Spatio-Temporal Object Relational for Biodiversity System (STORe-Biodi)

Mohd Taib Wahid[1], A.Z.M Kamruzzaman[2], Abdul Rashid Mohamed Shariff[3]

1

[1,2]Department of Information Systems
Faculty of Computer Science & Information Systems
University Technology of Malaysia, K.B. 791
81310 Skudai, Johor, Malaysia

[3]Department of Biological & Agricultural Engineering,
Faculty of Engineering, University Putra Malaysia

Summary
A Spatio-temporal data model for biodiversity is growing importance to the biodiversity data management, forest and environment control. Spatio-temporal data models have received much attention in the database research community because of their practical importance and interesting challenges they pose. This paper is discussed upon the research activities of selecting, designing, implementing the data model. The paper can be divided by two major parts, first: discuss about biodiversity data model and secondly: spatio-temporal conceptual framework design of biodiversity data model for long terms stewardship of biodiversity information. In this paper our main objective to minimize the extension required in SQL language. This paper also focuses on the unified models of space and time using object-relation approach. In particular, we propose a conceptual object-relational spatial temporal data model based on Donna J. Peuquet's pyramid framework. Standard and user management queries are applied to test the model. After extensive testing, the data model performed admirably in managing biodiversity data.

Keywords: Spatio-temporal, Data model, Biodiversity data, Object-relational, SQL.

1. Introduction

After two decades of research, representation of space and time in databases and functional applications are still problematic [1]. This paper presents a universal object-relational framework for spatial temporal data modeling. Spatial temporal data modeling aims to extend the existing data models to include space and time in order to better describe our dynamic real world.

Problem with previous data models is that spatial and temporal aspects of databases are modeled separately [2], [3]. Spatial database focuses on supporting geometries [4], while temporal databases focus on the past state [5]. But in many circumstances, such as environmental monitoring, resource management, transportation scheduling, etc, spatial and temporal attributes should be connected together. Many current systems can handle only one aspect of space and time. Spatial systems always fail to cater for many temporal aspects in a dynamic environment [6]. Though many researchers have found the necessity of integration of space and time in one environment, by far, little such work has been done.

Another problem is representation of data should be natural to human. The structures of space and time are identified as essential for the realization of cognitive systems [10]. According to Donna J. Peuquet and her group [11], models of spatial temporal data in geographical database representations must incorporate human cognitive principles. Human knowledge of the dynamic geographical world comprises of three different (and interrelated) subsystems that handle *what*, *where* and *when* aspects of object properties [8]. Object relational approach with its characteristic of inheritance and aggregation is capable of capturing the various notions of space and time and reflecting them into a single framework extensible to different applications.

Many database systems concentrate on the definition of a particular spatial temporal model that is related to certain application. The result is that more and more different models appeared. Each model focuses on a specific set of spatial temporal features [7]. When encountering other features and applications, the model doesn't work. So what we should do is to build an overall framework that can be extended to various applications, not driven by various applications generating different models [8]. Furthermore, application specific modeling will be more efficient if it is based on a generic model [9].

To design a biodiversity data model to support spatio-temporal data, we have chosen pyramid frame work to support a biodiversity data model design. The challenge consists in the design of a spatio-temporal biodiversity data model, in parallel with the development of appropriate databases, the development of efficient spatio-temporal biodiversity query languages

2. Data Modeling Approaches

2.1 Pyramid framework and event based approach

In the design of object-relational biodiversity data model to support spatio-temporal data, pyramid frame can make enhancement to support a biodiversity data model design. A conceptual framework (also known as pyramid framework) was designed which guide the implementation of the semantic geographical information system (GIS) data model [21]. Conceptual frame work (pyramid) is interrelated with two separate parts one is data component and another is knowledge component (Figure 1). space' concept commonly cited in analysis of remote sensing imagery.

The taxonomy structures groups similar objects within a category and stores a rule-base that describes how those objects may be identified within the data space. These rules may be derived from expert knowledge or from inductive analysis of the observation data. In an object-relational data model consists of a set of *object classes* (of different *types* or *schemas*). Each object class has an associated set of *objects*; each object has a number of *attributes* with values drawn from certain *domains* or *atomic data types*. Of course, there are additional features, such as object valued attributes (Oid), methods, object class hierarchies, etc. Besides objects, attributes describing *geometries* including *time* are of particular interest. Hence we would like to define collections of *abstract data types*, or in fact *many-sorted algebras* containing several related types and their operations, for spatial values changing over time. This section presents a simple and expressive system of abstract data types, comprising data types and encapsulating operations, which may be integrated into a query language, to yield a powerful language for querying spatio-temporal data.

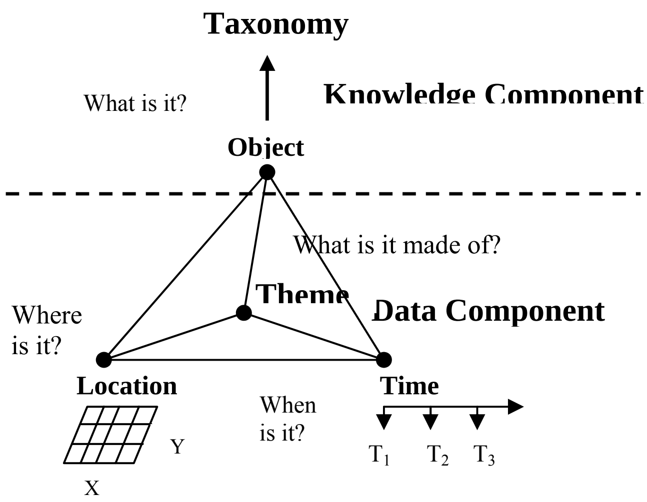


Figure 1: Pyramid Frame work: Data Component and Knowledge Component

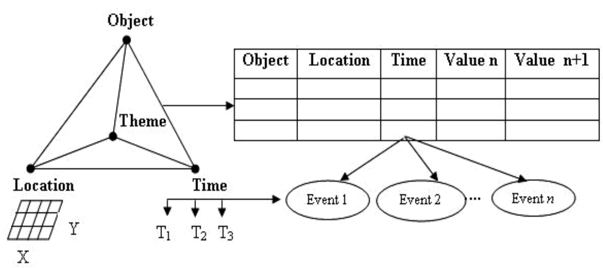


Figure 2: From the frame work data transform to the table and time modified as events

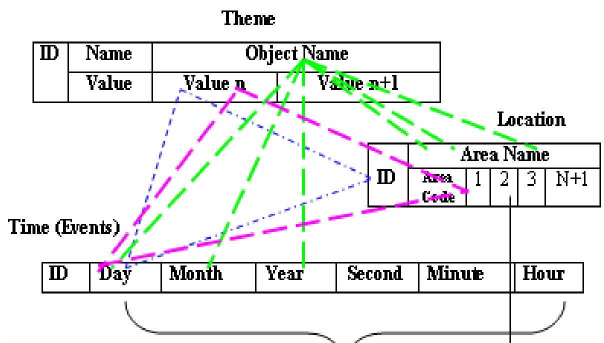
To support temporal data, we have applied another model such as event-based model. Pyramid system is a model to support multidimensional, spatio-temporal and geo-graphic objects (such as location and space). All the objects are also placed within two hierarchical relationship structures central to cognitive knowledge representational and object-oriented modeling, plant taxonomy (generalization). The plant taxonomy structures groups of similar objects within a category and stores a rule-base that describes how those objects maybe identified within the data space. Pyramid framework consists of main three components; objects, location and time. These features will be referred to accessed data from the database. This framework can be converted to table in stated Figure 2

After the framework conversion as a table, we can see that each object has attributes of theme, space (location) and time. Theme attributes record the information of "what is this object" and other related property information. For example, if the theme of this object is "Forest", a Forest must have attributes such as Area or Province it belongs to, Forest Size, and Forest Type, etc. Theme can be inherited according to certain classification hierarchy. For example, theme can be classified as natural resources (forest, earth quake, and rainfall), transportation (rail and highway), settlement (country, city), water system (river, lake, and pool), and cadastre etc.

Location attributes record space information of the object, such as whether it is a point, line, or polygon, and its location. Location attributes can be inherited into subclasses such as the point class, line class and polygon class.

Time attributes are represented through timestamps associated with theme and location information. When properties or forest information of an object changed, a new timestamp will be recorded for that object. That means only initial information and changed information will be recorded. Each timestamp is an instant. Interval is the time between two timestamps.

In this model, theme, spatial, time attributes are recorded as 3 tables. Each attributes use one list. For example, a flora object has theme attributes as *"Name"*, *"FloraID"*, *"FloraType"* and spatial attributes as *"Area"*. During different season and time of a year, the volume of the lake will change, and the border of the lake expands or contracts. Whenever the volume or the border of the lake changes, a timestamp will be recorded. Thus, as shown in Figure 3, each object has a name list, a value (*1 to N+1*), a location has area code (*1, 2, 3 and n+1*) and a timestamp or events (*Day, Month, Year, Second, minute, hour)*. Data can be collected same day same time from different area by different data collectors. Sometimes data collect from different forests for several days, months and years. Time stores as an events records as EventID, StartTime, Duration and Attributes (types of data). Diagram bellow indicates fields that can be used to associate objects between tables.



ID	ObjectID	Location	DateTime	Area Code
1	F001	Johor	02/01/1990 10:30 AM	01
2	F002	Skudai	10/02/1991 01:00 PM	02
3	F003	Skudai	10/02/1990 01:05 PM	02
4	F004	Skudai	15/01/1990 01:10 PM	02
5	F005	Johor	02/01/1990 10:30 AM	01

Figure 3: Relationship between theme, space and time

There are 3 kinds of relationships among theme, space and time: one-one, one-many, and many-many. In a one-one relationship, theme, space and time all changed, such as Value 1- *Area 1*. In one to many relationships, one theme attribute may be related to more spatial attributes and many timestamps. That means that theme attribute doesn't change during this interval while the spatial attributes changed, such as the attribute "*name*". Also, in some circumstances, spatial attributes of an object are static while theme attributes change. Another special case is that the timestamp list may have just one element. That means this object is static.

3. Structure and Conceptual Spatio-temporal Framework for Biodiversity data

Some research and development of spatio-temporal object-relational model is experimentally conducted in universal and several commercial software companies. A few experimental projects have been conducted based on TGIS by ESRI, IBM and several universities all over the world. The TGIS has been developed based on a relational data model that tries to overcome the problems of temporal representation being based on an area event system. This describes event area, what types of event occurred, when and where the event occurred. Event stores all kind of information in one table (Montgomery, 1995). Temporal attributes in this table are planned event start date, event end date and date of event is superseded by another event. In term of temporal expressions, there are two main methods of implementing time in database. The first is to leave time out of the data model and include it as an attribute of the data entities. The second is to include time as an integral part of the data model. With this implementation of time, the semantics of the data model are extended to incorporate time directly in spatio-temporal architecture. In principle, each spatio-temporal model is based upon three kinds of data, namely spatial, temporal and attributes data.

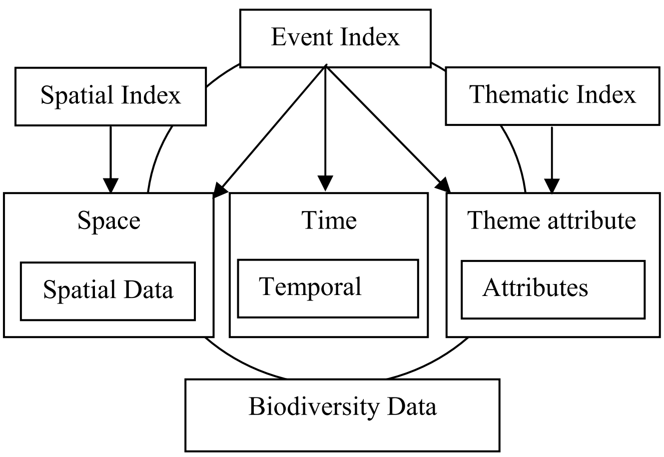


Figure 4: Spatio-temporal structure and biodiversity data model framework

Proposed Data model will be allowed support biodiversity data alone with temporal data. Spatio-temporal structure and biodiversity data model framework (Figure 4) is a combination of Spatial data, Attribute data and Temporal data. The three data together represents the values of attributes on an object occupying a certain position (spatial data) in the world at a certain time (temporal data). When dealing with temporal data, users look for casual relationships between objects. Events are used for an object that has both spatial and temporal locations and extends.

Three data together represents the values of the attributes of an object occupying a certain position (spatial data) in the world at a certain time (temporal data). When dealing with temporal data, users look for casual relationships between objects. Events are used for an object that has both spatial and temporal locations and extends.

The proposed framework separate themes, time, and space. The structure's links describes events explicitly and can be applied to spatio-temporal object models. Our proposed model might be new approach to spatio-temporal biodiversity information systems, called Spatio-temporal Object relational Plant Biodiversity (Stop-BDM) data model that is an increasing controversy in forest management. Meanwhile, there are few explanations concerning a temporal object models associated with temporal objects for bitemporal database. The proposes data models is describes by means of collection of one or more tables that form into a object model and connected with its relational tables to form an object-relational data model.

4. Implements the model

To implement and testing our proposed model, we have used an object-relational database management system. As a part of the testing, implementation and evaluation, a biodiversity data set covering more than 5 years data from different private and government organizations such mardi plant genebank data, pahang forest data and johor national park's data loaded onto the database. The physical model is developed by defining spatial and temporal versions of data types. For spatial and temporal data types as combinations of spatio-temporal and standard data types, queries can be differentiated by which dimensions the desired results are restricted to, and whether they query for a range in a dimension or a point. Therefore experiments on different spatio-temporal queries are performed with several different selectivity.

4.1 Querying with Spatio-temporal Operations

Our intention is not to devise a new spatio-temporal query language from scratch but to appropriately extend the widespread database query language SQL. The main focus is on integrating developments. We profit from the fact that the underlying data model rests on the Abstract Data Type (ADT) approach which necessitates only conservative extensions to SQL which are: (i) a set of basic spatiotemporal predicates, and (ii) an extension mechanism for new, more complex spatio-temporal predicates. We call this extended query language *STQL* (*Spatio-Temporal QueryLanguage*). All added functionality is captured by ADT objects and operations. The benefit of this approach is the preservation of well known SQL concepts, the high-level treatment of spatio-temporal objects, and the easy incorporation of spatio-temporal predicates.

The query facility of SQL is provided by the well known SELCT FROM WHERE clause. The integration of predicates like “<” or “<>”” for standard data types such as integers or strings is well understood. In particular, there are only a few of them which allow one to include them as built-in predicates. When considering more complex and more structured data such as points, lines, or regions, one can try to systematically derive all reasonable predicates. Temporally enhanced object-relational system can express temporal queries as powerful as those of TSQL2 with only minimal extensions to standard SQL by using construct such as user-defined functions and table expressions supported in Object-Relational systems.

We will consider queries from two (simplified) application scenarios. The first scenario is related to biodiversity data (such as flora, fauna) management which pursues the important goal of learning from past biodiversity data and their evolution. We assume a database containing relations with schemas

Flora (FloraName: VARCHAR2, Territory: Species name)
Flora TypeType (ForestArea: VARCHAR2, Extent: species)

Area (ForestArea: VARCHAR2, Location: Point)
Collectors (CollectorName: VARCHAR2, Actor: Person)

The relation flora records the location and the development of different species (attribute Territory) growing and shrinking over time through clearing, cultivation, and destruction processes, for example. The relation of flora collects from different places and it growth of different data by different people to their extinction (attribute Extent). The relation collectors describes about each flora being on duty from their start at the collection data up to their return (attribute Location). Figure 5 shows the results of biodiversity data (flora) has been collected different times (i. e., spatio-temporal data) by different persons.

Page 1 of 4

NAME	LOCALNAME	ENT_DATE	AREA_NAME
Andy	lidah kuching	2/1/2006	sarawak
Edie	ubi nyarum	3/1/2006	sarawak
Zeti	akar lapis	2/1/2006	KL
Azrul	akar lapis	2/1/2006	perak
David	tapak kuda	4/1/2006	KL
Elvis	kembang gajah	2/1/2006	perak
Henry	belong tanah	1/1/2006	perak
Azizul	belong tanah	2/1/2006	KL
Rahman	plot empty	1/1/2006	sarawak
Andy Lee	lidah kuching	4/1/2006	KL
Edie Lang	payung ali	3/1/2006	perak
Fatimah Abu	payung ali	3/1/2006	pahang

1 2 3 4 [Next >>]

Retrieval Time : 0.015625 's

Figure 5: Query to find out all flora name, colleted time and collector names.

The second scenario, finally, relates to a database about the migration of birds in order to explore their behavior patterns over the years.
Birds (Swarm: VARCHAR2, Movement:Point)

4.1 Temporal Selections

The first queries refer to the simple temporal range query. A *temporal selection* extracts the value of an object at a certain instant or the temporal development over a certain period. We can then ask queries like “what happened to the data between days January 10, 1995 to January 20, 1995?”

SELECT Flora_ID FROM flora
WHERE (((Flora_collection_year) Between # 01/10/1990# AND #01/20/1995#));

This query shows the functional character of a spatio-temporal object by determining the value of the object at a certain time through a simple function application. A more general version of this query asks what happened to the data between January 1, 2000 and Dec 20, 2006.

SELECT flora.localname, flora.updt_date, COUNT(*) FROM (flora INNER join observer ON flora.observer_id = observer.observer_id)INNER JOIN observation ON observation.observer_id = observer.observer_id WHERE ((flora.ent_date) BETWEEN '1-Jan-2000' AND '30-Dec-2006') GROUP BY flora.localname, flora.updt_date

LOCALNAME	UPDT_DATE	COUNT(*)
akar lapis	01-FEB-06	13
payung ali	01-MAR-06	3
payung ali	01-APR-06	1
tongkat ali	30-DEC-05	11
belong tanah	01-FEB-06	6
kuku bengkau	01-FEB-06	11
kuku bengkau	01-APR-06	2
kacip fatimah	01-FEB-06	12
kembang gajah	01-FEB-06	35
lidah kuching	01-FEB-06	27
lidah kuching	01-APR-06	1
mata pelanduk	01-MAR-06	2

Figure 6: query measure the data collected between two specific dates

The "Between" notation specifies a range of time values, that is, a time interval. If a spatio-temporal object is applied to a time interval (or a collection of disjoint time intervals separated by AND), this expression yields a spatio-temporal object restricted to that time interval (function restriction).

4.2 Projections to Space and Time

Projection operations on moving objects map either to their spatial or to their temporal aspect. Assume that we are interested in the geometric locations where the data was changed at the year January 20, 1990. These can be obtained by:

SELECT Flora_ID, AreaName
FROM Flora, Area
WHERE (((ForestArea =#Johor# AND #01/20/1990#));

This operation computes the *spatial projection* of a spatio-temporal object for the Johor forest area. For an evolving region the trajectory operation returns an object of the spatial type *region* which results from projecting the union of the region values for the Date 20th January, 1990. The next query asks for the lifespan of a spatio-temporal object: "How many times data have been collected from Johor before 1995?"

SELECT COUNT (Flora_ID) AS Ex1, COLLECTION_DATE AS Ex2
FROM Flora
GROUP BY COLLECTION_DATE
HAVING (COLLECTION_DATE <CONVERT (DATETIME, '1990-30-12'))
WHERE Area = "Johor"

The count operator collects the times when the area Johor is defined (*temporalprojection*). In this way inverse temporal functions can be computed. The duration operation computes the length of an interval or of several intervals.

4.3 Aggregations

The following query inquires about the largest collection of flora areas.

SELECT Area (max (Extent)) FROM ForestArea
WHERE Type = "Flora"

The query demonstrates an example of a *spatio-temporal aggregation* operation max which is an extension of the well known aggregation operator in SQL of the same name. It is here applied to a collection of evolving regions contained in a relation column and computes a new evolving region. Internally, this operator is based on a binary function $\boldsymbol{MAX_{ST}}$ applied to two evolving regions $\boldsymbol{R_1}$ and $\boldsymbol{R_2}$ 'and yielding a new evolving region in the following way:

$MAX_{ST}(R_1, R_2) := \{(t, r) \mid t \in time \wedge r \; MAX_{geo}(R_1(t), (R_2(t))\}$

This definition uses a function $\boldsymbol{MAX_{geo}}$ which is applied to two regions R_1 and R_2 and which returns larger of both regions.

$$MAX_{geo}(R_1, R_2) = \begin{cases} r_1 & \text{If area } (r_1) > \text{area } (r_2) \\ r_2 & \text{otherwise} \end{cases}$$

Altogether this means that for 2 evolving regions $R_1, \ldots, R_n$ we first compute the evolving region R = $MAX_{st}(R_1, \ldots MAX_{st}(R_{n-1}, R_n) \ldots)$. Afterwards, we apply the raise area of R, which computes the area of R at all times as a temporal real number.

To investigate sustainability, suitability of our proposed model for handling spatio-temporal data, we have also made comprehensive comparison among different feature of STORe-Biodi, PAHANG database model, MARDI, BODHI and JOHOR database models. To fulfill our comparison, we have stored and manipulated same data that collected from different organizations through STORe-Biodi, MARDI, JOHOR and BODHI database model to perform a set of defined queries on different data set for performing different tests such as checking the model's suitability for handling the spatial, temporal and spatio-temporal data. After applying the certain set of queries, we have obtained results. To examine the model we have used a set of user defined queries for support spatial, temporal and spatio-temporal data. After experiments and results, we have found that defined queries results indicate that fast query response times than other models. This also meant to present a comprehensive comparison between the STORe-Biodi and existing BODHI, MARDI, PAHANG and JOHOR database models. It has been analyzed that JOHOR, MARDI database model provides worst performance regarding handling spatio-temporal biodiversity data; according to our experiments we couldn't get positive results from JOHOR, and MARDI database models, but we received a bit better results from PAHANG and BODHI models, hence we got sufficient quality results applying the same data using our proposed model. After applying the certain set of queries, we have obtained results.

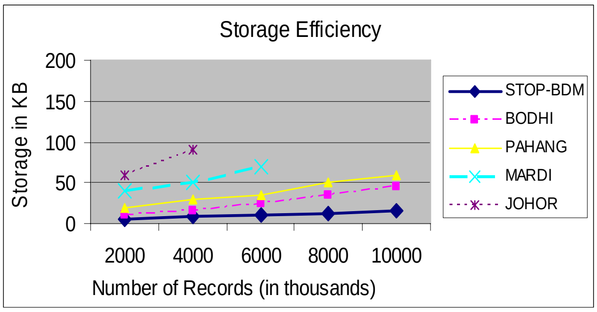


Figure 7: Spatio-temporal queries evaluation among data models

The graphical representation of those results carried out from those database models are shown in the figure 7:

Figure 7 shows that different spatial, temporal and spatio-temporal queries applied on those database models and the above graph results shown that our proposed models (blue line) shows the better performance among those data models for the handling spatio-temporal biodiversity data. To obtain the results, we have applied 1000 to 10,000 records and received the better performance. A pink and yellow line shows a bit better performance than aqua color. Purple color line shows the worst performance during query execution time. During our query execution not all models support temporal or spatio-temporal queries. Johor and mardi database model cannot execute query when data storage records become higher than 4000 and 6000 accordingly. While our proposed models not only able to provide the space to run queries but also produces best results with a very less time.

5. Discussion

Data model has been developed either spatial and temporally. In this paper we have discussed issue of spatio-temporal and data modeling and query language. A number of data models have been suggested for a spatio-temporal system. Some of them are based on the raster approach, others on the vector approach and yet other concepts can be applied to both vector and raster data. Some models are based on the object-oriented paradigm, an approach that is capable of integrating both vector and raster data into one data model. This paper takes into consideration the development of spatio-temporal data model for biodiversity data and minimizes those problems. For the first experimental investigation purpose, we have tried to applied johor forest data for the suitability of the model for handling spatio-temporal data. For this purposes, we inserted and manipulated the biodiversity data through proposed model and perform a specific set of pre-defined temporal and spatio-temporal queries on it as discussed above. The proposed model supports well the spatio-temporal data to perform spatio-temporal queries.

6. Future Works

In the light of spatio-temporal issues, there are great numbers of latent questions associated with the spatio-temporal cognitive representation, temporal taxonomies, topological and metric spatial knowledge, including in implementation issues for temporal biodiversity, etc. The performance of the model should be checked with different and much bigger in size data sets and more time series data. It would provide better evaluation if this model could be compared with other spatio-temporal models in terms of database size and data accessing time. Object-Relational database enabling us to trace biodiversity data with fully function of graphical visualization and should have an ability to detect and monitor small changes of forest data

References

[1] Langran, G. (1992), "Time in Geographic Information Syetems:, Taylor & Francis, London
[2] Sellis, T. (1999). Research issues in Spatiotemporal Database Systems, Springer-Verlag Berlin Heidelberg.
[3] Martin Erwig, R. H. G., Markus Schneider, Michalis Vazirgiannis (1999). "Spatio-Temporal Data Types: An Approach to Modeling and Querying Moving Objects in Databases." Geoinformatica 3(3): 269-296.
[4] Guting, R. H. (1994). "An Introduction to spatial Database Systems." VLDB Journal 4: 357-399.
[5] Tansel, A. U. and J. Clifford (1993). Temporal Databases: Theory, Design, and Implementation, Benjamin/Cummings Publishing Company.
[6] Abraham, T. and J. F. Roddick (1999). "Survey of Spatio-Temporal Databases." Geoinformatica 3(1): 61-69.
[7] Goralwalla, I. A., M. T. Ozsu, et al. (1998). An Object-Oriented Framework for Temporal Data Models, Springer-Verlag ABerlin Heidelberg.
[8] Mennis, J. L., D. J. Peuquet, et al. (2000). "A Conceptual Framework for Incorporating Cognitive principles into geographical database representation." International Journal of Geographical Information System 14(6): 501-520.
[9] Raza, A. and W. Kainz (1999). Cell tuple based spatio-temporal data model. Proceedings of the 7th international symposium on Advances in geographic information systems, ACM Press.
[10] Freksa, C. (1998). Spatial and Temporal Structures in Cognitive Processes, Springer Verlag Heidelberg.
[11] Peuquet, D. J. (2001). "Making Space for Time: Issues in Space-Time Data Representation." GeoInformatica 5(1): 11-32.
[12] Dieter Pfoser, Nectaria Tryfona, (1998), "Requirements, definitions, notations for spatio-temporal application environment", Geographic Information Systems, Proceedings of the 6th ACM international symposium on Advances in geographic information systems, Washington, D.C., United States, Pages: 124 - 130
[13] M. Bouzeghoub, "Morse: a functional query language and its semantic data model", proc. of int. Conf. on Trends an applications of databases, IEEE-NBS, Gaithersburg, USA, 1994
[14] Xiaoyu Wang, Xiaofang Zhou and Sanglu Lu. "Spatio-temporal data modeling and management: A survey". IEEE TOOLS-Asia'00, 2000.

Mohd Taib Wahid (URL:http://www.is.fsksm.utm.my/taib) received first degree in Computer Science (Information System) from University Technology of Malaysia (UTM) in 1996 and Master in Computer Science (Real Time Software Engineering) in 1998 from University Technology of Malaysia and Thomson Campus, France. Currently he is holding the position of lecturer in the Department of Information Systems, University Technology of Malaysia. He is also been actively involved in many research projects related data model, contact research project as a project leader with various public and private sector. His research interest includes database design and data modeling, data retrieval, query optimization and enterprise resource planning.

A.Z.M Kamruzzaman received his first degree in Bachelor of Science (Com) in 2003 from the University Technology Malaysia and currently waiting for the VIVA towards his Master of Science (Com) at the same university. His field of study is spatio-temporal data modeling for biodiversity data. From October 2003 to December 2005 he worked as a research officer with the GIS and biodiversity data model development research group at the Department of Information Systems, University Technology Malaysia. He is the web administrator and Social and Welfare Secretary of International Postgraduate Student Society (IPSS).

Abdul Rashid Mohamed Shariff (URL:http://eng.upm.edu.my/~rashid/work.htm) has a PhD in Spatial Information Science and Engineering from the University of Maine, USA. He is currently the program manager for GIS and Spatial Modeling at the Institute of Advanced Technology in the Universiti Putra Malaysia (UPM). Dr. Rashid is an active researcher, having a wide area of interest. Among his current areas of interest are Spatio-Temporal Data Models, Biodiversity Databases and Agricultural Database Management.