Design and Implementation of Double Cube Data Model for Geographical Information System

Mohd Shafry Mohd Rahim, Daut Daman, Harihodin Selamat
Faculty of Computer Science and Information System, University of Technology Malaysia, Malaysia

Abstract: *In general, Geographical Information System (GIS) has four main components, namely data input, data model, data analysis and manipulation, and data presentation. These components are vital in order to make GIS fully functional. This paper is centred upon the research activities of selecting, synthesis, designing and implementing the data model components. The primary research issue is to develop a data model that encompasses the capability of storing and managing changes in geographic features. A new perspective approach on the current data modelling is proposed in order to alleviate the current issue plaguing the GIS data management. In developing this new perspective, the feature based approach system cube method is synthesized to produce a new data model. Consequently, the combination of these approaches resulted in the design of double cube data model, which integrates temporal information of geographic features. The double cube data model has been implemented using relational database system. After extensive testing, the double cube data model performed admirably in managing the dynamic changes of geographic features. In conclusion, temporal information is the prime importance in managing geographic data. In this paper, the author has proved that temporal information can be integrated into a single GIS data model.*

Keywords: *Data model, database, temporal GIS.*

Received August 3, 2003; accepted December 5, 2003

1. Introduction

In developing the Geographical Information System (GIS), there are four aspects that need to be taken into consideration from the science computer aspect which are data input, data model, data manipulation and analysis and data presentation. Each of them has its function to ensure the GIS functions perfectly [12].

Time is an important element in the development of GIS. Most of the geographical disciplines use time as a basis of geographical history data management; medical geography, cultural geography and physical modelling [15]. By taking into consideration the element of time, analysis and data presentation process will be more effective and realistic. Other issue is the integration between geographical data that is spatial data and non-spatial data.

The mission of this research is to produce a data model that is concerned with the element of time. Executing the data model to see the effectiveness and the ability of data manipulation and analysis besides the geographical data presentation. While the objectives are to produce a data model design that can integrate the geographical data with time element, implementing the data model with a database system design and to make a test on the database system ability for data accessibility and represent the data using two dimensional map.

2. Time in Geographical Data

Geographical data is a data that is correlated directly with the state of the earth feature. The data that is involved is spatial data and non-spatial data. Spatial data is data that represents the space and location for geographical data [2, 9]. The non-spatial data is additional information that describes the spatial data from geographical features aspect, nature's phenomenon and from human activity. Time is an important element in geographical data [5, 20]. The element of time should be taken into consideration in analyzing the geographical data because in future, user needs are focused to historical data for geographical features on a map [6]. Figure 1 shows the importance of time in geographical data.

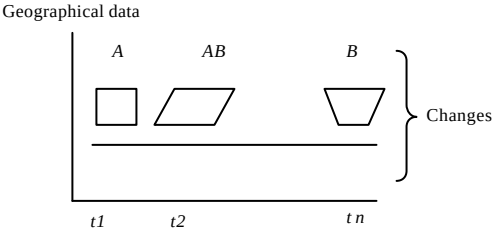

Figure 1. Relationship time and geographical data.

3. Geographical Data Model

Generally, there are four techniques that can be used; object based, field based, features based and direction based approach.

Object based approach is a technique that defines the earth surface and consistent of objects with its own features for example mountains and rivers [11, 12, 19]. Each object has its particular information such as name, capacious, shape and space. The bases of spatial data is a dot, a line or curve, a polygon or surface and solid object or three dimensional object. The operation that can be done in this technique is the relationship between the object that can be integrated to a set of notation, topology and matrix.

Fields based approach define the earth surface that consists of a vast carpet with its geographical features [11, 12]. The three components that must be identified are spatial work plan, spatial function and appropriate spatial operation set. The spatial work plan F is a limited space such as magnitude and latitude as shown in Figure 2.

Mathematical definition

fi: Work space Domain Attribute (*Ai*)

Objects in the space are the *x* attributes and are defined as

$$F(x) = x1+x2\ldots+xn$$

F is the work space, x is the attribute and geographical features. The operation that can be done is according to the function theory which is $f(x)$ and $g(x)$, $f(x)+g(x)$ and $f(x).g(x)$. Operation can be grouped into three, which are local operation, focus operation and zone operation [11]. For local operation, the value of space is dependant of the object in that workspace. While for focus operation, the operation is done in a focused area, zone operation takes place at zone where differentiation process towards spatial work plans function

$$Z = ?f(x)\,dx$$

in spatial work plan space $f(x)$.

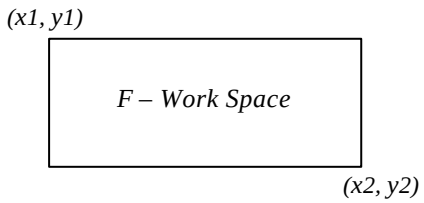

Figure 2. Work space plan.

In this approach, geographic data was defined continuously, which is all of geographic data has semantic relationship within the data. That mean, the changes of the spatial data will be effected to the non-spatial data and other way around. All of the analysis must be done by encompass all component of geographic data. For that purposed function theory as a basic for the geographic analysis.

Features based approach are defined as a real world geographical entity and presentation in digital and graphical [3, 18]. Geographical entity is referred as geographical features on a surface of the earth. It has specific information to understand a map. The graphical presentation is a technique to present the geographical features. For example, symbols and geographical information about a map are also presented on the same map.

Direction based approach was pioneered by [13]. It is according to the geographical space that is closed to direction. It is used to process a request that involved direction. Three important elements are vector, dot and degree that are close to each other.

There are a few models that have been gathered and researched such as TLDM data model [18], event based data model [20], GISER data model [12], SAND data model [10], IGMX data model [7], object oriented data model [19], and object-relational data model [17].

Three level data method is a data model that was pioneered by Weng Feng and friends [3]. This model solves the presentation issue of geographical features in geographical form. It is implemented in the transportation information system to represent the movement of spatial data that deputize the point for vehicle and manages the changes for each movement of the vehicle. The content of even based data model consists of changes entity, human, space information and its features and changes factor [20].

Geographical information systems entity relational was developed according to the gathering of space (field based) and object oriented design method [12]. This model is built according to 4 main features which are space and time, features, coverage and object space. This model takes into consideration spatial and non spatial data.

Object-Oriented technique consists of spatial data and non-spatial data [19]. This model was developed to overcome the problem of integrating the spatial data and non-spatial data that is complex. This model is divided into graphic object, thematic object and geometry object. The issue concerning the object oriented technique is the relationship with the object. It can be solved by using the object-relational technique [16, 17, 19].

Malaysian Hydrology Information System (MHIS) data model was developed by using the cube system approach [1, 14]. Three important coordinate in this model are Features Identity (FID), Description Identity (DID) and Time and Data (TID). These three coordinate are used to determine the data that is kept. FID is a value of data that has been measured or the key value of spatial data. DID is a key that describes the FID while the TID is a key where the value of time is taken.

As a conclusion, many researchers [5, 10, 11, 12, 17, 19] focused on the object based to design the data model because this approach is easier to understand. It does have some constraints in making complex analysis which involves the integration of the geographic feature. This is because the object was

separated by class of object. Field based analysis, is a better approach to design a data model but the approach is more complex and has a great relation ship within the geographic data. Theoretically, more analysis can be done using the discrete and continuous analysis [11, 12]. It needs to be simplified based on the GIS application. Features based approach have a similar way of the object based approach but classify geographic data based on the class of features. Based on [19], data can be integrated between the spatial and non-spatial data. A few issues have been discussed in research at the early 2000, to design a more dynamic GIS that can better represent the actual geographic event [16].This consist of improving the data model GIS [8, 16, 17], manipulation and analysis, increase speed of the process and visualize geographic data in the four dimension [4, 8]. To solve this issue GIS need to incorporate with the computer graphic approach such as animation and simulation.

4. Research Methodology

Researchers have discussed the various methods that are often used in the database system of a GIS. In this research, we will use a method that combines the feature base approach with the approach used by the MHIS. We added the time entity to the feature base approach to solve the research issues and achieve the objective.

4.1. Cube System

Cube system is an approach to produce data model. This system eases the data management that is related to time. It has been used in data model for MHIS for geographical attribute management that is always changing. Cube system is used directly to manage the changes of spatial data and non-spatial data. The coordinate structure and concept is still the same with the old system. The change is, the keyword for non-spatial data is changed to spatial data. Thus, the new coordinates are FID, TID and GID where the GID is referred to spatial data. As a conclusion, double cube is used for the purpose of managing the geographical features which is spatial and non-spatial data.

4.2. Model Integration Features Based Approach and Cube System

The connecting point between these two techniques is the element of time. Thus, the integration research is to produce a data model. Modelling technique based on geographical features is mapped to the cube system to produce a data model that will fulfill the requirement of the technique. The integration process is done based on four elements in modelling technique based on geographical features. The elements are divided into two, which are spatial and non-spatial data with time element. The integration is done using the same keyword in both cubes, which is FID.

This model is used to interpret the cube into table form. The cube has three coordinate (*w*, *y*, *z*) for each cell that is also the keyword for data reference. Cube will become a table that has three key that control the changes and additional data.

5. Data Model Design

5.1. Double Cube Data Model Design

This model is based on a combination of modelling technique according to the graphical features with new entity, which is time and cube technique. As a result, one data model has been produced and namely double cube data model, as shown in Figure 3.

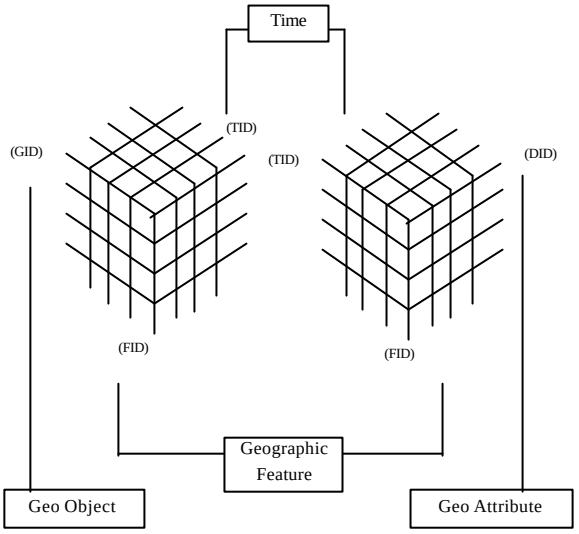

Figure 3. Double cube data model Concept.

Four important keywords as a relationship between data:

- FID (Feature Identification): Main keyword, furthermore a reference for each geographical feature on the earth surface. The keyword that can be found at both cubes ensures the integration between spatial and non-spatial data happen. FID has its own table* that describes the information about geographical features in detail.
- TID (Time Identification): Keyword for reference from time to time when the data is recorded. TID can be found in both cubes that are used to record the time for each data. The function is more to manage the changes of data.
- DID (Description Identification): This keyword only can be found in non-spatial data cube and is referred to the description data taken.
- GID (Geometry Identification): This keyword can be found in spatial data cube that is referred to the geometry object for geographical features. It is completed with table that stores a set of dots that creates the spatial object.

The component inside the data model consists of five important components which are space, attribute data, changed non-spatial data, changed spatial data and presentation element. Space is a component that contains a set of data that create feature and location for geographical features information. It consists of a set of vector data and the data are in a form of dots set (x, y, z). Attribute data is a component that describes the geographical features on the earth surface. The changed non-spatial data is a component that records the changes that happened towards the attribute data. There are two types of changes, which are the changes based on nature's phenomenon and the changes that happened based on human activity. The changed spatial data is a component that depends on time changes. It will record the changes that involved the location feature. The presentation element is a component that allows the data to be presented in digital form together with certain colours.

5.3. Database Design

Relational database is used to implement double cube data model. Both cubes are connected to table form to design the database. The first cube is for spatial data table while the second cube is to manage the non-spatial data. In spatial cube, FID, TID and GID become a reference while for non-spatial cube, these three keywords are a reference for the data.

5.4. Data Model Discussion

The purpose of using double cube in designing the data model is to reduce the high memory use in database system and to make sure that the conflict will not happen in storing the spatial and non-spatial data. This happens because of the changes between the parallel spatial data and non-spatial data. If the cube is combined as one, the column for the keyword must be added with different value between both cubes, which is the GID value. As a result, the database system memory will become higher from time to time. Double cube data model is dynamic where the increases can be done depends on the use necessity. This model can is also be used in all GIS application especially the Temporal GIS application.

6. Results of the Implementation

Double cube data model is implemented using the relational database model. There are ten entities that have been determined based on double cube data model that will be used in the database. The design of the relational database concept has been produced physically. Data has been defined as a field name, size and type of data, while the table will be defined as a relationship between tables. Figure 4 shows the relationship between the tables. Both cubes in data model is interpreted using each table that has three main keyword to ensure the data can be integrated and retrieved to presentation and testing towards time entity.

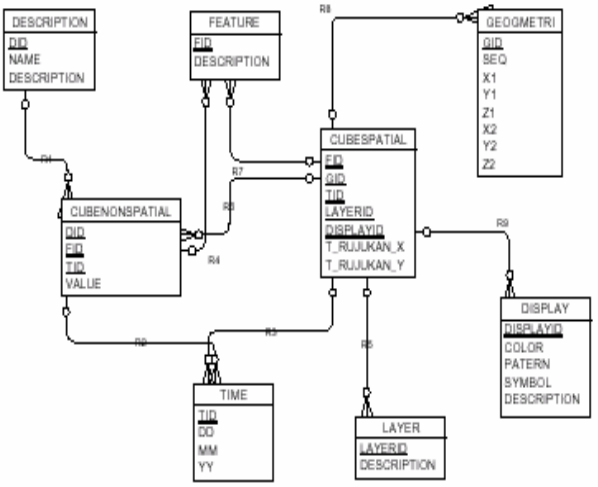

Figure 4. Database design based on the double cube data model concept.

6.1. Data Visualization and Attributes Searching

The data has visualized in the thematic map by layer from a single database. One set of query has been developed to retrieve the data from database. Besides, some simple query has been developing to search attribute. Figure 5 shows the result of the searching geographic attribute.

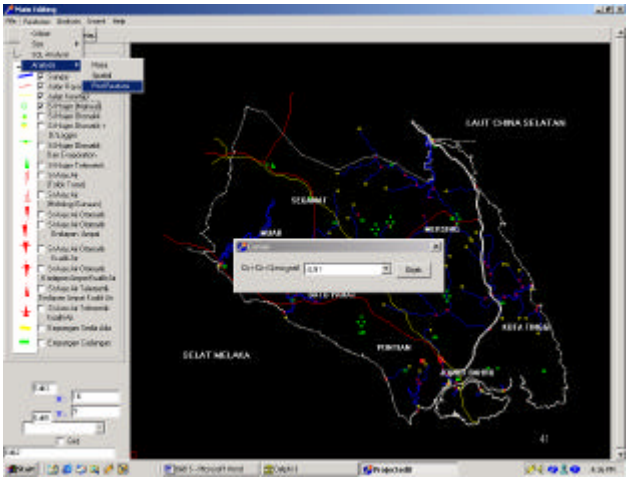

Figure 5. Geographical features searching process output.

6.2. Data Integration

Spatial data is presented in 2 dimensional map while for non-spatial data is presented as an additional information for spatial data. Both data are shown on the same interface. This proves that the integration of data happened at the database system.

The integration of the data is happening at the databases layer. As a result the query was programmed in our system to make sure the data will be displayed as shown in Figure 6. However in the process query

optimization is very important to make sure that data will be displayed at a very practical rate.

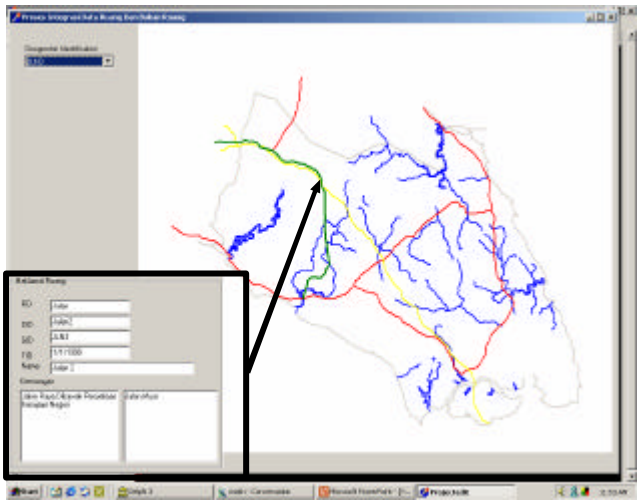

Figure 6. Output of the integration process between spatial data and non-spatial data.

6.3. Time Quantum Analysis

The testing of time analysis process for spatial data is done to show the changes of a spatial data according to its date and year. The result is shown in Figure 7.

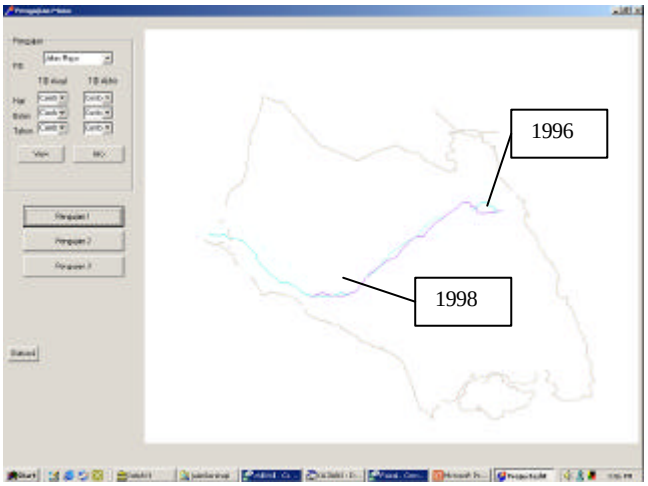

Figure 7. The changes of spatial data for muar-mersing road.

From the result, data model is able to supply the data for the purpose of time analysis. However this is only a simple temporal analysis which has been implemented and tested in this system. More analysis is required to make a more functional GIS for the user perspective. For more references please refer to [4, 6, 8, 16].

7. Conclusion

Double cube data model that is a combination of geographical features based on data modelling and the cube system. It has its own advantages that have been discussed in this article. The data model can integrate spatial and non spatial data in a single database and support the temporal analysis. The author has suggested a new framework for developing a GIS database system. More research is needed to be done in improving the double cube data model to fully support the temporal GIS analysis. To fully support the temporal GIS analysis with the double cube data model, the database base design need to be revise for optimisation query and a more practical way of displaying the information for a ore interactive and user friendly system. The solution in the future, GIS need to incorporate with the computer graphic discipline for increasing interactivity of the system and realistic process [4]. This is a challenging work in making the GIS which can become a very powerful tool in an environmental issue.

Acknowledgment

This research was sponsored by Ministry of Science and managed by Research Management Centre, Universiti Teknologi Malaysia through IRPA Funding RMK 7 Vot No: 72192 and RMK 8 Vot No: 74074. Special thanks to Faculty of Computer Science & Information System, and Department of Irrigation Malaysia on the cooperation in this research program.

References

[1] Bellamy S. P., "MHIS Development, Database/ArcInfo Link-Height Level Design," *Technical Report*, Issue V1.R1.M0 (24/7), Ministry of Agriculture, Malaysia, 1996.

[2] Environmental System Research Institute, *Understanding GIS: The ARC/INFO Method*, 1995.

[3] Feng W., Shuqiang Y., Huowang C., and Jichang S., "Spatial Data Model for Feature-Based GIS," *in Proceedings of Workshop on Computer Science and Information Technologies (CSIT'99)*, Russia, pp. 107-111, 1999.

[4] Hogeweg M., "Spatio-Temporal Visualization and Analysis," *Master Thesis*, University of Salford, 2000.

[5] Langran G. "A Review of Temporal Database Research and its Use in GIS Applications," *International Journal of Geographical Information Systems*, vol. 3, pp. 215-232, 1989.

[6] Langran G., *Time in Geographic Information Systems*: *Technical Issues in GIS*, Taylor and Francis, London, UK, 1992.

[7] Murillo A., "A GIS Data Model Prototype," *International Journal of Geographical Information Systems*, vol. 6, no. 3, pp. 332-343, 1995.

[8] Narsico F. E., "A Spatiotemporal Data Model for Incorporating Time in Geographic Information System," *PhD Thesis*, College of Engineering, University of South Florida, 1999.

[9] Rahim S., "Geographical Information System: Databases for Selection of New Catchments Area in Johor, Malaysai," *Degree Thesis*, University of Technology Malaysia, 1999.

[10] Samet H. and Aref W. G., *Spatial Data Models and Query Processing in Modern Database Systems, the Object Model, Interoperability, and beyond*, in Kim W. (Ed), Addison-Wesley, ACM Press, pp. 338-360, 1995.

[11] Shekar S., Chawla S., Ravada S., Fetterer A., Liu X., and Liu C. T., "Spatial Databases: Accomplishments and Research Needs," *IEEE Transactions on Knowledge and Data Engineering*, 1999.

[12] Shekar S., Coyle M., Goyal B., Liu D. R., and Sarkar S. S., "Data Models in Geographic Information Systems," *Communications of the ACM*, vol. 40, no. 4, 1997.

[13] Shekar S., Liu X., and Chawla S. "An Object Model of Direction and its Implications," *GeoInformatica*, vol. 3, no. 4, pp. 357-379, 1999.

[14] SMHB ITP, "Establishment of Hydrological Information System for Water Resources Planning, Development and Management," *Technical report*, Ministry of Agriculture, Malaysia, 1999.

[15] The Standards Committee of the Association of Geographic Information, *AGI GIS Dictionary*, University of Edinburgh, 1991.

[16] Thomas O., *Time Integrative Geographic Information Systems, Management and Analysis of Spatio-Temporal Data*, Springer-Berlin, Heidelberg, Germany, 2001.

[17] Twumasi B. O., "Modelling Spatial Object Behaviours in Object-Relational Geodatabase," *Master Thesis*, International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands, 2002.

[18] Usery E. L., "A Feature-Based Geographic Information System Model," *Journal of Photogram Metric Engineering and Remote Sensing*, vol. 62, no. 7, pp. 833-838, 1996.

[19] Wachowicz M., "Object-Oriented Design for Temporal GIS," *Research Monograph*, Taylor and Francis, London, UK, 1999.

[20] Wang D., "An Event Based Data Model for Transportation Information Systems," *Master Thesis*, Florida Atlantic University, 1998

Mohd Shafry Mohd Rahim received his Diploma in computer science in 1997, BSc in computer science in 1999 from University of Technology Malaysia and MSc in GIS and visualization in 2002 from University of Technology Malaysia. He is currently a lecturer in Department of Computer Graphic and Multimedia, Faculty of Computer Science and Information System at University of Technology Malaysia. He has over 5 years of experience in the field of computer graphics & visualization and GIS data model. He has also been actively involved in many research projects related to GIS and visualization, GIS data management and has published more than 20 publications.

Daut Daman received his BSc form Universiti Sains Malaysia and MSc from University of Cranfield, United Kingdom. He is currently an associate professor in the Faculty of Computer Science and Information System at University of Technology Malaysia. He has over 24 years of experience in the field of computer graphics and visualization. He taught subjects in the area of computer graphics, scientific visualization and software technologies. He has also been actively involved in many research projects related to computer graphics and visualization and has published more than 50 publications.

Harihodin Selamat received his MSc from University of Cranfield, and his PhD from University of Bradford, United Kingdom. He is currently an associate professor in the Faculty of Computer Science and Information System at University of Technology Malaysia. He has over 25 years of experience in the field of information systems and has also taught in the areas of system analysis and design, and database systems. He has also been actively involved in many consultancy and contract research projects as project leader with various public and private sectors in the areas of information strategic planning, information system development, and enterprise resource planning. He has also published more than 70 publications. His research interest includes database design and modeling, and enterprise resource planning. He is heading a group of more than 20 researchers working on various consultancy and research projects.