Automatic Slide Generation System

Snehal A. Patil[1], Apeksha J. Todkar[2], Lina S. Sonawane[3], Mahesh A. Saindane[4]

[1] Student, Dept. of Computer Engineering, SSBT COET, Jalgaon, Maharashtra, India.
[2]Student, Dept. of Computer Engineering, SSBT COET, Jalgaon, Maharashtra, India.
[3]Student, Dept. of Computer Engineering, SSBT COET, Jalgaon, Maharashtra, India.
[4]Student, Dept. of Computer Engineering , SSBT COET, Jalgaon, Maharashtra, India.

---***---

Abstract - *In most of the areas for sharing information, slide presentation plays an important role. The slides for presentation are traditionally prepared using various tools. The traditional way of presenting slides is labor-intensive. Labor-intensive nature leaves scope for human-errors. Also, for lengthy documents there is a chance of some important information being missed out. Preparing slides manually consume more time. The drawbacks of the traditional way lead to need for intelligent system. The intelligent system needs to be capable of generating slides with minimum human interference. The existing automatic tools fail to fetch the graphical elements from a given input. Hence the paper proposed an Automatic Slide Generation System. The proposed system fetches the graphical elements as well as text from a document. The proposed system is more reliable than the existing system.*

Key Words: Slide Generation, Tokenization, NLP, ILP, Text Mining

1. INTRODUCTION

Slide presentations are an effect making way to deliver key-concepts, convey and share information to the audience not only at professional but also educational meetings. The task of producing presentation slides manually from available data is both effort-taking and time-consuming.

Proposed system, Automatic Slide Generation System (ASGS) automatically generates slides containing graphical elements and text data. Developing the data mining technique is focused in the proposed system, which helps in giving scores to sentences and producing slides with graphical elements. The proposed system is designed by using Natural Language Processing (NLP) for giving scores to the paper, the Integer Linear Programming method (ILP) is used to generate well-structured slides. It is done by selecting and bringing into line key phrases and sentences along with diagrams. System uses key phrases as bullet points, and sentences applicable to the phrases are positioned below these bullet points. In order to extract key phrases, chunking implemented by the Open NLP library is applied to the sentences and noun phrases are extracted as the candidate key phrases. Images are extracted from the whole document by using the PDF box and PDF documents. Finally the slides are generated and images are added to the slide.

2. EXISTING SYSTEM

Manual systems put pressure on people to be correct in all details of their work at all times, the problem being people are not perfect, many of us wishes were. With the mannual systems the level of service is dependent on individuals and puts a requirement on management to run training continuously for staff to keep them motivated and to ensure they are following the correct procedures. It can be all too easy to accidentally switch details and end up with inconsistency in data entry. It has the effect of not only causing problems with customer service but also making information unable be used for reporting or finding trends with data discovery. Reporting and checking data is robust can be timely and expensive. It is often an area where significant money can be saved by automation.

2.1 Disadvantages of Manual Slide Generation System

Many tools help the presenter to generate the slides. These tools only help them in the formatting of the slides, but not in the content. It still takes presenters much time to produce the slides from the available material. The traditional tools thus require a lot of investment, both in terms of time and efforts. Duplication of data entry may happen. The task of generating presentation slides from one or many written materials is both tedious and time-consuming. It takes more effort and physical space to keep track of paper documents, to find information and to keep details secure. When mistakes are made or changes or corrections are needed, mostly a manual editing must be completely done again rather than just updated. With manual or partially automated systems information often has to be written down and copied or entered more than once. Systemization can reduce the amount of duplication of data entry.

3. PROPOSED SYSTEM

Paper proposed a method of automatically generating presentation slides. Automating the task of creating slides from a given source would save presenters valuable time allowing him/her to concentrate on other aspects of the presentation such as preparing their speech. Project extends the ones knowledge of Artificial Intelligence, focussing heavily on the field of Natural Language Processing. The primary aim is to generate slides for the user so as to reduce their time and efforts in setting up presentation slides.

The project named Automatic Slide Generation System is a window based application created in Core Java. In system, it proposes a framework to naturally produce slides have great structure and content quality from scholarly paper and the pdf. The construction modelling of framework is appeared It 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 point, it produces slides from the given paper by utilizing ILP.

The main purpose of the proposed system is to improve the efficiency and reduce the time required for automatic slide generation while maintaining the graphical elements with the text elements.

4. SYSTEM ARCHITECTURE

The following figure shows the architectural view of the proposed system. The description of the system is as follows:

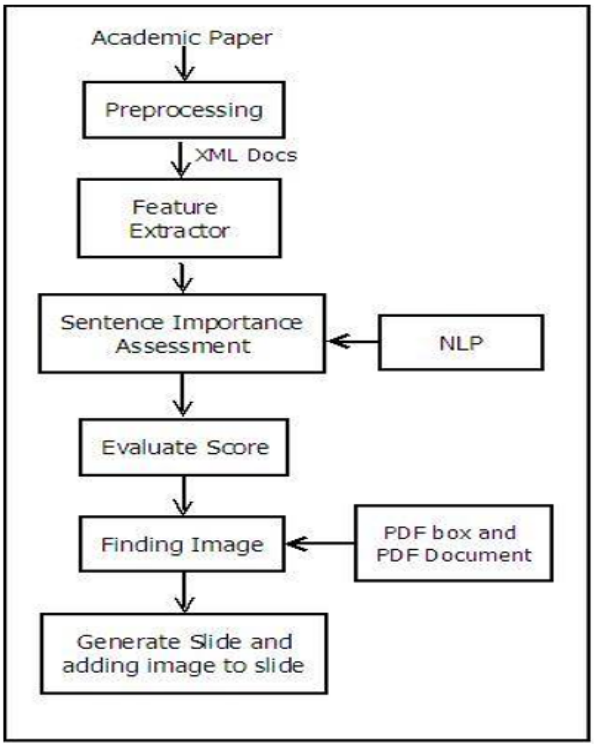

Fig -1: System Architecture

Proposed system primarily uploads the input file and preprocesses the input file in which the process of tokenization is done wherein sentences are splits up into tokens. Stemming is done in which the sentence is into their first normal form. Stop words such as like but, also, etc are removed from the data. Postagging of procedure is applied and score is evaluated by using the NLP procedure. Image is extracted by using PDFbox and PDF document from the input file. Finally, the slides with graphical images are generated.

5. ALGORITHM

The proposed system implements following algorithm for automatic slide generation along with graphical elements.

5.1 Algorithm 1: Proposed System Algorithm

Input: PDF file

Output: Slide Generated with graphical element and text.

Process:

1. Read PDF File.
2. Apply Parscit to detect their physical structures of paragraphs, sections and sections.
3. Token generation and apply stemmer.
4. Remove Stop words and apply post tagging.
5. Calculate sentence scoring using NLP.
6. Add Image into Slide (See detail in Algorithm 2).
7. Add scoring sentences and image using ILP method.

5.2 Algorithm 2: Algorithm for Including Graphical Elements into Slides

Input: PDF file

Output: Adds Graphical element in slide.

Process:

1. Read PDF File.
2. Load PDF file into PDF document.
3. List of all pages in the document.
4. Find image of each single page and store into map data.
5. Repeat Step 4 for all pages.
6. Check label of image and add image according to label.

6. MATHEMATICAL MODEL

Mathematical model of proposed system is stated below. The system S is represented as:

S = {U, P, T, R, F, PP, SP}

Input

Browse Dataset

U = {u1, u2, u3, ..., un}

Where

U is a set of number of related papers.

u1; u2; u3, ..., un are the number of papers.

Process

Parscit Process

P = {p1, p2, p3, ..., pn}

Where,

P is represented as a set of Parscit Process

To detect their physical structures of paragraphs, sections and sections by using Parscit.

p1, p2, p3, ..., pn are the number of Parscit process.

Preprocessing Method

T= {t1, t2, t3,……, tn}

Apply token generation, stemming, removing stop words and post tagging.

t1,t2,t3…, tn are the number of preprocessing process.

NLP Model

R = {r1, r2, r3,…., rn}

Where,

R is represented as a set of NLP model In NLP method, calculate score of the sentence.

r1, r2, r3,…..,rn are the number of NLP model process.

Finding Image

F = {f1, f2, f3,…., fn}

Where,

F is represented as a set of Finding Image

Finding image using PDF document and add all image into data Map.

f1, f2, f3, ..., fn are number of Finding Image process.

ILP Method

PP = {pp1, pp2, pp3,…., ppn}

Where,

PP is represented as a set of ILP processing

In ILP method, important sentences are added to the slide along with the images.

pp1, pp2, pp3,….., ppn are number of ILP processing process.

Output

Slide Transitions

SP = {sp1, sp2, sp3,…., spn}

Where,

SP is represent as a set of slide generation

sp1, sp2, sp3,….., spn are number of final output.

7. RESULTS AND ANALYSIS

This system automatically generates the slides from the research paper. This section shows the result obtained the proposed system.

The Table -1 shows the Average Pyramid Scores of TF-IDF, for the generated presentation slides. Overall research on the paper shows that proposed method can generate slides with better quality than the baseline methods. Using the ROUGE toolkit and the Pyramid evaluation, the slides generated by our method gets better ROUGE scores and Pyramid scores. Therefore, our slides are considered a better basis for preparing the final slides.

Table -1: Average Pyramid Scores Comparison

Method	Average Pyramid Score
TF-IDE	0.819
SVR-ILP	0.792
SVR-NLP	0.849

The graph shows that the slides generated by our proposed system shows higher pyramid score than that of SVR-ILP and TF-IDP because the proposed system not only contains the text elements but also graphical element like figures from the given input paper.

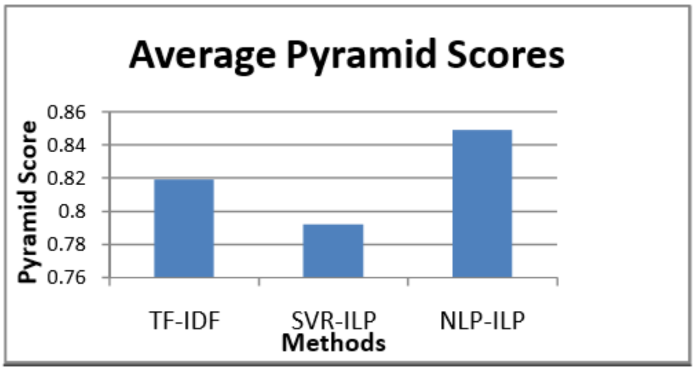

Fig -2: Average Pyramid Score Graph

The object function not only considers the count of the diagrams but also the weights of the diagrams. The author-written slides generally contain important sentences and important diagrams. By introducing the weights of the diagrams, the ILP solution tends to select more important diagrams and the generated slides can match the reference slides better and thus get better performance.

8. CONCLUSION AND FUTURE SCOPE

The system is proposed for automatically generating the slides from the pdf. The system generates the slides which includes text and graphical element. The graphical elements along with the text data make the generated slides look more comprehensible and vivid. System initially finds the graphical elements of each page from the paper and after that system stores the image to map data. System performs the same operation for each page and finally checks the label of each image and adds that image to the slide according to the label of each image. NLP method is used for sentence scoring and the ILP method is used for slide generation which contain key phrases and the relevant sentences. Presently, the system generate slides on the basis of only one given pdf, but in future extra information like additional relevant pdfs and information of citation can be utilized for improving the slides generation.

REFERENCES

[1] Ektaa Meshram, D. A. Phalke,"Technique for Generating Automatic Slides on the basis of Paper Structure Analysis", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 5, Issue 6, June 2016.

[2] Yue Hu and Xiaojun Wan, "PPSGen: Learning-Based Presentation Slides Generation for Academic Papers, in IEEE transactions on knowledge and data engineering, vol. 27, no. 4, april 2015.

[3] V. Qazvinian, D. R. Radev, S. M. Mohammad, B. J. Dorr, D. M. Zajic, M. Whidby, and T. Moon, Generating extractive summaries of scientic paradigms, J. Artif. Intell. Res., vol. 46, pp. 165 201, 2013.

[4] U. Masao and H. Koiti, Automatic slide presentation from semantically annotated documents, in Proceedings of the Workshop on Coreference and its Applications, pp. 2530, Association for Computational Linguistics, 2012.

[5] Y. Yasumura, M. Takeichi, and K. Nitta, A support system for making presentation slides, Transactions of the Japanese Society for Artificial Intelligence, vol. 18, pp. 212220, 2011.

[6] T. Hayama, H. Nanba, and S. Kunifuji, Alignment between a technical paper and presentation sheets using a hidden markov model, in Active Media Technology, 2005.(AMT 2005). Proceedings of the 2005 International Conference on, pp. 102106, IEEE, 2005.

[7] T. Shibata and S. Kurohashi, Automatic slide generation based on discourse structure analysis, in Natural Language ProcessingIJCNLP 2005, pp. 754766, Springer, 2005.

[8] M.-Y. Kan, Slideseer: A digital library of aligned document and presentation pairs, in Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries, pp. 81 90, ACM, 2007.

[9] 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. 240246.

BIOGRAPHIES

Student, Shram Sadhana Bombay Trust's College of Engineering and Technology, North Maharashtra University, Jalgaon, Maharashtra, India.

Student, Shram Sadhana Bombay Trust's College of Engineering and Technology, North Maharashtra University, Jalgaon, Maharashtra, India.

Student, Shram Sadhana Bombay Trust's College of Engineering and Technology, North Maharashtra University, Jalgaon, Maharashtra, India.

Student, Shram Sadhana Bombay Trust's College of Engineering and Technology, North Maharashtra University, Jalgaon, Maharashtra, India.