# Material Management Information System through the Use of Digital Database with OCR Technology

Rahmad Harjono [1], Lanny Panjaitan[2], Isdaryanto Iskandar [3]
Faculty of Engineering, Atma Jaya Catholic University of Indonesia

**Abstract:- Material management system is an important part of the construction stages. This system is generally used to control materials systematically, both in purchasing and storing sothey can be used regularly and timely. The complexity of an EPC project often requiresthat the material management in the fieldshould be delivered properly, in terms of their quantity, quality, andinformation. The use of an integrated database with Optical Character Recognition (OCR) technology which is programed by Python Programming Language at this case is intended to collect data and to identify materials so that theinformation is related to these materials can be accessed, processed, and displayed in real time and comprehensively.**

*Keywords: Material Management System, Digital Database, Optical Character Recognition*

## I. INTRODUCTION

Digital transformation is a must in the industrial era 4.0.The opportunity for that isvery wide, especially in theconstruction industry where the implementation for itself is relatively lower than other industries today. One example ofdigital transformation, especially in the field of information technology, is the digitization of information systems, which became popular since the 1980s through the invention of Personal Computer (PC). Digitization of information systems through the use of PCs, software, and Artificial Intelligence (AI) makes computerscapable of carrying out a certain task so that a complex problem can be solved more effectively and efficiently.

One example of the use of artificial intelligence in the digitalization of information is Optical Character Recognition (OCR)technology. OCR is a technology that can scantext objects,then process them intodigital data by machine learning. Asoftware to process these text objects is 'Tesseract' which is currently being developed by 'Google'. Through OCR technology,text objects that are processed into digital data can be used for various tasks, especially related to processinglarge and complex data.

This research which will be carried out here is how to create an application to integrate of an OCR technology with a digital database (SQL) to identify materials, spools, equipment, etc. so that the information desired byuser can be displayed timely and informatively through the Graphical User Interface (GUI).

## II. LITERATURE REVIEW

In the era of digital technology, human life is more complex and the use of technology is almost unstoppable. Artificial Intelligence technology where there isa Machine Learning (ML) inside them is one of the several technologies that are often used today.

An example of machine learning technologies to recognize optical characters is Optical Character Recognition (OCR). OCR is a technology that canchange text objects into digital data so thatthis data can be read by a machine / computer.A software used to support OCR is 'Tesseract'. The Tesseract software wasoriginally designed by Hewlett-Packard between 1985 and 1995 and became open source in 2005.As a complex application, OCR has several main stages torecognizecharacters accurately, which consists of: 1.) Pre-Processing Stage, 2.) Feature Extraction Stage, 3.) Introduction ProcessStage, and 4.) Post Processing Stage.

In general, some of the software and librariesthat will be used for this research are as follows:
- Python
  Python is a high-level programming language which is often used for various needs. Python, which was created by Guido Van Rossum in 1991, is a programming language widely used for data science, machine learning, computer vision, etc. Python also hasa lot of library resources to develop an application.
- Pytesseract
  Pytesseract is a wrapper for Google's OCR engine. Pytesseract is one of the python librarieswidely used to process image data into digital text data (image to string) that can be read by computers with machine learning.
- Open-CV
  Open-CV is a library which is used to extract information from images and videos. This library is generally used for real-time dynamic image processing.
- PIL
  Python Image Library (PIL) is an additional open-source library for Python whose main function is the manipulationof image files. Some types of manipulation that are commonly conducted includemasking, filtering, enhancing, adding text, and others.
- Tkinter
  Tkinter is a standard python graphic user interface (GUI) which is used to create an application display by components in the tkinter module such as Buttons, Textboxes, Labels, Frames, Windows, etc. The use of Tkinter is intended to create an input-output interface between user and computer to make it more interactive.

- SQLite
SQLite is a software library that implements the SQL database engine independently. SQLite has many advantages over other databases such as being easier to manage, more stable, more extensible, faster, and having the Write-Ahead Log (WAL) mode feature.

## III. RESEARCH METODOLOGY

This research was conducted at PT RekayasaIndustri especially at CDU Revamp Project. This research uses design thinking as research methodology to find the root cause analysis commencing with finding problems which often occur in the field, up to developing and testing of idea through benchmarking of market and building of prototype.

### A. Empathize Stage

Empathize is the first stage in Design Thinking. At this stage, we have toplace ourselves as a customer so that we can understand problems more closely from the customer's side. The description of empathy map in this research is as follows:
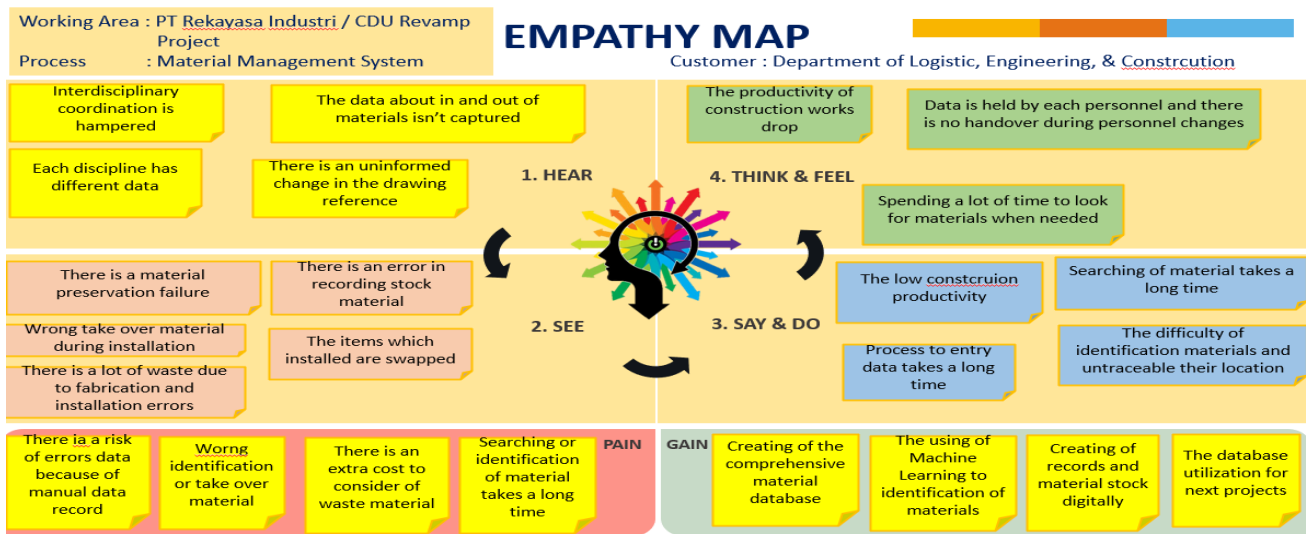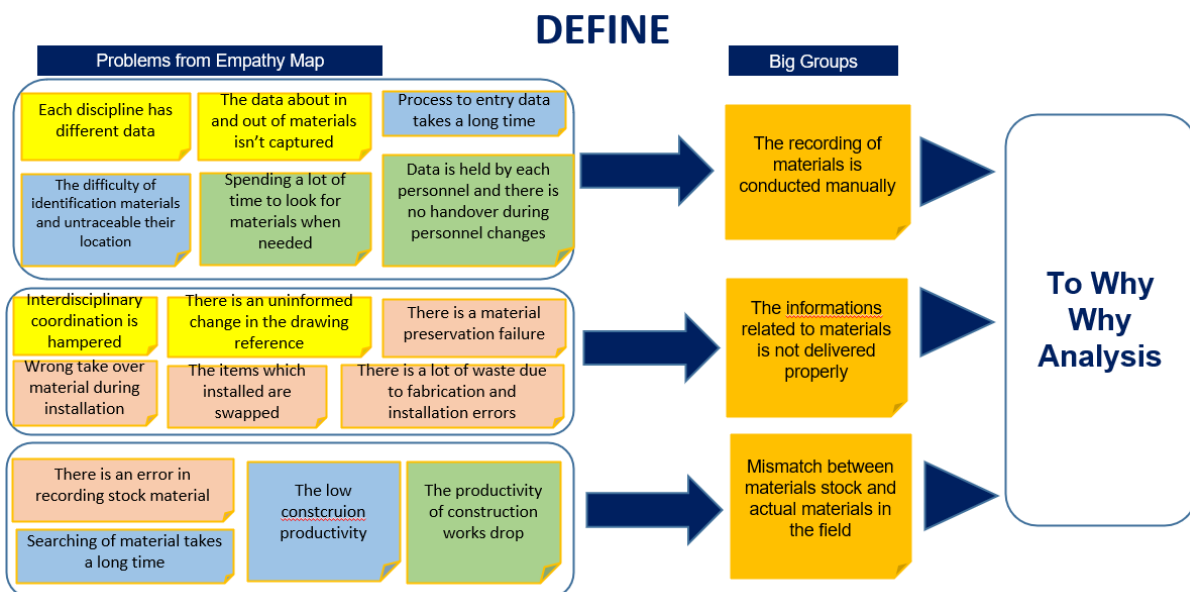


Fig. 1: Empathy mapping of material management system

### B. Define Stage

At this stage, some of the actionthat is conducted is to make a large number of groups ofproblems that have been described in the empathy map, then find the root cause by why-why analysis tools. The description of the define stage in this research is as follows:
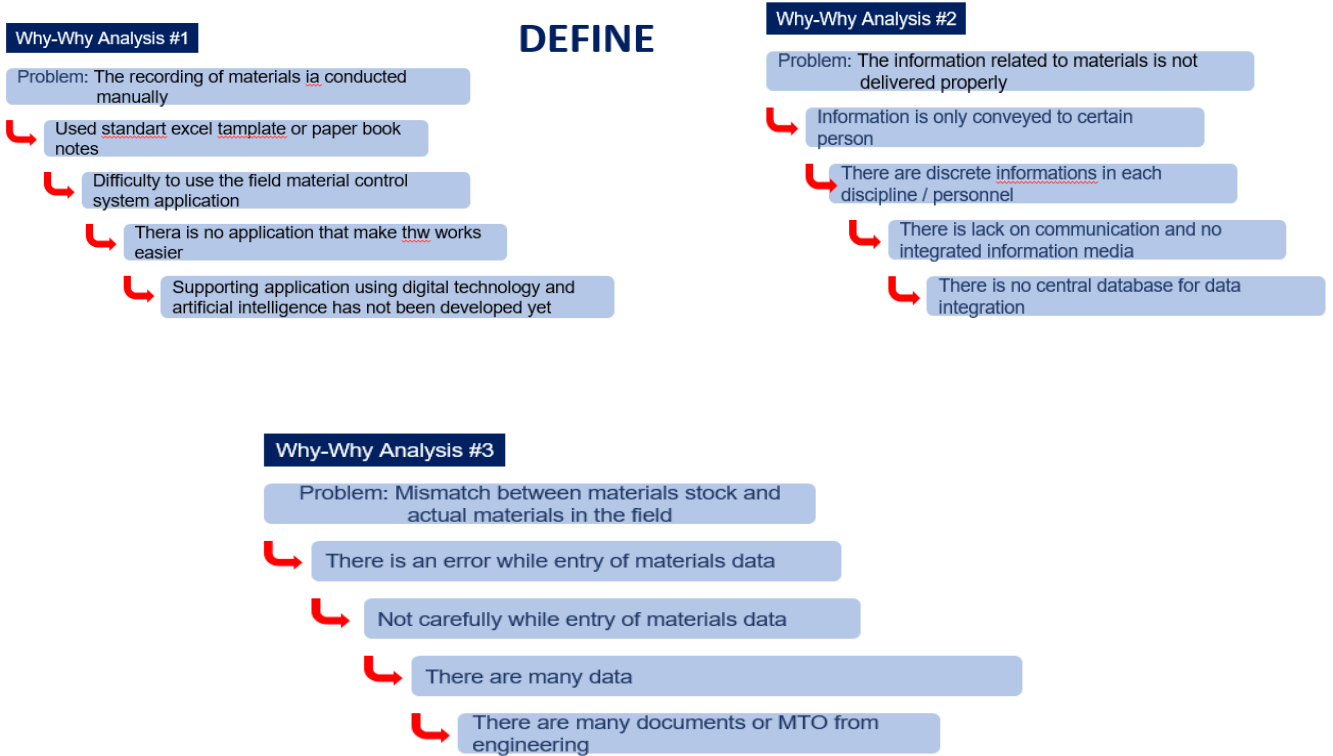
**DEFINE**

**Why-Why Analysis #1**

Problem: The recording of materials ia conducted manually

Used standart excel tamplate or paper book notes

Difficulty to use the field material control system application

Thera is no application that make thw works easier

Supporting application using digital technology and artificial intelligence has not been developed yet

**Why-Why Analysis #2**

Problem: The information related to materials is not delivered properly

Information is only conveyed to certain person

There are discrete informations in each discipline / personnel

There is lack on communication and no integrated information media

There is no central database for data integration

**Why-Why Analysis #3**

Problem: Mismatch between materials stock and actual materials in the field

There is an error while entry of materials data

Not carefully while entry of materials data

There are many data

There are many documents or MTO from engineering

Fig. 3: Why-why analysis diagram

The recording of materials is conducted manually

The informations related to materials is not delivered properly

Mismatch between materials stock and actual materials in the field

**The Topic Selection**

Supporting application using digital technology and artificial intelligence has not been developed yet

There is no central database for data integration

There are many documents or MTO from engineering

There is no central database which is integrated with the use of digital technology and Artificial Intelligence to support the material management system at site
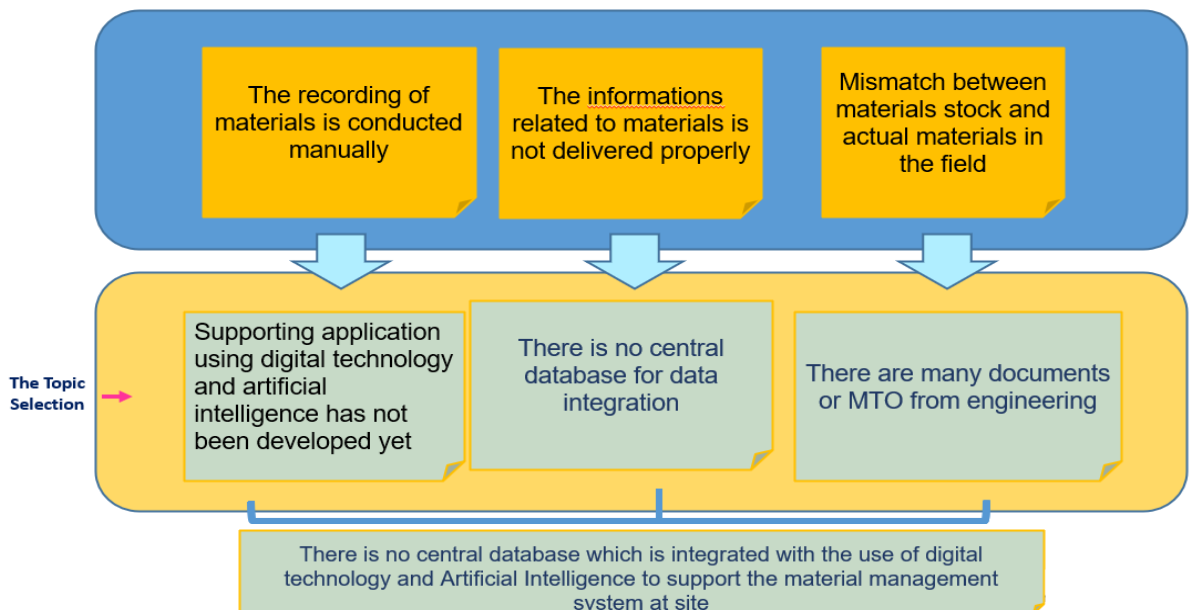
Fig. 4: Main topic selection diagram

Based on the description above, there are several major groups of problems consisting of:the recording of materials is conducted manually, the information related to materials is notdelivered properly, and mismatch between material stock and actual material in the field. From the root cause analysis, these problems are caused by: supporting applications using digital technology and artificial intelligence has not been developed yet, there is no central database for data integration, and there are many documents or MTO from engineering. The root cause of these problems isnothing central database that is integrated with the use of digital technology and AI to support material management systems in the field.

*C. Ideate Stage*

At this stage, ideas are breakdown as much as possible,then summarized to get 3 (three) main ideas. After that,the idea selection table is used to compare based on the desirability, feasibility, and viability,so that the best idea which hashighscore, will be developed to bea prototype. The details of thesestages in this researchare as follows:
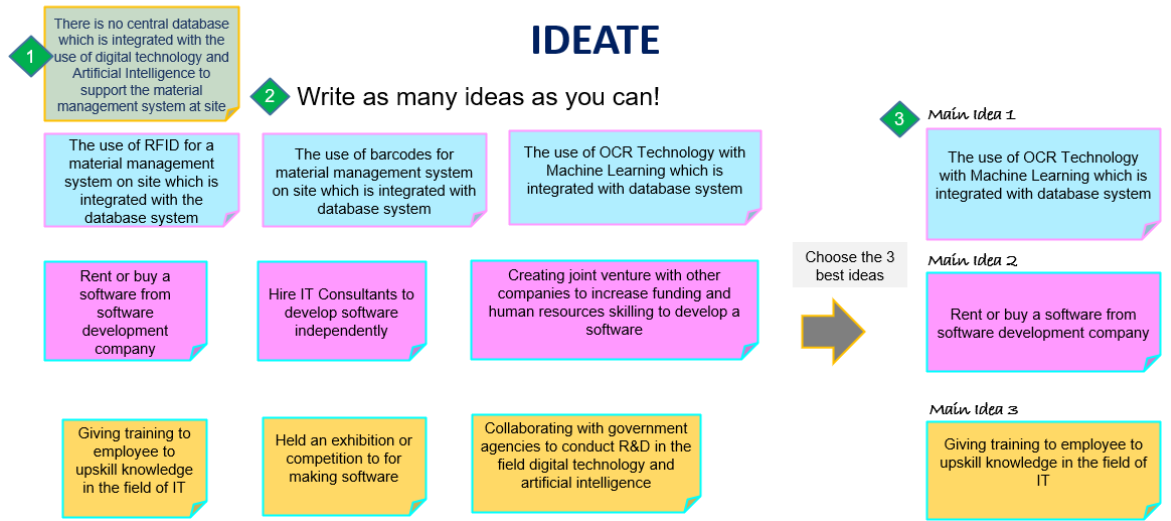
Fig. 5: Ideas breakdown diagram



Fig. 6: Ideas selection diagram

From the idea selection table above, we have got the best idea to be developed is the use of OCR Technology with Machine Learning which is integrated with database system. Through prototype and test stage, author tries to develop a graphical user interface (GUI) according to the description below.

### D. Prototyping Stage

As an initial description, the first product in this development stage consists of:

a) Main Dashboard

The main dashboard of this prototype consists of Material Code, Material Description, Discipline, PO No., and Vendor Name, where the results of data input will be displayed in the Material Info Dashboard. The main dashboard is also equipped with buttons as follows:

- **Add Button** : Used to add input data into database
- **Delete Button** : Used to delete data records that have been entered in database
- **Clear Button** : Used to delete entries of data that have not been saved in database
- **Exit Button** : Used to exit this application

b) Dropdown Menu – Search

The dropdown menu is used to select the appropriate material identification method, which consists of:

- **Search by Material Code**    :    Used to identify material data through input of material code into application
- **Search by Video Capture**   :    Used to identify material data through direct capture by camera
- **Search by Image File**      :    Used to identify material data through selection of image file data stored in the device

*E. Test Stage*
   a) Synchronization Test between Tkinter Module and SQLite Database
   The synchronization test between Tkinter module and SQLite database is carried out by testing the input of material data through Tkinter Graphic User Interface (GUI). As per the trial phase, it was found that the input material through the Tkinter module can be synchronized with the SQLite database according to the following figure:
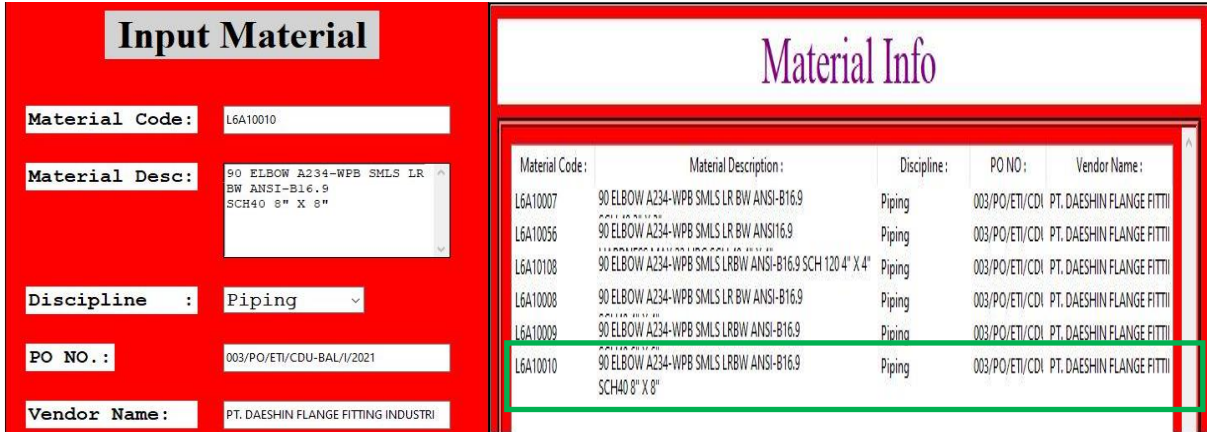


Fig. 7: The result of synchronization test between Tkinter Module and SQLite Database

   b) Material Identification Test with Search by Material Code
   Material identification test with Search by Material Code is conducted by entering a unique material code into the application to filtersearch results according to data that has been stored in the database. The results of this trial shown in the following picture:
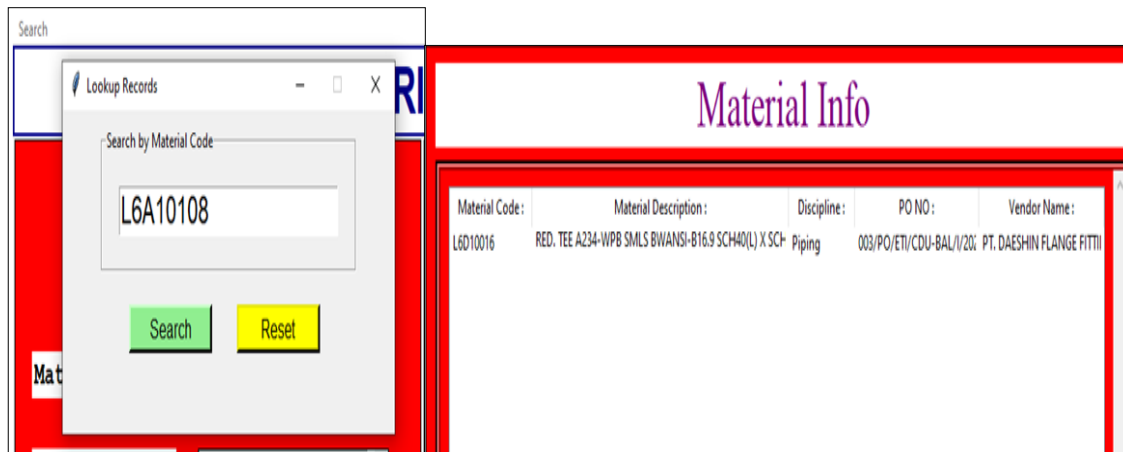


Fig. 8: The result of material identification with Search by Material Code

   c) Material Identification Test with OCR (Video Capture & Select Image) Technology
   Material identification test with OCR Technology is carried out using 2 (two) methods through direct capture by camera (Video Capture) or select of photo files from device (Select Image). Furthermore, through the "Capture" or "Select File" button, the application will convert the photo object (text) into digital text data,then automatically insertit into the "Lookup Records". The digital text data that has been inserted into the "Lookup Records" then ischecked by the "search" button to identify material data from the database so that information that matches with code of material will appear.

The results of this trial shown in the following picture:

In addition, an example of implementing the use of paper tags for tagging material code on materials in the field can be seen in the following example:
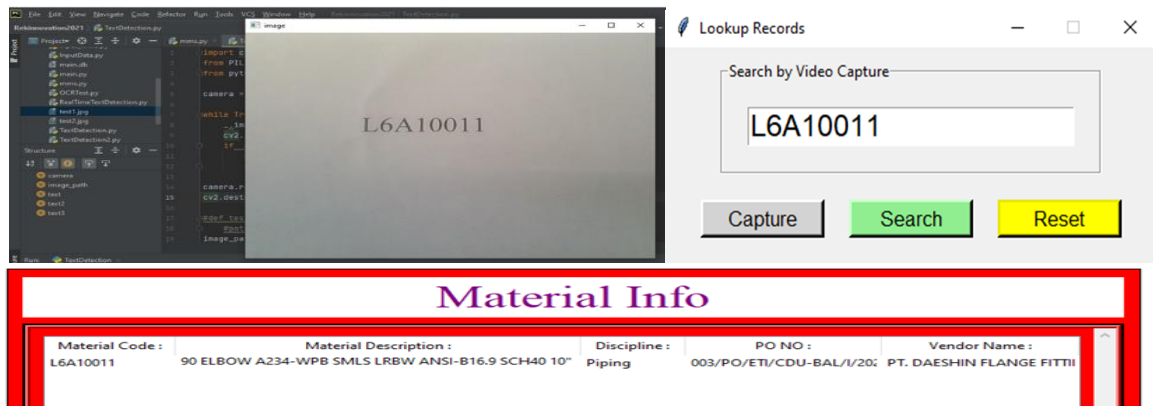
Fig. 9: The result of material identification with OCR Technology



Fig. 10: The example of placing materials nroduct the field

## IV. RESULTS AND DISCUSSION

From research methodology above, we have proven that the use of OCR Technology with Machine Learning which is integrated with database system is the best idea to be implemented compared to other options. Furthermore, the use of this technologyprovides many advantages including control of data redundancy, increasing of data integrity, increasing of security, andalso improving data backup and recovery services. This will be very beneficial for the engineering, procurement, construction, and other supporting divisions in accessing and processing data to support their work.

In terms of cost, the use of OCR technology is also cheaper than similar technologies such as RFID. The following is a comparison of investment and development costs between this technology:

| No. | Description | OCR Technology | RFID |
|---|---|---|---|
| 1 | Software Development | Self-Development: CostUSD 0 | By Third-Party: Cost +/ USD 15,000 |
| 2 | Software License | Open Source | Third Party License |
| 3 | Hardware | Server + Camera + CCTV + HP: Cost +/- USD 3,000 | RFID Tag + RFID Reader + Server + etc: Cost +/- USD 5,000 |
| 4 | Maintenance | Self-Maintenance: CostUSD 0 | Third-Party: Cost +/- USD 1000 |
| 5 | Return on Investment | Quick | Slow |
| | TOTAL | USD 3,000 | USD 21,000 |

Table 1: The different of developing cost between OCR Technology and RFID

The use of an integrated database system with OCR technology is also intended to assist stakeholders in overcoming current weaknesses, especially in terms of recording and identifying materials,to be faster and more accurate.

## V. CONCLUSION

The use of databases allows us to process large data setsand data integration between disciplines. The use of OCR technology also allows us to identify materials quickly with low budget. The combination of these two advantages is expected to add value,especially in EPC Company to facilitate the execution of large projects in the future.

In addition, this application will be developed through the Python programming language, which currently has more than 137,000 libraries and isthe most popular programming language in 2021 with ease of use and the support of a very large community. Python libraries which have an important role in developing machine learning, data science, data visualization, image and data manipulation, and many others, will play an important role in the development of this application in the next stages.

## REFERENCES

[1.] Raswaty and Nuryuliani, (2021), "Implementation Of Optical Character RecognitionAnd Voice Recognition Of House Of Words (How)Dictionary Application On Android Platform", Journal of Engineering, Mathematics and Computer Science, Vol.3 No.3, 93-101.

[2.] Anugrah and Bintoro, (2017), "Latin Letters Recognition Using Optical CharacterRecognition To Convert Printed Media Into Digital Format", Journal of Electronica and Telecommunication, Vol.17, 56-62.

[3.] Foster, Mary K , (2019), "Design Thinking: A Creative Approach To Problem Solving", Management Teaching Review.

[4.] Razzouk,Rim and Shute, Valerie, (2012), "What Is Design Thinking And Why Is ItImportant?", Review Of Educational Research, Vol. 82, No. 3, pp. 330–348.

[5.] Thanki, Jay Dilipbhai, (2021), "A Review on OCR Technology, JETIR, Volume 8, Issue 4.

[6.] Isheai and Hasan, H (2015), "Optical Character Recognition (OCR) System, IOSR Journal of Computer Engineering (IOSR-JCE), Volume 17, Issue 2, pp. 22-26.