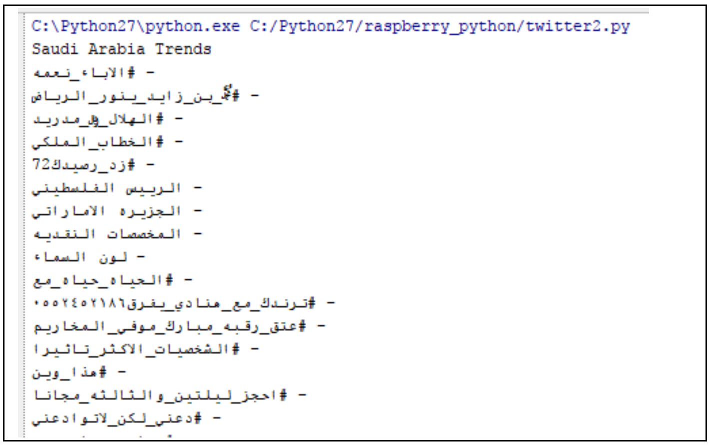

```
C:\Python27\python.exe C:/Python27/raspberry_python/twitter2.py
Saudi Arabia Trends
- #الاباء_نعمه
- #محمد_بن_زايد_ينور_الرياض
- #الهلال_ريال_مدريد
- #الخطاب_الملكي
- #زد_رصيدك72
- الرييس الفلسطيني
- الجزيره الاماراتي
- المخصصات النقديه
- لون السماء
- #الحياه_حياه_مع
- #ترندك_مع_منادي_يفرق٠٥٥٢٤٥٢١٨٦
- #عتق_رقبه_مبارك_موفي_المخاريم
- #الشخصيات_الاكثر_تاثيرا
- #هذا_وين
- #احجز_ليلتين_والثالثه_مجانا
- #دعني_لكن_لاتوادعني
```

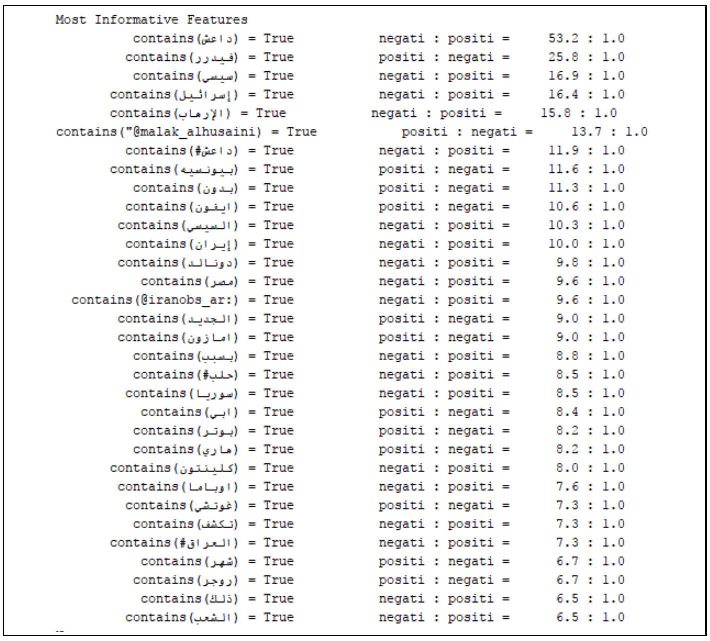
Fig. 7. Sample of universal hashtag list obtained on 13 Dec. 2017

C. *Results from Opinion Mining*

We started our experimental works by training Naïve Bayes classifier of positive and negative sentiments. In order to train the classifier, the wordlist of distinct words and their frequencies have been built. Fig. 8 shows the resulting most informative features in our dataset. The accuracy of the classifier using the training and test sets was 0.992.

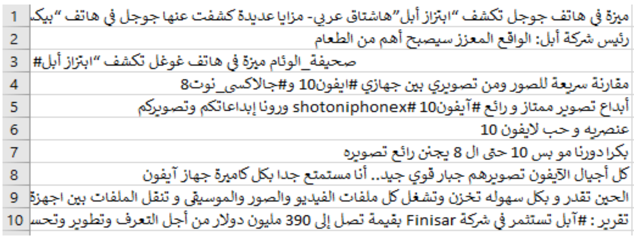
D. *Results Visulation using IoT*

For experimental works on our IoT machine, we streamed tweets from the hashtag #iPhone10; in Arabic ايفون١٠#. After cleaning, we used around 300 tweets for opinion prediction. Sample of the resulting tweets are shown in Fig. 9. Each tweet was predicted to be either positive or negative. Because among of the topics in the training set were Apple, iPhone, etc., we expected the sentiment analyzer to work well with the tweets appeared in this hashtag. We proposed a pie graph view showing the percentages of the predicted sentiments as iPhone 10. As demonstrated in Fig. 10, 92% of the tweets shows positive opinion regarding "ايفون١٠#". By manually analyzing the tweets, people expressed their opinion regarding different features such as the camera, color, battery, the home button, face ID, etc. Their expressions include for instance; ممتاز ورائع مدة الشحن تحسنت بشكل كبير ,ابداع في اللقطات ,أهم من الطعام ,للتصوير.



```
Most Informative Features
          contains(داعش) = True          negati : positi =     53.2 : 1.0
         contains(فيدرر) = True          positi : negati =     25.8 : 1.0
          contains(سيسي) = True          negati : positi =     16.9 : 1.0
       contains(إسرائيل) = True          negati : positi =     16.4 : 1.0
       contains(الإرهاب) = True         negati : positi =     15.8 : 1.0
contains("@malak_alhusaini) = True          positi : negati =     13.7 : 1.0
         contains(#داعش) = True          negati : positi =     11.9 : 1.0
       contains(بيونسيه) = True          positi : negati =     11.6 : 1.0
          contains(بدون) = True          positi : negati =     11.3 : 1.0
         contains(ايفون) = True          positi : negati =     10.6 : 1.0
        contains(السيسي) = True          negati : positi =     10.3 : 1.0
         contains(إيران) = True          negati : positi =     10.0 : 1.0
        contains(دونالد) = True          negati : positi =      9.8 : 1.0
           contains(مصر) = True          negati : positi =      9.6 : 1.0
  contains(@iranobs_ar:) = True          negati : positi =      9.6 : 1.0
        contains(الجديد) = True          positi : negati =      9.0 : 1.0
        contains(امازون) = True          positi : negati =      9.0 : 1.0
          contains(بسبب) = True          negati : positi =      8.8 : 1.0
          contains(#حلب) = True          negati : positi =      8.5 : 1.0
         contains(سوريا) = True          negati : positi =      8.5 : 1.0
           contains(ابي) = True          positi : negati =      8.4 : 1.0
          contains(بوتر) = True          positi : negati =      8.2 : 1.0
          contains(هاري) = True          positi : negati =      8.2 : 1.0
       contains(كلينتون) = True          negati : positi =      8.0 : 1.0
         contains(اوباما) = True          negati : positi =      7.6 : 1.0
         contains(غوتشي) = True          positi : negati =      7.3 : 1.0
          contains(تكشف) = True          negati : positi =      7.3 : 1.0
       contains(#العراق) = True          negati : positi =      7.3 : 1.0
           contains(شهر) = True          positi : negati =      6.7 : 1.0
          contains(روجر) = True          positi : negati =      6.7 : 1.0
           contains(ذلك) = True          negati : positi =      6.5 : 1.0
         contains(الشعب) = True          negati : positi =      6.5 : 1.0
```

Fig. 8. Example of two-way smart mirror for displaying results from tweets sentiment analysis and other IoT data



1	ميزة في هاتف جوجل تكشف "ابتزاز أبل"هاشتاق عربي- مزايا عديدة كشفت عنها جوجل في هاتف "بيكس
2	رئيس شركة أبل: الواقع المعزز سيصبح أهم من الطعام
3	صحيفة_الوئام ميزة في هاتف غوغل تكشف "ابتزاز أبل#
4	مقارنة سريعة للصور ومن تصويري بين جهازي #ايفون10 و#جالاكسي_نوت8
5	أبداع تصوير ممتاز و رائع #آيفون10 #shotoniphonex وارونا إبداعاتكم وتصويركم
6	عنصريه و حب لايفون 10
7	بكرا دورنا مو بس 10 حتى ال 8 يجنن رائع تصويره
8	كل أجيال الآيفون تصويرهم جبار قوي جيد.. أنا مستمتع جدا بكل كاميرة جهاز آيفون
9	الحين تقدر و بكل سهوله تخزن وتشغل كل ملفات الفيديو والصور والموسيقى و تنقل الملفات بين اجهزة
10	تقرير : #آبل تستثمر في شركة Finisar بقيمة تصل إلى 390 مليون دولار من أجل التعرف وتطوير وتحس

Fig. 9. Sample of tweets streamed from hashtag ايفون١٠# after cleaning

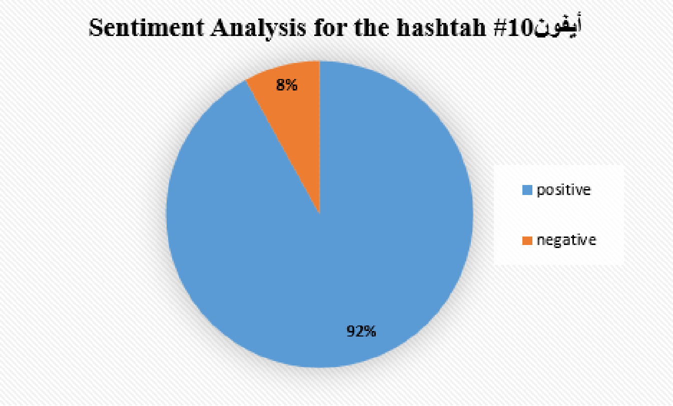


Fig. 10. Percentages of opinions about #iPhone10

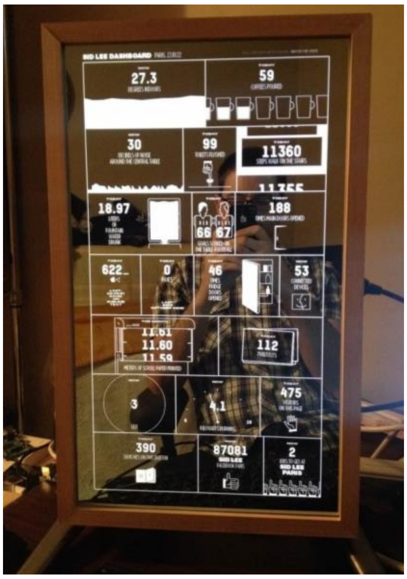


Fig. 11. Example of two-way smart mirror for displaying results from tweets sentiment analysis and other IoT data

Finally, we presented the results from IoT machine using a two-way smart mirror. The Pi graphs of tweets sentiment analysis in different trending hashtags can be updated every half-day and other IoT data, as shown in Fig. 11.

V. Conclusion and Future Works

In this paper, we presented the idea of integrating IoT domain with the computational linguistics domain. We proposed an IoT mining machine using Raspberry Pi pre-programmed for sentiment analysis and opinion generation on a smart mirror. The experimental works were conducted using Arabic Tweets Dataset released by SEM-Eval 2017 and Naïve Bayes classifier which yields an outstanding accuracy on this dataset. Future works includes training other classifier as well as using other datasets. The IoT mining machine will also improved to show the most important tweets in each hashtag and the phrases used for the dominant sentiment. We will also include more sentiment classes such as positive, negative, neutral, and others. Our IoT machine is expected to be commercialised and used in companies and homes.

REFERENCES

[1] Alzahrani, S.M.: ‘Sensing for the Internet of Things and Its Applications’. Proc. The IEEE 5th International Conference on Future Internet of Things and Cloud Prague, Czech Republic 21-23 August 2017 2017 pp. Pages

[2] Soro, A., et al.: ‘Designing the Social Internet of Things’. Proc. Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, Colorado, USA2017 pp. Pages

[3] Kim, J.E., et al.: ‘Empowering End Users for Social Internet of Things’. Proc. Proceedings of the Second International Conference on Internet-of-Things Design and Implementation, Pittsburgh, PA, USA2017 pp. Pages

[4] Okada, M., et al.: ‘Autonomous Cooperation of Social Things: Designing a System for Things with Unique Personalities in IoT’. Proc. Proceedings of the 6th International Conference on the Internet of Things, Stuttgart, Germany2016 pp. Pages

[5] Wang, Y.: ‘Integration of Data Mining with Game Theory’, in Wang, K., et al. (Eds.): ‘Knowledge Enterprise: Intelligent Strategies in Product Design, Manufacturing, and Management’ (Springer US, 2006), pp. 275-280

[6] Bramer, M.: ‘Principles of Data Mining’ (Springer-Verlag, 2007. 2007)

[7] Howland, P., and Park, H.: ‘Cluster-Preserving Dimension Reduction Methods for Efficient Classification of Text Data’, in Berry, M.W. (Ed.): ‘Survey of Text Mining: Clustering, Classification, and Retrieval’ (Springer New York, 2004), pp. 3-23

[8] Cai, Y., and Sun, J.-T.: ‘Text Mining’, in Liu, L., and ÖZsu, M.T. (Eds.): ‘Encyclopedia of Database Systems’ (Springer US, 2009), pp. 3061-3065

[9] Saleiro, P., et al.: ‘TexRep: A Text Mining Framework for Online Reputation Monitoring’, New Gener. Comput., 2017, 35, (4), pp. 365-389

[10] Wang, W., et al.: ‘The Impact of Sentiment Orientations on Successful Crowdfunding Campaigns through Text Analytics’, IET Softw., 2017, 11, (5), pp. 229-238

[11] Lin, J.J., et al.: ‘Personality-based refinement for sentiment classification in microblog’, Knowledge-Based Systems, 2017, 132, pp. 204-214

[12] Rodriguez-Esteban, R.: ‘Text Mining Applications’: ‘Reference Module in Life Sciences’ (Elsevier, 2017)

[13] Shatkay, H.: ‘Biomedical Text Mining’: ‘Reference Module in Life Sciences’ (Elsevier, 2017)

[14] Kumar, B.S., and Ravi, V.: ‘A survey of the applications of text mining in financial domain’, Knowledge-Based Systems, 2016, 114, (Supplement C), pp. 128-147

[15] : ‘Sentiment Analysis’, in Sammut, C., and Webb, G.I. (Eds.): ‘Encyclopedia of Machine Learning and Data Mining’ (Springer US, 2017), pp. 1152-1152

[16] Takahashi, S., et al.: ‘A Method for Opinion Mining of Coffee Service Quality and Customer Value by Mining Twitter’, in Kunifuji, S., et al. (Eds.): ‘Knowledge, Information and Creativity Support Systems: Selected Papers from KICSS’2014 - 9th International Conference, held in Limassol, Cyprus, on November 6-8, 2014’ (Springer International Publishing, 2016), pp. 521-528

[17] Golande, A., et al.: ‘An Overview of Feature Based Opinion Mining’, in Corchado Rodriguez, J.M., et al. (Eds.): ‘Intelligent Systems Technologies and Applications 2016’ (Springer International Publishing, 2016), pp. 633-645

[18] : ‘Microblog Sentiment Analysis’, in Alhajj, R., and Rokne, J. (Eds.): ‘Encyclopedia of Social Network Analysis and Mining’ (Springer New York, 2014), pp. 893-893

[19] : ‘Twitter Opinion Mining’, in Alhajj, R., and Rokne, J. (Eds.): ‘Encyclopedia of Social Network Analysis and Mining’ (Springer New York, 2014), pp. 2259-2259

[20] Li, G., et al.: ‘Twitter Microblog Sentiment Analysis’, in Alhajj, R., and Rokne, J. (Eds.): ‘Encyclopedia of Social Network Analysis and Mining’ (Springer New York, 2014), pp. 2253-2259

[21] Bao, Y., et al.: ‘The Role of Pre-processing in Twitter Sentiment Analysis’, in Huang, D.-S., et al. (Eds.): ‘Intelligent Computing Methodologies: 10th International Conference, ICIC 2014, Taiyuan, China, August 3-6, 2014. Proceedings’ (Springer International Publishing, 2014), pp. 615-624

[22] Kaiser, C., et al.: ‘Bridging the Gap between Qualitative and Quantitative Analysis of Opinion Forums’. Proc. Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 012008 pp. Pages

[23] Popescu, A.-M., and Pennacchiotti, M.: ‘Detecting controversial events from twitter’. Proc. Proceedings of the 19th ACM international conference on Information and knowledge management, Toronto, ON, Canada2010 pp. Pages

[24] Supriya, B.N., et al.: ‘Twitter Sentiment Analysis Using Binary Classification Technique’, in Vinh, P.C., and Barolli, L. (Eds.): ‘Nature of Computation and Communication: Second International Conference, ICTCC 2016, Rach Gia, Vietnam, March 17-18, 2016, Revised Selected Papers’ (Springer International Publishing, 2016), pp. 391-396

[25] Gonzalez-Marron, D., et al.: ‘Exploiting Data of the Twitter Social Network Using Sentiment Analysis’, in Sucar, E., et al. (Eds.): ‘Applications for Future Internet: International Summit, AFI 2016, Puebla, Mexico, May 25-28, 2016, Revised Selected Papers’ (Springer International Publishing, 2017), pp. 35-38

[26] Garg, Y., and Chatterjee, N.: ‘Sentiment Analysis of Twitter Feeds’, in Srinivasa, S., and Mehta, S. (Eds.): ‘Big Data Analytics: Third International Conference, BDA 2014, New Delhi, India, December 20-23, 2014. Proceedings’ (Springer International Publishing, 2014), pp. 33-52

[27] You, Q.: ‘Sentiment and Emotion Analysis for Social Multimedia: Methodologies and Applications’. Proc. Proceedings of the 2016 ACM on Multimedia Conference, Amsterdam, The Netherlands2016 pp. Pages

[28] Brooks, M., et al.: ‘Collaborative Visual Analysis of Sentiment in Twitter Events’, in Luo, Y. (Ed.): ‘Cooperative Design, Visualization, and Engineering: 11th International Conference, CDVE 2014, Seattle, WA, USA, September 14-17, 2014. Proceedings’ (Springer International Publishing, 2014), pp. 1-8

[29] Paltoglou, G., and Thelwall, M.: ‘Twitter, MySpace, Digg: Unsupervised Sentiment Analysis in Social Media’, ACM Trans. Intell. Syst. Technol., 2012, 3, (4), pp. 1-19

[30] Tan, C., et al.: ‘User-level sentiment analysis incorporating social networks’. Proc. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, San Diego, California, USA2011 pp. Pages

[31] Lee, J., et al.: 'Sentiment Analysis of Twitter Users Over Time: The Case of the Boston Bombing Tragedy', in Sugumaran, V., et al. (Eds.): 'E-Life: Web-Enabled Convergence of Commerce, Work, and Social Life: 15th Workshop on e-Business, WEB 2015, Fort Worth, Texas, USA, December 12, 2015, Revised Selected Papers' (Springer International Publishing, 2016), pp. 1-14

[32] Dinkić, N., et al.: 'Using Sentiment Analysis of Twitter Data for Determining Popularity of City Locations', in Stojanov, G., and Kulakov, A. (Eds.): 'ICT Innovations 2016: Cognitive Functions and Next Generation ICT Systems' (Springer International Publishing, 2018), pp. 156-164

[33] Nakov, P., et al.: 'Developing a successful SemEval task in sentiment analysis of Twitter and other social media texts', Language Resources and Evaluation, 2016, 50, (1), pp. 35-65

[34] Saif, H., et al.: 'Semantic Sentiment Analysis of Twitter', in Cudré-Mauroux, P., et al. (Eds.): 'The Semantic Web – ISWC 2012: 11th International Semantic Web Conference, Boston, MA, USA, November 11-15, 2012, Proceedings, Part I' (Springer Berlin Heidelberg, 2012), pp. 508-524

[35] Saif, H., et al.: 'Semantic Patterns for Sentiment Analysis of Twitter', in Mika, P., et al. (Eds.): 'The Semantic Web – ISWC 2014: 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part II' (Springer International Publishing, 2014), pp. 324-340

[36] Bhattacharya, S., and Banerjee, P.: 'Towards the Exploitation of Statistical Language Models for Sentiment Analysis of Twitter Posts', in Saeed, K., et al. (Eds.): 'Computer Information Systems and Industrial Management: 16th IFIP TC8 International Conference, CISIM 2017, Bialystok, Poland, June 16-18, 2017, Proceedings' (Springer International Publishing, 2017), pp. 253-263

[37] Voutyras, O., et al.: 'Achieving Autonomicity in IoT systems via Situational-Aware, Cognitive and Social Things'. Proc. Proceedings of the 18th Panhellenic Conference on Informatics, Athens, Greece2014 pp. Pages

[38] Blackstock, M., et al.: 'Uniting online social networks with places and things'. Proc. Proceedings of the Second International Workshop on Web of Things, San Francisco, California, USA2011 pp. Pages

[39] Wakkary, R., et al.: 'Morse Things: A Design Inquiry into the Gap Between Things and Us'. Proc. Proceedings of the 2017 Conference on Designing Interactive Systems, Edinburgh, United Kingdom2017 pp. Pages

[40] Yao, L., et al.: 'Exploring recommendations in internet of things'. Proc. Proceedings of the 37th international ACM SIGIR conference on Research &#38; development in information retrieval, Gold Coast, Queensland, Australia2014 pp. Pages

[41] Crabtree, A., and Tolmie, P.: 'A Day in the Life of Things in the Home'. Proc. Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, San Francisco, California, USA2016 pp. Pages

[42] Noergaard, T.: 'Chapter 1 - A Systems Approach to Embedded Systems Design': 'Embedded Systems Architecture (Second Edition)' (Newnes, 2013), pp. 3-19

[43] Molano, J.I.R., et al.: 'Internet of Things: A Prototype Architecture Using a Raspberry Pi', in Uden, L., et al. (Eds.): 'Knowledge Management in Organizations: 10th International Conference, KMO 2015, Maribor, Slovenia, August 24-28, 2015, Proceedings' (Springer International Publishing, 2015), pp. 618-631

[44] Kumar, A., and Rani, R.: 'Sentiment analysis using neural network', in Editor (Ed.)^(Eds.): 'Book Sentiment analysis using neural network' (2016, edn.), pp. 262-267

[45] Rana, S., and Singh, A.: 'Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques', in Editor (Ed.)^(Eds.): 'Book Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques' (2016, edn.), pp. 106-111

[46] Edward, L., and Steven, B.: 'NLTK: the Natural Language Toolkit', in Editor (Ed.)^(Eds.): 'Book NLTK: the Natural Language Toolkit' (Association for Computational Linguistics, 2002, edn.), pp. 63-70

[47] Goel, A., et al.: 'Real time sentiment analysis of tweets using Naive Bayes', in Editor (Ed.)^(Eds.): 'Book Real time sentiment analysis of tweets using Naive Bayes' (2016, edn.), pp. 257-261

[48] Rosenthal, S., et al.: 'SemEval-2017 Task4 : Sentiment Analysis in Twitter ', in Editor (Ed.)^(Eds.): 'Book SemEval-2017 Task4 : Sentiment Analysis in Twitter ' (Association for Computational Linguistics, 2017, edn.), pp.