# A Study on Information Extraction Method of Engineering Drawing Tables

Riza Sulaiman
Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia, Selangor, Malaysia

Mohd Fahmi Mohamad Amran
Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia, Selangor, Malaysia

Nur Amlya Abd Majid
Faculty of Computer Science and Information Technology,
Universiti Selangor, Malaysia

## ABSTRACT
In getting an existing engineering drawing to be reuse, the first step to be done is to search for the appropriate file. To get the desired drawing, an engineer will have to access every file. This procedure consumed much time and energy as there might be lots of files and folders related to the drawings. To ease engineers search for old engineering drawings, the extraction of the table structure has to be done. This paper describes review on recent research in extraction of engineering drawing tables. Along the way of this reviewing process, the weakness and downside of previous researches has been identified.

## General Terms
Information Extraction

## Keywords
Computer Aided Design, Engineering Drawing Tables, Reuse Design

## 1. INTRODUCTION
When a designer designs a new engineering drawing, they will refer to the previous valid design documents that are the same or nearly match with their new design specification. This is due to the fact that the previous designs have been reviewed, analyzed and proven to be successful. By referring to the previous designs, it likely saves significant resources and manpower [1]. In getting an existing engineering drawing to be reuse, the first step to be done is to search for the appropriate file. To get the desired drawing, an engineer will have to access every file. This procedure consumed much time and energy as there might be lots of files and folders related to the drawings. The best way to search for an old drawing is by applying an electronic-based search in the operation system compared to the manual search which required an engineer to read the content of the drawings. The limitation in the search process based on the operation system, the information states only the names of the files. In engineering practices, the name of the files does not necessary match the drawings. It might not state information on the drawings such as the name of the drawings, the designer, the date, the version number and other information. By reading only the files name, it could not visualize the real picture to engineers about the content regarding the engineering drawings. It becomes more complicated if the files name is written by using the project code and engineers do not aware of the certain project code representing to related project. Engineers also face difficulties to obtain the latest version of the drawing especially which are related and dependent to other several drawings [2]. Therefore, to gain information in the table, information extraction need to be done to access to the required information. Besides, by extracting information consists in the table itself, the information gained is accurate and free from problems.

To ease engineers search for old engineering drawings, the extraction of the table structure has to be done first, for text is not only written in the tables, but also in the whole of engineering drawings including the text which reflected the dimension, text notation and other component. On top of that, extraction also has to be done in differentiating the tables. The search information process is facing with difficulty if there are mistakes in the process of the extraction tables. In engineering drawings, there is a possibility of the existence of more than one table. Besides the title block, the table in engineering drawings is a table representing information such as the Bill of Materials (BOM), revision block, product structure and so on. Every schedule prepared different information and convey different meanings. For example, certain table explains about data regarding the engineering object whilst the other table explains data regarding the evaluation of the object [3].

The purpose of this paper is to review recent research in extraction of engineering drawing tables. This paper has been divided into two parts. The first part deals with review on related works in extraction methods of engineering drawings tables. For the second part, the weakness and the downside of the extraction methods will be analyzed and discussed.

## 2. OVERVIEW OF INFORMATION EXTRACTION IN ENGINEERING DRAWING
According to [4] the process of extraction is the process of extracting of CAD data in order to obtain details of products and processes that contribute to the production of the product. Basically, CAD data contains data on geometry and non geometry. Geometry data is a collection of objects described by CAD drawing of lines, arcs and circles. While non-geometry data are related to object attributes such as length, thickness, radius, diameter and label [5]. In this research, tables that will be studied is consider as non-geometry information. Geometry and non geometry data is closely related to information products [5]. [6] classified the automatic recognition of engineering drawings problems into three levels at which the first is the identification of low-level graphics primitives, the second stage of identification of high-level graphical objects, and the third level of high-level intelligence interpretation and analysis. There are several