FUNDAMENTALS OF GEOGRAPHICAL INFORMATION SYSTEM

P.L.N. Raju
Geoinformatics Division
Indian Institute of Remote Sensing, Dehra Dun

Abstract : The handling of spatial data usually involves processes of data acquisition, storage and maintenance, analysis and output. For many years, this has been done using analogue data sources and manual processing The introduction of modern technologies has led to an increased use of computers and information technology in all aspects of spatial data handling. The software technology used in this domain is Geographic Information Systems (GIS). GIS is being used by various disciplines as tools for spatial data handling in a geographic environment. This article deals with history of GIS and fundamentals of GIS such as elements of GIS; data models; data structures & data base; and elementary spatial analysis.

INTRODUCTION

We are presently positioned at the beginning of the twenty first century with the fast growing trends in computer technology information systems and virtual world to obtain data about the physical and cultural worlds, and to use these data to do research or to solve practical problems. The current digital and analog electronic devices facilitate the inventory of resources and the rapid execution of arithmetic or logical operations. These Information Systems are undergoing much improvement and they are able to create, manipulate, store and use spatial data much faster and at a rapid rate as compared to conventional methods.

An Information System, a collection of data and tools for working with those data, contains data in analog form or digital form about the phenomena in the real world. Our perception of the world through selection, generalization and synthesis give us information and the representation of this information that is, the data constitute a model of those phenomena. So the collection of

data, the database is a physical repository of varied views of the real world representing our knowledge at one point in time. Information is derived from the individual data elements in a database, the information directly apparent i.e. information is produced from data by our thought processes, intuition or what ever based on our knowledge. Therefore in a database context the terms data, information and knowledge are differentiated. It can be summarized that the data are very important and add value as we progress from data to information, to knowledge. The data, which has many origins and forms, may be any of the following:

- Real, for example the terrain conditions etc.

- Captured, i.e. recorded digital data from remote sensing satellites or aerial photographs of any area.

- Interpreted, i.e. land use from remote sensing data.

- Encoded, i.e. recordings of rain-gauge data, depth of well data etc.

- Structured or organized such as tables about conditions of particular watershed.

CONCEPTS OF SPACE AND TIME IN SPATIAL INFORMATION SYSTEMS

Spatial information is always related to geographic space, i.e., large-scale space. This is the space beyond the human body, space that represents the surrounding geographic world. Within such space, we constantly move around, we navigate in it, and we conceptualize it in different ways. Geographic space is the space of topographic, land use/land cover, climatic, cadastral, and other features of the geographic world. Geographic information system technology is used to manipulate objects in geographic space, and to acquire knowledge from spatial facts.

Geographic space is distinct from small-scale space, or tabletop space. In other words, objects that are smaller than us, objects that can be moved around on a tabletop, belong to small-scale space and are not subject of our interest.

The human understanding of space, influenced by language and cultural background, plays an important role in how we design and use tools for the processing of spatial data. In the same way as spatial information is always

related to geographic space, it relates to geographic time, the time whose effects we observe in the changing geographic world around us. We are less interested in pure philosophical or physical considerations about time or space-time, but more in the observable spatio-temporal effects that can be described, measured and stored in information systems.

Spatial information systems

The handling of spatial data usually involves processes of data acquisition, storage and maintenance, analysis and output. For many years, this has been done using analogue data sources, manual processing and the production of paper maps. The introduction of modern technologies has led to an increased use of computers and information technology in all aspects of spatial data handling. The software technology used in this domain is geographic information systems (GIS).

A general motivation for the use of GIS can be illustrated with the following example. For a planning task usually different maps and other data sources are needed. Assuming a conventional analogue procedure we would have to collect all the maps and documents needed before we can start the analysis. The first problem we encounter is that the maps and data have to be collected from different sources at different locations (e.g., mapping agency, geological survey, soil survey, forest survey, census bureau, etc.), and that they are in different scales and projections. In order to combine data from maps they have to be converted into working documents of the same scale and projection. This has to be done manually, and it requires much time and money.

With the help of a GIS, the maps can be stored in digital form in a database in world co-ordinates (meters or feet). This makes scale transformations unnecessary, and the conversion between map projections can be done easily with the software. The spatial analysis functions of the GIS are then applied to perform the planning tasks. This can speed up the process and allows for easy modifications to the analysis approach.

GIS, spatial information theory, and the Geoinformatics context

Spatial data handling involves many disciplines. We can distinguish disciplines that develop spatial concepts, provide means for capturing and processing of spatial data, provide a formal and theoretical foundation, are application-oriented, and support spatial data handling in legal and

management aspects. Table 1 shows a classification of some of these disciplines. They are grouped according to how they deal with spatial information. The list is not meant to be exhaustive.

Table 1: Classification of disciplines involved in spatial analysis

Characteristics of disciplines	Sample disciplines
Development of spatial concepts	Geography Cognitive Science Linguistics Psychology
Means for capturing and processing spatial data	Remote Sensing Surveying Engineering Cartography Photogrammetry Formal and theoretical foundation Computer Science Expert Systems Mathematics Statistics
Applications Legal Sciences	Archaeology Architecture Forestry Geo-Sciences Regional and Urban Planning Surveying Support Economy

The discipline that deals with all aspects of spatial data handling is called Geoinformatics. It is defined as:

> Geoinformatics is the integration of different disciplines dealing with spatial information.

Geoinformatics has also been described as "the science and technology dealing with the structure and character of spatial information, its capture, its classification and qualification, its storage, processing, portrayal and dissemination, including the infrastructure necessary to secure optimal use of this information". It is also defined as "the art, science or technology dealing with the acquisition, storage, processing, production, presentation and dissemination of geoinformation."

A related term that is sometimes used synonymously with geoinformatics is Geomatics. It was originally introduced in Canada, and became very popular in French speaking countries. The term geomatics, however, was never fully accepted in the United States where the term geographical information science is preferred.

There is no clear-cut definition for GIS. Different people defined GIS according to capability and purpose for which it is applied. Few of the definitions are:

- "A computer - assisted system for the capture, storage, retrieval, analysis and display of spatial data, within a particular Organization" (Stillwell & Clarke, 1987).

- "A powerful set of tools for collecting, storing, retrieving at will, transforming and displaying spatial data from the real world" (Burrough, 1987).

- A GIS is also defined as follows (Aronoff, 1989):

    - A GIS is a computer-based system that provides the following four sets of capabilities to handle geo-referenced data:

    - *Input,*

    - *data management (data storage and retrieval),*

    - *manipulation and analysis, and*

    - *Output.*

- "An internally referenced, automated, spatial information system".

- "A system for capturing, storing, checking, manipulating, analyzing and displaying data which are spatially referenced to the Earth".

- "An information technology which stores, analyses and display both spatial and non–spatial data".

- "A database system in which most of the data are spatially indexed, and upon which a set of procedures operated in order to answer queries about spatial entities in the database".

- "An automated set of functions that provides professionals with advanced capabilities for the storage, retrieval, manipulation and display of geographically located data".

- "A decision support system involving the integration of spatially referenced data in a problem solving environment".

- "A system with advanced geo – modeling capabilities".

Although the above definitions cover a wide range of subjects, the activities best refer to geographical information. Some times it is also termed as Spatial Information Systems as it deals with located data, for objects positioned in any space, not just geographical, a term for world space. Similarly, the term 'a spatial data' is often used as a synonym for attribute data (i.e. rainfall/ temperature/ soil chemical parameters/ population data etc.).

Frequently used technical terms in spatial data handling are:

- Geographic (or geographical) Information System (GIS),

- Geo-information System,

- Spatial Information System (SIS),

- Land Information System (LIS), and

- Multi-purpose Cadastre.

Geographic information systems are used by various disciplines as tools for spatial data handling in a geoinformatics environment.

Depending on the interest of a particular application, a GIS can be considered to be a data store (application of a spatial database), a tool- (box), a technology, an information source or a science (spatial information science).

Like in any other discipline, the use of tools for problem solving is one thing; to produce these tools is something different. Not all are equally well suited for a particular application. Tools can be improved and perfected to better serve a particular need or application. The discipline that provides the background for the production of the tools in spatial data handling is spatial information theory (or SIT).

Geographic information technology is used to manipulate objects in geographic space, and to acquire knowledge from spatial facts. Spatial information theory provides a basis for GIS by bringing together fields that deal with spatial reasoning, the representation of space, and human understanding of space:

Spatial reasoning addresses the inference of spatial information from spatial facts. It deals with the framework and models for space and time, and the relationships that can be identified between objects in a spatio-temporal model of real world phenomena.

Scientific methods for the *representation of space* are important for the development of data models and data structures to represent objects in spatial databases. Spatial databases are distinguished from standard databases by their capability to store and manage data with an extent in space and time (spatial data types).

The *human understanding of space*, influenced by language and culture, plays an important role in how people design and use GIS.

The theory is used for the design of high-level models of spatial phenomena and processes. They are then mapped into conceptual, logical and physical models of spatial databases. The database stands central in the geoinformatics environment. It is the database that holds the data; without it, no useful function can be performed. Data are entered into the database in input processes. Later, they are extracted from the database for spatial analysis and display.

The processes of data management, analysis and display are often supported by rules that are derived from domain experts. Systems that apply stored rules to arrive at conclusions, are called rule-based or knowledge-based systems. Those that support decision making for space-related problems are known as spatial decision support systems (or SDSS). They are becoming increasingly popular in planning agencies and management of natural resources.

All these activities happen in a social, economic and legal context. It is generally referred to as spatial information infrastructure. Everything within this infrastructure and all concepts of space and time in turn are shaped and determined by the cultural background of the individuals and organizations involved.

HISTORY OF GIS

The GIS history dates back to 1960 when computer based GIS have been used and their manual procedures were in life 100 years earlier or so. The initial developments originated in North America with the organizations such as US Bureau of the Census, The US Geological Survey and The Harvard Laboratory for computer graphics and Environmental Systems Research Institute (commercial). Canadian Geographic Information Systems (CGIS) in Canada, Natural Experimental Research Center (NREC), Department of Environment (DOE) and other notable organizations in U.K. were involved in early developments. The laboratory for Computer Graphics and Spatial Analysis of the Harvard Graduate School of Design and the State University of New York at Buffalo achieved worldwide recognition. Commercial agencies started to develop and offer GIS software. Among them were today's market leaders ESRI, Intergraph, Laserscan, Autodesk etc.

A sound and stable data structure to store and analyze map data became dominant in the early 1970's. This has lead to the introduction of topology into GIS. Topology and the related graph theory proved to be effective and efficient tools to provide logically consistent two-dimensional data representations. Another significant break through occurred with the introduction and spread of personal computers in 1980's. It was possible to have a computer on the desk that was able to execute programs that previously could only be run on mainframe computers. At the same time minicomputers, and later, workstations became widely available. Relational database technology became the standard. Research on spatial data structures, indexing methods, and spatial databases made tremendous progress. The 1990's can be characterized as a period of the breakthrough of object-orientation in system and database design, recognition of geoinformatics as a professional activity, and spatial information theory as the theoretical basis for GIS. Potentiality of GIS is realized in the recent past and now it has become popular among many users for a variety of applications.

In India the major developments have happened during the last one-decade with significant contribution coming from Department of Space emphasizing the GIS applications for Natural Resources Management. Notable among them are Natural Resource Information System (NRIS), Integrated Mission for Sustainable Development (IMSD) and Bio-diversity Characterization at National Level. IIRS is also playing a major role in GIS

through education and training programs at the National and International level. Recently the commercial organizations in India have realized the importance of GIS for many applications like natural resource management, infrastructure development, facility management, business/market applications etc. and many GIS based projects according to the user organization requirements were developed.

GIS OBJECTIVES

- Maximize the efficiency of planning and decision making

- Provide efficient means for data distribution and handling

- Elimination of redundant data base - minimize duplication

- Capacity to integrate information from many sources

- Complex analysis/query involving geographical referenced data to generate new information.

For any application there are five generic questions a GIS can answer:

Location	-	What exists at a particular location?
Condition	-	Identify locations where certain conditions exist.
Trends	-	What has changed since?
Patterns	-	What spatial pattern exists?
Modeling	-	What if ………. ?

Elements of A GIS:

The GIS has been divided into four elements. They are hardware, software, data, and liveware. Table-2 gives complete details of different elements.

Table 2: Details of Elements of GIS

S. No.	Elements of GIS	Details
1.	Hardware	Type of Computer Platforms Modest Personnel Computers High performance workstations Minicomputers Mainframe computers Input Devices Scanners Digitizers Tape drivers CD Keyboard Graphic Monitor Output Devices Plotters Printers
2.	Software	Input Modules Editing MRP Manipulation/ Analysis Modules Modeling Capability
3.	Data	Attribute Data Spatial Data Remote Sensing Data Global Database
4.	Liveware	People responsible for digitizing, Implementing using GIS Trained personnel

DATA MODELS

Conversion of real world geographical variation into discrete objects is done through data models. It represents the linkage between the real world domain of geographic data and computer representation of these features. Data models discussed here are for representing the spatial information.

Data models are of two types: Raster and Vector. In raster type of representation of the geographical data, a set of cells located by coordinate is used; each cell is independently addressed with the value of an attribute. Each cell contains a single value and every location corresponds to a cell. One set of cell and associated value is a LAYER. Raster models are simple with which spatial analysis is easier and faster. Raster data models require a huge volume of data to be stored, fitness of data is limited by cell size and output is less

beautiful. Figure 1 shows vector and raster data representation of the real world phenomena.

Vector data model uses line segments or points represented by their explicit x, y coordinates to identify locations. Connecting set of line segments forms area objects. Vector data models require less storage space, outputs are appreciable, Estimation of area/perimeter is accurate and editing is faster and convenient. Spatial analysis is difficult with respect to writing the software program.

The vector model is extremely useful for describing discrete features, but less useful for describing continuously varying features such as soil type or accessibility costs for hospitals. The raster model has evolved to model such continuous features. A raster image comprises a collection of grid cells rather like a scanned map or picture. Both the vector and raster models for storing geographic data have unique advantages and disadvantages. Modern GIS packages are able to handle both models.
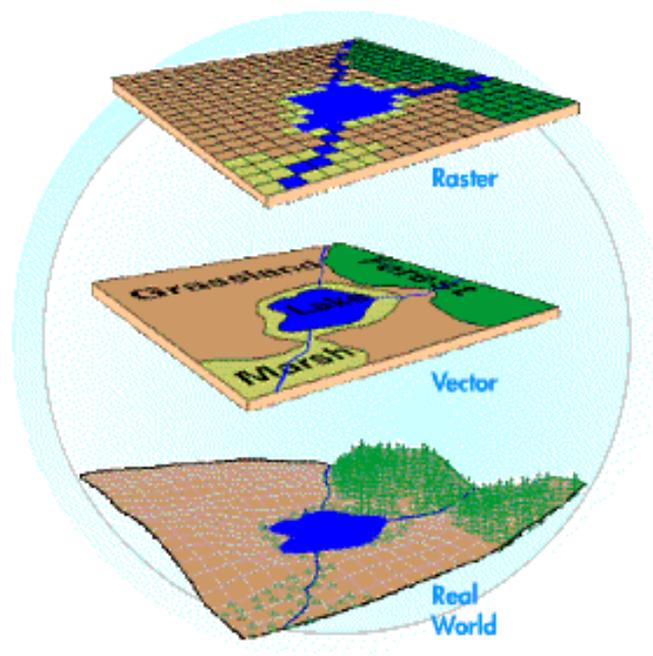

Figure 1: Vector and Raster data examples

Layers and Coverages

The common requirement to access data on the basis of one or more classes has resulted in several GIS employing organizational schemes in which all data of a particular level of classification, such as roads, rivers or vegetation types are grouped into so called layers or coverages. The concept of layers is to be found in both vector and raster models. The layers can be combined with each other in various ways to create new layers that are a function of the individual ones. The characteristic of each layer within a layer-based GIS is that all locations with each layer may be said to belong to a single Arial region or cell, whether it be a polygon bounded by lines in vector system, or a grid cell in a Raster system. But it is possible for each region to have multiple attributes. The Figure 2 shows layers and coverage concept in GIS.
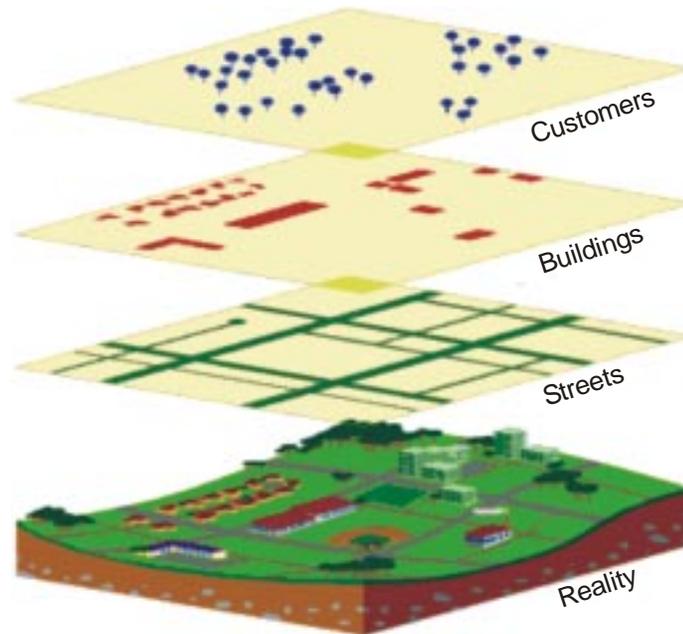

Figure 2: Layers and Coverage concept in GIS

Data Structures

There are number of different ways to organize the data inside the information system. The choice of data structure affects both Data storage volume and processing efficiency. Many GIS software's have specialized

capabilities for storing and manipulating attribute data in addition to spatial information. Three basic data structures are – Relational, Hierarchical and Network.

Relational data structure organizes the data in terms of two-dimensional tables where each table is a separate file (Table 3). Each row in the table is a record and each record has a set of attributes. Each column in the table is an attribute. Different tables are related through the use of a common identifier called KEY. Relation extracts the information, which are defined by query.

Table 3: Example of Relational Database

Settlement name	Settlement status	Settlement population	County name
Gittings	Village	243	Downshire
Bogton	Town	31520	Downshire
Puffings	Village	412	Binglia
Pondside	City	112510	Mereshire

Hierarchical data structure (Fig. 3) stores the data in a way that a hierarchy is maintained among the data items. Each node can be divided into one or more additional node. Stored data gets more and more detailed as one branches further out on the tree.
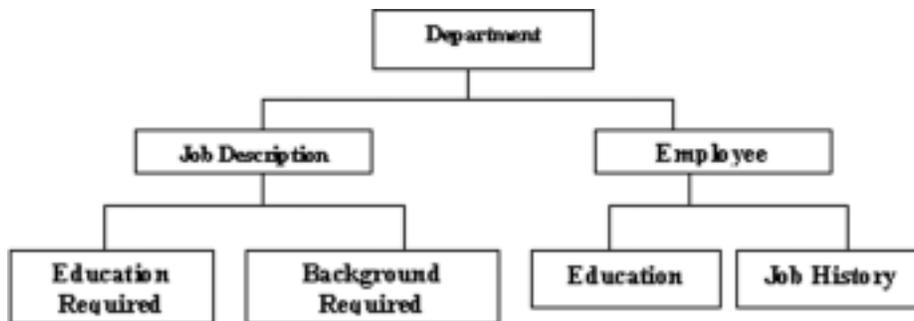

Figure 3: Hierarchical Data Structure

Network data structure (Fig. 4) is similar to hierarchy structure with the exception that in this structure a node may have more than one parent. Each node can be divided into one or more additional nodes. Nodes can have

many parents. The network data structure has the limitation that the pointers must be updated everytime a change is made to database causing considerable overhead.
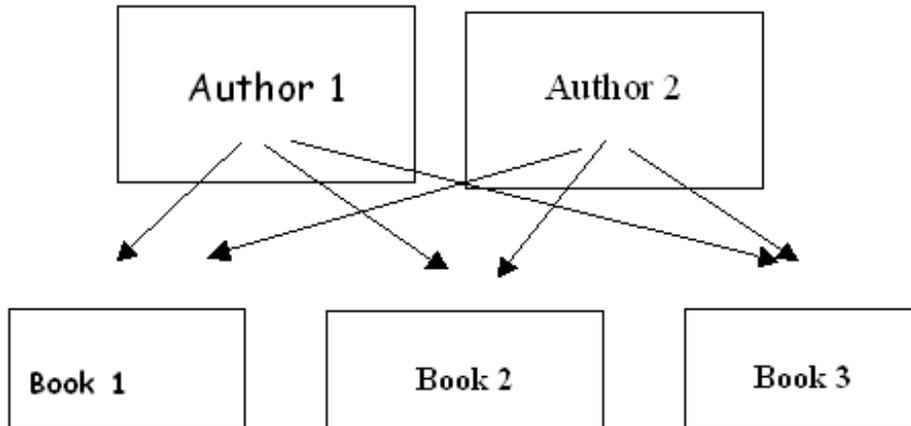

Figure 4: Network Structure

Errors in GIS

Uncertainties and errors are intrinsic to spatial data and need to be addressed properly, not sweeping away the users by high quality colour outputs. Data accuracy is often grouped according to thematic accuracy, positional accuracy and temporal accuracy occurring at various stages in spatial data handling. Given below are some of them while creating the spatial database and analysis.

(i) Errors in GIS environment can be classified into following major groups:

Age of data	-	Reliability decreases with age
Map scale	-	Non-availability of data on a proper scale or Use of data at different scales
Density of observation	-	Sparsely distributed data set is less reliable
Relevance of data	-	Use of surrogate data leads to errors
Data inaccuracy	-	Positional, elevation, minimum mapable unit etc.
Inaccuracy of contents	-	Attributes are erroneously attached

(ii) Errors associated with processing :

Map digitization errors	-	due to boundary location problems on maps and errors associated with digital representation of features
Rasteurization errors	-	due to topological mismatch arising during approximation by grid
Spatial Integration errors	-	due to map integration resulting in spurious polygons
Generalization errors	-	due to aggregation process when features are abstracted to lower scale

Attribute mismatch errors.

Misuse of Logic

The errors are also added from the source of data. Care must be taken in creating spatial databases from the accurate and reliable sources of data. Realizing the importance the users may demand in future to provide them the desired data with a tag showing how much of the data is accurate, spatial data must be presented in quantitative terms.

Spatial Analysis

Whether it is effective utilization of natural resources or sustainable development or natural disaster management, selecting the best site for waste disposal, optimum route alignment or local problems have a geographical component; geoinformatics will give you power to create maps, integrate information, visualize scenarios, solve complicated problems, present powerful ideas, and develop effective solutions like never before. In brief it can be described as a supporting tool for decision-making process. Map making and geographic analysis are not new, but a GIS performs these tasks better and faster than do the old manual methods. Today, GIS is a multibillion-dollar industry employing hundreds of thousands of people worldwide.

GIS is used to perform a variety of Spatial analysis, including overlaying combinations of features and recording resultant conditions, analyzing flows or other characteristics of networks; proximity analysis (i.e. buffet zoning) and

defining districts in terms of spatial criteria. GIS can interrogate geographic features and retrieve associated attribute information, called identification. It can generate new set of maps by query and analysis. It also evolves new information by spatial operations. Following are the analytical procedures applied with a GIS. GIS operational procedure and analytical tasks that are particularly useful for spatial analysis include:

    #   Single layer operations

    #   Multi layer operations/ Topological overlay

    #   Geometric modeling

        Calculating the distance between geographic features

        Calculating area, length and perimeter

        Geometric buffers.

    #   Network analysis

    #   Surface analysis

    #   Raster/Grid analysis

There are many applications of Geoinformatics, viz. facility management, planning, environmental monitoring, population census analysis, insurance assessment, and health service provision, hazard mapping and many other applications. The following list shows few applications in natural resource management:

- Agricultural development

- Land evaluation analysis

- Change detection of vegetated areas

- Analysis of deforestation and associated environmental hazards

- Monitoring vegetation health

- Mapping percentage vegetation cover for the management of land

- Crop acreage and production estimation

- Wasteland mapping

- Soil resources mapping

- Groundwater potential mapping

- Geological and mineral exploration

- Snow-melt run-off forecasting

- Monitoring forest fire

- Monitoring ocean productivity etc.

- GIS application in Forestry

With the rise of World Wide Web, new Internet protocols such as the Hypertext Transfer Protocol (HTTP), as well as easy to use interfaces (browsers), tools and languages (HTML, XML, and Java), the Internet has become a hub for GIS functionalities from the client side without even any GIS software. The GIS field is still evolving and it will be the major force in various walks of life dealing with geographic information.

CONCLUSIONS

Geographic Information System (GIS) is used by multi-disciplines as tools for spatial data handling in a geographic environment. Basic elements of GIS consist of hardware, software, data and liveware. GIS is considered one of the important tool for decision making in problem solving environment dealing with geo-information.

REFERENCES

Aronoff, S. 1989. Geographic Information Systems: A Management Perspective. Ottawa, Canada : WDC Publications.

Burrough, P.A. 1987. Principles of Geographical Information Systems for Land Resource Assessment. Oxford : Claredon Press.

Stillwell John and Clarke Graham (ed.) 1987. Applied GIS and Spatial Analysis. West Sussex : John Wiley and Sons, 2004.

http://www.ncgia.ucsb.edu/~spalladi/thesis/Chapter3.html

http://www.sli.unimelb.edu.au/gisweb/menu.html

http://www.gislinx.com/Software/Programs/MicroStation/index.shtml

http://geog.hkbu.edu.hk/geog3600/ (Hongkong Baptist University)

http://geosun.sjsu.edu/paula/137/ppt/lecture13/sld008.htm

http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/tutore.html

http://www.cla.sc.edu/gis/avshtcrs/handouts.html

http://www.gisdevelopment.net

http://www.planweb.co.uk/

http://www.sbg.ac.at/geo/idrisi/wwwtutor/tuthome.htm

http://www.ed.ac.uk/

http://www.geoplace.com