Data Extraction from DXF File and Visual Display

Han Zhang and Xueqing Li

School of Computer Science and Technology, Shandong University, China
zhanghan800418@mail.sdu.edu.cn, xqli@sdu.edu.cn

Abstract. DXF File is a graph exchanging file for CAD data exchange between other softwares provided by Autodesk Company. The house drawing of Shandong University is designed by AutoCAD software, and the graph information of house is recorded by the DXF file. Based on the DXF file, this paper will conduct an analysis on DXF file and extract the information from it. Then this paper will relate those extracted information with the information of house saved in the database by the automatic information retrieval algorithm and manual information retrieval algorithm. Finally, the information from database can be reclassified into four layers: base graph layer, room name layer, room area layer and room user layer.And those reclassified information will be displayed at the end of CS.

Keywords: parse DXF file, extract information, information retrieval, visualization.

1 Introduction

DXF file is a graph exchanging file for CAD statistics exchange between other softwares developed by Autodesk Company. It is a kind of file format for saving graph and data.

Housing drawings is made by the AutoCAD software, with the information recorded in the DXF files.But the information recorded in the DXF files is not related with the information stored in database.For this reason,this paper proposes two algorithms: the automatic information retrieval algorithm and manual information retrieval algorithm.Based on the two algorithms,we can relate the information recorded in DXF file with the information stored in the database.Then we can visually display more information stored in the database on the end of CS.

2 Related Work

2.1 Analysis of DXF File Structure

The key point to analyzing the DXF file is to know its structure intimately. A complete DXF file is composed of 7 segments, including the followings:

- HEADER SEGMENTS

C. Stephanidis (Ed.): HCII 2014 Posters, Part I, CCIS 434, pp. 286–291, 2014.

- CLASSES SEGMENTS
- TABLES SEGMENTS
- BLOCK SEGMENTS
- ENTITIES SEGMENTS
- OBJECTS SEGMENTS
- THUMBNAILIMAGE SEGMENTS

ENTITIES SEGMENTS includes all the graphic objects (graphic primitives) appeared in the graph, including the reference of block. Thus, this paper will extract all graph information from the segment.

Each segment of DXF file begins with group code "0" and character string "SECTION", and ends with group code "0" and character string "ENDSEC". Each segment is composed of several groups, each of which has a Group Code and a Group Value. The Group Code is used to record the character of data, and the Group Value records the value of data

2.2 Parse of DXF File

By parsing the DXF file structure, two lines of data, one is the group code and another the group value, will be read from the DXF file in the program. According to the value of group code, the corresponding value of group value is recorded. And during the process of the parse, the bound of DXF data elements is also recorded. This is the basis of information retrieval algorithms introduced later.

2.3 Proposed Algorithm

- Automatic information retrieval algorithm

The core concept of automatic information retrieval algorithm can be described as follows:

In the process of parsing DXF file, all text elements will be recorded. Then this paper will divide the text elements into three categories: room number, room area and others. The room number is matched to room number saved in the database.If it succeeds, the bound of room number will be considered as a standard. Then it finds out that all the line elements recorded in the DXF File elements are not contained in the bound. Then all the line elements are divided into horizontal lines and vertical lines. Based on the horizontal lines,this paper can determine the minimum value of Y and maximum value of Y and based on the vertical lines, it can determine the minimum value of X and the maximum value of X.Finally,this paper makes use of the value of X and Y to determine the bound of room

- Manual information retrieval algorithm

The core concept of manual information retrieval algorithm can be described as follows:

The first step is to select the bound manually. In this selected bound, the room number is supposed to be found out. Then this paper will match the room number to room number saved in the database. If it succeeds, the selected bound will be considered as a

standard. Then this paper finds out that all the line elements recorded in the DXF File elements are contained in the bound. In this paper all the line elements are divided into horizontal lines and vertical lines. Based on the horizontal lines, this paper can determine the minimum value of Y and maximum value of Y. And based on the vertical lines, this paper can determine the minimum value of X and the maximum value of X. Finally, this paper makes use of the value of X and Y to determine the bound of room.

2.4 Visual Display

Through the above information retrieval, the house information which has been successfully matched should be reorganized. This paper will divide information of house into four layers: base graph layer, room name layer, room area layer and room user layer. The information of different layers is differentiated by color and the visual display at the end of CS. On the basis of different focuses, this can also close different layers to show different house information.

2.5 Outcome of Experiment

The Figure 1 is displayed by AutoCAD software and the figure 2 shows that the house drawing is displayed by our system.

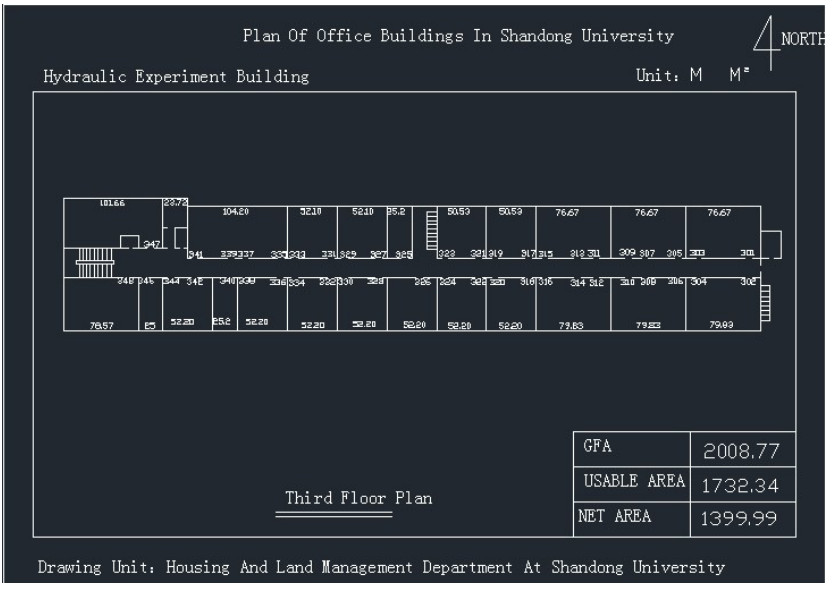

Fig. 1. The original AutoCAD drawings

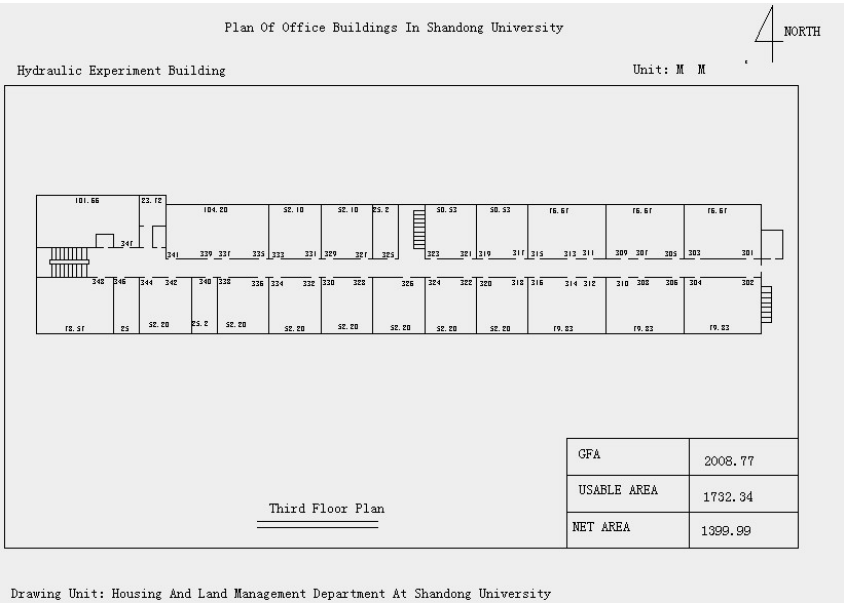

Fig. 2. House drawing displayed by our system

Figure 3 shows that the borders marked by red color are extracted by automatic information retrieval algorithm.

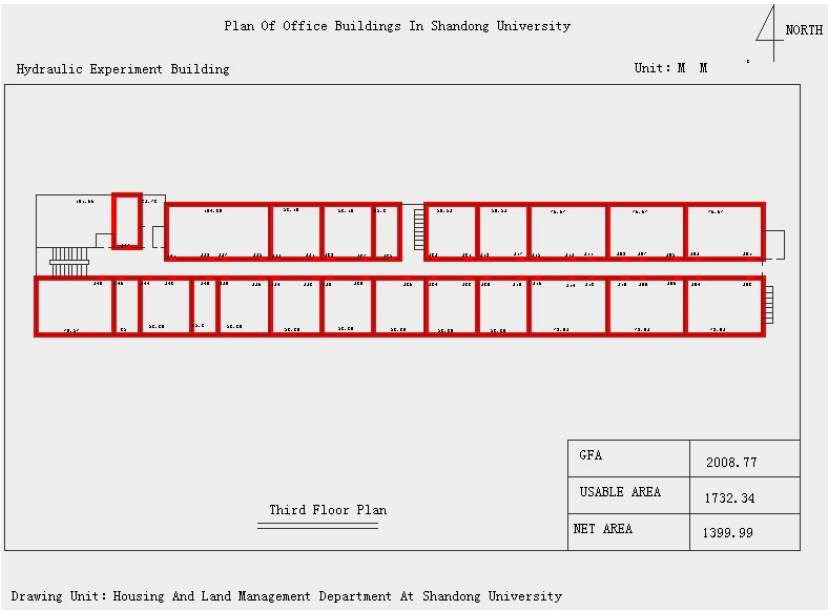

Fig. 3. Result of border extraction by automatic information retrieval algorithm

Figure 4 shows that more house information saved in the database is displayed in house drawing by our system. The information is divided into four layers: base graph layer, room name layer, room area layer and room user layer.And the different layer is marked by different color.

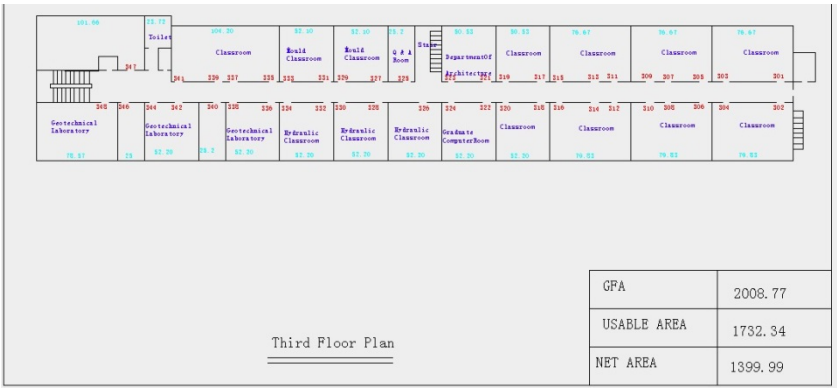

Fig. 4. Visual display after border extraction

We can find upper left corner of the border is not marked correctly.Figure 5 shows the final border by manual information retrieval algorithm and displays more information stored in database about house.

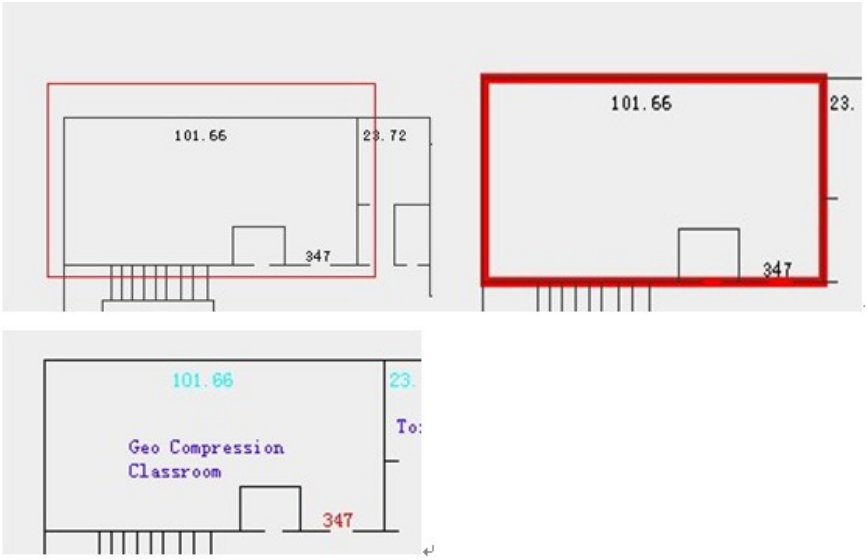

Fig. 5. The process of manual information retrieval algorithm and visual display

3 Conclusion

On the basis of conducting a complete analysis of File DXF and extracting data information of different segments of DXF file, this paper takes advantage of two retrieval algorithms to implement the association between information of house recorded in DXF file and information of house stored in the database. It is possessed of three advantages:

- More service information of the house can be displayed in the drawing, which makes the system more flexible.
- Keep the consistency of the house information in the drawing and the database and make sure the accuracy of the drawing display.
- The user can see through the drawing by the system without additional AutoCAD software.