EmotionFinder: Detecting Emotion From Blogs and Textual Documents

Shiv Naresh Shivhare
School of Computing Science and Engineering
Galgotias University
Uttar Pradesh, India
Shiv827@gmail.com

Shakun Garg
Department of Computer Science and Engineering
BBDIT Ghaziabad
Uttar Pradesh, India
shakunbah@gmail.com

Anitesh Mishra
Amdocs Development Centre
Gurgaon
Haryana, India.
anitesh.mishra@yahoo.co.in

***Abstract*—Emotion Detection is one of the most emerging issues in human machine interaction. Detecting emotional state of a person from textual data is an active research field along with recognizing emotions from facial and audio information. Several methods were given to recognize emotion from text in previous years. This paper proposed a new architecture (a keyword based approach) to recognize emotions from text. In case of recognizing emotion from a piece of text document or a blog, any human can do this better than a machine only problem is he/she takes time. Proposed emotion detector system takes a text document and the emotion word ontology as inputs and produces one of the six emotion classes (i.e. love, sadness, joy, fear and surprise, anger) as the output. Every input text contains some short stories which are firstly read and assigned an emotion class manually and then that emotion class is compared to the output of the proposed system to check the accuracy of the Proposed Emotion Detector System. It is found that the Proposed Emotion Detector System produces output with the accuracy of more than 75%.**

Keywords—Human-Computer Interaction; Textual Emotion Recognition; Emotion Word Ontology

I. INTRODUCTION

Human emotion recognition by analyzing written documents appear challenging but many times essential due to the fact that most of the times textual expressions are not only direct using emotion words but also result from the interpretation of the meaning of concepts and interaction of concepts which are described in the text document. Emotion detection from text plays a key role in the human-computer interaction [1]. Human emotions may be expressed in many ways like person's speech, face expression and written text known as speech, facial and text based emotion respectively [14]. In human computer interaction, human emotion recognition from text is becoming increasingly important from an applicative point of view.

Methods being used for text based emotion detection are classified into keyword spotting technique, lexical affinity method, learning based method and hybrid approach however each method has its own limitations [19]. A proposed architecture which contains the emotion ontology and emotion detector algorithm is explained in Section 3. In section 4, algorithm is implemented and results are shown. Conclusion is given in Section 5.

II. RELATED WORK

The role of human computer interaction is proposed by Picard in the concept of affective computing [3]. Many researchers from computer science, biotechnology, psychology, and cognitive science are attracted by this domain. From another point of view, research in the field of emotion detection from textual data emerged to determine human emotions. Emotion detection from text can be formulated as follows: Let A be the set of all authors, T be the set of all possible representations of emotion-expressing texts, and E be the set of all emotions. Let r be a function to reflect emotion e of author a from text t, i.e., $r: A \times T \rightarrow E$, then the function r would be the answer to the problem [4].

The concept of emotion recognition systems lies in fact that, although the definitions of T and E may be straightforward, the definitions of individual element, even subsets in both sets of T and E would be rather confusing. As the languages are constantly emerging new elements may add on one side, for the set T. Due to the complex nature of human minds, any emotion classifications can only be seen as "labels" annotated afterwards for different purposes, whereas on the other side, currently there are no standard classifications of "all human emotions". Methods used for text based emotion recognition system [4], [5] are:

A. Keyword Spotting Technique

The keyword spotting technique can be described as the problem of finding occurrences of keywords (love, anger, joy, sadness, surprise and fear) from a given text document. Many algorithms to analyze sentiment or emotion have been suggested in the past. In the context of emotion detection this method is based on certain predefined keywords. These emotion words are categorized into keywords such as disgusted, sad, happy, angry, fearful, surprised etc. Occurrences of these keywords can be found and based on that an emotion class is assigned to the text document.

B. Lexical Affinity Method

Detecting emotions based on related keywords is an easy to use and straightforward method. Keyword spotting technique is extended into *Lexical affinity approach* which assigns a

ISBN:978-1-4799-8890-7/15/$31.00 ©2015 IEEE

probabilistic 'affinity' for a particular emotion to arbitrary words apart from picking up emotional keywords. These probabilities are part of linguistic corpora but have some disadvantages also; firstly the assigned probabilities are biased toward corpus-specific genre of texts, secondly it misses out emotional content that resides deeper than the word-level on which this technique operates e.g. keyword 'accident' having been assigned a high probability of indicating a negative emotion, would not contribute correctly to the emotional assessment of phrases like 'I met my girlfriend by accident' or 'I avoided an accident'.

C. *Learning-Based Methods*

Learning-based methods are being used to analyze the problem in a different manner. Initially problem was to determine sentiment from input text data but now the problem is to classify the input texts into different emotion classes. Unlike keyword-based detection methods, learning-based methods try to recognize emotions based on previously trained classifier, which apply theories of machine learning such as support vector machines [8] and conditional random fields [9], to determine that input text belongs to which emotion class.

D. *Hybrid Methods*

Since keyword-based technique with thesaurus and naïve learning-based method could not acquire satisfactory results, both the methods are combined to improve accuracy.

E. *Limitations*

From above discussion there are few limitations [7]:

1) Ambiguity in Keyword Definitions

Keyword based emotion detection is a simple way to detect associated emotions but the meanings of keywords could be multiple and vague, as several keywords could change their meanings according to different usages and contexts. Moreover, even the minimum set of emotion labels (without all their synonyms) could have different emotions in some extreme cases such as ironic or cynical sentences.

2) Incapability of Recognizing Sentences without Keywords

Keyword-based approaches always search for some specific set of keywords. Therefore, sentences which does not contain any keyword would imply that they do not belong to any emotion at all, which is obviously wrong e.g. "I passed my qualify exam today" and "Hooray! I passed my qualify exam today" imply the same emotion (joy), but the sentiment in former sentence without "hooray" could not get detected if "hooray" is the only keyword to find the emotion.

3) Lack of Linguistic Information

Keyword-based approaches always search for some specific set of keywords. Therefore, sentences which does not contain any keyword would imply that they do not belong to any emotion at all, which is obviously wrong e.g. "I passed my qualify exam today" and "Hooray! I passed my qualify exam today" imply the same emotion (joy), but the sentiment in former sentence without "hooray" could not get detected if "hooray" is the only keyword to find the emotion.

4) Difficulties in Determining Emotion Indicators [10]

Learning-based approaches can determine the probabilities between features and emotions but these approaches still need keywords in the form of features. Emoticons are the most intuitive features which can be seen as author's emotion annotations in the texts. The cascading problems would be the same as those in keyword-based methods.

III. PROPOSED ARCHITECTURE

The proposed architecture is very simple and easy to understand. This model is based on keyword spotting technique apart from that it also uses the concept of ontology. Use of ontology makes this model more efficient than other methods in recognizing emotions from text input.

Proposed framework is divided into two main components: Emotion Ontology, Emotion Detector.

A. *Emotion Ontology*

Ontology is an explicit specification of conceptualization. Ontologies have definitional aspects like high level schemas and aspects like entities and attributes interrelationship is between entities, domain vocabulary. Ontology allows a programmer to specify, in an open, meaningful way the concepts and relationships that collectively characterise some domain [16].

Emotion can be expressed as joy, surprise, sadness, hate, fear, anger and so on. Since various emotion hierarchies are developed by researchers and there is not any standard emotion word hierarchy, focus is on the related research about emotion in cognitive psychology domain. In 2001, W. Gerrod Parrot [2], published a book named "Emotions in Social Psychology".

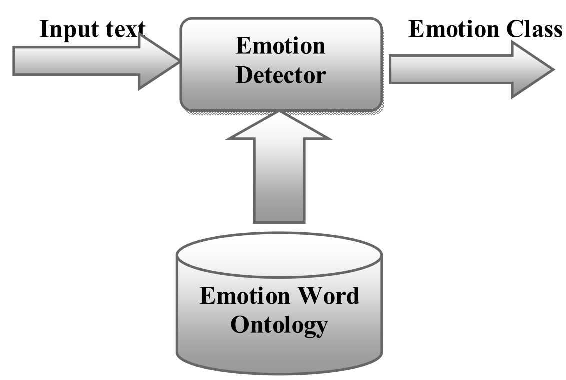

Fig 1. Proposed Architecture

In this book author explained the emotion system and formally classified the human emotions through an emotion hierarchy in six classes. Primary level contains classes i.e. Love, Joy, Anger, Sadness, Fear and Surprise. Certain other classes of emotion words fall in secondary and tertiary levels. Emotion word hierarchy is then converted into emotion word ontology showing class and subclass relationship format. Primary level emotion classes are shown at top in the emotion word ontology and emotion classes at the tertiary level are at the bottom of ontology. Emotion ontology is developed by *Protégé* (an open source ontology editor and knowledge-base framework).

Emotion ontology is an ontology representing "emotion" domain. A schematic diagram of emotion ontology is shown above in figure 2. Emotion ontology is displayed in left to right order. Only two levels (primary and secondary) are displayed in the image of emotion ontology while the tertiary level is hidden. Every emotion word in the ontology is an emotion class. "Thing" is the root and the super most class of the ontology. Every child class is related to its parent class by "is-a" relationship (child is-a parent) e.g. "anger" is a "thing" and "pride" is a "joy".

Protégé is a free, open source ontology editor and a knowledge acquisition system. Like Eclipse, Protégé is a framework for which various other projects suggest plug-in. Protégé application is written in Java and heavily uses Swing to create the rather complex user interface [15].

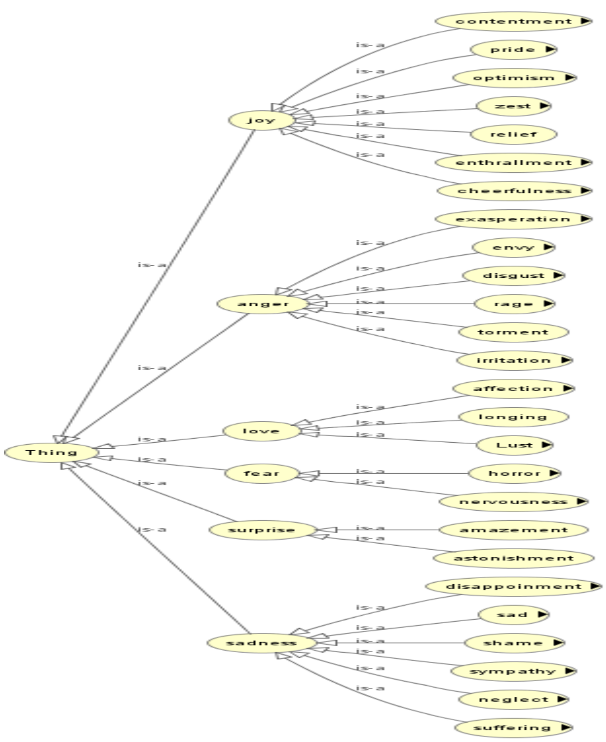

Fig. 2. Emotion Ontology

Protégé provides a growing user community with a suite of tools to construct domain models and knowledge-based applications based on ontologies. Protégé presents a rich set of knowledge base structures and actions that create, visualize, and manipulate ontologies in various representation formats. Protégé can be customized to provide domain-friendly support for creating knowledge models and entering data. Protégé is being extended and used by way of a plug-in architecture and a Java-based Application Programming Interface (API) for building knowledge-based tools and applications.

B. Emotion Detector Algorithm

Sentiment or emotion class of the text data can be recognized with the help of proposed emotion detector algorithm. The algorithm calculates score for every emotion class of primary level available in the emotion ontology by adding the scores of its respective secondary and tertiary levels' emotion classes. In final step scores of all the primary level classes are compared and emotion class having maximum score will be declared as the "emotion" of the input text document.

1) Parameters Used

Algorithm calculates score for all the emotion words in the emotion ontology and compare among them. Proposed algorithm uses certain parameters for this purpose. Jena library is used to calculate parameters which allow traversal and parsing of ontology. Required parameters are as follows:

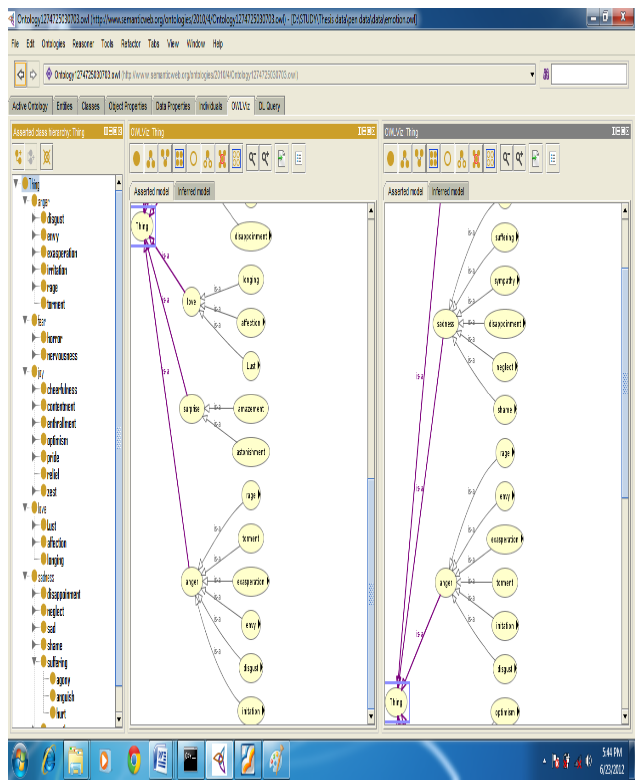

Fig. 3. Snapshot of Protégé Tool

a) Parent –Child Relationship

If the emotion keywords found in a text document belongs to the secondary or tertiary level of emotion ontology; it also indirectly refers to the parent (primary level) of it. Hence if a certain value is added to the secondary or tertiary level emotion classes' scores, corresponding parent's score also need to be modified. This is implemented by traversing the emotion word ontology in a breadth first manner using Jena API. When any node (emotion word) is encountered all of its children are retrieved. Then same method is applied to every child of the corresponding node.

b) Depth in Ontology

Depth is also needed as it gives an idea about how specific is the emotion word with respect to its corresponding ontology structure and at what level that word falls. More Specific smaller the depth, greater the weight age. Weight age value is calculated simultaneously while traversing the ontology tree.

c) Frequency in Text Document

If an emotion keyword appears multiple times, it is emphasizing to belonging to that emotion class. Frequency is an important criterion as more is the frequency more will be the importance of that emotion class. Frequency of emotion words is calculated by parsing the input text and searching for occurrences of emotion words that belong to emotion ontology.

2) Algorithm

Algorithm given below is proposed to calculate the score for each emotion word of emotion ontology with the help of certain parameters explained in previous step. This score will be directly proportional to the frequency of the term and inversely proportional to its depth in the ontology. A formula devised for the mth terminology:

```
for  j ← 1 to No. of Nodes [Ontology]
      do parent [j] ← parent of node j
          child [j] ← child of node j
       for m ← 1 to No. of Nodes [Ontology]
             do freq [m] ← frequency of occurrence of m^th
                 depth [m] ← depth of m^th node in ontology
```

Calculate (x):

```
  for m ← 1 to No. of Nodes [Ontology]
        score (x) ← freq [root] / depth [root]
         for m ← 1 to No. of parent nodes [Ontology]
             score (parent) = score (parent) + score (child)
             return score (parent)
          for m ← 1 to No. of parent nodes [ontology]
            emotion class ← High score [parent]
            return emotion class
```

In the above algorithm, Nodes [Ontology] denotes emotion classes (emotion words) in the ontology, Parent [j] denotes parent classes in the ontology, Child [j] denotes child classes in the ontology, Freq [m] denotes frequency of mth class in text document, Depth denotes depth of particular class in emotion ontology starting from the root class, Score [parent] denotes score of parent class in emotion ontology.

Proposed algorithm defines a method to calculate the score of primary level emotion classes. Primary level emotion class with highest score will be decided as the final “emotion class” for input text document.

IV. RESULTS AND IMPLEMENTATION

The proposed system is implemented using eclipse framework. Eclipse is a multi-language software development environment which is having an integrated development environment (IDE) and an extensible plug-in system. To implement the proposed system first of all we need to collect the data i.e. texts documents and preprocess the text data. The preprocessed text data is used as input in the proposed system. The design of the proposed system is shown in figure 1 which is composed of two main functional parts. The prototype is implemented using eclipse framework which uses Jena API.

Blogs from different sites [17][18] e.g. personal blogs and short stories are collected manually. After collecting blogs, preprocessing in terms of data cleaning and data transformation is done to format textual data in order to give input to the proposed framework. More than a hundred blogs are collected to process in the proposed system and determine the emotion as the output of the system. Input text is transformed in the XML format.

XML Documents Form a Tree Structure. A root element is contained in XML documents. Root element is "the parent" of all other elements. A document tree is formed by the elements in XML document. The tree starts at the root and branches to the lowest level of the tree. All elements can have sub elements (child elements). Relationship between the elements is described by the terms parent, child, and sibling. Parent elements have children. Children on the same level are called siblings (brothers or sisters). XML elements must be properly nested. XML documents must have a root element. Some XML tags are set to use in the proposed system which are <dmoz_doc>, <dmoz_subdoc>, <title> and <end> and pair tags </dmoz_doc>, </dmoz_subdoc>, </title>. Each text document is enclosed in a pair of these tags <dmoz_doc> and </dmoz_doc>. XML tag <dmoz_doc > is used to identify that from where the document starts and </dmoz_doc> indicates that from where the document finishes. <dmoz_subdoc> is used to identify that from where a subdocument or blog starts and </dmoz_subdoc> indicates that from where a blog finishes. <end> assigned where the whole input text finishes.

For the validation phase, the blogs are collected manually from different web locations and read out to manually assign a primary emotion class to every particular blog. The assigned primary emotion class is written in between the pair of XML tags <title> and </title> which can be one of the six primary emotion classes i.e. love, joy, sadness, fear, surprise and anger.

Experiment 1

All the blogs are read and classified by manual assignment and their summary is shown in table 1. Total 135 blogs are taken and analyzed to test the Emotion Detection System. Table 1 shows number of blogs which fall in particular primary emotion class by manual assignment. In the experiment 2 the same blogs are used in Emotion Detector System as input to find the out in form of primary emotion classes.

TABLE I. CLASSIFICATION SUMMARY OBTAINED MANUALLY

Love	Joy	Sadness	Anger	Fear	Surprise
26	31	17	31	28	2

Experiment 2

The Java project which contains the Emotion Detector System code in Eclipse IDE is named as "TextAnalyser". At every execution of the project TextAnalyser, an output file "output.txt" is generated. Output file contains the emotion class of every blog as an output of the Emotion Detector System. Output of Emotion Detector System is shown in table 2 in the form of particular emotion class.

TABLE 2. CLASSIFICATION SUMMERY OF THE BLOGS BY USING THE PROPOSED FRAMEWORK

Love	Joy	Sadness	Anger	Fear	Surprise
30	27	21	25	30	2

After collecting number of blogs per emotion class from both experiments, we make a comparison bar chart shown in figure 4 between results obtained from Emotion Detection System and the results previously assigned by manual assignment. In the graph blue bar represents the number of blogs classified manually in primary emotion class and red bar represents the number of blogs classified by Emotion Detector System in primary emotion class. Horizontal axis represents the name of primary emotion class and vertical axis represents the number of blogs which fall in a particular primary emotion class. In the next section these results will be analyzed in form of precision and accuracy.

A. Analysis of the Result

To test the detection capability and analyze the results obtained from Emotion Detector System, we need to check the accuracy of proposed framework. Binary classification test is the best way to check that how correct results are produced by Emotion Detector System. The decision matrix is used to calculate the classification accuracy, shown in table 3.

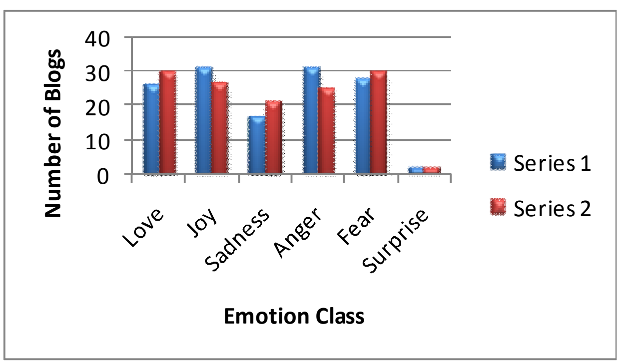

Fig. 4. Comparison of the experiments

TABLE 3: DECISION MATRIX FOR CALCULATING THE ACCURACY OF THE PROPOSED SYSTEM

	YES	NO
YES	a	b
NO	c	d

The meaning of a, b, c and d are defined below:

a: Proposed System and the manual assignment agree with the assigned emotion class,
b: Proposed System disagree with the assigned emotion class but the manual assignment agree,
c: Proposed System agree with assigned emotion class but the manual assignment disagree,
d: Proposed System and manual assignment both disagree.

Measurements are qualitative or quantitative. Descriptive and non-numerical results are given by qualitative measurements whereas quantitative measurements give results that are definite, usually as numbers and units. Measurements work best when they are accurate and precise. There are two important criteria through which we can check the performance and the correctness of any system which are as follows:

Accuracy and Precision

Accuracy is a measure of how close a measurement comes to the actual or true value of whatever is measured. Accuracy meant by something that is "capable of providing a correct reading or measurement". A measurement is accurate if it correctly reflects the size of the thing being measured. To evaluate the accuracy of a measurement, it must be compared to the correct value.

Accuracy = (a + d) / (a + b + c + d)

Precision defines a measure of how close a series of measurements are to one another. Precise means "repeatable, reliable and getting the same measurement each time". The precision of a measurement depends on more than one measurement.

Precision = (a) / (a + c)

Poor accuracy results from procedural or equipment flaws whereas poor precision results from poor technique. Accuracy shows the correctness while precision shows the reproducibility.

TABLE 4: CALCULATING THE ACCURACY AND PRECISION PER EMOTION CLASS

Emotion Class	a	b	C	d	Precision (%)	Accuracy (%)
Love	24	2	6	0	80.00	75.00
Joy	24	7	3	0	88.88	70.58
Sadness	17	0	4	0	80.95	80.95
Anger	23	8	2	0	92.00	69.69
Fear	26	2	4	0	86.66	81.25
Surprise	2	0	0	0	100	100
Average					**88.08**	**79.57**

B. Result Discussion

Experiments have been done to test the proposed system in different condition to prove the effectiveness of the system and successfully completed all the tests with required results. Table 4 shows the accuracy and precision calculated for the emotion detector system.

The test cases include running the Emotion Detector System and comparing the results generated in the form of emotion class with the emotion class of the blog assigned previously as titles. Finally number of blogs which emotion classes found same by Emotion Detector System as previously assigned as titles is called as "correct". Total 135 blogs are tested and 116 blogs' emotion classes are found correct by Emotion Detector System. These results proved that the proposed system and the implementation and testing of the Emotion Detector System are successful and it satisfies all the requirements.

V. CONCLUSION

Text-based input is the most common way for humans to interact with computers while writing letters or giving feedback to any product in the era of web 2.0. Thus emotion detection from text focuses as an important research issue in affective computing.

In this paper, existing research of emotion recognition based on textual data is surveyed and limitations of existing methods are reviewed. Emotion recognition system architecture is proposed to improve detection capabilities in an efficient manner. Proposed system is based on keyword spotting technique that is having rich features of ontology. Not all the limitations of existing methods are overcome by this architecture but use of ontology improves the detection capability by applying semantic approach.

REFERENCES

[1] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, "Recognition of Emotional States in Natural human-computer interaction," in IEEE Signal Processing Magazine, vol. 18(1), Jan. 2009.

[2] Parrott, W.G, "Emotions in Social Psychology," in Psychology Press, Philadelphia 2001

[3] C. Maaoui, A. Pruski, and F. Abdat, "Emotion recognition for human machine communication", Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 08), IEEE Computer Society, Sep. 2008, pp. 1210-1215, doi: 10.1109/IROS.2008.4650870

[4] Chun-Chieh Liu, Ting-Hao Yang, Chang-Tai Hsieh, Von-Wun Soo, "Towards Text-based Emotion Detection: A Survey and Possible Improvements ",in International Conference on Information Management and Engineering,2009.

[5] N. Fragopanagos, J.G. Taylor, "Emotion recognition in human–computer interaction", Department of Mathematics, King's College, Strand, London WC2 R2LS, UK Neural Networks 18 (2005) 389–405 march 2005

[6] C. Elliott, "The affective reasoner: a process model of emotions in a multiagent system," in Doctoral thesis, Northwestern University, Evanston, IL, May 1992.

[7] D. S´anchez, M.J. Mart´ın-Bautista, I. Blanco, "Text Knowledge Mining: An Alternative to Text Data Mining" in 2008 IEEE International Conference on Data Mining Workshops.

[8] C.-H. Wu, Z.-J. Chuang, and Y.-C. Lin, "Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models," ACM Transactions on Asian Language Information Processing (TALIP), vol. 5, issue 2, Jun. 2006, pp. 165-183, doi:10.1145/1165255.1165259.

[9] Feng Hu and Yu-feng Zhang, "Text Mining Based on Domain Ontology", in 2010 International Conference on E-Business and E-Government.

[10] Z. Teng, F. Ren, and S. Kuroiwa, "Recognition of Emotion with SVMs," in Lecture Notes of Artificial Intelligence 4114, D.-S. Huang, K. Li, and G. W. Irwin, Eds. Springer, Berlin Heidelberg, 2006, pp. 701-710, doi: 10.1007/11816171_87.

[11] C. Yang, K. H.-Y. Lin and H.-H. Chen, "Emotion classification using web blog corpora," Proc. IEEE/WIC/ACM International Conference on Web Intelligence. IEEE Computer Society, Nov. 2007, pp. 275-278, doi: 10.1109/WI.2007.50.

[12] C. M. Lee, S. S. Narayanan, and R. Pieraccini, "Combining Acoustic and Language Information for Emotion Recognition," Proc. 7th International Conference on Spoken Language Processing (ICSLP02), 2002, pp.873-876.

[13] C.-H. Wu, Z.-J. Chuang and Y.-C. Lin, "Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models", ACM Transactions on Asian Language Information Processing (TALIP), vol. 5, issue 2, Jun. 2006, pp. 165-183, doi:10.1145/1165255.1165259.

[14] C. Elliott, "The affective reasoner: a process model of emotions in a multiagent system," in Doctoral thesis, Northwestern University, Evanston, IL, May 1992

[15] Protégé tool, www.protege.stanford.edu/

[16] Nicu Sebea, Ira Cohenb, Theo Geversa, and Thomas S. Huangc "Multimodal Approaches for Emotion Recognition: A Survey", USA

[17] http://sail.usc.edu/~kazemzad/emotion_in_text_cgi/DAL_app/index.php ?overall=bad&submit_evaluation=Submit+Query".

[18] www.wikipedia.org/

[19] Shiv Naresh Shivhare, Saritha Khethawat, "Emotion Detection from Text", Second International Conference on Computer Science, Engineering and Applications(CCSEA-2012), May 26-27, 2012, Delhi, India, ISBN: 978-1-921987-03-8.