Increasing Performance through Information Retrieval System for Design and Build Company

Panot Khongkhaluang[1] Phannachet Na Lamphun[2], and Adisorn Kheaksong[3]

[1]Index International Group PCL, Pathumthani, Thailand
[2,3]Faculty of Engineering and Technology, Panyapiwat Institute of Management
Nonthaburi, Thailand
E-mail: 6272100116@stu.pim.ac.th, phannachetnal@pim.ac.th, adisornkhe@pim.ac.th

Received: February 8, 2021 / Revised: March 26, 2021 / Accepted: March 31, 2021

***Abstract*—Developing a progress report is a requirement in many types of business, including design and build business. This is because the purpose is to provide information on the progress of the project and other necessary information, such as problems, solutions, etc., for the project owner to be aware of the status of the project. The progress report is also linked to the payment of the project budget based on the progress. However, the report preparation of project might require various information from more than one parties; thus, the process would certainly be time- consuming. Moreover, the process might be redundant and prone to error when information is extracted from one source and incorporated in other. The Objective of this research were to analyzed, design and develop the automatic document report system for Design and Build Business and to evaluate the performance of the system. In the design of this prototype system, information is automatically extracted from submitted reports prepared by various working parties and, finally, generated the progress report for the contractor to submit to project owner in an effective and efficient way. The prototype system could help reduce the process time and work load on the staff, while ensuring the correctness of the information and document.**

***Index Terms*—Information Retrieval, Automatic document Generate, Data Extraction, Design and Build, Progress Report**

I. Introduction

Collaboration is an important key to success for Design and Build business. Design and Build business is a unique construction business that provides a wide range of services including design service for buildings, facilities, utilities, etc., and construction services for the project. The service also includes management, monitoring and controlling of the project. The size of the Design and Build projects, such as construction building, airport, mass rapid transit, requires the system to incorporate sub-contractors working together in their parts to complete the projects. For example, the 1st sub-contractor is responsible for the ground clearing and pavement, the 2nd sub-contractor for office building construction, the 3rd sub-contractor for depot station operation and construction, the 4th sub-contractor for providing and integrating the system for operation between the office building and depot station, etc. By involving many parties to complete the project, a collaboration between parties is very important to ensure that each party would know the status of the project and what to be carried on.

Collaborating to complete the work may seem simple, but it can be a challenge to achieve because sub-contractors can be from different companies and from various places. Normally, even people working with other people in one company, communication among them could be complicated. Thus, this level of people working with other people in different companies to complete the project could be even more challenging. Document format is another problem when working together as well. The document must be accessible to all parties and must comply with the project owner. The document report is certainly crucial in the project because it states the progress of the work, and is used to withdraw the budget from the project owner. However, the preparation of the document must be well prepared and this can be difficult when working with more than one parties. The main contractor must gather necessary information from sub-contractors and prepare a final report to present to the project owner. In collaboration, time is one factor that effects many aspects of the project, significantly, when it concerns payment withdrawing schedule. Each month the main contractor must submit a progress report of the work to the project owner, and if the project progress is up to a certain percentage, then the main contractor will

be able to withdraw the budget to carry on the project. However, failure in report submission would result in the delay of payment, causing delay of the whole system and then back to more delay of payment.

Therefore, collaboration is crucial in working together for Design and Build business. By sending paper reports to the main contractor to generate one whole report is not effective because the process can be time consuming. The information from sub-contractors might be incomplete with some details missing can affect the whole report. In this aspect, technology becomes a key to success in increasing the collaboration performance. The purpose of this research, thus, aims to study the need of such system to help increase collaboration and reduce time consumed in developing report for Design and Build Business. The system of automatic information retrieval is applied to gather information from the progress reports of sub-contractors. Then, the information thus incorporated is automatically generated as a final progress report, as well as a presentation. This system can help reduce the time of work, while monitoring the progress and informing each party of any necessary information.

II. Literature Review

Design and Build business is a unique business that contains all services of a construction project for customers. It can be a new construction project or a renovation of the existing one. Design and Build business provides design and construction services under a single design and build contract [1]. The benefit of this type of business is that the customer signs a single contract with the service provider, then the service provider provides all services including engineering, architectural work, labor, materials, etc., while still focusing on the time, cost, and quality. The main advantage of Design and Build includes faster delivery due to collaborative project management, cost saving by integrated team toward efficiency and better quality. Other advantages include singular responsibility, by having one company as the key contract for all the work; thus, it can reduce burden of contract and management. It also reduces risk from management and miscommunication [2]. However, in a very large scale of project, the sub-contractor is sometimes required to be able to complete the project in a shirt time. So, the main design and build business needs to outsource labor, engineering, technology, etc. from sub-contractors in order to manage and complete the project efficiently and in specified time.

Document automation can be used to support work process collaboration by reducing the time and redundant process. The key factors for collaborative document are the motivation competence and the trustworthiness [3]. Though there are various technologies that provide and support communication in various ways, documents are the most preferred and effective ways to communicate engineering information [4]. In addition, many managing software documents are available in the market; however, due to the different processes, formats, and natures of work, this often results in many of processes still requiring intensive manual operations. Therefore, for the system to be able to support work process, the system has to be smart enough to know which information is required for use in other parts of the report. There are two aspects of document automation [5]:

- Automating processing of received documents: this is the process where the system automatically collects required data from different types of documents and make use of them.
- Automatic generation of documents: this is the process where the system creates new documents like invoices, contracts, and order forms from the data extracted in the previous stage.

In order for the system to be able to understand the content, the document need to be structured. The structured document is well organized document that the system can understand to be able to process and create output.

Once document format is structured, the system can process the data. The research on Inter-Governmental Collaboration through E-Document Computation: A Case Study of Provincial Budgeting in Thailand shows how documents could be extracted and automatically generated to support government work [6]. The system was developed using semantic web and ontology to extract data from a number of submitted projects in the province. The system checked the similarity of submitted projects to prevent duplicated ones, and then automatically generated a final report in the Word file. The research shows that, with structured documents, the content can be extracted and reused to create a final report. Ontology is also needed for the system to understand the relation of individual words to compare similarity. Semantic web has been used widely in document retrieval [7]. By creating the document through the semantic process, documents are managed by both people and procedure. This helps the system able to understand and represent knowledge. In addition, the semantic information can be applied to multimedia data [8].

Other example of document automation is using document automation system in law firms' knowledge management [9]. The research focused on implementation of knowledge management and resourcing. The study shows that a well-prepared document can enable the automation effectively. Moreover, through document automation, the system becomes a powerful tool for internal knowledge development and enable collaboration. The result

from this research study helps ensure that the developed system for project progress report can help with collaboration and knowledge management for design and build companies.

III. Methodology

The system is developed based on the offline process, where each sub-contractor creates their own progress report and submits it to the main contractor. After the main contractor has received the report, the contractor would review the work progress and summarize the information to generate the main progress report to present to the project owner as presented in Fig. 1.

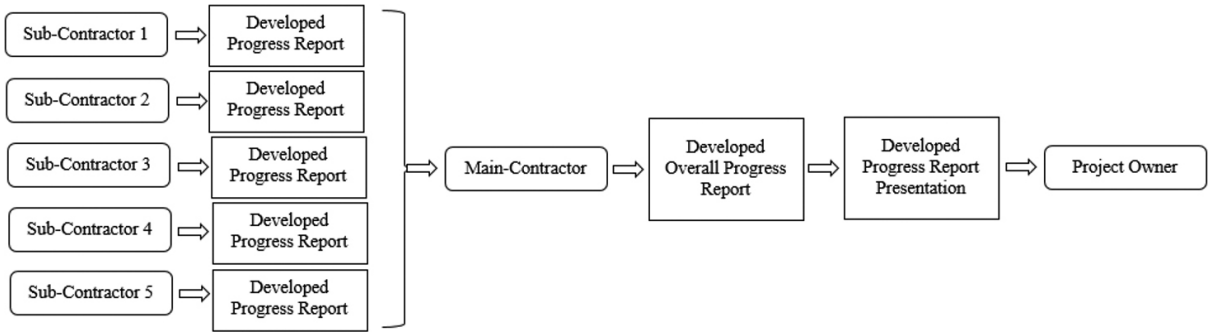

Fig. 1. Process of developed progress report

The sub-contractors develop their progress reports based on the performance of that month and submit them to the main contractors. After reviewing and inspecting the progress, the main contractor would create the overall progress report of that month to submit to project owner. The contractor would receive the budget payment if the project progress were on track. The problem could arise when the sub-contractors submit the wrong information. As a result, the main contractor would have to wait for the sub-contractor to resubmit the revised progress report, so the main contractor could continue the process.

The system is developed in the offline system to protect information of the business. To reduce the process, the system is developed to accept report documents as input data. Then the system will go through the document and extract data as assigned and put in the overall progress report. The system is designed to use the keyword on the subject to as a checkpoint to which section of the document the information would go. After completing the first sub-contract progress report, the system would proceed to the next one and continue the process until completing all sub-contract progress reports as shown in Fig 2.

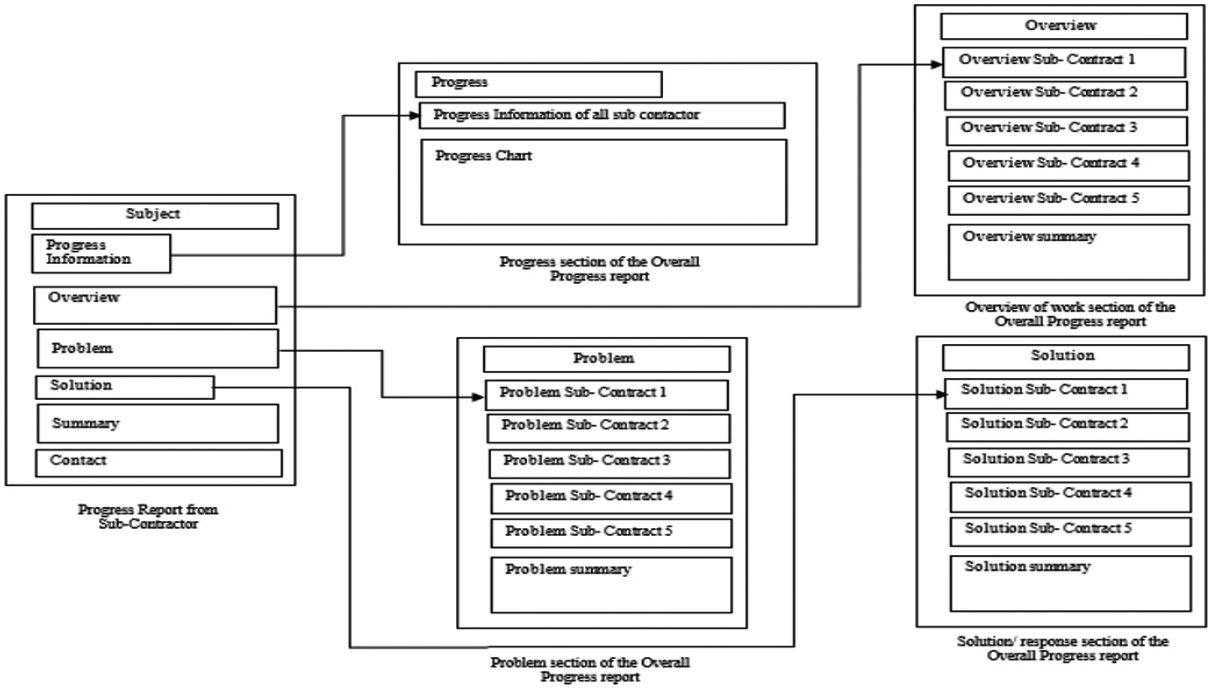

Fig. 2. Process of automatic data retrieval

The system would then take the progress inputs from all the sub-contracts to calculate the overall progress and plot the data in the graph form. This would account for how the business progresses in that month. The system would also compare the actual progress of that month to the progress planned at the beginning. This is to inform all the parties that the project in on track, ahead of the track, or behind the track. The information would show which sub-contact is behind the schedule, so the main contractor could discuss and plan to have the project back on track.

The next step is to extract content from sub-contractor's progress report and put in the correct section of the overall progress report. Since the format of the report follow the format from the project owner, the subject or title of each section could be used as a checkpoint for the system to recognize where the following content would go. For example, in the Problem and Threat section where the sub-contractors describe problems and threats during the work progress and how the sub-contractors handle the issues. The system would search the keyword "Problem and Threat", and then extract information of that section and put in the overall progress report's problem section. The same process would be repeated for the next subject/title until the whole progress report document is completed. After the completion of the first document, the system would continue to extract contents from other progress report documents of all sub-contractors. However, only extracting and incorporating information from one document to the rest still would not support collaboration. The System would also check the contents for similarity. The purpose of similarity check is to reduce redundant data that could appear in the overall progress report. This features aims to help the main contractor to collaborate with sub-contractors more efficiently. If the similarity of two contents reach the assigned percentage of similarity, the system would identify such similarity, so the main contractor could choose which content to put in the overall progress report. When this occurs, the main contractor would know that there are several sub-contractors possibly with similar problems, and that would allow the main contractor to collaborate to solve the problem. The system would also inform the main contractor whether such problems have occurred before and how the main contractor has chosen to solve it. This is to reduce the time and to apply the previous knowledge gained as a guideline to solve similar issues.

After the overall progress report document is generated, if there is no further change or adjustment, the user can create the PowerPoint presentation through the system. The process is the same by extracting information from overall progress report and put such information in PowerPoint presentation as shown in Fig. 3.

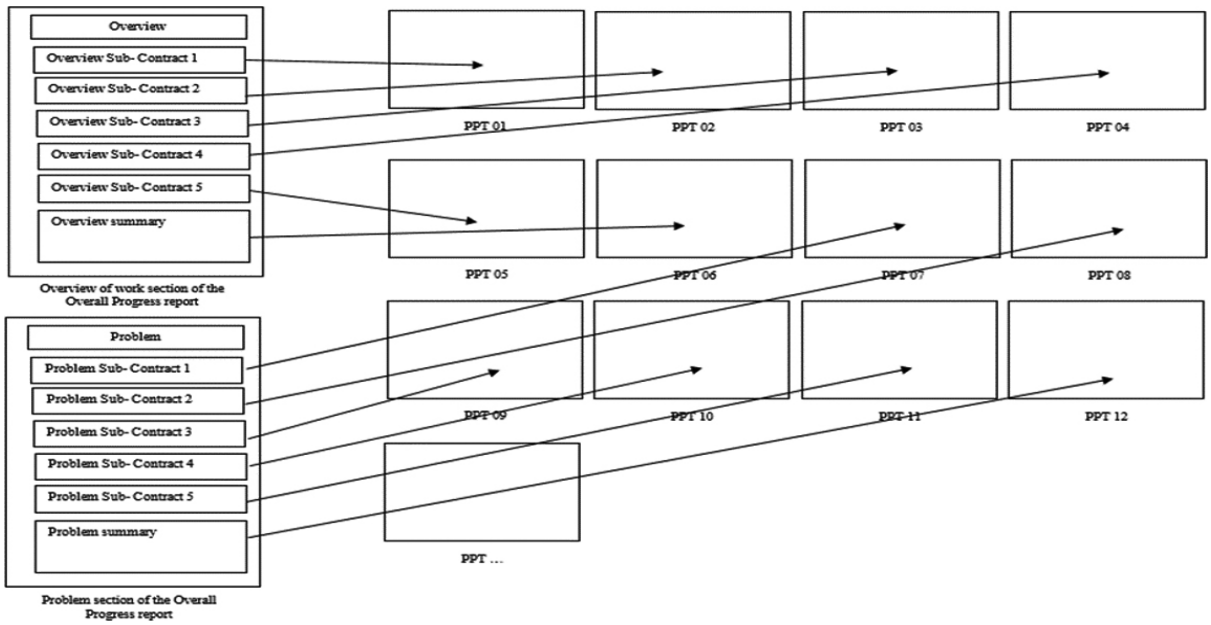

Fig. 3. Process of extracting information on the report and generate presentation file

IV. Experiment

Design and Build Business operates on a high budget scheme, for management and construction operations. However, the budget would be paid based on the performance and project progress. This would be reported for the project owner's inspection. This would be the first problem because means any errors in the document might cause a delay of payment. The process of developing overall progress report can sometimes take a long time because all other

sub-contracts are required to submit their own progress reports to the main contractor. Then, the main contractor would inspect the report and create an overall progress report to submit to the project owner. This process of sub-contractors creating the progress report normally vary based on the nature of each sub-contractor's work. If the submission of progress report to the main contractor were late on schedule, then the main contractor would have little time to review and create the overall progress report. The second problem is that if there was a mistake or error on the progress report, then the sub-contractor would have to review and correct the information before resubmitting the progress report to the main contractor. Such process can be time consuming, in particular, when the main contractor need to make any corrections based on the revised information obtained from the sub-contractor. Any delays might result in late payment from the project owner, which might affect the work plan of the company.

In the experiment, the process started after the sub-contractors submitted their progress reports to the main contractor. The main contractor started the process by uploading the document files into the system as shown in Fig. 4.

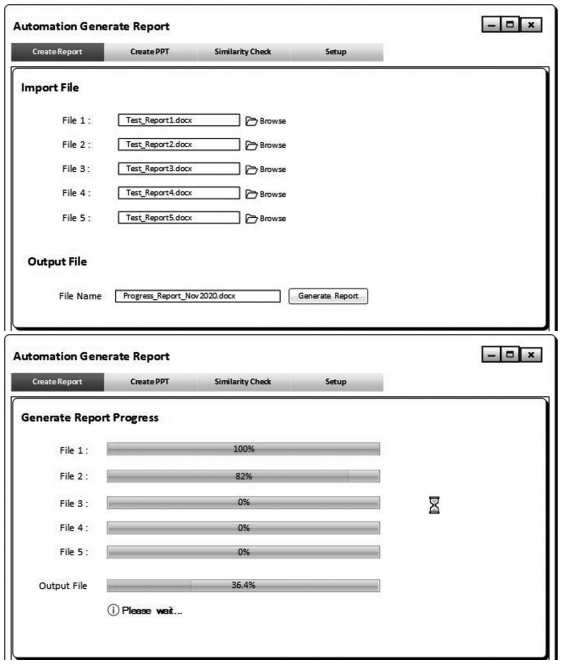

Fig. 4. Main screen for upload file

The system automatically extracted information from submitted documents and stored the information in the database for use when creating the overall progress report. The system used two-keyword search; the first keyword was used as a beginning point where the information was initially extracted, and the second keyword was used as an end for the system to stop extracting information for each section as shown in Fig. 5.

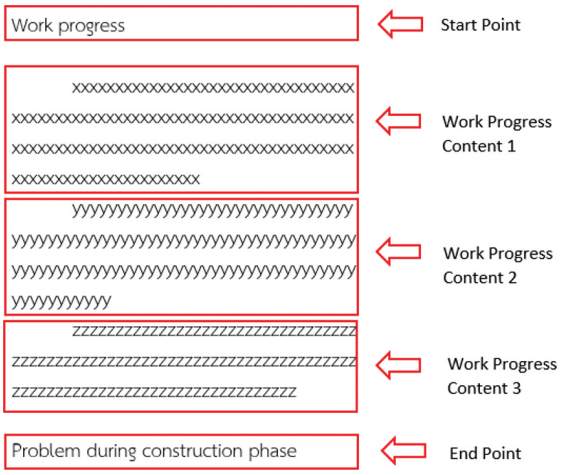

Fig. 5. Use the keyword as a start point and end point for content extraction

The System continued extracting information from submitted document until it reached the end of document. Then the system applied the process to the following documents. However, for the other submitted documents beside the first one, the system applied the keyword and the content similarity check to avoid putting similar contents in the same report. The process of this part started from separating words from extracted contents, then comparing the later contents with the content from the first submitted document. If the similarity exceeded the setup percentage, then both contents with similarity were highlighted in red to indicate the section that was similar to the previous one. The process continued until all the documents were completed as shown in Fig. 6.

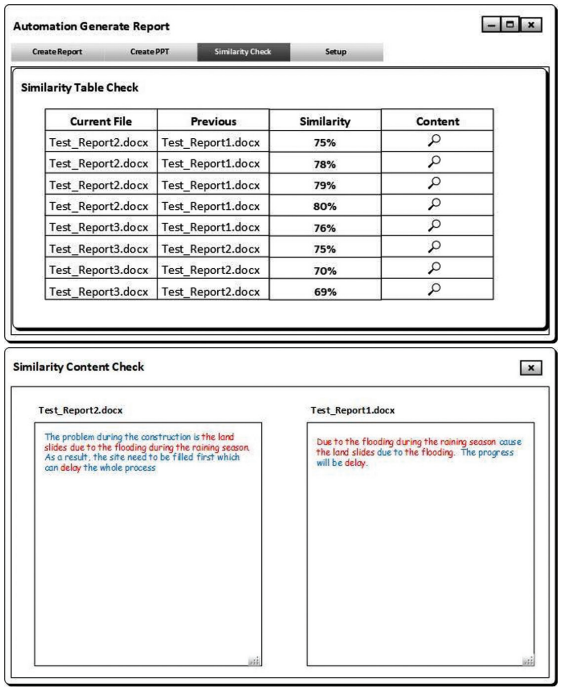

Fig. 6. Content similarity check

After overall progress report was generated, the system presented the option for the user by uploading contents in the document so that preferred contents could be chosen for a PowerPoint presentation as shown in Fig. 7. Thus, the system ran the same process of extraction, but put the information content in PowerPoint file, instead of in a Word file.

The system was also capable of creating a summary of the progress report document automatically, which helped reduce the time and workload for staff. The system Instead of staff reviewing each document and summarizing the content into the report, the system extracted the contents in the same category, putting them together and highlighting similar contents. Therefore, the time needed to perform the task was reduced. Moreover, once the content in the overall progress report was settled, the system automatically generated the PowerPoint file to reduce the process for staff.

V. Evaluation

The system is capable of automatically creating a summary of the overall progress report. The system uses keywords to determine which sections to start extracting the contents and where the information should be incorporated in the report. The system was evaluate based of the performance whether the Collaborative technology has improve outcome performance while reducing coordination, communication, and process costs [10]. The experiment started with having the staff review five test reports and then create a summary report based on those five report. The data collected included the time for human to complete the task, the correction of contents, the correction of format, as well as ability to distinguish similar contents and to summarize it. Thus, the evaluation of this research was the outcomes of monitoring the performance of the system and data from users as shown in Table.

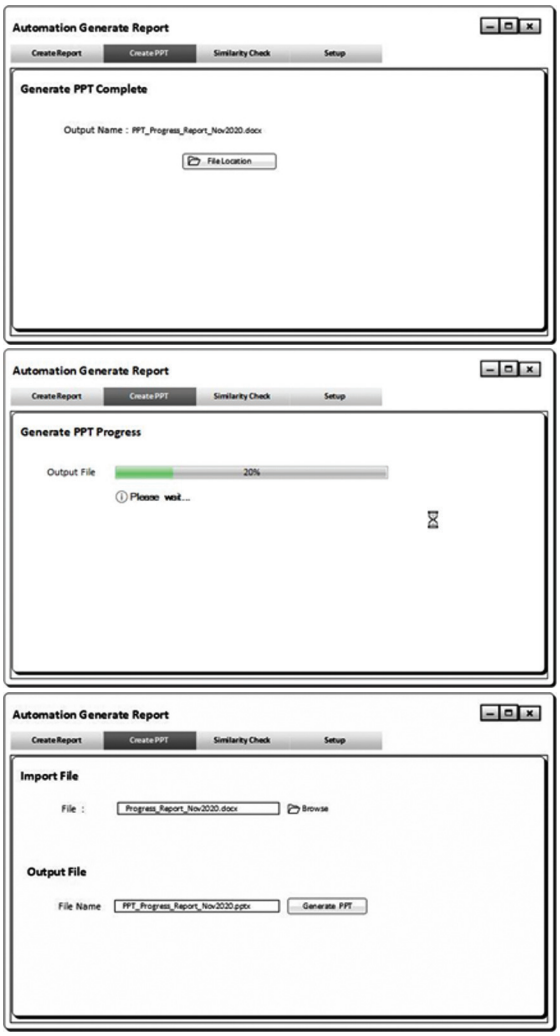

Fig. 7. Create PowerPoint file from Word content

TABLE I
THE PERFORMANCE OF SYSTEM ON AUTOMATIC DOCUMENT GENERATOR

Criteria	Traditional Process	System Process
Create Summary report from 1 document • Process Time • Format Correction • Content Correction • Similarity Check	 18 Minutes 100% 95%	 3 Minutes 100% 100%
Create Summary report from 2 documents by combining similar contents • Process Time • Format Correction • Content Correction • Similarity Check	 42 Minutes 90% 90% 3 of 3 Similarities Found	 6.5 Minutes 100% 100% 3 of 3 Similarities Found
Create Summary report from 3 documents by combining similar contents • Process Time • Format Correction • Content Correction • Similarity Check	 73 Minutes 90% 75% 5 of 7 Similarities Found	 18 Minutes 100% 105% 7 of 7 Similarities Found
Create Summary report from 4 documents by combining similar contents • Process Time • Format Correction • Content Correction • Similarity Check	 104 Minutes 75% 70% 7 of 11 Similarities Found	 33 Minutes 100% 100% 10 of 11 Similarities Found
Create Summary report from 5 documents by combining similar contents • Process Time • Format Correction • Content Correction • Similarity Check	 145 Minutes 75% 70% 9 of 13 Similarities Found	 47 Minutes 100% 100% 11 of 13 Similarities Found
Create PowerPoint Presentation • Process Time • Format Correction • Content Correction	 55 Minutes 80% 70%	 27 (+10) Minutes 100% 100%

From collected data in Table I, when working with a single file, the process time was slow because staff had to copy the contents from the submitted file and put it in the correct format of the summary report file. When working with one document, the format and content did not have to be recheck, so the percentage of correction was 100% for staff. However, as the number of submitted documents increased, the time to process increased as well, resulting in staff completing the task in more than two hours. When working with multiple documents, the format and content correction rate tended to drop because staff had to focus on many contents at one time; moreover, exhaustion could create human error. As for the system, it could still perform, as it should, even with the increasing number of documents. The processing time of content extraction, of course, increased with each additional document because there was more work to process and more similarity to recheck. As for content similarity, staff tended to have more difficulty in spotting similarity as the number of documents increased. As for the system, the similarity found rate dropped as well, in particular when inspecting the case of two similarities that the system was unable to spot. This was because the similarity rate was lower than the setting, as the keyword was different word but had similar or relate meaning. However, the processing time was still lower than human process. As for PowerPoint file creation, the system performed with a low processing time and higher rate of correction than traditional process (human staff). However, human recheck was required to check some of the slides because the content extracted from the system might be too large to put in one slide. In the performance test, it took the system 27 minutes to create the PowerPoint file and the staff required around 10 minutes to go over the slides and make some corrections. [11]

In addition, it was necessary to evaluate the effectiveness and benefit of the system it had for the staff and organization. This required a set of questionnaire containing two parts: the first part with a set of Likert-scale questions and the second part open for comments. The questionnaire was distributed to seven staff members involved in the progress report. The result of the effectiveness evaluation is shown in Table II.

TABLE II
THE EFFECTIVENESS OF SYSTEM ON AUTOMATIC DOCUMENT GENERATER

Criteria	Score	Level	Comment
The system reduced the process and time of creating Overall Progress Report.	4.43	Agree	• The System processing time should be shorter. • The System should automatically extract content-related pictures from reports and incorporated them in the report.
The format of the Overall Progress Report generated by the system was accurate.	4.71	Strongly Agree	-
The content of the Overall Progress Report that generated by the system is accurate.	4.57	Strongly Agree	-
The system helped organize the contents to reduce similar contents to appear in multiple places in the report.	4.29	Agree	• User still needed to check similarity contents.
The Similarity check help staff to be able to seek the solution of the same problem that appeared before.	4.29	Agree	• The similarity check helped find solution from other cases • The system should store content, which can be used to find solution for similar problem; this would be even more helpful.
The system reduced the process and time of creating Overall Progress Report Presentation.	4.43	Agree	• The system should have the capability to separate long contents into multiple slides.
Over all the system met the needs of users to complete the work.	4.71	Strongly Agree	-

From the data collected, all staff agreed on the effectiveness of the system. They all agreed that the system helped reduce the process and time in developing the Overall Progress Report. However, the staff felt that the system should complete the task much faster. However, the system generated contents in the documents without installing them in the system. If the system could save data and recall them when needed, the process could be faster. The staff strongly agreed that the format and content of report were accurate as it should be. They also agreed that the similarity function was useful in helping staff to spot similar contents and to summarize those points. The staff reported that the system reduced time and process in creating PowerPoint file from the content in MS Word file. However, the system should improve performance by separating one long contents so that several smaller content files could fit in the Slides. Overall, the staff strongly agreed that the system met their needs and helped them to complete the task more efficiently and effectively.

VI. Conclusion

From this research study, the prototype system can help reduce process and time, which is necessary for organizations that have to complete the task in time for withdrawing payment phase of the project. The information extraction is necessary to the task through replacing manual extraction and generating report. The system could do this automatically to reduce workload for the staff. By doing so, the system can generate the report with the correct format and content for the staff. The process that might require more than two hours or even a day can be reduced into less than an hour. Besides helping the staffs to complete the report, the system helps notify the staff of the similar contents, such as similar problems that occur among different sub-contractors. The main contractor can spot the same problem and help resolve such problem for sub-contractors. This can also alert the main contractor if similar problems appear among multiple sub-contractors. These might be important or urgent problems that have to be solved before they can cause damage or delay to an organization.

VII. Future Work

This research study has developed a prototype system to help reduce the process of developing the report, which can sometimes be time consuming. The result met the expectation of people involved in the process. However, there is still much room for improvement, in particular, how to enhance the system to extract and incorporate pictures in the final report. The system should be capable of extracting data from submitted documents and store them in database instead of extracting and creating the final report instantly. This will reduce the time of creating the final report. If there is an adjustment in one of the submitted documents, then only one document is needed to extract the information, not redoing all the documents. Storing data in the system becomes information source for staff to find the solution or

process in previous similar problems. Of course, the current system can point out the similarity of content within the documents submitted for process at the time. However, in reality, the problem that occurred in one sub-contractor report might happen in the past operation or in the future operation. So if the system is stored the database, this will enable staff and people involved to react to the problem more effectively.

References

[1] Design-Build Institute of America. (2021, Jan 10). *What is Design-Build*? [Online]. Available: https://dbia.org/wp-content/uploads/2018/05/Primers-What-is-Design-Build.pdf

[2] Design and Build Institute of America Rocky Mountain Region. (2021, Jan 10). *What is Design-Build*? [Online]. Available: https://www.dbiarockymountain.org/what_is_design build.php

[3] J. Beel and B. Gipp, "Collaborative document evaluation: An alternative approach to classic peer review," in *Proc. World Academy of Science, Engineering and Technology, 2008*, pp. 353-356.

[4] L. Stein and D. Sharon, "Trends in automating document generation," *IEEE Software*, vol. 12, no. 5, pp. 116-118, Sep. 1995.

[5] AIMultiple. (2021, Jan 10). *The Ultimate Guide to Document Automation in 2020.* [Online]. Available: https://research.aimultiple.com/document-automation/#what-is-document-automation

[6] P. N. Lamphun and V. Wuwongse, "Inter-Governmental Collaboration Through E-Document Computation: A Case Study of Provincial Budgeting in Thailand." *International Journal of Information Communication Technologies and Human Development*, vol. 4, no. 1, pp. 1-23, Jan. 2012.

[7] A. Celentano, M. Grazia Fugini, and S. Pozzi, "Knowledge-based document retrieval in office environments: the Kabiria system," *ACM Transactions on Information Systems (TOIS)*, vol. 13, no. 3, pp. 237-268, Jul.1995.

[8] G. Amato, G. Mainetto, and P. Savino, "An Approach to a Content-Based Retrieval of Multimedia Data," *Multimedia Tools and Applications*, vol. 23, no. 4, pp. 9-36, Jul. 1998.

[9] R. Lankester, "Implementing Document Automation: Benefits and Considerations for the Knowledge Professional," *Legal Information Management; Cambridge*, vol. 18, no. 2, pp. 93-97, Jun. 2018.

[10] G. L. Klein and L. Adelman, "A Collaboration Evaluation Framework," in *Proc. 2005 International Conference on Intelligence Analysis*, McLean, Virginia, 2005, pp.1-6.

[11] P. Preelakha, P. Na Lamphun, and J. Qu, "An Effect of Gamification on Education under Online Lives Teaching. (2020, Jul.-Dec.). *International Scientific Journal of Engineering and Technology* [Online]. *4*(2), pp. 20-27. Available: https://ph02.tci-thaijo.org/index.php/isjet/article/view/241736/164589

Panot Khongkhaluang received the Bachelor of Science (Information Technology), Panyapiwat Institute of Management, Thailand. Past work experience on acting IT Manager at Index International Group Public Company Limited. (IND).

Phannachet Na Lamphun received the Master of Science (Computer Engineering), Polytechnic institute of New York University, New York, USA and Doctor of Engineering (Information and Communications Technologies), Asian Institute of Technology. Past work experience is at Index International Group Co.,Ltd. as a system engineer for Project Management Consultation of MRTA Blue Line, Dustfree Road Project.

Adisorn Kheaksong is an assistant professor with Computer Engineering and Artificial intelligence program, Panyapiwat Institute of Management (PIM), Thailand. He received his B.Ind. Tech. degree from Saint John's University in 2001 and his Ms.CS degree from Rangsit University, and Ph.D. degrees from King Mongkut's University of Technology North Bangkok (KMUTNB), Thailand. His research interests include wireless sensor network, cognitive radio communication and Artificial Intelligence (AI) on wireless communication.