ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 10(24), DOI: 10.17485/ijst/2017/v10i24/96498, June 2017*

Sentiment Analysis of English Literature using Rasa-Oriented Semantic Ontology

D. Sreejith[1], M. P. Devika[1], Naga Santosh Tadikamalla[2] and Sanju Varghese Mathew[3]

[1]Department of English, SRM University, Kattankulathur – 603203, Tamil Nadu, India;
sreejith.d@ktr.srmuniv.ac.in, devika.m@ktr.srmuniv.ac.in
[2]Amtex Software Solutions Pvt. Ltd., Chennai – 603103, Tamil Nadu, India;
nagasantosh.t@live.com
[3]Greenway Health, Florida, USA;
sanjumathew4293@gmail.com

Abstract

Objectives: Sentiment Analysis analyses people's opinion, sentiments, attitudes and emotions towards entities such as products, services, literature and their attributes. Since literature has got an exponential growth in digital format recently, it will help the readers to choose the genre according to their interest as well. **Methods/Statistical Analysis:** Finding and monitoring such opinions present in the Internet and filtering the required information is a formidable task for an average reader because of the huge amount of data available online. In such difficult situations, Sentimental Analysis can play a big role in helping the user. In this study, the sentiment analysis of a literary work is done using ontology of 'Navarasa'. **Findings**: This study does, for the first time, the sentiment analysis of a short story using the 'Navarasa' ontology created by the researcher. The sentiment polarity of the work could be derived with a better accuracy using the emotion lexicons generated. **Application/Improvements:** Thus this paper provides a novel method of sentiment analysis of English literature and throws light on new avenues for future research work in this domain.

Keywords: Emotion Lexicon, Literature, Navarasa, Ontology, Polarity, Sentiment Analysis

1. Introduction

The term "literature" generally refers to any written or spoken material in any language. But in this paper the term refers to the works of the creative imagination on the likes of poetry, short story, drama, novel etc. Literature is the most effective mode of expression to represent this world. It encompasses every sphere of human life like culture, tradition, history, psychology etc. To depict the human life in all its richness, it used diction expressing various emotions and feelings. These varied emotional expressions are called 'Rasa' (flavor) in Sanskrit Language.

In this paper, we deal with the Sentiment Analysis of English Literature by using Rasa-Oriented semantic ontology which comes under the broad area of Natural Language Processing (NLP). Natural Language Processing is the ability of a computer to understand human language. It is a component of Artificial Intelligence. NLP consists of several researched tasks like Machine Translation, Natural Language Generation, Morphological Segmentation, Part-of-Speech Tagging, Parsing, Sentiment Analysis etc[1]. Sentiment Analysis is, at present, widely applied in the areas of product and movie reviews[2], whereas for this paper, we have tried to use the

*Author for correspondence

Navarasa Ontology for the sentiment analysis of a short story from English Literature. In this ontology, OWL provides a machine readable ontology which can then be processed by Semantic Web applications[3].

The term ontology refers to a set of distinct objects resulting from an analysis of a domain. "The main aim of ontology is to provide knowledge about specific domains that can be understood by both, the computer and developers. It also helps to interpret a text review at a finger granularity with shared meanings and provides a sound semantic ground of machine understandable description of digital context" (Haider, 10).

Ontology consists of classes, the subsumption relations between them, object properties that relate to instances of classes, and restrictions on what properties may hold for these instances. In a well-designed ontology, one can make inferences about classes and the types of individuals. We used SPARQL to query our Rasa ontology to produce triples, i.e., subject, object and predicate, present in this ontology. SPARQL is an RDF query language which is used to retrieve and manipulate data stored in RDF files. For instance, 'bewilderment has expression horror' is a 'triples', retrieved from this ontology by using SPARQL Query. Here, 'bewilderment' is the subject and 'has expression horror', the predicate.

2. Problem Statement

Sentiment Analysis or Opinion Mining refers to the application of Natural Language Processing to identify and extract subjective information from source materials. Its application can be seen in the area of product and movie reviews in social media.

2.1 Previous Research

An analysis of nineteenth century novelist Jane Austen's writing was compiled by KatrinOlmann as part of a university project which presented selected results from her study on irony in Austen's writing complemented by findings from some additional queries based on Jane Austen Corpus (JAC) and a large corpus of 18th and 19th century English novels. According to Olmann, keywords in Jane Austen's novels could be used as pre-reading activity or as a preliminary step in the interpretation of any literary work.

3. Methodology

3.1 Navarasa

'Navarasa' means nine emotions in which 'nava' signify nine and 'rasa' signifies emotions. The nine emotions included in 'Navarasas' are Adbhuta Rasa (Wonder), Beebhatsa Rasa (Disgust), Bhayanaka Rasa (Terror), Hasya Rasa (Comedy), Karuna Rasa (Pathos), Roudra Rasa (Fury), Shanta Rasa (Quietism), Sringara Rasa (Eroticism) and Veera Rasa (Heroism). These are the emotions that humans show according to the situation. All emotions and feelings are said to originate from these navarasas and by applying these nine flavors, we discovered that it was possible to analyze the sentiment nature of a literary work. For the purpose of this study, we utilized the concept of the navarasas as provided in the Sanskrit text "SahithyaDarpana" (The Mirror of Composition) which is considered to be one of the most authentic texts on Indian Aesthetics[4]. Each 'rasa' has a permanent mood and several accessories (fluctuating moods). The term 'expression' in our ontology refers to the various English words used in literature to express the above mentioned 'rasa' as shown in the Table 1.

In this research, we have tried to analyze the sentiment nature of an English short story by using Rasa-Oriented Semantic Ontology. As the first requirement, we created the ontology of Navarasa. We drew the elements of this ontology from the Rasa concept in Indian Aesthetics. This ontology asserted the classes and subclasses as well as the disjointedness of classes and its individuals. The scope of the paper was restricted to the identification of the main ontological elements of cases including classes, object properties, class restrictions, and individuals; the whole ontology was more easily examined using the OWL file and an ontology editor such as Protege. We aimed at extracting a rich emotional semantics of tagged resources through an ontology driven approach. This was done by exploiting and combining available computational

Table 1. Rasa Table

Rasa	Permanent Mood	Accessories	Expression
Sringara Rasa	Desire	Desolation, Frenzy	Adore, Dear
Hasya Rasa	Joy	Break Up, Snicker	Bliss, Cheer
Karuna Rasa	Sorrow	Anxiety, Distress	Gloom, Grief
Roudra Rasa	Anger	Impatience, Sternness	Rage, Rave
Veera Rasa	Energy	Pride, Reasoning	Dignity, Fairness
Bhayanaka Rasa	Fear	Terror, Death	Horror, Panic
Beebhatsa Rasa	Aversion	Bafflement	Dismay, Hatred
Adbhutha Rasa	Surprise	Confusion, Flurry	Miracle, Sensation
Shantha Rasa	Tranquility		Felicity, Reverie

and sentiment lexicons with ontology of emotional categories. We checked to see if the tags of a given resource are "emotion-denoting" words directly referring to some emotional categories of the ontology[5].The novelty of this paper is the creation of a fuller, more explicit ontology of 'navarasas' which follow current methodologies of ontology design and presentation of the ontology in rich, graphic and linguistic modes. And in the second part of the paper, we present how this ontology was used for the sentimental analysis of a short story by using programs like Java's Apache Jena, SPARQL, Ruby and Python Language.

3.2 Ontology Generation

As the first part of the research, we created an ontology of Navarasas (Nine Flavours) by using Protege 4.3 ontology editor and knowledge base framework developed by Stanford University. Protege is a free, open source ontology editor and framework for building intelligent systems. We provided graphical and textual representations of the case, which served different presentational purposes. In this navarasa ontology, rasa and expressions are the two main classes and rasa is further divided into nine subclasses namely Adbhuta Rasa (Wonder),

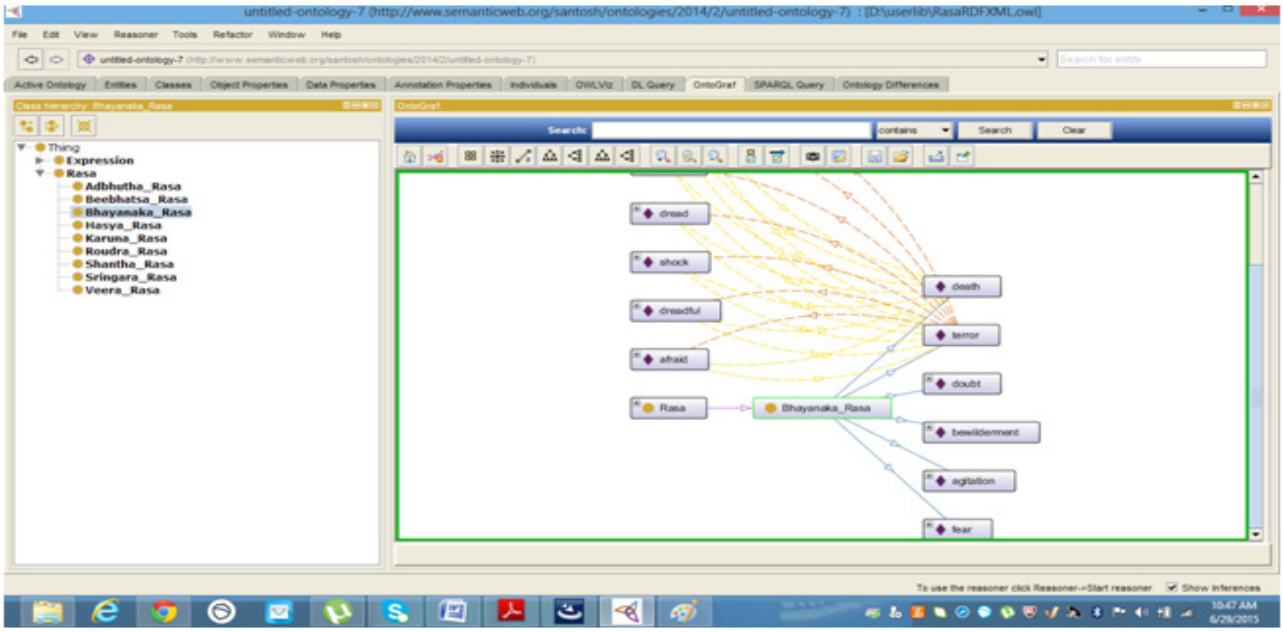


Figure 1. Depiction of Rasa-Oriented Semantic Ontology.

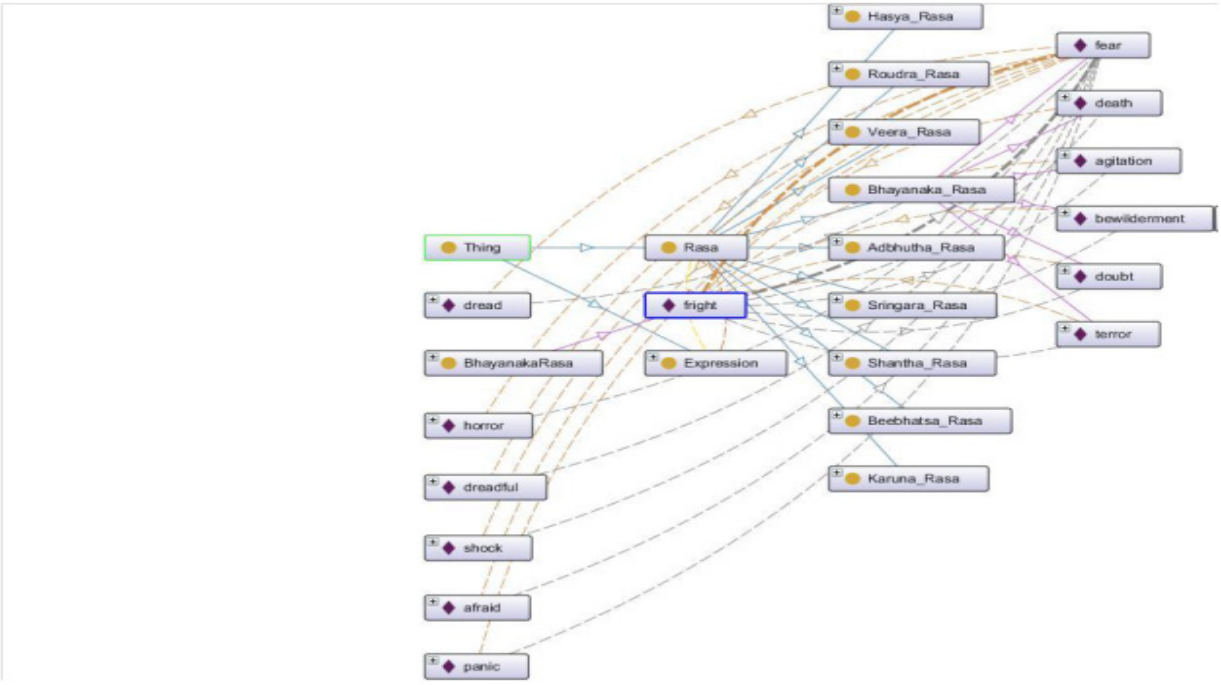


Figure 2. Inter-relations of words in the Ontology.

Beebhatsa Rasa (Disgust), Bhayanaka Rasa (Terror), Hasya Rasa (Comedy) etc. In the category of 'expressions', we employed under each 'rasa' the English words generally used in literature to express the corresponding emotion which is given as a member of that respective expression. We then specified the classes with respect to the properties that define them. For example, one of the subsidiary moods of Sringara (Eroticism) is desolation. This property defines the class. The ontology is shown in Figure 1.

In the 'rasa' class, each 'rasa' has one permanent mood and several accessory moods which are given as members of that particular class. In a well-designed ontology, one can make inferences about classes and the types of individuals (i.e., the classes to which they belong); for example if 'horror' is the expression of 'Bhayanaka Rasa', then it is inferred that it is the expression for the permanent mood and all the accessory moods of Bhayanaka Rasa. Idioms and phrases have not been included within the scope of this ontology. They can however, be added later on, in future ontological developments. The graphical representation is given in Figure 2.

4. OWL to RDF/XML Conversion

After creating the Ontology, we converted the OWL format into RDF/XML using Java's Apache Jena API. The purpose of converting into RDF/XML is to extract triples from the ontology since it doesn't work in OWL as shown in Figure 3.

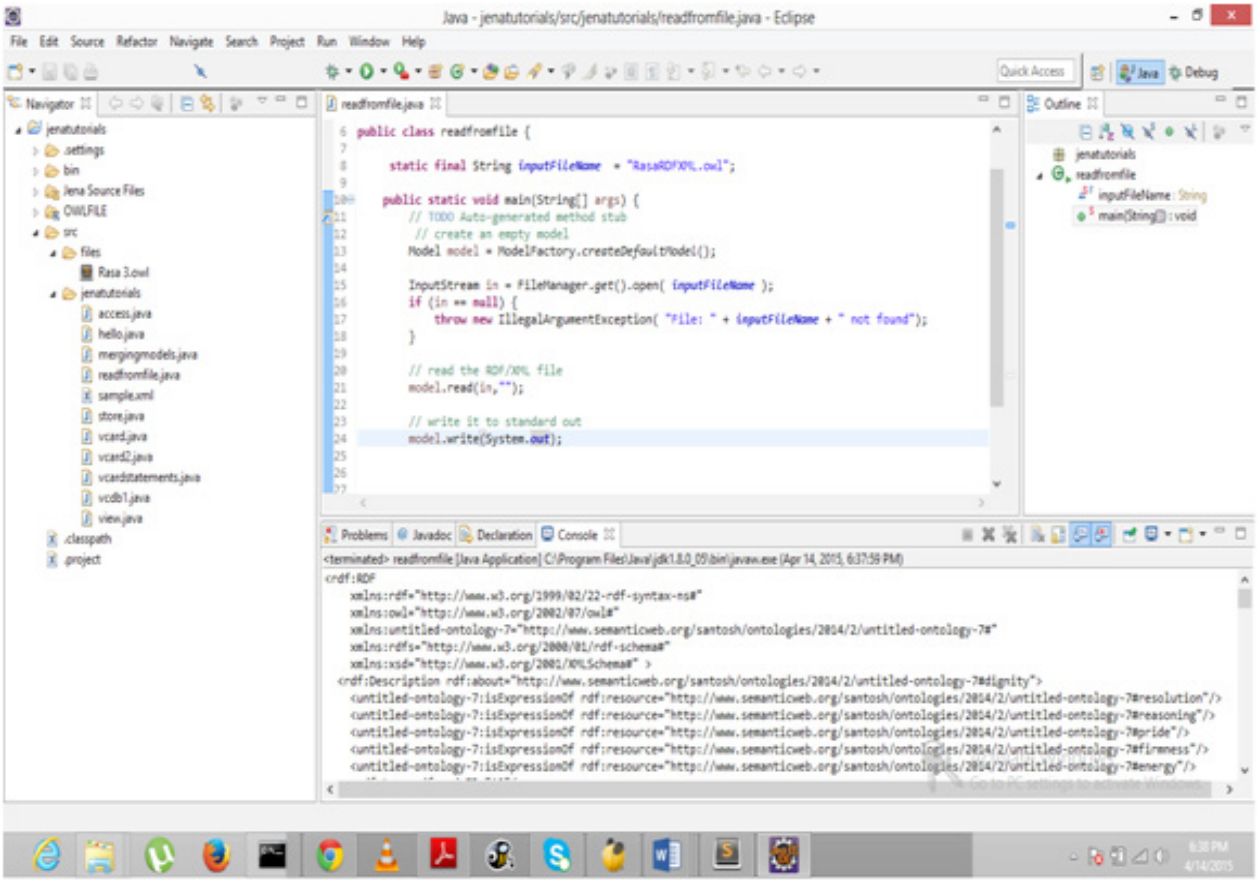


Figure 3. Conversion of OWL file to RDF/XML format.

5. SPARQL Querying

We used SPARQL to Query our Rasa ontology to produce triples, i.e., Subject, Object and Predicate, present in this ontology. For instance, 'bewilderment has expression horror' is a triples retrieved from this ontology by using SPARQL Query. Here, 'bewilderment' is the subject and 'has expression horror' is predicted in Figure 4.

6. Emotion Lexicon Generation

A list of words that express each emotion is termed as an emotion lexicon. After creating triples through SPARQL Query, these triples are filtered to generate emotion lexicon. It involves removing unnecessary blank nodes, redundancies etc. By analyzing the emotion lexicons, we tried to assess the hierarchical relationship derived from the various 'Rasa' emotions. We also tried to identify the sentimental nature of the text by assessing the polarity (positive, negative or neutral), type and intensity of the emotion lexicons present[6].

7. Synset Generation

Synset, according to Wordnet, is a set of one or more synonyms that are interchangeable in some contexts without changing the true value of the proposition in which they are embedded[7]. To generate a Synset, we used Python's

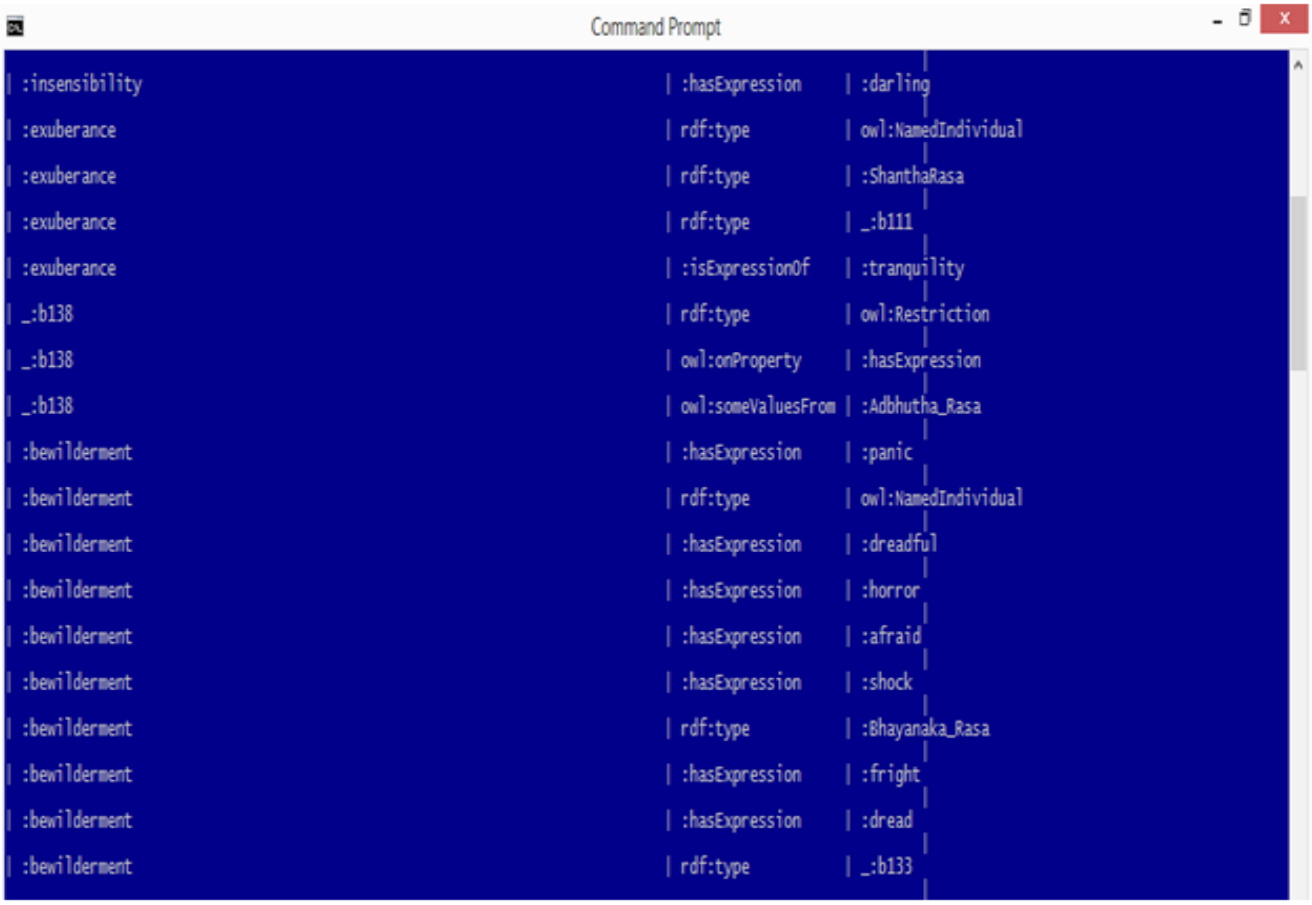


Figure 4. Execution of a SPARQL Query.

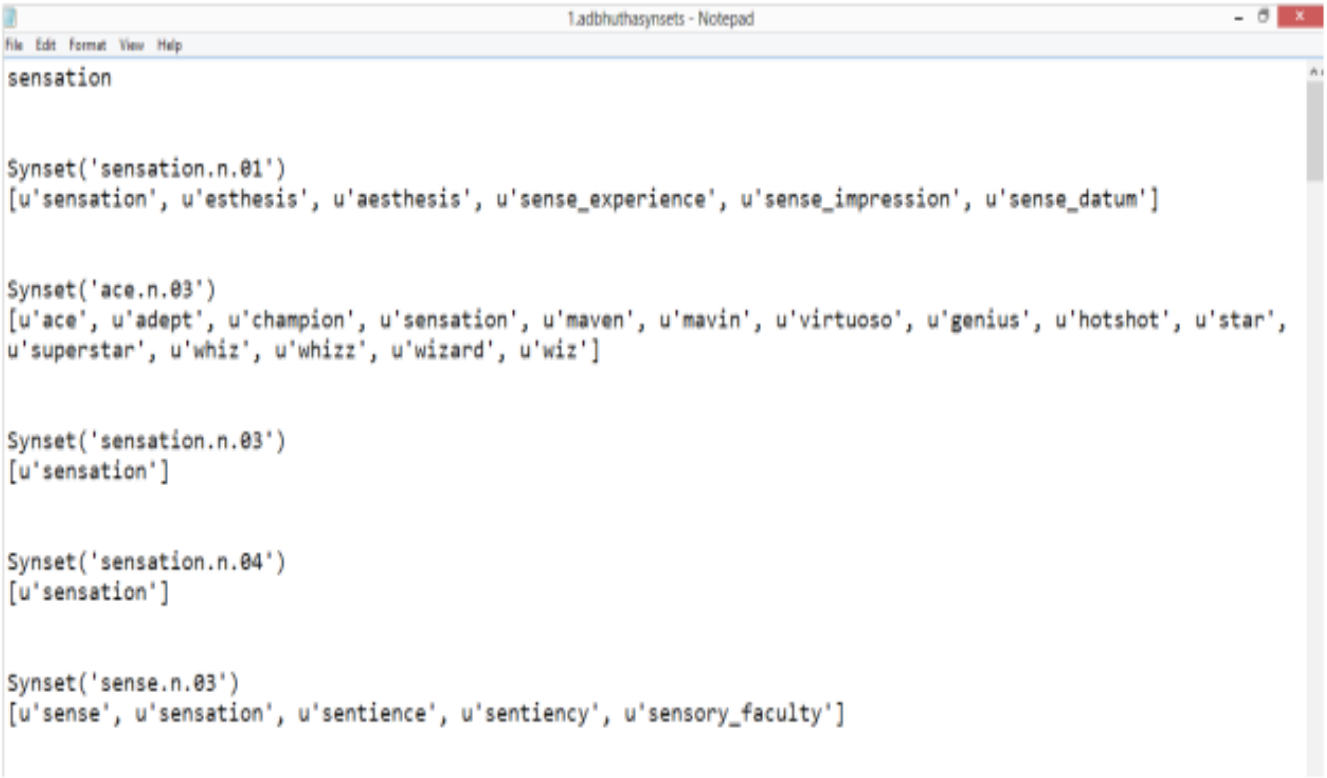


Figure 5. WordNetSynset Generation.

NLTK (Natural Language Tool Kit) library. Python code runs a loop through each emotion keyword producing synsets as shown in Figure 5.

8. Producing Relevant Synonyms

We filtered redundancies to obtain a list of relevant synonyms for each context of an emotion keyword present in the original list. That is, for each synset for an emotion keyword, we enlisted all relevant synonyms.

9. WEBSTER'S Dictionary

In the initial stage of sentiment analysis using limited emotion lexicons in our Rasa ontology, we felt the expressions given in the ontology were not sufficient to analyze the text. Hence, we decided to attach the ontology with Webster's dictionary so that a greater number of words were available to detect the sentiment nature of the work as shown in Figure 6. We condensed the Webster's dictionary into a simple text file for assessing word validity.

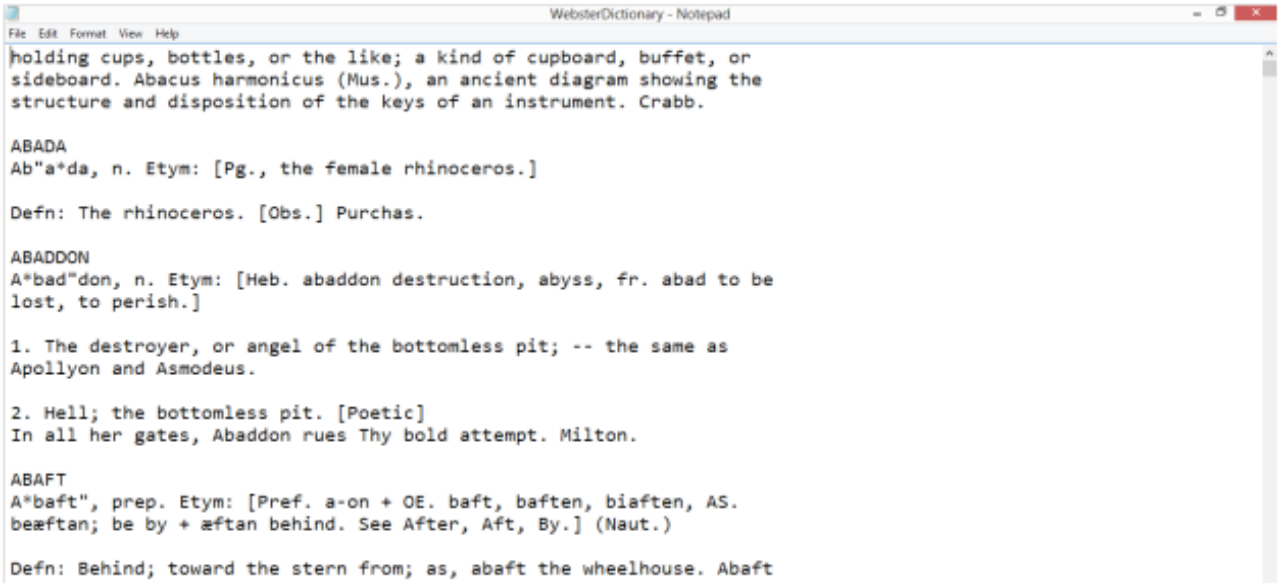


Figure 6. Portion of full Webster Dictionary.

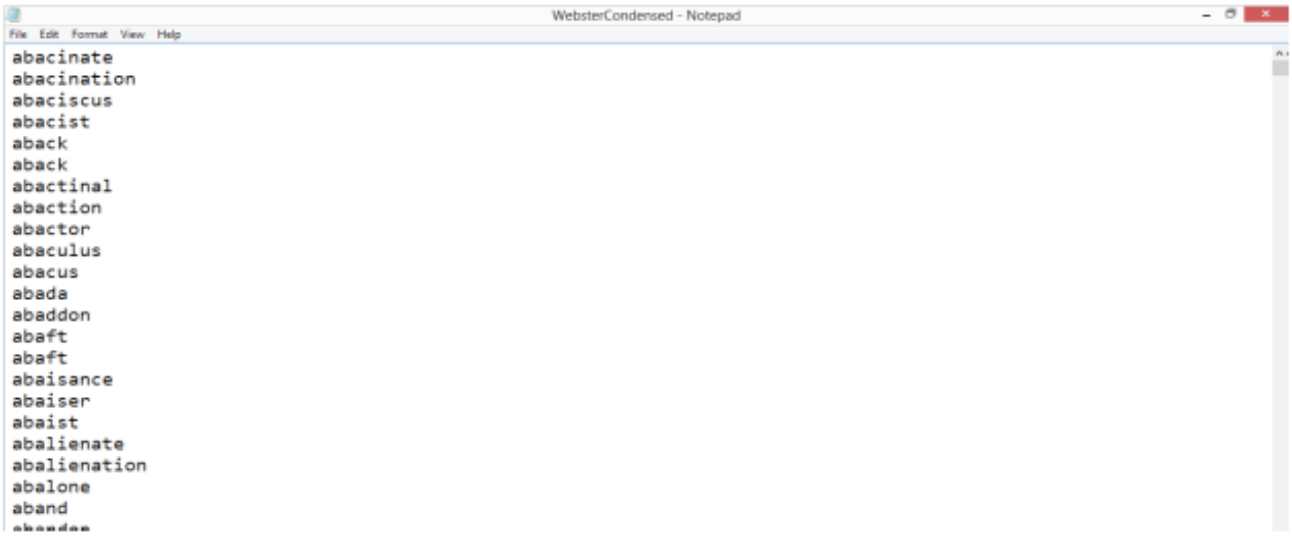


Figure 7. Portion of condensed list of Webster Dictionary words.

As shown in Figure 7. Conversion was done through pattern matching in Ruby language.

10. Word Variations

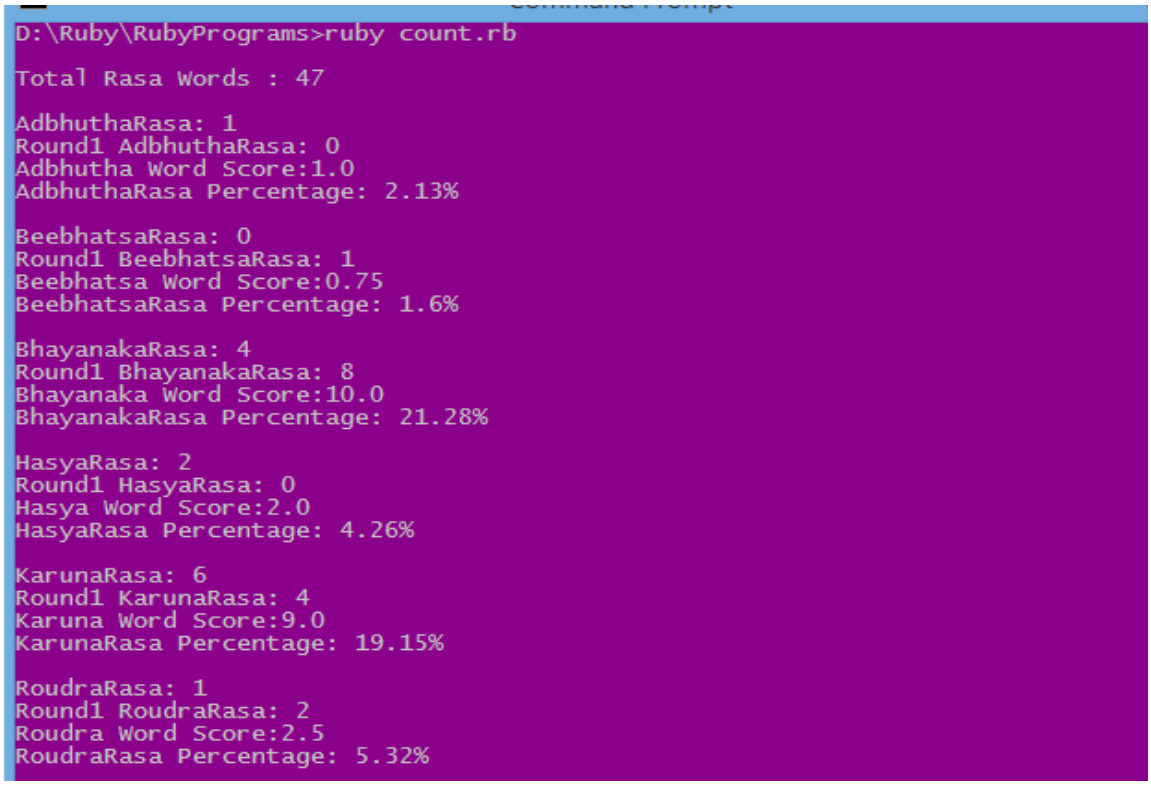
As the next step, we generated all word variations of an emotion keyword - adjectives, adverbs, verbs etc, by adding suffixes and verifying with the condensed Webster's dictionary[8]. This was done in Ruby language and once a potential suffix was added, the Ruby script was used to check and see if the newly generated word existed in the dictionary. If yes, it was added to the original emotion lexicon. If not, it was discarded. Newly generated words were mapped to a particular Rasa and inserted back into emotion lexicon.

11. Identify Emotion Keywords Present in Raw Text

Once we were able to map the emotion lexicons to the 'Rasas' given in the ontology, we proceeded to the implementation stage. With the emotion lexicons mapped to the 'Rasas', we tried to identify the emotion keywords present in the raw text[9].

12. Perform Analysis on Generated List of Emotion Keywords

This constituted the final part of our research. In this stage the raw text was analyzed with respect to the emotion keywords present and mapped with the respective Rasas. We took a count of the number of words belonging to each rasa. First we counted the emotion keywords of the raw text which belong to the ORIGINAL ontology list of emotion keywords and gave them a score of 1. Then we checked words that have come from the derived list of emotion keywords. Those words were given a score of 0.75. An overall rasa score was generated by obtaining



```
D:\Ruby\RubyPrograms>ruby count.rb

Total Rasa Words : 47

AdbhuthaRasa: 1
Round1 AdbhuthaRasa: 0
Adbhutha Word Score:1.0
AdbhuthaRasa Percentage: 2.13%

BeebhatsaRasa: 0
Round1 BeebhatsaRasa: 1
Beebhatsa Word Score:0.75
BeebhatsaRasa Percentage: 1.6%

BhayanakaRasa: 4
Round1 BhayanakaRasa: 8
Bhayanaka Word Score:10.0
BhayanakaRasa Percentage: 21.28%

HasyaRasa: 2
Round1 HasyaRasa: 0
Hasya Word Score:2.0
HasyaRasa Percentage: 4.26%

KarunaRasa: 6
Round1 KarunaRasa: 4
Karuna Word Score:9.0
KarunaRasa Percentage: 19.15%

RoudraRasa: 1
Round1 RoudraRasa: 2
Roudra Word Score:2.5
RoudraRasa Percentage: 5.32%
```

Figure 8 Perform Analysis on Generated List of Emotion Keywords.

```
ShanthaRasa: 0
Round1 ShanthaRasa: 0
Shantha Word Score:0.0
ShanthaRasa Percentage: 0.0%

SringaraRasa: 1
Round1 SringaraRasa: 16
Sringara Word Score:13.0
SringaraRasa Percentage: 27.66%

VeeraRasa: 0
Round1 VeeraRasa: 1
Veera Word Score:0.75
VeeraRasa Percentage: 1.6%


The sample text has traces of BhayanakaRasa
The sample text has traces of SringaraRasa
Polarity: Neutral

D:\Ruby\RubyPrograms>
```

Figure 9. Computational Output of Sample Text.

the sum of two. The percentage of a particular rasa was computed by comparing it with the total number of Rasa words that have been identified as shown in the Figure 8 and 9. Algorithm for assessment was done in Ruby script.

We made an attempt to analyze the short story 'War' written by Luigi Pirandello in order to find out how far the sentiment mood of the story corresponded with our newly created 'navarasa' ontology[10]. And the output says that the predominant sentiment of this short story is combinely Karuna Rasa (sorrow) and 'Bhayanaka Rasa' (Terror) which are inter related and inter mixed in this story.

13. Conclusion

With this ontology as a standard representation of knowledge, knowledge acquisition and information retrieval is facilitated. This ontology can also be used to analyze explicit details of the implicit mood of literary works.

14. Future Work

This ontology can be expanded to include idioms and phrases which will enable its utilization in the genre of literary criticism. In this paper, we have tried to detect the sentiment nature of a literary work by identifying the emotional keywords associated with the various 'rasas'. It can be expanded to the analysis of all the elements of a sentence by analyzing the neighboring words of the emotion lexicon, pronouns, articles etc. This will enable us to use Sentiment Analysis to perform literary criticism with better accuracy.

15. Acknowledgement

We gratefully acknowledge the support of Dr. Ranjani Parthasarathy (Professor, Department of Information Science and Technology, Anna University, Chennai, India) who has given valuable suggestions to write this paper.

16. References

1. Marcu D. Cambridge: MIT Press: The Theory and Practice of Discourse Parsing and Summarization. 2000; p. 1-3.
2. Hu M, Liu B. Mining and summarizing customer reviews. New York, USA: Proceedings of the tenth ACM SIGKDD. 2004; p. 168-77. CrossRef.
3. Daniel J, James MH. New Jersey: Prentice Hall Series: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 2009.
4. Ballantne JR, Pramadadasa M. Delhi: Motilal Banarasidas Publishers: The Sahitya Darpana or the Mirror of composition of visvanatha a treatise on Poetical Criticism. 1994.
5. Haider SZ. Sweden: University of Skoyde: An Ontology Based Sentiment Analysis: A Case Study. 2012.
6. Esuli A, Sebastiani F. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. Genoa: Proceedings of International Conference on Language Resources and Evaluation (LREC). 2006; p. 417-22.
7. Balamurali AR, Aditya J, Pushpak B. Harnessing WordNet Senses for Supervised Sentiment Classification. Edinburgh: Proceedings of EMNLP. 2011; p. 1081-91.
8. Hatzivassiloglou, Wiebe V. Effects of Adjective Orientation and Gradability on Sentence Subjectivity. Germany: Proceedings of the 18th conference on Computational linguistics. 2000; 1:299-305. CrossRef.
9. Janyce WM. Identifying Subjective Characters in Narrative. Finland: Proceedings of the COLING. 1990; p. 401-08.
10. War. Available from: http://www.mpsaz.org/stapley/staff/jkmiller/ elp8/8thhonorsassignments/ files/war_by_luigi_pirandello.pdf . Date accessed: 13/01/2016.