AN AUTOMATIC PPT GENERATION USING DATAMINING

T.K.P RAJAGOPAL[1],S.SHANTHINI[2],R.SUDARVIZHI[3],S.THANGARAJU[4]

[1]*computer science ofengineering,Kathir college of engineering*
[2,3,4] *information Technology,Kathir college of engineering*

Abstract-In this,we investigate a very challenging task of automatically generating presentation slides for academic papers. The generated presentation slides can be used by the presenters to prepare their formal slides in a easier way. It first employs a model called SVR model(Support Vector regression model). System use Support Vector regression model for sentence importance assessment by checking importance score of each sentences in an academic paper.Another model named ILP model is used.After getting importance score for each sentence from academic word document with the help of SVR and ILP method is used to generate fine draft slides with content as key sentences. Evaluation results on a test set of pairs of slides collected on the web demonstrate that our proposed system can generate slides with better quality.

Keywords-presentation,slidegeneration,sentence,document,analysization,accuracy

I. INTRODUCTION

Data mining is the process of retrieving patterns and useful information from data. In the context of web search, some data mining techniques are more applicable than others. The most used data mining techniques in the field of web search are: document clustering and document classification techniques.

Data Mining is known as Knowledge Discovery Database refers to the nontrivial extraction of implicit previously unknown ,potential useful information from data in database..Data mining the extraction of hidden predictive information from large database is a powerful technology with great potentiality to help companies focus on the most important information in the data warehouse. Data mining predicts future trend allowing business to make proactive knowledge decisions. The automated prospective offered by data mining move to the analysis of past events provided by tools typical of decision support systems. Data mining techniques can be implemented on the software and hardware platform to handle the value of existing information resource can be integrated with new products and systems.

II. PROCESS OVERVIEW

The document is given as input and it is given into the pre-processing sector.The pre-processor sector consist of tokenizing and stop words.These two process is used to check the document in the given process.Tokenization is used to allocate space and check every single word in the document after analysing it . Another important method is Stop Words.

The stop words are called as elimination process.This stop word allows the software to use the words like "is" "as" wherever it is needed in the document. If not, it will eliminate the words in the unnecessary places in the document.After reaching out from the pre-processing sector,the document will be sent for "Frequency count" process.The frequency count is a process of counting the frequently repeated words in a document.It checks them for processing.

System use Support Vector regression method for assessment by calculating importance score of each statement. Highest score terms, sentences,and phrases will be used as key statement, phrases for slide generation.We also implement some methods to compute the important score of sentence.The SVR Model uses a method named Keyword construction. This is used to construct the keyword in a document.The SVR model refers this keyword construction method if they need

it.After completing the above processes, the document finally enters the slide generation method.The slide starts to generate for execution.

III.   FEATURES

The graphic device interface (GDI) is the version GDI+ can take advance of gamma correction and 3D interfaces in high color depth. Windows XP is more based on HTML than previous versions. The system control was designed complete in HTML with the new CD-R/CD-RW software. It is possible to create easy and simple CDs. The Windows terminal service makes the access to shared Windows XP desktop with an terminal client like the VNC solution.

The **Service Pack 2** (SP2) for Windows XP needs about 900 mega byte of free storage space. Another system modification is the Security Center, which shows the status and settings of the firewall, and automatically updates and one additional anti virus program. The new memory function called data execution prevention protects the software code in the memory in front of manipulation like the insert of malicious program code to be executed, the protection works only with 64-bit processors.

The improved firewall now can detect waiting ports for connections and the definition of exception rules for various networkservices.

The browser was extended by a pop-up blocker; the file execution protection with information about the used download zone was revised . Downloaded programs inherit the zone information of the browser and warn before the execution from Internet files as well as with NTFS file system also before   executing   on   local   partitions.
 Versions are,
1.   Service Pack1
2.   Service Pack2
3.   Service Pack3

IV.   PROBLEM DEFINITION

The problem definition for this method is duplicated contents or duplicate pages may present in the document.Same content will be added in the presentation.Integrity checking is necessary before making the presentation.

V.   EXISTING SYSTEM

Existing System has many software such as Microsoft Power- Point and Open Office to help researchers prepare their slides. Tools only help them in the formatting of the slides, but not in the content. It takes presenters much time to write the slides from scratch. User have to knows the importance of each sentence in a paper is learned after that only they can prepare Slide.Slides generation for academic papers is a very challenging task.Thedrwabacks are:
Efficiency less,take more time read and prepares Slide,it is very challenging and difficult task preparing Slide.

VI.   PROPOSED SYSTEM

Our proposed PPSGen system, sentence importance assessment is one of the two key steps, which aims to assign an importance score to each sentence in the given paper.The score of each sentence will be used in the slides generation process.Sentences in the paper which are more similar to the sentences in the slides should be considered more important and higher scores should be assigned to them using the scoring method.Every paragraph can be generally divided into several parts and each part may be relevant to one section in the paper.The advantage are: reduce Manual Work,HighEfficiency,Time Consumption ,Slide Generate score of each sentence so more accuracy.

VII.   TESTING PROCESS

The purpose of testing is the discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a   product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or finished product It is process of exercising software with the intent of ensuring that the Software system meets requirement and user expectations and does not fail in unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

VIII.   SYSTEM IMPLEMENTATION

 Implementation is the most crucial stage   achieving a successful system and giving the user's confidence that the new system can be   workable and effective. Implementation of a modified application to replace an existing one. This type of conversation is easy to handle, provide there are no major changes in the system.

 Each program is tested individually at the time of development using the data has verified that this program linked together in the way specified in the program specification, the computer system and environment is tested to the satisfaction of the user. The system that has been developed is accepted and proved as satisfactory for the user. And so the system is going to be implemented very soon. A simple operating procedure is included so that user can understand the different functions clearly and quickly Initially the first step the executable form of the application is to be created loaded in the common server machine which is accessible to all the user and the server is to be connected to a network. The final stage is to document the entire system that   provides components and the operating procedures of the system.

IX.   CONCLUSION

 It is concluded that application works well and satisfy the needs. The application is tested very well and errors are properly debugged. Finally our proposed PPSGen system to automatically generate slides that have good structure and content quality from academic papers. Finally Our method attempts to generate draft slides of the typical type mentioned above and helps people to prepare their final slides.

REFERENCES

1. M.Y. Kan, "SlideSeer: A digital library of aligned document and presentation pairs," in Proc. 7th ACM/IEEE-CS Joint Conf. Digit. Libraries, Jun. 2006, pp. 81–90.
2. 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.
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. H. P. Edmundson, "New methods in automatic extracting," J. ACM, vol. 16, no. 2, pp. 264–285, 1969.
5. S. Modgil, N. Faci, F. Meneguzzi, N. Oren, S. Miles, and M. Luck, "A framework for monitoring agent-based normative systems," in Proc. 8th Int. Conf. Auton. Agents Multiagent Syst., 2009, pp. 153–160.
6. P. Nakov, A. Schwartz, and M. Hearst, "Citation sentences for semantic analysis of bioscience text," in Proc. SIGIR'04 Workshop Search Discovery Bioinformatics, 2004, pp. 81–88.
7. 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
8. 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.
9. S. Modgil, N. Faci, F. Meneguzzi, N. Oren, S. Miles, and M. Luck, "A framework for monitoring agent-based normative systems," in Proc. 8th Int. Conf. Auton. Agents Multiagent Syst., 2009, pp. 153–160.
10. Q. Mei and C.Zhai, "Generating impact-based summaries for scientific literature," in Proc. ACL, vol. 8, pp. 816–824, 2008