



*Deepu S. Nair, Jisha P. Jayan,*

*Rajeev R.R, Elizabeth Sherly*

Virtual Resource Centre for Language

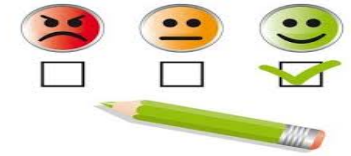
Computing

IITM-K

Technopark

# SentiMa – Sentiment Extraction For Malayalam

# Overview



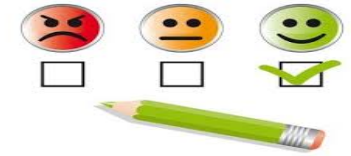
- Introduction
- Related Work
- Proposed Work
- Result
- Conclusion
- Reference



# Sentiment Analysis

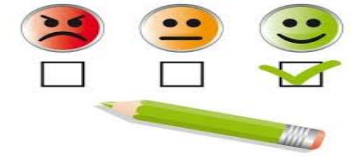
- It is a process by which one can extract the subjective or conceptual information from various sources.
- Sentiment Analysis is a cognitive process.
- It can be a mere judgement of a topic like writer's documents, speaker's speech or user's review.
- Sentiment Analysis is one of the hottest research areas in computer science.





## Problem Description

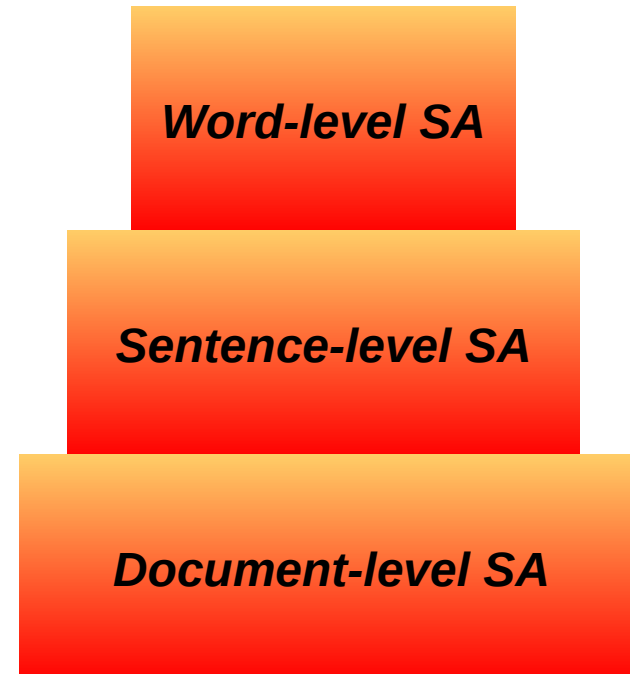
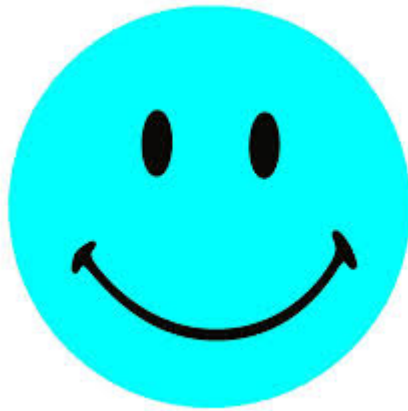
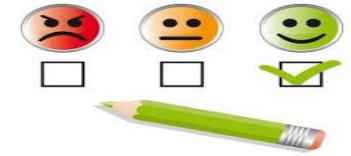
- Analysis of the sentiments in Malayalam text mainly focusing on film reviews and to check whether reviews are either positive, negative or neutral .
- Analysis of sentiments from smilyes.



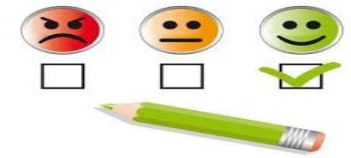
# Why Sentiment Analysis ?

- Movie : Is this review positive or negative?.
- Product : What do people think about the new iphone?.
- Aid in decision making.
- Business Strategies.
- Education.
- Government.

# Level of Sentiment Analysis?



# Challenges



- Unstructured Data.
- Ambiguous words.
- Noise.

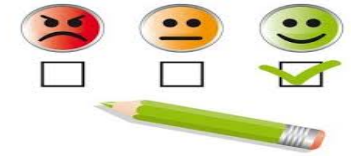


# Literature Review

Paper	Author	Content
Sentiment Analysis: A New Approach for Effective Use of Linguistic Knowledge and Exploiting Similarities in a Set of Documents to be Classified".	Alekh Agarwal, Pushpak Bhattacharya . Dept. of Comp Science & Engg.I.I.T. Bombay	Determines the overall polarity of a document, such as identifying for the appreciation or criticism of a movie . Using machine learning approach
Thumbs up? Sentiment classification using machine learning techniques,”	Bo Pang, Lillian Lee, and S Vaithyanathan,”	Finds which is the better way to classify the sentiment from text, either standard machine learning techniques or human-produced baseline.
”A Non-syntactic Approach for Text Sentiment Classification with Stopwords”	V Suresh, A Veilumuthu,Indian Institute of Science Bangalore, India,	Used supervised machine learning method for analysing the given text either positive, negative using SVM
“Interpreting sentiwordnet for opinion classification”.	H. Saggion and A. Funk.	Calculates positive and negative score for a review and based on the maximum score, the polarity of the review was assigned.
Domain Specific Sentence Level Mood Extraction In Malayalam Text	Geethu Mohandas, Janardhanan P S Nair, Govindaru V	Proposed a semantic orientation method for extraction of the mood from any sentence.

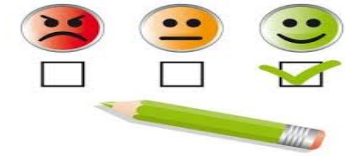


# Malayalam

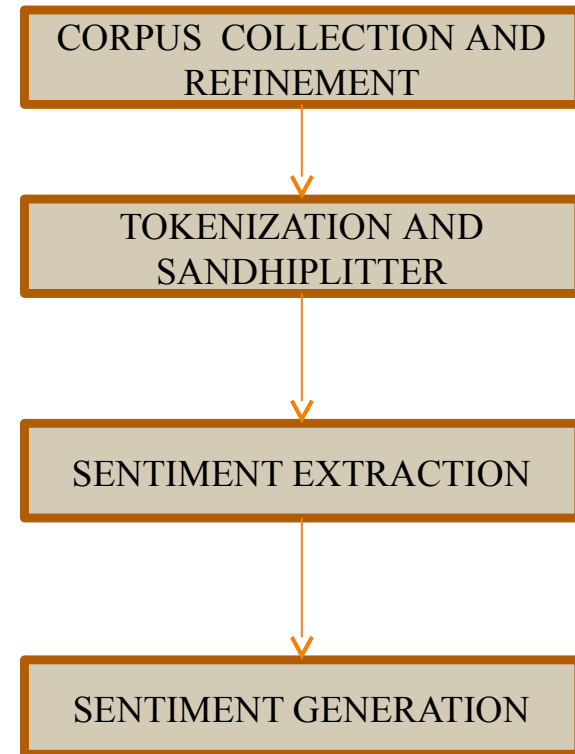


- Belongs to the Dravidian family of languages .
- The most characteristics feature of the Dravidian languages is that they are agglutinative and exhibit the inclusive and exclusive feature.
- Malayalam :Is rich with literary tradition, derived from Tamil and is influenced by Sanskrit and Prakrit.
- Requires many morphophonemic changes in the word formation.

# Proposed Work



- Focus film reviews as domain.
- Introducing new approach for extract the sentiment from user's review.
- Implementation of Rule based approach.
- Simple as well as complex Malayalam sentence can be analysed and check whether it is either positive, negative or neutral.



Block Diagram



# Tokenization and Sandhi-Splitting

## *Tokenization :*

- Tokenization mainly deals with the splitting of sentences into various tokens.

Eg: വളരെ നിലവാരം പുലർത്തിയ സിനിമ ആയിരുന്നു = വളരെ ,നിലവാരം, പുലർത്തിയ, സിനിമ ,ആയിരുന്നു.

## *Sandhi Splitter :*

Sandhi splitting deals with splitting the combinational words in to more than one unique words for assigning the sentiment.

Eg: നല്ലതാണ് = നല്ലത് , ആണ്

# Sentiment Extraction



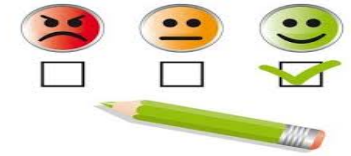
- This is the process for finding the sentiment of individual words.
- Mentioned some positive and negative data .

Example: ന + ല + ൾ + ല + ത + ൾ = Positive

ച + ൾ + ത + ൾ + ത = Negative

- If the word is not in positive or negative category then it considered as neutral.

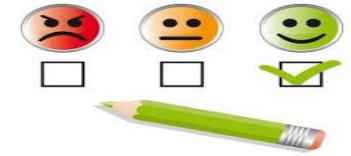
## Example



Input: പടം നിരാശ ഉളവാക്കിയില്ല എന്ന് പറയാൻ പറ്റില്ല.

പടം	-	Neutral
നിരാശ	-	Negative
ഉളവാക്കിയില്ല	-	Negative
എന്ന്	-	Neutral
പറയാൻ	-	Neutral
പറ്റില്ല	-	Negative

# Method For Analysing Sentiment



- To find the overall sentiment of the entire sentence.
- Use Negation Rule.

```
if word = 'Positive' then
```

```
    pos++
```

```
else if word = 'Negative' then
```

```
    neg++
```

```
else if word = 'Neutral' then
```

```
    neu++
```

```
if  $neg \% 2 = 0$  then return Positive
```

```
else if  $neg \% 2 \neq 0$  then return Negative
```

```
else if  $neg = 0 \ \&\& \ pos \geq 1$  then return Positive
```

```
else if  $neg = 0 \ \&\& \ pos = 0$  then return Neutral
```

# Result



*Input* : സിനിമയിലെ ഓരോ സന്ദർഭങ്ങളും വളരെ മികച്ചതായിരുന്നു.

*Output* : Feeling Happy (Positive)

*Input* : പടം വളരെ നിലവാരമില്ലാത്തതാണ് എന്നാണ് എന്റെ അഭിപ്രായം.

*Output* : Feeling Sad (Negative)

*Input* : പടത്തിൽ കുറേ പാട്ടുകൾ ഉണ്ടായിരുന്നു.

*Output* : Neutral.

*Input* : ഇതാണ് പടം 

*Output* : Feeling Happy (Positive)

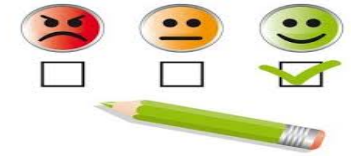


## Result Analysis

പടം മികച്ചതായിരുന്നു.	Correct
സിനിമ മോശമല്ലായിരുന്നു.	Correct
പാട്ടുകൾ മനോഹരമല്ലായിരുന്നു.	Correct
കാണാവുന്ന ഒരു ചിത്രമാണ് എന്ന് കൊണ്ടും ലോപോയിന്റ് .	Correct
വളരെ നിലവാരം പുലർത്തിയ സിനിമ ആയിരുന്നു.	Correct
ചിത്ര പടമായിരുന്നു.	Correct
വൃത്തികെട്ട പടമായിരുന്നു.	Wrong
സെക്കന്റ് ഹാഫിൽ സ്ക്രീനിൽ നിന്നുകണ്ണെടുക്കാനേ തോന്നില്ല അത്രയ്ക്ക് മനോഹരമായ വിഷ്വൽസ്.	Wrong



# Conclusion & Future Works



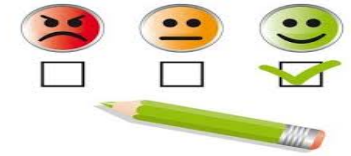
## *Conclusions :*

- This proposed work has been tested with 100 unique sentence and their combination successfully.
- It should be helpful for giving rank to the new arrived film.
- Users can write their suggestion and opinion about the film.
- It will help to improve their product or services.
- The System performs with an accuracy of 85% .

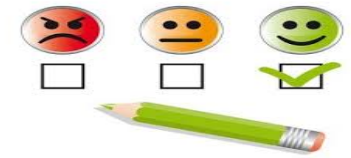
## *Future Works :*

- Machine learning approach can be applied in this work.

# Reference



- [1] Alekh Agarwal, Pushpak Bhattacharya "Sentiment Analysis: A New Approach for Effective Use of Linguistic Knowledge and Exploiting Similarities in a Set of Documents to be Classified". Dept. of Comp Science & Engg.I.I.T. Bombay.
- [2] B Pang, L Lee, and S Vaithyanathan,"Thumbs up? Sentiment classification using machine learning techniques," Proc. EMNLP-02, the conference on Empirical Methods in Natural Language Processing.
- [3] Franky and R Manurung,"Machine Learning-based Sentiment Analysis of Automatic Indonesia Translations of English Movie Reviews",University of Indonesia,ICACIA:2008
- [4] V Suresh, A Veilumuthu,"A Non-syntactic Approach for Text Sentiment Classification with Stopwords," Indian Institute of Science Bangalore, India, in Proceedings of the 20th International Conference Companion on World Wide Web: 2011.
- [5] H. Saggion and A. Funk. Interpreting sentiwordnet for opinion classification. In *LREC, 2010*.
- [6] Piyush Arora. "Sentiment Analysis For Hindi Language", MS Thesis IIIT-H 2013,
- [7] Geethu Mohandas , Janardhanan P S Nair, Govindaru V, "Domain Specific Sentence Level Mood Extraction From Malayalam Text " Advances in Computing and Communications (ICACC), 2012 .



***THANK YOU.....***