A TAXONOMIC MODEL SUPPORTING HIGH PERFORMANCE SPATIAL-TEMPORAL QUERIES IN SPATIAL DATABASES

Gregory Vert, Rawan Alkhaldi, Sara Nasser, Frederick C. Harris. Jr., Sergiu M. Dascalu
Department of Computer Science and Engineering
University of Nevada, Reno
Reno NV 89509, USA
E-mail: {gvert, alkhaldi, sara, fredh, dascalus}@cse.unr.edu

KEYWORDS: spatial data storage, spatial taxonomies

ABSTRACT

Spatial data has become more important everyday in decision-making and planning processes. As such, it needs to be stored and retrieved in information systems that often require high performance due to the voluminous nature of spatial data. Typically this is not much of a problem unless one considers the effect of spatial extent as a function of time in information retrieval. Taxonomies of spatial objects can be useful in suggesting a storage model that addresses spatio-temporal queries. This research develops such a taxonomy and then proposes how the taxonomy might lend itself to a high performance binary tree model for query and storage of spatial data that considers the relationship of time on the shape of objects in storage. The approach has the potential to retrieve data for certain types of queries much more quickly than a linear search of the same types of spatial objects. Comparative evaluation will be the subject of future work.

INTRODUCTION

Spatial databases have become increasingly more important in everyday life. Although they were first used in government and military operations, they have become a basic commodity for almost every business and home. Today, almost everyone knows how to run a computer. Spatial earth resource data has become increasingly important in everyday decision-making and planning processes, such as resource management and urban planning. As such, there is a need to store, retrieve and manage spatial information in an efficient fashion. Due to the volume of information associated with spatial data, geographic information systems (GIS) need to focused on high performance methods for information management. Some methods to date such as Z curves and Hilbert curves attempt to organize spatial data so that it can be retrieved and stored in a linear fashion that reflects the spatial adjacency of geographic objects. These methods however do not address the temporal aspects of spatial data and the effect of time on geometric shape. Additionally, little work has been done on how to retrieve, store and order spatial information that has geometries that change as a function of time. Taxonomies of spatial data can organize information in a fashion that can be utilized for high performance storage and retrieval of temporally oriented spatial data and may have other applications such as spatial data authentication (Vert et al. 2003).

A taxonomy is typically defined in literature as theories and techniques of naming, describing, organizing, and classifying organisms. For example, the biological taxonomic hierarchy describing life is, from top to bottom: kingdom, phylum (for animals) or division (for plants and fungi), class, order, family, genus, species. Data taxonomies are useful because they can simplify and organize access to data. However, taxonomies have been typically developed in the past for a special purpose or application. The goal of this research has been to develop a method and taxonomy that classifies the spatial data contained in maps based on their type, relationship to other spatial objects, and the time effect on such objects. Implied is that the classification of spatial data should be based on similarities of structure or origin as a function of time. This research proposes the concept of classification of spatial data into taxonomies that can then be utilized for rapid retrieval of spatial information within a spatial extent affected by time.

PREVIOUS WORK

A search of the literature for previous work with spatial taxonomies came up with limited results. The University of California, Santa Barbara has developed geographical data taxonomy for the Alexandria Digital Library (ADL). They called the taxonomy “Feature Type Thesaurus,” in which a thesaurus is defined as a “set of terms representing concepts and the relationships among the terms, including hierarchy, equivalence, and associative relationships” (Jensen and Snodgrass, 1994). The library uses the thesaurus to organize the geographic information based on the nature of the place. This taxonomy was not based on spatial data objects that exist in maps; it rather classifies almost all geographical terminology that is used to describe places in nature. They used “MultiTes” software to create their “Feature Type Thesaurus.” MultiTes is software that has many tools that make it easy to create and manage thesauri, taxonomies, and other types of controlled vocabularies.

Proceedings 21st European Conference on Modelling and Simulation
Ivan Zelinka, Zuzana Oplatková, Alessandra Orsoni ©ECMS 2007
ISBN 978-0-9553018-2-7 / ISBN 978-0-9553018-3-4 (CD)

The following is an example of how the ADL thesaurus (Jensen and Snodgrass, 1994) can be used:

"Sample thesaurus entry with explanations: *Canals* is a feature type category for places such as the Erie Canal. This category is used instead of any of the following:

- Canal bends
- Canalized streams
- Ditch mouths
- Ditches
- Drainage canals
- Drainage ditches
- … More …

Broader terms: Canals is a sub-type of "hydrographic structures."

The following is a list of other categories related to canals (non-hierarchical relationships):

- Channels
- Locks
- Transportation features
- Tunnels

So, canals can be defined as manmade waterway used by watercraft or for drainage, irrigation, mining, or water power."

Calkins and Obermeyer developed a taxonomy of spatial data use and value that aims at enhancing the general understanding of importance and use of geographic information in decision making (Calkins and Obermeyer, 1991). Onsrud and Rushton have suggested a taxonomy of spatial data sharing based on the characteristics of organizations, data, constraints, and exchange that is intended to differentiate the different activities of spatial data-sharing to build future relationship models (Onsrud and Rushton, 1995).

Jensen and Snodgrass have attempted to address time in the spatial context by creating a temporal taxonomy for events and classified them based on a valid time and a transaction time into fifteen different categories. These categories review generalized temporal relations in order to be able to query a predecessor relation from a successor relation for the purposes of information retrieval (Jensen and Snodgrass, 1994).

Peuquet classified temporal objects (Heywood, et al. 2002), based on the type of event that caused the change to the entity (point, line, or polygon), into four types:

- Continuous - events that occur over a period of time.
- Majorative – events that go on most of the time.
- Sporadic – events that occur some of the time.
- Unique – events that occur only once.

For example, for swimmers who go to the beach, swimming would represent a majorative event during a single day because this event (swimming) occurs most of the time during the day. But, for swimmers to visit restaurants and cafes this event could be considered sporadic because it will occur for limited times during the day for breakfast, lunch, and dinner. The existing approaches to spatial taxonomies seem to be created for special purposes and do not actually classify spatial objects but aspects and events related to spatial objects. What has been needed is a new approach that classifies spatial objects in a robust fashion, hierarchically, and considers the nature of time on such objects. The development of such taxonomy can potentially be applied to rapid, high performance wild card queries of spatial databases by suggesting a tree structure for data organization

APPROACH

In developing of this taxonomic approach, several challenges had to be identified and addressed. The first of these was that of trying to collect and list the spatial data objects used in maps. Firstly, there appeared to be an endless number of spatial object types. Upon examination of the data, these were determined to stem from the fact that there appears to be no naming standards for spatial objects. The same object may exist under a different name in a different map. For example, a small lake in one data file may be referred to as a pond in another file. Most of the spatial data examined in this research was from the USGS and Geography departments at the University of Nevada. Additionally, each taxonomic class needed the concept of a time dimension applied to it. For example there are bodies of water that flow occasionally above ground and occasionally below ground in the deserts of the western United States based on the time of year.

The first step of our research was to build a collection of spatial object identifiers and then look for structural classifiers. Because little previous work had been done in this regard we started by examining objects from the previously mentioned taxonomies (Heywood et al., 2002, Calkins and Obermeyer, 1991) and by examining USGS data for the state of Nevada and ESRI spatial objects (ESRI, 2007). The initial categories defined from this work are: Urban, Rural, Forest, Soil, Costal, and Polar. These categories were determined based on the Alexandria Digital Library Feature Type Thesaurus (Jensen and Snodgrass, 1994). For classification we examined the data found in the ESRI Data & Maps Media kit (ESRI, 2002, ESRI, 2007). After collecting the data for the taxonomy, it was easy to distinctly categorize the data types into the categories developed in previous work (Tables 1-7). In addition, it was realized that those groups, which had been developed for purposes other than taxonomies, would not work correctly for data classifications.

Table 1: Hydrography spatial datasets

Lake	River	Ocean
Sea	Bay	Strait
Fiord	Inlet	Sound
Pond	Stream	Swamp
Estuary	Lagoon	Harbor
Shore	Port	Waterfall
Creek	Cove	Dam/weir
Brook	Ditch	Anchorage
Beach	Island	Ice mass
Geyser	Reef	Hot spring
Reach	Desert water	Oases
Gulch	Gulf	Channel
Reservoir	Marsh	Slough
Cape	Canal	Bayou
Bog	Playa	Desert Lakes

Table 2: Urban spatial datasets

Building	Home	Gauging Stations
Fort	University	Telephone Lines
City	Parcel Data	Transmission Lines
Estate	Power Lines	Sewer Sys.& Water Sys.
House	Castle	Communications & Electricity
College	Voting District	National Monument
County	Major Cities	

Table 3: Transportation spatial datasets

Road	Trail	Railroad
Gate	Bridge	Dead end
Lane	Highway	
Crossing	Interchange	Service facility
Blvd	Avenue	Air force base
Bypass	Street	Interstate
Transit	Station	Launch complex
Junction	Fork	Airport
Freeway	Residential	Dirt
Train yard	Route	Drive
Tunnel	Exit	Way
Jeep Trail	Train tracks	Landing strip

Table 4: Vegetation spatial datasets

Park	Lawn	Garden
Land	Timber	Farm
Reserves	Grove	Field
Bushes	Brushwood	Vineyard
Pasture	Cultivated areas	Ranch
Croplands	Zoo	Forest
Wood	Trees	Meadow
Orchard	Plantation	

Table 5: Rural spatial datasets

County	Septic sys. & Wells	Parcel Data
Village	Transmission Lines	
Barn	Gauging Station	Dairies
Corral	Communications & Electricity	Voting Districts
Estate	Telephone Lines	House
Building	Power Lines	
Hamlets	Town	

Table 6: Soil spatial datasets

Clay	Sand sea	Bedrock outcrop
Sand	Silt	Alluvial fans
Rocks	Desert soil	Sand sheet

Table 7: Elevation spatial datasets

Hill	Valley	Mountain
Summit	Gap	Canyon
Mound	Crest	Coulee
Butte	Volcano	Crevasse
Mine	Ridge	Quarry
Mount	Dale	Pike
Foothill		

Grouping of spatial data into these categories was done by structural or semantic basis. For example, man-made structures found in urban areas became members of the urban spatial dataset. Whereas, the hydrography category of spatial objects contains objects that can have varying geometries and thus are classified by the semantics of being objects that are characterized semantically by water. With the spatial data categorized, we then classified each type of data object based on its geometric stability over a given unit of time. Tables 8 (a) and (b) show this the result of this classification. The abbreviations and symbols used in these tables are as follows:

Legend

I ➔ Innate
RE ➔ Relative Extent
S ➔ Seasonal
C ➔ Continual
Sing. ➔ Singular
Y ➔ Yes
L ➔ Large
V ➔ Vary
Blank ➔ N/A
S ➔ Static
ST ➔ Static-Temporal
TC ➔ Temporal-Continuous
TS ➔ Temporal-Sporadic

Table 8 (a): Temporal classification of spatial data classes from the taxonomy

Class	I	RE	S	C	Sing.	TG
Ocean	Y	L				**S**
Ice Mass		L	Y			**TC**
Sea	Y	V				**ST**
Lake		V	Y			**TC**
River		V	Y			**TC**
Desert Water		V	Y			**TC**
Island	Y					**S**
Shore	Y			Y		**TC**
Port					Y	**TS**
Dam/Weir					Y	**TS**
Sand	Y			Y		**TC**
Silt			Y			**TC**
Clay			Y			**TC**
Rocks	Y					**S**
Land		V			Y	**TS**
Forest	Y	L				**S**
Farm		V			Y	**TS**
Park		V			Y	**TS**
Bushes		V	Y			**TC**
Summit	Y					**S**
Mountain	Y	L				**S**
Hill	Y	V				**S**
Valley	Y	V				**S**
Mine		V			Y	**TS**

Grouping spatial objects by temporal attributes produces the classification shown in Table 9. This classification can then be used to order the objects in a category by their spatial extent and the effect of time on the spatial extent. For example, an ice mass is large and is continuously changing spatial extent over a period of time.

In the classification process used for the taxonomy, temporality was studied based on relative features amongst the classes within the same natural category. By mapping classes from each group onto the spatial-temporal plots, classes could be compared on the basis of spatial extend and time. The spatial-temporal plots are logarithmic representations of the relative size and temporality of spatial objects that can show the relative relationship between different objects.

The spatial-temporal plots use the log scales of area and time as their axis. For example, classes of the static-temporal group had been studied on the spatial-temporal plot shown in Figure 1. The Ocean class has been placed on the top right corner because it is large in size and changes in its extent happen over a very long period of time. Island, mountain, forest, hill, and valley classes have been clustered in the middle of the spatial-temporal plot. The classes of rocks and summit have been placed at the lower left corner. The presented mapping is based on our interpretation of the different classes; other interpretations may produce different outcomes. For simplicity purposes, Figure 1 can be represented in aggregate form as shown in Figure 2.

Table 8 (b): Temporal classification of spatial data classes from the taxonomy

Class	I	RE	S	C	Sing.	TG
County		L			Y	**TS**
City		L			Y	**TS**
University		V			Y	**TS**
Building		V			Y	**TS**
National Monument					Y	**TS**
Voting District		L			Y	**TS**
Parcel Data		L			Y	**TS**
Sewer and Water Systems					Y	**TS**
Communic ation and Electricity					Y	**TS**
County		L			Y	**TS**
Village		L			Y	**TS**
Town		L			Y	**TS**
Building		V			Y	**TS**
Corral		V			Y	**TS**
Voting District		L			Y	**TS**
Parcel Data		L			Y	**TS**
Septic Systems and Wells					Y	**TS**
Communic ation and Electricity					Y	**TS**
Road					Y	**TS**
Trail					Y	**TS**
Crossing					Y	**TS**
Railroad					Y	**TS**
Airport		V			Y	**TS**
Bypass					Y	**TS**
Service Facility		V			Y	**TS**

All classes from the developed taxonomic classification with similar temporal classifications have been represented on the spatial-temporal plot as shown in Figures 1, 2, 3, 4, and 5.

Table 9: Temporality classification of the taxonomy

Static	Static-Temporal	Temporal-Continuous	Temporal-Sporadic
Ocean	Sea	Ice mass	Port
Island		Desert water	Land
Rocks		Shore	Farm
Forest		Sand	Park
Summit		Silt	University
Mountain		Clay	Parcel data
Hill		Bushes	Corral
Valley		Lake	Dam/Weir
		River	Mines
			County
			Building
			National Monument
			Voting Districts
			Road
			Crossing
			Railroad
			Bypass
			Service Facility
			Septic sys/ Wells
			City
			Airport
			Village
			Town
			Trail
			Sewer sys/ water sys.
			Communications & Electricity

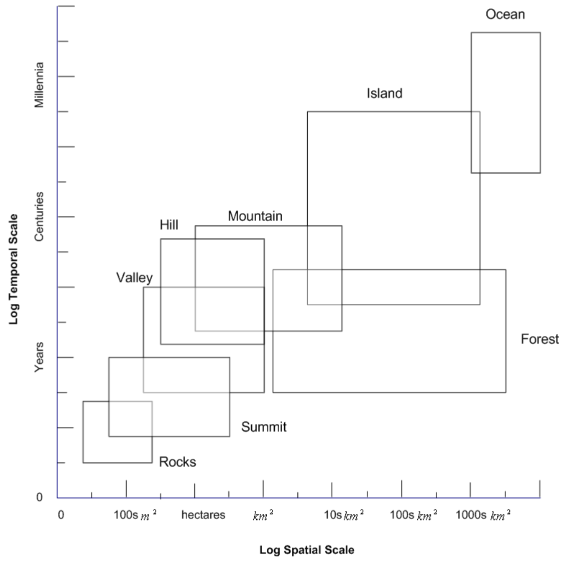

Figure 1: Spatial temporal plot for classes with Static (S) classification

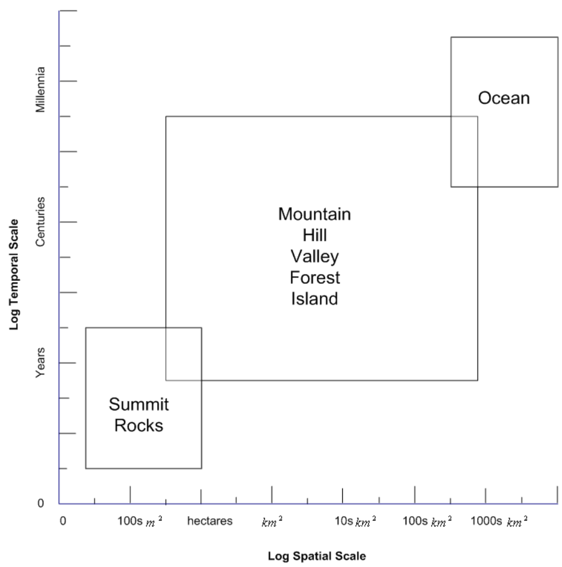

Figure 2: Aggregated representation of spatial temporal plot for static group in Figure 1.

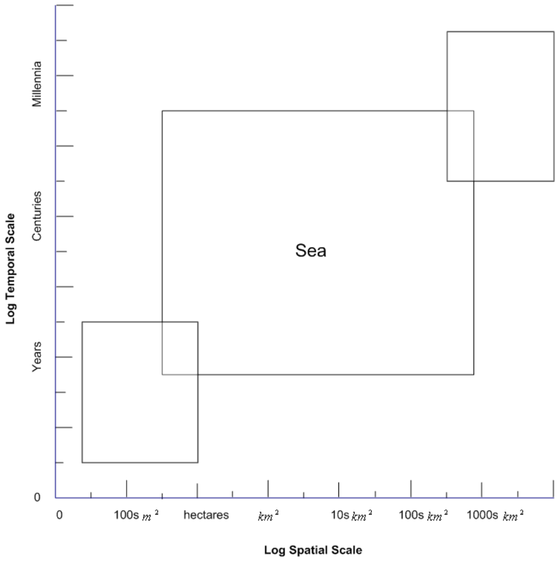

Figure 3: Aggregated spatial temporal plot for static temporal (ST) classification

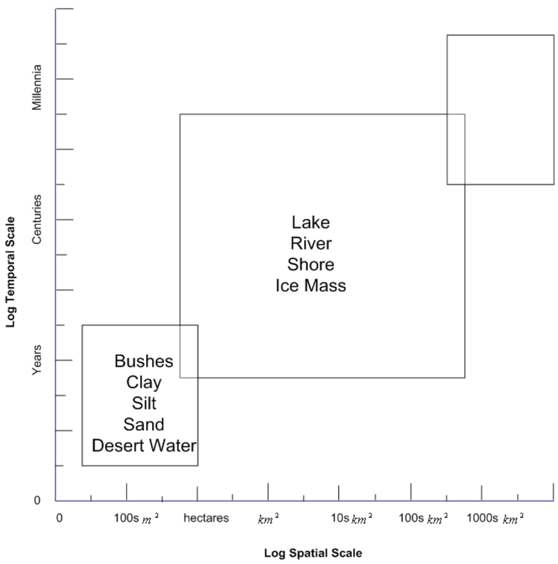

Figure 4: Aggregated spatial temporal plot for temporal continuous (TC) classification.

APPLICATION

The temporal spatial plots developed in Figures 1-5 can be used for the design and creation of binary storage structures that can be used to support high performance range and bound queries of spatial data. To illustrate the application, imagine a query of the form:

*Retrieve * Static objects with spatial extent > 1 hectacre on time scale > n years*

Using Figure 2, this query would traverse a tree structure for Static object and return Mountain, Hill Valley, Forest, Island and Ocean objects. While the subject of future research, the spatial temporal taxonomies seems to suggest a binary tree structure (Figure 6) whose worst search time would be `O(log n)` versus `O(n)` for a linear search of the same data. Such a tree structure might look like the one shown in Figure 6, where objects are organized by spatial extent and trees are characteristic of the temporal classification of objects in the tree.

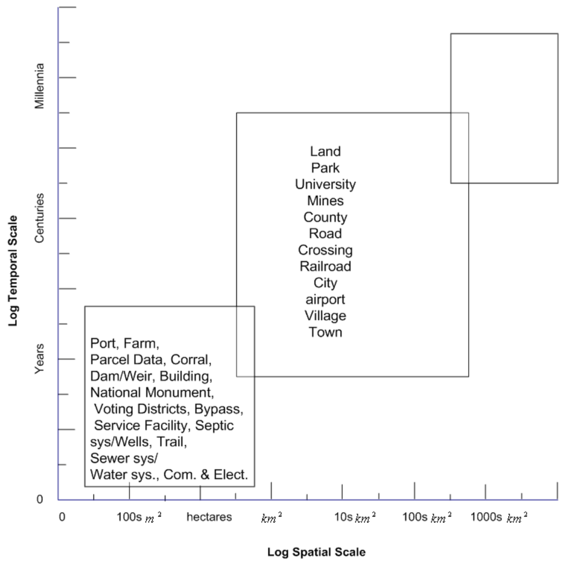

Figure 5: Aggregated spatial temporal plot for temporal sporadic (TS) classification

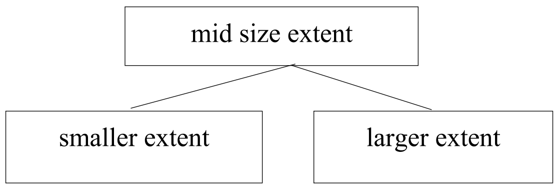

Figure 6: Architectural search tree based on taxonomies

When candidate objects are found in a node in the above search tree, a similar binary search is run to meet the temporal taxonomic criteria of the query as shown in Figure 7.

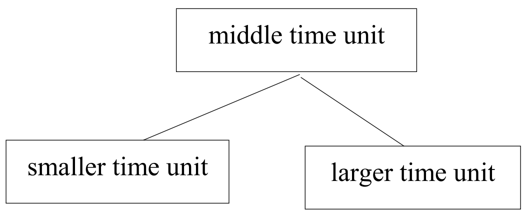

Figure 7: Temporal taxonomic search tree

This can then produce a search time for spatial temporal queries of `O(lg (n + m))` where n is the number of nodes in the spatial tree and m is the number of elements in the spatial tree. In contrast a linear search would run in O(n+m) time. Consequently, taxonomies based on spatial extent as a function of time, appear to be useful in developing data structures and organization that can support high performance queries. Additionally, such an approach incorporates the concept of temporality in a way that previous approaches do not. Evaluation and empirical measurement of exactly how much faster this approach is will be the subject of future research.

FUTURE WORK AND CONCLUSIONS

The spatio-temporal taxonomy developed in this research classifies spatial data objects by their physical extents and by their temporal properties. This classification is then turned into a taxonomy which leads to the organization of the taxonomy into spatio-temporal plots. These plots suggest a way to organize spatial data that supports queries to a database based on time and the physical extent of a spatial object as a function of time. The suggested performance improvement is dramatic versus a linear search and provides support not found in spatial data queries such as Z curves and Hilbert curves. Future work will include more precise techniques for quantification of relationships in the taxonomy. Additional work will be done to empirically measure the performance of this technique versus other storage structures.

REFERENCES

Calkins, H.W. and Obermeyer, N.J., 1991, Taxonomy for surveying the use and value of geographical information. *International Journal of Geographic Information Systems* **5**(3): 341-351.

ESRI, 2007, Introduction to ArcView 3.x., ESRI Virtual Campus, GIS Education and Training on the Web. Retrieved March 19, 2007 from: http://campus.esri.com/

ESRI, 2002, Data & Maps, Media Kit, ESRI ArcGIS. http://www.esri.com

Heywood, I., Cornelius, S. and Carver, S., 2002, *An Introduction to Geographical Information Systems*, 2nd Edition, Prentice Hall.

Jensen, C.S. and Snodgrass, R., 1994, Temporal specialization and generalization. *IEEE Transactions on Knowledge and Data Engineering* **6**(6): 954-974.

Onsrud, H.J. and Rushton, G., 1995, Sharing Geographic Information. Center for Urban Policy Research, New Brunswick, N.J. pp. 510.

Takeyama, M. and Couclelis, H., 1997, Map dynamics: integrating cellular automata and GIS through Geo-Algebra. *International Journal of Geographical Information Science* **11**: 73-91.

Vert, G. Yuan, B. Cole, N., 2003, "A Visual Algebra for Detecting Port Attacks on Computer Systems," Proceedings of the Intl. Conf. on Computer Applications in Industry and Engineering (CAINE-2003), Las Vegas, NV, pp 131-135.

AUTHOR BIOGRAPHIES

GREGORY VERT was born in Fairfield California. He received his BS in Geography with a specialization in GIS from the University of Washington in 1985, his MS in Information Systems Management, Seattle Pacific University, in Seattle, Washington in 1988, and his PhD in Computer Science from the University of Idaho in Moscow Idaho in 2000. He has been an Assistant Professor at the University of Nevada, Reno in the Computer Science and Engineering Department since 2002. His research is in the areas of GIS, Computer Security, Fuzzy System, Database and Bioinformatics. His email address is gvert@cse.unr.edu

RAWAN ALKHALDI was born in Jordan. She received her MS in Computer Science from the University of Nevada in 2006. Her research is in the area of Spatial Databases and Spatial Taxonomies. Her email address is alkhaldi@cse.unr.edu

SARA NASSER was born in Hyderabad, India. She received her BE in Computer Science and Engineering from Muffakham Jah College of Engineering and Technology, Osmania University, and her MS in Computer Science from the University of Nevada in 2003. She is currently a PhD student in Computer Science and Engineering and is expecting to graduate in 2008. Her research is in the area of Bioinformatics and Fuzzy Systems. Her e-mail address is sara@cse.unr.edu

FREDERICK C. HARRIS, JR. was born in San Jose, California, USA. He received his BS in Mathematics and MS in Educational Administration from Bob Jones University in 1986 and 1988, and his MS and PhD from Clemson University in 1991 and 1994. He has been at the University of Nevada, Reno since 1994, and is currently a Professor in the Computer Science and Engineering Department. His research is in the areas of Parallel and Distributed Computing, as well Graphics and Virtual Reality. His e-mail address is: fredh@cse.unr.edu

SERGIU M. DASCALU was born in Bucharest, Romania. He received his MS in Automatic Control & Computers from the Polytechnic of Bucharest, and his PhD in Computer Science from the University of Dalhousie, Halifax, Nova Scotia, Canada in 2001. He has been at the University of Nevada since 2002 and is currently an Assistant Professor in the Computer Science and Engineering Department. His research is in the areas of Software Engineering and Human-Computer Interaction. His email address is: dascalus@cse.unr.edu