

Ontology-based Emotion Detection in Arabic Social Media

Sawsan N. Cassab
Faculty of Information Technology
Syrian Virtual University
Damascus, Syria

Dr. Mohamad-Bassam Kurdy
Faculty of Information Technology
Syrian Virtual University
Damascus, Syria

Abstract: With the spread of social media and its content that is often driven by user's emotions or their opinions on some topic, efforts have been made to provide effectiveness mechanisms for automatic emotion detection, to employ it in various fields.

With the spread of Arabic language on those sites, however Arabic studies in emotion detection are still shy somewhat, perhaps, due to the difficult dealing with Arabic language especially with slang that is used mainly in social media, in the absence of unified representation of emotion in Arabic, this was motivation to build an ArEmontology ontology, for conceptual representation of emotions in Arabic with standard language based on emotional theories, in addition to propose an effective mechanism to detect emotion from Arabic text by using the classification and semantic relations in ArEmontology, this mechanism was applied on Facebook's posts, as it is the most site used by Syrian, and got reasonable accuracy of 65% in detecting one of the emotional categories: joy, sadness, fear, anger, surprised and disgust.

Keywords: Ontology, Emotion detection, text annotation, Arabic social media, Levantine dialect

I. INTRODUCTION

As the text is the most common way in social media that used by individuals to express their opinions and emotional states, many researches have been done to monitor those social media's content, for opinion's analyzing and classifying into three main categories, which are positive, negative and neutral, or for emotion detection and classify the content into several main emotional categories, such as joy, sadness, fear, anger, disgust and surprising, according to the emotional model adopted by each research.

Emotion detection in text can be employed in many important applications. In the area of business development, emotion detection can help marketers to develop strategies for customer satisfaction, new product development and service delivery. Psychologists can benefit from being able to infer people's emotions based on the text that they write which they can use to predict their state of mind. In the field of education, the ability of computers to automatically track attitudes and feelings with a degree of human intuition has contributed to the development of Text-to-Speech systems and Intelligent Tutoring Systems[1].

With the spread of Arabic language on social media, some studies have been introduced to analyze Arabic text in those sites, most of them have focused on classifying these texts, comparing classification methodologies, or inferring emotional status by relying on matching with emotional dictionaries, without taking the context into account.

In this paper, to overcome the absence of semantically emotion detection in Arabic text, an Arabic emotional ontology has built, called ArEmontology, and used to annotate posts by the right emotion category, depending on classifying and semantic relations that has defined in the ontology, and approved by an expert in the emotional field.

Our approach was applied on Facebook posts, as it is the most used by Syrian, according to new statistics which have resulted that 86% of people in Syria use Facebook, while about 1.7% use Twitter [2].

II. PROPOSED APPROACH

This section discusses the design methodology of ArEmontology, as well as the methodology used for detecting emotion in Arabic social media.

A. Ontology Development Model

ArEmontology was developed using an application-independent design methodology, which was given in [3]. The steps for this methodology are shown in Fig.1, and described as follow:

1) **Identify Domain:** ArEmontology covers six basic emotional categories: joy, sadness, anger, fear, surprise and disgust, the main purpose is to identify emotion in Arabic classical or slang text to annotate it with the right emotional category.

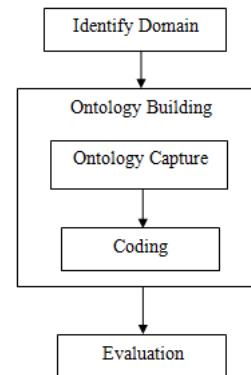


Fig. 1. Methodology for ontology development

Re-use an existent ontology: no another ontology was used as ArEmontology is the first one that represents emotional concepts in Arabic.

2) **Ontology building:**

The ontology was constructed over emotions and two languages using top-down method, its structure comprises of

three main classes: 'Emotion', 'Language', and 'Intensity', also contains three object properties: 'trustFactor', 'isOppositeOf', and 'isBelongTo'.

Ontology capture:

A thorough knowledge and understanding of the domain was obtained via different sources, then the ontology was developed by trying to integrate the emotional hierarchy structure of Parrot, with the Plutchik's wheel of emotions.

a) Define Classes and Class's hierarchy:

ArEmontology has three main classes:

- **Emotion class:** from this class, two levels of hierarchy are branched as follow:

The top level contains emotion classes at the primary level in the Parrot's hierarchy, except 'Love'. 'Disgust' is a subclass of 'Anger' in this hierarchy, but we represent it as a top level class according to the Plutchik's wheel, and emotion classes at the second level are subclasses in the ontology. Both structures (Parrot and Plutchik) were translated into Arabic using English-English¹ and Arabic-Arabic² dictionaries, and the translation was verified by an Arabian expert in the emotional field.

- **Intensity class:** According to ascending emotional intensity classification, that arranges the emotional vocabularies into intensities belonging to the integer numerical domain [-10, +10], a nominal classification was proposed as shown in table 1.

TABLE 1 PROPOSED INTENSITY CATEGORIES

| Intensity category | | |
|--------------------|--------------|------------------|
| Numerical domain | Arabic name | English name |
| [+8, +10] | إيجابي مرتفع | High positive |
| [+4, +7] | إيجابي متوسط | Normal positive |
| [0, +3] | إيجابي منخفض | Low positive |
| [-3, -1] | سلبي منخفض | Low negative |
| [-7, -4] | سلبي متوسط | Normal negative |
| [-9, -8] | سلبي مرتفع | High negative |
| [-10] | سلبي تحذيري | warning negative |

This proposal was approved by an expert in the emotional field, and represented as subclasses of Intensity class.

- **Language class:** it has two subclasses, 'العربية_الفصحي' (classical Arabic) and 'اللهجة_الشامية' (Levantine dialect), in order to associate the emotional words to their corresponding language.

b) Define relations: Three object properties were defined, they are:

- **Language class:** it associates individuals in emotional classes with intensity ones,

for example: 'قلق' for 'عامل_الثقة' (worry in English) is 'سلبي_منخفض'.

- **('isBelongTo' in English):** it associates each individual in emotional classes to its corresponding language, e.g. 'ينتهي_للغة' ('isBelongTo' in English) is 'العربية_الفصحي'.
- **('isOppositeOf' in English):** it connects contrasting emotional individuals, as 'قلق' is 'سلبي_منخفض' and 'اطمئنان' is 'معاكس_لـ' ('contentment' in English).

Also, **Disjoint With** is defined to connect two contrasting emotions on class level, as described in Plutchik's wheel, for example: 'حزن' (sadness) Disjoint With 'متعة' (joy).

c) create Instances: they are entities in tertiary level in the Parrot's hierarchy in addition to some of entities in Plutchik's wheel.

d) Enrich ontology with synonyms and stemming: Levantine emotion words were collected by a questionnaire was posted on some Facebook Syrian groups, they added as synonyms, including the associated words that have emotional connotations, which is a group of words that come together to indicate a different meaning from the meaning indicated by each word separately, for example: 'فأير قلبي' (my heart is boiling) denotes anxiety, 'فأير دمي' (my blood is boiling) denotes angry, and different forms of the same word also added as label annotation properties due to lack of stemmers specialized in Levantine, for example: 'ملل', 'ملاين', 'مالي' all denote 'bored', a light stemmer for classical Arabic only was used.

Coding:

The ontology was represented in the formal language OWL (Web Ontology Language). It is the standard and recommended language by W3C (World Wide Web Consortium), Protégé 5.5 is used as an ontology editor for building ArEmontology.

3) Evaluation:

A golden standard was used with the help of domain expert to evaluate ArEmontology according to the following metrics:

$$\text{Precision: } P = \text{#correct guesses} / \text{\#total guesses} \quad (1)$$

$$\text{Recall : } R = \text{#correct guesses} / \text{\#total} \quad (2)$$

#correct guesses is the number of correct concepts (or individuals) in the ontology

#total guesses is the number of total concepts (or individuals) in the ontology

#total is the number of possible concepts (or individuals)

The results of evaluation are:

For concepts: $P = 36/37 = 0.97$, $R = 36/36 = 1$.

For individuals: $p = 329/350 = 0.94$, $R = 329/362 = 0.90$.

Also, In order to verify and validate the ontology with regards to competency questions, the Description Logic Query (DL-Query) was used which is standard Protégé plug-in and it is based on the Manchester OWL syntax with HermiT OWL Reasoner.

An example of the querying function that answers the questions that were asked in the development process of the ontology is: *What are the individuals refer to Disgust and used in Levantine dialect?*, which is illustrated in table 2 in DL-Query format.

¹<https://www.oxfordlearnersdictionaries.com/definition/english/>

²<https://www.almaany.com/>

B. Approach Architecture:

For the purpose of this work, components of a General Architecture for Text Engineering (GATE)³ have used, the proposed approach that is show in Fig.8, consists of the following components:

1) Preprocessing unit: In this unit which is shown in Fig.3, a set of language processing techniques is applied, in order to eliminate the noisy and refine posts early and this will lead to cost reduction throughout the emotion detection process:

TABLE 2 QUESTION IN DL QUERY FORMAT

| Lang | DL Query | | | |
|------|----------|------------|------------|---------------|
| | اشمئزاز | and | يتنمي_للغة | اللغة_الشامية |
| Eng | Disgust | isBelongTo | value | Levantine |

Fig.2 depicts the result of this DL-Query,



Fig. 2. The result of DL-Query

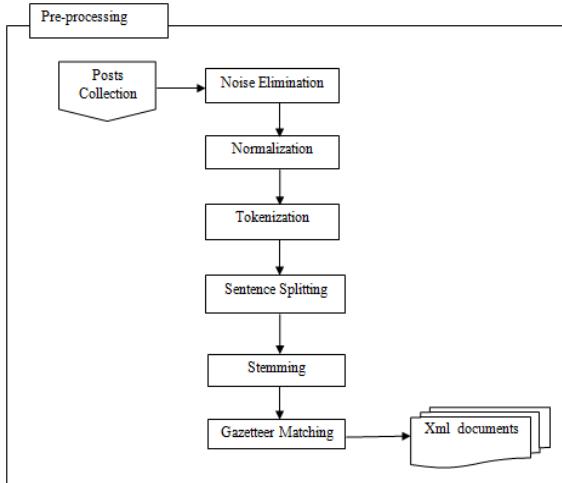


Fig. 3 Preprocessing unit

a) Noise Elimination: by using regular expression, Arabic and English numbers are removed, as well as non Arabic characters, URLs, hashtags, Mentions, Diacritics (e.g. ﻭ، ﻻ) Tatweel (“—”), and repeated letter (e.g. ﻭ، ﻻ) are all removed.

b) Normalization: (such as Hamza ‘ا’ standardization).

³ <https://gate.ac.uk/>

c) Stemming: a number of stemmers that implemented in the Java, namely Light8, Khoja, Lucene, and Aranlp, were tested on a random set of word’s corpus in this research, but the results were not accurate enough with words that belong to the Levantine dialect, so Light8 was chosen after some modification to stem classical Arabic words only.

d) Tokenization: by using an ArabicTokenizer, it is a processing resources from GATE, text is broken into words and elements that are called tokens.

e) SentenceSplitting: by using an ANNIE SentenceSplitter, which is a processing resource from GATE, text is split into sentences, so the inferring of emotion doesn’t interfere between sentences.

f) Gazetteer Matching: the role of the gazetteers is to identify entity names in the text based on lists, two modifier gazetteers were defined, they contain words that may affect the intensity of the emotional significance of the subsequent or previous word in the post, “Emphasis gazetteer” contains words like (‘كثير’) means ‘much’), and “Reducing gazetteer” contains words like (‘شوي’) means ‘ little’), in addition to a “Negation gazetteer” which contains negative words such as (‘مراح’) means ‘not’), and “Levantine_maxims gazetteer” contains phrases that denote emotion like “زاد الطين بلة”, it means “add fuel to the fire” and denotes to angry, and the last one is “Stopwords gazetteer” which contains words that don’t add any meaning to the text.

2) Annotator unit: in this unit, a mapping is done between words in the text and ontology’s concepts to capture the emotion that a word refers to, this mapping is done by “Onto Root gazetteer” which is called by a flexible one, both of them are processing resources in GATE.

The input of this unit is the xml documents that returned by the “Preprocessing” unit while the output is represented by ‘Lookup’ annotation, with number of features that will be processed later in the next unit, one of the most important features is ‘ClassURIList’, it contains emotional classes related to that word according to the ontology’s hierarchy.

3) JAPE emotion assignment unit: it is totally depending on JAPE transducer, a processing resource in GATE, it takes the output of “Annotator” unit as input and returns a special emotion annotation with its intensity,

Fig.4 shows the work flow of this unit, where the accompanying number for each process refers to the degree of priority in implementation:

In this unit, depending on the gazetteer matching results (from “Preprocessing” unit), negation (single negation or repeated) is checked, also emphasis/reducing words in the text, then any ‘Lookup’ annotation related to emotional concepts, is turned into ‘EMOTION’ annotation after traversing in the ontology and finding the parent class of those in ‘Lookup’, this parent is one of the six basic emotions (joy, sadness, fear, anger, surprise and disgust), and get the associated trust factor, those findings are stored in a Tree Map (Fig.5) to input to the Emotional Category Assignment sub-unit, that returns the total or dominant emotion in the text. The traversing algorithm in the tree map is as follow:

For each entry in the tree:

Find the dominant entry which is related to the maximum counter (it is calculated by summing of iteration times of trust factor in the sub-tree associated with that entry), and it will represent a dominant emotion..

The trust factor is the key with the maximum value in the sub-tree.

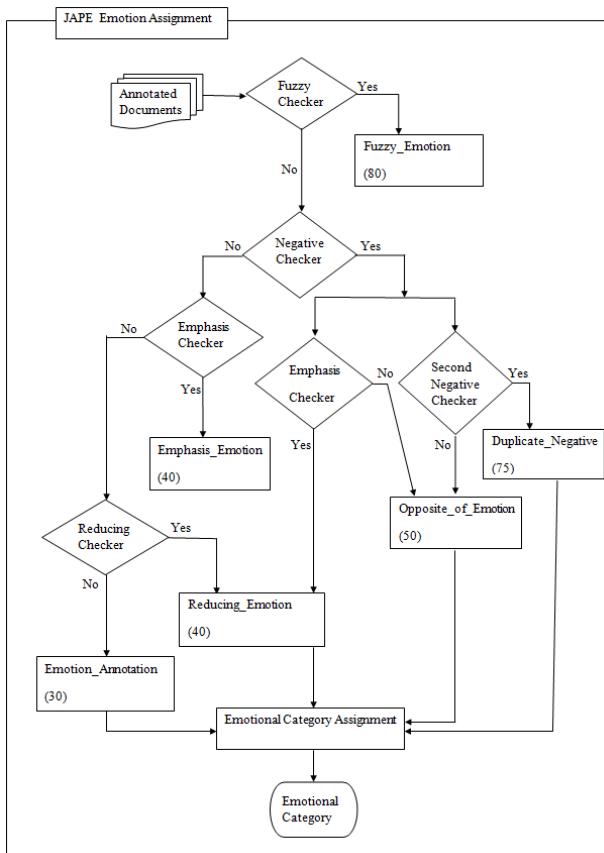


Fig. 4 workflow of JAPE emotion assignment unit

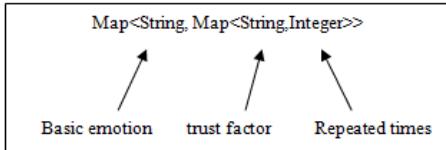


Fig. 5 Tree Map

In the case of equivalence for all trust factor's values associated in the output entry, one of them will be chosen randomly.

In the case of the equivalence of trust factors associated to all tree entries, one of them will be the dominant emotion.

For example: for the following tree map:

```

Map<String, Map<String, Integer>> EmoMap=
{Sadness, <<1, low negative> , <2, high negative>> },
{Disgust, <<1, low negative> , <1, normal negative>> }
  
```

The result is calculated as follow:

Sadness counter = 1+2 = 3, Disgust counter = 1+1 = 2

So, the dominant emotion is sadness (as it associates with the highest counter), the trust factor is high negative (as it has highest frequency).

III. EXPERIMENTS AND RESULTS

The proposed approach approved that it is able to process even special issues, such as:

1) *Negation* : when a negative word appears in the post and associate to the emotional word, the contrast emotion for that word is returned.

2) *Emphasis/ Reducing*, when an emphasis/reducing word appears in the post, and associated to the emotional word, a trust factor for that emotion is increased/decreased by one step, even when those modifier words placed at different locations in the sentence (e.g. I am so happy, or I am happy so much).

3) *Negation with emphasis*: it equivalent to reducing (e.g. 'not so much' equivalent to 'little').

4) *More than one negation*: this case is equivalent to a positive one, (e.g. 'I don't think I am not happy' equivalent to 'I think I am happy').

Fig.6. shows the emotions in the post 'دایق خلقی بهالحجر و مافی شی مسلی' (in English: 'my mood is malaised and nothing is amusing') :

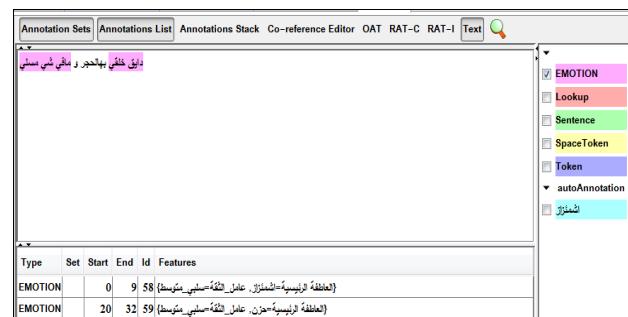


Fig. 6 emotions in the post

'Disgust' and trustFactor is 'Normal negative'

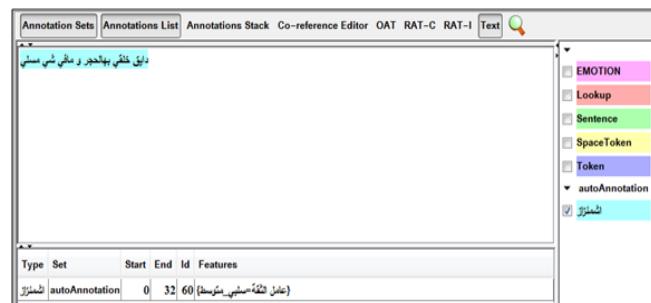


Fig. 7 total emotion of the post

Also, the approach can assign 'no emotion' to posts that have no emotion connotation.

For evaluation, we depended on a golden standard which is 100 manually annotated posts, to compare the output annotation of our algorithm with it, an IAA(inter-annotator agreement) component from GATE was used, and achieved 65% accuracy.

IV. CONCLUSION AND FUTURE WORK

This research presented the first Arabic emotional ontology which is ArEmontology, it will be the base for new studies in the field of automatic emotion detection in Arabic text, especially as it was approved by an expert in the emotional field, and its designing which allow it to expand by adding new dialects, or even languages.

Also, the research presented an ontology-based effective mechanism to detect emotion and its intensity from Arabic text that is written in Classical or Levantine dialect, with the help of GATE tool.

The new method has achieved 65% accuracy when tested on Facebook posts.

Depending on this work, it will be possible to analyze the emotional state of individuals in Syria, it is known that the crisis negatively affected the emotions and lives of many of them and their stability and mood, and this work may have a helpful role in providing follow-up directives on reliable and selected websites that provide scientific material as psychological books commensurate with the results, bearing in mind that these results may not be final (but they are sufficient according to the expert's opinion), especially since some individuals may have wrong dealings with social media, and may write posts that do not necessarily express their true emotional state.

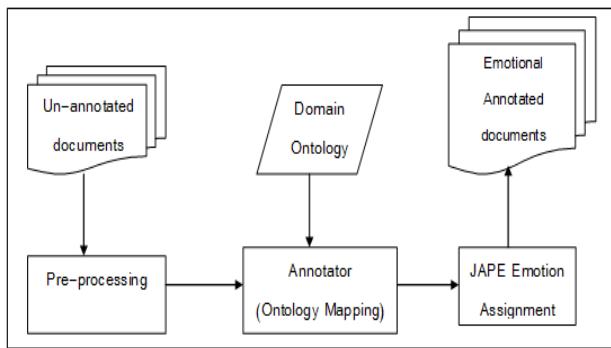


Fig. 8 Proposed architecture

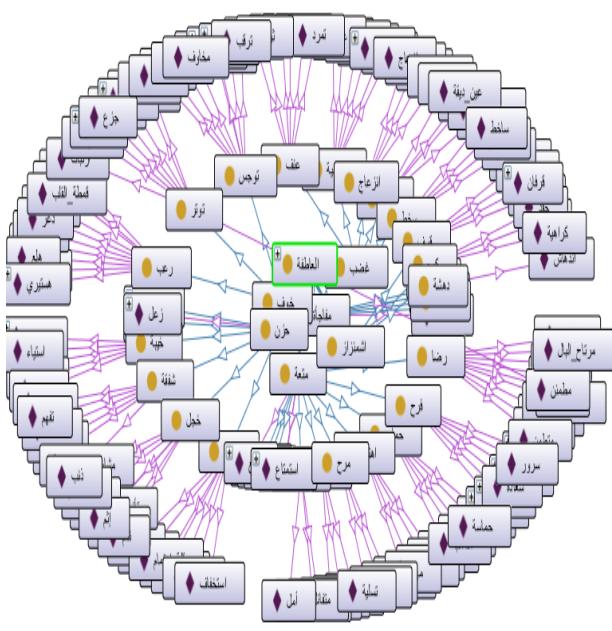


Fig. 9 ArEmontology Ontology

It is worth noting the diversity in the use of tools and their integration to build the practical application of this research, which was distinguished by its achievement of the desired results, the ability to re-use as independent units (plugins),

scalable, and easy to use, in addition to a user-friendly displaying.

Future work is:

1. Extend ArEmontology by :
 - Adding a new object property, which is “isComposedOf” that describes which two basic emotions constitute a complex emotion, for example: ‘Contempt’ = ‘angry’ + ‘Disgust’.
 - Adding concepts of ‘Love’ and ‘Trust’ emotions, and their individuals.
 - Adding another Arabic dialect or another Languages and relate each other by ontology relations to be a basis for translation projects.
2. Add a new gazetteer for emoji.
3. Use a POS Tagger that is able to pares both classical Arabic and Levantine dialect, so that it is possible to identify personal nouns that have emotional connotations, to overcome any ambiguity associated with such cases where some emotion names used as person's name (it is usual in Arabic) , for example: "سعید": the adjective and noun are similar in this text.
4. Developing of an Emotional Recommendation System to reply to each post with comments appropriate to the emotion extracted from it, with the aim of showing support for the publisher and understanding his emotions. These comments are stored in the system's database and should be reviewed by an emotional expert. and such system will be useful for customer relationship management in business enterprises.

ACKNOWLEDGMENT

This work has done with a help of Dr. Alaa Morad, a professor in department of Psychology at Damascus university

REFERENCES

- [1] El Gohary, A. F., Sultan, T. I., Hana, M. A., & El Dosoky, M. M. (2013). A Computational Approach for Analyzing and Detecting Emotions in Arabic Text. *International Journal of Engineering Research and Applications* , pp. 100-107.
- [2] Social Media Stats Syrian Arab Republic. (2020). Retrieved 2020, from StatCounter Global Stats: <https://gs.statcounter.com/social-media-stats>
- [3] Uschold, M and King, M, 1995, “Towards a methodology for building ontologies” Workshop Held in Conjunction with IJCAI on Basic Ontological Issues in Knowledge Sharing (<http://citeseer.nj.nec.com/uschold95toward.html>) 19-1 to 19-12.
- [4] Ahmed, S., & Tabassum, H. (2016). EmotiOn: An Ontology For Emotion Analysis.
- [5] Almanie, T., Aldayel, A., Alkanhal, G., Alesmail, L., Almutlaq, M., & Althunayan, R. (2018). Saudi Mood: A Real-Time Informative Tool for Visualizing Emotions in Saudi Arabia Using Twitter. *2018 21st Saudi Computer Society National Computer Conference (NCC)*. 2018 21st Saudi Computer Society National Computer Conference (NCC).
- [6] Bani-Hani, A., Majdalweih, M., & Obeidat, F. (2017). The Creation of an Arabic Emotion Ontology Based on EMOTIVE. *ScienceDirect* , 1053-1059.
- [7] Daoood, A. A., Salman, I., & Ghneim, N. (2017). Comparision study of automatic classifiers performance in emotion recognition of Arabic social media users. *Jait* , pp. 5172-5183.
- [8] Francisco, V., Peinado, F., Hervás, R., & Gervás, P. (2010). Semantic web Approach to the extraction and presentation of emotions in texts.
- [9] Panger, G. T. (2017). *Emotion in Social Media*. Berkeley.
- [10] Parrott, W. G. (2001). Emotions in Social Psychology.
- [11] Plutchik, R. (1980). *Theories of Emotion*.
- [12] Ghady A., Mohamad-Bassam K. (2016) Web Opinion Mining For Arabic Language

- [13] Seal, D., Roy, U., & Basak, R. (2019). Sentence-Level Emotion Detection from Text Based on Semantic Rules. In *Advances in Intelligent Systems and Computing*.
- [14] Shivhare, S. N., Garg, S., & Mishra, A. (2015). EmotionFinder: Detecting Emotion From Blogs and Textual Documents. *International Conference on Computing, Communication and Automation (ICCCA2015)*, (pp. 52-57).
- [15] Suet Yan Liew, J. (2016). Fine-grain emotion detection in microblog text. Semantic Scholar.