# A SEMANTIC APPROACH TO SUMMARIZATION


DIVYANSHU BHARTIYA (10250)

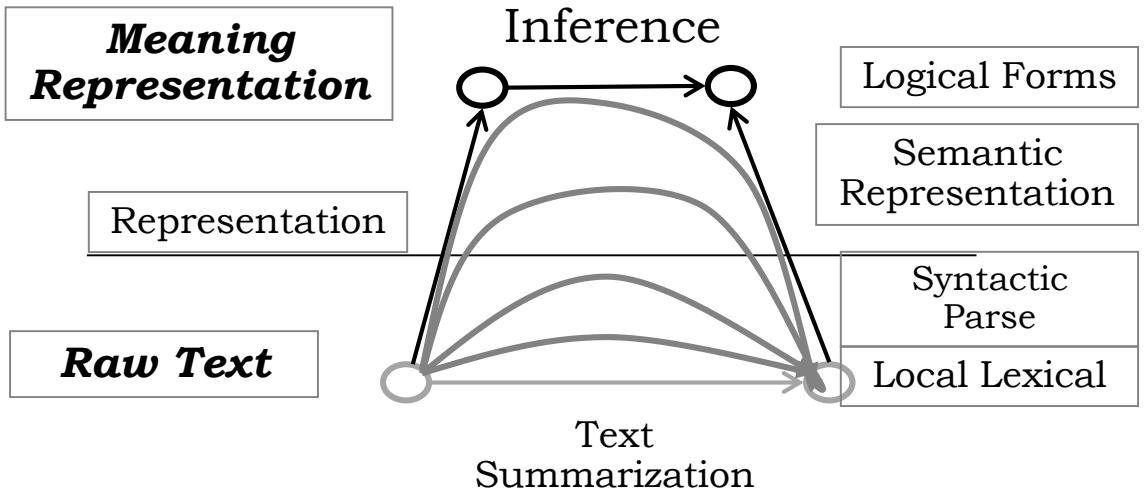
ASHUDEEP SINGH (10162)

*UNDER THE GUIDANCE OF*

PROF. HARISH KARNICK

# INTRODUCTION AND MOTIVATION

- The increasing availability of online information has necessitated intensive research in the area of automatic text summarization within the Natural Language Processing (NLP) community.
- Our main motivation was to follow the same footprints that a human takes while creating a summary. A human understands the document, and links up the parts of documents that try to convey a piece of information.
- We hereby introduce the same approach, i.e. identifying the meaning of the document, linking them up, getting the best representation and creating a concise version of it.




Text representation can be handled at various levels of representations:

- Lexical
- Syntactic
- Semantic
- Logical

# PREVIOUS WORK

- Earliest Works on Summarization:
  - Luhn, 1958 – Proposed word frequency as the measure of importance of a word
  - Baxendale, 1958 – discussed importance of words position
  - Edmundson, 1969 – explained the way key phrases help extract useful summaries

But there were no rich NLP tools available then.

- In the last decade, work on summarization has been on 3 kinds of systems- Single Document summarization, Multi-document Summarization, Query-based summarization.
- Machine Learning methods for Summary generation:
  - Naïve Bayes to determine probabilities to determine whether a sentence should be in a summary or not. (Kupiec ’95)
  - Decision Trees on various textual features. (Lin ‘99)
- Multi-Document Systems:
  - SUMMONS

# TECHNIQUES AND TECHNOLOGIES

i. **Pronominal Resolution (Anaphora Resolution)**
   - Resolving the pronouns and other dependencies amongst neighboring sentences.
   - *"John helped Mary. She was happy for the help provided by him."*
   - Most of the information is contained in the second line, but the nouns there are referenced from the first line. We need to resolve the dependencies
     *"John helped Mary. She was happy for the help provided by him."* → *"John helped Mary. Mary was happy for the help provided by John."*

ii. Part of Speech Tagging

iii. Semantic Role Labelling

iv. WordNet

# TECHNIQUES AND TECHNOLOGIES

i. Pronominal Resolution (Anaphora Resolution)

**ii. Part of Speech Tagging**

- *"Mary was happy for the help provided by John."*
- Represented as :

```
(ROOT
  (S
    (NP (NNP Mary))
    (VP (VBD was)
      (ADJP (JJ happy)
        (PP (IN for)
          (NP
            (NP (DT the) (NN help))
            (VP (VBN provided)
              (PP (IN by)
                (NP (NNP John)))))))))
    (. .)))
```

iii. Semantic Role Labelling

iv. WordNet

# TECHNIQUES AND TECHNOLOGIES

i. Pronominal Resolution (Anaphora Resolution)

ii. Part of Speech Tagging

**iii. Semantic Role Labelling**

- A shallow semantic parsing technique in NLP
- It detects the predicates associated with a verb in a sentence. It is the task of finding the arguments and the modifiers associated with a verb
- *"Mr.Bush met him privately, in the White House, on Thursday."*

  Relation: *met*

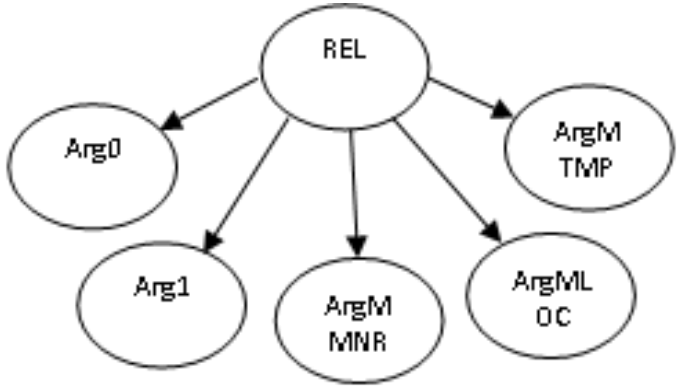
  Arg0: *Mr. Bush*

  Arg1: *him*

  ArgM-MNR: *privately*

  ArgM-LOC: *in the White House*

  ArgM-TMP: *on Thursday.*




iv. WordNet

# TECHNIQUES AND TECHNOLOGIES

i. Pronominal Resolution (Anaphora Resolution)

ii. Part of Speech Tagging

iii. Semantic Role Labelling

**iv. WordNet**

- WordNet is a freely available standard corpus from Princeton University,put to the use of Natural Language tasks.
- WordNet is a lexical database of English, comprising and arranged in Nouns, Verbs, Adjectives and Adverbs.
- They are grouped into sets of cognitive synonyms. These sets are called Synsets as they describe the semantic and lexical relations between words.

# OUR APPROACH

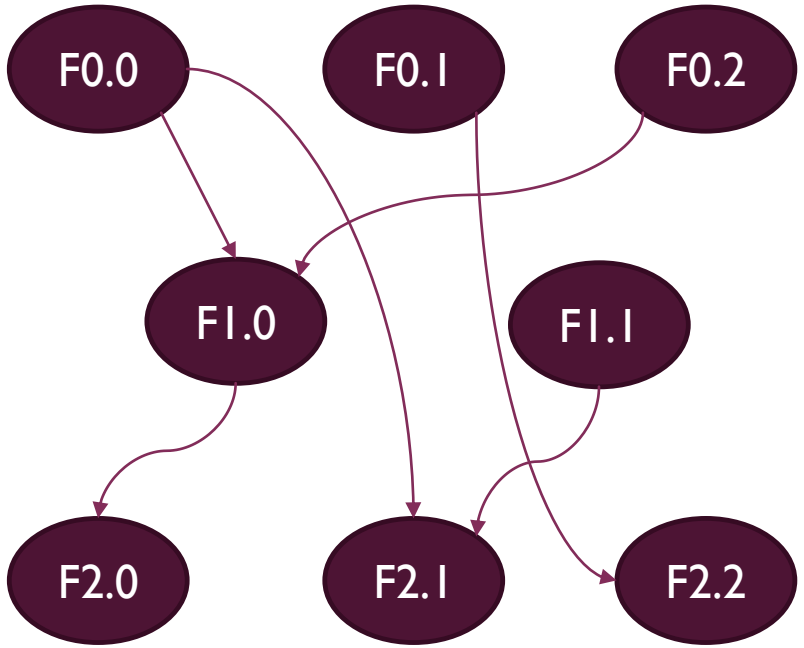
- Pronominal resolution
  - Resolve pronouns with nouns and remove ambiguity
- POS tagging
  - identify nouns and verbs
- Semantic role labelling
  - Increases granularity of document
  - Model representing the semantics of sentence
  - SENNA : neural network based semantic role labelling
  - Construction of frames with verb as root and argument as branches
  - PropBank annotation
  - $D = \cup SRL(Si) \; \forall \; i \in N$

# OUR APPROACH

- Wordnet : find semantics
  - Find synsets of arguments and verbs
  - $Frame_i.nounSynsets = \bigcup Synsets(n), \forall\ n \in Frame_i.Args$
  - $Frame_i.verbSynsets = Synsets(v), v = Frame_i.root$
  - $Synsets(n) = Hyponyms(n) \cup Hypernyms(n)$
  - $Synsets(v) = Hypernyms(v) \cup Troponyms(v)$
  - Each frame now describes the possible meanings and has a semantic representation to it

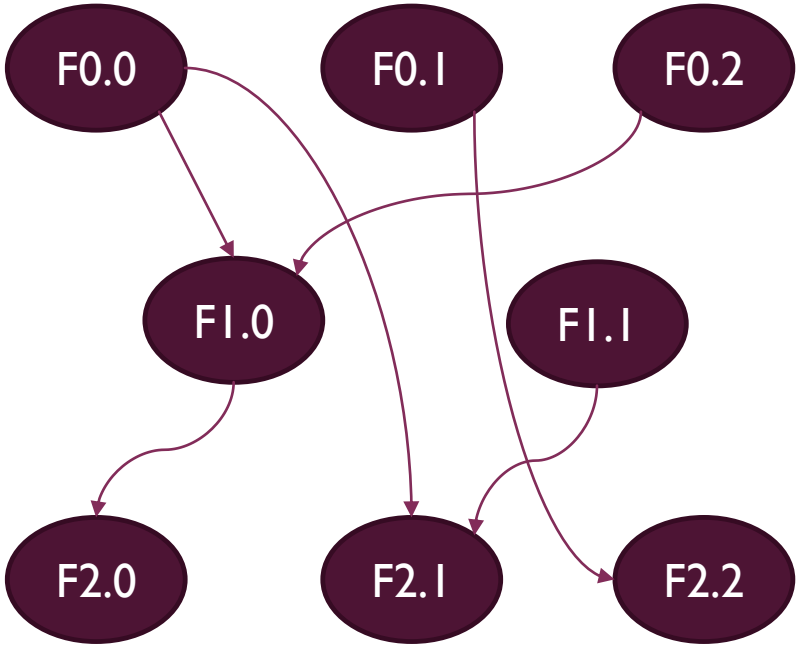
# OUR APPROACH

- Graph formation
  - Textual entailment
  - Each frame a node and each frame match a weighted directed edge
  - Priority 1: A ($F_i$, $F_j$) > 0 and V ($F_i$, $F_j$) > 0
  - Priority 2: A ($F_i$, $F_j$) > 0 and V ($F_i$, $F_j$) = 0
  - Priority 3: A ($F_i$, $F_j$) = 0 and V ($F_i$, $F_j$) > 0
  - Priority 4: A ($F_i$, $F_j$) = 0 and V ($F_i$, $F_j$) = 0
  - Weights based on the match score



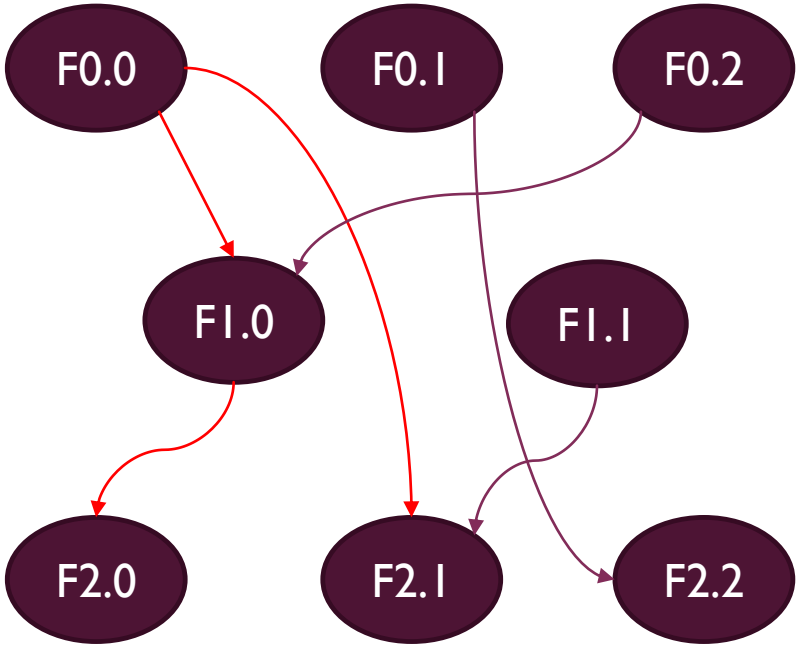
# OUR APPROACH

- Segmentation
  - Join all connected frames
  - All frames accessible from a node
  - Generates clusters in document, having similar type of information



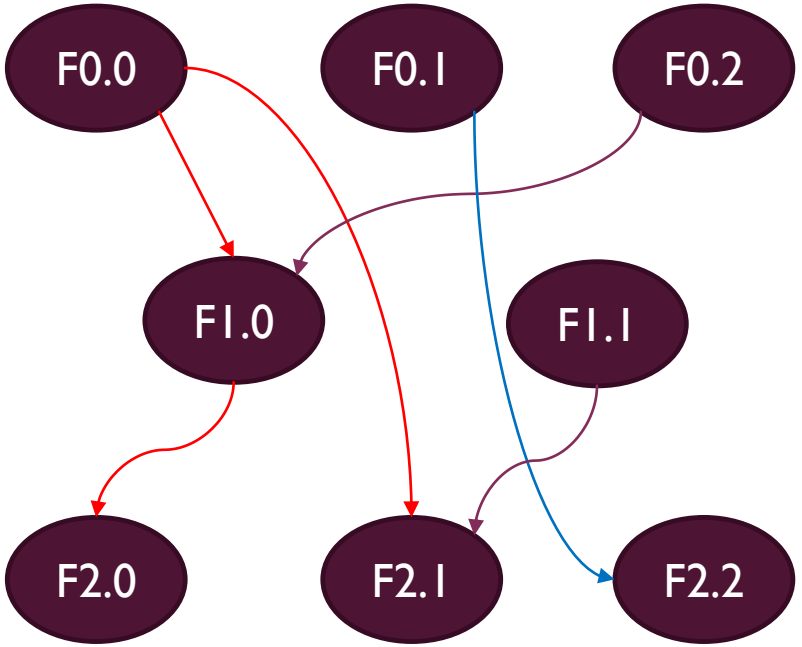
# OUR APPROACH

- Segmentation
  - Join all connected frames
  - All frames accessible from a node
  - Generates clusters in document, having similar type of information
- Segment 1 : { F0.0, F1.0, F2.0, F2.1 }



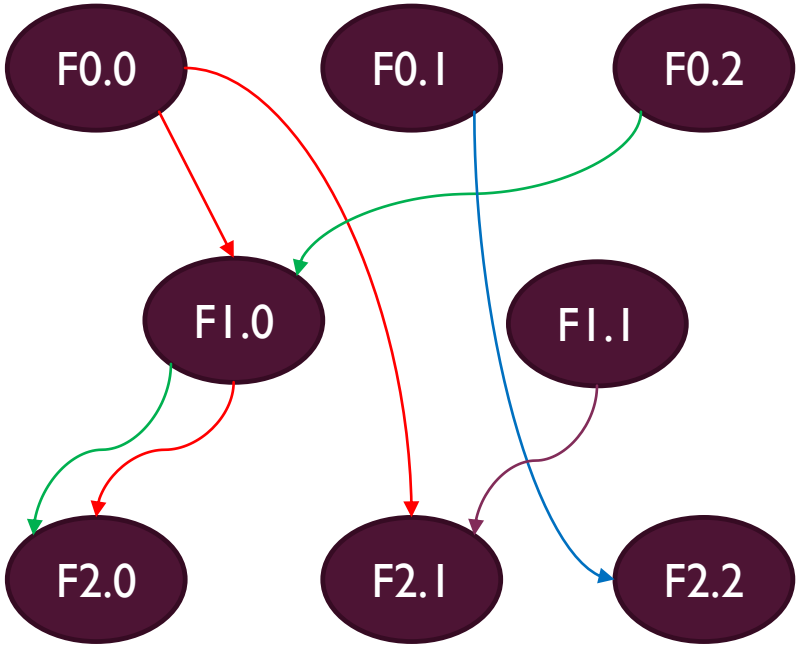
# OUR APPROACH

- Segmentation
  - Join all connected frames
  - All frames accessible from a node
  - Generates clusters in document, having similar type of information
- Segment 1 : { F0.0, F1.0, F2.0, F2.1 }
- Segment 2 : { F0.1, F2.2 }



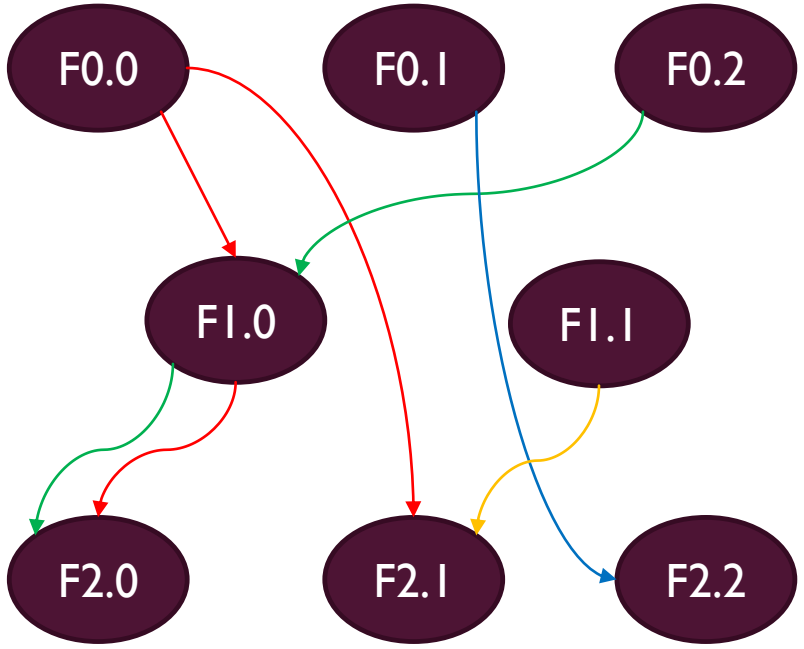
# OUR APPROACH

- Segmentation
  - Join all connected frames
  - All frames accessible from a node
  - Generates clusters in document, having similar type of information
- Segment 1 : { F0.0, F1.0, F2.0, F2.1 }
- Segment 2 : { F0.1, F2.2 }
- Segment 3 : { F0.2, F1.0, F2.0 }



# OUR APPROACH

- Segmentation
  - Join all connected frames
  - All frames accessible from a node
  - Generates clusters in document, having similar type of information
- Segment 1 : {F0.0, F1.0, F2.0, F2.1}
- Segment 2 : {F0.1, F2.2}
- Segment 3 : {F0.2, F1.0, F2.0}
- Segment 4 : {F1.1, F2.1}



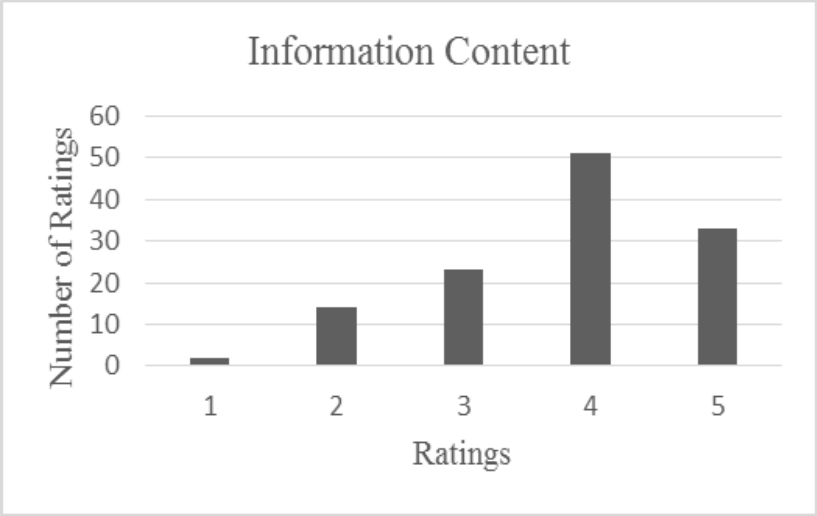
# OUR APPROACH

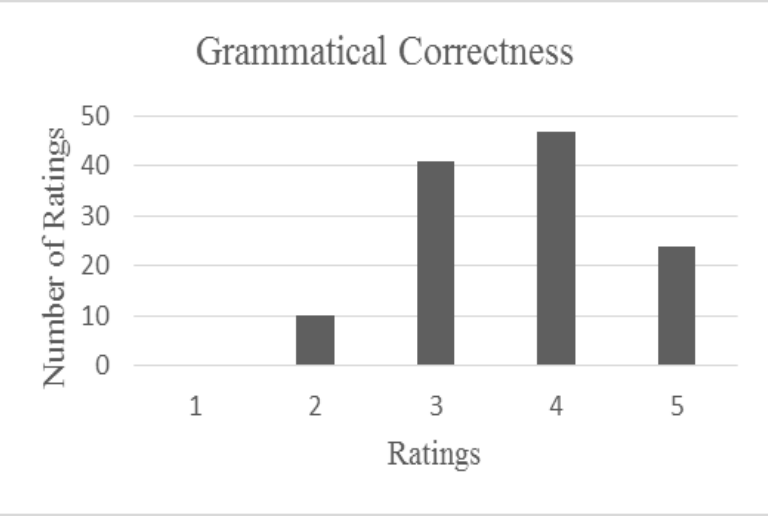
- Centroids Extraction
  - These frames will represent the information contained in a segment in a concise and representative manner
  - Find the weighted score of each frame in a segment based on certain features
  - Number of incoming edges, number of outgoing edges, frame position in document, frame length
  - The frames with highest score in each segment are picked out
- Sentence generation
  - Sentence formed by concatenating [Arg0] [verb] [Arg1] [Arg2].
  - May be grammatically incorrect

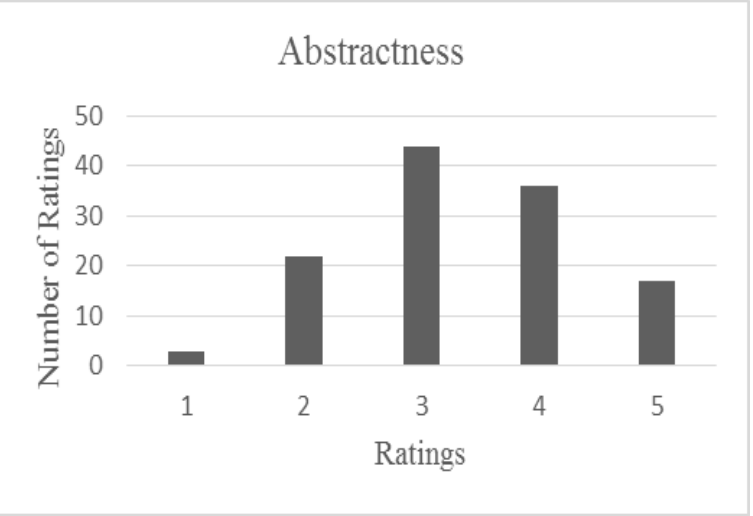
# OBSERVATIONS AND RESULTS

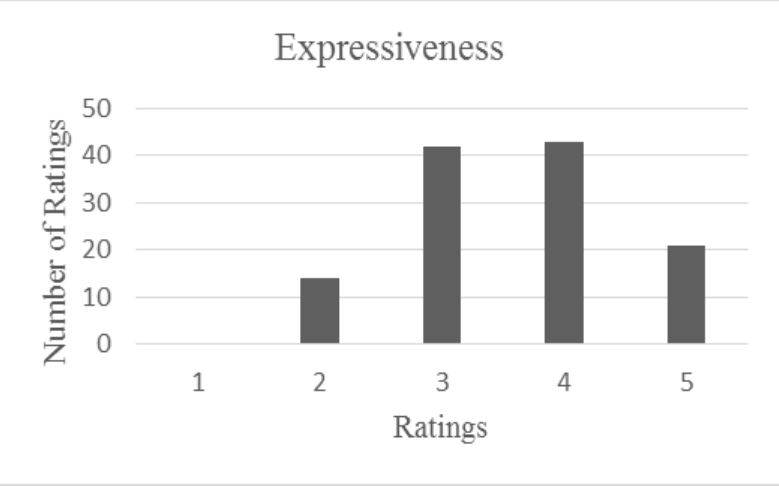
- Measured similarity between system generated summary and human written summary
- Same idea as in using wordnet and graph formation
- Edges are formed between frames of our summary and frames of human composed summary
- $Sim(S, S') = \Sigma_{i \in S.Centroids} (\text{argmax}_{j \in S'}.\text{frames sim } (i, j))$
- Similarity about 50%

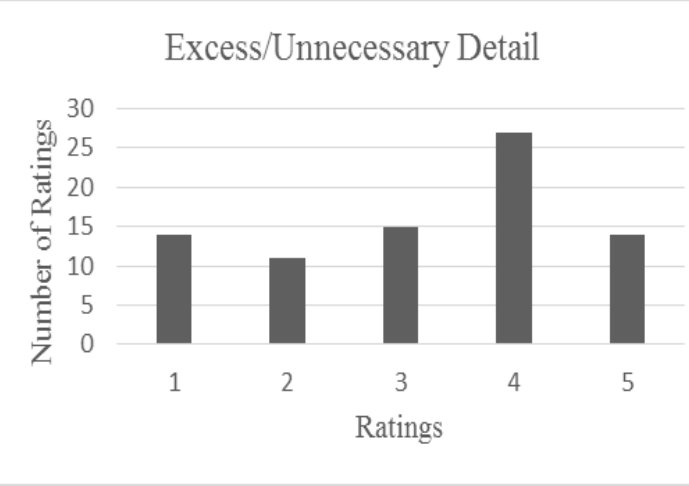
# HUMAN EVALUATION OF ATTRIBUTES



# CONCLUSION

- Further modifications that can be involved are
  - Better heuristics and features for centroid selection
  - Improved sentence generation system module
  - Involving logical entailment
- Summarization is used in news mining, customer retrieval, intelligent tutors and question-answering systems

# REFERENCES


- Das, D., & Martins, A. F. (2007). A survey on automatic text summarization. Literature Survey for the Language and Statistics II course at CMU, 4, 192-195.
- Collobert, R., & Weston, J. (2007, June). Fast semantic extraction using a novel neural network architecture. In Annual meeting-association for computational linguistics (Vol. 45, No. 1, p. 560).
- Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 168-177). ACM.

# EXAMPLE

## Reference Summary

The various components of an ecosystem are interdependent. The producers make the energy from sunlight available to the rest of the ecosystem. There is a loss of energy as we go from one trophic level to the next, this limits the number of trophic levels in a food-chain. Human activities have an impact on the environment. The use of chemicals like CFCs has endangered the ozone layer. Since the ozone layer protects against the ultraviolet radiation from the Sun, this could damage the environment. The waste we generate may be biodegradable or non-biodegradable. The disposal of the waste we generate is causing serious environmental problems

## Computer generated Summary

All organisms such as plants , animals , microorganisms and human beings as well as the physical surroundings maintain a balance in nature. All the interacting organisms in an area together with the non-living constituents of the environment form an ecosystem. An ecosystem consists of biotic components comprising living organisms and abiotic components comprising physical factors like temperature ,rainfall , wind , soil and minerals.We have seen that organisms can be grouped as producers , consumers and decomposers according to the manner in which they obtain their sustenance from the environment.Let us recall what we have learnt through the self sustaining ecosystem created by us above .Consumers classed as herbivores ,carnivores, omnivores and parasites. The microorganisms, comprising bacteria and fungi break-down the dead remains and waste products of organisms.They fix the solar energy.Food chains consist of only three or four steps. Energy moves through the various trophic levels. Several pesticides and other chemicals protect our crops from diseases and pests. from the soil these are absorbed by the plants along with water and other minerals. They enter the food chain. This process is called biological magnification. Ozone performs an essential function. It shields the surface of the earth from ultraviolet (UV) radiation from the Sun.It cause skin cancer.