Legal Case Summarization: An Application for Text Summarization

Kanika Agrawal
Department of Computer Science
Roorkee Institute of Technology
Roorkee, Uttarakhand, India
ka841995@gmail.com

***Abstract*— Today in the realm of developing advances, there is a gigantic development in technologies. These days every single theme identified with research is accessible on the web. Consequently to profit the clients, it is required to condense the substance as short as possible. Different research networks are proposing enormous number of methods to abridge the substance. So in this exploration an extractive methodology is executed so as to save Indian lawyer's time in dealing with the big cases. This research uses web scrapping technique that helps in direct extraction of data from online resources hence results in accurate results. This is done using python on ANACONDA to summarize the IT LAW cases which are extricated using web scrapping strategy. After the extraction, different techniques such as LUHN, LSA, LEXRANK and SUMBASIC are used to summarize the content and calculate the ROUGE scores. The exploration just compares these four techniques on the basis of their performance metrics of ROUGE-1 and ROUGE-2 scores and according to the precision, recall and f-measure scores finally concludes that LUHN and LSA are the two best algorithms according to the ROUGE-1 and ROUGE-2 scores.**

Keywords—Text summarization; LUHN; LSA; LEXRANK; SUMBASIC.

I. INTRODUCTION

In this advanced universe of quick progress, today we are having a gigantic measure of information in our storehouses. Since the information is enormous it needs the system by which the information can be proficiently investigated in a short period of time. Henceforth it is the text mining methods which aides in doing as such. It utilizes a few sorts of examples for information retrieval which aides in further examination. The determination of right strategy helps in diminishing the time and expands the throughput. The process of removing extraneous content from the document while retaining the major points of the text in order to shorten the length of the paragraph is called text summarization.

This research focuses on to help various lawyers to help in their cases. Apart from long descriptive cases, lawyers have other works to deal with. Hence to ease their burden this research is done to save their time and to increase the efficiency of their work. The work uses web scrapping technique to extract the IT Legal cases that are present online. In other researches usually people directly apply the algorithms on a given dataset, but here the work is directly carried on the online resources that help in accurate extraction and analyses of the data. Based on their Rouge scores and depending on their various performance metrics various algorithms are compared. Finally the comparison is done to predict the best of two algorithms .The technique to help lawyers is a new innovative work to be applied in our country. Hence this is just a step to carry forward the work and help our Indian lawyers. Usually text summarization is the last step of text mining techniques. It benefits the clients to abridge the substance as short as conceivable. Different types of text summarization are given as below:

A. Extraction-based summarization

In extraction based summarization, the content objects are extracted without any modification in themselves.

B. Abstraction-based summarization

Abstraction based summarization usually involves paraphrasing of the words and the sentences from the document.

C. Aided summarization

Apart from the Fully Automated Summarizers (FAS) that automatically summarizes the text, there are many machine systems that help the people to summarize the text such as Machine Aided Human Summarization (MAHS) and Human Aided Machine Summarization (HAMS). Generally there are two types of extractive summarization:

- Generic Summarization, which produces generic type of summary.
- Query Based Summarization, in which document is summarized depending on some specific query.

Other types of summarization problems includes: Single source document summarization, Multi-document summarization, Image summarization, Video summarization, Supervised and unsupervised type of summarization , Multi-lingual or cross-lingual summarization, Email based summarization, Web–based summarization, Personalized summarization and Sentiment based summarization.

978-1-7281-4514-3/20/$31.00 ©2020 IEEE

II. Applications of Text Summarization

Nowadays text summarization has various applications. Some of them are listed below:

- Media controlling
- Newsheadlines creation
- Internal document management in a company
- Legal cases analysis
- Question answering
- Video scripting
- Medical cases Analysis
- Books and literature
- Email content
- Making class assignments
- Research and Development
- Video-conferencing and meetings
- Helping disabled people

III. Literature Review

With the growing technologies and the data, it is required to shorten the content as much as possible to benefit the users. Till today many new approaches are developed which are as discussed. One of the research articles [1], proposed extractive summarization algorithms over the datasets having critical attributes like gender and political leaning merged with the text, and finally generated the summaries. They basically worked on fairness of summarization. Research article [2], describes an approach which tells about the technique to summarize different communication sources into one big summary using the improved text summarization approaches. Research article [3], proposed a technique called Topic Aspect-Oriented Summarization (TAOS) which is based on topic factors. These topic factors help in describing different topics. Various topics have various aspects and to represent various aspects, various preferences of features are used. Research article [4], proposed a Folksonomy-based system which summarizes multiple documents for extracting important sentences. It is a system used for creating and managing tags assigned by users. Research articles [5] and [6], proposed Reader-Aware summarization system for Multiple Documents (RA-MDS). It generates summaries using sparse-coding technique that considers the reader comments also to generate the report at that particular time. Research article [7], used generic operators and guided local search for single documents. It uses a memetic algorithm which has combined the guided local search strategy with that of population based search of evolutionary algorithm. This type of summarization is also treated as binary optimization problem. Single and multi-document summarizations are the two important categories of summarization which are based on number of documents [8][9]. Task to summarize multiple documents is more tedious than the task of shortening single documents. Because of redundancy issue faced during summarization there are some techniques which initially selects the sentences at the beginning of the paragraph and then measures the similarity of the next sentence with the already chosen above sentences and if the later sentence consists of one new matter , then only it is selected [10]. Maximal Marginal Relevance (MMR) approach is suggested in research article [11] for reducing redundancy. In research article [12], the authors have compared different machine learning algorithms such as Naïve Bayes, Decision Tree and Support Vector Machines on Cardiotocography data to predict the best algorithm out of them. In research article [13], authors showed the different techniques, applications and challenges faced by text analysis.

IV. Techniques Used

A. LEXRANK

It is a technique which uses eigenvector centrality on graph to calculate the importance of textual units based on similarity of the units.

B. LSA

It is a technique which is based on SVD on the matrix to estimate the importance of the textual units by constructing terms-by-units matrix.

C. LUHN

It is a technique used for each textual unit to derive ‘significance factor’ based on placements and occurrences of frequent words within the unit.

D. SUMBASIC

It is a technique which uses re-weighting of the word probabilities and frequency-based selection of textual units to minimize redundancy.

V. Performance Measures Used

Recall-Oriented Understudy for Gisting Evaluation (ROUGE) is a combination of certain metrics that are used to evaluate the different texts that are summarized and various machine translations. It compares an automatically produced summary against a set of reference summary (summary produced manually by human). Different types of ROUGE scores are there. Below description gives a brief view over them:

A. ROUGE-N

According to the n-gram overlap it helps in measuring unigram, bigram and higher order overlap.

B. ROUGE-L

Using longest common subsequence technique, it measures the sequence of words that are longest matching.

C. ROUGE-S

It is usually called as skip gram co-occurrence technique that helps in allowing arbitrary gaps.

In this article, using Rouge package, ROUGE-1 and ROUGE-2 scores are calculated. ROUGE-1 depends on unigram overlaps of the words while ROUGE-2 depends on bigram overlaps of the words. Precision, Recall and F-

measures are calculated on the data and the better of two algorithms are suggested for the dataset of IT LAW legal cases.

A. RECALL

It tells about the percentage coverage of human produced summary by the system summary. Formula in terms of words overlapping is given as:

Recall = (No. of overlapping words) / (total words in reference summary)

B. PRECISION

It measures the percentage of system summary that is relevant for the work. Formula for precision in terms of overlapping of words is measured as:

Precision = (No. of overlapping words) / (total words in system summary)

C. F-MEASURE

It is a harmonic average of both recall and precision. The best value of f-measure is calculated as 1 and worst value as 0.

F-measure = (2*recall*precision) / (recall + precision)

VI. IMPLEMENTATION

Figure below displays the flowchart of automatic text summarization system used in the research. Initially, web scrapping technique is used to extract the data from the website. The website used here is 'indiankanoon.org'. Here the legal documents related to "IT LAW" are extracted. After extracting the documents the direct links were made available to see the documents. The code is done in such a way that it asks for the direct query related to the documents to be fetched. All the legal documents are then saved in MyDocuments folder as .txt files in the folder. Then preprocessing of the source document like segmentation of sentences, removal of stop-words, removal of punctuation marks, stemming, etc are done using the python libraries, pandas and nltk. Segmentation helps in dealing with the text to divide them into sentences. Then elimination of stop-words and noisy terms i.e. the elimination of the words that occur frequently in the text but have no contribution in selecting the important sentences, for example prepositions, articles, pronouns, etc is done.

Their removal helps in natural processing task execution. Stemming is used for reducing the words with the same root or stem to a common form, thus removing the variable suffixes. For the summarization process LUHN, LSA, LEXRANK and SUMBASIC algorithms are used. The python code has direct libraries for all the summarization to be executed. The 'Sumy' package is used to summarize the text files and saves all this data in the separate folders of these algorithms. Now when the summarization is done, ROUGE library is imported to compute the rouge scores for different text files. Rouge-1 score and Rouge-2 scores are computed for all the text files. Here ten text files are used to compute the scores. F-measure, recall and precision are calculated for all the six combinations of the algorithms. The code done here finds all the possible combinations of the four algorithms. Using the combinatorics logic for four algorithms, six combinations are generated. For generating the combinations, 'itertools.combinations' command is used to pair them up. After pairing them, rouge scores are calculated and the results are saved. Finally on the basis of precision, recall and f-measure the two best algorithms are calculated. "Fig. 1" shows the flowchart depicting the steps involved in the summarization of IT LAW legal cases.

VII. RESULTS AND DISCUSSION

In this research basically the focus is to help various lawyers to help in their cases. Apart from reading huge descriptive cases, lawyers have other work too in their life. Hence to ease the burden of reading and spending so much time in the cases , a research is done to save their precious time and to summarize the cases that they deals with. An approach used in this research to extract the IT Legal cases is web scrapping. This is an advanced version of work where the contents from the web pages are directly extracted. In other researches usually people directly apply the algorithms on a given dataset, but here the work is directly applied on the online resources. It helps in accurate extraction of data and analyses the data using various text summarization algorithms. To know which algorithm gives the better efficiency algorithms are compared based on their Rouge scores and their various performance metrics. Finally the comparison is done to predict the best of two algorithms .The technique to help lawyers is a new innovative work to be applied in our country. Hence this is just a step to carry forward the work and help our Indian lawyers. This work focuses on helping our Indian lawyers and to give them the accurate results. Steps in this research basically deals with web scrapping which is done to extract the IT LAW legal cases directly from the website. Then various modules like os, re, pd, nltk and subprocess are imported. Then all the files for summarization are stored in data folder in MyDocuments.

All files are extracted are in the .txt format. Before the summarization process to occur all the data is first cleaned, stemmed and selected. Then summarization process is done using four techniques i.e LUHN, LSA, LEXRANK and SUMBASIC. Then performance measures are calculated for all these techniques on the summarized data. Finally the better of two algorithms are suggested on the basis of precision, recall and f-measure of ROUGE-1 and ROUGE-2 scores. Given below are the screenshots displaying the results of various performance metrics. "Fig. 2" displays the precision values according to Rouge-1 score for the text files for the combinations of different algorithms. "Fig. 3" displays the precision values according to Rouge-2 score for the text files for the combinations of different algorithms. "Fig. 4" displays the recall values according to Rouge-1 score for the text files

for the combinations of different algorithms. “Fig. 5” displays the recall values according to Rouge-2 score for the text files for the combinations of different algorithms. “Fig. 6” displays the f-measure values according to Rouge-1 score for the text files for the combinations of different algorithms and “Fig. 7” displays the f-measure values according to Rouge-2 score for the text files for the combinations of different algorithms.

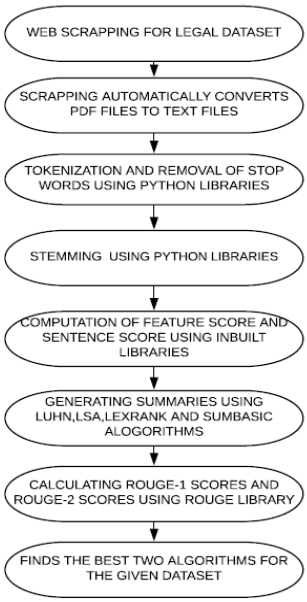

Fig. 1. Flowchart of text summarization system on legal dataset

	SUMBASIC & LUHN	SUMBASIC & LSA	SUMBASIC & LEXRANK	LUHN & LSA	LUHN & LEXRANK	LSA & LEXRANK
textfile_1	0.496598639	0.517006803	0.510204082	0.846560847	0.478835979	0.46361186
textfile_2	0.519230769	0.557692308	0.673076923	0.664658635	0.25502008	0.288288288
textfile_3	0.595959596	0.323232323	0.45959596	0.418886199	0.571428571	0.429657795
textfile_4	0.602209945	0.458563536	0.569060773	0.730769231	0.403846154	0.340909091
textfile_5	0.509036145	0.412650602	0.445783133	0.444940476	0.470238095	0.522260274
textfile_6	0.573863636	0.596590909	0.596590909	0.761020882	0.844547564	0.695175439
textfile_7	0.627692308	0.581538462	0.612307692	0.552922591	0.682464455	0.651933702
textfile_8	0.641891892	0.587837838	0.648648649	0.806267806	0.666666667	0.593984962
textfile_9	0.895604396	0.43956044	0.868131868	0.377300613	0.641104294	0.497797357
textfile_10	0.533227848	0.537974684	0.455696203	0.628486056	0.550796813	0.454949944
AVERAGE	0.599531517	0.50126479	0.583909619	0.623181333	0.556494867	0.493856871

Fig. 2. Screenshot displaying the precision values according to Rouge-1 score for the text files for the combinations of different algorithms

	SUMBASIC & LUHN	SUMBASIC & LSA	SUMBASIC & LEXRANK	LUHN & LSA	LUHN & LEXRANK	LSA & LEXRANK
textfile_1	0.187192118	0.275862069	0.270935961	0.829953198	0.372854914	0.333868379
textfile_2	0.265625	0.2421875	0.4140625	0.594877765	0.115250291	0.140646976
textfile_3	0.351851852	0.133333333	0.303703704	0.314325452	0.514603616	0.287015945
textfile_4	0.358565737	0.262948207	0.410358566	0.699481865	0.328151986	0.236797274
textfile_5	0.311608961	0.150712831	0.254582485	0.314417178	0.405674847	0.473366834
textfile_6	0.397435897	0.401709402	0.444444444	0.692086331	0.78705036	0.637720488
textfile_7	0.416494845	0.391752577	0.408247423	0.454914703	0.597075548	0.544278607
textfile_8	0.38	0.29	0.365	0.690298507	0.574626866	0.531847134
textfile_9	0.859205776	0.25631769	0.812274368	0.260707635	0.592178771	0.347701149
textfile_10	0.287722587	0.304592315	0.277413308	0.530271399	0.491649269	0.323308271
AVERAGE	0.381570277	0.270941592	0.396102276	0.538133403	0.477911647	0.385655106

Fig. 3. Screenshot displaying the precision values according to Rouge-2 score for the text files for the combinations of different algorithms

	SUMBASIC & LUHN	SUMBASIC & LSA	SUMBASIC & LEXRANK	LUHN & LSA	LUHN & LEXRANK	LSA & LEXRANK
textfile_1	0.193121693	0.204851752	0.245098039	0.86253369	0.591503268	0.562091503
textfile_2	0.108433735	0.130630631	0.243055556	0.7454955	0.440972222	0.444444444
textfile_3	0.285714286	0.243346008	0.303333333	0.65779468	0.786666667	0.376666667
textfile_4	0.299450549	0.235795455	0.381481481	0.75568182	0.544444444	0.444444444
textfile_5	0.251488095	0.234589041	0.340229885	0.5119863	0.726436782	0.701149425
textfile_6	0.234338747	0.230263158	0.246478873	0.71929825	0.854460094	0.744131455
textfile_7	0.322274882	0.348066298	0.350970018	0.64456722	0.761904762	0.624338624
textfile_8	0.270655271	0.218045113	0.309677419	0.70927318	0.75483871	0.764516129
textfile_9	0.5	0.352422907	0.600760456	0.54185022	0.794676806	0.429657795
textfile_10	0.335657371	0.378197998	0.388140162	0.70189099	0.745283019	0.551212938
AVERAGE	0.280113463	0.257620836	0.340922522	0.68503718	0.700118677	0.564265343

Fig. 4. Screenshot displaying the recall values according to Rouge-1 score for the text files for the combinations of different algorithms

	SUMBASIC & LUHN	SUMBASIC & LSA	SUMBASIC & LEXRANK	LUHN & LSA	LUHN & LEXRANK	LSA & LEXRANK
textfile_1	0.059282371	0.08988764	0.105769231	0.853932584	0.459615385	0.4
textfile_2	0.039580908	0.043600563	0.116483516	0.718706048	0.217582418	0.21978022
textfile_3	0.132127955	0.082004556	0.147217235	0.514806378	0.66427289	0.226211849
textfile_4	0.155440415	0.112436116	0.234090909	0.689948893	0.431818182	0.315909091
textfile_5	0.117331288	0.074371859	0.15842839	0.412060302	0.670468948	0.596958175
textfile_6	0.13381295	0.127544098	0.149210904	0.652645862	0.784791966	0.674318508
textfile_7	0.164094232	0.189054726	0.182488479	0.55721393	0.677419355	0.504147465
textfile_8	0.141791045	0.092356688	0.146	0.589171975	0.616	0.668
textfile_9	0.44320298	0.204022989	0.531914894	0.402298851	0.75177305	0.286052009
textfile_10	0.160229645	0.187969925	0.194480946	0.587622903	0.61892247	0.367279895
AVERAGE	0.154689379	0.120324916	0.19660845	0.597840772	0.589266466	0.425865721

Fig. 5. Screenshot displaying the recall values according to Rouge-1 score for the text files for the combinations of different algorithms

	SUMBASIC & LUHN	SUMBASIC & LSA	SUMBASIC & LEXRANK	LUHN & LSA	LUHN & LEXRANK	LSA & LEXRANK
textfile_1	0.059282371	0.08988764	0.105769231	0.853932584	0.459615385	0.4
textfile_2	0.039580908	0.043600563	0.116483516	0.718706048	0.217582418	0.21978022
textfile_3	0.132127955	0.082004556	0.147217235	0.514806378	0.66427289	0.226211849
textfile_4	0.155440415	0.112436116	0.234090909	0.689948893	0.431818182	0.315909091
textfile_5	0.117331288	0.074371859	0.15842839	0.412060302	0.670468948	0.596958175
textfile_6	0.13381295	0.127544098	0.149210904	0.652645862	0.784791966	0.674318508
textfile_7	0.164094232	0.189054726	0.182488479	0.55721393	0.677419355	0.504147465
textfile_8	0.141791045	0.092356688	0.146	0.589171975	0.616	0.668
textfile_9	0.44320298	0.204022989	0.531914894	0.402298851	0.75177305	0.286052009
textfile_10	0.160229645	0.187969925	0.194480946	0.587622903	0.61892247	0.367279895
AVERAGE	0.154689379	0.120324916	0.19660845	0.597840772	0.589266466	0.425865721

Fig. 6. Screenshot displaying the recall values according to Rouge-2 score for the text files for the combinations of different algorithms

	SUMBASIC & LUHN	SUMBASIC & LSA	SUMBASIC & LEXRANK	LUHN & LSA	LUHN & LEXRANK	LSA & LEXRANK
textfile_1	0.278095234	0.293436289	0.331125823	0.854472625	0.529239761	0.508124072
textfile_2	0.17940199	0.211678829	0.357142853	0.70276008	0.323155212	0.349726771
textfile_3	0.386252041	0.277657262	0.365461843	0.511834315	0.66199158	0.401420954
textfile_4	0.399999996	0.311444648	0.456762745	0.743016755	0.463722393	0.385852085
textfile_5	0.336653382	0.299126633	0.385919161	0.476114645	0.570912371	0.598626099
textfile_6	0.33278418	0.332278477	0.348837205	0.739571585	0.849474907	0.718820857
textfile_7	0.425887261	0.436483866	0.446188336	0.59523809	0.719999995	0.637837833
textfile_8	0.380761519	0.318098716	0.419213969	0.754666662	0.708018149	0.668547245
textfile_9	0.641732279	0.391198039	0.710112355	0.444846288	0.709677414	0.461224485
textfile_10	0.411980435	0.444154143	0.419213969	0.663163421	0.633447876	0.498476534
AVERAGE	0.377354832	0.33145569	0.423997826	0.648568447	0.616963966	0.522865693

Fig 7. Screenshot displaying the f-measure values according to Rouge-1 score for the text files for the combinations of different algorithms

	SUMBASIC & LUHN	SUMBASIC & LSA	SUMBASIC & LEXRANK	LUHN & LSA	LUHN & LEXRANK	LSA & LEXRANK
textfile_1	0.09004739	0.135593217	0.152143841	0.841772147	0.411714036	0.363954501
textfile_2	0.068895641	0.073897494	0.181818178	0.650956409	0.150684927	0.171526582
textfile_3	0.192113242	0.101551476	0.19830713	0.390328147	0.579937299	0.253012043
textfile_4	0.216867466	0.157517896	0.298118664	0.694682671	0.372914617	0.270691329
textfile_5	0.170473534	0.099596227	0.195312495	0.356676811	0.505494501	0.528026901
textfile_6	0.200215281	0.193614826	0.223415678	0.671787705	0.785919535	0.655509061
textfile_7	0.235431231	0.255033553	0.252229295	0.500894449	0.634715021	0.523444971
textfile_8	0.206521735	0.140096615	0.208571424	0.635738827	0.59459459	0.592198577
textfile_9	0.58476658	0.227199995	0.642857138	0.316384176	0.662499995	0.313878075
textfile_10	0.205833049	0.23247496	0.228659709	0.55747599	0.547993014	0.343894181
AVERAGE	0.217116515	0.161657626	0.258143355	0.561669633	0.524646753	0.401613622

Fig. 8. Screenshot displaying the f-measure values according to Rouge-2 score for the text files for the combinations of different algorithms

In the work, Rouge-1 measures for the legal dataset are as follows:

A. For Precision

LUHN and LSA with maximum value of 0.623 are considered best.

B. For Recall

LUHN and LEXRANK with maximum value of 0.700 are considered best.

C. For F-measure

LUHN and LSA with maximum value of 0.648 are considered best.

In the work, Rouge-2 measures for the legal dataset are as follows:

A. *For Precision*

LUHN and LSA with maximum value of 0.538 are considered best

B. For Recall

LUHN and LSA with maximum value of 0.597 are considered best.

C. For F-measure

LUHN and LSA with maximum value of 0.561 are considered best.

TABLE I and TABLE II below shows the techniques used and their corresponding ROUGE-1 and ROUGE-2 scores.

TABLE I. ROUGE-1 measures showing best two techniques for the legal dataset

	TECHNIQUES	VALUES
PRECISION	LUHN & LSA	0.623
RECALL	LUHN and LEXRANK	0.700
F-MEASURE	LUHN & LSA	0.648

TABLE II. ROUGE-2 measures showing best two techniques for the legal dataset

	TECHNIQUES	VALUES
PRECISION	LUHN and LSA	0.538
RECALL	LUHN and LSA	0.597
F-MEASURE	LUHN and LSA	0.561

VIII. FUTURE SCOPE

Text summarization is an advance field in terms of text mining. Day by day new technologies are being developed so as to ease the process of summarization. But today in the world of internet, the existing technologies are not helping the user in an appropriate and efficient manner. There are still other and new chances of developing the technologies so as to lessen the human effort in terms of summarization. Some newer approaches can be developed which can help in summarizing the online videos and meetings into the text form in the required format. For those who are not able to listen, for them new software can be developed which can directly convert the given data either in the form of text or in the form of videos or audios, summarized and then be converted again into the audio format. Also new visualization techniques can be developed which can help in visualizing the results either in form of graphs, tables or by statistics. This helps in reducing the time of the user. New ways can be developed to increase the summary quality using summarization techniques. Text summarization is about fifty two years old technique, but today also the summaries produced by the techniques are not so far perfect. Thus newer methodologies should be developed or combined with the existing technologies so to produce high quality summaries.

IX. CONCLUSION

Text summarization is an interesting research field and it has wide range of applications. This research basically deals with the extraction of online cases using web scrapping technique. It is used to give the accurate results and to analyse various algorithms and their performance metrics onto the data. This is done so as to help various lawyers to help gain about the information in an abstract manner for the large documents. Based on the results of summarization techniques the comparison between them is done on the basis of various performance metrics and the better of two algorithms out of them are found.

REFERENCES

[1] Shandilya, Anurag, Kripabandhu Ghosh, and Saptarshi Ghosh. "Fairness of Extractive Text Summarization." Companion of the the Web Conference 2018 on the Web Conference 2018. International World Wide Web Conferences Steering Committee, 2018.

[2] Käfer, Verena. "Summarizing software engineering communication artifacts from different sources." Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering. ACM, 2017.

[3] Fang H, Lu W, Wu F et al (2015) Topic aspect-oriented summarization via group selection. Neurocomputing 149:1613–1619. doi:10.1016/j.neucom.2014.08.031.

[4] Heu JU, Qasim I, Lee DH (2015) FoDoSu: multi-document summarization exploiting semantic analysis based on social Folksonomy. Inf Process Manag 51(1):212–225.

[5] Li C, Liu Y, Zhao L (2015a) Using external resources and joint learning for bigram weighting in ilp-based multi-document summarization. In: Proceedings of NAACL-HLT, pp 778–787.

[6] Li P, Bing L, Lam W, Li H, Liao Y (2015b) Reader-aware multi-document summarization via sparse coding.arXiv preprint arXiv: 1504.07324.

[7] Mendoza M, Bonilla S, Noguera C et al (2014) Extractive single-document summarization based on genetic operators and guided local search. Expert Syst Appl 41:4158–4169. doi:10.1016/j.eswa.2013.12.042.

[8] Fattah MA, Ren F (2009) GA, MR, FFNN, PNN and GMM based models for automatic text summarization.Comput Speech Lang 23:126–144. doi:10.1016/j.csl.2008.04.002.

[9] Zajic DM, Dorr BJ, Lin J (2008) Single-document and multi-document summarization techniques for e-mail threads using sentence compression. Inf Process Manag 44:1600–1610.

[10] Sarkar K (2010) Syntactic trimming of extracted sentences for improving extractive multi-document summarization. J Comput 2:177–184.

[11]Carbonell JG, Goldstein J (1998) The use of MMR, diversity-based re-ranking for re-ordering documents and producing summaries. In: Proceedings of the 21st annual international ACM SIGIR conference on research and development in information retrieval, pp 335–336.

[12]K. Agrawal and H. Mohan, "Cardiotocography Analysis for Fetal State Classification Using Machine Learning Algorithms," *2019 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, Tamil Nadu, India, 2019, pp. 1-6.

[13]K. Agrawal and H. Mohan, "Text Analysis: Techniques, Applications and Challenges," presented in *2019 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, Tamil Nadu, India, 2019.