DEVELOPMENT OF BUILDING AND ROAD SPATIAL DATABASE FOR PARIT BUNTAR NEW TOWNSHIP

By

TENGKU NOOR ARBAEE BINTI TG. AZHAR

This dissertation is submitted to

UNIVERSITI SAINS MALAYSIA

as partial fulfillment of requirements for the degree of

BACHELOR OF ENGINEERING (CIVIL ENGINEERING)

School of Civil Engineering,
Universiti Sains Malaysia

May 2006

ABSTRACT

The general scope of this project is the development of spatial database for road network system and building information of Parit Buntar New Township. The data capture process applies the utilization of satellite image of high resolution where road and building features can be extracted by means of screen digitization. Spatial database development was carried out using IDRISI GIS database workshop capability where all attributes regarding road network and building lots are obtained through field sampling process. The geo-relational data model that joins spatial data and attributes data was adopted in the development of database, as every spatial data are stored in graphic files while the attribute data are stored in table. There were linked through the features ID which is differentiated by color's unique value. The final output of this project is a spatial database system that provides attribute and spatial information of Parit Buntar New Township. There seem to be a demand in the development of database as it will became the mechanism for residents to get information and as reference for future development which can be up dated easily.

ABSTRAK

Skop umum bagi projek ini ialah pembangunan pangakalan data reruang bagi maklumat rangkaian jalan raya dan bangunan di Pekan Baru Parit Buntar. Proses pengumpulan maklumat ini mengaplikasikan penggunaan imej satelit beresolusi tinggi dimana cirri-ciri jalanraya dan bangunan boleh diekstrak dengan kaedah pendigitan di skrin. Pembangunan pangkalan data reruang dilakukan dengan menggunakan kebolehan Bengkel Pangkalan Data dalam IDRISI GIS. Semua atribut tentang rangkaian jalan raya dan lot bangunan adalah diperolehi melalui persampelan di tapak. Model data geo-hubungan menghubungkan data ruang dan data atribut diaplikasikan dalam pembangunan pangkalan data tersebut., di mana data ruang disimpan dalam fail grafik manakala data atribut disimpan dalam jadual. Kedua-dua data tersebut akan dihubungkan melalui ID yang akan dibezakan dengan memberikan nilai unik berdasarkan warna untuk setiap ID. Output akhir bagi projek ini ialah sistem pangkalan data yang membekalkan maklumat atribut dan reruang bagi Pekan Baru Parit Buntar. Terdapat permintaan terhadap pembangunan pangkalan data ini memandangkan ia merupakan sebuah mekanisme untuk penduduk setempat mendapatkan maklumat dan sebagai rujukan untuk pembangunan dimasa hadapan selain dapat dikemaskinikan dengan mudah.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background

Geographical Information Systems or GIS is a system comprised of computer software, hardware and data, and technical expertise to help manipulate, analyze and present information that is tied to a spatial or geographical location. GIS are now widely used in areas such as commercial, local government, defense and academic research. GIS may then be seen to be taking over and greatly extending the role of spatial data storage which was previously played by maps. GIS can provide tools for data manipulation and analysis. The visualization are far more versatile than any associated with traditional cartography.

In modern map production, a shift has taken place from maps that stored in analogue form on paper of film to a digital database map which contain information about each places such as name of shop, type of business and also the contact number. In order to cope with the higher product demands, the automated tools in the production should be employed. This will increase the productivity and reduce cost as well as less time consuming option compared to updating analogue map.

Feature extraction from high resolution satellite images has drawn considerable attention due to the need for the efficient acquisition and updating of the data. Feature extraction of satellite images produces spatial data which is one of the main components in geographically referenced data. Besides spatial data, attribute data is another component in geographically referenced data. For example, the description of a building refers to its location which is represented by spatial data and the name of the building as the attribute data.

The spatial data describe the location of spatial features which may be discrete or continuous. Discrete features are individually distinguishable features that do not exist between observations such as lines, areas and points. In the other hand, the continuous features are features that exist spatially between observations such as the elevation and precipitation. The locations of spatial features on the Earth's surface are based on geographic coordinate system with longitude and latitude value, whereas the locations of maps features are based on a plane coordinate system with x and y coordinates.

Attribute data describe the characteristic or quality of spatial data that are being represented. Heights, land use, name of shop and vegetation type are example of attribute data that are normally related to features in map. Attribute data are stored in features attribute tables which is organized by row and column. For vector data, the amount of attribute data to be associated with a spatial feature can vary significantly, whereas for raster data, each cell has a value that corresponds to the attribute of spatial feature at that location.

In this project, geo-relational data model that joins spatial data and attribute data was adopted. This model will store the spatial data in graphic files whilst the attribute data are store in tables and there are linked through the features ID. A database is a collection of information that is related to a particular subject or purpose which stored in tables. There are four types of database design; which are flat file, hierarchical, network, and relational.

The relational database model is normally used for GIS database management system. The relational database is a collection of tables (relations) that can be connected through a key; common fields whose values can be uniquely identify the record in a table. A relational database is efficient and flexible for data search, data retrieval, data editing and creation of tabular reports. Thus it is very suitable in GIS approach as each table in the database can be prepared, maintained, and edited separately from other tables.

1.2 Location of study area

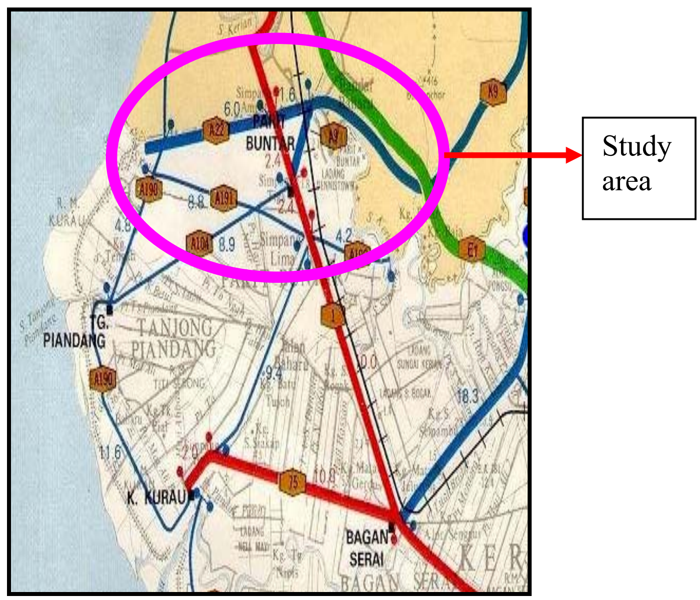

Figure 1.0 Location of Parit Buntar

The study area of this project is focused on the new town commercial lot area of Parit Buntar, Perak. This town is situated in the north of the Kerian district and is bordered by the towns,Bandar Baharu, Kedah and Nibong Tebal,Penang. Kerian is an administrative district in Perak. Local Administration is known as Majlis Daerah Kerian or Kerian District Council based at Parit Buntar.

According to history, the name Parit Buntar originated from a well-known leader, Tok Buntar, who was famous at the time when he and his followers built had successfully the drainage system which used water from the Kerian River to irrigate the surrounding paddy fields. In honor of Tok Buntar, the main drain was named "Parit Tok Buntar" meaning the drain of Tok Buntar. Later on, the town where the main drain is located was named Parit Buntar.

The Clock Tower which is situated in the old town of Parit Buntar had been the landmark for this town. The tower was built in early 1961 and was officiated by the First Prime Minister of Malaysia, the Almarhum Tunku Abdul Rahman Al-Haj on 24th August 1961.

1.3 Problem statement

Parit Buntar is popular as a shopping destination especially for residents in Northern Perak, Southern of Seberang Perai and Southern of Kedah. There are many competitive bargains for the customer as there are so many shops open especially in the new town of Parit Buntar which is the study area. There are many branches of various major financial institutions in Parit Buntar, thus it can be said as important commercial centre in the northern region. Besides that, there are also many medical clinics, automobile centre, fashion boutique and telecommunication center.

There seem to be demand in the development of digital spatial database informatioj for Parit Buntar Township. The development of an up to date digital map of commercial lots in Parit Buntar new town will provide the mechanism for information provision to the residents as well as making it good tourist destination. The database information will also be used as reference for future planning and development of infrastructure.

1.3 Objective of Study

The main objectives of the project are:-

(i) To develop up to date digital map of commercial area in Parit Buntar new town by means of road feature extraction method from high resolution satellite image.

(ii) To build a spatial database information regarding road feature and commercial sector in study area.

1.4 Scope of Report

The output of this project is a spatial database for Parit Buntar New Township.

The literature was first reviewed in Chapter 2 of this report.

In Chapter 3, the complete methodology flowchart for this project was shown. Then all the GIS activities involved in this project were explained.

In Chapter 4 of this report, the results of this project which are the digitized plan, vector file in IDRISI32 and the database table was presented.

The conclusion and suggestions to improve the result were given in Chapter 5.

CHAPTER 2

LITERATURE REVIEW

2.1. Introduction of Geographic Information System

A geographic information system (GIS) is a computer system for capturing, storing, querying, analyzing, and displaying geographically referenced data as described in Chang (2004). Geographically referenced data is data that describe both the location and characteristic of spatial features such as roads and land parcels on Earth's surface. The ability of GIS to handle and process geographically referenced data distinguishes GIS from other information systems and it has become new Information Age Paradigm. GIS technologies have been broadly applied in many fields such as urban planning, environment management, agriculture, transportation, utilities etc.

According to Haithcoat and Song (2001), millions dollars are invested to transform the paper maps into digital format, but many of these GIS data are outdated especially the cultural features such as roads and buildings which have change frequently because humans build new roads and buildings constantly to improve quality of life. Thus, a current road database is needed and automated data extraction from remotely sensed imagery can be used in this updating process.

Unfortunately, as discussed by Davis (2001) there is no real "automatic digitizing" because human intervention is necessary during the process. For example, the line-following software cannot make decision about the features either they are features or annotations. Hence, automatic digitizing is really a semi-automatic process as human monitoring and intervention are necessary during those digitizing and editing works. In order to achieve the fully automated technology, the road feature extraction was one of the popular research topics in computer vision, photogrammetry, remote sensing and GIS communities nowadays. Feature extraction is expected to be an area of significant improvement and advance during the next decade, particularly in the extraction of spatial features (Haithcoat and Song, 2001).

2.2. Automatic Road Feature Extraction

Automated feature extraction is a technology with broad applicability for analysis and interpretation of data obtained through remote sensing (satellite and aerial photo imagery) which can be used for a wide variety of tasks and applications such as in urban planning management and development of roads database. In addition, this technology has drawn considerable attention as it can significantly reduce the cost of data acquisition and database development. Besides that, it is also time consuming especially for roads because they are the most time consuming features to digitize in photogrammetry. A great amount of researches has been targeted on fulfilled the need for the efficient acquisition and updating of road data for geo-databases, therefore the automated road features extractioj has been key research issues in GIS communities recently.

The process of extraction are useful in segregating human-made features and it is intended to isolate features for inteppretation, classification, counting and determining spatial distribution. According to Ed Granzow (2001) the procedures of automated road feature extraction allow relevant features and their outlines to be identified through digital imagery to enhance and isolate the feature attributes. In the other hand, automated feature extraction allows the identification of relevant features and their outlines by post-processing the digital imagery through technique to enhance and isolate feature definition. There are some examples of types of post processing techniques utilized that had been given in Ed Granzow (2001) which include mathematically strengthening feature to background contrast, eliminating image noise and pattern recognition.

In (Heipke et al., 1998), approaches for road extraction use one or both of the following two properties of roads: in low resolution images road are usually modelled as lines, whereas in high resolution imagery methods are considered as homogeneous, elongated area with parallel roadsides. Referring to Chang (2004), the spatial resolution of the satellite image relates to ground pixel size. For example, a spatial resolution of 20 meter means that each pixel on the satellite image corresponds to a ground pixel of 20 meters by 20 meters. High spatial resolution is defined as being less than 2 meters and low resolution being greater than 2 meter per pixel.

The pixel value in a satellite images represents light energy (based on spectral bands from continuum of a wavelengths knows as electromagnetic spectrum) that reflected or emitted from Earth’s surface. In this study, QuickBird’s high resolution satellite imagery data is use for feature extraction process. QuickBird collects panchromatic images (comprised of a single spectral band) with 61cm spatial resolution and 2.44m resolution for multispectral images (comprised of multiple bands).

There are several factors influencing the quality of the road extraction such spatial resolution, spectral information (contrast), and context (rural, urban) as state in Haithcoat and Song, (2001). The research had attempted to quantify these factors and examine their inter-relationships. The imagery's spatial resolution is a key determinant in choosing which procedure is most applicable. Various extraction processes have been developed, applied and evaluated for both spatial resolutions.

It is critical that when applying any feature extraction methodology, the physical characteristics of feature types had to be clearly defined. Algorithms have been formulated that compensate for shadows, variation in width and geometric problems associated with width, steepness and curvature. These algorithms will use the characteristics to recognize the roads during evaluation and mapping processes from the imagery. We take account of roads characteristics as discussed in Eker and Seker, (2004) which can be classified into five groups:

- geometry (e.g. : elongation, curvature etc)
- radiometry (e.g. : assume road surface is homogenous)
- topology (e.g. : road intersect and build a network)
- functionality (e.g. : roads connect cities)
- contextual (e.g. : fly-over that may cast a shadow)

In fully automatic road extraction, the computer algorithms find and extract roads without human interaction. Detecting and following of lines are the most common techniques for road extraction in low-resolution imagery (resolution >1m). For road extraction in high-resolution imagery which is resolution between 0.2-0.5m, roads are considered as elongated homogeneous regions and are detected by matching profiles and parallel edges, and multi-scale images are often used and combined to produce a better result.

The automatic road extraction process includes three basic procedures: (1) image preprocessing, (2) road extracting, and (3) GIS processing as state in Haithcoat and Song, (2001). The process was illustrated as in Figure 2.1. Image preprocessing has to be conducted in order to facilitate or enable further operations, but this process varies depending on the image characteristics. In this preprocessing stage, an automatic brightness threshold method was applied through elimination of brighter/darker area slightly improved final result (Hasegawa, 2004).

Then the road extraction stage, algorithms were applied on these reduced resolution images. Images were convolved with derivatives of Gaussian kernels. In current research, (Hasegawa, 2004) correctness constantly increase as Gaussian kernels parameter's become bigger and the research suggested that if the kernel is too small, it picks up much noise and results in many false road segment. A Hessian matrix was the second partial derivatives in row and column direction and the mixed second derivative (Haithcoat and Song, 2001). A raster-to-vector conversion was conducted after threshold the second derivative image, to extract the vector roads. GIS operations were then applied to refine the results both topologically and aesthetically.

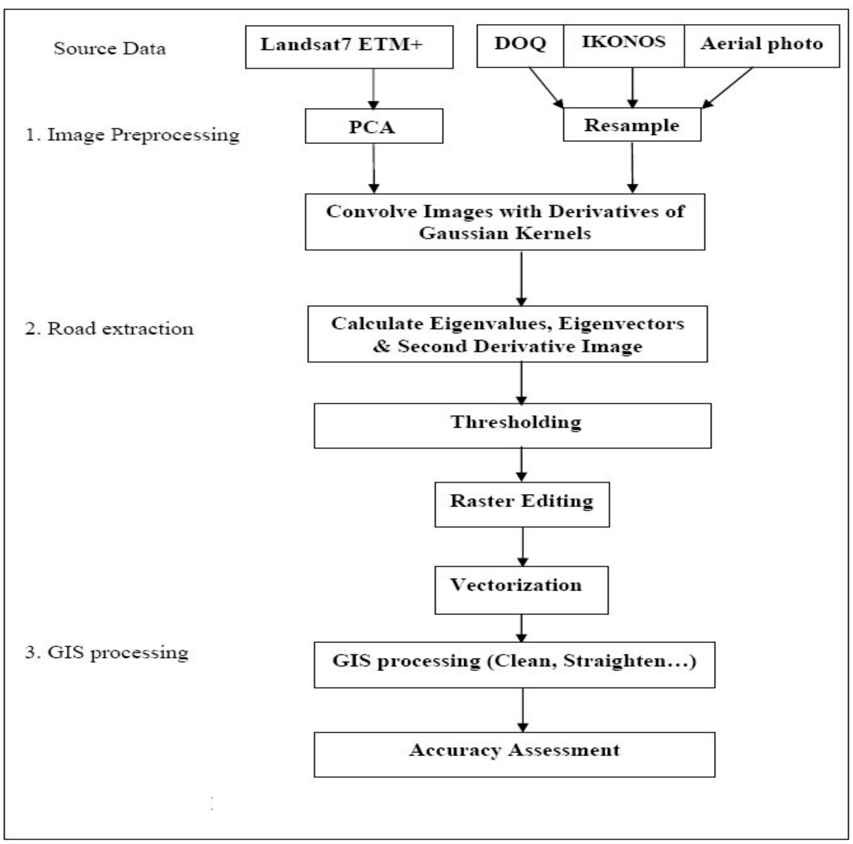

Figure 2.1: Automatic road extraction procedure. (Haithcoat and Song, 2001)

In extraction process of roads from high-resolution images, we need to identify the elongated homogeneous areas within the image which have parallel step edges and a semi-consistent width that need to be included. Treash and Amaratunga (2000) had proposed an automatic road detection system using high resolution images which consists of three parts; edge detection, edge thinning, and edge pairing. The result from that is good in rather simple image while it show poor performance in dense resident area. For that reason, those system cannot be used in this study as Parit Buntar is a quite dense area.

One of the quality measurements that will be used for the evaluation of road extraction results is Root Mean Square (RMS) difference. The RMS difference expresses the average distance between the extracted and the reference network, thus the geometrical accuracy of the extracted road data (Heipke, 1998).

$$RMS = \sqrt{\frac{\sum\left[\left(X_{extracted} - X_{reference}\right)^2 + \left(Y_{extracterd} - Y_{reference}\right)^2\right]}{n}} \quad \textbf{(2.1)}$$

n =number of pieces of matched extraction

The value depends on the buffer width. If an equal distribution of the extracted road data within the buffer around the reference network is assumed, it can be shown that:

$$RMS = \frac{1}{\sqrt{3}} \times buffer\ width \quad \textbf{(2.2)}$$

The optimum value for RMS is 0.

2.3. Development of Database

The result's map from the automated road feature extraction will be used in development of up to date database for digital road map of Parit Buntar. The example of geographically referenced data to describe a road is its location (i.e., where it is) and its characteristics (i.e., length, name, class, speed limit, and direction). Spatial data describe the location of the spatial features, which usually has specific location according to some world geographic referencing system (such as Latitude-Longitude) or address system. In the other hand, the attribute data describe the characteristics of the map features.

Attribute data are store in table as state in Chang (2004) and each row of table represent a map feature and each column represent a characteristic. Therefore intersection of a column and a row shows the value of a particular characteristic for a particular map feature. A row is also called a record and a column is also called a field. Each column describes an attribute of the map feature (Figure 2.2). Each row represents a map feature and each column represents a property or characteristic of the map feature.

Label - ID	pH	Depth	Fertility
1	6.8	12	High
2	4.5	4.8	Low

Row

Column

Figure 2.2: A feature attribute table consists of roe and column. (Chang 2004)

According to Chang (2004), a database is a collection of interrelated tables on digital format. Without the database GIS would be little more than computer cartography. There are four types of database design; flat file, hierarchical, network, and relational, but the best types for GIS database is relational database (Davis, 2001). The relational database design is simple and flexible compared to other database design. In (M. Rahimi, 2004), state that the GIS interface of the system is coupled to an attribute information base residing in an RDBMS (Relational Database Management System), thereby providing for extensive database querying in a user-friendly environment.

A relational database is a collection of tables (relations), which can be connected to each other by keys. It is an integrated table which had been constructed so that each record and its attributes are linked and related (cross-referenced) to all records and their attribute. For example, in Figure 2.3; the key that connecting zoning and parcel is zonecode and the key that connecting parcel and owner is the PIN (parcel ID number).

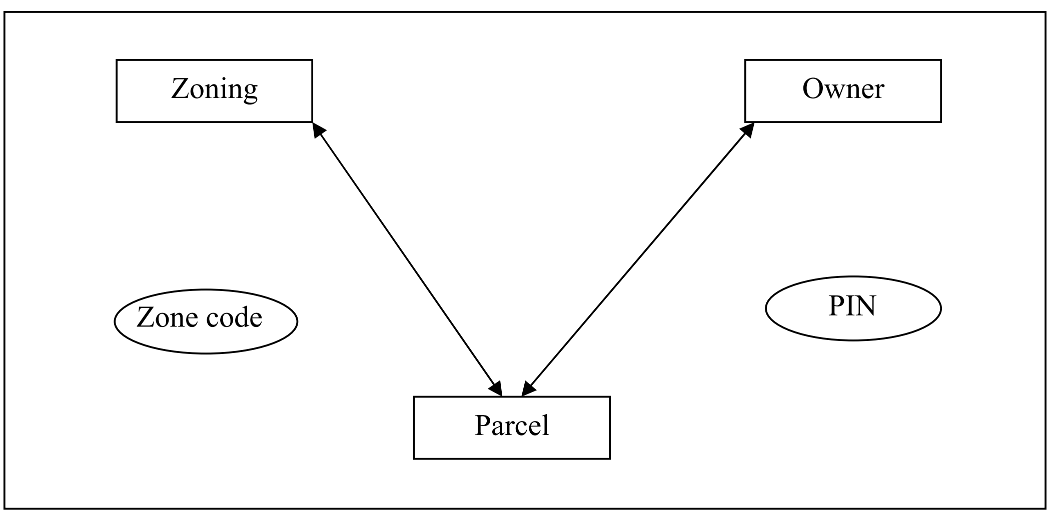

Figure 2.3: Relational database design. (Chang, 2004)

Referring to Chang (2004), a relational database may contain four types of relationship between tables. Those relationships are (Figure 2.4):

- **One-to-one relationship**: one and only one record in a table is related to one and only one record in another table.
- **One-to-many relationship**: one record in a table may be related to many records in another table.
- **Many-to-one relationship**: many records in a table may be related to one record in another table.
- **Many-to-many relationship**: many records in a table may be related to many records in another table.

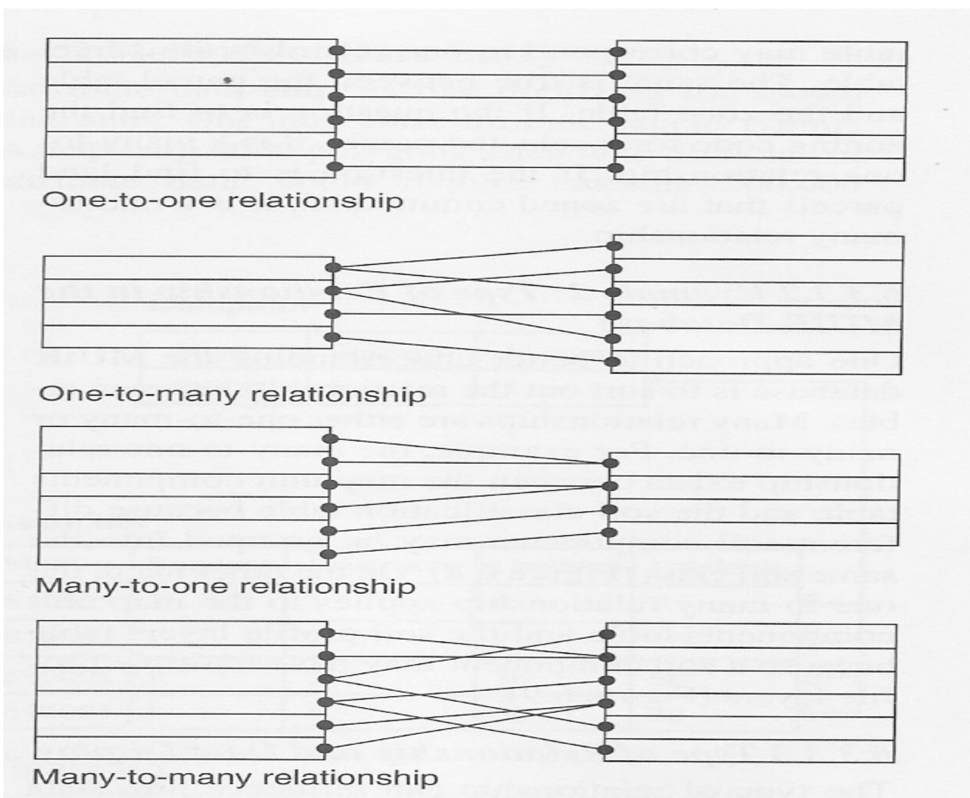

Figure 2.4: Types of data relationship between tables. (Chang, 2004)

CHAPTER 3

METHODOLOGY

3.1 Introduction

Basically, the GIS activities involved in this project are concentrated on spatial data input, attribute data management, data display and data exploration which is summarized as follow:

(i) Spatial Data Input

The spatial data input covers the most expensive part of a GIS project which is data acquisition. It has to basic option; either uses the existing data or creates new data. New digital spatial data can be created from satellite images, GPS data, field surveys, street addresses and text files with *x* and *y* referenced coordinates but paper maps remain as the dominant data source. Paper maps and satellite image can be converted into digital format by manual digitizing or scanning. In this project, the main source of data input is digitizing using satellite image as the base image. Editing can be done later to remove the digitizing error and the geometric transformation will convert newly digitized map into real-world coordinate system. Here, the satellite image of Parit Buntar was obtained from the School Of Civil Engineering. It has 6 cm cell resolution which makes it possible for feature extraction through digitizing.

(ii) Attribute Data Management

The attribute data management was performed using the relational database concept. There are two basic elements in the design of a relational database, which are the key and type of data relationship. The key will established a connection between corresponding records in two tables and the type of data relational represent how the tables are actually joined.

(iii) Data Display

Map display is a routine of GIS operation for data visualization, query, analysis and presentation as maps are the most effective in delivering spatial information. A well-designed map can help the mapmaker communicate spatial information to the map user although the user is not so familiar with GIS environment. The data display method in this project was the IDRISI32 version which is linked to a simple database workshop capability.

(iv) Data Exploration

Data exploration is data-centred query and analysis that will allow user to explore the general trends in the data. User can take a close look at the data subsets and focus on possible relationship between datasets. Interactive and dynamic linked visual tool is required for effective data exploration. Besides that, interactive features in data exploration can increased our capacity for information processing and synthesis.

3.2 The Flowchart of Methodology

The complete methodology flowchart for this project describe in Figure 3.0., while tha complete GIS activities that was carried out in this project is describe in Figure 3.1. A detail procedure of each activity is explained in the following sub-sections.

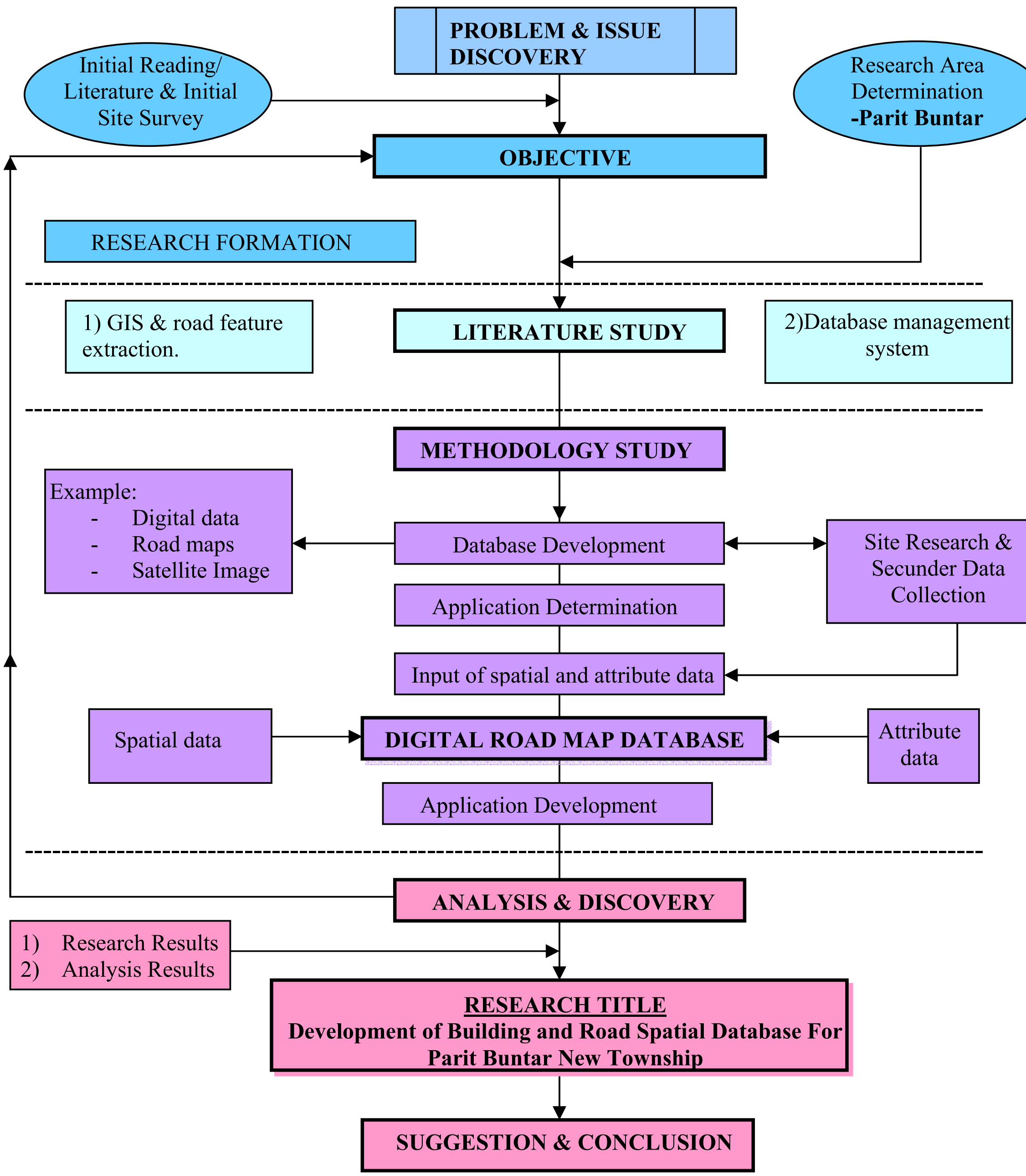

Figure 3.1: Methodology flowchart

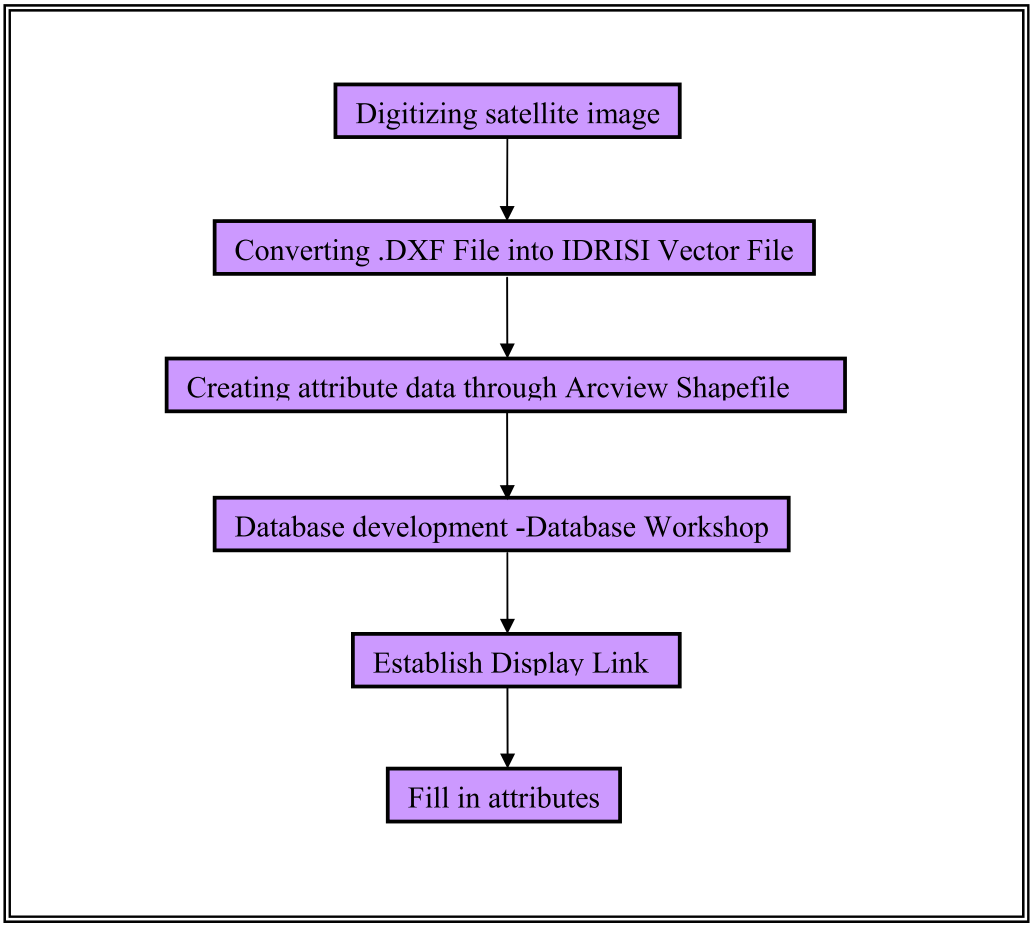

Figure 3.2: Flowchart for GIS activities

3.1.1 Digitization

In this project, the AutoCAD 2005 software tool was used for on-screen digitizing. The data source is the satellite image of Parit Buntar as the background. The satellite image was viewed in Autocad window using the raster image selection file, as shown in Figure 3.3.

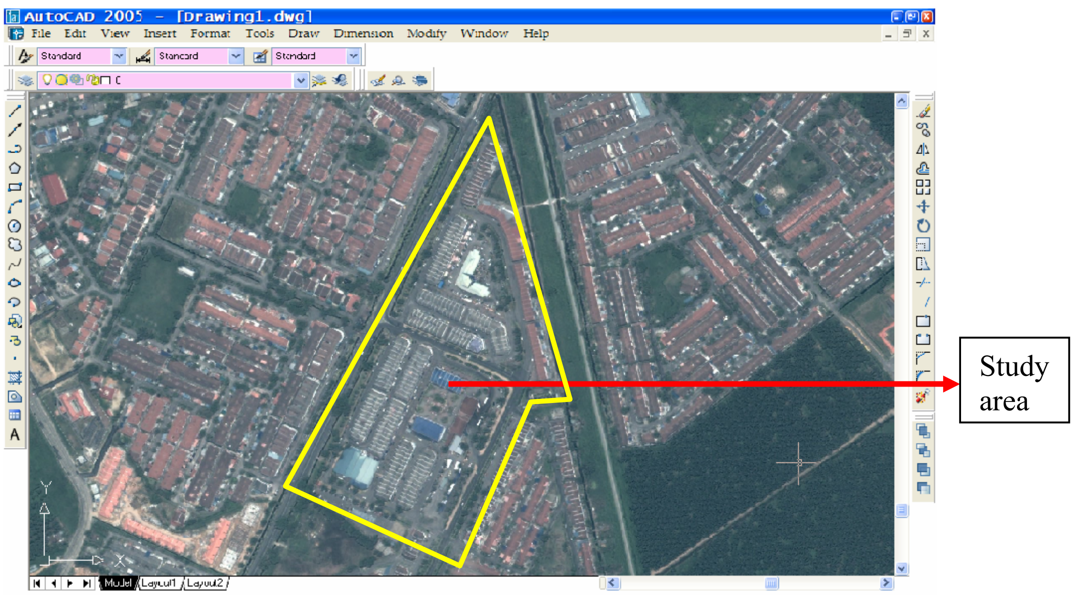

Figure 3.3 Satellite image as background image

The construction lines were used as references for creating other objects. In this project, construction lines were used to estimate the edge of the building and in calculating the exact dimension of the building before the actual digitize lines were built from the background image. Those construction lines were built in different layer from the actual digitize lines, so that it can be associated it with the actual digitize line's layer.

After all the construction lines were built, then digitize polygon of the building was made. The line colour of the polygon must be in unique value for each building so that it could be identified in IDRISI32 database workshop as different identifier. After all the buildings in the study area were digitized, then the Color Properties command was selected to assigned the correct color, as shown in Figure 3.4. Those steps were repeated for each building polygon so that the unique ID could be obtained. The completed digitized image with the background image is shown in Figure 3.5.

The process is completed by saving the file in AutoCAD R12/LT2 DXF [.dxf] which is in drawing interchange format, so that it can be exported to IDRISI32.

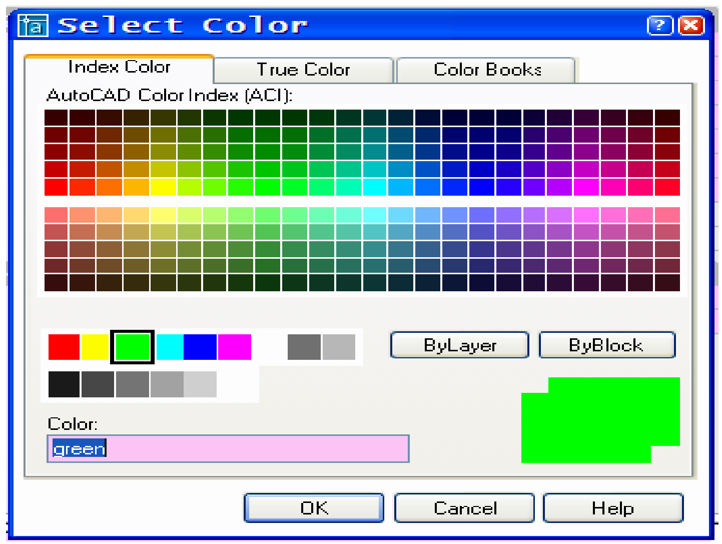

Figure 3.4: Colours selection

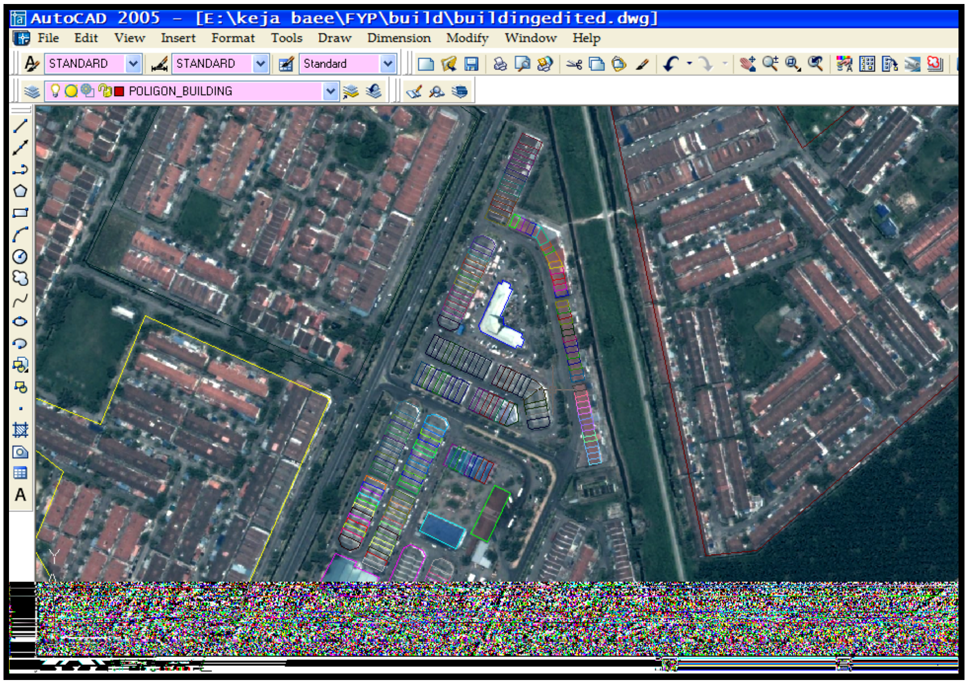

Figure 3.5: Digitize image

3.2.2 Converting AutoCAD .DXF File into IDRISI Vector File.

The conversion of DXF to IDRISI requires some knowledge of the DXF file such as the feature types in the file. The conversion uses, the Import menu option in IDRISI where the DXFIDRIS menu was chosen, as the export format. This process required the creation of new file and naming the attribute value file. In this project the attribute was stored as color code to represent the attribute value. Figure 3.6 shows the DXFIDRISI conversion module which successfully converted DXF file to IDRISI vector file.

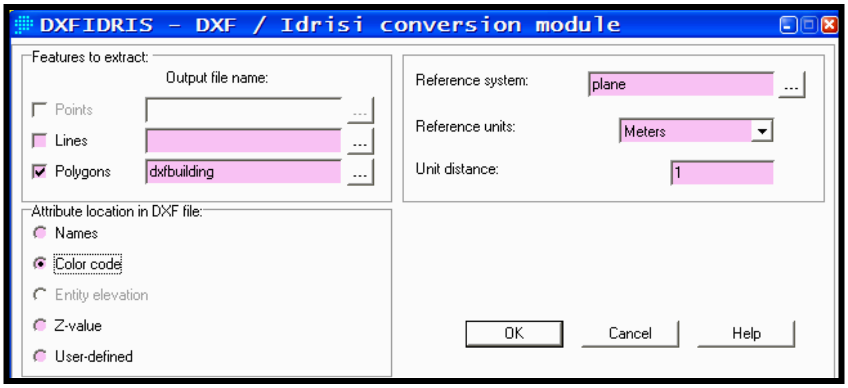

Figure 3.6: DXF / IDRISI conversion module

3.2.3 Attribute Data Conversion

In the initial stage of attribute data development, ArcView 3.1 was used. The conversion of DXF file to Shapefile was performed. The display of Shapefile in ArcView 3.1 was as shown in Figure 3.7. Shapefile in ArcView automatically creates a database file which is compatible with Microsoft ACCESS .mdb format. This format can be read in IDRISI32 Database Workshop tool. The identifier (ID) of attribute data is represented by unique colors values ranging from 0-256. The legend editor as shown in Figure 3.8 can be used to edit those unique ID. Once the process is completed; the Shapefile with attributes can be read or exported to IDRISI32.

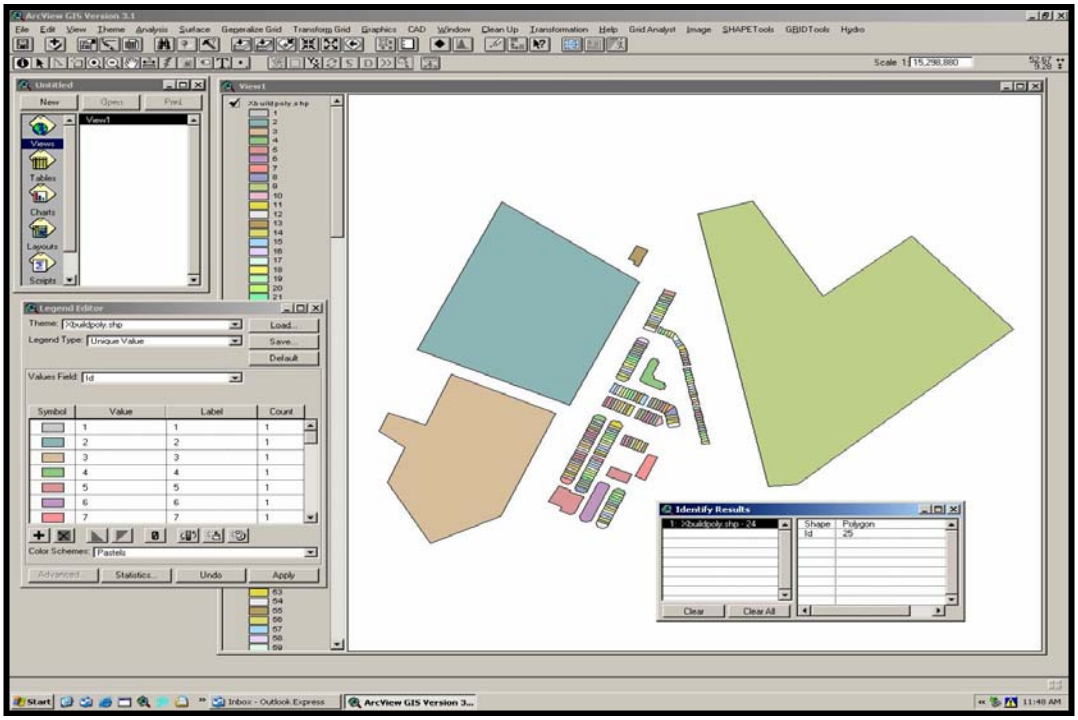

Figure 3.7: Shapefile in ArcView

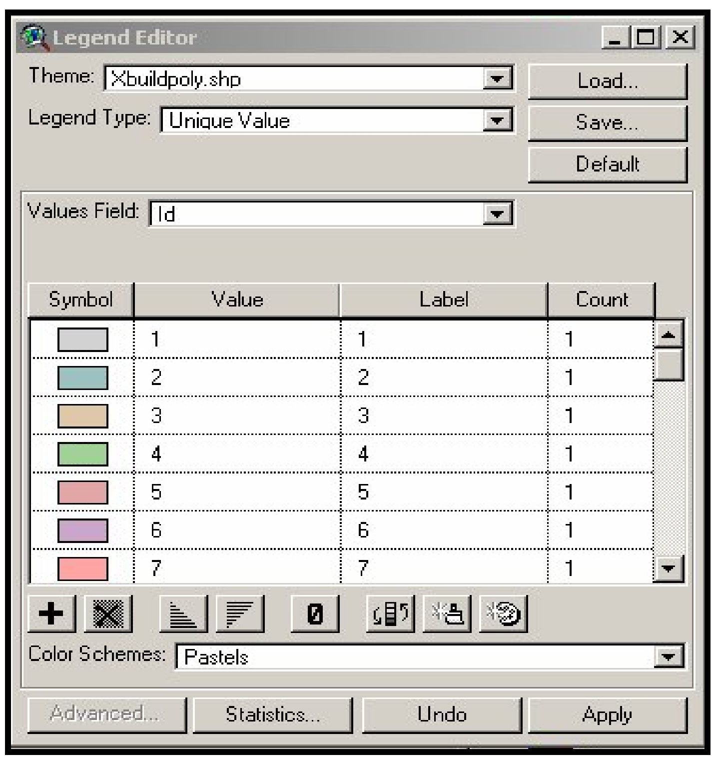

Figure 3.8: Legend Editor

3.2.4 Database Workshop

Database Workshop provides an integrated relational database management system. It has an extensive set of capabilities of creating, editing and analyzing database file in IDRISI32. Database Workshop assigned the IDRISI ID number "IDR_ID" during the conversion process which can be edited later. The database can be changed by adding a new field or record using the Edit command. The field name and its data type had to be specified. In this project, the type of data is text as it contains information of shops unit such as the names of the shop and the descriptions about that shop.

The addition of new record in database is possible uqing the New Record command, and it depends on single feature or entity. As each new record is added, the Status Bar will increase the reported number of records. All the changes that were made in Database Workshop were automatically saved. In this project, no new record were made as the attribute value from the Shapefile is enough and adequate. Figure 3.9 described the Database Workshop Display in IDRISI32, while Figure 3.10 was example of Database Workshop with the attributes data.

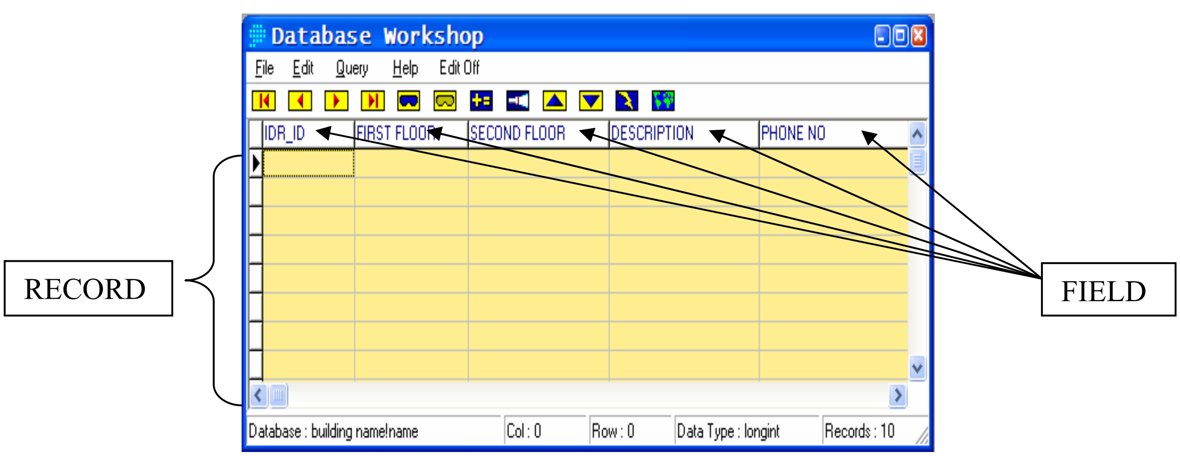

Figure 3.9: DATABASE WORKSHOP Interface.