

¹ Yotaka Seniwong
na Ayudhya

² Parames Chutima *

³ Panasin
Chungsawanant

Leveraging Text Mining Techniques to Disclose Material Errors in Engineering Build Request Forms



Abstract: - In the electronics industry, driven by continuous innovation, the preparation of precise engineering build request forms (EBRs) is critical for ensuring product quality and efficient production processes. Errors in these documents can lead to significant financial losses, including penalties and project cancellations. This research aims to enhance the error-checking process in EBRs by applying text-mining techniques through Python programming, integrating Optical Character Recognition (OCR) and automated text processing to improve inspection accuracy and efficiency. The methodology involves comparing the performance of human inspectors with the automated program using paired t-tests to assess performance. The results demonstrate that the automated program significantly reduces inspection time and increases accuracy, thereby minimizing production errors and associated costs. This advancement is anticipated to substantially improve the efficiency and reliability of document preparation in the production process, leading to more accurate and timely product outputs.

Keywords: Engineering Build Request Form, Text Mining Techniques, Automatic Detection integration, Optical Character Recognition technology, Text Processing.

I. INTRODUCTION

The electronics industry has introduced new technology and innovations to improve product quality and create new products. Hence, the preparation of engineering request documents is an important tool for communicating product details in terms of technical data, requirements, and limitations related to the manufacturing process, which will directly affect the accuracy of the product, efficient customer service, preparation of machine capacity in each process, and preparation of raw materials used in production such as Wafer, Lead frame, Mold Compound, and Epoxy, to prevent validation or qualification problems[1]. Therefore, this research aims to improve the inspection of errors in engineering production request documents by using text-mining techniques to improve the quality of the inspection process and prepare production documents more efficiently [2,3].

Based on our experience in verifying and refining programs for operating drones and satellites using an embedded system built with VHDL and Python, we believe that our research can be extended to detect errors in Engineering Build Request Forms (EBRs). Additionally, we plan to review other relevant research, the majority of which will utilize text mining theory to analyze, verify, and summarize documents. The details are as follows.

Miner et al. [4] leveraged Text Mining in Property and Casualty Insurance and applied the Text Mining method to analyze insurance claim data, with error detection and loss cause identification. The technique used is Machine Learning (ML) in the form of Natural Language Processing (NLP), which allows for in-depth data extraction from various data. Kim & Yoon [5] summarized multiple patent documents using the Transformer and Long Short-Term Memory models. The results indicated that the model could summarize multiple patent data accurately and easily. Djedović et al. [6] developed business processes and document systems using automation and simulated and compared the operation of the old and new systems, which resulted in solving the bottleneck of the system operation.

From the research mentioned above, tools are used to facilitate document processing to reduce work time. However, in the case of the manufacturing system, there is no document verification. If there is an error in the Engineering Requirements form (EBRs), it can cause serious damage such as fines, lost opportunities, and reputation, etc. Therefore, this research has introduced an automatic document management system in the EBR document section of the production process. It uses Optical Character Recognition technology (OCR), Text Processing, and Automatic Detection integration. Then, compare the results before and after improving the EBR document management system of the production process. This can solve the problem of document inspection errors caused by human error and reduce various impacts from errors. Also, it can reduce production errors and increase

¹ Department of Industrial Engineering, Faculty of Engineering, Chulalongkorn University, Thailand.
6670202721@student.chula.ac.th

² Department of Industrial Engineering, Faculty of Engineering, Chulalongkorn University, Thailand.
Academy of Science, The Royal Society of Thailand, Thailand. Parames.C@chula.ac.th

³ Graduate School of Human Resource Development, National Institute of Development Administration. Panasin.chu@nida.ac.th

* Corresponding Author Email: Parames Chutima

Copyright © JES 2024 on-line : journal.esrgroups.org

efficiency in the production process, resulting in accurate production of products in a shorter time and reducing production costs in the long term [7].

II. METHODOLOGY

From the study of the work process of those who prepare production documents and check errors in engineering build request form using human eyes, it was found that the average time spent checking documents is quite long and there are errors caused by human errors that enter the production line. Hence, the Python program and text mining techniques were used to detect errors in the documents, and tests were conducted to compare work efficiency by recording the time taken and accuracy between human document checkers and program document checkers. Statistical hypothesis testing was employed to analyze the results [8].

A. Text Programming

Since the customer's document is in PDF format, the document must first be converted from PDF to text. Then, the Python program will be used to collect important information from the document. The program will find the key that needs to be checked and then store the text after the key in the document. After extracting all important parts from the document, the data will be checked to see if the conditions are met. The conditions for each key are different, such as the size must be within a certain range and cannot be blank [9]. An example of the result is shown in Fig. 1. In the program performance testing section, real documents are imported into the program and the results are checked by recording both human visual document testing and program verification.

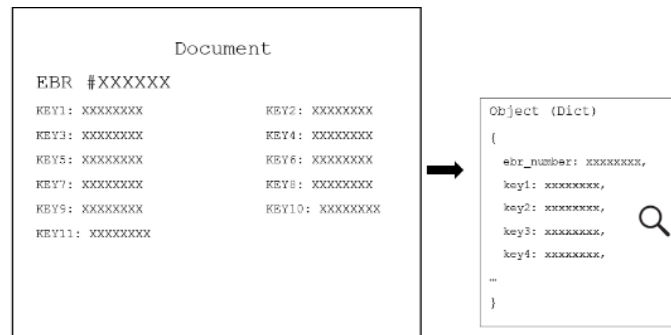


Fig. 1 The collection of important information from the program documentation.

III. CASE STUDY COMPANY

All The case study company is one of the leading independent companies in Southeast Asia. It has 7 manufacturing facilities in Thailand, China, USA, Cambodia, and South Korea and employs over 10,000 people worldwide. The company is a full-service electronic manufacturing company, "Electronic Manufacturing Service" (EMS), which operates under international standards.

The most important department in checking EBRs and preparing production documents to be able to produce new products and improve the quality of existing products correctly and completely is the Project Management NPI team in the Special Work Requirement section or the SWR team, which has the following duties and work processes as shown in Fig. 2.

- Receive Engineering Build Request forms (EBRs) from customers' systems, which will specify technical data and special work requirements.
- Download the documents to analyze and check the technical data in depth, including incomplete request data, the correctness of the request documents, production limitations, the number of raw materials in inventory (WIP), the process used in production, data collection, packaging conditions, and plan the time it will take to process from document inspection until the product is delivered to the customer, etc.
- Visually check the accuracy of the documents. If there are any errors in the documents, summarize the incorrect information and send it to the requester to correct and confirm the information via email.
- Take the complete Engineering Build Request form (EBRs) to prepare production documents, including Build sheet, Marking, Special Work Requirement and issue Run Card (R/C) documents to be entered into the production line.

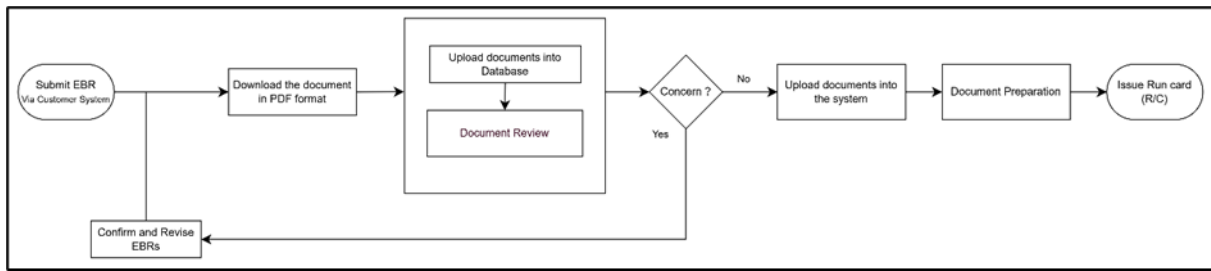


Fig. 2 Process flow of detecting incorrect information from EBRs

A. Define Problem

Currently, in the case study company, there are many engineering work request documents in the New Product Innovation (NPI) section when introducing new technologies into the product, reducing costs by using other raw materials to try to produce as a replacement, producing for product validation from changing the production site, etc. The EBRs will show details of the products, including in-depth technical engineering data, requirements, and limitations related to the production process to customers to assess the price, production feasibility, availability of machinery and raw materials, etc. The SWR team has found several problems that may cause errors in preparing production documents, as follows:

- 1) Problems in the EBR documents of small customers: Since small customers do not have their standard forms, they will send information via email, online meetings, etc., which causes some data to be incorrect, and incomplete. Hence, it takes longer to review the documents and requires multiple revisions before the documents are complete. However, later we made improvements by creating a standard form for small customers to fill in the details of the product request via Google Form, with clear data entry conditions, such as Wafer Size, which can only be entered as 5/6/8/12 in inches, etc. Therefore, no more errors were found in the requested documents.
- 2) Problems in the EBR documents of large customers: In some months, there were many document requests, causing the SWR team to be unable to complete the visual document inspection in time due to the long document review time and multiple document revisions. Moreover, large-sized customers have standard forms in their system, making it impossible to change/edit data. This results in many errors, such as incomplete document review, in terms of units, expired Raw Materials, and data entry in the wrong position, which leads to errors in the document being leaked into the production document preparation process, resulting in problems with products that deviate from the desired target. It will take more time to resolve the issue using principles such as CAR, 8-Disciplines (8D), and Why-Why Analysis, etc.

Processing with expired Mold Compound in mass production that produces many products will result in problems affecting a wide scale. The data from the past year showed that there was damage from documents with errors leaking into the production line 3 times, with direct damage from paying a fine of \$3K and indirect damage from the customer canceling the entire project, totaling 200K pieces, with damages totaling \$80K, refer details per the graph (Fig. 3).

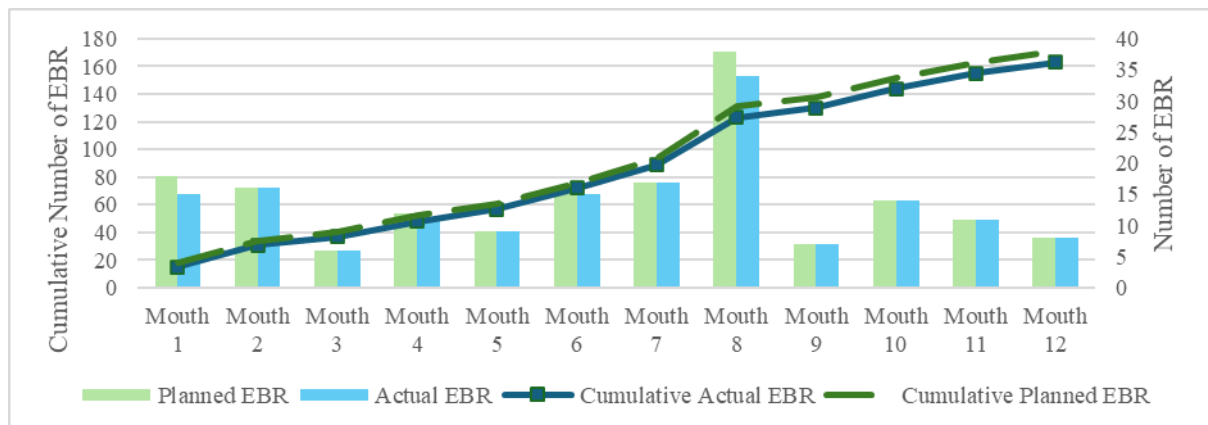


Fig. 3 Graph showing the number of planned EBRs and actual EBRs.

B. Experimental Method

After studying the error-checking process of the SWR team's engineering request form, it was found that there were several problems, such as not checking the documents in time and not checking the errors completely.

Therefore, the error-checking process of the engineering request form of a major customer was improved, as shown in Fig. 4. Then, 64 EBR documents were randomly selected from May 2023 to April 2024 for employees A, B, and C to test by checking the documents visually and using an automated program (Bot) to check their correctness. After that, the results were measured by timing the document checking, counting the number of errors found and using statistical hypothesis testing to test the results, including expected penalty and expected opportunity cost (Table 1).

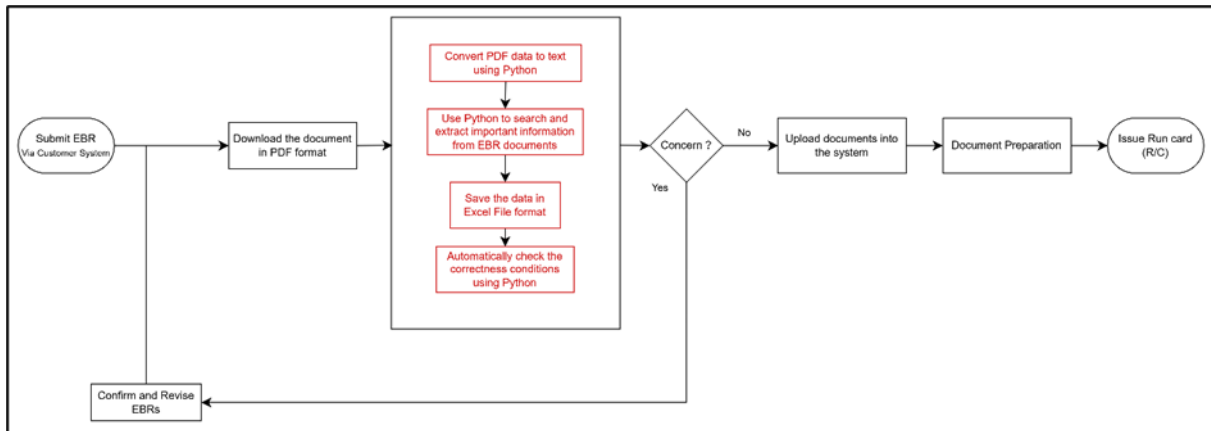


Fig. 4 Revised process flow of detecting incorrect information from EBRs

Table I Results from Detecting EBRs

	Detector			
	Human A	Human B	Human C	Bot
Time Average [Min/EBR]	30.61	29.72	30.72	0.0035
Time Median [Min/EBR]	16	15	16	0.0027
Standard Deviation of Time	31.17	31.38	30.53	0.0015
% Accuracy	92.61%	92.64%	92.72%	99.48%
% Error	7.39%	7.36%	7.28%	0.52%
Expected Opportunity Cost (\$40000 per 1 defect R/C)	-\$2,955.61	-\$2,944.44	-\$2,913.19	-\$208.33

C. Results and Discussion

From the experiment, humans A, B, C and an automated program (Bot) checked and considered the time spent checking the EBRs documents. Since the experiment was the same set of documents but checked separately, a paired t-test was used to test the hypothesis between humans and automated programs. It was found that using a Bot significantly reduced the checking time compared to using humans, ranging from 30 minutes to less than 1 minute. Considering the efficiency of the accuracy of the inspector, it was found that the average accuracy of humans was between 92%-93%. The accuracy of the Bot was around 99%. Moreover, the error occurred because 1 document with the same content was interrupted by a page change, causing the program to be confused in detecting the text. When using the percentage of errors as the chance that the EBR that did not meet the specifications was sent to the production line, causing damage and lost opportunities, the Expected Opportunity Cost can be calculated. For the review of request documents by humans, the Expected Opportunity Cost is approximately \$2900, while the review of request documents by the program is approximately \$200, which is 7% of the human review.

IV. CONCLUSIONS

In conclusion, this research has demonstrated the significant benefits of the application of text-mining techniques via Python programming to enhance the accuracy and efficiency of the inspection process of engineering build request forms (EBRs) in the electronics manufacturing industry. By utilizing Optical Character Recognition (OCR) and automated text processing, the research effectively identified and corrected errors that are often missed in manual inspections. The comparative analysis between human inspectors and the automated bot revealed that the latter not only reduces the time required for document review but also improves accuracy, thereby mitigating the

risk of costly production errors. This advancement suggests that the implementation of such automated systems can lead to substantial cost savings, reduced inspection times, and improved reliability in the production process.

REFERENCES

- [1] Supapan J, Chutima P.: Defect Reduction in the Manufacturing Process of In-Mould Decoration of Injection Moulded Components, *MSF*, Vol.962, (2019), pp. 181–188.
- [2] Polsuksiri, A., Chutima, P.: Simulation Analysis to Estimate Performances of Automated Parking System, *MSIE '23: Proceedings of the 2023 5th International Conference on Management Science and Industrial Engineering*, (2023), pp. 297-301.
- [3] Todd, J., Richards, B., Vanstone, B., Gepp, A.: Text Mining and Automation for Processing of Patient Referrals, *Appl Clin Inform* (2018), pp. 232-237.
- [4] Miner, G., Delen, D., Elder, J., Fast, A., Hill, T., Nisbet, R.: *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications* (Academic Press, 2012), pp. 967-982.
- [5] Kim, S. and Yoon, B.: Multi-document summarization for patent documents based on generative adversarial network, *Expert Systems with Applications*, Vol. 207 (2022).
- [6] Djedović, A., Žunić, E., Alić, D., Omanović, S., Karabegović, A.: Optimization of the Business Processes via automatic integration with the Document Management System, *International Conference on Smart Systems and Technologies (SST)*, Osijek, Croatia (2016), pp. 117-122.
- [7] Thongdonnoi, C., Chutima, P., Jiamsanguanwong, A.: Application of Collaborative Robots for Increasing Productivity in an Eyeglasses Lenses Manufacturer, *Eng. J.*, Vol. 27, No. 10 (2023), pp. 93-112.
- [8] Massey, A., Eisenstein, J., Antón, A., Swire, P.: Automated text mining for requirements analysis of policy documents, *21st IEEE International Requirements Engineering Conference (RE)*, Rio de Janeiro, Brazil (2013), pp. 4-13.
- [9] Friedman, C., Shagina, L., Lussier, Y., Hripsak, G.: Automated Encoding of Clinical Documents Based on Natural Language Processing, *Journal of the American Medical Informatics Association*, Vol. 11, Issue 5 (2004), pp. 392-402.