IJCLA VOL. 3, NO. 1, JAN-JUN 2012, PP. 25–39
RECEIVED 31/10/11 ACCEPTED 09/12/11 FINAL 13/06/12

Hindi Subjective Lexicon Generation using WordNet Graph Traversal

PIYUSH ARORA, AKSHAT BAKLIWAL, AND VASUDEVA VARMA

International Institute of Information Techonology, India

ABSTRACT

With the induction of UTF-8 unicode standards, web content in Hindi language is increasing at a rapid pace. There is a great opportunity to mine the content and get insight of sentiments and opinions expressed by people and various communities. In this paper, we present a graph based method to build a subjective lexicon for Hindi language, using WordNet as a resource. Our method takes a pre-annotated seed list and expands this list into a full lexicon using synonym and antonym relations. We show two different evaluation strategies to validate the Hindi Lexicon built. Main contribution of our work 1) Developing a Subjective lexicon of adjectives using Hindi WordNet. 2) Developing an annotated corpora of Hindi reviews.

KEYWORDS: *Hindi Language, Sentiment Analysis, Adjective Polarity Lexicon, WordNet, Graph Traversal*

1 INTRODUCTION

In past 8–10 years, we have seen an enormous increase in web content in Hindi language. This information is important to be mined for the use of/by researchers, industries and government(s). A large number of advertising industries and recommendation systems work on understanding the people likings and tastes from this content. Most of the earlier work targets sentiment and opinion analysis in resource rich languages like English. Our work addresses the problem of identifying sentiments and

opinions from user generated content in Hindi and builds a model (Subjective Lexicon) using Hindi WordNet.

Hindi language has approx. 409 million native speakers as in 1999[1] and with unicode (UTF-8) standards for Hindi introduced, web pages catering Hindi is increasing on a rapid pace. There are many websites which cater information in Hindi[2] and to the best of our knowledge there are very few works [1, 2] in this field for Indian languages. This part of the web hasn't been explored much in the direction of sentiment and opinion analysis.

In this paper, we present a method of building a subjective lexicon for Hindi language with dependency only on WordNet and a small pre-annotated seed list. Using WordNet and simple graph traversal method we construct the subjectivity lexicon. In our method, initially a small seed list of words is decided along with their polarity. Using WordNet this seed list is populated based on the synonyms and antonyms of the words in the list. Here, we make an assumption that synonyms possess similar polarity and antonyms show opposite polarity.

The road map for rest of the paper is as follows: Existing related works are presented in Section 2. Section 3 presents a comprehensive view of the approach proposed in this research work. Section 4 gives details about the lexicon generated using this proposed method for Hindi language. Section 5 gives a description of product review dataset for Hindi language. In Section 6, we describe the different methods of evaluation used in this research. In Section 7, we discuss the results and limitations of this system. Section 8 presents the conclusions along with directions for our future work.

2 Related Work

Research in the field of sentiment analysis is done at various levels: Document Level [3, 4], Sentence Level [5–9] and Word or Phrase Level [10, 11].

In 1966, IBM developed the General Inquirer system [12], which marked the beginning of sentiment extraction from plain text. This system was termed as content analysis research problem in behavior science and comprised of 11789 words, with each word having at least one instance.

[1] en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers

[2] hindiblogs.org, hindigear.com, bbc.co.uk/hindi

In 1998, the authors of [13] developed a method to predict semantic orientation of adjectives. Their idea consisted in predicting the semantic orientation of adjectives based on the nature of conjunctive joining the two adjectives. A log-linear regression model uses these constraints to predict whether conjoined adjectives are of same or different orientations, achieving 82% accuracy in this task when each conjunction is considered independently.

In 2002, Turney [4] extended the work [13] to other POS-tags. Turney used adverbs and nouns along with adjectives for performing opinion classification on reviews. He achieved 84% accuracy on automobile review classification compared to 66% on movie reviews.

For English, a good amount of work is done in the lines of generating subjective lexicon. SentiWordNet [14, 15] was developed in year 2006 by Esuli and Sebastiani. It contains four Part-of-Speech tags namely adjectives, adverbs, verbs and nouns with $\sim$2 million words out of which 3% are adjectives. Each word is assigned three scores positive, negative and objective (Equation 1). SentiWordNet was built using WordNet and a ternary classifier. Their classifier is based on "bag of synset" model which uses manually disambiguated glosses available from the Princeton WordNet Gloss Corpus.

$$positive\ score + negative\ score + objective\ score = 1. \quad (1)$$

Banea *et. al.* [16] proposed a bootstrapping method for building subjective lexicon for under-resourced languages. Their method build a subjective lexicon using a small seed list (60 words), an online dictionary (Romanian Dictionary) and a small annotated corpora. They used word level similarity (LSA and PMI) to filter words. In their bootstrapping method the initial seed list was manually selected and contained 60 words, which were evenly distributed among adjectives, adverbs, nouns, and verbs.

Kamps *et. al.* [17] tried to determine sentiments of adjectives in WordNet. In this work, they divided adjectives into four major categories and used base words (to measure relative distance) depending on the category. For category Evaluative their base words were "good" and "bad", for category Activity their base words were "active" and "passive", etc. The polarity orientation of a word 'w' belongs to range [-1,1], -1 for words on bad side and 1 for words on good side. Based on this method, they populated a total of 1608 words in all four categories with avg. correctness of 67.18% for English.

Kim and Hovy [18] proposed a method of identifying and analyzing judgement opinions. This was a four step process in which first step was recognizing the opinion. For identifying the opinion they introduced an algorithm to classify a word as positive, negative or objective which was based on WordNet.They made an assumption which was to add synonyms of a word with the same polarity as the source word. To avoid words with multiple meaning (dual nature) they applied a method to identify closeness of a word to each category (positive, negative, objective). For their proposed method to give high recall the initial seed list should be large enough and with wide variety of words.

Rao and Ravichandran [19] presented an extensive study on the problem of detecting polarity of words. They considered bi-polar classification of words i.e. a word can be either positive or negative. They performed semi-supervised label propagation in graph for polarity detection of words. Each of these words represent a node in the graph whose polarity is to be determined. They focused on three languages mainly English, French and Hindi but claim that their work can be extended to any other language for which WordNet is available.

As far as Indian Languages are concerned, we can see small amount of work done in Hindi and Bengali. Das and Bandhopadyay [1] developed SentiWordnet for the Bengali language. They applied word level lexical-transfer technique to each entry in English SentiWordNet using an English-Bengali Dictionary to obtain a Bengali SentiWordNet. This process resulted in 35,805 Bengali entries. In [20], authors devised four strategies to predict the sentiment of a word. First approach, an interactive game which in turn annotated the words with their polarity. Second approach, using Bi-Lingual dictionary for English and Indian Languages. Third approach, wordnet expansion using synonym and antonym relations, but their article missed the approach they followed for this expansion. Fourth approach, learning from pre-annotated corpora.

Joshi *et. al.*[2] created H-SWN (Hindi-SentiWordNet) using two lexical resources namely English SentiWordNet and English-Hindi WordNet Linking [21]. Using WordNet linking they replaced words in English SentiWordNet with equivalent Hindi words to get H-SWN.

Our work is directed towards the Hindi Language. It is related to works by Kim and Hovy [18] and Rao and Ravichandran [19]. Kim and Hovy restricted their assumption to synonyms; we extend the relation to antonyms. Rao and Ravichandran performed bi-polar classification; we extend it to a third level: objectivity. In this work, we use Hindi WordNet [22] to obtain the polarity of adjectives for Hindi Subjective Lexicon.

3 ALGORITHM

Our algorithm makes a hypothesis of traversing WordNet like a graph where every word in WordNet is imagined as a node in graph. This graph will be an undirected graph and will be highly connected but not fully connected. In this graph, nodes are connected to each other on synonym and antonym relations. Kim and Hovy, 2006 [18] made a simple assumption that synonyms carry the same sentiment as the word. We extend his assumption to antonyms, we assume antonyms carry opposite polarity. Each node will be connected to many other nodes i.e. each node will have many in-links and many out-links. This graph has three basic connected components (positive, negative and objective). Words can be connected using two kinds of edges:

1. Simple Edge: An edge which connects two words in the same (different) domain and represents a synonym (antonym) relation with a given condition that each word should belong to non overlapping region.
2. Cross Edge: An edge which connects two words based on synonym (antonym) relation and atleast one among these words lies in the overlapping region.

Cross Edges among these connected components produce words which have ambiguous (dual) nature.

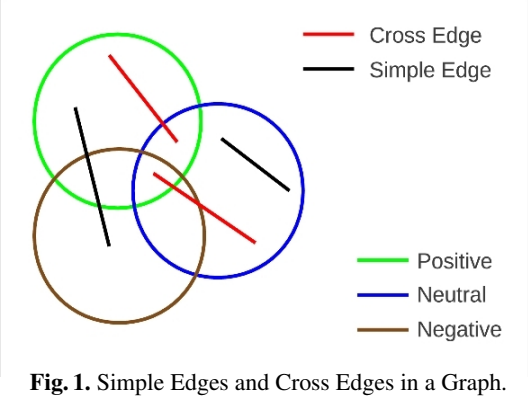

Fig. 1. Simple Edges and Cross Edges in a Graph.

Fig. 1 explains Simple Edges and Cross Edges pictorially. In Fig. 1 each circle represents a connected component and overlapping zone contains words which are ambiguous by nature.

Here, we use Hindi WordNet and a list of words (seed list) which is pre-annotated based on polarity. Our seed list contains 15 objective, 15 positive and 15 negative words. Each word in this seed list will be expanded on synonym and antonym relations. We can consider this expansion as a Breadth First Traversal of a graph. In this traversal method all the siblings (nodes at any depth d) are expanded before any node at the next level (depth d+1) is expanded. We make use of queue data structure to maintain the order in which the nodes (words) are introduced or expanded. This method helps us to ensure that each node is expanded only once and all the synonyms of a word are traversed at the same time.

In our method we have 2 lists, one is temporary and the other is final list. The initial seed list which contains 45 words is copied to temporary seed list with the polarity. Now every time we get a word (a structure which contains a pair of seed and polarity) from the temporary seed list by de-queuing it from the list, we check for this word if it exists in the final seed list or not. If this word is in the final seed list then we don't populate this word further, we just add the current polarity of this word to the polarity in the final list. But if this word is not in the final list, we do three things

1. Add this word to the final list with the current polarity
2. Find out all the synonyms of this word and en-queue them in the temporary seed list with the polarity same as the source word.
3. Find out all the antonyms of this word and en-queue them in temporary seed list with opposite polarity. (P -> N, O -> O, N -> P).

We continue this process till all the words in the temporary seed list are explored or in other words till the temporary seed list becomes empty. When the temporary seed list becomes empty the final seed list contains adjectives and against each adjective we have string of P's, N's and O's. Based on this we decide the final polarity of the word. Say for a word 'x' in the final seed list we have string 's' made of P's, N's and O's.

$$\begin{aligned} Length\ of\ string\ (s) &= Len \\ Number\ of\ P's\ in\ s &= nP \\ Number\ of\ N's\ in\ s &= nN \\ Number\ of\ O's\ in\ s &= nO \\ Positive\ polarity\ of\ x &= nP/Len \\ Negative\ polarity\ of\ x &= nN/Len \\ Objective\ polarity\ of\ x &= nO/len \end{aligned} \tag{2}$$

For a pseudocode, refer to Algorithm 1.

4 HINDI SUBJECTIVE LEXICON

Lexicon[3] built using the above mentioned approach for Hindi language contains 8048 words in all. Out of 8048 words 2521 are positive, 3060 are negative and 2467 are neutral. For each word, our lexicon provides three scores: positive, negative and objective. The sum of these scores amounts to 1.

$$(Positive + Negative + Objective)Score = 1 \tag{3}$$

We validate this lexicon by two different methods which are explained in Section 6.

5 PRODUCT REVIEW DATA

This dataset[3] was translated from English to Hindi using Google[4] translate[5]. We translated pre-annotated Amazon product reviews [23] of length $\leq$ 25. We took this threshold of 25 words in order to avoid (reduce) translations errors. After translating the product reviews we asked human judges to manually validate the translation. Table 1 summarizes the data (reviews) generated by translation.

Table 1. Product Review Data Summary

Total Positive Reviews	900
Manually Corrected Reviews	350
Total Negative Reviews	900
Manually Corrected Reviews	350
Total Annotated Reviews	350 + 350

[3]This Resource is in the initial stage of development and is avaliable for non-commercial and research usage on request. Request should be made to any of the authors.

[4]http://translate.google.com/

[5]We made an assumption that while translation sentiment bearing words are translated correctly without any loss or modification of sentiments.

Algorithm 1 Algorithm for Populating SeedList using WordNet to generate Subjective Lexicon

$InitialSeedList = \{45\ words\}\ (15\ \times\ objective,\ positive,\ negative)$
$//\ Each\ word\ is\ a\ structure\ which\ contains\ a\ pair:\ Seed,\ Polarity$
$FinalSeedList = \{\}$
$TempSeedList = \{\}$
$TempSeedList = InitialSeedList$
while $TempSeedList \neq EmptyList$ **do**
  $Word\ =\ TempSeedList.pop()\ //\ pop\ the\ first\ word\ out\ of\ the\ list$
  $Seed = Word[0]$
  $Polarity = Word[1]$
  if $Seed \in$ FinalSeedList **then**
    $FinalSeedList[Seed] = FinalSeedList[Seed] + Polarity$
  else
    FinalSeedList[Seed]=Polarity
    SynonymSet = All the synonyms of Seed
    AntonymSet = All the antonyms of Seed
    for all $synonyms \in$SynonymSet **do**
      $TempSeedList.append(synonym : Polarity)$
      $//\ Polarity\ will\ be\ P/N/O$
    end for
    for all $antonyms \in$AntonymSet **do**
      $TempSeedList.append(antonym : OppPolarity)$
      $//\ OppPolarity\ will\ be\ P\ if\ Seed\ has\ Polarity\ N$
      $//\ OppPolarity\ will\ be\ N\ if\ Seed\ has\ Polarity\ P$
      $//\ OppPolarity\ will\ be\ O\ if\ Seed\ has\ Polarity\ O$
    end for
  end if
end while
$//\ Against\ Each\ adjective\ in\ the\ FinalSeedList\ we\ have\ a\ string$
$//\ of\ P's,\ N's,\ and\ O's,\ which\ contains\ the\ polarity\ of\ that\ word$
for all $adjectives \in$FinalSeedList **do**
  $S\ =\ FinalSeedList[i]\ //\ Here\ i\ is\ an\ adjective\ and\ S\ is\ the\ string$
  $//\ of\ polarity\ for\ that\ adjective$
  $nP\ =\ Number\ of\ P's\ in\ S$
  $nN\ =\ Number\ of\ N's\ in\ S$
  $nO\ =\ Number\ of\ O's\ in\ S$
  $Len\ =\ length\ of\ S\ //\ Note:\ nP\ +\ nN\ +\ nO\ =\ Len$
  $PositivePolarity\ =\ nP/Len$
  $NegativePolarity\ =\ nN/Len$
  $Objectivity\ =\ nO/Len$
  $//Note:\ PositivePolarity + NegativePolarity + Objectivity = 1$
end for

6 EVALUATION

One of the major task while proposing a new method is evaluation. In these kind of systems we mainly evaluate by human judgement or by classifying some pre-annotated text. These are a few methods which are commonly used for validation.

1. Human Judgement: This method is usually opted for languages which are scarce resource languages. In this method, some manual annotators are appointed whose task is to annotate the lexicon generated and later, taking the majority vote of annotators the system generated lexicon is validated.
2. Classification: In this method of evaluation, we classify pre-annotated reviews/blogs using our system generated lexicon and find precision, recall, F1 Scores, etc to show the correctness. This strategy is generally used for resource rich languages or for those languages for which we have pre-annotated data.
3. Validating Against Existing Resources: In this strategy of evaluation, we find the accordance of lexicon generated using our approach with a lexicon which is already proposed and accepted by the research community. This strategy of evaluation is used for languages which are resource rich.

Subsequent sub-sections explain two methods which we used to evaluate the lexicons generated by our system.

6.1 *Human Judgement*

In this method of evaluation, we hired five manual annotators[6] who are language experts in Hindi. We asked each annotator to tag the words generated by our system on the scale of 3 (negative:-1, neutral:0, positive:1). After getting the list annotated by all the annotators, we had five votes for each word and we took the majority call. Table 2 reports accordance of Hindi lexicon generated using our system with manual annotation.

Reason behind low mutual agreement among the annotators is that many words in Hindi show ambiguous nature. Their polarity depends on the sense in which they are used. This ambiguous nature is highlighted in Fig. 2.

[6]None of the authors were annotators for this task.

Table 2. Results for Manual Agreement for Hindi Lexicon

Mutual agreement among the annotators	**70.48%**
Agreement of each annotator with our lexicon	
Annotator 1	66.08%
Annotator 2	64.01%
Annotator 3	**68.45%**
Annotator 4	66.70%
Annotator 5	68.34%
Overall Agreement of our lexicon with the annotators	**68.80%**

6.2 *Review Classification*

For this evaluation strategy, we performed classification on product review dataset described in Section 5. On this data, we performed unigram presence and simple scoring method classification. In unigram presence method, we count unigrams of positive, negative and objective polarity and assigned the polarity for which the count was highest. In simple scoring method, we summed the positive, negative and objective scores of each adjective and assigned the polarity of the dominant score. From every review we identified adjective and scored those adjectives using our lexicon. If an adjective was missing from our lexicon we considered the stemmed variant[7] [8] of that word for scoring. In addition to stemming we also performed negation handling. We identified the words with tag "NEG" using sliding window of 6 words and swapped the polarity (positive and negative) of adjectives in this range. Our sliding window, looked upto 3 words in both the directions (left and right) of this word. Table 3 reports the results of classification.

7 DISCUSSION

Results in Table 3 highlights the point that our scoring method performs better than mere presence counting of the adjectives. The following example shows how the presence counting failed in classification while the scores assigned to each word by our method correctly classified the review. "बहुत ही प्रतिभाशाली संगीतकारो का एक बिलकुल बेकार प्रयास".

[7]We used the stemmer which is bundled with Hindi WordNet API 1.2

[8]cfilt.iitb.ac.in/wordnet/webhwn/index.php

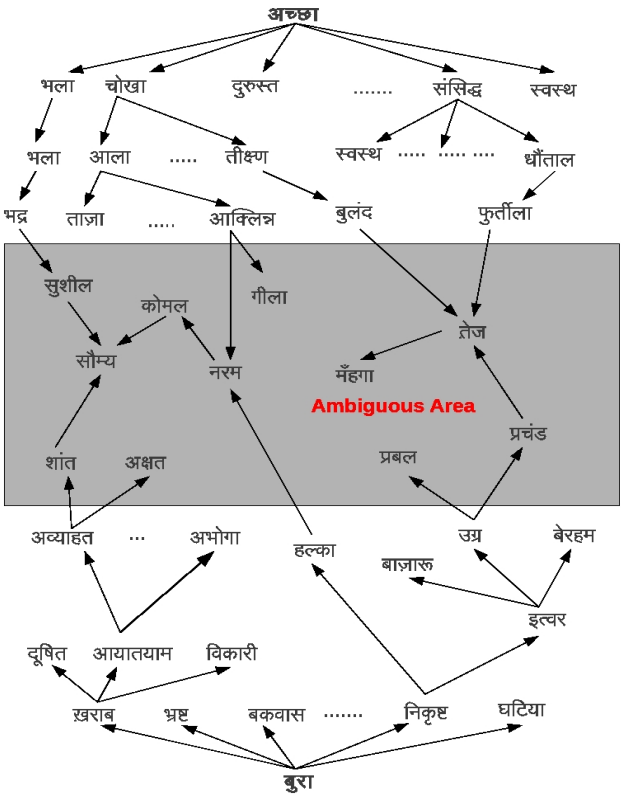

Fig. 2. The Graph Traversal of words in Hindi WordNet. The dark portion shows ambiguous words.

Adjectives in this example are "प्रतिभाशाली" (positive) and "बेकार" (negative). If we account for presence of adjectives for classification, this review becomes neutral. However, scores generated using our system $(+0.75, -0.0, 0.25)$ for "प्रतिभाशाली" and $(+0.0, -1.0, 0.0)$ for "बेकार" with overall score for the review as $-0.25([0.75 - 0.0] + [0.0 - 1.0])$, classify the review correctly (as negative).

Using the proposed strategy for negation handling, results in Table 3 show $\sim$3% improvement in classification of reviews. We proposed the use of stemmer to identify the root word for adjectives which were

Table 3. Results for Product Review Classification using Lexicon generated by our approach

Method	Accuracy
Adjective Presence	
Baseline	65.50
Baseline + Negation Handling	68.67
Baseline + Stem	67.17
Baseline + Stem + Negation Handling	70.80
Adjective Scoring	
Baseline	67.33
Baseline + Negation Handling	70.00
Baseline + Stem	71.00
Baseline + Stem + Negation Handling	**74.10**

present in the review but went missing from our lexicon. Stemming also showed an improvement of $\sim$3% in classification of reviews. Table 4 lists a few mapping of words to their stemmed form.

Table 4. Words and their stemmed (root) words

Word(s)	Stemmed Word
छोटे	छोटा
अच्छी, अच्छे	अच्छा
बड़ी	बड़ा
हल्के	हल्का
लंबे	लंबा

There are a few limitations and issues with the current version of algorithm proposed above.

- To handle adjectives which were present in reviews and were missing from our lexicon, we performed stemming. If an adjective was missing from our lexicon we stemmed the adjective to get its root word. Instead of using a stemmer if a morph is used, then we expect results to improve.
- The current version of this algorithm does not perform Word Sense Disambiguation (WSD).
- Scope of the system proposed above is dependent on the initial seed list used to populate the WordNet. If we choose the seed list in a

careful manner with the help of linguistic experts, the results and scope of the Lexicon thus generated would be better.

8 Conclusion and Future Work

We proposed a graph based method to generate the subjectivity lexicon for Hindi and explored how the synonym and antonym relations can be exploited using graph traversal. Our method is language independent and just uses only one resource (WordNet) for Lexicon generation. As a part of this research, we worked on Hindi language. The lexicon generated using our proposed algorithm contains 8048 words and it achieved 74% accuracy on classification of reviews and 69% in agreement with human annotators for Hindi.

In future, this work can be extended to incorporate Word Sense Disambiguation (WSD) to emphasize more on senses of a word. Another extension can be morphological variants which could result in better accuracy for words which have dual nature. We experimented only with adjectives and this work can be extended for other parts of speech.

References

1. Das, A., Bandyopadhyay, S. In: SentiWordNet for Bangla. Linguistic Data Consortium for Indian Languages (LDC-IL) (2010)
2. Joshi, A., R, B.A., Bhattacharyya, P.: A fall-back strategy for sentiment analysis in Hindi: a case study (2010)
3. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP). (2002) 79–86
4. Turney, P.: Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews (2002)
5. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: KDD. (2004) 168–177
6. Intelligent, T.W., Wilson, T.: Annotating opinions in the World Press. In: In SIGdial-03. (2003) 13–22
7. min Kim, S.: Determining the sentiment of opinions. In: In Proceedings of COLING. (2004) 1367–1373
8. Wiebe, J.M., Bruce, R.F., O'Hara, T.P.: Development and use of a gold-standard data set for subjectivity classifications. In: Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics. ACL '99, Stroudsburg, PA, USA, Association for Computational Linguistics (1999) 246–253

9. Yu, H., Hatzivassiloglou, V.: Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In: Proceedings of the 2003 conference on Empirical methods in natural language processing. EMNLP '03, Stroudsburg, PA, USA, Association for Computational Linguistics (2003) 129–136
10. Agarwal, A., Biadsy, F., Mckeown, K.R.: Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams (2009)
11. Wilson, T.: Recognizing contextual polarity in phrase-level sentiment analysis. In: In Proceedings of HLT-EMNLP. (2005) 347–354
12. Stone, P.J., Dunphy, D.C., Smith, M.S., Ogilvie, D.M.: The General Inquirer: A Computer Approach to Content Analysis. MIT Press (1966)
13. Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics. ACL '98, Stroudsburg, PA, USA, Association for Computational Linguistics (1997) 174–181
14. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Chair), N.C.C., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., Tapias, D., eds.: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta, European Language Resources Association (ELRA) (may 2010)
15. Esuli, A., Sebastiani, F.: SENTIWORDNET: A publicly available lexical resource for opinion mining. In: In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC-06). (2006) 417–422
16. Carmen Banea, R.M., Wiebe, J.: A bootstrapping method for building subjectivity lexicons for languages with scarce resources. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odjik, J., Piperidis, S., Tapias, D., eds.: Proceedings of the Sixth International Language Resources and Evaluation (LREC'08), Marrakech, Morocco, European Language Resources Association (ELRA) (may 2008) http://www.lrec-conf.org/proceedings/lrec2008/.
17. Kamps, J., Marx, M., Mokken, R.J., Rijke, M.D.: Using wordnet to measure semantic orientation of adjectives. In: National Institute for. (2004) 1115–1118
18. min Kim, S., Hovy, E.: Identifying and analyzing judgment opinions. In: Proceedings of HLT/NAACL-2006. (2006) 200–207
19. Rao, D., Ravichandran, D.: Semi-supervised polarity lexicon induction. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. EACL '09, Stroudsburg, PA, USA, Association for Computational Linguistics (2009) 675–682
20. Das, A., Bandyopadhyay, S. In: SentiWordNet for Indian Languages. Asian Federation of NLP (AFNLP) (2010)
21. Karthikeyan, A.: Hindi English WordNet linkage. (2010)

22. Dipak Narayan, Debasri Chakrabarti, P.P., Bhattacharyya, P.: An experience in building the Indo WordNet—a WordNet for Hindi. In: First International Conference on Global WordNet. (2002)
23. Blitzer, J., Dredze, M., Pereira, F.: Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification. In: ACL. (2007)

PIYUSH ARORA
SEARCH AND INFORMATION EXTRACTION LAB, LTRC,
INTERNATIONAL INSTITUTE OF INFORMATION TECHONOLOGY,
HYDERABAD, INDIA
E-MAIL: <PIYUSH.ARORA@RESEARCH.IIIT.AC.IN>

AKSHAT BAKLIWAL
SEARCH AND INFORMATION EXTRACTION LAB, LTRC,
INTERNATIONAL INSTITUTE OF INFORMATION TECHONOLOGY,
HYDERABAD, INDIA
E-MAIL: <AKSHAT.BAKLIWAL@RESEARCH.IIIT.AC.IN>

VASUDEVA VARMA
SEARCH AND INFORMATION EXTRACTION LAB, LTRC,
INTERNATIONAL INSTITUTE OF INFORMATION TECHONOLOGY,
HYDERABAD, INDIA
E-MAIL: <VV@IIIT.AC.IN>