SMART ARCHIVING MECHANISMS FOR ENERGY AND PETROLEUM PROJECTS USING BIG DATA

By

Mahmoud Mohamed ElMortada ElZahed

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
INTERDISCIPLINARY - MASTER OF SCIENCE
in
INTEGRATED ENGINEERING DESIGN IN CONSTRUCTION PROJECTS

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2020

SMART ARCHIVING MECHANISMS FOR ENERGY AND PETROLEUM PROJECTS USING BIG DATA

By

Mahmoud Mohamed ElMortada ElZahed

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
INTERDISCIPLINARY - MASTER OF SCIENCE
in
INTEGRATED ENGINEERING DESIGN IN CONSTRUCTION PROJECTS

Under the Supervision of

Prof. Mohamed Mahdy Marzouk

Professor of Construction Engineering and Management
Structural Engineering Department
Faculty of Engineering, Cairo University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2020

SMART ARCHIVING MECHANISMS FOR ENERGY AND PETROLEUM PROJECTS USING BIG DATA

By

Mahmoud Mohamed ElMortada ElZahed

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
INTERDISCIPLINARY - MASTER OF SCIENCE
in
INTEGRATED ENGINEERING DESIGN IN CONSTRUCTION PROJECTS

Approved by the Examining Committee

Prof. Mohamed Mahdy Marzouk ……Thesis Main Advisor
Professor of Construction Engineering and Management - Structural Engineering Department - Cairo University

---

Prof. Dr. Fouad Khalaf Mohamed…….…….Internal Examiner

Professor of Petroleum - Department of Mining Petroleum and Metallurgical Engineering - Cairo University

---

Dr. Mohamed Abdel-Latif Bakry……..…..External Examiner
Former Head of Planning and Control – Social Fund for Development

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2020

Engineer's Name: Mahmoud Mohamed ElMortada ElZahed
Date of Birth: 22/11/1993
Nationality: Egyptian
E-mail: mahmoud.elzahed@outlook.com
Phone: 01011972172
Address: 6th of October, Giza, Egypt

Registration Date: 1 / 10 / 2016
Awarding Date: 1 / 11 / 2020
Degree: Interdisciplinary - Master of Science
Department: Integrated Engineering Design in Construction Projects

Supervisors: Prof. Mohamed Mahdy Marzouk – Cairo University

Examiners:
Prof. Mohamed Mahdy Marzouk (Thesis main advisor)
Prof. Fouad Khalaf Mohamed (Internal examiner)
Dr. Mohamed Abdel-Latif Bakry – Social Fund for Development (External examiner)

Title of Thesis:
Smart Archiving Mechanisms for Energy and Petroleum Projects Using Big Data

Key Words:
Big Data, Energy and Petroleum Projects, Smart Archiving, Optical Character Recognition

Summary:
Complexity of the construction projects vary by the domain and type of the project. Due to the interaction between different disciplines and parties, EPP are considered among the most complex. This research proposes a framework that increases the efficiency of archiving the accumulated data without affecting the normal workflow of companies, overcoming the man-hours expenditure, and reducing the time of archiving while not affecting the accuracy of the outcome. The proposed framework integrates four modules to provide a complete solution to the problem. The first module is responsible for image processing to enhance the quality of the images. Then, OCR module converts the images to text to be processed, this data is then processed; where text cleansing and preparation is performed using big data tools to allow for large scale real-time implementation. Followed by text searching and results verification using regular expressions. The final module is responsible for archiving the verified data in a structured database to be available for users. The framework transforms the existent unstructured data into structured data which can be used in initial estimations and referencing.

Disclaimer

I hereby declare that this thesis is my own original work and that no part of it has been submitted for a degree qualification at any other university or institute.

I further declare that I have appropriately acknowledged all sources used and have cited them in the references section.

Name: Mahmoud Mohamed ElMortada Mohamed ElAnwar Azazy ElZahed
Date: .. /.. /…

Signature:

Acknowledgments

In the name of Allah, the Most Gracious and the Most Merciful, all praises to Allah for the strengths and His blessing in completing this thesis.

I would like to express my sincere gratitude to my supervisor Prof. Mohamed Mahdy Marzouk for his tremendous efforts and continuous support throughout the period of my master's degree. His words of wisdom, guidance and insightful suggestions were pivotal to make this thesis reach such level.

I would like to thank Eng. Haitham Badawy and Eng. Essam El-Eskandarany for providing me with their utmost support to obtain the necessary approvals and providing this research with their important business insights which helped make this research more relevant to the business.

I would like to thank Eng. Mohamed El-Desouky for supporting me with his technical expertise and availing such IT infrastructure required to proceed with the research required to complete this thesis.

Additionally, I would like to thank my friends Mohamed Abdelraouf and Mohamed Ezzat, for their support.

Finally, I would like to express my sincere appreciation and gratitude to my family for their help, support, patience, and words of encouragement during the preparation of this thesis.

Table of Contents

List of Tables

List of Figures

Nomenclature

AI; Artificial Intelligence

ANN; Artificial Neural Network

BD; Big Data

BI; Business Intelligence

EPP; Energy and Petroleum Projects

HSE; Health, Safety and Environment Department

I&C; Instrumentation and Control Engineering Department

KDD; Knowledge Discovery in Databases

KM; Knowledge Management

Ksize; Kernel Size

ML; Machine Learning

OCR; Optical Character Recognition

PEM; Project Engineering Management Department

PM; Project Management Department

PV; Pressure Vessels Engineering Department

QEHMS; Quality, Energy and Health Management Systems Department

SQL; Structured Query Language

W.r.t.; with respect to

Abstract

Complexity of the construction projects varies by the domain and type of the project. Due to the interaction between different disciplines and parties, Energy and Petroleum Projects (EPP) are considered among the most complex. This complexity produces a dense network of interrelated documents which are produced to cover the various aspects and details of the project. Analyzing this network is required in order to gain insights from old data. This task traditionally requires experience, knowledge, and awareness about the existence of the required data. Accordingly, a key asset of any company is the knowledge accumulated over the time from various projects. The main challenge of utilizing this asset is archiving such data and storing it in a structured manner.

This research proposes a framework that increases the efficiency of archiving the accumulated data without affecting the normal workflow of companies, overcoming the man-hours expenditure, and reducing the time of archiving while not affecting the accuracy of the outcome. Due to the large diversity of the EPP projects, the research focuses on five main commodities as the main data to be stored which are Tanks, Air Coolers, Pumps, Generators and Distributed Control Systems (DCS). The selection of these commodities is based on the frequency of their existence in projects in addition to their monetary value. The key attributes of each commodity are identified based on technical questionnaires with technical specialists to act as the basis for building the proposed framework. The proposed framework integrates four modules in order to provide a complete solution to the problem. The first module is responsible for image processing to enhance the quality of the images and remove artifacts due to scanning. The second module, Optical Character Recognition (OCR) module converts the images to text in order to be processed. The third and main module is responsible for data analysis; where text cleansing and preparation is performed using big data tools to allow for large scale real-time implementation. Followed by text searching and results verification using regular expressions. The final module is responsible for archiving the verified data to a structured database to be available for users. The proposed framework harnesses the power of big data analytics to transform the existent unstructured data into structured data ready to be used for ongoing business operations such as initial estimations and referencing. Additionally, the savings in time and money compared with conventional methods further support this conclusion. In order to properly implement the verification workflow, a case study project that has eleven purchase orders of the main commodities is worked out to illustrate the use of the proposed framework.

Chapter 1: Introduction

1.1. The Special Case of the Construction Industry

Construction industry is considered a key industry to many of the world's countries, being a labor-intensive industry requiring high investment; this industry contributes to about 13% of the Gross Domestic Product (GDP) in the U.S economy while employing 8% of the working force [1]. Over the last 10 years, the % GDP change averaged 3.5% globally, also the trend of the change in the construction industry resembles the change in GDP (as shown in Figure 1.1). Several supporting research associated between investment in the construction sector and economic growth [2], [3].

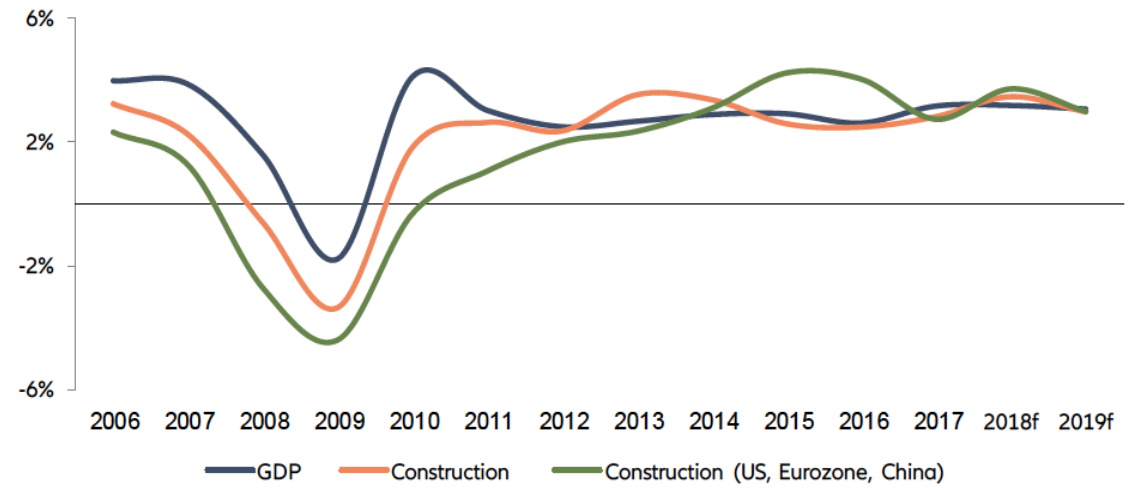

Figure 1.1: Global Construction Industry vs Global GDP (real USD, %change y/y) **[3]**

Complexity of the construction projects varies by the domain and type of the project. Due to the interaction between different disciplines and parties, Energy and Petroleum Projects (EPP) are considered among the most complex. This complexity produces a dense network of interrelated documents which are produced to cover the various aspects and details of the project. Analyzing this network is required in order to gain insights from old data. This task traditionally requires experience, knowledge, and awareness about the existence of the required data. Accordingly, a key asset of any company is the knowledge accumulated over the time from various projects. The main challenge of utilizing this asset is storing it in a structured manner.

1.2. Problem Statement

Complexity of EPP and the high number of interfaces resulted that most of the projects are done on an Engineering, Procurement and Construction (EPC) basis; EPC projects. In this type of projects, the main contractor is assigned the responsibility of designing the plant, purchasing the required material for the plant, erecting the equipment, and constructing the structures and buildings. EPC projects do exist in many industries such as aerospace, automotive, software and electronics. Typically, the EPC projects in the EPP sector are valued at millions of USDs, accordingly the competition between the contractors is high and requires that they prepare the best tender that balances between the quality, time, and cost of the project. Preparation of such tenders can be considered as big projects which requires a lot of effort from highly skilled personnel.

Usually, the process of EPC proposals in EPP projects starts by client preparing the basic engineering package of the plant and issuing the Instruction To Bidders (ITB), from this the contractors start analyzing the bid documents to prepare their tenders. The key activities that are performed by the contractors are:

a. Pre-award Engineering: basic package is analyzed against the specifications to identify key requirements that shall be considered and also to increase the level of details of the design to a sufficient level to allow for more confident pricing and risk assessment.
b. Budgetary Quotations: critical commodities are identified, and quotations are requested from vendors to enhance the accuracy of the pricing, and to identify any potential major risks.
c. Construction Estimates: accounting for a significant percentage of the contract price, construction Bill Of Quantities (BOQs) are prepared and priced. Also, considerations for safety requirements are taken into account in addition to temporary facilities for accommodation and construction required since projects are typically in remote areas with few services provided.

After completion of the above activities, usually contractors are ready to estimate the tender price and the associated risks and proceed with the final bid decision. In cases where contractors decide to bid, the typical industry percentage of success ranges between 30 and 50% of the tenders submitted. This probability of winning depends upon the accuracy of quotations received and evaluated within the usually tight proposal timeframe, the level of details reached in engineering and construction estimation accuracy.

The balance between the level of details and efforts devoted in tenders preparation shall be treated carefully; as Gardner et al. [4] found out that increasing the input variables during early estimation doesn't necessarily enhance the accuracy of the estimate [4]. Accordingly, these efforts shall be closely monitored and managed to ensure that they are not wasted. In a corporate context, these proposal preparation