

# Automated Essay Generation using Tensorflow

Nikhil Jawale

Department of Computer Engg,  
Vidyalankar Institute of Technology,  
Mumbai, India.

Bijali Mulgir

Department of Computer Engg,  
Vidyalankar Institute of Technology,  
Mumbai, India.

Vaishnavi Pednekar

Department of Computer Engg,  
Vidyalankar Institute of Technology,  
Mumbai, India.

Umesh Kulkarni

Department of Computer Engg,  
Vidyalankar Institute of Technology,  
Mumbai, India.

**Abstract**—Essay Generation is the process of creating a condensed form of a text document which maintains significant information and general meaning of source text. It is the process of creating a shorter version of one text document. Automatic essay generation becomes an important way of finding relevant information precisely in large text in a short time with little efforts. In this, the extracted information is achieved as a summarized report and conferred as a concise (shortly) summary to the user. [16]

**Keywords**— Tensorflow, embeddings, pre-processing, Decision making.

## I. INTRODUCTION

In a recent advance, the significance in automated essay generation is accomplished more attention due to data inundation (flooding) on the web. Hence this information overwhelms yields in the big requirement for more reliable and capable progressive text generator. Essay generator gains its importance due to its various types of applications just like the summaries of books, a summary of stories, the stock market, news, meetings, sports events. Abstract of scientific papers, newspaper articles, magazines, etc. To fully utilize these online documents effectively, it is crucial to be able to extract the summary of these documents, and even it is very crucial for humans to understand and to describe the content of the text. So, in this, Information overload is one of the major problems which require algorithms and tools for faster access. Essay generation system will be one of the solutions to the above problem. This project proposes a system which will be able to perform summarization of a single input document and generation of the essay. [15]

Automatic Essay Generation becomes an important way of finding relevant information precisely in large text in a short time with little efforts. In this, the extracted information is achieved as a summarized report and conferred as a concise (shortly) summary to the user. [16]

## II. EXISTING APPROACHES



Fig . 1. Existing approach

### A. Analyzing the source text (Preprocessing):

In the first phase, the preprocessing of the text document is done to obtain a structured representation of the original version. The preprocessing step includes:

- 1) Stop-word elimination.
- 2) Contradiction elimination.

### B. Determining the salient features (Restructuring).

C. Make decisions using some frequency base algorithms.[18]

## III. LITERATURE SURVEY

After analyzing the structure of the current recommendation system used by the text summarizer, we have aimed at working on our proposed idea. The first step towards it is to refer some of the papers related to the concept required in our proposed system. The research papers involved to determine various aspects of our proposed scheme are as follows:

Zhang and Li used a cluster of sentences using K means clustering algorithm.[1]

Patil and Mahajan extract and group representative sentence from a research article. [2]

Wu et al. proposed spectral clustering and LexRank approach that leads to maximum coverage and minimum redundancy. The sparse matrix of similar sentences is generated using a k-nearest neighbor method. LexRank score is calculated based on a common feature to create a summary. [3]

Ferreira et al. propose a clustering algorithm with a graph model. A document converted into the graph; important sentences are identified using TexRank cluster is formed based on the similarity between sentences.[4]

There are numerous researches in text mining area; specialized in the automatic text summarization. Years ago, pair of researchers listed the trends of Automatic Text Summarization over the years such as statistical approach , natural language processing (NLP), semantic analysis approach , fuzzy logic and swarmed intelligence .

We are analyzing the strengths and limitations of previous methods. The purpose is to outrank the prior techniques. Inside the category of semantic analysis approach, the existence of ontology knowledge tends to be critical proven by related researches [9].

The common ontology knowledge is WordNet as a lexical ontology word resource. Therefore, our research used WordNet for the Indonesian language as the lexical word resources. [17]

#### IV. PROPOSE SYSTEM

Automated Essay Generation is the process of creating condensed form of a text document which maintains significant information and general meaning of source text. It is the process of creating a shorter version of one text document. Automatic text summarization becomes an essential way of finding relevant information precisely in large text in a short time with little efforts. In this, the extracted data is achieved as a summarized report and conferred as a concise (shortly) summary to the user.

##### A. Tensorflow:

Tensor Flow software library for high-performance numerical computation. Its flexible architecture allows easy deployment of calculation across a variety of platforms (CPUs, GPUs, TPUs), and from desktops to clusters of servers to mobile and edge devices. It comes with strong support for machine learning and deep learning, and the flexible numerical computation core is useful across many other scientific domains. [19]

##### 1) REMOVE CONTRADICTION:

A contradiction is a kind of semantic relation between sentences. Disagreement occurs when penalties are unlikely to be correct at the same time. For example, the contradiction happens between the sentence pair "Some people and vehicles are on a crowded street" and "Some people and vehicles are on an empty street." Contradiction detection aims to recognize the different meanings between the two sentences, natural language processing, such as information integration, inconsistency discovery and sarcasm detection.

##### 2) STOP WORDS REMOVAL:

Stop words are commonly used words (as "the", "a", "an", "in"). Every search engine is programmed to ignore, the result of a search query and the indexing entries for searching. These words take processing time and space in database.

##### 3) TOKENIZATION

The process of breaking the text into smaller components (terms, sentences, and bigrams).

##### C. SEQUENCE TO SEQUENCE (SEQ2SEQ):

Sequence-to-sequence learning (Seq2Seq) is about training models to convert sequences from one domain (e.g., sentences in English) to courses in another area (e.g., the same sentences translated to French).

##### D. LONG SHORT-TERM MEMORY (LSTM):

Long Short-Term Memory (LSTM) are units of Recurrent Neural Network (RNN). An RNN composed of LSTM units is often called an LSTM network. A common LSTM unit is consists of a cell, an input gate, an output gate and a forget gate. The battery remembers values over arbitrary time intervals, and the three gates regulate the flow of information into and out of the cell. LSTM networks are well suited for processing, classification, and prediction.

##### E. NUMBER BATCH METHOD:

Conceptnet Number batch is a set of semantic vectors:

It associates words and phrases in a variety of languages with lists of 600 numbers, representing the gist of what they mean.

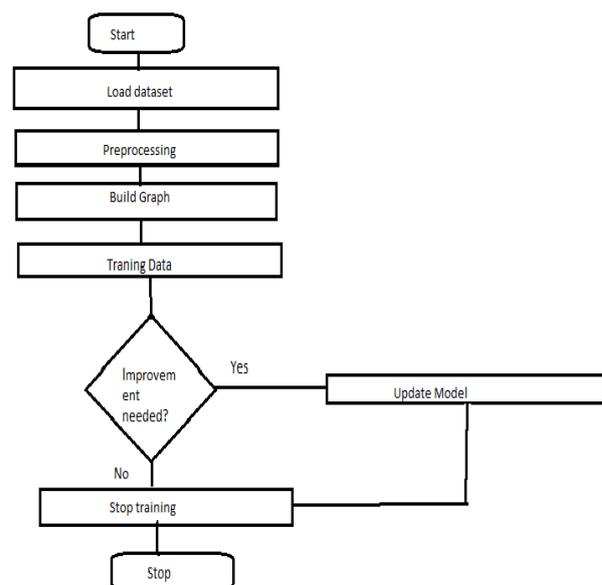


Fig. 2. The flow of the training data

#### V. OVERVIEW OF THE SYSTEM

The overall system, pre requirement includes, samples of text files from which you want to generate an essay. After the collection of raw data, this raw data is pass to the preprocessing module where stop words and contradictions are deleted. After removing stop words and

inconsistencies, tokenization is work on words. After that, we will use Sequence-to-sequence learning model to convert sequence from one domain to another domain. Where Sequence-to-sequence model consists of LSTM encoder and LSTM decoder.

The output of LSTM encoder forwarded to the LSTM decoder which predicts the next word depending upon the input provided by a user to LSTM encoder.

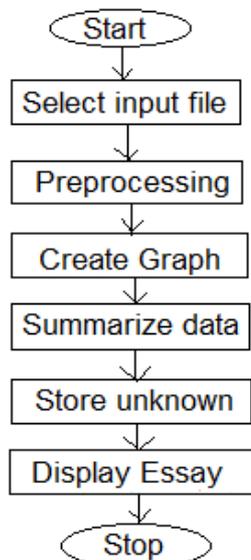


Fig. 3. Flowchart of system

## VI. CHALLENGES ANTICIPATED IN IMPLEMENTATION

After understanding the process of generating essay in detail, we have studied the requirements of implementation of the system. In due course, we have anticipated the following challenges in execution.

### A. Spelling mistakes

The existence of words representing some information. But whenever the system detects that the input text contains errors or spelling mistakes it ignores.

### B. Mis-arranging of words

There can be a scenario, where the worlds will be jumbled, or there can be no meaning of the sentence then the system will store the relevant data for further processing if required.

## VII. CONCLUSION AND FUTURE SCOPE

Although research on summarization started way long back, there is still a long way to go. Over time, focused has drifted from summarizing scientific articles to advertisements, blogs, electronic mail messages, and news articles. Simple eradication of sentences has composed satisfactory results in large applications. Some trends in automatic evaluation of summary system focused. [14].

This software shows that deep learning based approach are promising and give some hope in solving abstractive text summarization which had been largely unsolved till now. But in future we can use more effective framework for word embedding that may give more efficient and meaningful summary for essay.

## REFERENCES

- [1] P.-y. Zhang and C.-h. Li, "Automatic text summarization based on sentences clustering and extraction," in 2nd IEEE International Conference on Computer Science and Information Technology, vol. 1, pp. 167–170, IEEE, 2009.
- [2] S. R. Patil and S. M. Mahajan, "A novel approach for research paper abstracts summarization using cluster-based sentence extraction," in Proceedings of the International Conference & Workshop on Emerging Trends in Technology - ICWET '11, no. Icwet, p. 583, ACM Press, 2011.
- [3] K. Wu, P. Shi, and D. Pan, "An Approach to Automatic Summarization For Chinese Text Based on the Combination of Spectral Clustering and LexRank." In 12th International Conference on Fuzzy Systems and Knowledge Discovery, pp. 1350–1354, IEEE, 2015.
- [4] R. Ferreira, L. de Souza Cabral, F. Freitas, R. D. Lins, G. de Franc, a Silva, S. J. Simske, and L. Favaro, "A multi-document summarization system based on statistics and linguistic treatment," Expert Systems with Applications, vol. 41, pp. 5780–5787, Oct 2014.
- [5] H. P. Luhn. The automatic creation of literature abstracts. IBM Journal of Research Development, 2(2):1 59-165, 1958.
- [6] Rakesh Verma, Ping Chen, and Wei Lu. A semantic free-text summarization system using ontology knowledge. Document Understanding Conference DUC 2007, pages 1-5, 2007.
- [7] Farshad Kyoomarsi, Hamid Khosravi, Esfandiar Eslami, and Pooya Khosravayan Dehkordy, and Asghar Tajoddin. Optimizing text summarization based on fuzzy logic. ICIS '08 Proceedings of the Seventh IEEEIACIS International Conference on Computer and Information Science pages 347 - 352, 2008.
- [8] Mohammed Salem Binwahlan, Naomie Salim, and Ladda Suanmali. Fuzzy swarm based text summarization I, 2009.
- [9] Mohsen Pourvali and Mohammad Saniee Abadeh. Automated text summarization base on lexical chain and graph using wordnet and knowledge base. International Journal of Computer Science Issues (IJCSI), 9:343, 2012. ss
- [10] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller. Introduction to WordNet: an on-line lexical database. International Journal of Lexicography, 3(4):235-244, 1990.
- [11] Desmond Darma Putra, Abdul Arfan, and Ruli Manurung. Building an Indonesian word net, 2008.
- [12] ChukFong Ho, Masrah Azrifah Azmi Murad, Rabiah Abdul Kadir, and Shyamala C. Doraisamy. Word sense disambiguation-based sentence similarity. In Chu-Ren Huang and Dan Jurafsky, editors, COLING (Posters), pages 41 8-426. Chinese Information Processing Society of China, 2010.
- [13] Aristoteles, Yeni Hendriyeni, Ahmad Ridha, and Julio Adisantoso. Text feature weighting for summarization of documents in Bahasa Indonesia using a genetic algorithm. International Journal of Computer Science Issues, 9: 1, 2012.
- [14] N. Moratanch, S. Chitrakala. "A survey on extractive text summarization," 2017 International Conference on Computer, Communication and Signal Processing (ICCCSP), 2017
- [15] Angelin Florence, Vijaya Padmadas. "A summarizer system based on a semantic analysis of web documents," 2015 International Conference on Technologies for Sustainable Development (ICTSD), 2015, IEEE Xplore.
- [16] Narendra Andhale, L.A. Bewoor. "An overview of Text Summarization techniques," 2016 International Conference on Computing Communication Control and automation (ICCUBEA), 2016
- [17] Pandu Prakoso Tardan, Alva Erwin, Kho I Eng, Wahyu Muliady. "Automatic text summarization based on semantic analysis approach for documents in the Indonesian language," 2013 International Conference on Information Technology and Electrical Engineering.(ICITEE), 2013
- [18] <https://www.shodhganga.inflibnet.ac.in>
- [19] <https://www.tensorflow.org>