Analysis and Tracking of Emotions in English and Bengali Texts: A Computational Approach

Dipankar Das
Supervised by
Professor Sivaji Bandyopadhyay
Department of Computer Science & Engineering, Jadavpur University,
Kolkata, India
dipankar.dipnil2005@gmail.com

ABSTRACT
The present discussion highlights the aspects of an ongoing doctoral thesis grounded on the analysis and tracking of emotions from English and Bengali texts. Development of lexical resources and corpora meets the preliminary urgencies. The research spectrum aims to identify the evaluative emotional expressions at word, phrase, sentence, and document level granularities along with their associated holders and topics. Tracking of emotions based on topic or event was carried out by employing sense based affect scoring techniques. The labeled emotion corpora are being prepared from unlabeled examples to cope with the scarcity of emotional resources, especially for the resource constraint language like Bengali. Different unsupervised, supervised and semi-supervised strategies, adopted for coloring each outline of the research spectrum produce satisfactory outcomes.

Categories and Subject Descriptors
I.2.7 [**ARTIFICIAL INTELLIGENCE**]: Natural Language Processing, Text Processing

General Terms
Algorithms, Measurement, Performance, Design, Experimentation, Human Factors, Languages.

Keywords
Emotions, Expression, Holder, Topic, Tracking, BengWAL, Blogs, Syntactic Argument Structure, CRF, SVM.

1. INTRODUCTION
Major studies on Opinion Mining and Sentiment Analyses have been attempted with more focused perspectives rather than fine-grained emotions [1]. A wide range of Natural Language Processing (NLP) tasks such as tracking users' emotion about products or events or about politics as expressed in online forums, to customer relationship management are using emotional information. Question Answering (QA) systems [3] and modern Information Retrieval (IR) [4], [9] systems are also increasingly incorporating emotion analysis within their scope. Emotion analysis aims to determine the attitude of a speaker or a writer with respect to some topic [5].

WWW 2011, March 28–April 1, 2011, Hyderabad, India.
ACM 978-1-4503-0637-9/11/03.

2. PROBLEM
Human-machine interface technology has been investigated for several decades. Scientists have found that emotion technology can be an important component in artificial intelligence [6]. Recent research has placed more emphasis on the recognition of nonverbal information, and has especially focused on emotion reaction. Emotions, of course, are not linguistic things. However the most convenient access that we have to them is through the language [10]. Emotional stance, not open to objective observation or verification is difficult to convey in text as it does not always contain any direct, specific or conclusive hints. Hence, we need a comprehensive theory of what a human emotion is, and then we need to understand how the emotion is expressed and transmitted within the natural language. These aspects raise the need of syntactic, semantic, and pragmatic analysis of a text [8].

3. STATE OF THE ART
Started from beginning, characterization of the words and phrases according to their emotive tone was done [11]. The annotation of opinion and emotion related content of a language has been described in [3]. With respect to the traditional natural language domains such as News [15], Blogs [16] are also becoming the popular, communicative and informative repository of text based emotional contents in the Web 2.0. Opinion mining at word, sentence and document levels and opinion summarization on news and web blog articles are discussed in [13]. In order to estimate the affects in text, the model proposed in [17] processes symbolic cues and employs NLP techniques for word, phrase and sentence level analysis. Several machine learning techniques were used on blog data to identify the mood of the authors during writing [16]. The text-based emotion prediction using supervised machine learning based on the SNoW learning architecture is discussed in [12]. Most works cited above were carried out for English. Bengali is less privileged and less computerized than English. Thus, the language independent techniques are employed to develop and evaluate the systems for English and Bengali [32].

Some of the well known prior works in identifying opinion holders are described in [18] [19]. Based on the traditional perspectives, another work discussed in [20] uses an emotion knowledge base for extracting emotion holder. Though our adopted syntactic techniques are similar to the earlier attempts, the identification of emotion holders with respect to emotional verbs signifies the difference from others.

In opinion topic extraction, different researchers contributed their efforts [21], [22]. But, all these works are based on lexicon look up approach and applied on the domain of product reviews. A contributory topic annotation task on the MPQA corpus is described in [23]. In contrast, the incorporation of rhetorical structure, four similarity features and multi-engine voting technique significantly identifies the individual topic spans from the focused target span [38], [39].

Temporal Sentiment Analysis in Social Events has been studied in [14]. *MoodViews* also analyzes multiple sentiments [24]. Presently, an emotion tracking system is being developed [7]. Work using unlabelled data together with positive examples to predict the sentiment orientation of blogs is discussed in [25]. The preparation of labeled corpus from unlabeled data is being carried out to fulfill our partial research objectives.

4. PROPOSED APPROACH

Emotion analysis involves identifying the emotion holder, subject / topic / event detection on which emotions are expressed and classification of emotional / evaluative expressions at word / phrase / sentence / document level.

4.1 Classification of Evaluative Expressions

The input documents can be analyzed to identify its sentiment orientation (factual or emotion), the subject and aspects involved and the evaluation keywords of *positive* or *negative* sentiment type. The sentiment oriented documents can be decomposed to identify the fine grained Ekman's [26] six universal emotions (e.g. *happy, sad, anger, disgust, fear or surprise*) from different portions of the documents (e.g. word, phrase or sentence).

4.2 Holder Identification

An emotion holder is the person or organization that expresses the emotion [3]. In the case of product reviews and blogs, holders are usually the authors of the posts. Information about emotion holder is generally present in words or phrases with the semantic roles: *experiencer*, *agent*, *actor*, *patient* and *beneficiary*. The writer's and reader's emotion for a specific text is to be determined so that their underlying sense can be differentiated.

4.3 Subject-Topic-Event Detection

There are other relevant effort could found in literature [27]. The proposal aimed to identify the subject, topic or event in a sentence on which the emotion is expressed. This information is necessary for tracking of emotions expressed on a particular subject / topic / event by a holder with respect to time.

4.4 Emotion Tracking

Sometimes, the blog users comment on each other. The identification of such overlapped comments on a given topic is crucial for detecting emotion. This module aims to track a single user's comments on the same topic as well as on different topics to analyze the changes in emotion with respect to topic and time. Tracking of mass emotions on certain subject / topic / event over time will also be taken up in the proposed research.

4.5 Working with Unlabeled Data

An important task in the present research will be the development of emotion annotated news stories and blog corpora in Bengali as these are important resources for analyzing emotions.

5. METHODOLOGY

The whole activities in the proposed research are centered on news and blog corpora. The methodologies will be language independent in nature with English, Bengali as for case study. The methodologies developed will be useful for news agencies, news reporters, sociologists, psychologists and common people in general. Descriptions of supporting resources and developed systems are as follows.

5.1 Resources

5.1.1 Lexical Resource

We have incorporated two available sentiment resources, *SentiWordNet* [28] and *WordNet Affect* [10] in our present task. The collection of the *WordNet Affect* synsets [15] used in the present task was provided as a resource for the SemEval-2007 shared task of "Affective Text" [15]. Both of these resources are in English. Hence, the development of Bengali *WordNet Affect Lists* (*BengWAL*) has been carried out in four phases [31]. The first two phases, updation using *SentiWordNet* [28] and *VerbNet* [29] and translation through a synset based English to Bengali bilingual dictionary (being developed as part of the EILMT project[1]) are followed by a duplicate removal technique. The sense disambiguation algorithm [42] based on the similarity clue is applied on the translated Bengali synsets. Inter-translator evaluation achieves the moderate agreement.

5.1.2 Corpus

We have used two types of English corpus. One is the SemEval 2007 affect sensing corpus [15] which is a sentence level emotion annotated corpus containing only the news headlines. The scores for valence and six emotions are assigned to each of the sentences. Another one is an emotion-annotated blog corpus [43] that contains a rich set of emotion information such as category, intensity and word or phrase based expressions at sentence level. Additionally, a manual effort has been made for annotating the topic and target spans in the blog corpus [43]. If an emotional verb present in the *WordNet Affect* [10] is found as a member verb of any *VerbNet* [29] class, the sentences corresponding to that class have been retrieved to construct the emotion corpus. This corpus has been only used for identifying emotion holders.

To the best of our knowledge, there is no such available corpus annotated with detailed linguistic emotional expressions in Bengali or even in other Indian languages. Hence, an emotion annotated Bengali blog corpus has been developed manually [30]. Each sentence is annotated with the emotional components such as emotional expression (word/phrase), intensity, associated holder(s) and topic(s). Ekman's six emotions along with three types of intensities (*high*, *general* and *low*) were considered for sentence level annotation. Different types of fixed and relaxed strategies were employed to measure the agreement. Consequently, each of the 205 blog documents is annotated with single or multiple document level emotions for experiments [36].

5.2 Identifying Emotional Expressions

The system of identifying evaluative emotional expressions consists of four inter-connected modules at different levels of

[1] English to Indian Languages Machine Translation (EILMT) is a TDIL project undertaken by the consortium of different premier institutes and sponsored by MCIT, Govt. of India

granularities such as word (W), phrase (P), sentence (S) and document (D). The baseline system for word level emotion tagging has been developed to measure the performance with respect to each emotion class without integrating any prior knowledge of word features [32]. Each of the words of the corpus is passed through these six modules separately to tag with the appropriate class label. On the other hand, the Conditional Random Field (CRF) [44] and Support Vector Machine (SVM) [45] based machine learning classifiers are employed for word level emotion tagging [33], [34], [35]. Different singleton features (e.g. *Part of Speech (POS) of the words, Question words, Reduplication, Colloquial / Foreign words, Special Punctuation Symbols, Negations, Emoticons* etc.), context features (e.g. *unigram, bigram*) at word and *POS* tag level along with their different combinations are used for training and testing. Measuring the classifier's accuracy using confusion matrix, the equal distribution of emotion tags with the non-emotion tag improves word level emotion tagging systems [32].

The assignment of sentence level emotion tags and scores based on the word level constituents is carried out using corpus based and sense based *tag weights* of six emotions [33] followed by post-processing techniques for handling negations [32].

The baseline system at phrase level is developed based on *WordNet Affect* lists [10] and parsed dependency relations [40]. SVM based supervised framework is also employed by incorporating different word and context level features. Information Gain based pruning, application of admissible tag sequences and a class-splitting technique improves the system's performance by reducing the label bias problem of SVM. Phrase to sentence level emotion tagging is also mentioned in [40].

The document level emotion tagging from the emotion tagged sentences is carried out based on some combinations of heuristic features (e.g. *emotion tag of the title sentence or end sentence, emotion tags assigned to an overall topic, most frequent emotion tags expressed in user comment portions of a document, identical emotions that appear in the longest series of tagged sentences* etc.). The best two emotion tags are assigned to each document based on the ordered emotion scores obtained [36].

5.3 Holder Identification

The baseline model (BM) for identifying emotion holder in English has been developed based on the *subject* information of the emotional sentences parsed using Stanford Dependency Parser [37]. Similarly, the Bengali blog sentences are passed through an open source Bengali shallow parser [2] that produces different morphological information (e.g. *root, case, vibhakti, tam, suffixes* etc.). The lexical pattern based phrase level similarity clues containing different *POS* combinations, Name Entities (NEs) and noun phrases have been considered [41].

Another way to identify emotion holder is based on the syntactical argument structure of the emotional sentences. In English, the *head* of each chunk in the dependency-parsed output helps in constructing the syntactic argument structure with respect to the key emotional verb. Two separate techniques have been adopted for extracting the argument structure [37]. One is from parsed result directly and another is from the corpus that has been POS tagged and chunked separately. Similarly, the verb based argument structures are acquired from the chunk or phrase level lexical patterns of the shallow parsed Bengali blog sentences. The pivotal hypothesis considered in the syntactic model (SynM) is based on the hypothesis followed in [42]. But, the equivalent English verbs of identical sense for the Bengali verbs are retrieved using Bengali to English bilingual dictionary [3]. The available frames of the equivalent English verbs are extracted from English *VerbNet* [29].

For both English and Bengali, each acquired syntactic argument structure is mapped to all the possible frame syntaxes present for the corresponding verb in the *VerbNet* [29]. If the acquired syntactic argument structure of a sentence matches with any of the retrieved frame syntaxes of *VerbNet*, the holder roles (e.g. *Experiencer, Agent, Actor, Beneficiary* etc.) associated with the *VerbNet* frame syntaxes are then assigned in the appropriate slots in the syntactic arguments of the sentence [37] [41].

5.4 Topic Identification

The baseline topic identification model (BM) for English is developed based on the parsed constituents of the *object* related dependency parsed relations. Like emotion holder, the phrase segments containing topic related *Thematic Roles* (e.g. *Topic*, *Theme*, *Event* etc.) are extracted from the verb based syntactical argument structures of the sentences. On the other hand, a supervised model (SvdM) is adopted to identify multiple emotion topics along with their topic and target spans from each of the blog sentences [46]. CRF, SVM and Fuzzy Classifier (FC) [46] are employed by considering various features (e.g. the annotated emotional expressions along with *direct* and *transitive* dependencies, *causal verbs*, *discourse markers, Emotion Holder*, *Named Entities* and four types of similarity measures like *Structural Similarity, Sentiment Similarity, Syntactic Similarity* and *Semantic Similarity*) and their combinations. The incorporation of a special feature, *Structural Similarity* based on the Rhetorical Structure Theory [47] improves our system [38].

5.5 Emotion Tracking

Presently, tracking of sentiment events based on temporal relations in English [2] and tracking of emotion of Bengali bloggers [7] have been carried out successively.

6. RESULTS

Various experiments regarding symbolic feature, language, and domain-dependent features were carried out for evaluating the word level emotion classification system [32], [34], [35]. The lexical feature (e.g. POS, words of *SentiWordNet* and *WordNet Affect*) outperforms other features significantly. Different combination of context features also shows significant improvement in performance. The word level tagging system has demonstrated the average F-Scores of 83.65% and 70.23% on 1,500 word tokens on English news and Bengali blogs respectively [32]. The average F-Scores of 65% and 63.26% on 200 test sentences were achieved for news and blogs in sentential emotion tagging. A supervised system for English blogs [43] outperforms the baseline system and achieves the average F-Scores of 82.72%, 76.74% and 89.21% for emotional expressions, sentential emotions and intensities respectively on 565 gold standard test sentences [40].

[2] http://ltrc.iiit.ac.in/showfile.php?filename=downloads/shallow_parser.php

[3] http://home.uchicago.edu/~cbs2/banglainstruction.html

The baseline model for English suffers in identifying emotion holders from the passive sentences. Though the *recall* value decreases, the syntactic model outperforms over baseline significantly in terms of F-Score. The dependency parser based method achieves better F-Score (66.98%) than the other method (F-Score of 62.39%) on a collection of 4,112 emotional sentences as the second method fails to disambiguate mostly the arguments from adjuncts [37]. The maximum average F-Scores of the baseline and hybrid systems for emotion topic identification are 56.75% and 58.88% respectively on 500 sentences [39]. The supervised multi-engine voting system achieves the F-Scores of 70.51% and 90.44% for topic and target span identification respectively from the blog sentences [38].

For Bengali, the baseline system achieves the average F-Scores of 53.85% and 50.02% for identifying emotion holders and topics respectively on 500 test sentences [41]. The average F-Scores of the syntactic system are 66.03% and 61.98% for single as well as multiple emotion holders and topics on the same test set. But, the syntactic system suffers in resolving some errors (e.g. *appositive cases, co reference with emotional expression, multiple holders and topics, overlapping topic spans, anaphoric presence of the holders*). Thus, some simple rule based error reduction techniques based on rhetorical structure, emotional expressions are employed into the syntactic system [41].

Table 1. F-Scores (in %) of three emotional components

Topics	English	Bengali
Evaluative Expressions	83.65 (W, #1500) 76.74 (S, #565) 82.72 (P, #603)	70.23 (W, #1500) 82.72 (S, #200) 76.74(D, #110)
Emotion Holder	(# 4112) 64.83 (BM) 66.98 (SynM)	(#500) 53.85 (BM) 66.03 (SynM)
Emotion Topic	(#500) 56.75 (BM) 70.51(SvdM)	(#500) 50.02 (BM) 61.98 (SynM)

7. CONCLUSIONS AND FUTURE WORK

The information of emotional expressions, holders and topics acquired from the present systems are being embedded for other due systems. The resulting document level emotion tagger can be used in an emotion based Information Retrieval system. Since development of annotated data is time consuming and also prone to error, research efforts will be directed towards applying machine learning techniques that work with unlabeled data. Emotion analysis related to the effect of metaphors (especially in blogs) is another research area to be explored in future.

8. ACKNOWLEDGMENTS

The work reported in this paper was supported by a grant from the India-Japan Cooperative Programme (DSTJST) 2009 Research project entitled "Sentiment Analysis where AI meets Psychology" funded by Department of Science and Technology (DST), Government of India.

9. REFERENCES

[1] Quan, C., and Ren, F. 2009. Construction of a Blog Emotion Corpus for Chinese Emotional Expression Analysis. In *EMNLP-ACL*, pp. 1446-1454

[2] Das, D. and Bandyopadhyay, S. 2011. Temporal Analysis of Sentiment Events – A Visual Realization and Tracking. *CICLing- 2011, LNCS 6581*.

[3] Wiebe, J., Wilson, T. and Cardie, C. 2005. Annotating expressions of opinions and emotions in language. *LRE*, vol. 39(2-3), pp. 165-210.

[4] Pang B., and Lee, L. 2008. Opinion Mining and Sentiment analysis. *Foundations and Trends in Information Retrieval*. vol. 2(1-2), pp. 1-135

[5] Yang, C., Lin, K. H. Y., and Chen, H. H. 2009. Writer Meets Reader: Emotion Analysis of Social Media from both the Writer's and Reader's Perspectives. In *009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, pp.287-290

[6] Salovey, P. and Mayer, J. 1990. Emotional Intelligence. *Imagination, Cognition and Personality*, 9(3), pp. 185-211

[7] Das, D. and Bandyopadhyay, S. 2011. Incorporating Affects for Tracking Bloggers' Emotions. *FLAIRS*.

[8] Polanyi, L., and Zaenen, A. 2006. Contextual valence shifter. In *James G. Shanahan, Yan Qu, and Janyce Wiebe, editors, Computing Attitude and Affect in Text: Theory and Applications*, Springer chapter 1, pp. 1- 10

[9] Sood, S., and Vasserman, L. 2009. ESSE: Exploring Mood on the Web. In *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media (ICWSM)*

[10] Strapparava, C. and Valitutti, A. 2004. Wordnet-affect: an affective extension of wordnet. In *4th LREC*, pp. 1083-1086

[11] Turney, P.D. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the 40th ACL*, pp. 417- 424

[12] Alm, C. O., Roth, D. and Sproat, R. 2005. Emotions from text: machine learning for text-based emotion prediction. In *HLT-EMNLP*, pp. 579 – 586

[13] Ku, L. W., Liang, Y. T. and Chen, H. H. 2006. Opinion extraction, summarization and tracking in news and blog corpora. In *AAAI-CAAW2006,* pp. 100-107.

[14] Fukuhara, T., Nakagawa, H., and Nishida, T. 2007. Understanding Sentiment of People from News Articles: Temporal Sentiment Analysis of Social Events. In *ICWSM'2007*

[15] Strapparava, C. and Mihalcea, R. 2007. SemEval-2007 Task 14: Affective Text. In *45th ACL*

[16] Mishne, G., and Rijke, M. de. 2006. Capturing Global Mood Levels using Blog Posts. In *(AAAI-CAAW2006)*, pp. 145-152

[17] Neviarouskaya, A., Prendinger, H., and Ishizuka, M. 2007. Narrowing the Social Gap among People Involved in Global Dialog: Automatic Emotion Detection in Blog Posts. *ICWSM*.

[18] Bethard S., Yu, H., Thornton, A., Hatzivassiloglou, V., and Jurafsky, D. 2004. Automatic Extraction of Opinion Propositions and their Holders. In *AAAI Exploring Attitude and Affect in Text: Theories and Applications*.

[19] Kim S. M., and Hovy, E. 2006. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. In *ACL*.

[20] Hu, J., Guan, C., Wang, M., and Lin, F. 2006. Model of Emotional Holder. In *Shi, Z.-Z., Sadananda, R. (eds.) PRIMA 2006 LNCS (LNAI)*, 4088, 534–539

[21] Kobayashi, N., K. Inui, Y. Matsumoto, K. Tateishi, and T. Fukushima. 2004. Collecting evaluative expressions for opinion extraction. *IJCNLP*.

[22] Yi, J., T. Nasukawa, R. Bunescu, and W. Niblack. 2003. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Proceedings of ICDM*

[23] Stoyanov, V. and C. Cardie. 2008. Annotating topics of opinions. In *Proceedings of LREC*

[24] Mishne, G. and Rijke, M. de. 2006. MoodViews: Tools for Blog Mood Analysis. In *AAAI-CAAW2006*

[25] Suzuki, Y., Takamura, H. and Okumura, M. 2007. Semi-Supervised Learning to Classify Evaluative Expressions from Labeled and Unlabeled Texts. In *IEICE Transactions on Information and Systems*, pp.1516-1522.

[26] Ekman, P.1993. Facial expression and emotion. *American Psychologist*, vol. 48(4) 384–392.

[27] Qiaozhu, M., Ling, X., Matthew W, Hang, S., and ChengXiang Z. 2007. Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs. In *Proceedings of the 16th international conference on World Wide Web*, pp.171-180.

[28] Esuli, A and Sebastiani, F. 2006. SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. *LREC-06*

[29] Kipper-Schuler, K. 2005. VerbNet: A broad-coverage, comprehensive verb lexicon. Ph.D. thesis, Computer and Information Science Dept., University of Pennsylvania

[30] Das, D. and Bandyopadhyay, S. 2010. Labeling Emotion in Bengali Blog Corpus – A Fine Grained Tagging at Sentence Level. In *Proceedings of the 8th Workshop on Asian Language Resources (ALR8), COLING-2010*, pp. 47-55

[31] Das, D. and Bandyopadhyay, S. 2010. Developing Bengali WordNet Affect for Analyzing Emotion. *ICCPOL-2010*, pp. 35-40

[32] Das, D. and Bandyopadhyay, S. 2010. Sentence Level Emotion Tagging on Blog and News Corpora. *Journal of Intelligent System (JIS)*, vol. 19(2), pp. 125-134.

[33] Das, D. and Bandyopadhyay, S. 2009. Word to Sentence Level Emotion Tagging for Bengali Blogs. In *ACL-IJCNLP 2009 (Short Paper),* pp.149-152

[34] Das, D. and Bandyopadhyay, S. 2009. Sentence Level Emotion Tagging. In *the proceedings of the 2009 International Conference on Affective Computing & Intelligent Interaction.* pp. 375-380

[35] Das, D. and Bandyopadhyay, S. 2009. Emotion Tagging – A Comparative Study on Bengali and English Blogs. In *ICON-09*. pp.177-184

[36] Das, D. and Bandyopadhyay, S. 2010. Sentence to Document Level Emotion Tagging – A Coarse-grained Study on Bengali Blogs. In *proceedings of 2nd MCPR,* pp. 332-341

[37] Das, D. and Bandyopadhyay, S. 2010. Emotion Holder for Emotional Verbs – The role of Subject and Syntax. In *CICLing- 2010, A. Gelbukh (Ed.), LNCS 6008*, pp. 385-393

[38] Das, D. and Bandyopadhyay, S. 2010. Extracting Emotion Topics from Blog Sentences – Use of Voting from Multi-Engine Supervised Classifiers. In *2nd SMUC Workshop of 19th ACM International CIKM,* pp. 119-126

[39] Das, D. and Bandyopadhyay, S. 2010. Identifying Emotion Topic - An Unsupervised Hybrid Approach with Rhetorical Structure and Heuristic Classifier. In *proceedings of the 6th IEEE-NLPKE*, doi: 10.1109/NLPKE.2010.5587777

[40] Das, D. and Bandyopadhyay, S. 2010. Identifying Emotional Expressions, Intensities and Sentential Emotion Tags using A Supervised Framework. In *24th PACLIC***,** Japan

[41] Das, D. and Bandyopadhyay, S. 2010. Identifying Emotion Holder and Topic from Bengali Emotional Sentences. *In ICON 2010*, IIT Kharagpur, India

[42] Banerjee, S. Das, D. and Bandyopadhyay, S. 2010. Classification of Verbs –Towards Developing a Bengali Verb Subcategorization Lexicon. In *GWC 2010*, Mumbai.

[43] Aman, S. and Szpakowicz, S. 2007. Identifying Expressions of Emotion in Text. *V. Matoušek and P. Mautner (Eds.):TSD 2007, LNAI 4629*, 196–205

[44] Lafferty, J., McCallum, A.K., and Pereira, F. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *proceedings of 18th International Conference on Machine Learning*

[45] Joachims, T.1998. Text Categorization with Support Machines: Learning with Many Relevant Features. In *European Conference on Machine Learning* , pp. 137-142

[46] Liu, T, Moore, A., Gray, A. and Yang, K. 2004. An Investigation of Practical Approximate Nearest Neighbor Algorithms. *LIU-NIPS 04*

[47] Mann, W. C. and Thompson, S. 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization, *TEXT,* vol. 8, pp. 243–281.