iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** V **Month of publication:** May 2016

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 4 Issue V, May 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Automatic Slide Generation of Academic Paper

Aruna.M[1], Manju.A[2]
[1]PG Scholar, [2]Assistant Professor
Department Of Computer Science and Engineering, Saveetha Engineering College, Tamil Nadu, India

Abstract- Presentation slides have been a effective and easy means to present and transfer information, especially in academic conferences. Academic papers are difficult to read and summarize content within a short span of time. This consumes large amount of time and effort of the researchers to extract information from the papers. Existing system have tools such as Microsoft PowerPoint in windows and Open Office in open source to format slides but not in the content generation. It still take presenters time to prepare the content from scratch. In general, automatic slide generation for academic paper is a very challenging task than summarization. Current methods like GDA tag set, TF-IDF, Hidden Markov Model (HMM) scores generally extract any sentences from the paper to construct the slides. The proposed method extracts important sentence from the paper and automatically generate slides which can be used as basis and draft slide. It first employs the regression method to learn the important scores of the sentences in an academic paper, and then exploits the integer linear programming to generate well structured slides by selecting and aligning key phrases and sentences. The work focuses on both text and image elements such as figures into the slide and the goal of this system is to automatically generate well structured slides by considering both text and image elements in paper and provide such draft slides as a basis to reduce the presenter's time and effort when preparing their final presentation slides.

Keywords- Academic paper, Presentation slides, Integer Linear Programming, Support Vector Regression, Web Mining.

I. INTRODUCTION

Web Mining is a data mining technique to discover pattern from World Wide Web. Web Mining can be classified into three types such as Web Content Mining, Web Structure Mining and Web Usage Mining. Web Content Mining is used in extracting, integrating and mining useful information from Web page content. Web Structure Mining is used to identify and analyze the nodes in a network connection and it is based on graph theory. Web Usage Mining is used to estimate the weblogs. It is used to understand the browsing behaviour of a user and to discover pattern for serving the needs of a web based applications. Automatic slide generation is a method of Web Content Mining which is used in extracting vital points from the document and preparing slides.

A slide is a single page of a presentation. Collectively, a group of slides may be known as a slide deck. Presentation is the process of presenting a topic to audience. It is typically a demonstration, lecture, a speech meant to inform, persuade or build a good will. A collection of pages arranged in sequence that contain text and images for presenting to an audience. It often refers to a Microsoft PowerPoint presentation. In the digital age, a slide most commonly refers to a single page developed using a presentation program such as Microsoft PowerPoint or Apple Keynote. It is also possible to create them with a document mark-up language, for instance with the Latex class Beamer. Lecture notes in slide format are referred to as lecture slides, frequently downloadable by students in .ppt or .pdf format. Presentation slides can be done by different types of software such as Microsoft PowerPoint, Apple Keynote, Prezi, SlideDog, ClearSlide,etc.,

Slides have been an effective and popular means of presentation of information. Slides provide an easy way of sharing knowledge between different groups of people. A presenter takes the help of slides to present his work in a pictorial and more convenient way in many conference and meetings. In past years with the availability of many software tools like Microsoft PowerPoint supported in Windows, Open office Presenter in Linux is used only for formatting slides but not in content generation. More recently, Google Slide had been developed by Google for creating and formatting slide in open source but not in content. These tools help only in the formatting of content such as alignment, bulleting, font style and size etc., but not in preparing the content itself. A user has to start taking notes from the academic paper and then vital points in the paper is identified and used in presentation which is a time consuming task. But this concept propose a methodology that generates slides for the presentation with important points and all necessary figures, tables and graphs from a technical paper. It has been evident that proposed technique saves time and reduces the effort of presenter by providing a basic presentation, which can be further upgraded as final presentation.

Slides contain the summarized version of an academic paper and file associated with academics. They contain the vital points of the document arranged in a systematic way, including graphic elements like figures and tables for easy illustration of the idea. Image

elements are the core of an academic document. Given a document, automatically generating presentation slides becomes a nontrivial task because of challenges like segmentation of document into multiple topics, summarizing content of each topic and aligning these topics to one or more slides and placing necessary graphical content like figures, graphs and tables in appropriate slides at appropriate locations.

II. RELATED WORK

PPSGen investigated the method of automatically generating presentation slides for academic papers. This method first employs the regression method to learn the importance scores of the sentences in an academic paper, and then exploits the integer linear programming (ILP) method to generate well-structured slides by selecting and aligning key phrases and sentences. Evaluation results on a test set of 200 pairs of papers and slides collected on the web demonstrated that PPSGen system can generate slides with better quality [29]. Sentence assessment is done based on Support Vector Regression (SVR) [22]. Then by using extracted sentence from the paper slide is generated. Slide generation is based on Integer Linear Programming (ILP). Then after postprocessing generated output is the presentation slide. The disadvantage of the system is that it considers only text elements into the slide. Graph elements such as tables and figures are not considered.

Multi document summarization involves multiple aspects of content selection and surface realization. Multi summarization method is used to create summaries from multiple papers based only on integer linear programming (ILP). The summaries must be informative, succinct, grammatical, and obey stylistic writing conventions [19]. It learns individual aspects but optimized jointly using an integer linear programme. The ILP framework allows combining the decisions of the expert learners and to select and rewrite source content through a mixture of objective setting, soft and hard constraints. The disadvantage of this system is that it focuses on summarization of the content. It does not generate slides and it does not consider graph elements. Sentence importance assessment is not up to quality standards.

SlidesGen investigated automatic generation of presentation slides from technical papers in LATEX. A query specific extractive summarizer QueSTS is used to extract sentences from the text in the paper to generate slides. QueSTS transfers the input text to an integrated graph (IG) where a sentence represents a node and edges exist between the nodes that the sentences corresponding to them are similar. The weights of the edges are calculated as cosine similarity between the sentences [18]. SlidesGen framework includes steps such as Preprocessing, Configuration file generation, Extracting key phrases and QueSTS Summarizer [21]. Finally slides are generated for each section in the paper and graphics are rendered in presentation. The disadvantage of the system is that it takes input only as latex format, wherein all documents cannot be considered as input. Most of the documents are in PDF format so many documents cannot be processed.

Term Frequency-Inverse Document Frequency (TF-IDF) investigated a support system for making presentation slides from a technical paper. This system provides functions that assign slides to each section and put objects on a slide [28]. The system calculates weights of terms in the document by the TF-IDF method. Based on the term weights, objects in the document such as sentences, figures and tables are weighted. Finally, outputs of the system are generated as presentation slides in XHTML. An extractive multi-document summarizer [6] forms summaries by extracting or selecting sentences from the input documents, without modifying the selected sentences. The disadvantage of the system is that it calculates the weight of the sentences based on the term weight. This method is also used only for summarization and the sentence extracted using TF-IDF are less important when compared with Support Vector Regression.

Integer Linear Programming (ILP) method investigated a system to generate extractive multi-document summaries. The method uses Integer Linear Programming to jointly maximize the importance of the sentences it includes in the summary and their diversity, without exceeding a maximum allowed summary length [6]. To obtain an importance score for each sentence, it uses a Support Vector Regression[17] model trained on human-authored summaries, whereas the diversity of the selected sentences is measured as the number of distinct word bigrams in the resulting summary. Experimental results on widely used benchmarks show that our method achieves state of the art results, when compared to competitive extractive summarizers, while being computationally efficient as well [14].The disadvantage of the system is that it generates summaries depending on the human trained corpus. The area for this system is narrow, it generates summaries only for specific document depending on the corpus.

LexRank investigated method of generating extractive summaries using bibliometric text mining and summarization techniques to generate summaries of scientific literature [9]. This system shows how the system use citations to produce automatically generated, readily consumable, technical extractive summaries. This system shows how the system use citations to produce automatically generated, readily consumable, technical extractive summaries. This method first proposed C-LexRank, a model for summarizing

Volume 4 Issue V, May 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

single scientific articles based on citations, which employs community detection and extracts salient information-rich sentences. Next, experiments are done to summarize a set of papers, which cover the same scientific topic. This method generate extractive summaries of a set of Question Answering (QA) and Dependency Parsing (DP) papers, their abstracts, and their citation sentences and show that citations have unique information amenable to creating a summary.

Text mining is the discovery of interesting knowledge in text documents. It is a challenging issue to find accurate knowledge (or features) in text documents to help users to find what they want. Many data mining techniques have been proposed for mining useful patterns in text documents [16]. However, how to effectively use and update discovered patterns is still an open research issue, especially in the domain of text mining. Since most existing text mining methods adopted term-based approaches, they all suffer from the problems of polysemy and synonymy. This system presented an innovative and effective pattern discovery technique which includes the processes of pattern deploying and pattern evolving, to improve the effectiveness of using and updating discovered patterns for finding relevant and interesting information. Substantial experiments on RCV1 data collection and TREC topics demonstrate that the proposed solution achieves encouraging performance [26].

III. PPT GENERATION BASED ON SVR AND ILP

Academic Paper or any document is given as input to the system which does preprocessing which is based on stop word removal technique. Preprocessed document is given as input to calculate the frequency of word appearing in the document. Similarity clustering is done based frequency of words collected in the document. Topics are given to each cluster in the document using lexical analysis. Sentence importance assessment is done based upon Support Vector Regression (SVR) which extracts important sentence from the document and put it into slide. Slide is created by placing necessary text and graph into the slide which obeys slide writing conventions using integer Linear Programming (ILP). The output to the given document is in presentation (.ppt) format. It is easy for user to gain information from presentation rather than whole document.

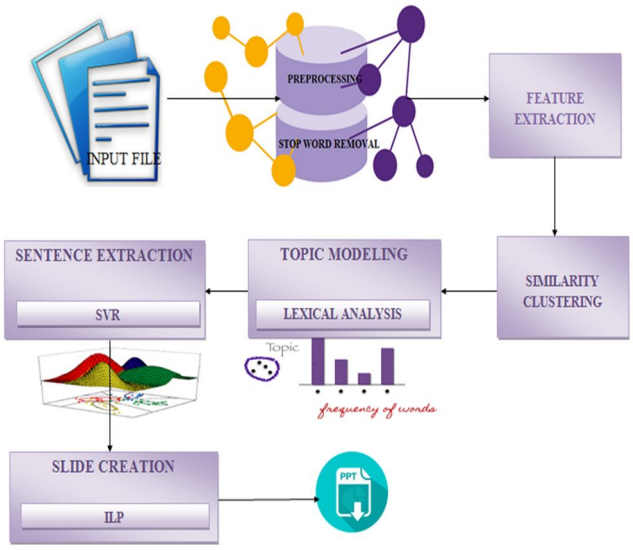


Fig 1.Architecture of PPT generation

A. *Support Vector Regression*

Support Vector Regression is used in maintaining all the main features which characterize the maximal margin of the algorithm. The Support Vector (SV) algorithm is a nonlinear generalization of the Generalized Portrait algorithm developed in Russia in the sixties. As such, it is firmly grounded in the framework of statistical learning theory, or *VC* theory, which has been developed over the last three decades by Vapnik and Chervonenkis [22]. In a nutshell, VC theory characterizes properties of learning machines which enable them to generalize well to unseen data. The key idea is to construct a Lagrange function from the objective function which is called as the primal objective function and the corresponding constraints, by introducing a dual set of variables. It can be shown that this function has a saddle point with respect to the primal and dual variables at the solution.

Volume 4 Issue V, May 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

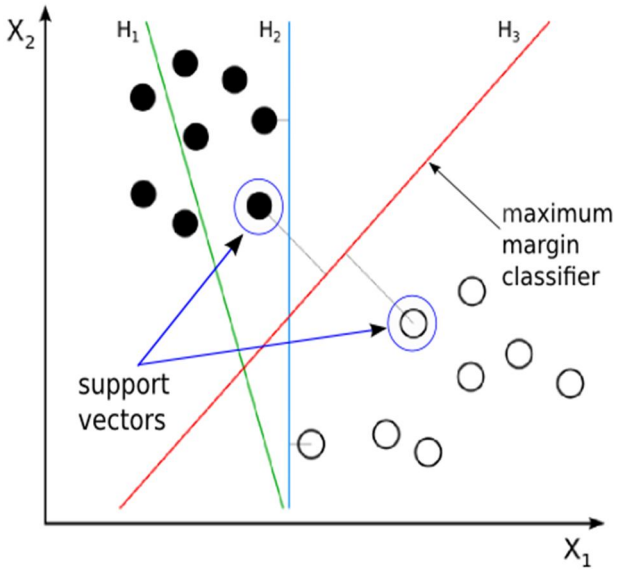


Fig 2. Support Vector Regression

B. Integer Linear Programming

An integer linear programming problem is a mathematical optimization or feasibility program in which some or all of the variables are restricted to be integers. In many settings the term refers to integer linear programming (ILP), in which the objective function and the constraints are linear.

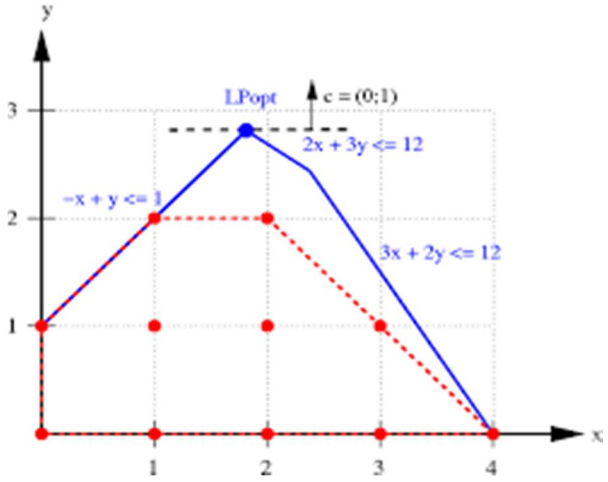


Fig 3. Integer Linear Programming

C. PDF Extraction

Portable Document Format(PDF) is a standard layout which cannot be altered or modified easily by the user. All the data and information should be extracted from the document without loss of any data. PDFBox Api is imported such that it is used to process the documents in PDF format. There are many tools to extract such as GreenStone, IcePDF, PDFLib etc., but every tool cannot extract the content accurately, it does not process image elements and it does not compatible with every platform. This disadvantage can be overcome using iText PDF This method can be used to extract text and images from the PDF Document which is used efficiently when compared with other tools. This method is used to bring documentation to the next level. This method can be customized by the user and it is compatible with many of the programming languages.

IV. SLIDE GENERATION TECHNIQUES

Slide Generation is defined as the method of preparing content and placing it in a slide which is called as presentation slides. This can be done manually or automatically. Different slide generation approaches [16] are natural language processing, latex format,

web mining, text summarization and machine learning which is based on semantic structure of the article. Generating slides using inference [3] is based on annotation method. Inference is done based on identification of topic. Text summarization [3] is method of summarizing the document based on preprocessing techniques such as stop word removal, stemming and case fold have been used. Then it uses a scoring scheme to generate summary of the given content. Slide generation using natural language processing depends on discourse structure of the paper. Generation of slides from latex format is based on the weighting score of each object in the sentence. Slide generation is also done based on the query provided by the user. This system takes the query as input and gathers information from the web pages and provides it to the user as summary and slide. This technique of slide generation is based on web mining. Slide generation is done based on corpus or collection of a text. This technique is based on machine learning methodology which needs training initially on a particular domain. Due to growth in learning technologies it is encouraged to generate slide automatically with less amount of time.

A. Stop Word Removal

PDFBox API is used to read and write pdf files. Preprocessing is done by stop word removal technique. Stop words are words which are filtered out before or after processing of natural language data. Though stop words usually refer to the most common words in a language, there is no single universal list of stop words used by all natural language processing tools, and indeed not all tools even use such a list. Some tools specifically avoid removing these stop words to support phrase search.Most frequently used words in English are useless in Text mining. Such words are called Stop words. Stop words are language specific functional words which carry no information. It may be of the following types such as pronouns, prepositions, conjunctions. Stop word list are contained in a list which contains maximum of stop words. Once the file is read it checks the uploaded file with the stop word dataset. If the words in the document and dataset are matched, then the corresponding word is removed from the document. Then the non-stop word is send as input to calculate word frequents.

B. Feature Extraction

Feature extraction is used to extract important features such as word frequent, sentence position, word overlap with the title and sentence parse tree information. In this system only word frequent is mined from the document. Here maximum number of repeated words in the file is considered for analyzing and is used to know which topics are explained in the paper. In this module, the system gets the non-stop words as input and calculates the count of words and finds the repeated occurrence of each and every word from the non-stop words. Repeated occurrence of the word is used to identify key and core words of the document. If a key word is identified it gives an idea to presenter to gain rough knowledge about the document.

Feature Selection Algorithm is used to find frequent words of the document. It involves frequent mining of the given input document. Feature Extraction involves calculating the total count of all words in the given input document, sorting the frequency of words depending on the highest value and finally sorting the words based on the alphabetical order. This type of mining and sorting is used for easy processing of the document. Then it sends topmost 20 important words of the document to the next module.

C. Similarity Clustering

In similarity clustering, from the maximum frequents word, weightage of every word is found. Then the weightage is calculated based on the similarity between the words and depending on the similarity words are grouped as cluster. Similarity is calculated based on wordnet tool. Wordnet tool is used to find the ontology of the given input document and it is compared with the frequent of the document. Generated ontology document and frequent document is compared and the important terms are clustered. WordNet is a machine readable dictionary built on the basis of psycholinguistic principles. It contains English nouns, verbs, adjectives, and adverbs organized on the basis of their word meanings, rather than word forms. Each word form in WordNet represents some underlying lexical meaning. Some word forms can represent several meanings and some meanings can be represented by various forms. Java API for WordNet Searching (JAWS) is an API that provides Java applications with the ability to retrieve data from the WorldNet database. It is a simple and fast API that is compatible with both the 2.1 and 3.0 versions of the WordNet database files and can be used with Java 1.4 and later. To use JAWS in your application following must be done:

1) Obtain a copy of the WordNet database files, which can be accomplished by downloading and installing WordNet.
2) Download the Java Archive (JAR) file containing the compiled JAWS code.
3) Include the Java Virtual Machine's class path the JAR file you downloaded from the internet and use the wordnet.database.dir system property to specify where the WordNet database files are located

Volume 4 Issue V, May 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

D. Topic Modeling

Topic models provide a simple way to analyze large volumes of unlabeled text. A topic consists of a cluster of words that frequently occur together. Using contextual clues, topic models can connect words with similar meanings and distinguish between uses of words with multiple meanings. There are different types of topic modelling techniques available whereas this system is implemented with unsupervised topic modelling algorithm. Topic modeling [5] is the suite of algorithms that uncover the hidden thematic structure in document collections. This method provides a new way of searching and browsing experience. It is also used to summarize large archives of text. Different topic modeling algorithms include Latent Dirichlet allocation, Latent Semantic analysis (LSA), Explicit Semantic Analysis and Hierarchial Dirichilet Process. Latent Dirichilet allocation (LDA) [11] is a generative model that allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar. For example, if observations are words collected into documents, it posits that each document is a mixture of a small number of topics and that each word's creation is attributable to one of the document's topics.

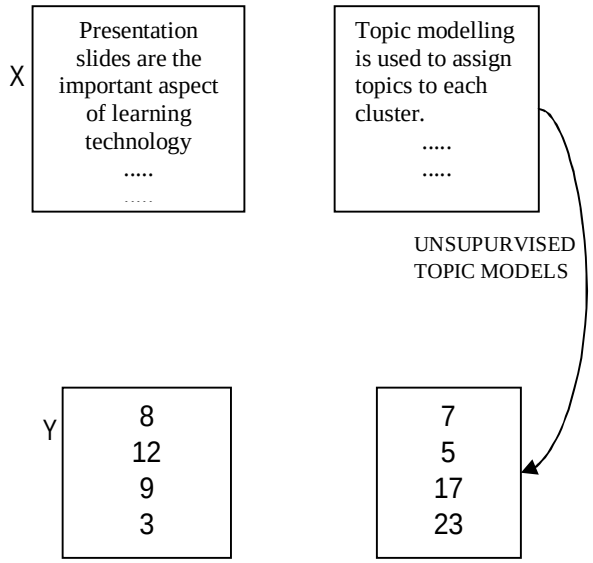


Fig 4. Topic Modeling Representation

E. Sentence Extraction

Sentence Extraction is done based on the topics generated during topic modelling. Sentence Extraction is done to create summary report for the given base paper. Summaries are used to provide an overview of the given base paper. Sentence Extraction is based on Support Vector Regression. It is used to predict the important points from the document. Important points are extracted by scanning the entire document. Summaries that are generated using this methodology splits the input line by line and extracts only the vital points from the document. Sentence extraction initially extracts the abstract from the given base paper and create summary which provides overview of the given paper.

F. Slide Creation

Using sentence extracted from the input document presentation is generated by obeying general stylistic conventions. Slide creation is done by integer linear programming (ILP). Generated slide can be downloaded from the web or it can be retrieved from the specific folder. Slide creation must incorporate PDFBox api and then it should set both the constraints and objectives in linear fashion.

V. EXPERIMENTAL SETUP

The system asks the user to upload a PDF file which then gets uploaded with the file name displayed on the web page. After uploading the PDF document, it is processed, images are extracted in a separate folder and images are preprocessed thereby stop words are removed from the document. It then calculates the frequent occurring of the words in a document and it calculates the topmost twenty important words which is then clustered based on wordnet tool and then topics are generated for the given document. It then extracts sentence for summary generation. After generating summaries presentation slides can be downloaded.

Volume 4 Issue V, May 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Generated slide appears as shown below:

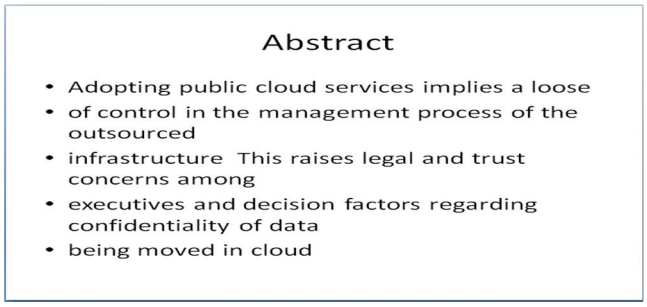


Fig 5. Sample Generated Slide

This system has an advantage of extracting image elements from the base paper. Images are the heart of a system which is used to illustrate the idea easily. Pictorial representation of the concept is efficient when compared with the text elements. This system extracts both the graph and text elements into the slide where the traditional method does not support this technique.

Fig 6. Generated Image Element

VI. CONCLUSION

Automatic slide generation for academic paper is an interesting technique. This method is used to automatically generate slides considering the important point from the academic paper. Slides provide a better way of understanding rather than document. This system can be further enhanced by using training data in which the slide generated will be efficient when compared with traditional methods. It can be further extended by creating slides for multiple documents. Multi document slide generation would gain popularity and advantage over the recent future.

REFERENCES

[1] 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.
[2] Alex J. Smola, Bernhard Schölkopf , “Tutorial on support vector regression”, Statistics and Computing, ,August 2004, Volume 14, pp 199-222.
[3] B. Beamer and R. Girju, “Investigating automatic alignment methods for slide generation from academic papers,” in Proc. 13th Conf. Comput. Natural Lang. Learn., Jun. 2009, pp. 111–119.
[4] Chong Tze Yuang, Chng Eng Siong,” An Empirical Evaluation of Stop Word Removal in Statistical Machine Translation”, Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, pages 30–37, Avignon, France, April 2012.
[5] Cong Duy Vu Hoang and Min-Yen Kan, (2010), Towards automated related work summarization. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters, COLING ’10, pages 427–435, Stroudsburg, PA, USA. Association for Computational Linguistics.
[6] D. Galanis, G. Lampouras, and I. Androutsopoulos, “Extractive multi-document summarization with integer linear programming and support vector regression,” in Proc. COLING, 2012, pp. 911–926.
[7] D. Gillick and B. Favre “A scalable global model for summarization,” in Proc. Workshop Integer Linear Program. Nat. Lang. Process., 2009, pp. 10–18.
[8] 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.
[9] G. Erkan and D. R. Radev, “LexPageRank: Prestige in multi-document text summarization,” in Proc. EMNLP, 2004, pp. 365–371.
[10] Gonenc Ercan, Ilyas Cicekli,” Lexical Cohesion Based Topic Modeling for Summarization”, in Computational Linguistics and Intelligent Text Processing, pg

no 582-192.

[11] Jiawei Han, Hong Cheng and Dong Xin Xifeng Yan, "Frequent pattern mining: current status and future directions", Data Min Knowl Disc (2007).

[12] 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, pg.233-243.

[13] List of English stopwords, Available: http://xpo6.com/list-of-english-stop-words.

[14] Masum, S. M. A., and Ishizuka, M. (2006) "Making topic specific report and multimodal presentation automatically by mining the web resources", in Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, 240–246. Washington, DC, USA: IEEE Computer Society.

[15] McDonald, R. (2007). A study of global inference algorithms in multi-document summarization in Proceedings of the European Conference on Information Retrieval, pages 557–564, Rome, Italy.

[16] 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.

[17] M. Utiyama and K. Hasida, (2003) "Automatic slide presentation from semantically annotated documents," in Proc. ACL Workshop Conf., pp. 25–30.

[18] 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.

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

[20] 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.

[21] R. Mihalcea and P. Tarau, "A language independent algorithm for single and multiple document summarization," in Proc. IJCNLP, 2005, pp. 19–24.

[22] Srimanta Pal, Debasish Basak and Dipak Chandra Patranabis,"Support Vector Regression" Neural Information Processing – Letters and Reviews Vol. 11, No. 10, October 2007.

[23] Standard set of english stopwords [online] Available: https://github.com/arc12/Text-Mining-Weak-Signals/wiki/Standard-set-of-english-stopwords .

[24] T. Berg-Kirkpatrick, D. Gillick, and D. Klein, "Jointly learning to extract and compress," in Proc. 49th Annu. Meeting Assoc. Comput. Linguistics: Human Lang. Technol., 2011, pp. 481–490.

[25] T. Hayama, H. Nanba, and S. Kunifuji, (2005) "Alignment between a technical paper and presentation sheets using hidden Markov model," in Proc. Int. Conf. Active Media Technol., pp. 102–106.

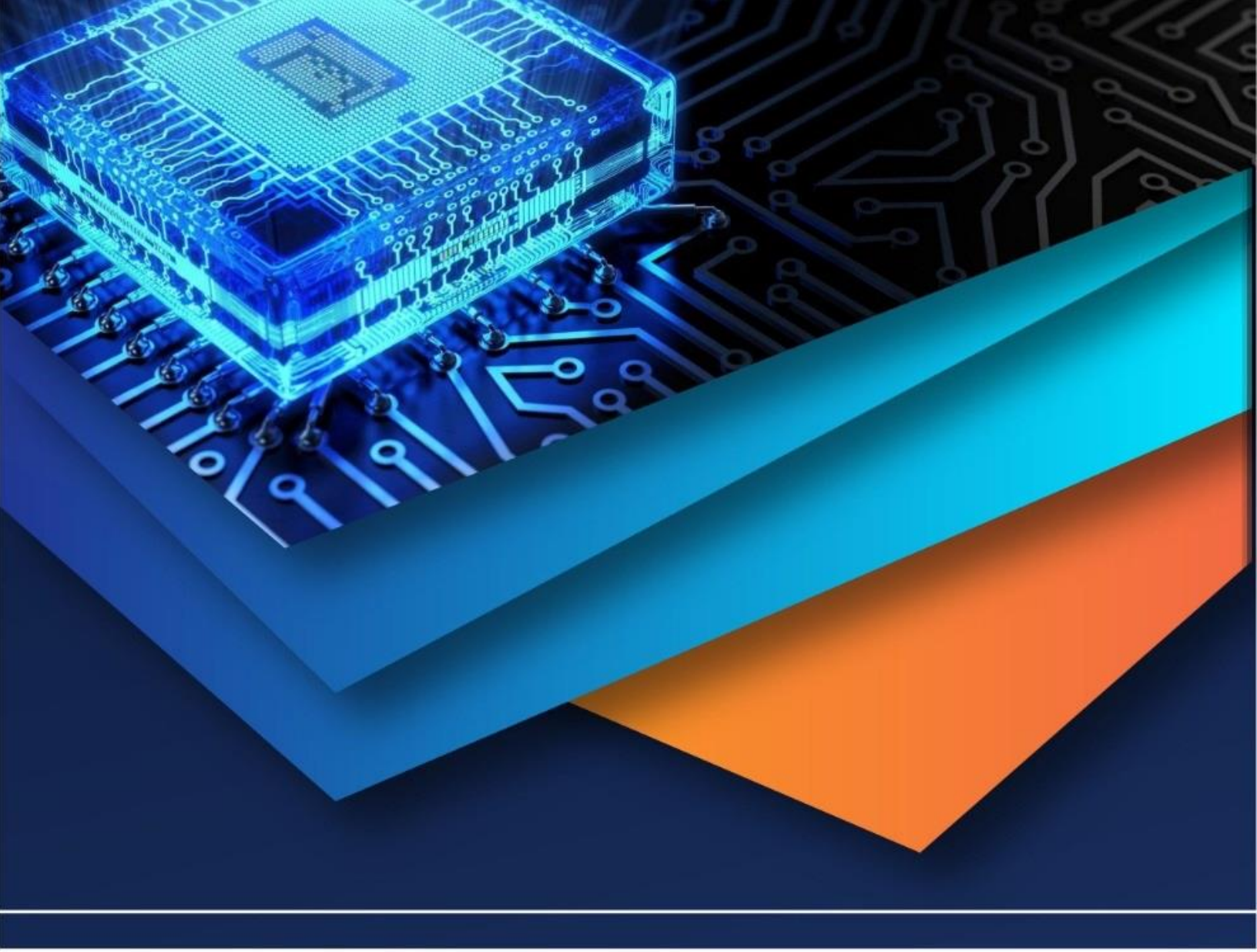
[26] T. Shibata and S. Kurohashi, (2005) "Automatic slide generation based on discourse structure analysis," in Proc. Int. Joint Conf. Natural Lang. Process., pp. 754–766.

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

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

[29] Yue Hu and Xiaojun Wan, (2015)" PPSGen: Learning-Based Presentation Slides Generation for Academic Papers" IEEE transactions on Knowledge and Data engineering, vol. 27, no. 4.

[30] Zajic, D. M., Dorr, B. J., Lin, J., & Schwartz, R. (2007). Multi-candidate reduction: Sentence compression as a tool for document summarization tasks. Information Processing and Management



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY