

AUTOMATIC GENERATION OF MINUTES OF MEETING USING DATA ANALYTICS

Mrs. Navya Damodar¹, Nandini R², Karthik J³, Gangu⁴, Malik Mahammed Shakir⁵

Assistant Professor, Department of ISE, AMC Engineering College, Bangalore, India¹

Final year B.E, Department of ISE, AMC Engineering College, Bangalore, India^{2,3,4,5}

Abstract: At meetings, a variety of outputs such as presentation materials, text minutes, video/audio data are generated. Both meeting participants and absentees can easily understand meeting minutes easily if these data are mutually referred and extracted important points. Therefore, focusing on participants' importance, we have proposed and developed the system that automatically produces Web pages by collecting and editing meeting data with much labor saving so as to easily understand a sequence of meeting. The importance submitted by participants are automatically collected and showed their status to participants in this system.

Keywords: Google Docs, OneNote, Lucid Meetings, Evernote, Agreedo, Minutes.io.

I. INTRODUCTION

Meeting minutes are documentation of oral communication in meeting in written form. The large number of discussions that occur in the meeting, may cause the reader cannot take essence of the meeting, whether participants or nonparticipants of the meeting. Particularly at parliamentary meetings, the large number of speeches and the long time in a meeting makes meeting minutes more elusive. The meeting minutes will be easier to understand if there is a structured summary of meeting minutes. However, writing manual summary of meeting minutes by stenographers and professional secretaries is time consuming and costly. This is a challenge of making an automatic summarization of meeting minutes.

II. LITERATURE SURVEY

[1] Development and Evaluation of a Minutes System Focusing on Importance in a Meeting: The study presented the Development and Evaluation of a Minutes System Focusing on Importance in a Meeting. This paper shows a variety of outputs such as presentation materials, text minutes, and video/audio data are generated. The main drawbacks It requires enormous efforts to collect data which are located in participants' PCs and to edit them to be mutually referred.

[2] A key phrase based approach to interactive meeting summarization: This paper proposes a summary can be either abstractive – expressing the content in newly formulated sentences or extractive by selecting relevant parts. The main Drawbacks of this key phrase based approach to interactive meeting summarization is that, the topic focus which might all be different per user. The other drawback of this project is, it often contains unimportant chit chat.

[3] Two-Step Sentence Extraction for Summarization of Meeting Minutes: Many or most readers want to grasp the important point of a meeting or the opinion of each member. For instance, the readers want to know who is positive or negative about a specific issue. In order to satisfy needs. The research for automatic summarization of meeting minutes was discovered. The main drawback is Recently digitalized and it is a very time consuming task to read whole contents of document. This is because most minutes are very long and even contain unnecessary utterances.

[4] A Virtual Meeting System for the New Age: The paper proposed, Virtual meetings and virtual collaboration are becoming commonplace in organizations Facilitation is an important factor especially for electronic meetings. In some factors that can improve the productivity. The main drawback of a virtual meeting system for the New Age is considering that the facilitation of virtual teams can be extremely difficult.

[5] Automatic Extractive Summarization on Indonesian Parliamentary Meeting Minutes: The paper proposed, Distill the most important information from a meeting minutes into a short textual passage. Most systems select relevant segments, which have proven successful in document summarization. The main drawback is writing manual summary of meeting minutes by stenographers and professional secretaries is time consuming and costly.

III. METHODOLOGY

BLOCK DIAGRAM:

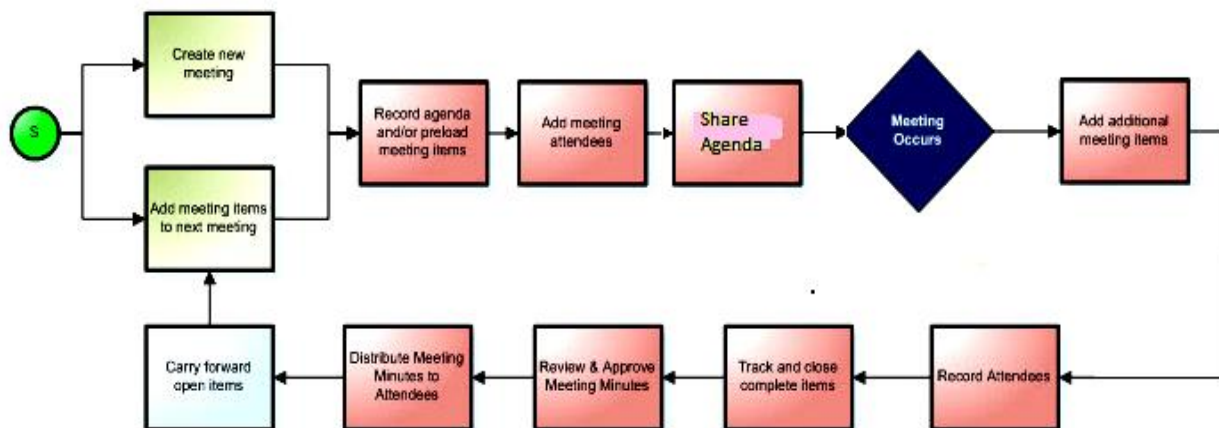


Fig.1 Block Diagram

The Meeting Minutes component allows the minute taker to record discussion topics within e-Builder. The review and approval mechanism ensures that all participants agree with notes and action items before minutes are published. Optionally, you can assign action items using an assigned form type. Action items are updated in real time, so all participants are aware of item status before the next meeting. The Meeting Minutes component also gives you the ability to carry open items forward from previous meetings, allowing you to keep track of your progress. The diagram above illustrates how meeting minutes can be used.

DETAILS OF COMPONENT:

Hardware Component:

1. Processor- intel® core™ i5-4200



Fig. 2 Processor- intel® core™ i5-4200

The Intel Core i5-4200U is an ULV (ultra low voltage) dual-core processor for ultrabooks launched in Q2 2013. It is based on the Haswell architecture and is manufactured in 22nm. Due to Hyper-Threading, the two cores can handle up to four threads in parallel, leading to better utilization of the CPU. Each core offers a base speed of 1.6 GHz, but can dynamically increase clock rates with Turbo Boost up to 2.6 GHz for 1 active core or 2.3 GHz for 2 active cores.

Haswell is the successor to the Ivy Bridge architecture with improvements on both GPU and CPU performance. The CPUs are produced in 22nm and offer an optimized branch prediction as well as additional execution ports, improving performance per clock by almost 10 percent. Furthermore, new features like AVX2 and FMA should increase the performance in future applications.

2. RAM 8GB



Fig. 3 RAM 8GB

With 8 GB of RAM, you will have enough memory to run several programs at once. You can open lots of browser tabs at once, use photo or video editing programs, stream content, and play mid-to-high-end games. Many Windows 10 and macOS computers or laptops come with 8 GB of memory installed these days. So, 8 GB of memory should be more than enough to run most productivity programs. It's also the minimum amount of memory recommended by Adobe to run Creative Cloud programs like Photoshop. While 8 GB of RAM is more than enough for most people, you can easily use it up if you have dozens of tabs open on your web browser, and you are running several programs at the same time. For example, we used up more than 8 GB of memory on a laptop by opening 50 tabs in Chrome while running Photoshop, PowerPoint, Word, and Spotify at once.

3. Storage requirements-64-bit OS



Fig. 4 64-bit OS

In principle, a 64-bit microprocessor can address 16 EiB ($16 \times 1024^6 = 2^{64} = 18,446,744,073,709,551,616$ bytes, or about 18.4 exabytes) of memory. However, not all instruction sets, and not all processors implementing those instruction sets, support a full 64-bit virtual or physical address space. The x86-64 architecture (as of 2016) allows 48 bits for virtual memory and, for any given processor, up to 52 bits for physical memory.^{[26][27]} These limits allow memory sizes of 256 TiB (256×1024^4 bytes) and 4 PiB (4×1024^5 bytes), respectively. A PC cannot currently contain 4 pebibytes of memory (due to the physical size of the memory chips), but AMD envisioned large servers, shared memory clusters, and other uses of physical address space that might approach this in the foreseeable future. Thus the 52-bit physical address provides ample room for expansion while not incurring the cost of implementing full 64-bit physical addresses. Similarly, the 48-bit virtual address space was designed to provide 65,536 (2^{16}) times the 32-bit limit of 4 GiB (4×1024^3 bytes), allowing room for later expansion and incurring no overhead of translating full 64-bit addresses.

SOFTWARE COMPONENTS

4. Python 3.9.0 version

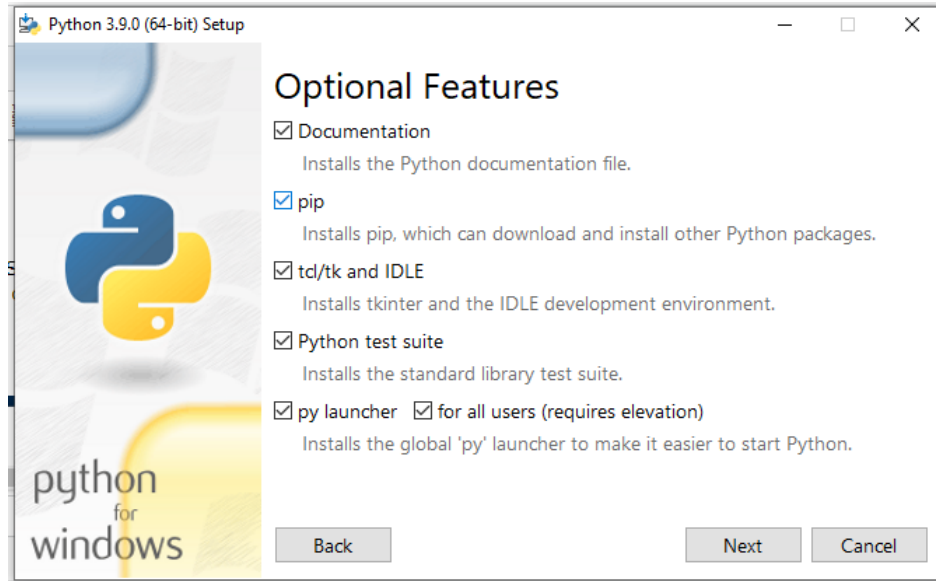


Fig 5. Python 3.9.0 version

The current CPython parser is LL based. Subsequently, the grammar is LL based to allow it to be parsed by the LL parser. The LL parser is a top-down parser. Furthermore, it parses the inputs from left to right. The current grammar is context-free grammar hence the context of the tokens is not taken into account. The Python 3.9 version is proposing to replace it with a new PEG-based parser which means it will lift the current LL grammar Python restrictions. Additionally, the current parser has been patched with a number of hacks that are going to be removed. As a result, it will reduce the maintenance cost in long run. As an instance, although the LL parsers and grammars are simple to implement, the restrictions do not allow them to express common constructs in a natural way to the language designer and the reader. The parser only looks at one token ahead to distinguish possibilities.

5. Anaconda-Jupyter Notebook

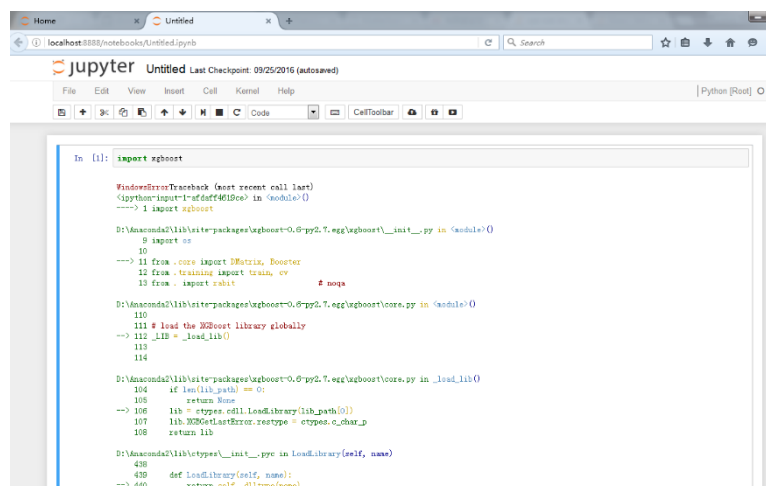


Fig 6. Anaconda-Jupyter Notebook

An open-source, web-based IDE with deep cross-language integration that allows you to create and share documents containing live code, equations, visualizations, and narrative text. Data scientists and engineers love using Jupyter for data cleaning and transformation, statistical modelling, visualization, machine learning, deep learning, and much more. Jupyter Notebook's format (ipynb) has become an industry standard and can be rendered in multiple IDEs, GitHub, and other places. Jupyter has support for over 40 programming languages, including Python, R, Julia, and Scala. Notebooks can be shared easily with others, and your code can produce rich, interactive output, including HTML, images, videos, and custom MIME types. It allows you to leverage big data tools such as Spark and explore that same data with pandas, scikit-learn, Tensor Flow, and ggplot2.

6. Open CV



Build a Web App With Streamlit

Fig 7. Open CV

OpenCV is the huge open-source library for the computer vision, machine learning, and image processing and now it plays a major role in real-time operation which is very important in today's systems. By using it, one can process images and videos to identify objects, faces, or even handwriting of a human. When it integrated with various libraries, such as NumPy, python is capable of processing the OpenCV array structure for analysis. To Identify image pattern and its various features we use vector space and perform mathematical operations on these features. The first OpenCV version was 1.0. OpenCV is released under a BSD license and hence it's free for both **academic** and **commercial** use. It has C++, C, Python and Java interfaces and supports Windows, Linux, Mac OS, iOS and Android. When OpenCV was designed the main focus was real-time applications for computational efficiency.

7. PyCharm (Accuracy)

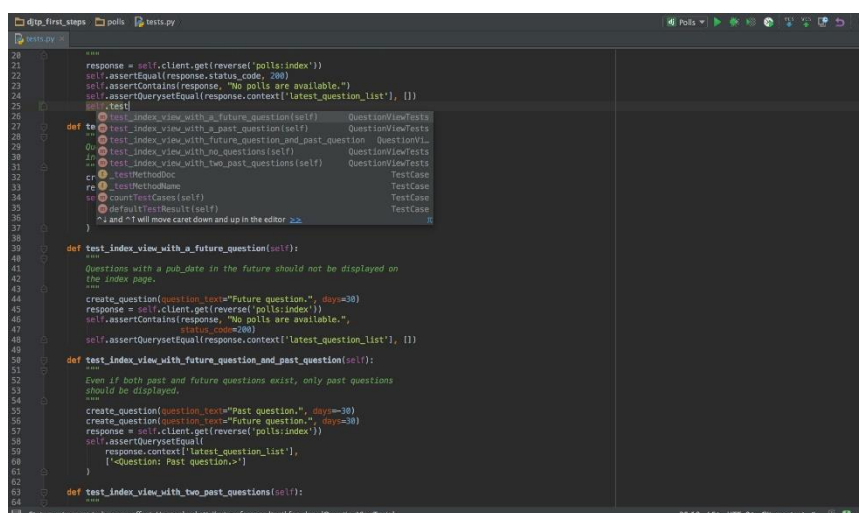


Fig 8. Pycharm (Accuracy)

This will use PyCharm Community Edition 2019.1 as it's free and available on every major platform. Only the section about the professional features will use PyCharm Professional Edition 2019.1. The recommended way of installing PyCharm is with the JetBrains Toolbox App. With its help, you'll be able to install different JetBrains products or several versions of the same product, update, rollback, and easily remove any tool when necessary. You'll also be able to quickly open any project in the right IDE and version. To install the Toolbox App, refer to the documentation by JetBrains. It will automatically give you the right instructions depending on your OS.

IV. PROPOSED SOLUTION

- Having a backup plan
- Send out a meeting agenda before the time
- Have a meeting plan with set instructions
- Make sure that you set a routine for your meetings
- Try to make all the meeting participants feel comfortable
- Allocate time during the meeting to ensure that everyone is keeping up with the events

IV. CONCLUSION

Data process from automatically collecting to uploading. Next, we consider the relation between generated text minutes importance and submitted participants' importance. In addition, we explain improved system for clearly showing what is important. Moreover, we state evaluation results and show the validity of our proposed system. We have not yet considered about the meeting minutes originally from the result of speech. Storage of member names in the database can be done to avoid errors. In further work, it will be considered to achieve good performance of meeting minute summarization. Moreover, the data set should be added more also to get higher performance.

REFERENCES

- [1] J. Carbonell and J. Goldstein, "The Use of MMR, DiversityBased Reranking for Reordering Documents and Producing Summaries," *Research and Development in Information Retrieval*, pp. 335– 336, 1998.
- [2] A. Janin, D. Baron, J. Edwards, D. Ellis, D. Gelbart, N. Morgan, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke, and C. Wooters, "The ICSI Meeting Corpus," in *Proc. ICASSP*, 2003.
- [3] E. Shriberg, R. Dhillon, S. Bhagat, J. Ang, and H. Carvey, "The ICSI Meeting Recorder Dialog Act (MRDA) Corpus," in *Proc. 5th SIGDAL*, 2004.
- [4] K. Riedhammer, D. Gillick, B. Favre, and D. Hakkani-Tur, " "Packing the Meeting Summarization Knapsack," in *Proc. Interspeech*, 2008, pp. 2434–2437.
- [5] X. Zhu and G. Penn, "Summarization of Spontaneous Conversations," in *Proc. Interspeech*, 2006, pp. 1531–1534.