International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 01 | Jan-2016 www.irjet.net p-ISSN: 2395-0072

SLIDEGen: Approach to automatic Slides Generation

Miss. Autade Dhanshri P [1], Prof. Raut S.Y[2]

Student Of Computer Engineering, Pravara Rural Engg. College Of Loni, India[1]

Asst Professor Of Computer Engineering, Pravara Rural Engg. College Of Loni , India[2]

Abstract :-In this paper, we explore an exceptionally difficult assignment of consequently creating presentation slides for scholarly papers. The created presentation slides can be utilized as drafts to offer the moderators some assistance with preparing their formal slides quickerly. A novel framework called PPSGen is proposed to address this undertaking. It first utilizes the relapse strategy to take in the significance scores of the sentences in a scholarly paper, and afterward misuses the whole number straight programming (ILP) system to produce very much organized slides by selecting and adjusting key expressions and sentences. Assessment results on a test set of 200 sets of papers and slides gathered on the web show that our proposed PPSGen framework can produce slides with better quality. A client study is likewise shown to demonstrate that PPSGen has a couple of apparent focal points over gauge systems.

Index Terms:-Abstracting methods, text mining

1.1 INTRODUCTION

PRESENTATION slides have been a mainstream and powerful intends to display and exchange data, particularly in scholarly gatherings. The analysts dependably make utilization of slides to show their work pictorially on the gatherings. There are numerous virtual products, for example, Microsoft Power- Indicate and OpenOffice offer specialists some assistance with preparing their slides. Then again, these devices just help them in the arranging of the slides, however not in the substance. Despite everything it takes moderators much time to compose the slides without any preparation. In this work, we propose a technique for naturally producing presentation slides for scholarly papers. We intend to consequently produce all around organized slides and give such draft slides as a premise to lessen the moderators' opportunity and exertion while setting up their last presentation slides.

Scholarly papers dependably have a comparable structure. They by and large contain a few segments like conceptual, presentation, related work, proposed technique, tests and conclusions. Despite the fact that presentation slides can be composed in different routes by distinctive moderators, a moderator, particularly a tender foot, dependably adjusts slides consecutively to the paper areas while setting up the slides. Every segment is adjusted to one or more slides and one slide for the most part has a title and a few sentences. These sentences may be incorporated into some visual cues. Our system endeavors to create draft slides of the run of the mill sort said above and people groups to set up their last slides.

Programmed slides era for scholarly papers is a extremely difficult assignment.

Current routines for the most part concentrate objects like sentences from the paper to develop the slides. Rather than the short outline removed by a rundown framework, the slides are required to be much more organized and any longer. Slides can be separated into a requested succession of parts. Every part addresses a particular theme and these subjects are additionally pertinent to each other. As a rule, programmed slide era is a great deal more troublesome than synopsis. Slides typically have content components as well as chart components such as figures and tables. Be that as it may, our work concentrates on the content components just.

In this study, we propose a novel framework called PPSGen to create all around organized presentation slides for scholastic papers. In our framework, the significance of every sentence in a paper is found out by utilizing the bolster vector relapse (SVR) model with various helpful components, and after that the presentation slides for the paper are created by utilizing the whole number direct programming (ILP) model with intricately outlined target capacity and limitations to choose and adjust key expressions and sentences.

Investigates a test set of 200 paper-slides sets demonstrate our strategy can create slides with preferable quality over the standard strategies. Utilizing the ROUGE toolbox and the pyramid assessment, the slides created by our strategy can improve ROUGE scores and pyramid scores. Additionally, in light of a client study, our slides can get higher rating scores by human judges in both substance and structure angles. In this way, our slides are viewed as a superior premise for setting up the last slides.

2.1 RELATED WORK

2.1.1 Slides Generation

Programmed slides era for scholarly papers stays far under-researched these days. Few concentrates specifically examine on the subject of programmed slides era. Utiyama and Hasida [2] endeavored to naturally create slides from information reports explained with the GDA tagset.1 GDA labeling can be utilized to encode semantic structure. The semantic relations incorporate linguistic relations, for example, subject, topical relations, for example, operators, quiet, and expository relations for example, reason and elaboration. They first identify subjects in the data records and after that concentrate essential sentences significant to the subjects to create slides

Yasumura et al. [3] presented an emotionally supportive network for making slides from specialized papers. The inputs of the framework are scholarly papers in LATEX position. The framework computes the weights of the terms in the paper utilizing TF-IDF scores. Utilizing the term weights, objects in the paper like sentences, tables and so forth are likewise weighted. In light of the weights of the articles, the framework chooses the quantity of the items like sentences to be separated for every area in the paper and after that produce the slides utilizing a slide arrangement format which can be altered by the clients.

2.2.2 Scientific Article Summarization

[4]The objective of exploratory article synopsis is to produce a short outline for a given experimental article or article set. Early works including attempted to utilize different elements particular to exploratory content (e.g., explanatory pieces of information components).

[5] Agarwal et al. presented an unsupervised methodology to the issue of multi-record investigative article rundown. The data is a rundown of papers refered to together inside the same source article. The key purpose of this methodology is a subject based grouping of parts separated from each cocited article.

2.2.3 Document Summarization

[7]Woodsend and Lapata additionally received routines based on ILP to concentrate synopsis. The article capacity of the ILP model consolidates the significance of the bigrams in the synopsis' sentences, the striking nature of the parse tree hubs of the synopsis' sentences and a unigram dialect model which punishes sentences containing words that are likely to show up in outlines.

The above methodologies all arrangement with the undertakings of conventional rundown and exploratory article synopsis. On the other hand, slides era is very different from customary rundown and exploratory synopsis. They just select a few sentences from the records, while slides era is a great deal more confused. Our proposed approach not just chooses various vital sentences be that as it may, likewise the expressions relating to the sentences. After the choice of sentences and expresses, we can develop all around organized slides.

3.1 PROBLEM DEFINITION AND CORPUS

3.1.1 Problem Definition

In our work, we plan to naturally produce presentation slides for scholastic papers. We have to create very much organized slides as the draft slides for a moderator to set up the last slides. There are different sorts of slides which are made by Microsoft PowerPoint and OpenOffice. They can be much diverse in styles and we clearly can't consider all sorts of styles. So before presenting our technique, we require to address the style of slides we produce.

A novice more often than not gets ready slides which are consecutively adjusted to the paper. One area in the paper is by and large adjusted to one or more slides. One slide more often than not incorporates a few visual cues and sentences that clarify the relating visual cues. It is sensible to utilize that style of slides that novices dependably use to make draft slides and we respect it all around organized in light of the fact that it employments sets of visual cues and sentences to address imperative focuses and makes it simple for the peruser to handle the focuses. From Fig. 1, we can have a look at the style of the slides we produce. Here, key expressions "Operators Behaviors" furthermore, "Authorization Approach" are set as the projectile focuses. The sentences significant to the key expressions are set beneath the relating visual cues.

In this work, we just consider the content components in the paper. Different components, for example, tables and figures are definitely not incorporated into the created slides. Despite the fact that tables and figures are valuable in the

slides, we disregard them to improve the issue what's more, better concentrate on the era of the content components.

3.3.2 Corpus and Preprocessing

To figure out how people produce slides from scholastic papers, we fabricate a corpus that contains sets of scholastic papers and their relating slides. Numerous scientists in the software engineering field put their papers and the comparing slides together in their landing pages. The landing pages' URLs are acquired by creeping Arnetminer.2 After downloading the landing pages, we utilize a few strict examples to remove the connections of the papers and the related slides what's more, download the documents to construct the dataset. We gather more than 2,000 sets. In the wake of tidying up the inaccurate sets, we have 1,200 paper-slides sets

The papers are all in PDF design and the slides are in either PDF or PowerPoint design. For the papers, we separate their writings by utilizing PDFlib3 and distinguish their physical structures of passages, areas and segments by utilizing ParsCit.4 A custom XML arrangement is utilized to depict this structure. For the slides, we likewise extricate their writings and physical structures like sentences, titles, visual cues, and so on. We utilize xpdf5 and the Programming interface gave by Microsoft Office to manage the slides in PDF and PowerPoint designs, individually. The slides are changed to a predefined XML position too.

4.1 OUR PROPOSED METHOD

4.1.1 Overview

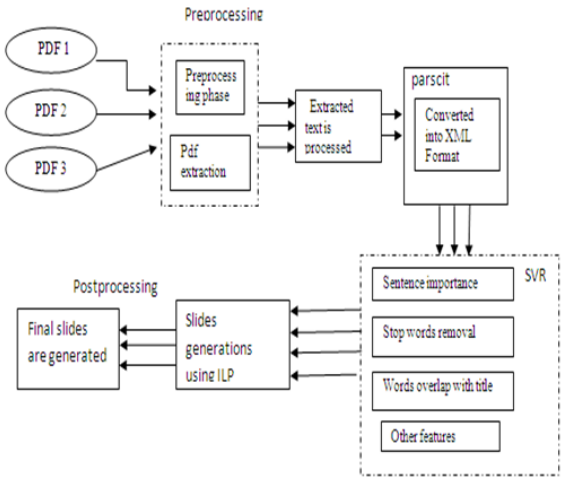

Figure 1: Generic framework for automatic slides generator for academic papers

In this paper, we propose a framework to naturally produce slides that have great structure and content quality from scholarly papers. The construction modeling of our framework is appeared We utilize the SVR-based sentence scoring model to appoint a significance score for every sentence in the given paper, where the SVR model is prepared on a corpus gathered on the web. At that point, we produce slides from the given paper by utilizing ILP. More subtle elements of every part will be talked about in the accompanying areas.

4.1.2 Sentence Importance Assessment

In our proposed PPSGen framework, sentence significance appraisal is one of the two key steps, which expects to dole out a significance score to every sentence in the given

paper. The score of every sentence will be utilized as a part of the slides era process. In this study, we present a couple of helpful components also, propose to utilize the bolster vector relapse model to accomplish this objective.

CONCLUSION

This paper proposes a novel framework called PPSGen to produce presentation slides from scholastic papers. Sentence scoring model is prepared taking into account SVR and use the ILP strategy to adjust and separate key expressions and sentences for producing the slides. Trial results demonstrate that our system can make immensely enhanced slides than standard schedules. scoring model taking into account SVR and utilize the ILP strategy to adjust and separate key expressions and sentences for creating the slides. Exploratory results demonstrate that our technique can create vastly improved slides than conventional strategies.

REFERENCES

[1] S. M. A. Masum, M. Ishizuka, and M. T. Islam, "Auto-presentation: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.

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

[3] 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.

[4] 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.

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

[6] A. Abu-Jbara and D. Radev, "Coherent citation-based summarization of scientific papers," in Proc. 49th Annu. Meeting Assoc. Comput.Linguistics: Human Lang. Technol.-Volume 1, 2011,pp. 500–509.

[7] K. Woodsend and M. Lapata, "Multiple aspect summarization using integer linear programming," in Proc. Joint Conf. Empirical Methods Nat. Lang. Process. Comput. Nat. Lang. Learn., 2012,pp. 233–243.

[8] P. B. Baxendale, "Machine-made index for technical literature: an experiment," IBM J. Res. Develop., vol. 2, no. 4, pp. 354–361,1958.

[9] V. Qazvinian and D. R. Radev, "Identifying non-explicit citing sentences for citation-based summarization," in Proc. 48th Annu. Meeting Assoc. Comput. Linguistics, Jul. 2010, pp. 555–564. 1096.

.