

A Survey on Opinion Mining Of Associations for Related Entities from Hindi Text: Future Research Opportunities

¹Arpana Prasad, ²Neeraj Sharma

¹Department of Computer Science, Punjabi University, Patiala, Punjab, India

²Department of Computer Science, Punjabi University, Patiala, Punjab, India

E-mail: ¹arpanaprasad2013@gmail.com, ²sharma_neeraj@pbi.ac.in

ABSTRACT

Opinion mining is a field dedicated towards classification of writer's opinion. Considerable work in opinion mining has been done for deriving writer's opinion from English unstructured text appearing on the web and in mobile applications. There is an urgent need to develop resources and identify approaches that facilitate to scale up opinion mining solutions for Hindi unstructured web and mobile application contents. This survey has been conducted with an objective to trace the development in this field and to identify gaps and derive research opportunities for opinion mining from unstructured text in Hindi language. Highlights of the analysis are the identified research opportunities in polarity classification of associated entities. Various areas that can be explored for opinion mining from web contents in Hindi and languages other than English are suggested in this paper. A proposed software framework for polarity classification of associated entities is presented in this paper. With slight domain specific modifications, the proposed framework can be re-used for developing opinion mining system for texts with cause and effect relationship between entities in other domains and languages also.

Keywords: *Opinion Mining, Hindi Unstructured Text, Research Opportunities, Association Polarity Classification*

1. INTRODUCTION

News highlight, "Hindi content consumption on internet growing at 94%: Google"¹, and similar highlights from renowned sources, gives adequate motivation, that varied computational solutions for web contents that exist for input text in different popular languages like English, must be extended at proportional scale to support Hindi also. One such opportunity lies in the computational field of opinion mining (OM) which also has task specific synonyms; sentiment analysis, opinion extraction, sentiment mining, subjectivity analysis, effect analysis, emotion analysis and review mining. This field has vivid applications in the domains of - business analytics, social reforms, politics, agriculture, youth psychology analysis and highlighting future research directions etc. It is considered a very challenging field of Information Retrieval. Approaches of Information Retrieval (IR), Artificial Intelligence (AI), and Natural Language Processing (NLP), Data Mining, Web Mining, and Deep Learning are used for opinion mining. These underlying fields of opinion mining have a long history but specific task of deriving writer's opinion and sentiment has gained interest of the researchers after the year 2000. Lot of algorithms to aggregate opinions and classify opinions is available for text written in various languages. Linguistic differences amongst various languages make an opinion mining algorithm written for one language unusable for deriving opinion from text of another language. Lots of innovative applications have been developed for opinion mining from text in English. Research in this field for

deriving opinion from text in other languages has also gained a lot of momentum. With the availability of text on the Web in Asian languages, research in this field for Asian languages has caught the attention of several researchers. It is not only the variation of language that has caught the attention of the researchers; the researchers have innovatively explored various domains for the purpose. This field which had initially witnessed research in the domain of movie-review, customer's product review has gradually witnessed research in very challenging domains like legal blogs, newspaper quotations, politics, bio-medical text, opinion mining on citation, finding accuracy of voting approach, opinion on nuclear power, opinion about people etc. Learners are generally curious to know the difference amongst 'Data Mining', 'Text Mining', 'Web Mining' and 'Opinion Mining'. Data Mining is a field that is used to process and analyze large datasets in order to discover hidden patterns, behavior, and information. The datasets generally picked for the purpose are in the form of records that are usually transactional in nature. Data Mining is an inter-disciplinary field involving statistical techniques, machine learning² and artificial intelligence. For example, transactional data of an online shopping site can be analyzed using data mining tools to find out gender specific buying patterns, like items belonging to which categories are purchased together by customers of which gender. Text Mining is a field related to data mining. In contrast to data mining, it mainly deals with mining information from textual unstructured data. It does not deal with transactional records. It is an interdisciplinary field involving statistical techniques, machine learning and computational linguistics. Computational linguistics is not used in Data Mining. Text Mining uses complex techniques of NLP. An application

¹ <http://economictimes.indiatimes.com/tech/internet/hindi-content-consumption-on-internet-growing-at-94-google/articleshow/48528347.cms>

² An algorithm that automatically learns how to make predictions based on the past observations.

©2012-20 International Journal of Information Technology and Electrical Engineering

developed to comprehend the feedback mails written by customers of an online shopping site and finding patterns of relationships amongst those mails, is an example of an application of this field. Web Mining involves automatic crawling of the Web to find facts, patterns, and activities. Web Mining works by sending intelligent agents to target uniform resource locators (URLs). For example, an online shopping site may use tools of web mining to track customer's online behavior by tracking cookies and correlations of hyperlinks. Opinion Mining is specific task of mining opinions using tools of data mining, text mining, web mining or their combinations. The online shopping site may use an opinion mining system to aggregate opinions from product review section and weblogs. Keen interest of the industry and wide applications of commercial value has made opinion mining a popular field amongst researchers. Opinion mining task involves evaluation on some popular dimensions. The popular dimensions are document level, sentence level and aspect level opinion mining.

The paper surveys the field with five key objectives: i) *to trace the application domains, approaches used, challenges encountered, resource(s) developed and trends of research in the field of opinion mining for English input unstructured text*; ii) *to analyze the state of research in this field for Hindi input unstructured text*; iii) *to identify gaps and derive research opportunities in opinion mining for Hindi unstructured input text*; iv) *to analyze the state of research in polarity classification of associations in related entities mentioned in unstructured web contents*. The paper presents major highlights of the analysis on the research in the field. Section 2 presents detail of our survey findings on state of significant research in this field. Findings about the research opportunities in this field, especially for Hindi input text data are presented in Section 3. Section 4 discusses an ongoing research on polarity classification of relationship of associated entities from Hindi unstructured text in the domain of home remedies. This section also presents the software architecture framework designed for this research. The conclusion drawn with major highlight of the survey is presented in Section 5.

2. REVIEW OF LITERATURE

This section highlights exemplary research in the field of opinion mining. Since this field has seen interest of researchers from the fields of NLP, Information Extraction and AI, contributions from all domains are included in the review. Survey of literature has been divided into the following sub-sections; Review of work in opinion mining from unstructured text in languages other than Hindi ; Review of research in opinion mining from unstructured input text in Hindi and other digital resource scarce languages ;Summary of reviewed literature.

2.1 REVIEW OF WORK IN LANGUAGES OTHER THAN HINDI

An early work in the field of Opinion Mining on movies review database used the Internet Movie Database

(IMDb) archive³ of the rec.arts.movies.reviews newsgroup as the data source for experimentation. Linguistic techniques specific to English were used to classify the sentiments. Various machine learning techniques were applied on the dataset developed by the researchers. The dataset is available online⁴. The evaluation results on the machine learning algorithms for sentiment classification problem highlighted that Naïve Bayes Algorithm tend to perform worst and Support Vector Machines (SVM) classification algorithm tends to perform best [1]. Boot strap approach by aggregating the use of several machine learning algorithms for sentiment classification of customer reviews that are in the form of free text. Annotation was done for a random sample of 3000 sentences. Sentence level sentiment analysis has been done in this research [2]. Significant survey paper with a maximum number of citations as per Google scholar search as accessed in March 2017 lists the techniques and approaches that enables opinion oriented information seeking[3]. Different methods used to handle challenges; benchmark datasets and list of evaluation campaigns are included in this survey [4]. Linguistic techniques have been used to identify opinion expressions from noisy text⁵. Semi-automatic approach for cleaning text using domain knowledge has been used. An approach has been proposed to deal with adverbial modifiers⁶. An opinion mining framework in the domain of vehicle and movie review has been built. Context based spelling correction has been developed as an important sub module [5]. Various versions of popularly used lexical resources for opinion mining i.e. SentiWordNet⁷ are developed in a significant research in this field. The latest version SentiWordNet 3.0 has sentiment mapping for all synsets⁸ appearing in WordNet⁹[6]. Investigation on key issues in blog mining in terms of (i) sentiments as classified by the writer (ii) mood of the reader (iii) content analysis of the blogs has been done in a research . Experiments on the performance of the approach have been conducted on the data from financial domain. The researcher reports high correlation between the sentiment of contents on blogs, message boards, and knowledge sources. The work is domain specific and extension to other domain is left as a future work. 'SenticNet with emotion labels' is contributed in a research in the field. Information on how International Survey for Emotion Antecedent and Reaction (ISEAR), an emotion related dataset, may be leveraged for the purpose of opinion mining is provided in the research [7].

³ www.imdb.com

⁴ Available on <http://www.cs.cornell.edu/people/pabo/movie-review-data/>

⁵Noisy text is an electronically stored text that cannot be processed directly by the text mining algorithm without pre-processing.

⁶ Examples of adverbial modifiers are very, mildly etc.

⁷ <http://sentiwordnet.isti.cnr.it/>

⁸Set of one or more synonyms.

⁹Word Net available for download at <https://wordnet.princeton.edu/wordnet/> is a lexical database for the English language i.e. a combination of dictionary and thesaurus.

Researchers have proposed a methodology for enhancing WordNetAffect¹⁰(WNA) and similar resources. Usage of multi-lingual corpora is left as future scope. Usage of syntactical and psychological concepts is also left as the future scope. A corpus from news articles has been formed. These articles are cleaned, stemmed, tagged, and classified as good or bad using an algorithm devised using Support Vector Machine approach [8]. Domain ontology pairing method for feature-opinion pair identification of product review in Chinese is presented in research [9]. Automatic domain specific lexicon construction to overcome the domain dependency in opinion mining to an extent has been proposed in a research [10]. Bi-lingual knowledgebase has been used for analyzing Chinese social media for Opinion. N-Gram and SVM classifier has been for text classification. The research highlights that the SVM classifier outperforms the N-Gram classifier [11]. Ontology based approach for automatic feature based opinion classification has been developed. The approach is domain independent but language dependent. The research is specific to English and Spanish [12]. Research has been done on mining of reader's implicit and explicit emotions from social media by devising two readers oriented sentiment topic models i.e. multi-label supervised topic model and sentiment latent topic model. These models connect latent topics with evoked emotion of the reader [13]. An algorithm using Fuzzy C Mean Clustering approach to mine telecommunication data record in order to understand human behavior for user segmentation has been proposed. Sixteen week long data from Mobile Service Providers has been used for the purpose. Fuzzy C mean clustering approach has been used in the research [14]. Knowledge based concept level sentiment analysis for emerging political topics has been used [15]. Opinion mining from Big Data has been done for Big Data Social Analytics. The work is specific to English text [16]. An exhaustive survey on approaches and methodology of opinion mining used by researchers from year 2000 to year 2015 is presented in a research. Resource creation for opinion mining task for ontology and non-ontology based approaches is discussed in the research. An exhaustive review on concepts and technique utilized and type of datasets is presented in the survey. Set of approaches that has not been used by researchers till date is an important derivative of the survey [4]. Extraction of features for analysis of online reviews for books using human intelligence to facilitate prospective buyers is an opinion mining application developed for a recommendation system. The domain of this research is product review specifically book review and the language of the unstructured text is English [17]. Systematic literature review on opinion mining studies from user app store review is presented in a research. Various parameters and effective parameters of mobile app store review is derived from this survey [18]. Analytical mapping of opinion mining research from the year 2000 to 2015 with year wise research pattern in the field on various parameters like year wise usage distribution of datasets,

year wise application domains, year wise journal details in which the publication in this field has been published, most productive author, approach used and distribution of the research work is presented in a recent survey [19]. Ontology on six emotions is developed in a research based on Google ranked top five social media sites. Researchers have built an ontology for multilingual environment using manually built classes in the domain of weather, books and shopping. Machine learning approach in development of ontology is a future direction of this research [20]. A novel work in the field in the domain of health care is presented in a research paper. The researchers have proposed opinion influencing factors which they justify affects the opinion of people. Methodology to discover semantic rules that can find the change in opinion is developed in the research. An aspect based network is drawn based using graph based methods in this research [21]. Opinion fusion and grouping to generate reliable reputation value for a recommendation system is the central idea behind a research. Researchers have developed reputation classification algorithms for the recommendation system based on opinion fusion and grouping of data compiled on feedbacks [22]. For cross domain opinion mining an entropy based classification methodology is proposed in a recent research. Researchers have proposed a semi-supervised machine learning approach for developing an opinion mining classifier in one domain that may be applied to another domain also. Opinion bearing words from one domain are identified and used to find opinion bearing words for another target domain. The key objective of the research is to overcome domain dependency. [23]. A methodology to calculate importance of review on the basis of author's credibility, author's recommendation score, author's representativeness, review helpfulness, timeline of review and content similarity is discussed in a research paper. Practical and theoretical implication of the methodology after implementation in the hotel review domain is presented in the research. Exhaustive review on NLP techniques for opinion mining task has been presented in a the survey paper [24]. Review of research in the subarea of determining association polarity from entity relationship bearing sentences or documents is the focal area of the current survey. An algorithm using a statistical and manual approach is developed that creates a graph between related geographical entities. Manually compiled text in Hindi is used as input to the algorithm [25]. A classification algorithm using a machine learning approach was used to determine the polarity of the author's sentiment of the citation in his/her document. A manually annotated corpus in English is used for the purpose of the development of the algorithm. The corpus is built from the Public Library of Science (PLOC) journal papers [26]. Algorithm for food-disease name identification using a hybrid of lexicon-based, linguistic features based, machine learning-based, and ontology-based approach is developed in research where related biomedical entities are identified along with polarity classifications on the relationship. PUBMED referenced database of medical journals which is an extensive text in English are used as input text in the research [27]. A corpus in English using a linguistic approach is developed as part

¹⁰ Set of affective concepts correlated with affective words. This is an extension of WordNet.

of research to identify polarity in relations between entities from the domain of drugs and diseases. Drug reviews on the social media website are used for the purpose[28]. A classification algorithm that determines a summarization that x% of the document have positive polarity about a food entity's effect on a disease entity and y% of the document have negative polarity about the mentioned entities for Biomedical Text in the Chinese language is developed as part of a research. Dictionary-based named entity recognition, verb-based relationship extraction, semantic-based polarity detection, and intuitive data interpretation approaches are used to develop the algorithm [29]. Annotated corpus and classification algorithm for relationship polarity classification of German text in the domain of food and disease are developed in research [30]. Corpus of manually annotated probes and classification algorithms for new probes for relationship polarity extraction using a machine learning approach for English text from social media is developed in research[31]. Dynamic and static aspect based lexicons give better performance for aspect based polarity classification. Part of speech-based filters, stop word removal and a window of the lexicon in the text based on the distance of an aspect from polarity lexicon are used to improve the performance of the classifier. Pre-existing customer review datasets available in English are used to test the performance of the classifier [32].

2.2 REVIEW OF RESEARCH IN OPINION MINING FROM UNSTRUCTURED INPUT TEXT IN HINDI AND OTHER LOW RESOURCE LANGUAGES

In this subsection an effort is made to highlight work that is done in the field of opinion mining from text in Hindi and other resource deficient languages. Traditional NLP approach has been used to develop a system that can identify subjective sentences and successively identify the polar phrases from the sentences. Multipurpose Question Answering (MPQA) news corpus has been chosen for the same. A translation system has been used to translate English text to Hindi. Dependency parsing has been used for subjectivity analysis and polarity detection [33]. A fallback strategy for sentiment analysis of Hindi documents using three approaches. In first approach corpora is developed in the Hindi movie review domain. A Hindi document is classified using a classifier trained on the developed annotated corpora. In the second approach, the Hindi document is translated into English and then an English text classification algorithm is used to classify the sentiment of the text. In the third approach, the Hindi-SentiWordNet (H-SWN)¹¹ is developed using its English counterpart and majority based approach is used to classify the document based on the majority of the score of each word in the document both with and without the stop words. The research concludes that the sentiment analysis done using first approach gives better results than the

approach that uses machine translation. Word sense disambiguation is left as a future scope of this research [34]. Sentiment subjective lexicons and a product review pre-annotated corpus in Hindi is a contribution of research for Hindi. The corpus is created by translating product review corpus from English to Hindi. A seed list of subjective lexicons is generated and adjectives from the corpus are extracted, then the graph based method is used to extend the seed list of subjective words. This is done by creating a simple edge between a word in the seed list and its synonym from WordNet and a cross edge between the word and its antonym in the WordNet [35]. Prime objective of mining opinion about people from news articles in resource scarce language has been achieved in another research. To fulfill the initial requirements a comparable corpora in English has been translated and transliterated into the resource scarce language. Adjectives, verbs and named entities have been extracted using Part of Speech (POS) tagger in English and named entity recognition tools in English. The same has been translated into Hindi. Opinion and opinion targets have been extracted using statistical techniques [36]. Datasets of Amazon reviews have been translated into Hindi for building corpora to deal with resource scarcity for a sentiment analysis system for Hindi language. Domains chosen for the work are movie review domain and product review domain. Subjective lexicons, n-grams and weighted n-gram classification approach have been used for opinion classification. The entire work has been done for Unicode format UTF 8.0. Work on transliterated formats for Hindi has been left as a future scope [37]. A methodology for fine grained opinion mining in resource scarce language has been proposed and tested on datasets in Hindi and Telugu. From a comparable corpus in English, subjectivity words have been extracted and translated to the resource scarce language. These words have been used to identify subjective sentence in the corpus in Hindi language. It has been established that feature weighing method along with Naïve Bayes Multinomial classifier works well for subjective sentence classification for Hindi language [38]. Two linguistic approaches have been used for sentiment classification. Clustering based on word sense and clustering based on co-occurrence has been used to formulate word clusters for sentiment classification. Experiments have been conducted on four publicly available datasets in Hindi, English, and Marathi. The domains chosen for the datasets are travel and product domains. A knowledgebase of words along with their hyponyms and synonyms have been used to find word sense. The result of experiments conducted in the research showed that best accuracy is achieved by SVM classifier for document level classification i.e. positive or negative [39]. Word and character level n gram features are used for classification of text which are in mixed languages[40]. Dictionary based approach for polarity classification for Hindi text in the domain of movie review [41]. An algorithm to build context specific lexicons for Hindi reviews in the domain of hotel reviews and movie reviews is developed in a research. Researchers have built the lexicon based on 5000 reviews with equal distribution of positive review and negative review. 200 reviews are used

¹¹ Resource available for download on request at http://www.cfilt.iitb.ac.in/resources/senti/HSWN_download_erInfo.php

as the test data set. Results published from this research illustrates that a considerably good accuracy is achieved when algorithm is trained on context specific polarity lexicons that are extended using the synonym set in Hindi WordNet. As per the researchers the results are constrained as the reviews collected are translated from English to Hindi [42]. Word sense disambiguation techniques are applied and analyzed by developing a corpus in Urdu. Character n gram and word n gram models are used for comparative analysis[43]. Word-Level feature, character level feature, and machine learning techniques are applied for building a corpus and sentiment lexicons for low resource language, Roman Urdu. Combination of character based and word based unigrams, bigrams and trigrams are used to enhance the performance of features set used for classifications [44]. Classification into genres of children's stories in the form of text in Hindi and other resource deficient languages is done in a research. Linguistic features like part of speech and keywords with weighing mechanisms are used for classification[45]. A study of opinion mining systems for Indian languages is highlighted in a survey paper [46]. Lexicons in Persian are developed as part of a research for effective classification of Persian text [47]. Resources available for Opinion Mining task in Hindi language are available on websites that have: Tools for Hindi¹², Hindi WordNet¹³, Indian Language Technology Proliferation and Deployment Centre¹⁴ and Hindi Stop Word List¹⁵. Still there exist a lot of opportunities to scale up the tools and resources to their English counterpart.

2.3 SUMMARY OF REVIEWED LITERATURE

Opinion Mining has gained attention of the researchers as thumbs down and thumbs up approach in year 2002. The basic parameters involved in opinion mining across various researches over the years are: initial requirements, opinion classification approach, degree of automation, sentiment classification level, domain of the opinion mining problem, and the language of the input text considered for opinion mining. The text for opinion mining which was traditionally considered as bag of words is now considered as bag of concepts. The initial requirements for opinion mining are the resources or back ground knowledge required to obtain the opinion mining solutions. Literature review suggests the use of lexicons in the same language as the input text of sentiment words for the purpose. Term weighing techniques like term frequency are also used for the purpose. Classification algorithm is an important part of an opinion mining framework. Machine learning approaches have made several advances in the field. Use of labelled training datasets is one of the requirements of these approaches. Dictionary based approaches provide an advantage, that once they are built they do not require any training data.

Hybrid approach i.e. using combination of two or more approaches for forming an opinion mining system is gaining popularity lately. The various classification approaches used by researchers are based upon the degree of automation. Degree of automation means whether intervention of an expert is required as part of the approach or not. An approach requiring no human intervention is an automatic approach; those requiring moderate human intervention are semi-automatic approach. The use of both approaches is in the reviewed literature depending on the requirement of the opinion mining problem. Text at different level of granularity is considered for mining opinion in the literature. The various granularity levels are document level, sentence level, and feature or aspect level. Feature or aspect level granularity has become popular lately. Combination of static and dynamic lexicons helps in the better performance of aspect based opinion mining algorithms. Most of the research in the field is domain specific. The domains in which the opinion mining frameworks have been developed are product review, movie review, politics, food-disease-gene associations, nuclear power, legal, financial, agriculture, journal reviews etc. Most of the novel work in this field has been done for input text in English. The work is mostly language dependent. Linguistic features are taken into account for the purpose of opinion mining. Each language has morphological differences; use of different POS for different purposes is a language dependent concept; word sense also differs from language to language. Hence in the literature most of the work is language dependent. Limited work in this field has been done for web contents in Hindi. The domain specific research in Hindi is accomplished in domains of movie review and product review. From the literature review it is quite evident that the challenges in opinion mining are partially addressed. The challenges are: handling negation in input text, sarcasm in input text, thwarted expectations, noisy input text, domain independence and language independence. Data extraction for opinion mining has been done from social networking sites, weblogs, newspapers, eBook and other web contents. Automatic and semi-automatic procedures have been adopted for data retrieval. Trending currently are the opinion mining solutions for Big Data¹⁶ and application of Deep Learning algorithms.

3. RESEARCH OPPORTUNITIES

This section mainly presents the research opportunities in development of opinion mining system for unstructured input text in Hindi. Systematic reviews on the various data-sources and approaches used across various researches in past fifteen years are presented in a recent research. The various data-sources as reported in the research are; reviews, twitter, news articles, forums, videos, blogs, web

¹² Available at http://sivareddy.in/downloads#hindi_tools

¹³ Available at <http://www.cfilt.iitb.ac.in/wordnet/webhwn/>

¹⁴ Available at <http://tdil-dc.in/index.php?lang=en>

¹⁵ Available at http://www.isical.ac.in/~fire/data/stopwords_list_hin.txt

¹⁶A term coined for huge amount of structured and unstructured data with 3Vs i.e. large volume of data, heterogeneous or variety of data and velocity or high rate of generation of data.

pages, messaging services, speeches medical data, news, key tone patterns/ mouse touches, images, questionnaires and psychological data. The most popular data-source has been the reviews. For Hindi language there are several data-sources that remain unexplored by the current researchers. Most researchers working on Hindi language in this field have used movies and product reviews only. Hence there is a research opportunity to explore data-sources like videos and messaging services on variety of domain in Hindi. From the review of literature conducted in the current survey it is clear that to develop solution for any opinion mining problem there is an initial requirement to develop domain specific resources and back ground knowledge. So development of datasets from data-sources in Hindi is a task for researchers. Supporting context specific knowledge base is an initial requirement in development of an opinion mining application. There is a scarcity of resource available currently to develop opinion mining solutions for several domains in Hindi. For example, there is a lack of availability of resources for developing opinion mining solutions in domains like politics, drug reviews, legal blogs, newspaper quotations, bio-medical text, opinion, finding accuracy of voting approach, opinion on nuclear power, opinion about people, astrology, agriculture, travel and tourism etc. Hence development of domain specific resources in Hindi that can support development of opinion mining systems will be a significant research contribution in this field. Development of domain specific corpus for training and testing of classification algorithms is the first step towards development of an opinion mining system for a novel domain. Rich resources are needed for development of a corpus. E-books in Hindi which can be rich source for text in few of the above mentioned domains may be downloaded from the link <https://sites.google.com/site/hindiebooks/>. Domain specific corpus may be developed from Rich Site Summary (RSS) feeds from various e-newspapers and web pages of news channels in Hindi. These can be accessed from the link <http://www.w3newspapers.com/india/hindi/>. Google समाचार¹⁷ is also a rich source of trending news in Hindi.

Opinion bearing sentences and documents exists on several weblogs in Hindi. There is a need to aggregate the opinions to facilitate decision making of the reader. The need for aggregation is explained with the following example. Contradictory weblogs are available on different sites about effect of some common medicines on an ailment. For example, there is a review on a website¹⁸ regarding the usage of 'Paracetamol', "शरीर में दर्द, बुखार और सर्दी लग जाने पर हम में से कई लोग एंटीबायोटिक का सेवन करते हैं। ऐसे में प्रयोग होने वाली पैरासीटामोल दवा, ऑस्टियोआर्थराइटिस में दर्द को कम करने वाले मेडिकल

इफेक्टिविटी के सैंडर्ड टास्क को पूरा नहीं करती है।"
[Transliteration: 'Shareer mein dard, bukhaar aur sardi lag jaane par hum mein se kain log antibiotic ka sevan kartein hain. Aise mein prayog hone vali paracetamol dawa, osteoarthritis mein dard kam karne wale medical effectivity ke standard ko poora nahin karti hai.' Translation: 'Many of us consume antibiotics when the body feels pain, fever and cold. Paracetamol drug, used in such a way, does not meet the task of medical effectiveness or reduces pain in osteoarthritis.']. As per this review, there is a negative opinion regarding cause and effect relationship between use of 'Paracetamol' and 'Arthritis pain'. Whereas as per a weblog on another website¹⁹, "पैरासिटामोल(Paracetamol) को सिर दर्द(headache), मांसपेशियों में दर्द (muscle aches), गठिया (arthritis), पीठ में दर्द backache, दांत में दर्द toothaches, जुकाम colds, और बुखार fevers में दर्द(pain) से राहत(relief) और आराम पाने के लिए प्रयोग किया जाता है।"

[Transliteration: 'Paracetamol ko sar dard, maspeshiyon mein dard, gathiya, peeth mein dard, daant mein dard, jukaam, aur bukhaar mein dard se rahat aur aaram pane ke liye prayog kiya jaata hai.' Translation: 'Paracetamol is used for relief and rest from headache, muscle aches, arthritis, backache, toothache, and pain in fever', there is a positive association between the related entities 'Paracetamol' and 'Arthritis pain'. An opinion mining system to classify and aggregate opinion from Hindi opinion bearing documents can be developed in the domain of effects of medicines. Another example, web pages dedicated to mythology and religion forms a basis of belief formation amongst the readers. Contradictory beliefs are posted on social networking sites and weblogs in various languages. Just like product reviews and movies reviews are available with ratings; an opinion mining system can be developed to rate different beliefs appearing on websites across various languages. The derivatives would be of social importance and an insight for domain specific researchers in future. An example of one such opinion bearing sentence from a weblog on a website²⁰ is "घर में पूजा करने वाले व्यक्ति का मुंह पश्चिम दिशा की ओर होगा तो बहुत शुभ रहता है। [Transliteration: 'Ghar mein puja karne wale vyakti ka moonh paschim disha ki ore hoga toh bahut shubh rehta hai.' Translation: 'If the person who performs puja in the house will face on the west side then it is very auspicious']. "An opinion mining system can derive association polarity i.e. positive for two related entities 'पूजा करने वाले व्यक्ति का मुंह(person who is worshipping)' and 'पश्चिम दिशा (west direction)'. An example of variation in opinion on these related entities found on another weblog²¹ is "पूजा करते समय भक्त का मुख किस दिशा में हो

17

https://news.google.co.in/?edchaged=1&ned=hi_in&authser=0

¹⁸ <https://khabar.ndtv.com/topic/paracetamol> accessed on 16/4/2017

¹⁹ <http://www.bimbim.in/general/paracetamol/418> accessed on 16/4/2017

²⁰ religion.bhaskar.com accessed on 15/04/2017

²¹ <http://www.astroyantra.com/> accessed on 15/4/2017

यह एक महत्वपूर्ण विषय है वस्तुतः पूजा करते समय व्यक्ति का मुख पूर्व या उत्तर दिशा (East or North Direction) में ही होना चाहिए। [Transliteration: 'Pooja karte samay bhakt ka much kis disha mein ho yeh mahatvapoorna vishe hai vastutah puja karte samay vyakti ka much purva ya uttar disha mein hee hona chahiye' Translation: 'It is important that while praying the face should be in which direction, actually while worshipping person's face should be on East or North direction.' It is clear from this example that whereas one weblog states that it is good to keep face towards west while praying whereas another weblog emphasises that prayers must be performed facing in east or north direction only. Hence there is a variation in opinion on the related entities 'face' and 'direction'. Similarly opinion bearing sentences are there in Hindi language on social networking sites too in the domain of home remedies, hotel reviews, tourist spots reviews etc. There is a need to aggregate the opinions in different domains just as it is done for product and movie reviews. Association of a rating with an opinion will not only help the readers to form statistically validated opinion but will also give inputs to future researchers who are working in several fields. Web links like <http://www.hindijournal.com/archives/>, <http://www.shodh.net/>, <http://road.issn.org/issn/2455-2232-international-journal-of-hindi-research-#.WPORIWmGOpo> have archive of research journals in Hindi language. A corpus may be developed from these archives and an opinion mining system that determines polarity of author's sentiment of the citation in his/her research publication may be developed like a similar work done in English.

Determining association polarity from entity relationship bearing sentences or documents is one aspect of opinion mining that has not been explored exhaustively. As per the current study, most of the work done in this area is domain dependent and language dependent. Scope to extend the existing work for various languages and across various domains exists. Opinion mining can be done by developing corpus from weblogs that has text specifying cause and effect relationships between entities. Through the various opinion mining approaches these cause and effect relationships can be extracted and the polarities can be specified. Weblogs have contents on dream analysis, gem stone analysis, recipes, agriculture, gardening tips, yoga postures, astrological effects, superstitions, public opinion on government decisions, health beliefs etc. Such domains may be explored further for opinion mining for relationship extraction and relationship polarity classification. An ongoing research in an identified subarea of relationship polarity classification from Hindi unstructured text is discussed in detail in the subsequent section [48]. A software development framework designed for the identified research problem is also discussed in detail in the next section.

Researchers in this field have reported several challenges in development of opinion mining system. These challenges are: dependency on domain knowledge, demarcating the text as fact bearing or opinion bearing, hindered polarity. Existing opinion mining solutions can be extended so that optimized solutions may be developed in order to overcome these challenges. Development and enhancement of domain

wise linguistic resources is a contribution that will help the future researchers and application developers. In lexical based approaches of opinion mining it has been noticed that opinion bearing words of one domain do not match the opinion bearing words of another domain. For example consider an expression 'red rose', the word 'red' has positive polarity score in this context whereas in the expression 'red eyes', the word 'red' has negative polarity score. For English language source domains have been identified such that their opinion bearing words have similar polarity scores across various target domains [23]. Similar work to achieve domain independence and overcome the cost of resource development may be done for Hindi language also.

Most of the work for input Hindi text is done for UTF-8 encoding formats. Researchers have left the use of other encoding formats as future scope. Opinion mining for unstructured Hindi contents appearing on mobile applications is a research area that can be explored further. On social networking sites several opinion bearing sentences in Hindi and other Indian languages are part of an image. Such images are available for various domains. Researchers may apply image processing approach to classify the text appearing in the images. Several approaches have not been used or compared for accuracy to mine opinion from Hindi. Examples of such approaches are fuzzy rule based approach, conditional random field approach, ontology based approach. Apart from these approaches there are several intelligent system approach which are reported in a survey [4]. The approaches may be compared for their appropriateness across several domain specific opinion mining applications for Hindi input text. Similarly comparative analysis of performance of classification algorithms developed using various NLP toolkits and their appropriateness for opinion mining applications for Hindi unstructured text may also be done. The various NLP toolkits as listed in a research are Weka using Java, NLTK using Python, CoreNLP using Python, Gensim using Python [24]. Opinion mining solutions for Big Data remains unaddressed for Hindi and several Indian languages. Existing solutions may be extended so that opinion may be derived and suitably aggregated for text appearing on images.

4. OVERVIEW OF AN ONGOING RESEARCH ON POLARITY CLASSIFICATION OF ASSOCIATION

In this section an overview is given on the methodologies formulated for a research problem identified in the field of opinion mining. The research is being carried by the researcher's of this study. It is dedicated towards the classification of polarity of associations. The details of the research may be referred to in our previous research contribution [48]. The research is inspired by the following. Lot of beliefs related to food benefits for health issues persists amongst the population of the world. Knowledge regarding benefits derived from consumable parts of plant extracts, herbs has its origin in several part of the world. A huge population in the world speaks, reads and writes in Hindi

language. There are several websites related to health benefits of herbs, home remedies that are available in Hindi. Many newspapers also contain articles related to health values of food, herbs etc. This research is to facilitate mining of opinion from domain specific text in Hindi appearing on the Web. The text is restricted to the domain of home remedies and impact of food on health of a human being. Whenever the term 'food' is referred, it means a single food item or a food component or food combination or any consumable item for human beings. The term 'health issue' refers to any health problem, disease or an ailment associated with human beings. Table 1 illustrates example of the input text and the output that is desired by the opinion mining system being developed in the ongoing research. Resource scarcity, linguistic flexibility and diversity of Hindi language make the opinion mining task challenging. The outcomes of this research may be put to practical use to public health services and social reformists working at national and international levels. They may contribute towards strategic plan of actions to address wrong beliefs and hence improve health conditions of human beings. Contradictory beliefs may be put to scientific validations by the biomedical researchers. Scientifically validated novel or unique beliefs may be used for drug discovery. In the area of strategic planning for business development there is an interesting application. A supermarket or an e-shopping site may display related food items next to one another on the basis of certain health beliefs to facilitate sales. Since the domain specific corpus and supporting lexical knowledge base did not exist in Hindi before the research, hence this research is dedicated towards developing the same. Annotated corpus and the lexical knowledge base that is being developed in the current research will facilitate future researchers in the field of NLP, Information Retrieval and Processing, and Artificial Intelligence. Not only the researchers willing to do research for Hindi language will benefit from the work but researchers doing research for languages grammatically similar to Hindi will also benefit. Specifically researchers who develop research solutions for text in Punjabi may use the derivatives with linguistic modifications. As evident from the discussion, this research is significant for its application in the area of health services, drug discovery, biomedical research, social reforms, business intelligence and future research in computer science. Some important derivatives of the research may be obtained from our previous study[48]. Majorly, there are four tasks to the development of an opinion mining system in this research. The tasks may be listed as: development of domain specific corpus; development of domain specific lexical knowledge base; development of the polarity classifier; and development of the opinion mining framework. The procedures that are being adopted to accomplish the various tasks are illustrated in figures given in this paper. Procedure being used for the development of the domain specific corpus is illustrated in Figure I. Procedure adopted for the development of the lexical knowledge base is illustrated in Figure II.

Algorithm developed using a lexicon based approach is given in our previous study[48]. Figure III illustrates the complete procedure adopted for the development of the opinion mining framework.

Table 1: Input and Outputs of the proposed system

Examples of Input and Expected Outputs
<p>Example 1 Input Text: आंवला और अंगूर को पीस कर खाने से बुखार की प्यास और बेचैनी ठीक हो जाती है। ;Transliteration: 'Anvla aur angoor pees kar khane se bukhar ki pyaas aur bechaine theek ho jaati hai.' ;Translation: 'Eating grinded gooseberries and grapes will help to reduce the fever and restlessness of the fever.' Food Item : 'आवला' और 'अंगूर';Transliteration: 'Anvla' and 'Angoor';Translation: 'Goose Berry' and 'Grapes' Health Issue: 'बुखार की प्यास' और 'बुखार की बेचैनी';Transliteration: 'Bukhar ki pyaas' and 'bukhaar ki baichaini'; Translation: 'fever' and 'restlessness of the fever' Association polarity: Positive 'ठीक';Transliteration: 'Theek'; Translation: 'Correct(semantically means cure)' Polarity Strength: Medium</p>
<p>Example 2 Input Text: हिचकी आना- दस ग्राम राई को पाओ भर पानी में उबाल कर छान कर पीले, आराम मिलेगा।; Transliteration: 'Hichki aana- Das gram rai ko pao bhar paani mein ubaal kar chaan kar peelein, aaram milega.'; Translation: Hiccups - Take ten grams of mustard seeds and boil them in water and drink, it will give relief. Food Item : 'राई' और 'पानी' ;Transliteration: 'Rai' and 'paani'; Translation: 'Mustard seeds' and 'water' Health Issue: 'हिचकी';Transliteration: 'Hichki'; Translation: 'Hiccups' Association polarity: Positive 'आराम'; Transliteration: 'Aaram'; Translation: 'relief' Polarity Strength: Medium</p>

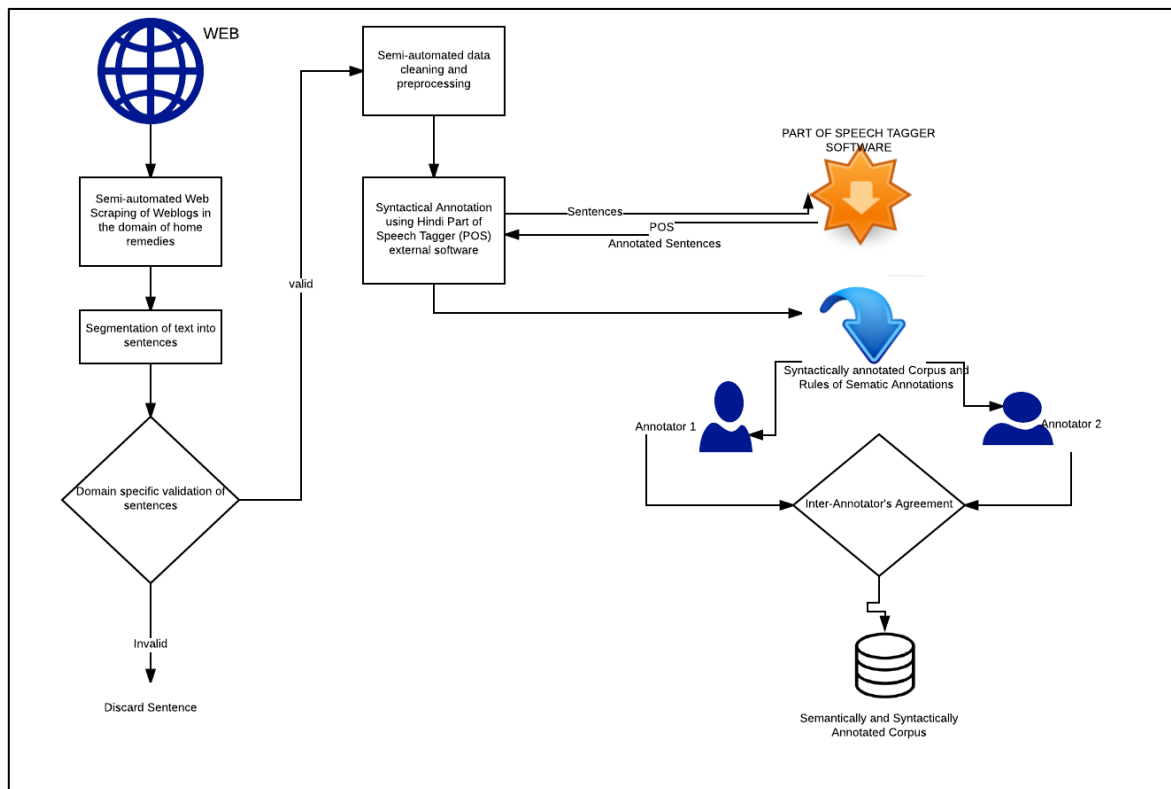


Figure I. Block diagram for development of domain specific corpus with semantic and syntactic annotations.

The scope of this research is as follows. The web contents appearing on the high ranked sites by the Google search engine²² with contents in Hindi language are being considered in the identified ongoing research. The domain of the chosen text is home remedies, impact of food on health of human beings, herbs and their effect on health of human beings. Input texts considered are in the form of a single sentence. Only the web contents related to food and its impact on health of the human beings are being considered. Those unstructured text is considered for which copy/paste operation is allowed. The text that is part of an image is not being considered in the research. The opinion mining system that will be an outcome of this research will be engaged in only extracting and classifying the association polarity of related entities, not validating them.

5. CONCLUSION

This paper presented a systematic report based on survey conducted on some significant research publications in the field of opinion mining traced from year 2002 to year 2020. The study included parameters like: language of the input unstructured text, underlying domain of the text, classification approach used, challenges encountered, major outcomes, and resources developed, if any. Research opportunities derived from this survey are mostly focused to identifying required opinion mining systems from Hindi unstructured input text and in the subarea of extraction of

opinion regarding relationship of associated entities. India is culturally rich country and a considerably huge growth of Hindi text production and consumption is being witnessed on the WebPages and mobile application pages lately. Most of the supportive resources for the development of opinion mining system are available in English. These resources are in the form of corpus, knowledge bases and tools for sentiment analysis/ opinion mining. Stanza²³, a Python package for NLP, that has tools for several languages, does not provide NER tool for Hindi. NER is an important component of an OM system. The resources that are available for English cannot be used directly for building opinion mining systems for Hindi and other Indian languages. Hence there is a requirement to develop resources across various domains in Hindi and other Indian languages. This would facilitate researchers in this field to develop opinion mining solutions for Hindi text at par with other languages. Development of opinion mining solutions for Hindi text is challenging as it is morphologically rich language. Research scope in the field exists to handle sarcastic text in Hindi. Deriving opinion from documents that have text with thwarted expectations is also a challenging task that awaits better accomplishment. Rather it is left as a future scope in several research works.

²² www.google.com/blogsearch

²³ <https://stanfordnlp.github.io/stanza/>

©2012-20 International Journal of Information Technology and Electrical Engineering

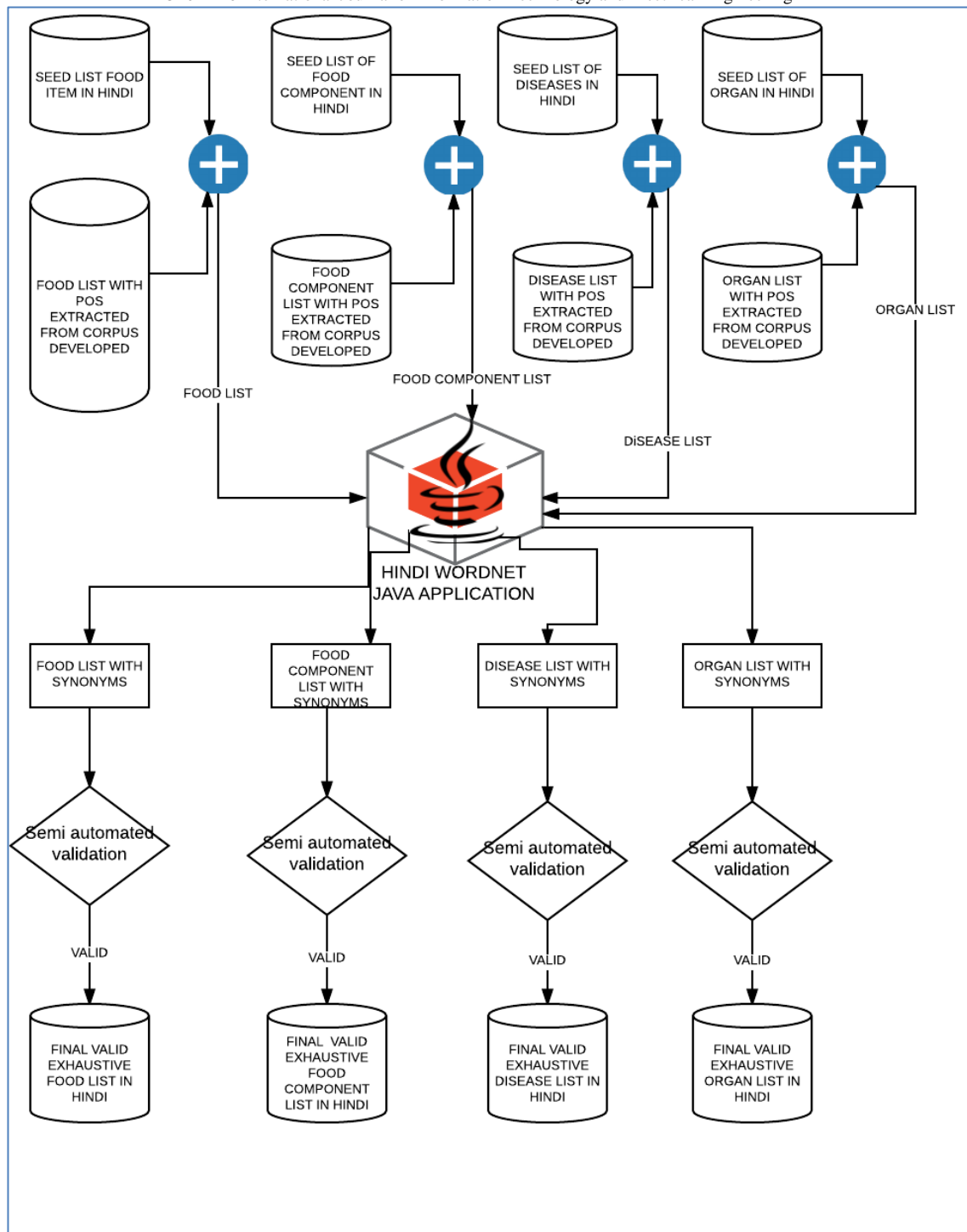


Figure II. Block diagram for development of syntactic and semantic database for named entity recognition

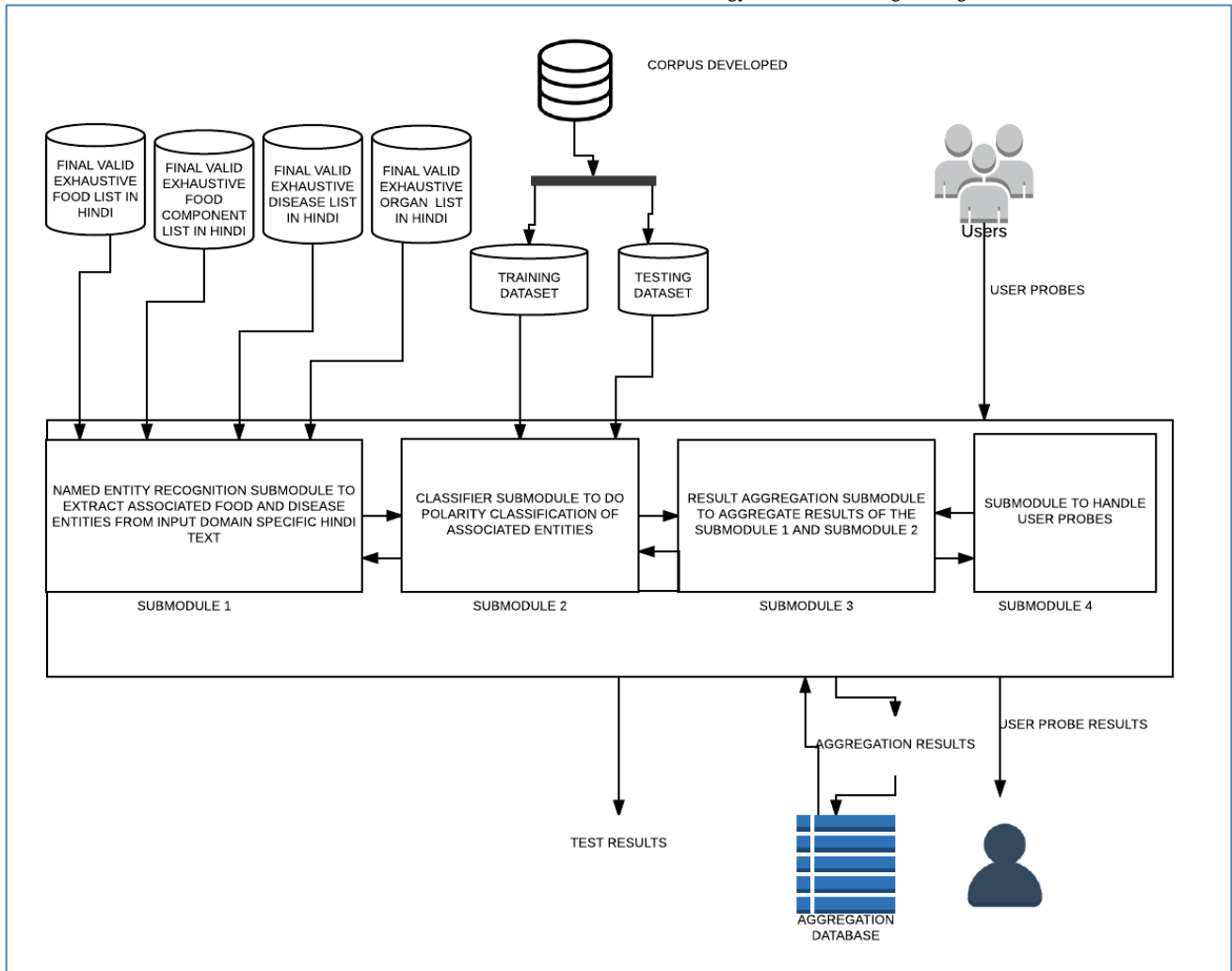


Figure III. Opinion mining system for the identified research problem

An overview of the ongoing research in the area of relationship polarity extraction and classification from Hindi unstructured text in the domain of home remedies is presented in this paper. Significance and scope of this research is also included in this paper. A software framework designed for the identified research is also illustrated. The limitation of the proposed framework is domain dependency. The expected solution is domain and language dependent. However, the solution with little domain and language based modifications may be extended to other domains and languages. It is expected that the outcomes of this research may be significant for future researchers. Two major challenges identified in the development of the opinion mining system for the identified problem are; (i) lack of supportive resources available in Hindi (ii) the input text is highly skewed. Input text is highly skewed means that the ratio between domain centric positive opinion bearing sentences and negative opinion bearing sentences is skewed towards positive sentiment. Availability of skewed supportive data sets a limit on the approach that can be used for polarity classification. Work is in progress to improve solution for this limitation.

REFERENCES

- [1] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," *Proc. Conf. Empir. Methods Nat. Lang. Process.*, pp. 79–86, 2002.
- [2] M. Gamon, A. Aue, S. Corston-Oliver, and E. Ringger, "Pulse: Mining Customer Opinions from Free Text," in *International Symposium on Intelligent data analysis*, 2005, pp. 121–132.
- [3] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Found. Trends® InformatioPang, B., Lee, L. (2006). Opin. Min. Sentim. Anal. Found. Trends® Inf. Retrieval, 1(2), 91–231. doi10.1561/1500000001n Retr.*, vol. 1, no. 2, pp. 91–231, 2006.
- [4] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowledge-Based Syst.*, vol. 89, pp. 14–46, 2015.
- [5] L. Dey and S. M. Haque, "Opinion mining from noisy text data," *Int. J. Doc. Anal. Recogniti[1] L.*

©2012-20 International Journal of Information Technology and Electrical Engineering

- Dey S. M. Haque, "Opinion Min. from noisy text data," *Int. J. Doc. Anal. Recognit.*, vol. 12, no. 3, pp. 205–226, 2009.
- [6] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining SentiWordNet," *Analysis*, vol. 0, pp. 1–12, 2010.
- [7] S. and G. Poria, A. and Hussain, A. and Howard, N. and Das, D. and Bandyopadhyay, and Sivaji, "Enhanced SenticNet with Affective Labels for Concept-Based Opinion Mining," vol. 28, no. April, pp. 31–38, 2013.
- [8] K. S. Doddi, "Sentiment Classification of News Article," College of Engineering, Pune, 2013.
- [9] P. Yin, H. Wang, and K. Guo, "Feature–opinion pair identification of product reviews in Chinese: a domain ontology modeling method," *New Rev. Hypermedia Multimed.*, vol. 19, no. 1, pp. 3–24, 2013.
- [10] S. Huang, Z. Niu, and C. Shi, "Automatic construction of domain-specific sentiment lexicon based on constrained label propagation," *Knowledge-Based Syst.*, vol. 56, pp. 191–200, 2014.
- [11] G. Yan, W. He, J. Shen, and C. Tang, "A bilingual approach for conducting Chinese and English social media sentiment analysis," *Comput. Networks*, vol. 75, no. PB, pp. 491–503, 2014.
- [12] I. Peñalver-Martinez *et al.*, "Feature-based opinion mining through ontologies," *Expert Syst. Appl.*, vol. 41, no. 13, pp. 5995–6008, 2014.
- [13] Y. Rao, Q. Li, X. Mao, and L. Wenyin, "Sentiment topic models for social emotion mining," *Inf. Sci. (Ny)*, vol. 266, pp. 90–100, 2014.
- [14] Z. Wang, L. Tu, Z. Guo, L. T. Yang, and B. Huang, "Analysis of user behaviors by mining large network data sets," *Futur. Gener. Comput. Syst.*, vol. 37, pp. 429–437, 2014.
- [15] S. Rill, D. Reinel, J. Scheidt, and R. V. Zicari, "PoliTwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis," *Knowledge-Based Syst.*, vol. 69, no. 1, pp. 24–33, 2014.
- [16] F. Bravo-Marquez, M. Mendoza, and B. Poblete, "Meta-level sentiment models for big social data analysis," *Knowledge-Based Syst.*, vol. 69, no. 1, pp. 86–99, 2014.
- [17] S. S. Sohail, J. Siddiqui, and R. Ali, "Feature Extraction and Analysis of Online Reviews for the Recommendation of Books using Opinion Mining Technique," *Perspect. Sci.*, vol. 8, pp. 4–6, 2016.
- [18] N. Genc-Nayebi and A. Abran, "A Systematic Literature Review: Opinion Mining Studies from Mobile App Store User Reviews," *J. Syst. Softw.*, vol. 125, pp. 207–219, 2016.
- [19] R. Piryani, D. Madhavi, and V. K. Singh, "Analytical mapping of opinion mining and sentiment analysis research during 2000–2015," *Inf. Process. Manag.*, vol. 53, no. 1, pp. 122–150, 2017.
- [20] W. Tao and T. Liu, "Building ontology for different emotional contexts and multilingual environment in opinion mining," *Intell. Autom. Soft Comput.*, pp. 1–7, 2017.
- [21] E. Bilici and Y. Saygin, "Why do people (not) like me?: Mining opinion influencing factors from reviews," *Expert Syst. Appl.*, vol. 68, pp. 185–195, 2017.
- [22] Z. Yan, X. Jing, and W. Pedrycz, "Fusing and mining opinions for reputation generation," *Inf. Fusion*, vol. 36, pp. 172–184, 2017.
- [23] J. S. Deshmukh and A. K. Tripathy, "Entropy based classifier for cross-domain opinion mining," *Appl. Comput. Informatics*, 2017.
- [24] S. Sun, C. Luo, and J. Chen, "A review of natural language processing techniques for opinion mining systems," *Inf. Fusion*, vol. 36, pp. 10–25, 2017.
- [25] K. Dutta, N. Prakash, and S. Kaushik, "Hybrid framework for information extraction for geographical terms in hindi language texts," in *IEEE International Conference on Natural Language Processing and Knowledge Engineering, IEEE NLP-KE'05*, 2005, pp. 577–581.
- [26] S. S. Piao, S. Ananiadou, Y. Tsuruoka, Y. Sasaki, and J. Mcnaught, "Mining Opinion Polarity Relations of Citations," *Int. Work. Comput. Semant.*, pp. 366–371, 2007.
- [27] H. Yang, R. Swaminathan, A. Sharma, V. Ketkar, and J. D'Silva, "Mining biomedical text towards building a quantitative food-disease-gene network," *Stud. Comput. Intell.*, vol. 375, pp. 205–225, 2011.
- [28] L. Goeuriot *et al.*, "Textual and Informational Characteristics of Health-Related Social Media Content: A Study of Drug Review Forums," in *Asia Pacific Conference Library & Information Education & Practice, 2011*, 2011, pp. 548–557.
- [29] Q. Miao, S. Zhang, B. Zhang, and H. Yu, "Extracting and Visualizing Semantic Relationships from Chinese Biomedical Text," *Proc. 26th Pacific Asia Conf. Lang. Information, Comput.*, pp. 99–107, 2012.
- [30] F. Gotti, P. Langlais, and A. Farzindar, "Proceedings of the Workshop on Language Analysis in Social Media," in *Proceedings of the Workshop on Language Analysis in Social Media, 2013*, no. June, pp. 80–89.

©2012-20 International Journal of Information Technology and Electrical Engineering

- [31] S. Bhattacharya, "Computational methods for mining health communications in web 2.0," University of Iowa, 2014.
- [32] M. E. Mowlaei, M. S. Abadeh, and H. Keshavarz, "Aspect-Based Sentiment Analysis using Adaptive Aspect-Based Lexicons," *Expert Syst. Appl.*, p. 113234, 2020.
- [33] D. Bhattacharyya, S. Biswas, and T. hoon Kim, "A review on natural language processing in opinion mining," *Int. J. Smart Home*, vol. 4, no. 2, pp. 31–38, 2010.
- [34] A. Joshi, B. A. R., and P. Bhattacharyya, "A Fall-back Strategy for Sentiment Analysis in Hindi: a Case Study," no. October 2015, 2010.
- [35] P. Arora, A. Bakliwal, and V. Varma, "Hindi Subjective Lexicon Generation using WordNet Graph Traversal," *Int. J. Comput. Linguist. Appl.*, vol. 3, no. Jan-Jun 2012, pp. 25–29, 2012.
- [36] A. Mogadala and V. Varma, "Retrieval approach to extract opinions about people from resource scarce language news articles," *Proc. First Int. Work. Issues Sentim. Discov. Opin. Min. - WISDOM '12*, no. August, pp. 1–8, 2012.
- [37] P. Arora, "Sentiment Analysis For Hindi Language," International Institute of Information Technology, 2013.
- [38] M. Aditya, "Fine-Grained Opinion Mining in Resource-Scarce Languages with Sentence-level Subjectivity Analysis," Search and Information Lab, International Institute of Information Technology, Hyderabad, India, 2013.
- [39] K. Popat, A. B. R., P. Bhattacharyya, and G. Haffari, "The Haves and the Have-Nots: Leveraging Unlabelled Corpora for Sentiment Analysis," *Proc. 51st Annu. Meet. Assoc. Comput. Linguist.*, no. October 2015, pp. 412–422, 2013.
- [40] B. G. Patra, D. Das, and A. Das, "Sentiment Analysis of Code-Mixed Indian Languages," *ICON -2017: 14th International Conference on Natural Language Processing*, 2017. [Online]. Available: <http://www.aclweb.org/anthology/W/W17/W17-75>.
- [41] R. Sharma, S. Nigam, and R. Jain, "Polarity Detection of Movie Reviews in Hindi Language," *Int. J. Comput. Sci. Appl.*, vol. 4, no. 4, pp. 49–57, 2014.
- [42] D. Mishra, M. Venugopalan, and D. Gupta, "Context Specific Lexicon for Hindi Reviews," *Procedia Comput. Sci.*, vol. 93, no. September, pp. 554–563, 2016.
- [43] A. L. I. Saeed, R. A. O. Muhammad, and A. Nawab, "A Sense Annotated Corpus for All-Words Urdu Word," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 18, no. 4, 2019.
- [44] K. Mehmood, D. Essam, and K. Shafí, "Sentiment Analysis for a Resource Poor Language — Roman Urdu," *ACM Trans. Asian Low Resour. Lang. Inf. Process.*, vol. 19, no. 1, pp. 1–15, 2019.
- [45] D. M. Harikrishna and K. S. Rao, "Children's Story Classification in Indian Languages Using Linguistic and Keyword-based Features," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 19, no. 2, 2019.
- [46] D. T. Miranda and M. Mascarenhas, "A Study of Opinion Mining in Indian Languages," *Adv. Intell. Syst. Comput.*, pp. 71–77, 2018.
- [47] R. Dehkharghani, "SentiFars: A Persian Polarity Lexicon for Sentiment Analysis," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 19, no. 2, 2019.
- [48] A. Prasad and N. Sharma, "Lexicon Based Extraction And Opinion Classification Of Associations In Text from Hindi Weblogs," *Int. J. Adv. Res. Eng. Technol.*, vol. 11, no. 8, pp. 439–451, 2020.

AUTHOR PROFILES

Arpana Prasad is a Master in Computer Applications, and is pursuing Ph.D. from Department of Computer Science, Punjabi University, Patiala, India. Her field of study is Opinion Mining. She is currently engaged in developing lexical resources and exploring computing techniques appropriate for development of a domain centric Opinion Mining System for Hindi text available on Weblogs. She has an experience of more than 16 years in teaching undergraduate and postgraduate courses in Information Technology and Computer Applications. She also has a work experience of over 6 years in the Learning and Development Industry. She has been engaged in doing a few instructional writing projects in Computer Science as a freelancer.

Neeraj Sharma received the degree in Computer Science from Punjabi University, Patiala, Punjab, India in 1995. He obtained his Ph.D. in Computer Science from Punjabi University, Patiala, Punjab, India in 2011. Currently, he is a Professor at Department of Computer Science, Punjabi University, Patiala, Punjab, India. He has been a guide to several M.Tech. and Ph.D. students. Under his guidance, one of his Ph.D. students is developing an Opinion Mining System for unstructured Hindi text from Weblogs.