

ENTITIES FILTERING TECHNIQUE FOR INFORMATION EXTRACTION IN ENGINEERING DRAWING FILES

M. F. M. Amran^{1,*}, R. Sulaiman², R. M. T. R. L. Ahmad², N. M. M. Yusop¹,
N. A. A. Jamaludin¹ and H. Mohamed¹

¹Computer Science Department, Faculty of Science and Defense Technology, National Defence University of Malaysia, Sungai Besi Camp, 57000 Kuala Lumpur, Malaysia

²Institute of Visual Informatics, Universiti Kebangsaan Malaysia 43600, Selangor, Malaysia

³Information Technology Department, Faculty of Communication, Visual Arts and Computing, Universiti Selangor, Malaysia

Published online: 10 September 2017

ABSTRACT

Engineering drawing has become essentials to most organization be it consultants, contractors or even government sectors. To get the information from an engineering drawing, entities must be filtered in order to extract the accurate information. This research presents the entities filtering technique for information extraction in engineering drawing files. The technique filtered the entities based on four attributes which are layer, color, entity type and line type. Ten engineering drawings files was used for testing, showing that the entities filtering technique is able to extract all the entities.

Keywords: information extraction; computer aided design; filtering technique; engineering drawing.

Author Correspondence, e-mail: fahmiamran@upnm.edu.my

doi: <http://dx.doi.org/10.4314/jfas.v9i3s.17>



1. INTRODUCTION

The initial step in the extraction of information contained in an engineering drawing is to analyze the existing file format. There are various CAD systems that can be chosen by engineers which suit the function and type of drawing that is being produced. Consultants involved with engineering projects uses different type of CAD system in executing task [1]. Different types of CAD system create problems when it is required to change to different format which involves high costs. File format that is open access and can be used by all CAD system must be chosen, thus Drawing Exchange Format (DXF) that has been developed by Autodesk have been selected due to the fact that it meets the requirement. Despite differences regarding content from engineering drawing from various fields such as mechanical and manufacturing, this type of file format is preferred because open access is widely used in most engineering sector. In [2-3] has proposed a study on information extraction technique on engineering drawing. To get the information from an engineering drawing, entities must be filtered in order to extract the accurate information. Previous studies [4-9] highlighted the importance of this study which will be used in future research. This research discusses about entity filtering technique using Drawing Exchange Format (DXF) type of file format that functions as eliminating entity in engineering drawings based on respective categories. Entity filtering technique is an extensive process because geometric model, texts and lines are some of the elements that exist in engineering drawing [10]. Therefore, algorithm or certain technique has to be introduced so that process of filtering entity can be executed accurately. This paper describes a research and development of an entities filtering technique in the field of engineering drawing. Entities that will be emphasis in this research are line and text.

Drawing Exchange Format (DXF) type of file format is a CAD data file which has been developed by Autodesk. DXF was introduced in 1982 as part of AutoCAD 1.0 with the purpose to represent data file in AutoCAD. It enables the data to function with various CAD systems.

DXF file type is one of CAD files that is widely being used to support different type of file format [11]. This type of file format is considered as vector type and operates in digital mode. Engineering drawing in raster scan image or pixel and drawn using pen or pencil does not

belong to this category. Information contains in engineering drawing is represented in DXF format which is in the form of tagged data. Each tag is match with specific integer. These integers are known as group code. The value of the group code represents type of data element for a specific type of object. Arrangement technique for this type of file starts with data element and followed by integer.

DXF file is consists of group codes and associated values [12]. These group codes highlight the value type which is match with the group codes. Using the group codes and associated values, DXF file is divided into few sections which consist of records of group codes and information of the data itself. Few section of DXF file structure has information that relates to entity that represents information for geometry and non-geometry elements. Each group codes and its values are located at each line in DXF file. The DXF file can be divided into two types of format which are ASCII and binary format. Basically, DXF file is widely use in ASCII format rather than binary format [13].

There are six sections in DXF file which are HEADER, CLASSES, BLOCKS, TABLES, OBJECT and ENTITIES. Each section starts with '0' values of the group codes followed by section. Group code '2' followed after that and it shows section as well. Each section consists of group codes and values that depict its element. Each section end with group codes with the value of '0' followed by ENDSEC. Entities that is part of DXF file is located under Entity section [14]. Therefore, this research focuses on extracting information which only is under the Entity section.

2. RESULTS AND DISCUSSION

In order to prove that the developed technique able to carry out its task, one simple test has been conducted based on comparison with the result obtain from a different application.

This test is important since it ensure that the developed technique can filter the entity correctly. Quick Select dialog box in AutoCAD 2010 will be used to choose a specific entity. 10 engineering drawings have been used to assist with the testing process. This test is conducted to extract the right number of entity that can be filtered from the two techniques.

Table 1. Comparison for line entity

Number of Engineering Drawing	Software Technique (AutoCAD 2010)	Entities Filtering Technique	Difference
1	293	293	0
2	242	242	0
3	711	711	0
4	856	856	0
5	607	607	0
6	454	454	0
7	663	663	0
8	357	357	0
9	738	738	0
10	195	195	0

In order to verify the result of the developed technique, the result will be compared with the result manually extracted from AutoCAD 2010 software. Table 1 shows the comparison for line entity, whereas Table 3 shows comparison for the text entity.

Table 2. Comparison for text entity

Number of Engineering Drawing	Software Technique (AutoCAD 2010)	Entities Filtering Technique	Difference
1	263	263	0
2	513	513	0
3	413	413	0
4	679	679	0
5	428	428	0
6	692	692	0
7	569	569	0
8	567	567	0
9	684	684	0
10	231	231	0

The result for each engineering drawing based on the two technique used have been recorded. Based on Table 1 and Table 2, it is proven that for all ten engineering drawing, the result is same either using AutoCAD 2010 or entities filtering technique. This proves that the developed entities filtering technique is precise and reliable.

3. EXPERIMENTAL

DXF file is consists of match able group codes and certain values. These codes are read through CAD system and translated into two dimensional vector drawings. Users are able to read DXF file through applications such as Notepad and Textpad since it is able to open in text mode. Each codes in DXF file are located in their own line. Group codes refer to type of values that it represents. Integers are use as representative for group codes.

```
0
SECTION
2
ENTITIES
0
LINE
8
0
6
Continuous
62
1
100
AcDbLine
10
2.0
20
3.0
```

```

30
0.0
11
5.0
21
7.0
31
0.0
0
ENDSEC
    
```

Fig.1. DXF code for line entity

The '0' group codes shows the beginning of a section. Group codes '1' shows values of the main text for an entity. Group codes '2' represent a name such as section name, block and tables. Other function of '0' group codes includes the start of an entity. Example of an entity is line. Line is located at layer 0, based on highlights showed by group code '8'. Group code '6' shows type of line such as continuous line. Group code '62' shows the color of the entity. The choice of color is represented using integer. Based on the following example, integer 1 means its red. Group code '10' until '59' shows starting and ending points for a specific line.

Fig. 1 shows the DXF codes represent the line entity and Fig. 2 shows the text entity.

```

0
SECTION
2
ENTITIES
0
TEXT
8
0
100
AcDbText
    
```

```
10
72
20
143
30
0.0
40
3
1
Contoh Kod DXF
50
0
0
ENDSEC
```

Fig.2. DXF code for text entity

As previously discussed, there are four attribute for entities which are involve in filtering process. These attributes have specific entities, which attach to them.

DXF file that is represented in tag data is isolated based on class of AutoCAD format. In this research, class is considered as entity. Few steps must be conducted in order to extract entities from engineering drawing. During the early stages, all the entities that are attached to DXF file will be stored in variable named `entity_list`. `Entity_list` is a collection of `System.Collections.Generic.List` that stores list of entities that exist in DXF file and have been opened. Each entity has basic attribute such as color, line type, entity type and layer. At this stage, all entity has yet to be filtered and stored as `entity_list` variable.

Fig 3 shows the flow chart of entity filtering technique which is used in order to filter the needed entities on engineering drawing.

After the entire entities are assigned to `entity_list`, these entities will be tested with a specific filter attribute through 'for' loop. Each entity will be compared with different type of attribute such as layer, color, entity type and line type. If the same value obtained same with the filter

attribute, the entity will be added to `filtered_entities` variable. The result of `filtered_entities` which consist of entities that have been filtered based on attribute can be displayed and numbers for each entity can be identified. Entity filtering technique must be developed in order to carry out entity filtering process for engineering drawing based on certain categories. Entities which have been filtered can be used as inputs in process of extracting information. For each attribute, it is represented by few entities.

There are four attributes namely layer, color, entity type and line type. Each entity is consisting of certain attributes such as Arcs, Lines, Texts, Circles and Ellipses which are under the Entity Type attribute. Various colors such as Red, Green and Blue are under Color attribute. Example of Line Type attribute includes Continuous and Center, whereas for the Layer attribute is represented by numbers such as 1, 2 or 3. These attributes simplify process of classification of entity where each attribute is represented by a specific entity. Table 3 highlights type of entities for each attributes.

Table 3. Type of entities for each attributes

Layer	Color	Entiti Type	Line Type
0	Red	Arcs	ByLayer
1	Blue	Lines	ByBlock
2	Cyan	Texts	Continous
3	Black	Circles	Center
4	Dark Grey	Ellipses	Dashdot
5	Light Grey	NurbsCurves	Dashed
6	Yellow	Points	Dot
7	Green	Faces 3D	

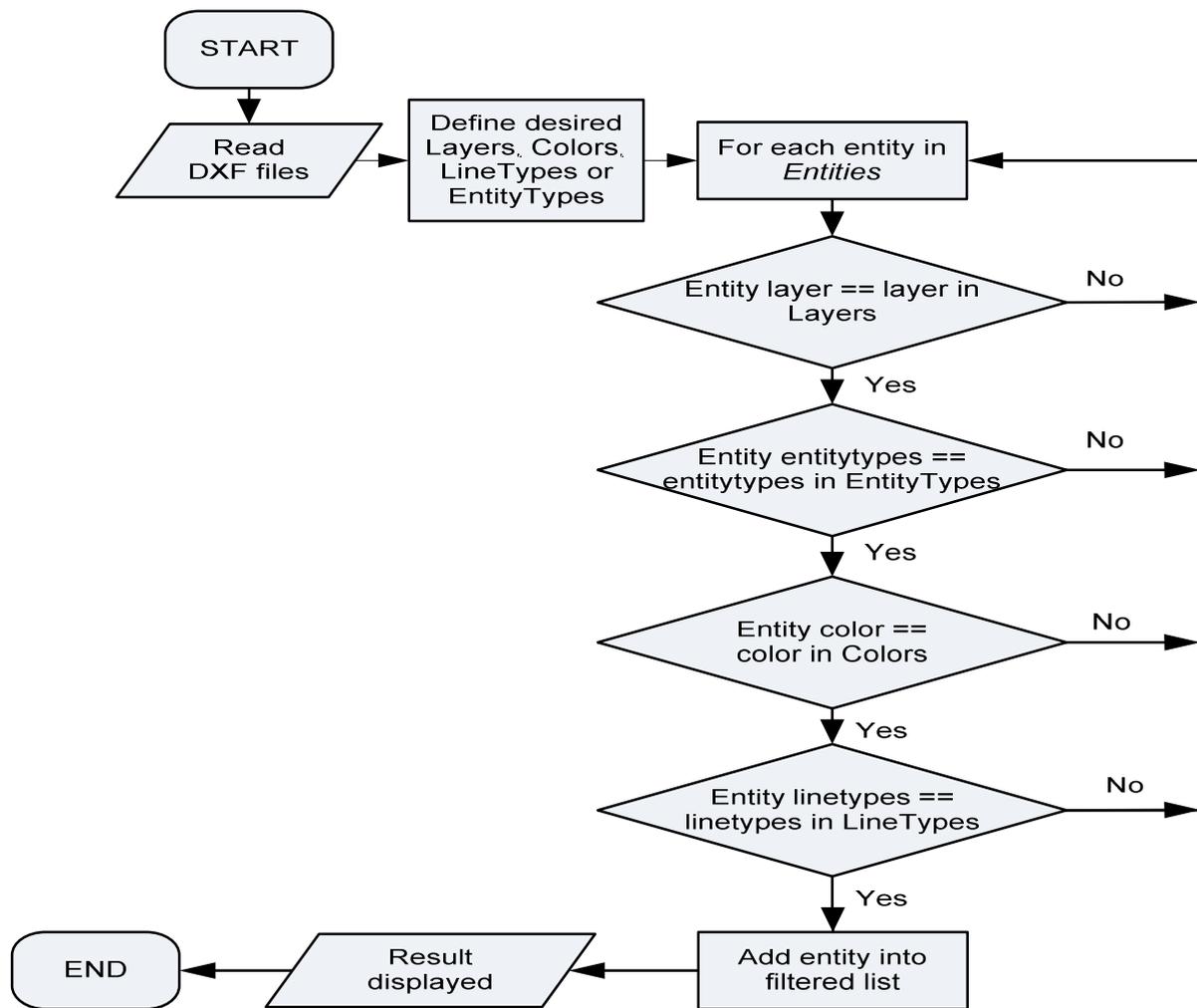


Fig.3. Flow chart of entity filtering technique

4. CONCLUSION

DXF files was developed by AutoDesk is a format that has been widely used in engineering. DXF file structure consisting of several sections contains information regarding the entities that represent the geometric and non-geometric information. In this paper, we propose a new entity filtering technique to extract the accurate information from engineering drawing files. Testing was done by comparing between developed technique and software AutoCAD 2010. The test results found that the technique is able to extract all the entities.

5. REFERENCES

[1] Xiantao C, Fazhi H, Xiaoxia L, Xin C. A Direct feature retrieval and reuse method for

heterogeneous collaborative CAD systems. In 3rd International Conference on Pervasive Computing and Applications, 2008, pp. 718-723

[2] Mohd F M A, Riza S, Saliyah K S, Muhammad F A R. A proposed information extraction technique in engineering drawing for reuse design. *International Journal of Social and Human Sciences*, 2012, 6(2):265-269

[3] Mohd F M A, Riza S, Suziyanti M, Saliyah K. A study on extraction method of non geometry information in engineering drawing title block. *Advanced Materials Research*, 2011, 383-390:995-999

[4] Mohd F M A, Riza S, Suziyanti M, Saliyah K, Zuraidy A. Extraction method of version number in engineering drawing. *Applied Mechanics and Materials*, 2012, 110:4856-4860

[5] Riza S, Mohd F M A, Nur A A M. A study on information extraction method of engineering drawing tables. *International Journal of Computer Applications* 2012, 50(16):43-47

[6] Yang M, Chao H, Zhang C, Guo J, Yuan L, Sun J. Effective clipart image vectorization through direct optimization of bezigons. *IEEE Transactions on Visualization and Computer Graphics*, 2016, 22(2):1063-1075

[7] Dew N, Grichnik D, Mayer-Haug K, Read S, Brinckmann J. Situated entrepreneurial cognition. *International Journal of Management Reviews*, 2015, 17(2):143-164

[8] Stone-Sundberg J, Kaminsky W, Snyder T, Moeck P. 3D printed models of small and large molecules, structures and morphologies of crystals, as well as their anisotropic physical properties. *Crystal Research and Technology*, 2015, 50(6):432-441

[9] Onoue Y, Kukimoto N, Sakamoto N, Koyamada K. E-grid: A visual analytics system for evaluation structures. *Journal of Visualization*, 2016, 19(4):753-768

[10] Weber N, Henrich A. Retrieval of technical drawings in DXF format-Concepts and problems. In *LWA*, 2008, pp. 213-220

[11] Sukimin Z, Haron H. geometric entities information for feature extraction of solid model based on DXF file. In *International Symposium on Information Technology*, 2008, pp. 1-5

[12] Autocad. General DXF file structure. California: Autodesk Inc., 2000

[13] Autocad. AutoCAD services and support-DXF reference. California: Autodesk Inc., 2017

[14] Ahmad N, Haque A. Manufacturing feature recognition of parts using DXF files. In International Conference on Mechanical Engineering, 2011, pp. 111-115.

How to cite this article:

Amran MFM, Sulaiman R, Ahmad RMTRL, Yusop NMM, Jamaludin NAA, Mohamed H. Entities filtering technique for information extraction in engineering drawing files. *J. Fundam. Appl. Sci.*, 2017, *9(3S)*, 199-209.