

A Survey on Implementation of Presentation Slide from Articles

Anuja A. Bhandare¹, Mr. Chetan J. Awati², Sonam S. Kharade³

¹Post Graduate Student, Department Of Technology, Shivaji University, Kolhapur, Maharashtra, India

²Assistant Professor, Department Of Technology, Shivaji University, Kolhapur, Maharashtra, India

³Assistant Professor, Department Of Technology, Shivaji University, Kolhapur, Maharashtra, India

Abstract – The automatic generation of presentation slides from specialized articles is a standout amongst the most wanted however under-looked into range in the field of processing. Mechanized generation of slide substance from specialized articles is much troublesome than a run of the mill content synopsis process, since it requires the ID of all the urgent substance from the article and their game plan in a precise way, subsequently making it a non-paltry assignment. The procedure is thought to be one of the center applications of content mining. Automatic slide generators can be comprehensively characterized in view of NLP, Statistical Methods, and Machine Learning. A point to point survey of the absolute most essential automatic slide generation procedures from scholastic articles is exhibited. What's more, a brief correlation among the talked about strategies is given.

Key Words: text mining; support vector regression (SVR); Integer linear programming (ILP); Natural Language Processing (NLP).

1. INTRODUCTION

Presentation slides have been broadly utilized over quite a while for the theoretical transport of thoughts. It helps clients to exchange their thoughts and hypotheses viably. Analysts introduce their speculation with the assistance of presentation slides and henceforth, presentations have turned out to be to a great degree useful in gatherings and meetings. Devices like Microsoft PowerPoint, Open Office, Libre and so forth help in building and arranging the slides as indicated by one's sufficient needs. The creators themselves need to make the slides from scratch, which is a monotonous and tedious technique. The self-generation of slides is past the abilities of the above virtual products, prompting the need for devices that does programmed production of slides from specialized articles, which is to a great degree troublesome and nontrivial. Since robotization in all angles has turned into a key slant today, the current computational abilities weight on complete substitution of the manual methodology of slide generation. Research articles have a pretty much comparable structure. All the specialized articles have areas like abstract, introduction, related work and so forth. The principle methodology behind a mechanized slide generator is to

adventure this likeness and subsequently, outline the area to one or more slides in the presentation. For this reason, distinctive formats might be utilized by the generator. One ordinarily utilized format a structure is to keep up the consistent structure of the exploration article what's more, offer an outline for every segment in the yield slides.

To produce the slide substance, mechanized slide generators depend vigorously on content summarizers. They can be ordered in light of their working methodologies. The outline technique of a generator can work taking into account measurements, which may utilize the TF-IDF scoring technique, Natural Language Processing, which examinations the talk structure of the article, Machine Learning which relies on upon area information procured through broad preparing. Multi-archive extractive summarizers receive an eager system to assemble synopses from related reports by selecting sentences having high pertinence components. Then again, abstractive summarizers assemble synopses by either shortening or revamping the chose sentences. It is conceivable to produce the synopses that might be sorted out on to the slides utilizing the reference data from examination articles. [1], [2] are outline plots that adventure the reference data from insightful articles. The motivation behind utilizing the reference data is to accumulate critical ideas, which can be utilized to extricate the applicable sentences from the article. [3], [4] examined plans wherein the outlines were created by separating components from various related articles. It improves the differing qualities of the created outlines. A Query particular summarizer [5] includes the outline of an article concerning the watchwords determined by a client. Other outline plans include the utilization of Shrouded Markov Models [6] and Conditional Random Fields [7] for record outline. Since the inclusion of images and tables to the slides offer an enhanced understandability to the viewers, automated slide generators should be capable of adding graphical elements also to the output slides. The efficiency of a slide generator can be measured as the degree to which all the important information has been map perform the article to the output slides.

The slides that are generated must have the following **traits**:

- All the vital ideas must be masterminded in an organized furthermore, precise way and must be fair.
- The association from the client part for slide era must be insignificant.
- The slides must have great substance quality, differences and adequate length when contrasted with its comparing research article.

The slides must be editable freely so that the creator can change the substance to suit the client's needs.

2. REVIEW OF AUTOMATIC SLIDE GENERATION TECHNIQUES

Most slide generators require the distinguishing proof of critical points also, sentences from the article. A wide assortment of methods has been utilized to recognize the applicable parts that shape the substance of an article. A survey of the methods for slide era is exhibited to toss light on how the strategies recognize and compose the critical parts of the specialized record to shape a presentation slide.

2.1 Era of Slides Based on Inference of Basic Semantic Structure of Articles

The surmising of semantic structure of the article helps in comprehension the relations between sentences, recognizable proof of imperative themes and co-references effectively. **Utiyama et al.**, [8] built up a framework that could naturally create presentation slides from an article and powerfully modify the presentation in light of the inquiries issued by the group of onlookers. The articles were at first commented on by the GDA Tagset. Utilizing the comment, the framework extricated the basic semantic structure also, connections in the article. The framework utilized the information of semantic conditions and co-references to extricate out the subject parts from the article. From the rundown of subjects, the framework sifted through the most applicable subjects, which later turns into the underlying presentation. The framework arranged a slide for every subject utilizing the condensing sentences removed from the article. In light of the questions, the slides were overhauled on the fly and if required, the framework alluded the article again on the off chance that the record contained extra data pertinent to the current inquiry.

Despite the fact that the framework was dialect/area/style autonomous, it couldn't include graphical components to the yield slides. In addition, viable point recognizable proof what's more, crowd connection plans were not proficiently made utilization of in the framework.

2.1.2. Generation of Slides from the LATEX Manuscript of an Article

LATEX composition of an exploration article offers high auxiliary data about the record. It offers a simple extraction office of content and graphical substance from the record. The LATEX source of an article is considered as the beginning stage for slide era. **Y. Yasumura et al.**, [9] actualized an answer for the issue of naturally creating presentation slides from specialized articles. The framework required a client to enter the LATEX composition of the specialized archive. The strategy included the count of weights of all the terms in the record utilizing TF-IDF scoring strategy. The term weights served as a procedure to decide a importance score for all the items in the report. The measure of the outline to be extricated for every area could be dictated by utilizing the term weights. The arrangement utilized the utilization of slide formats in order to create presentation slides. The yield slides could be tweaked by the client. The execution required the client to indicate catchphrases to the framework. **Sravanthi et al.**, [10] contributed another answer for self-create presentation slides of a specialized archive. The procedure started with the deduction of the nitty-gritty coherent structure of the article from the LATEX original copy of the record. Every segment was arranged to fall under Introduction, Related Work, Model, Experiments on the other hand Conclusion. The framework extricated critical key expressions and the QueSTS Summarizer [11] was utilized to condense the Model, Tests and the Conclusion segments. The QueSTS summarizer considered the segment to be condensed as an incorporated chart comprising of vertices speaking to sentences. Hub weights what's more, edge weights were computed and were utilized for picking the synopsis sentences comparing to every segment. The framework separated the graphical components from the LATEX original copy and annexed them to the yield slides as required. It is obvious that the proficiency of the framework could accomplish a much more elevated amount by fusing NLP based procedures to build the nature of the produced presentation slides.

2.1.3. Generation of Slides Using Natural Language Processing

Normal Language Processing, one of the spearheading ranges of calculation, concentrates on the investigation of the talk structure, connections between content units, syntax investigation and so forth. NLP based summarizers [12] are equipped for examining the basic talk structure of the reports and to utilize this learning to compress the substance proficiently. Slide generators can make the slides

by utilizing the separated substance. Examined beneath are systems that produce slides taking into account NLP methodologies. **Shibata et al.**, [13] depicted a technique to produce presentation slides from content by investigating the talk structure of the article. The framework considered a provision and a sentence as a talk unit furthermore, intelligence relations, for example, contrast, list, added substance, elaboration and so forth were extricated and broke down. The framework extricated point and non-theme parts from the article in light of the fundamental talk structure. The separated subject and non-point parts were put on the slides by giving legitimate indents in light of the examination of their syntactic structure. The framework assembled the slides by interfacing applicable sentences to the most comparative going before sentences. The framework likewise pruned the non-subject parts in light of some heuristic measures to give an improved lucidness.

K. Gokul Prasad et al., [14] actualized a framework that was engaged on the instructive space to make presentation slide for courses also, addresses. The framework took a shot at the premise of 2 modules - Information Extractor and Slide Generator individually and included the introductory extraction of content substance from the information. The framework utilized the utilization of center NLP operations like content division and lumping to distinguish sections and additionally thing phrases. The execution assigned fragments and their segment sentences with weightage values. Slides were based on the premise of expressions having high pertinence elements. A metaphysics tree was worked for everything phrase identified utilizing a chunker. The framework utilized the metaphysics tree to deduce the semantic relations between sentences. Taking into account the produced philosophy and weightage esteem, the framework distinguished the imperative key expressions, which were utilized for visual cue ID lastly, the development of presentations. The framework's exactness could be helped by utilizing an area particular philosophy.

2.1.4 Slide generation using Text Summarization

Content rundown can be characterized as the extraction of the most important delegate parts of a content piece. Summarizers utilizes techniques like positioning, irregular choice, area based selection, HMM and so on to choose sentences to be incorporated into the synopsis. **Tulasi Prasad Sariki et al.**, [15] introduced a novel plan to outline a record and henceforth utilize the condensed form to fabricate presentation slides. The framework at first got and made the archive to be outlined and connected various

fundamental content preprocessing procedures, for example, sentence division; case collapsing, stop word evacuation, stemming and lemmatization. The framework then utilized a scoring plan which was a blend of a portion of the standard scoring philosophies, for example, prompt expression, key, title and area based scoring techniques to distribute an importance score for every sentence. Utilizing a sentence positioning plan, the rundown was made and from this rundown, presentation slides were produced. The framework was equipped for making rundowns particular to inquiry catchphrases and thus build a question particular presentation.

2.1.5. Slide generation using web mining

The web can be seen as a colossal store containing data which can be mined helpfully to suit the necessities. Web mining as a rule includes the bringing of pages containing the required data, extraction of information from them lastly, the utilization of mining systems. A couple of strategies that undertaking web mining to recover theme particular data and their association as presentation slides are examined.

ShaikhMostha Al Masum et al., [16] explained another specialist based plan to assemble presentation slides by mining question based data from the web. The information was accumulated from Wikipedia alternately by utilizing famous web crawlers like Google, Yahoo and Alta Vista in light of the accessibility of required data. The framework added pictures additionally to the yield slides for giving better clarity and understanding on the subject. The strategy fabricated the presentation information utilizing a mix of strategies like web information getting, site page parsing and outline extraction, each of which were performed by operators. The determination of a presentation layout was done taking into account the decision of the data vault. The proposed plan utilized the factual technique to discover a significance score for every sentence amid synopsis. The technique additionally utilized devoted calculations for site page parsing and presentation era. The framework made MPML scripts lastly produced slides in HTML what's more, Java script and the points were disclosed to various headings by specialist characters. The calculations for mining should have been made strides in order to build the general execution of the framework.

Mistsuru Ishizuka et al., [17] dealt with another plan that created a compact report and a presentation by mining the web assets in light of the question issued by a client.

Every progression in the creation of slides was finished by programming specialists. The framework utilized 6 distinctive programming operators. Despite the fact that the above strategy is a comparable method, it didn't produce a report relating to the issued question. As an underlying stride, the info questions were preprocessed by the framework. Ambiguities if the present were evacuated by adding to the theme, its top disambiguated faculties. In view of the inquiry subject the framework utilized framework operators to utilize methods of web slithering, what's more, information extraction to download the pages and subsequently extricate the headings and in addition the content substance from the downloaded papers. From the extricated information, compact synopses were created utilizing an outline unit. For a synopsis, the closeness between writings was broke down by the estimation of a vector separation. For every subject, a report was produced and from every report, the proposed framework produced the presentation scenes. The scenes are at that point changed over to MPML scripts lastly to a Java script code based presentation. The drawback of the framework was that it could not handle numerous client collaborations and was delicate to higher loads.

3. COMPARISON OF TECHNIQUES

The distinctions in the working procedures of all the talked about strategies should be seen adequately to pick up a knowledge on to how slides can be naturally created from exploration articles.

Table -1: Comparison of automatic slide generation techniques

Basis of Procedure	Type
Surmising of semantic conditions in the article, recognizable proof of points, and element customization of presentation.	Annotation based
Count of weights for the terms in the report, rundown of every area and era of slides.	Statistical Method
Extraction of contents from LATEX Manuscript, categorization	Statistical Method
Analysis of discourse structure, detection of topic/non-topic parts, slide generation based on intending, pruning of non-topic parts	NLP, Discourse Analysis
Gathering theme particular information from the web, compress them and produce a question particular presentation.	Statistical Method, Agent based
Gathering subject particular information from the web, abridge them and create a question particular report and presentation	Agent based

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Experimental Results of Semantic Structure Inference Based Method

Even though the framework could progressively adjust the presentations taking into account questions, the trial results demonstrated that:

- Heuristic measures received by the framework could fizzle in a few situations.

4.2 Experimental Results of Methods using LATEX Manuscript of an article

The framework was given 8 articles and the comparing writers were chosen for rating the framework in light of different quality parameters. The outcomes set forward the accompanying focuses:

- Coherence of the slides was great.
- Coverage of the slide was extraordinary.
- Better quality would be accomplished if sentences were compacted

4.3 Experimental Results of NLP Based Methods

The assessment consequences of [13] uncovered the accompanying territories:

- Errors could happen because of word-chain unrecognizable proof
- Recognition mistakes of difference relations amongst provisions and sentences were experienced.
- The surmising from the outcomes was to change unique writings to Multimodal presentations to upgrade the presentation quality.

The test investigation of [14] demonstrated the accompanying results.

- Assessment parameters of Precision, Recall and F1-Score were utilized for this reason.
- The framework had higher efficiencies on working with nontechnical records.

4.4 Experimental Results of Text Summarization Based Method

The framework was contrasted and benchmark outline plans what's more, the effectiveness of the plan can be envisioned in Figure 2 as follows:

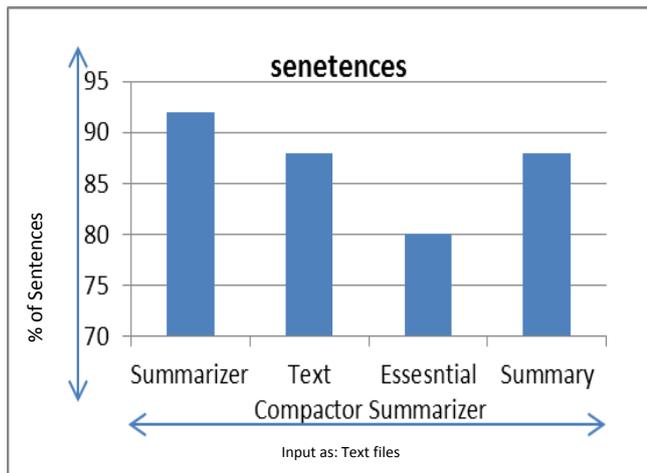


Chart -1: Efficiency comparison of Summarization based slide generation technique

5. CONCLUSIONS

Programmed era of presentation slides from an article includes the distinguishing proof and association of all the significant data in a requested and orderly grammar. The examination of some of the late patterns in the field of programmed slide era is finished that originates from greenhorn methodology to the techniques that utilize the current capacities of calculation. A brief examination of the talked about procedures is likewise made and all the key methods for programmed era of slides have been overviewed. It is apparent from the study that there is sufficient opportunity to get better in each of the talked about strategies. It would be more productive if mixes of procedures are utilized to manufacture the yield slides. The objective of the slide era procedure ought to be to take in the complexities behind how scientists manufacture slides from their articles, and utilize this information to produce slides that are much easy to understand and very adaptable.

ACKNOWLEDGEMENT

I express my sincere thanks to Chetan J. Awati whose supervision, inspiration and valuable guidance helped me a lot to complete my review paper. His guidance proved to be the most valuable to overcome all the hurdles in the fulfillment of this paper.

REFERENCES

[1] V. Qazvinian, D. R. Radev, S. M. Mohammad, B. J. Dorr, D. M. Zajic, M. Whidby, and T. Moon, "Generating

extractive summaries of scientific paradigms", *J. Artif. Intell. Res.*, vol.46, pp. 165- 201, 2013

[2] V. Qazvinian and D. R. Radev, "Scientific paper summarization using citation summary networks", in *Proc. 22nd Int. Conf. Comput. Linguistics-Volume 1*, Aug. 2008, pp. 689- 696.

[3] N. Agarwal, K. Gvr, R. S. Reddy, and C. P. Rose, "Towards multidocument summarization of scientific articles: Making interesting comparisons with SciSumm", in *Proc. Workshop Autom. Summarization Different Genres, Media, Lang.*, 2011, pp. 8-15

[4] O. Yeloglu, M. Evangelos, and Z.-H. Nur, "Multi-document summarization of scientific corpora", in *Proc. ACM Symp. Appl. Comput.*, 2011, pp. 252-258

[5] R. Jha, A. Abu-Jbara, and D. Radev, "A system for summarizing scientific topics starting from keywords", *ACM Comput. Surv*, vol. 40, no. 3, p. 8, 2013

[6] M. J. Conroy and D. P. O'leary, "Text summarization via hidden Markov models", in *Proc. 24th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2001, pp. 406-407.

[7] D. Shen, J. T. Sun, H. Li, Q. Yang, and Z. Chen, "Document summarization using conditional random fields", in *Proc. 20th Int. Joint Conf. Artif. Intell.*, 2007, vol. 7, pp. 2862-2867

[8] M. Utiyama and K. Hasida, "Automatic slide presentation from semantically annotated documents", in *Proc. ACL Workshop Conf. Its Appl.*, 1999, pp. 25-30.

[9] Y. Yasumura, M. Takeichi, and K. Nitta, "A support system for making presentation slides", *Trans. Japanese Soc. Artif. Intell.*, vol. 18, pp. 212-220, 2003.

[10] M. Sravanthi, C. R. Chowdary, and P. S. Kumar, "SlidesGen:Automatic generation of presentation slides for a technical paper using summarization", in *Proc. 22nd Int. FLAIRS Conf.*, 2009, pp. 284-289.

[11] M. Sravanthi, C. R. Chowdary, and P. S. Kumar, "QueSTS: A query specific text summarization approach", in *Proc. 21st Int. FLAIRS Conf.*, 2008, pp. 219-224

[12] D. Marcu, "From discourse structures to text summaries", in *Proc. ACL Workshop Intell. Scalable Text Summarization.*, 1997, vol. 97, pp. 82-88.

[13] T. Shibata and S. Kurohashi, "Automatic slide generation based on discourse structure analysis", in *Proc. Int. Joint Conf. Natural Lang. Process.*, 2005, pp. 754-766.

[14] Gokul Prasad, K., Mathivanan, H., Jayaprakasam, M., and Geetha, T. V., "Document summarization and information extraction for generation of presentation slides", *Advances in Recent Technologies in Communication and Computing*, 2009. ARTCom'09. International Conference on. IEEE, 2009.

[15] Sariki, Tulasi Prasad, Bharadwaja Kumar, and Ramesh Ragala. "Effective Classroom Presentation Generation Using Text Summarization".

[16] S. M. A. Masum, M. Ishizuka, and M. T. Islam, "Autopresentation: A multi-agent system for building automatic multi-modal presentation of a topic from world wide web information", in *Proc. IEEE/WIC/ACMInt. Conf. Intell. Agent Technol.*, 2005, pp. 246-249

- [17] S. M. A. Masum and M. Ishizuka, "Making topic specific report and multimodal presentation automatically by mining the web resources", in Proc. IEEE/WIC/ACM Int. Conf. Web Intell., 2006, pp. 240-246
- [18] Hu, Yue, and Xiaojun Wan. "Ppsgen: learning to generate presentation slides for academic papers", Proceedings of the Twenty-Third international joint conference on Artificial Intelligence. AAAI Press, 2013
- [19] C. C. Chang and C. J. Lin. (2001), LIBSVM: A library for support vector machines, [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [20] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web," Stanford Digital Libraries, Stanford, CA, USA, Tech. Report: SIDLWP- 1999-0120, 1999.
- [21] D. Radev, T. Allison, S. Blair-Goldensohn, J. Blitzer, A. Celebi, S. Dimitrov, E. Drabek, A. Hakim, W. Lam, D. Liu, J. Otterbacher, H. Qi, H. Saggion, S. Teufel, M. Topper, A. Winkel, and Z. Zhang, "MEAD - A platform for multidocument multilingual text summarization," in Proc. 4th Int. Conf. Lang. Resources Eval., 2004, pp. 14.
- [22] C. Y. Lin "ROUGE: A package for automatic evaluation of summaries" in Proc. Workshop Text Summarization Branches Out, Post- Conf. Workshop ACL, 2004, pp. 2526.