G/Technology

► Intergraph: Real Time Operational Geospatial Applications

Gaussian

► Hurricane Wind Fields, Multivariate Modeling

Gaussian Process Models in Spatial Data Mining

NAREN RAMAKRISHNAN[1], CHRIS BAILEY-KELLOGG[2]
[1] Department of Computer Science,
Virginia Tech, Blacksburg, VA, USA
[2] Department of Computer Science,
Dartmouth College, Hanover, NH, USA

Synonyms

Active data mining

Definition

Gaussian processes (GPs) are local approximation techniques that model spatial data by placing (and updating) priors on the covariance structures underlying the data. Originally developed for geo-spatial contexts, they are also applicable in general contexts that involve computing and modeling with multi-level spatial aggregates, e. g., modeling a configuration space for crystallographic design, casting folding energies as a function of a protein's contact map, and formulation of vaccination policies taking into account social dynamics of individuals. Typically, we assume a parametrized covariance structure underlying the data to be modeled. We estimate the covariance parameters conditional on the locations for which we have observed data, and use the inferred structure to make predictions at new locations. GPs have a probabilistic basis that allow us to estimate variances at unsampled locations, aiding in the design of targeted sampling strategies.

Historical Background

The underlying ideas behind GPs can be traced back to the geostatistics technique called kriging [4], named after the South African miner Danie Krige. Kriging in this literature was used to model response variables (e. g., ozone concentrations) over 2D spatial fields as realizations of a stochastic process. Sacks et al. [12] described the use of kriging to model (deterministic) computer experiments. It took more than a decade from this point for the larger computer science community to investigate GPs for pattern analysis purposes. Thus, in the recent past, GPs have witnessed a revival primarily due to work in the statistical pattern recognition community [5] and graphical models literature [3]. Neal established the connection between Gaussian processes and neural networks with an infinite number of hidden units [8]. Such relationships allow us to take traditional learning techniques and re-express them as imposing a particular covariance structure on the joint distribution of inputs. For instance, we can take a trained neural network and mine the covariance structure implied by the weights (given mild assumptions such as a Gaussian prior over the weight space). Williams motivates the usefulness of such studies and describes common covariance functions [14]. Williams and Barber [15] describe how the Gaussian process framework can be extended to classification in which the modeled variable is categorical. Since these publications were introduced, interest in GPs has exploded with rapid publications in conferences such as ICML, NIPS; see also the recently published book by Rasmussen and Williams [11].

Scientific Fundamentals

A GP can be formally defined as a collection of random variables, any finite subset of which have a (multivariate) normal distribution. For simplicity, we assume 2D spatially distributed (scalar) response variables t_i, one for each location $\mathbf{x_i} = [x_{i1}, x_{i2}]$ where we have collected a data sample. Observe that, in the limiting case, each random variable has a Gaussian distribution (but it is not true that any collection of Gaussian random variables will induce a GP). Given a dataset $\mathcal{D} = \{\mathbf{x_i}, t_i\}$, $i = 1 \dots n$, and a new data point $\mathbf{x_{n+1}}$, a GP can be used to model the posterior $P(t_{n+1} \mid \mathcal{D}, x_{n+1})$ (which would also be a Gaussian). This is essentially what many Bayesian modeling techniques do (e. g., least squares approximation with normally distributed noise), however, it is the specifics of how the posterior is modeled that make GPs distinct as a class of modeling techniques.

To make a prediction of t_{n+1} at a point $\mathbf{x_{n+1}}$, GPs place greater reliance on t_i's from nearby points. This reliance is specified in the form of a covariance prior for the process. One example of a covariance prior is:

$$\mathrm{Cov}(t_i, t_j) = \alpha \exp\left(-\frac{1}{2}\sum_{k=1}^{2} a_k (x_{ik} - x_{jk})^2\right) . \tag{1}$$

Intuitively, this function captures the notion that response variables at nearby points must have high correlation. In Eq. 1, α is an overall scaling term, whereas a_1, a_2 define the length scales for the two dimensions. However, this prior (or even its posterior) does not directly allow us to determine t_j from t_i since the structure only captures the covariance; predictions of a response variable for new sample locations are thus conditionally dependent on the measured response variables and *their* sample locations. Hence, we must first estimate the covariance parameters (a_1, a_2, and α) from $\mathcal{D}$, and then use these parameters *along with* $\mathcal{D}$ to predict t_{n+1} at $\mathbf{x_{n+1}}$.

Before covering the learning procedure for the covariance parameters (a_1, a_2, and α), it is helpful to develop expressions for the posterior of the response variable in terms of these parameters. Since the jpdf of the response variables $P(t_1, t_2, \dots, t_{n+1})$ is modeled Gaussian (we will assume a mean of zero), we can write:

$$P(t_1, t_2, \dots, t_{n+1} \mid \mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_{n+1}}, \mathrm{Cov}_{n+1}) = \frac{1}{\lambda_1} \cdot \exp\left(-\frac{1}{2}\,[t_1, t_2, \dots, t_{n+1}]\,\mathrm{Cov}_{n+1}^{-1}\,[t_1, t_2, \dots, t_{n+1}]^{\mathrm{T}}\right)$$

where we ignore λ_1 as it is simply a normalizing factor. Here, Cov_{n+1} is the covariance matrix formed from the $(n+1)$ data values $(\mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_{n+1}})$. A distribution for the unknown variable t_{n+1} can then be obtained as:

$$\begin{aligned} &P(t_{n+1} | t_1, t_2, \dots, t_n, \mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_{n+1}}, \mathrm{Cov}_{n+1}) \\ &= \frac{P(t_1, t_2, \dots, t_{n+1} \mid \mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_{n+1}}, \mathrm{Cov}_{n+1})}{P(t_1, t_2, \dots, t_n \mid \mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_{n+1}}, \mathrm{Cov}_{n+1})} \\ &= \frac{P(t_1, t_2, \dots, t_{n+1} \mid \mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_{n+1}}, \mathrm{Cov}_{n+1})}{P(t_1, t_2, \dots, t_n \mid \mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_n}, \mathrm{Cov}_n)} , \end{aligned}$$

where the last step follows by conditional independence of $\{t_1, t_2, \dots, t_n\}$ w.r.t. $\mathbf{x_{n+1}}$ and the part of Cov_{n+1} not contained in Cov_n. The denominator in the above expression is another Gaussian random variable given by:

$$\begin{aligned} &P(t_1, t_2, \dots, t_n \mid \mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_n}, \mathrm{Cov}_n) \\ &= \frac{1}{\lambda_2} \exp\left(-\frac{1}{2}\,[t_1, t_2, \dots, t_n]\,\mathrm{Cov}_n^{-1}\,[t_1, t_2, \dots, t_n]^{\mathrm{T}}\right) . \end{aligned}$$

Putting it all together, we get:

$$\begin{aligned} &P(t_{n+1} | t_1, t_2, \dots, t_n, \mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_{n+1}}, \mathrm{Cov}_{n+1}) \\ &= \frac{\lambda_2}{\lambda_1} \exp\Bigg(-\frac{1}{2}[t_1, t_2, \dots, t_{n+1}]\,\mathrm{Cov}_{n+1}^{-1}[t_1, t_2, \dots, t_{n+1}]^{\mathrm{T}} \\ &- \frac{1}{2}[t_1, t_2, \dots, t_n]\,\mathrm{Cov}_n^{-1}\,[t_1, t_2, \dots, t_n]^{\mathrm{T}}\Bigg) . \end{aligned}$$

Computing the mean and variance of this Gaussian distribution, we get an estimate of t_{n+1} as:

$$\hat{t}_{n+1} = \mathbf{k}^{\mathrm{T}}\mathrm{Cov}_n^{-1}[t_1, t_2, \dots, t_n] , \tag{2}$$

and our uncertainty in this estimates as:

$$\sigma^2_{\hat{t}_{n+1}} = k - \mathbf{k}^{\mathrm{T}}\mathrm{Cov}_n^{-1}\mathbf{k} , \tag{3}$$

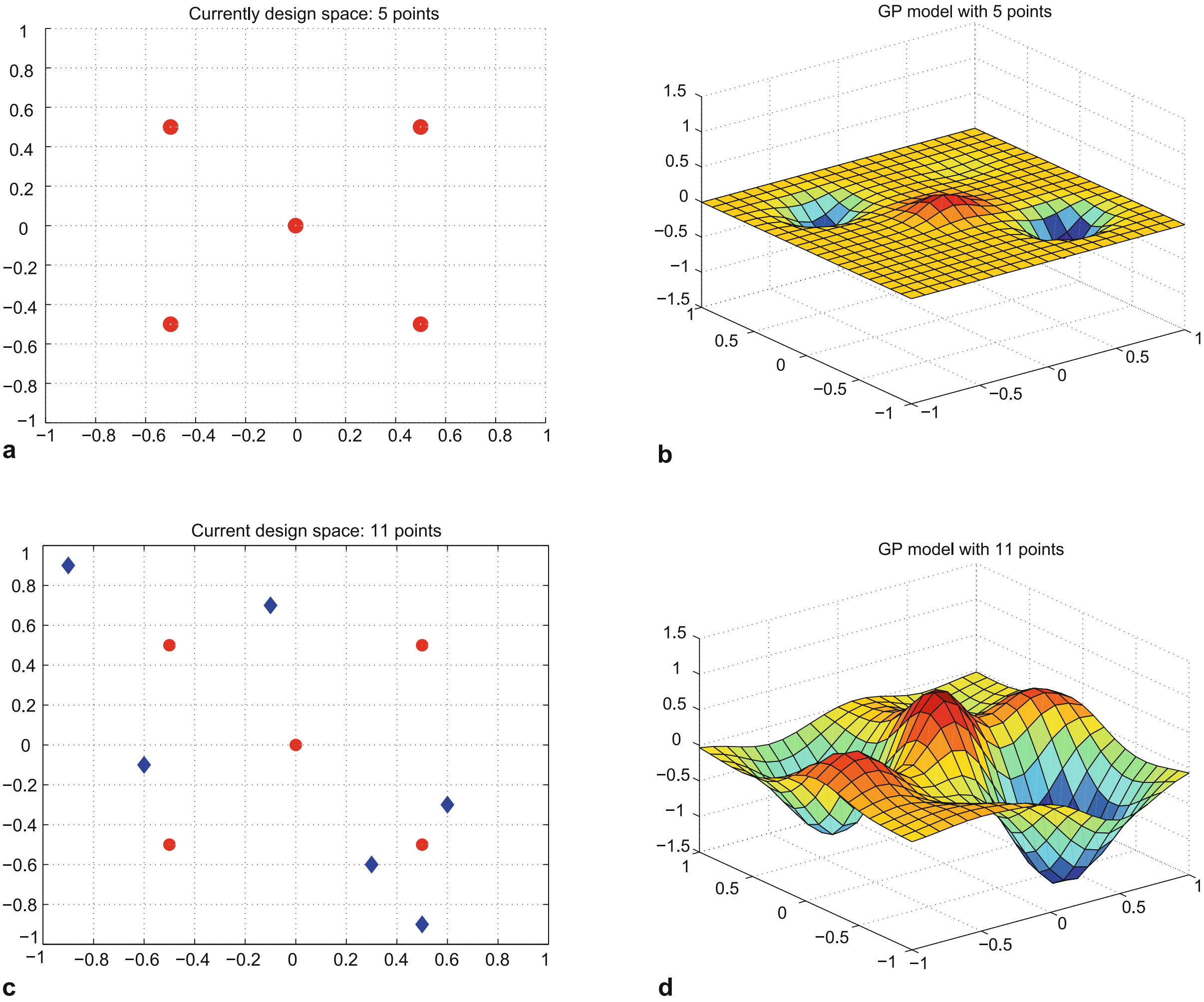
where $\mathbf{k}^{\mathrm{T}}$ represents the n-vector of covariances with the new data point:

$$\mathbf{k}^{\mathrm{T}} = [\mathrm{Cov}(\mathbf{x_1}, \mathbf{x_{n+1}})\ \mathrm{Cov}(\mathbf{x_2}, \mathbf{x_{n+1}})\ \dots\ \mathrm{Cov}(\mathbf{x_n}, \mathbf{x_{n+1}})] ,$$

and k is the $(n+1,\ n+1)$ entry of Cov_{n+1}. Equations 2 and 3, together, give us both an approximation at any given point and an uncertainty in this approximation; they will serve as the basic building blocks for closing-the-loop between data modeling and higher level mining functionality.

The above expressions can be alternatively derived by positing a linear probabilistic model and optimizing for the MSE (mean squared error) between observed and predicted response values (e. g., see [12]). In this sense, the Gaussian process model considered here is also known as the BLUE (best linear unbiased estimator), but GPs are not restricted to linear combinations of basis functions.

To apply GP modeling to a given dataset, one must first ensure that the chosen covariance structure matches the data characteristics. The above example used a stationary structure which applies when the covariance is translation invariant. Various other functions have been studied in the literature (e. g., see [7,9,12]), all of which satisfy the required property of positive definiteness of a covariance



Gaussian Process Models in Spatial Data Mining, Figure 1 Active mining with Gaussian processes. An initial sample of data points (**a**; shown as *red circles*) gives a preliminary approximation to the target function (**b**). Active sampling suggests new locations (**c**; *blue diamonds*) that improve the quality of approximation (**d**)

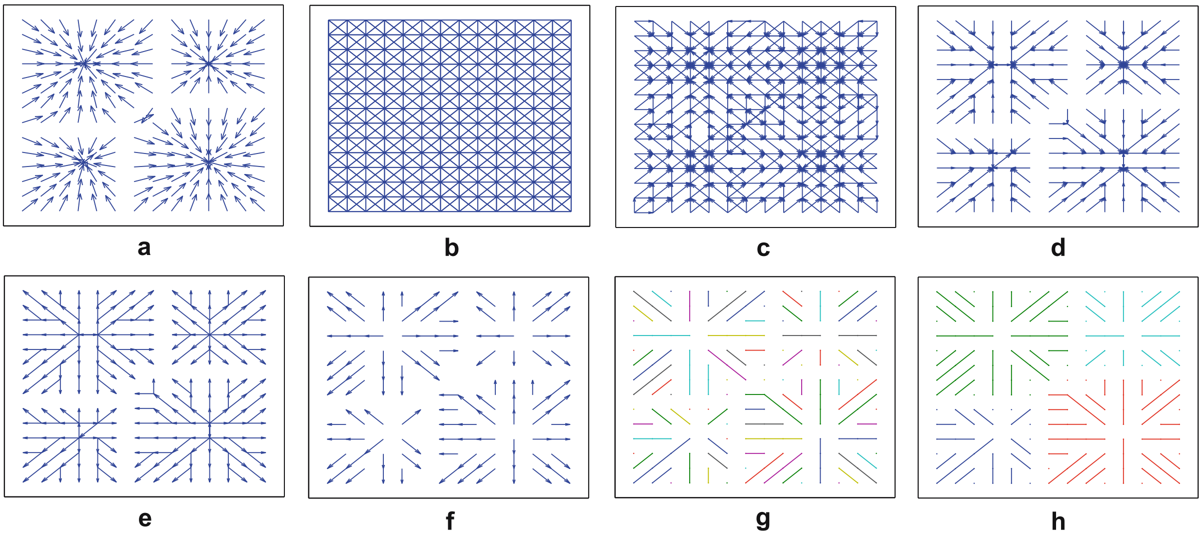
matrix. The simplest covariance function yields a diagonal matrix, but this means that no data sample can have an influence on other locations, and the GP approach offers no particular advantages. In general, by placing a prior directly on the function space, GPs are appropriate for modeling 'smooth' functions. The terms a_1, a_2 capture how quickly the influence of a data sample decays in each direction and, thus, the length scales for smoothness.

An important point to note is that even though the GP realization is one of a random process, we can nevertheless build a GP model for deterministic functions by choosing a covariance structure that ensures the diagonal correlations to be 1 (i. e., perfect reproducibility when queried for a sample whose value is known). Also, the assumption of zero mean for the Gaussian process can be relaxed by including a constant term (gives another parameter to be estimated) in the covariance formulation.

Learning the GP parameters $\theta = (a_1, a_2, \alpha)$ can be undertaken in the maximum likelihood (ML) and maximum a posteriori (MAP) frameworks, or in the true Bayesian setting where we obtain a distribution over values. The log-likelihood for the parameters is given by:

$$\begin{aligned}\mathcal{L} &= \log P(t_1, t_2, \ldots, t_n | \mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}, \theta) \\ &= c + \log P(\theta) - \frac{n}{2}\log(2\pi) - \frac{1}{2}\log \mid \mathrm{Cov}_n \mid \\ &\quad - \frac{1}{2}[t_1, t_2, \ldots, t_n]\ \mathrm{Cov}_n^{-1}\ [t_1, t_2, \ldots, t_n]^{\mathrm{T}} .\end{aligned}$$

To optimize for the parameters, we can compute partial derivatives of the log-likelihood for use with a conjugate



Gaussian Process Models in Spatial Data Mining, Figure 2 Computation of multi-level spatial aggregations. **a** Input vector field. **b** 8-adjacency neighborhood graph. **c** Forward neighbors. **d** Best forward neighbors. **e** Neighborhood graph transposed from best forward neighbors. **f** Best backward neighbors. **g** Resulting adjacencies redescribed as curves. **h** Higher-level aggregation and classification of curves whose flows converge

gradient or other optimization algorithm:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \theta} &= \frac{\partial \log P(\theta)}{\partial \theta}\\ &\quad - \frac{1}{2}\,\mathrm{tr}\left(\mathrm{Cov}_n^{-1}\,\frac{\partial\,\mathrm{Cov}_n^{-1}}{\partial \theta}\right)\\ &\quad + \frac{1}{2}\,[t_1, t_2, \ldots, t_n]\;\mathrm{Cov}_n^{-1}\;\frac{\partial\,\mathrm{Cov}_n^{-1}}{\partial \theta}\\ &\quad \mathrm{Cov}_n^{-1}\;[t_1, t_2, \ldots, t_n]^{\mathrm{T}}\,,\end{aligned}$$

where $\mathrm{tr}(\cdot)$ denotes the trace function. In our running example, we need only estimate three parameters for θ, well within the purview of modern numerical optimization software. For larger numbers of parameters, we can resort to the use of Monte Carlo Markov Chain (MCMC) methods [9].

Key Applications

Gaussian processes are applicable for spatial modeling tasks in a variety of application contexts.

Active Data Mining

In applications such as crystallographic design, where one must characterize a configuration space or design space in terms of spatial aggregates, data collection can become costly. In these applications, it is beneficial to collect data only at those locations that are deemed important to support a data mining objective. Toward this goal, we can use GPs to work with only a sparse set of samples and, based on the quality of approximation, provide objective criteria for choosing the next sample point. Figure 1 depicts a 2D example of 'seeding' a GP with an initial sample of data points (left two frames), thereby defining functionals over the unsampled region (not shown) which are then optimized to arrive at new locations to sample (right two frames).

Geostatistical Motion Interpolation

Gaussian processes have been used to solve the motion interpolation or 'in-betweening' task in computer graphics [6]. Given two frames denoting an individual in motion and a multi-parameter space of control variables, a GP model synthesizes smooth animations that emulate natural human movements and obey geographical constraints. GPs have also been used for robotic imitation by modeling data gathered from human motion capture devices [13].

Spatial Aggregation

GPs can be used to model the multi-layer construction of spatial aggregates from data. Figure 2 describes steps in aggregating individual vectors, first into streamlines and then into convergent flows, using a custom spatial aggregation algorithm. The qualitative nature of such aggregations can be summarized computationally using GPs to yield mathematical models of data mining algorithms.

Sensor Networks

GPs have been applied in sensor network contexts [2], e. g., monitoring physical variables over an environment using a number of sensing devices. By parametrizing the covariance distribution of the physical variable and determining where uncertainty of estimation is highest, one can design judicious sensor placement policies.

Future Directions

There are many open and promising directions for Gaussian processes research. There are new, overlapping, notions of spatiality that must be modeled in applications such as pandemic disease modeling [1]. In these contexts, the definition of nearby random variables is drawn both from geographical distance as well as social proximity considerations. From work that merely estimates parameters of covariance functions, new work has begun to learn the structure of covariance functions. These will undoubtedly become more critical as new applications of GPs are explored. Finally, as the sensor network application reveals, the development of new objective functions for active data mining is crucial, especially for those that are suited for distributed model building.

Cross References

► Kriging

Acknowledgments

The figures in this chapter were published previously in [10] and reproduced here with permission from SIAM Press.

Recommended Reading

1. Bailey-Kellogg, C., Ramakrishnan, N., Marathe, M.V.: Spatial Data Mining for Pandemic Preparedness. ACM SIGKDD Explorations **8**(1), 80–82 (2006)
2. Guestrin, C., Krause, A., Singh, A.P.: Near-optimal Sensor Placements in Gaussian Processes. In: Proceedings of the Twenty-Second International Conference (ICML 2005), pp. 265–272. (2005)
3. Jordan, M.I.(ed.): Learning in Graphical Models. MIT Press, Cambrige, MA (1998)
4. Journel, A.G., Huijbregts, C.J.: Mining Geostatistics. Academic Press, New York (1992)
5. MacKay, D.J.: Gaussian Processes: A Replacement for Supervised Neural Networks? In: Lecture Notes of Tutorial at Neural Information Processing Systems (NIPS'97) (1997)
6. Mukai, T., Kuriyama, S.: Geostatistical Motion Interpolation. ACM Trans. Graph. **24**(3), 1062–1070 (2005)
7. Nabney, I.T.: Netlab: Algorithms for Pattern Recognition. Springer-Verlag, New York (2002)
8. Neal, R.M.: Bayesian Learning for Neural Networks. Lecture Notes in Statistics No. 118. Springer-Verlag, New York (1996)
9. Neal, R.M.: Monte Carlo Implementations of Gaussian Process Models for Bayesian Regression and Classification. Technical Report 9702, Department of Statistics, University of Toronto, January 1997
10. Ramakrishnan, N., Bailey-Kellogg, C., Tadepalli, S., Pandey, V.N.: Gaussian Process for Active Data Mining of Spatial Aggregates. In: Proceedings of the SIAM International Conference on Data Mining (2005)
11. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. MIT Press, Cambrige, MA (2006)
12. Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P.: Design and Analysis of Computer Experiments. Stat. Sci. **4**(4), 409–435 (1989)
13. Shon, A.P., Grochow, K., Rao, R.P.N.: Robotic Imitation from Human Motion Capture using Gaussian Processes. In: Technical Report, Department of Computer Science and Engineering, University of Washington, Seattle, WA (2006)
14. Williams, C.K.I.: Prediction with Gaussian Processes: From Linear Regression to Linear Prediction and Beyond. In: Jordan, M.I.(ed.) Learning in Graphical Models, pp. 599–621. MIT Press, Cambridge, MA (1998)
15. Williams, C.K.I., Barber, D.: Bayesian Classification with Gaussian Processes. IEEE PAMI **20**(12), 1342–1351 (1998)

Gazeteer

► Retrieval Algorithms, Spatial

GDAL

► Open-Source GIS Libraries

GE Smallworld

► Smallworld Software Suite

Geary Coefficient

► Geary's *C*

Geary Ratio

► Geary's *C*

Geary's *C*

XIAOBO ZHOU, HENRY LIN
Department of Crop and Soil Sciences, The Pennsylvania State University, University Park, PA, USA

Synonyms

Geary's index; Geary ratio; Geary coefficient

Definition

Geary's *C* tests statistics for spatial autocorrelation by using the sum of squared differences between pairs of data of variable x as a measure of covariation

$$C = \frac{(n-1)\sum_i \sum_j w_{ij}(x_i - x_j)^2}{2nS^2 \sum_i \sum_j w_{ij}} .$$

Where x_i denotes the observed value at location i,

$$S^2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2 ,$$

$\bar{x}$ is the mean of the variable x over the n locations and w_{ij} are the elements of the spatial weights matrix, defined as 1 if location i is contiguous to location j and 0 otherwise. Other spatial weights matrices can also be used.

Main Text

Geary's C ranges from 0 to a positive value. The value of C is 1 in the absence of spatial autocorrelation. A low value of C $(0 < C < 1)$ represents a positive spatial autocorrelation and approaches zero for strong autocorrelation. A high value $(C > 1)$ represents negative spatial autocorrelation with greater values corresponding to a strong negative spatial autocorrelation. Geary's C is more sensitive to the variation of neighborhoods than to the global variation.

Cross References

- ► Autocorrelation, Spatial

Geary's Index

- ► Geary's *C*

Generalization

- ► Map Generalization
- ► Privacy Threats in Location-Based Services

Generalization and Symbolization

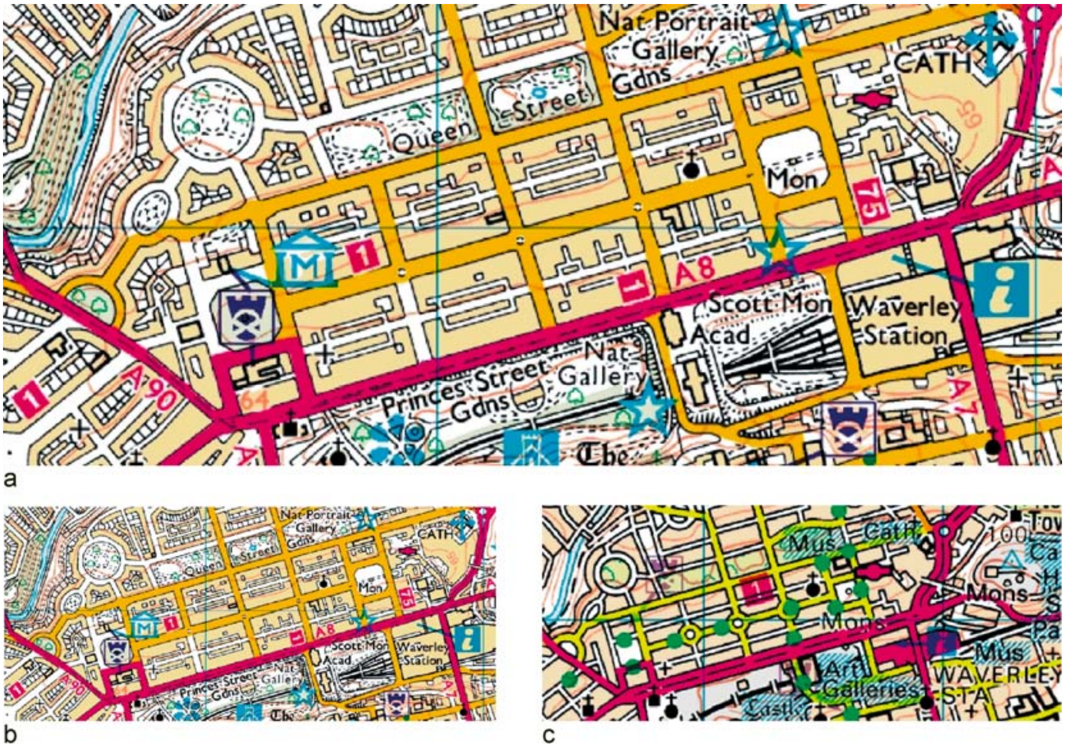
William A. Mackaness, Omair Chaudhry
Institute of Geography, School of GeoSciences,
The University of Edinburgh, Edinburgh, UK

Definition

Map generalization is a process concerned with the application of a set of algorithms to geographic data (represented in vector form) in order to control the optimal representation of geographic phenomenon at a range of different scales or levels of detail. In that sense, generalization seeks to mirror the process of map design previously undertaken by the human cartographer. In the context of geographical information systems (GIS), this process is modeled as two sets of operations: the first is a set of database operations (model generalization) and the second is a set of visualization operations (cartographic generalization). Model generalization is concerned with simplifying the representational form in order to achieve efficiencies in data storage, selecting classes of objects according to some specified scale and map theme, and aggregating groups of objects in accordance with scale constraints. Cartographic generalization (a compliment to model generalization) is concerned with the optimal portrayal of those selected and aggregated features. Cartographic generalization involves selecting appropriate symbols, giving emphasis to some of the feature's defining characteristics, and where there are dense regions of features, omitting some features or making small displacements to features in order to resolve ambiguity. Figure 1 seeks to demonstrate the need for generalization. Simple photographic reduction is not sufficient (Fig. 1b); thus the aim of map generalization is to derive smaller scale mapping (Fig. 1c) from detailed, large scale mapping (Fig. 1a).

Historical Background

All geographical processes are imbued with scale [1:214], thus issues of scale are an essential consideration in geographical problem solving. The scale of observation governs what phenomena can be viewed, what patterns are discernible, and what processes can be inferred. Study in the geosciences is focused both on the detail of those phenomena, as well as the broad linkages across regional and global space. Choosing scales of analysis, comparing output at different scales, describing constructions of scale [2] are all common practices in the geosciences. Traditionally it has been the cartographer's responsibility to select a scale, to symbolize the phenomena, and to give meaning through the addition of appropriate contextual information. The paper map was the basis of geographical inquiry. Indeed it was argued that if the problem 'cannot be studied fundamentally by maps – usually by a comparison of several maps – then it is questionable whether or not it is within the field of geography' [3:249]. Information technology has not devalued the power of the map, but it has driven a series of paradigm shifts in the storage, representation and interaction with geographical information. Early work in automated mapping focused on supporting the activities of the human cartographer who remained central to the map design process. Current research is focused more on ideas of autonomous design – systems capable of selecting optimum solutions among a variety of candidate



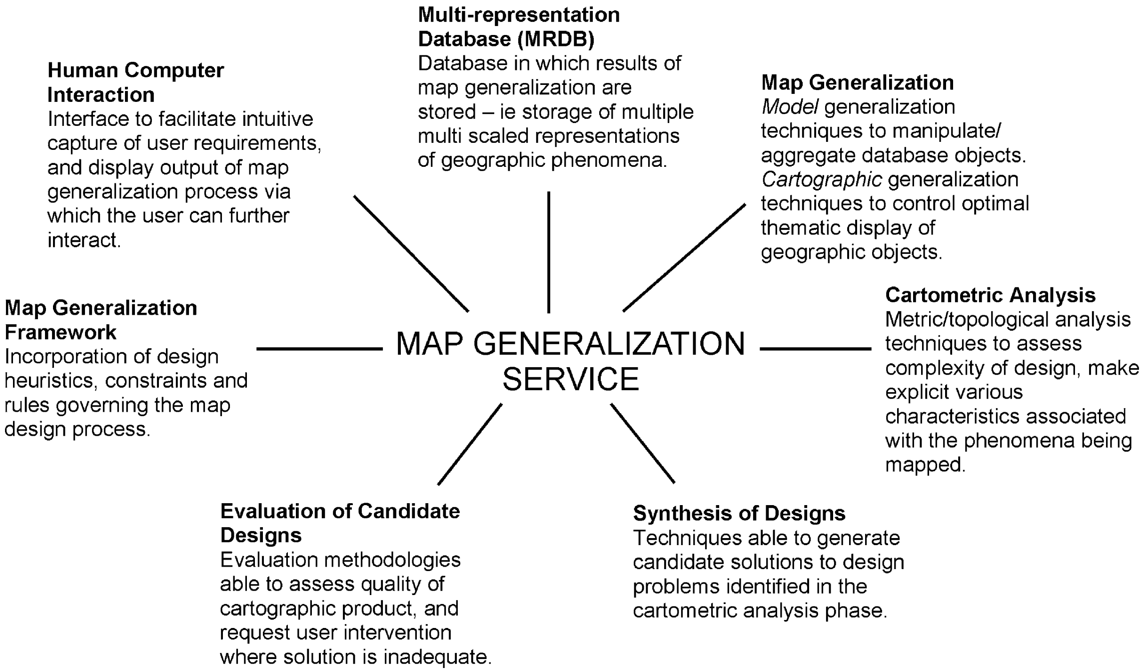
Generalization and Symbolization, Figure 1 Map generalization – creating different geographies of space (Mapping is Ordnance Survey © Crown Copyright. All rights reserved)

solutions delivered over the web, in a variety of thematic forms, in anticipation of users who have little or no cartographic skill. Historically the paper map reflected a state of knowledge. Now it is the database that is the knowledge store, with the map as the metaphorical window by which geographic information is dynamically explored. In these interactive environments, the art and science of cartography is being extended to support the integration of distributed data collected at varying levels of detail, whilst conforming to issues of data quality and interoperability. With respect to map generalization, the challenge is in developing a set of algorithms and methodologies that mirror the service traditionally provided by the human cartographer, yet takes advantage of the paradigm shift afforded by information science in interacting with, and exploring geographic information.

Scientific Fundamentals

The human cartographer provides a service that involves interpreting the requirements of the user, creating and executing a design to a very high quality and clarity according to a theme and scale, and one that is void of ambiguity. Over the past thirty years huge advances in database technology, together with developments in geo-visualization [4,5] and interactive and web based mapping has disrupted and further displaced the role of the cartographer. The digital map now acts as a window by which to search and explore the underlying database, and the cartographer has supposedly been replaced by symbol libraries and color ramps that may, in the absence of cartographic expertise, facilitate 'the creation of cartographic monstrosities with unprecedented ease' [6].

Within this paradigm shift, the requirement to view the world at different scales (or multiple levels of detail) has remained, as has the requirement to produce high quality cartographic products. Initially paper maps at different scales were digitized and stored in different databases. However there is huge redundancy in this model as changes in the real world have to be reflected in changes in each of the databases. A new line of thinking has emerged which asks whether it is possible to store the phenomenon once (at a very high level of detail), and then apply a range of algorithms in order to control the selection and representation of the phenomenon in a form appropriate to the intended scale. There are significant benefits to this line of thinking; maintaining a single database is more cost effective than maintaining multiple databases; a high level of consistency can be maintained between different datasets; duplication of storage can be avoided thus obviating the



Generalization and Symbolization, Figure 2 The components of a Map Generalization Service

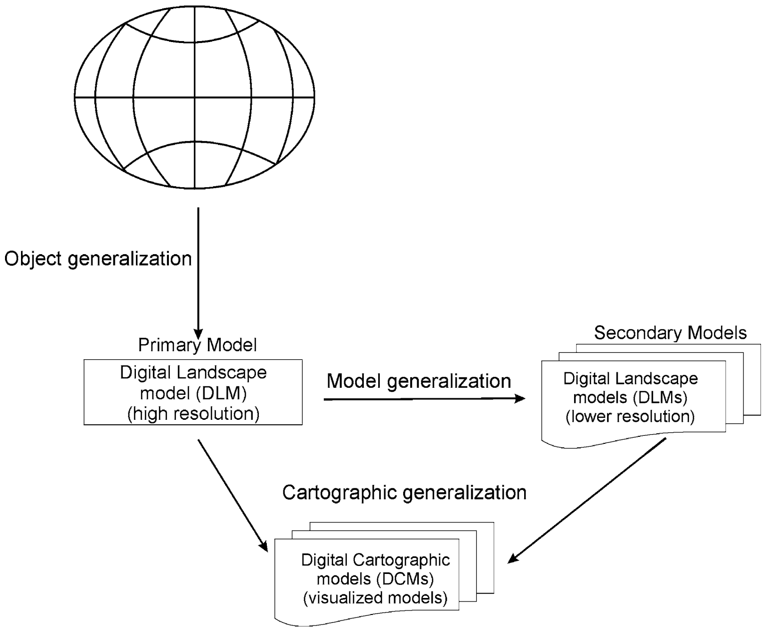
need to make multiple updates across separate databases each time a change occurs in the real world. Most importantly it offers the opportunity to share data, enabling integration of data from disparate sources, captured at different levels of detail.

These benefits are premised on the existence of a set of algorithms that can, with minimum intervention from the user, control the selection and representation of geographic phenomenon according to a specified scale and theme. The science of 'map generalization' is all about designing such algorithms; algorithms that manipulate and symbolize the geometric primitives stored in the database. Map generalization can also be viewed as a service that anticipates users unfamiliar with cartographic concepts, and with poor evaluation skills. Such a service must contain the following components: a database capable of storing multiple representations of geographic phenomena, a set of model and cartographic generalization techniques to create such multiple representations, and design heuristics that govern the appropriate choice and sequencing of generalization techniques. The evaluation of any candidate design requires the system to create alternate candidate designs (synthesis), and to evaluate and select the best solution (which in turn requires a set of cartometric analysis tools). Interpreting the map requirements of the user, and presenting solutions in response requires an interface that can 'translate' straightforward requests into rich specifications and parameter setting. These are deemed to be the essential components of a Map Generalization Service (Fig. 2).

This chapter begins by describing the techniques used to manipulate objects within the database. It then describes some of the frameworks designed to support their application in the overall design of the map. The discussion that follows this, argues that high levels of automation can only be achieved if the automated environment includes methods of evaluation. The entry concludes with a brief discussion of the changing context of map generalization within developing applications (such as exploratory data analysis and location based services).

Tools and Techniques for Map Generalization

The goal of map generalization is to give emphasis to salient objects and their properties whilst omitting less important qualities with respect to the scale and the purpose of a map. Therefore a system is needed that supports manipulation of map objects and their relationships, and more generally supports the representation of phenomena at different scales. For example at the finest scale each individual building, street light and pavement might be represented. But at a coarse scale, all of this might be subsumed by a single 'dot' (with say, the word 'London' next to it), representing the idea of 'city' in which all those buildings are contained. Therefore the requirements for a map generalization system are: 1) a database



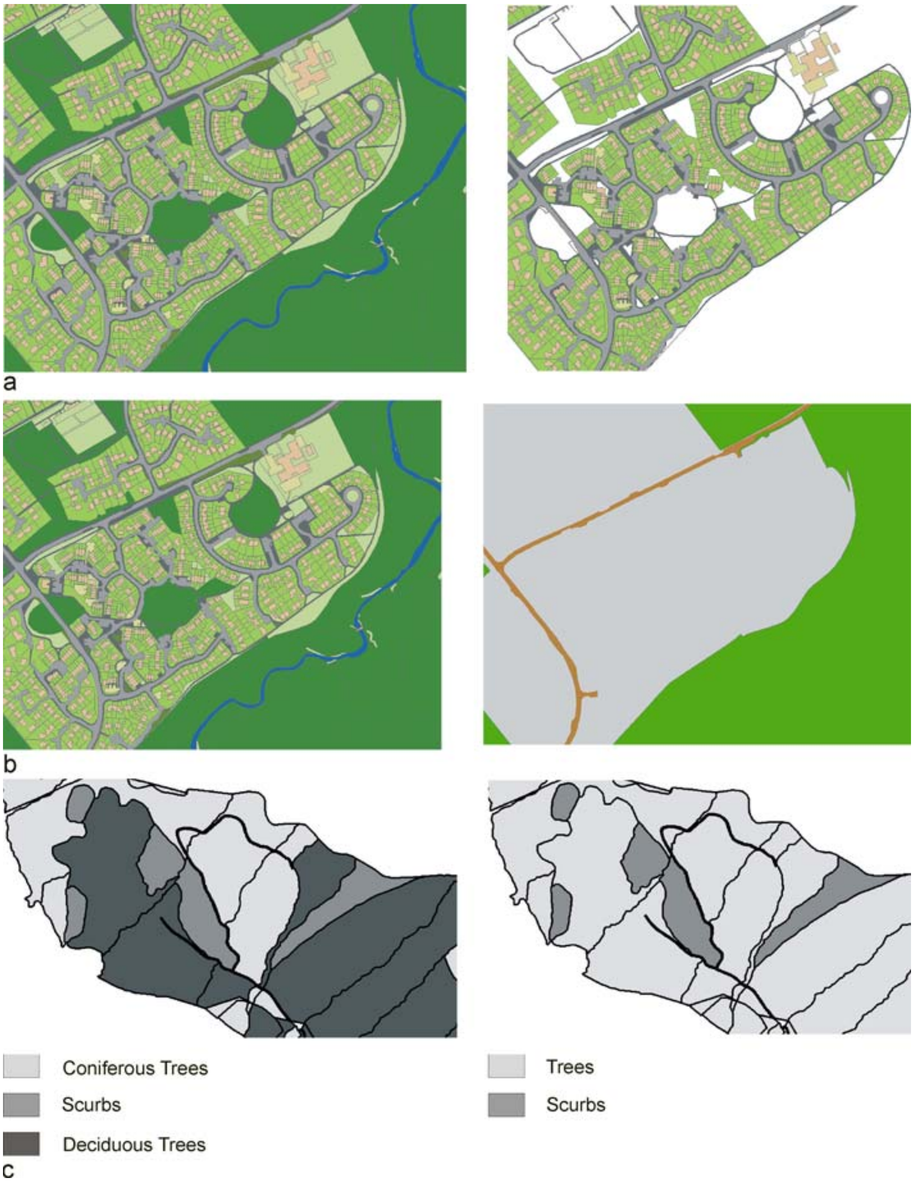
Generalization and Symbolization, Figure 3
DLM, DCM, Model and Cartographic Generalization

containing some abstraction of the real world, 2) a set of algorithms for aggregating objects in that database (model generalization), 3) a library of symbols with which to render the objects according to various themes, and 4) a set of algorithms focusing on improving the legibility of those symbolized objects (cartographic generalization). The database containing that first abstraction is typically called a digital landscape model (DLM – Fig. 3) [7]. The DLM might be created by digitizing paper maps, or from photogrammetric techniques applied to remotely sensed imagery. Typically a notional scale is associated with the DLM database though it is more apposite to talk of level of detail. Data from the database can be symbolized and visualized directly via cartographic techniques. Alternatively a database of lower semantic and geometric resolution can first be derived (via model generalization) – creating different digital cartographic models (DCM – Fig. 3) before cartographic generalization techniques are applied to produce different maps.

Altering the theme, and level of detail enables different phenomena and different properties to be portrayed. Sometimes the emphasis is on precision of location, or of shape (important in the map interpretation process). In other circumstances, the emphasis may be on connectivity at the expense of other properties and qualities. Maps of transportation networks (such as the London Underground) are a nice example of the need to emphasize connectivity over geographical location. Irrespective of theme, in all cases a map (digital or paper) reflects a compromise in design – a compromise between wanting to convey information unambiguously but not having enough room (given the minimum size of symbology) to show all that information. In this sense the process of design is about making sense of things – the cartographer perhaps working from a mental thumbnail sketch by which their solution reflects the requirements of the user in terms of their needs, which in turn governs and constrains the representation of each feature in the map.

Various methodologies have been proposed that try to capture this design process within an automated environment. Considerable research effort has gone into creating algorithms that mimic these human techniques. These techniques are not applied in isolation, but rather in concert, and in varying degree, across the map, depending on the density of information, and the type of phenomenon being mapped, and of course, the theme and scale. Therefore in addition to algorithms that mimic these techniques, a framework is required that can orchestrate this whole design process, together with some evaluation methodologies required to assess the quality of the solution produced within such a framework. Next is a review of generalization techniques under the headings of model and cartographic generalization.

Model Generalization The objective of model generalization techniques is to reclassify and reduce down the detail, thus giving emphasis to entities associated with the broader landscapes – enabling us to convey the extent of the forests rather than see the trees, or to see the island chain along the plate margin, rather than the individual island. The model generalization process is not concerned with issues of legibility and visualization. It is more useful

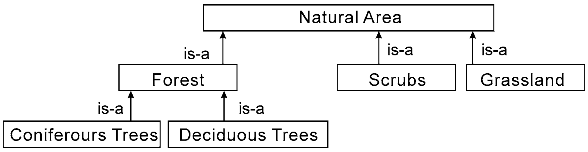


Generalization and Symbolization, Figure 4 **a** Selection, **b** Aggregation and **c** Classification. (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved)

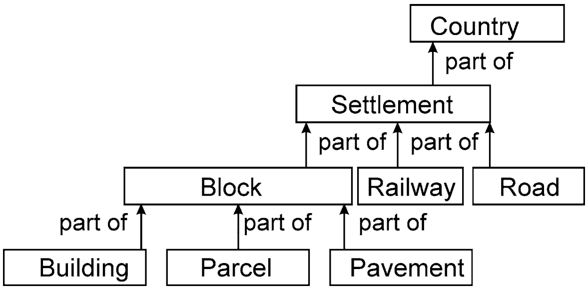
to view it as a filtering process; a set of techniques concerned with 1) selection of phenomena according to theme, and 2) the classification and aggregation of phenomena. As the name suggests, selection is the (straightforward) process of selecting a subset of all classes of objects falling within a specified region (Fig. 4). The selection process is governed by task, which in turn tends to define both the intended scale and theme. The long tradition of topographic and thematic mapping often acts as a basis for specifying content, and thus which classes of objects are selected.

Typically model generalization precedes cartographic generalization. It may also be required in response to a non-visual query, or as a prerequisite to data analysis. For example the question 'what modes of travel exist between the cities of Edinburgh and Glasgow?' requires us to aggregate together phenomena at the fine scale (in this case dense regions of buildings) in order to define the extent and general location of these two entities. Only then can the major roads connecting these two urban centers be identified.

Composite or higher order objects are formed via the process of thematic and spatial abstraction. In thematic abstraction the number of distinct attributes of objects in the database is reduced. In spatial abstraction the number objects is reduced by means of aggregation or elimination. Thematic abstraction often triggers spatial abstraction. For instance objects having similar attribute structure can be categorized into classes under the process of classification.



Generalization and Symbolization, Figure 5 Example of a taxonomy



Generalization and Symbolization, Figure 6 Example of a partonomy

Each object then becomes an instance of a particular class and that class defines an object's properties in terms of its attribute structure. If different classes share some attributes then a super class or parent class can be created whose attributes are the common attributes of its child classes. This creates a hierarchy where complex classes are present at the detailed (low end of a hierarchy) and increasingly abstracted classes are present as one travels up the hierarchy. This type of hierarchy is called a taxonomy or classification hierarchy (Fig. 5) and can be used as a basis for classification of data ('classification' Fig. 4).

Another complimentary hierarchy useful in the creation of composite objects is a partonomy. Whereas a taxonomy refers to a 'is-a' relationship, a partonomy refers to 'part-of' relationships between parent and child classes – reflecting more of a functional and conceptual division of geographic space (Fig. 6) [8]. Over large changes in scale it is necessary to aggregate objects belonging to different classes in order to create composite objects. A prototypical view of a city might be defined as a dense collection of municipal and industrial buildings, and multi modal transportation infrastructures. Once represented in partonomic form, it can be used as a basis for combining such objects together ('aggregation' Fig. 4).

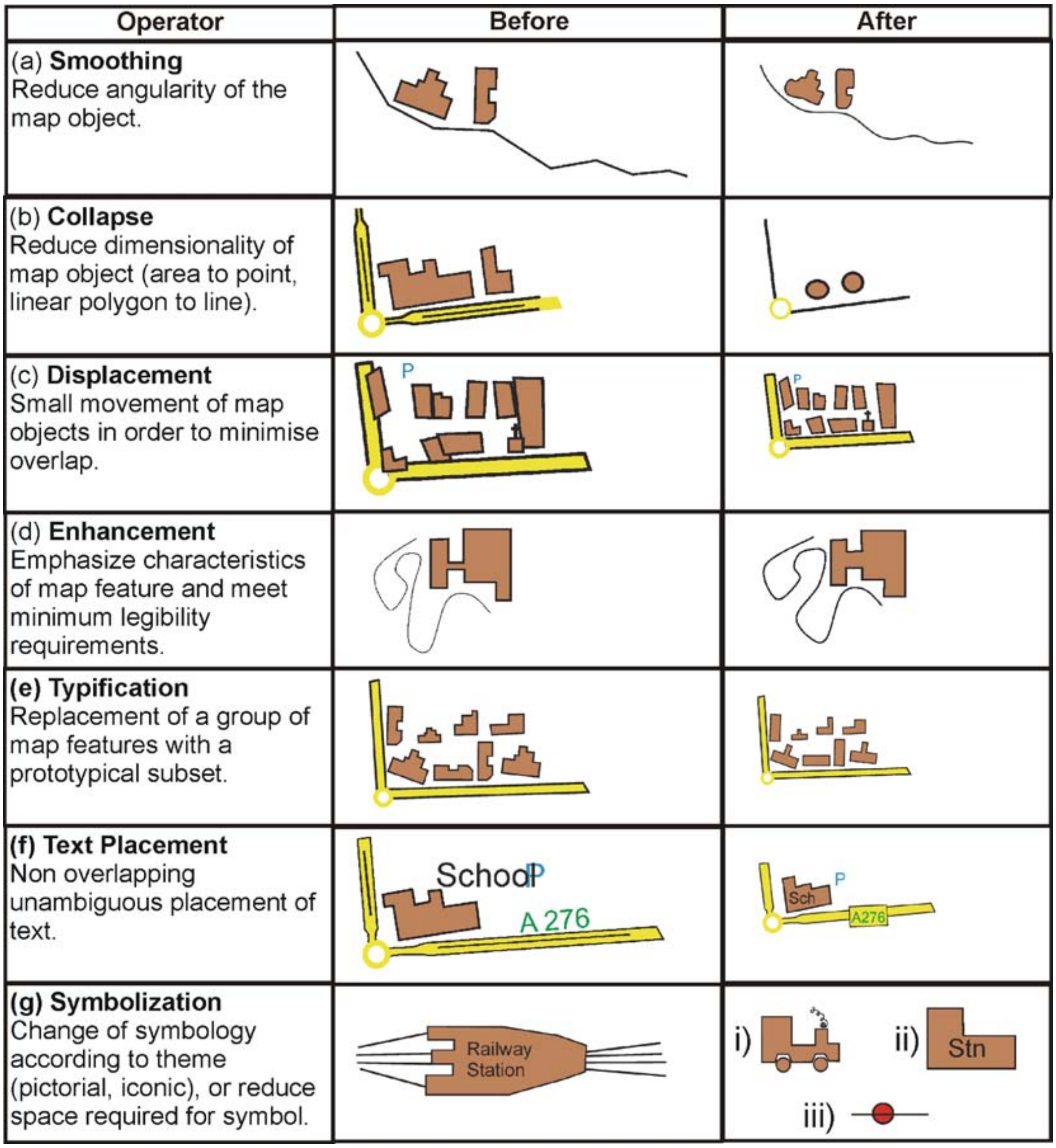
In addition to the techniques of selection and aggregation, there is 'simplification' – which is defined as the process of reducing the number of geometric points used to store the physical location or extent of a geographic object. One can envisage many points being used to record the detail of the outline of a gothic cathedral, or the sinuous path of a low lying river. The challenge of simplification is to reduce the number of points used to store the representation of such features, but in a way that still conveys their essential shape and location. Successful simplification reduces storage requirements and processing time. Once the model generalization process is completed, the challenge is then to render those objects into some map space (whether it is for paper production, or as part of a digital interactive environment – in either a desktop or mobile environment).

Cartographic Generalization Cartographic generalization involves symbolizing the selected data, and applying a set of techniques that optimally convey the salient characteristics of that data, including careful placement of associated text. Symbols used to represent spatial objects from the source database need to be visible to the naked eye. As the scale reduces the amount of space available decreases thus creating competition for space among the symbology. To retain clarity and to represent the information effectively a range of techniques are applied such as symbolization, smoothing, simplification, grouping, enhancement, displacement, and text placement (Fig. 7). These techniques (often applied in combination), seek to give prominence to the essential qualities of the feature portrayed (that rivers retain their sinuous and connected form, and buildings retain their anthropogenic qualities – such as their angular form). Different combinations, amounts of application, and different orderings of these techniques can produce different yet aesthetically acceptable solutions. The focus is not on making changes to information contained in the database, but to solely focus upon avoiding ambiguity in the interpretation of the image. The process is one of compromise reflecting the long held view among cartographers that making maps involves telling small lies in order to tell the truth!

Analysis, Synthesis and Evaluation of Cartographic Solutions

For any given cartographic conflict, one can envisage a number of viable solutions. The choice of solutions will depend on: the density of features, their position relative to one another, and their importance relative to the intended theme. Trying to create alternate viable solutions (synthesis), and then choosing a solution amongst that choice requires two things: 1) an initial analysis phase in which the conflicts are identified (analysis) and a form of evaluation such that the quality of the solution can be assessed (evaluation). Failure to find an adequate solution might either result in further analysis of the conflict or flagging unresolved conflicts and drawing these to the attention of the user.

The analysis phase is akin to the eyes of the cartographer and involves making assessment of the degree of severity

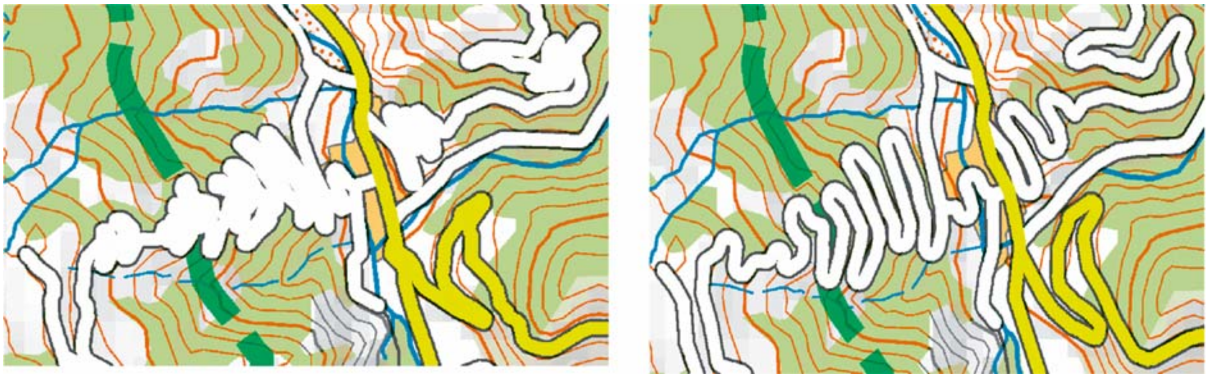


Generalization and Symbolization, Figure 7 Cartographic generalization operations

of the conflict (extent and complexity and composition). A broad and extensive set of cartometric techniques have been developed to measure the various qualities inherent among a set of map objects. This analysis is required because the goal is to ensure minimum disruption in those qualities during the cartographic generalization process. Many shape and pattern metric techniques have been proposed to measure and minimize the effects of cartographic generalization [9,10]. These are often applied in the analysis phase, and again in the evaluation phase. The best solution among a set of candidate solutions might be the one that has resolved the conflict (improved its legibility), whilst producing the least amount of change among the various cartometric measures (in terms of topology, orientation, area, shape and distance).

Modeling the Generalization Process

The selection and application of generalization techniques, the creation of candidate solutions and their evaluation requires some framework in which this can all take place. Because of the interdependent nature of geographic phenomena, it is rare that changes can be made without having to consider the broader context. For example the solution in Fig. 7c is only appropriate because there is sufficient space for the objects to be displaced into. If buildings have



Generalization and Symbolization, Figure 8 Example output from the IGN's agent based system

to be aggregated in one part of the map (perhaps because of the density of features) then for reasons of consistency, this needs to be applied in other similar instances. Procedural and heuristic knowledge needs to be incorporated within these frameworks so that the solutions most likely to be successful can be applied first. Among the various 'frameworks' explored, two are worthy of mention: rule based approaches, and constraint based approaches.

Since the cartographic design process involves decision making and heuristics ('rules of thumb'), it was assumed that knowledge based approaches (expert systems) could be used to model the process – using a rule based approach. These systems used either a predetermined rule execution sequence or an inference engine to control the execution sequence in applying various techniques. They consisted of three main parts: a knowledge base, an inference engine and a user interface. The knowledge base contained a set of rules, facts or procedures. The inference engine controlled the generalization process by making use of the rules and procedures in the knowledge base. The user interface supported the process of data selection and a mechanism for adding or updating rules in the knowledge base [11].

More recently generalization research has focused on an holistic view of the process acknowledging the knock on effects of generalization and the interdependent nature of the solution. Currently there is much interest (and promise) in using constraint based approaches – where the aim is to find a state whereby the maximum number of constraints can be satisfied. In this context, much research effort has been devoted to agent based methodologies – in which each object in the database is modeled as an agent – an object oriented concept in which the object has goals, behaviors, and a capacity to communicate with other agents. These are referred to as 'multi agent systems'. The goals reflect those of the generalization process – namely to efficiently render the object without ambiguity. The agent makes decisions about its representation based on its own goals whilst considering the goals and constraints of its neighbors. Ideas have included a hierarchy of agents in which higher order agents are concerned with broader contexts and distribution of agent classes, whilst agents at the individual object level are concerned with the specific representation of individual objects. The AGENT [12] project is one project which has been developed into a commercial system that now supports a number of national mapping agencies, notably the National Mapping Agency of France (IGN). Figure 8 shows the result from the Carto2001 project [13].

By partially incorporating the decision making process within both rule based and agent based systems, the balance of decision making has shifted away from the human to the machine. This has presented some real challenges in the design of interfaces that are intuitive to use, allowing the user to specify their mapping requirements in a simple and efficient manner within a very complex system. Researchers have challenged the idea of totally autonomous solutions, arguing that interaction is critical to ensuring that the user remains very much part of the design process. The idea of semi autonomous generalization techniques, involving the user in critical evaluation tasks reflects a more collaborative approach to design. Coupled with machine learning techniques, this scenario might enable capture of design heuristics – thus gradually improving the sophistication of proffered solutions.

Key Applications

The idea that map generalization is some 'cartographic end process' belies its importance in supporting five key activities:

Cartographic Assistant

The existence of many different generalization techniques means that a 'cartographic toolbox' is available for use by a trained cartographer. Research efforts have yielded a set of algorithms able to analyze map content, and to consistently generalize classes of objects in clearly defined ways. In this collaborative environment, such systems have the capacity to improve the quality of cartographic training, ensure quality control in the design process and enable refinement in the adjustment of parameters used to control generalization techniques.

Map Generalization Service

In the absence of the cartographer, and in the context of GIS, users (with limited cartographic knowledge) require assistance in the rapid design and delivery of cartographic products – often via the Internet, that can vary in theme and scale according to task. Completely autonomous solutions (with no user intervention) have proved to be very difficult to design, but in any case are not desirable where meaning is often derived through interaction and exploration of the data. The idea of a map generalization service is that maps can be delivered over the Internet in response to user requests – which in turn has led to a focus on the pre-processing of solutions, in which intermediate solutions are stored in a multiple representation database (MRDB).

Populating Multiple Representation Databases

There currently exist multiple, often disconnected ‘silo’ databases containing data at different levels of detail. The vision is that model generalization techniques are applied to data captured at the finest detail in order to create a hierarchical framework of increasingly aggregated geographic phenomena (from house, to suburb, to city to region, to country) – in effect a semantically indexed structure from which different scale linked phenomena can be extracted and queried. The benefit of this approach is consistency and ‘lineage’ (provenance) by which the source objects from which the higher order geographies have been created can be identified. This can support both data integration, and hugely facilitate the data update process. The existence of MRDB can also support on-the-fly generalization and instantaneous delivery of geographic data over the Internet and mobile devices [14,15].

Spatial Data Integration Service

Considerable ‘value add’ comes from the sharing and integration of data. Integration of geographic data is beset by a host of challenges receiving considerable attention – notably in development of shared data schemas, and addressing ontological issues linked to culture, original purpose and conceptual understandings of place. Many of these issues relate to the notional scale at which the data was originally captured. Model generalization techniques can play a critical role in aggregating data according to shared partonomic and taxonomic classification methodologies.

Future Directions

Generalization in the context of geographical information science has an importance beyond traditional cartographic lines. It has everything to do with revealing and giving emphasis to properties inherent among geographic phenomena – and therefore has important cross over with ideas of design (making sense of things), data mining and geo-visualization [16]. The aggregation of phenomena is dependent on taxonomic and partonomic hierarchies, which themselves reflect complex functional and contextual interdependencies inherent among geographic phenomena. In this sense, issues of generalization are central to meaningful interrogation and analysis of all geographic information.

Cross References

- ▶ Conflation of Geospatial Data
- ▶ Data Analysis, Spatial
- ▶ Data Warehouses and GIS
- ▶ Exploratory Visualization
- ▶ Generalization, On-the-Fly
- ▶ Hierarchical Spatial Models
- ▶ Hierarchies and Level of Detail
- ▶ Map Generalization
- ▶ Mobile Usage and Adaptive Visualization
- ▶ Scale, Effects
- ▶ Web Mapping and Web Cartography

Recommended Reading

1. Taylor, P.J.: Is there a Europe of cities? World cities and the limitations of Geographical Scale Analyses. In: Sheppard, E., McMaster, B. (eds.) Scale and Geographic Inquiry: Nature, Society, and Method, pp. 213–235 Blackwell Publishing, Malden, MA (2004)
2. Leitner, H.: The Politics of Scale and Networks of Spatial Connectivity: Transnational Interurban networks and the Rescaling of Political Governance in Europe. In: Sheppard, E., McMaster, B. (eds.) Scale and Geographic Inquiry: Nature Society and Method, pp. 236–255 Blackwell Publishing, Malden, MA (2004)
3. Hartshorne, R.: The Nature of Geography: A Critical Survey of current thought in the light of the past. PA: Association of American Geographers, Lancaster (1939)
4. Dykes, J.: Exploring Spatial Data Representation with Dynamic Graphics, Computers and Geosciences. **23** 345–370 (1996)
5. Fisher, P., Dykes, J., Wood, J.: Map Design and Visualization. The Cartographic Journal. **30** 136–142 (1993)

6. Monmonier, M.: How to Lie with Maps. University of Chicago Press, Chicago (1991)
7. Brassel, K.E., Weibel, R.: A review and conceptual framework of automated map generalization. International Journal of Geographical Information Systems. **2** 229–244 (1988)
8. Molenaar, M.: An Introduction to the Theory of Spatial Object Modeling for GIS. Taylor and Francis, London (1998)
9. Peng, W., Sijmons, K., Brown, A.: Voronoi Diagram and Delaunay Triangulation Supporting Automated Generalization. Paper presented at 17th ICA/ACI, Barcelona, Spain, 1995
10. Mackaness, W.A., Beard, M.K.: Use of Graph Theory to Support Map Generalization. Cartography and Geographic Information Systems. **20** 210–221 (1993)
11. Buttenfield, B.P., McMaster, R.B.: Map Generalization: Making Rules for Knowledge Representation. Longman, London (1991)
12. Lamy, S., Ruas, A., Demazeau, Y., Jackson, M., Mackaness, W.A., Weibel, R.: The Application of Agents in Automated Map Generalization. Paper presented at Proceedings of the 19th International Cartographic Conference, Ottawa 1999
13. Lecordix, F., Lemarie, C.: Managing Generalization Updates in IGN Map Production. In: Mackaness, W.A., Ruas, A., Sarjakoski, L.T. (eds.) Generalization of Geographic Information: Cartographic Modeling and Applications. Elsevier, Oxford (2007)
14. Elias, B., Brenner, C.: Automatic generation and application of landmarks in navigation data sets. Paper presented at Developments in Spatial Data Handling, Proceedings of the 11th International Symposium on Spatial Data Handling, Leicester, UK, 2004
15. Dias, E., Beinat, E., Scholten, H.: Effects of Mobile Information Sharing in Natural Parks. Paper presented at ENVIROINFO 2004, 18th International Conference Informatics for Environmental Protection, CERN, Geneva, Switzerland, 2004
16. Dykes, J., MacEachren, A.M., Kraak, M.-J.: Exploring Geovisualization. Elsevier, London (2004)
17. Barrault, M., Regnauld, N., Duchene, C., Haire, K., Baeijs, C., Demazeau, Y., Hardy, P., Mackaness, W., Ruas, A., Weibel, R.: Integrating multi agent, object oriented and algorithmic techniques for improved automated map generalization. In: Proceedings of the 20th International Cartographic Conference, **3**, 2110–2116, Beijing, China (2001)
18. Devogele, T., Trevisan, J., Ranal, L.: Building a Multi Scale Database with Scale Transition Relationships. In: Kraak, M.-J., Molenaar, M., Fendel, E.M. (eds.) Advances in GIS Research II, Proceedings of the 7th International Symposium on Spatial Data Handling, pp. 337–352. Taylor and Francis, Delft, London, (1996)
19. João, E.M.: Causes and Consequences of Map Generalization. Taylor and Francis Ltd., London (1998)
20. Kilpelainen, T., Sajakoski, T.: Incremental generalization for multiple representations of geographical objects. In: Muller, J.C., Lagrange, J.P., Weibel, R. (eds.) GIS and Generalization: Methodology and Practice, pp. 209–218. Taylor & Francis, London (1995)
21. Mackaness, W.A., Ruas, A., Sarjakoski, L.T.: Generalization of Geographic Information: Cartographic Modeling and Applications. Elsevier, Oxford (2007)
22. McMaster, B.R., Shea, K.S.: Generalization in Digital Cartography. Association of American Geographers, Washington, D.C. (1992)
23. Müller, J.C., Lagrange, J.P., Weibel, R.: GIS and Generalization: Methodology and Practice. In: Masser, I., Salge, F. (eds.) GISDATA 1. Taylor and Francis, London (1995)
24. Sheppard, E., McMaster, R.B.: Scale and Geographic Inquiry: Nature Society and Method. Blackwell Publishing, Malden, MA (2004)
25. Weibel, R., Dutton, G.: Generalizing Spatial Data and Dealing with Multiple Representations. In: Longley, P., Goodchild, M.F., Maguire, D.J., Rhind, D.W. (eds.) Geographical Information Systems, pp. 125–156. John Wiley, New York (1999)

Generalization, On-the-Fly

ROBERT WEIBEL, DIRK BURGHARDT
Department of Geography, University of Zurich,
Zurich, Switzerland

Synonyms

Real-time generalization; Dynamic generalization; Online generalization; Hierarchical data structures

Definition

Map generalization defines the process of producing maps at coarser levels of detail (LOD), while retaining essential characteristics of the underlying geographic information. On-the-fly generalization, then, denotes the use of automated generalization techniques in real time. According to [1], this process creates a temporary, generalized dataset exclusively for visualization, not for storage or other purposes. On-the-fly generalization is intimately linked to highly interactive applications of cartography such as web mapping, mobile mapping [e. g., in location-based services (LBS)], and real-time decision support systems (e. g., in disaster and evacuation management) that involve multiple spatial scales. As it takes place in a highly interactive setting, the cartographic quality requirements are typically relaxed compared to traditional, high-quality paper maps. On the other hand, (near) real-time behavior is imperative.

Solutions that satisfy the above requirements can generally be assigned to two groups. The first group of approaches relies on fast map generalization algorithms that generate coarser levels of detail in real time. The second group utilizes hierarchical spatial data structures. In both cases, the generalization operations that are implemented are generally rather simple from a functional point of view, compared to the cartographically more sophisticated, yet computationally more costly algorithms that are typically used in the production of high-quality paper maps. Closely related to on-the-fly map generalization is progressive vector data transmission (i. e., the transmission, over a network, of vector datasets at progressively finer detail).

Historical Background

For centuries, cartography was exclusively devoted to the production of static paper maps. Even with the introduction of the computer to cartography in the 1960s and the growing use of interactive computer systems in the 1980s the situation did not change much. Paper was still the main output medium, and screen maps were commonly only used for editing and proofing in paper map production, rather than as end products. Consequently, research in automated map generalization—despite the fact that it dates back to the early days of computer cartography and geographical information systems (GIS)—focused primarily on achieving high cartographic quality in the generalization process, while largely neglecting computational efficiency. While this preference for graphic quality over efficiency may sound odd to the computer scientist, it did make sense from a cartographic perspective, bearing in mind that firstly, map generalization is an ill-defined process and nontrivial to automate [2], and secondly, since the end products were static, improved cartographic quality at the expense of added computing time could easily be tolerated.

The advent of interactive personal computers in the 1980s and, more importantly, of the world wide web (WWW) in the early 1990s brought new requirements for cartography and map generalization. Since the usage and interaction with web mapping services are highly time-critical, scale-changing has to take place in real time, essentially demanding on-the-fly generalization. However, despite early solutions for on-the-fly line simplification and object selection using reactive data structures [3,4], researchers in automated map generalization continued to place little emphasis on on-the-fly generalization throughout the 1990s. Operational web map services such as mapquest.com, mapblast.com, or map24.com rely on pragmatic solutions, involving offline production of multirepresentation databases containing multiple LOD, restricting the on-the-fly part to the real-time retrieval and display of the LOD that matches the given zoom level as well as real-time placement of map labels and symbols (e. g., icons for points-of-interest).

In recent years, however, repeated calls have been made for a redefinition of cartography and map generalization [5,6]. New forms of mapping applications such as mobile mapping in LBS or real-time decision support systems go beyond web mapping as it's best known, in the form of the services mentioned above. In addition to the requirement of real-time map delivery, these new mapping applications demand adaptation and personalization of the thematic map content to the given user query and context. Thus, precomputing and storing potential visualizations as in "classical" web mapping is no longer a feasible solution; true on-the-fly generalization capabilities are needed. Researchers have started to respond to these new challenges in recent years.

Scientific Fundamentals

On-the-Fly Generalization Versus Multirepresentation Databases

As mentioned in the preceding section, true on-the-fly generalization is not to be equated with the mere retrieval and display of pregeneralized LOD from a multirepresentation database (MRDB). Hence, there is no need to dwell on MRDB further. However, it should be emphasized that MRDBs remain an active and important research area [7,8]. For instance, many public and private mapping organizations have large holdings of digitized maps at different scales and thus are interested in linking these together so that updates can be propagated automatically from detailed to less detailed representations, allowing incremental updates [9].

Characteristics

The main characteristics of on-the-fly generalization are (see also [10]):

- A temporary, reduced scale (generalized) dataset/map is generated for visualization purposes from a spatial database.
- The map has to meet the user preferences (e. g., personalized content) and the technical display specifications (i. e., typically low screen resolution and small screen size).
- The scale of the resulting map may vary (particularly due to zooming operations) and is not predefined.
- The generalization process must be accomplished automatically and no user interaction is possible, e. g., to check the result before publishing.
- The resulting map must appear on the display within a few seconds, as the user does not want to wait.
- On the web and on mobile devices, there is an additional problem of limited network bandwidth.

Techniques for on-the-fly generalization follow two main tracks, either making use of efficient algorithms that allow generation of coarser LOD in real time, or exploiting hierarchical spatial data structures.

On-the-Fly Generalization by Algorithms

Since on-the-fly generalization is a time-critical task, generalization algorithms that are suited for this purpose must be fast and/or they must be supported by precomputed data structures or attributes. In principle, all known generalization algorithms that run in linear or logarithmic time make candidates for on-the-fly generalization. One example of such fast algorithms is simple selection algorithms that merely rely on precomputed attributes, such as the Horton stream ordering scheme used for river network selection [10]. Attribute-based selection is also straightforward to implement, as exemplified by the system described in [11] that uses the extensible stylesheet language transformation (XSLT) mechanism to generate real-time, generalized maps. An extended version of this system [12] offers a range of well-known algorithms: feature selection by object class, area selection by minimum/maximum value, line selection by minimum/maximum length, contour line selection by interval, line simplification by the Douglas–Peucker algorithm, line simplification by the Lang algorithm, line smoothing by Gaussian filtering, and building outline simplification. Another system for on-the-fly generalization that makes use of a combination of simple (and efficient) algorithms is described in [13].

The algorithms discussed so far have in common that they were originally not specifically developed for on-the-fly generalization of potentially large datasets. They are merely useful for this purpose because they are so simple that they require relatively little computational effort. An algorithm that specifically targets dynamic generalization is presented in [14]. This algorithm performs line simplification of large map datasets through a novel use of graphics hardware (frame buffer, color buffer, stencil buffer, depth buffer) using a hybrid vector/raster-based approach. For interactive visualization, presimplified maps of different LOD are organized in a hierarchical data structure, the Data Tree. the solution presented in [14] thus represents a hybrid between algorithmic approaches to on-the-fly generalization and those methods that fully rely on hierarchical data structures (to be discussed next section).

The above examples implement only algorithms that are restricted to rather simple generalization operations without consideration of their spatial context, such as selection, line simplification, line smoothing and polygon aggregation (e. g., by the convex hull). More complex, contextual generalization operations such as feature displacement or typification—necessary to achieve high cartographic quality—commonly require iterative optimization techniques that are generally not suited to real-time applications (for an overview, see [15]). A possible solution to speed up displacement computation consists in using interpolation, or "morphing", between the geometries of two LOD [16]. This approach, however, requires at least two LOD whose vertices of corresponding map objects have been correctly matched and linked in an MRDB. A more realistic approach to achieving more complex generalization behavior that is nevertheless efficient is by using auxiliary data structures, as discussed in the next section.

On-the-Fly Generalization by Hierarchical Data Structures

Map generalization results in hierarchies of progressively coarser maps. Thus, it is only natural that hierarchical spatial data structures are exploited in map generalization, and even more prominently in speeding up on-the-fly map generalization. This section discusses selected examples of solutions that rely exclusively on the hierarchical representation of spatial data in tree data structures. These examples have in common that they try to establish variable scale data structures, thus avoiding the redundant data storage typical of multiple representations using a stack of LOD.

The earliest proposal of a tree data structure for on-the-fly generalization was already mentioned in the historical overview: the Binary Line Generalization (BLG) Tree [3]. It uses the classic of line simplification, the Douglas–Peucker algorithm, to precompute the order of elimination of the vertices of a line. The vertex numbers and associated tolerance values are then stored in a binary tree. At run time, the tree can be descended down to the level that matches the resolution of the target map and the corresponding vertices retrieved for rendering. As the BLG tree is restricted to organizing single line objects, it cannot be used for the spatial organization (e. g., indexing) of multiple map objects. This restriction is overcome by the Reactive Tree [4], an extension to the R-tree [17] that stores importance levels for map objects (with important objects stored higher in the tree). The Reactive Tree is dynamic, allowing inserts and deletes.

The BLG and Reactive Tree data structures are not suited to the generalization of polygonal maps [18], as they do not represent the special nature of an area partitioning of adjacent polygons. This deficiency led to the development of the Generalized Area Partitioning (GAP) Tree which defines successive hierarchies of aggregations of adjacent polygons in a polygonal map. A system which uses the BLG Tree (for line simplification), the reactive Tree (for map object selection), and the GAP Tree (for area aggregation) together is reported in [1], containing also a description of the GAP Tree data structure. Recently, a new, topological version of the GAP Tree was introduced [18] which combines the use of the BLG Tree and the Reactive Tree and avoids redundant storage and sliver polygons along the boundary of neighboring polygons, problems associated with the original GAP Tree.

The use of hierarchical data structures for on-the-fly generalization of point distributions commonly found in thematic maps (e. g., animal observation data, distributions of disease occurrences) and LBS (e. g., points-of-interest) is reported in [19,20]. Two methods are proposed. The first one uses a quadtree to index the original points to successively coarser, aggregated levels. At run time the original points are then replaced by the centroids of the quadtree cells corresponding to the appropriate resolution (i. e., scale of the target map). The disadvantage of this first solution is that the output point pattern will be aligned to the (regular) quadtree pattern, creating an unnatural arrangement. Hence, a second proposed solution uses a hierarchical tessellation of the map space that corresponds to the semantics of the data points. In the example shown in [19,20], animal observation data are mapped to the network hierarchy of drainage basins, as these are bounded by ridges that often also form physical barriers to animal movement.

Related Issues

In recent years, research interest has started to develop into methods for the progressive transmission of vector map data. This interest is motivated by the very same reason that prompted the earlier development of progressive techniques for the transmission of raster images over the WWW, implemented today in standard image formats: the need to access large datasets in distributed, bandwidth-limited computing environments. Progressive vector data transmission shares with on-the-fly generalization the aim to represent map data at successively coarser or finer LOD, respectively. While the aim of progressive vector data transmission is to ultimately transmit the *entire* dataset, the user will initially only receive a coarse representation of the map data, followed by progressive refinements, until the full map has been transmitted. Any of the intermediate refinement steps represents a generalization of the full map. Hence, there is also a strong similarity (or even congruence) of methods between progressive vector data transmission and on-the-fly generalization. In comparison to progressive methods for image data, equivalent methods for vector data are inherently more complex to achieve and hence still very much in the research stage. Starting from initial conceptual work [21], solutions have been proposed for the "continuous generalization" of buildings for LBS [22], for an MRDB architecture in the context of LBS applications [23], and for line and polygon data [24].

Label and icon placement on screen maps is a further issue that is closely related to online-generalization, for two reasons. First, the selection of map labels and/or map icons is driven by the same principles—scale, semantics, available map space—as the selection of other map objects. Second, the placement of map labels and map icons shares many similarities with displacement operations in map generalization. While many algorithms exist for offline placement of map labels and icons, real-time labeling (e. g., for mobile maps in LBS) has only rarely been addressed in the literature so far [25].

Finally, web generalization services should be mentioned. On-the-fly generalization is typically linked to web and/or mobile mapping applications, hence to applications that take place in distributed computing environments and client/server architectures. Therefore, the recently initiated move toward the exploitation of service-based architectures in map generalization [26,27] nicely corresponds to the evolution of on-the-fly generalization.

Key Applications

As has become obvious from the preceding discussion, on-the-fly map generalization is still very much a research area. The development of appropriate techniques targets a variety of applications which have in common that they are highly interactive and have requirements for (near) real-time visualization with adaptable scale and content. Following are a few examples of such applications.

Web Mapping

As mentioned in the Historical Background section the evolution of web mapping has provided the initial setting that prompted the need for on-the-fly generalization capabilities. For many years, web mapping largely defined the requirements for on-the-fly generalization. Today, however, it has been superseded as a trendsetter by less mainstream applications.

Adaptive Zooming

Adaptive zooming is a capability that is still sorely lacking in many interactive mapping systems. It denotes "the adjustment of a map, its contents and the symbolization to target scale in consequence of a zooming operation" [28]. As follows from this definition, adaptive zooming also requires some sort of on-the-fly generalization. A pragmatic solution that uses on-the-fly algorithms for the simple generalization operations in combination with LOD as substitutes for complex generalization operations is presented in [28].

Mobile Cartography and LBS

LBS have given a new direction to cartography and GIS. They place the user in the center of the map; the map display needs to adapt to the user's changing location; and the map display needs to be adapted (or personalized) to the user's specific information requirements. Furthermore, mobile devices are bound to impose more stringent technical limitations than commonly encountered in cartography (e. g., low resolution and small size of the display screen, low bandwidth, unreliable network connectivity). An overview discussion of the requirements and research perspectives of LBS, including the need for on-the-fly generalization, can be found in [29].

Real-Time Decision Support Systems

GIS are used a great deal as tools for decision support. While most uses of spatial decision support systems (SDSS) do not have real-time requirements, new applications have recently started to appear that do involve decision making in response to real-time data feeds. Examples include emergency service dispatching, evacuation route planning, and disaster management [30].

Future Directions

As [18] notes, data structures supporting variable scale datasets—and hence also solutions for on-the-fly map generalization—are still very rare. On the other hand, there are a growing number of applications that require functionality for real-time adaptation of spatial datasets and maps to the scale and purpose of the target display. Hence, it can be expected that increasingly more sophisticated solutions will complement or supersede the rather pragmatic, usually LOD-based techniques commonly used today. In addition to the development of new techniques, there is also room for improvement by *combining* existing methods. First, individual real-time algorithms may be combined to create more comprehensive solutions, as exemplified by [18]. In the future, this approach may also benefit from the current trend towards web-based architectures [26]. A second track may exploit the potential of combining MRDB- and LOD-based techniques and on-the-fly generalization, illustrated by the (still rather pragmatic) solution presented in [28].

Cross References

► Indoor Positioning

Recommended Reading

1. van Oosterom, P., Schenkelaars, V.: The development of an interactive multiscale GIS. Int. J. Geogr. Inf. Syst. **9**, 489–507 (1995)
2. Weibel, R., Dutton, G.: Generalizing spatial data and dealing with multiple representations. In: Longley, P.A., Goodchild, M.F., Maguire, D.J.,Rhind, D.W. (eds.) Geographical Information Systems: Principles, Techniques, Management and Applications, 2nd edn., pp. 125–155. Wiley, Chichester (1999)
3. van Oosterom, P., van den Bos, J.: An object-oriented approach to the design of geographic information systems. Comput. Graphics **13**, 409–418 (1989)
4. van Oosterom, P.: A storage structure for a multiscale database: the reactive-tree. Comput., Environ. Urban Syst. **16**, 239–47 (1992)
5. Jones, C.B., Ware, J.M.: Map generalization in the web age. Int. J. Geogr. Inf. Sci. **19**, 859–870 (2005)
6. Mackaness, W.A.: Automated cartography in a bush of ghosts. Cartogr. Geogr. Inf. Sci. **33**, 245–256 (2006)
7. Hampe, M., Sester, M.: Generating and using a multi-representation database (MRDB) for mobile applications. Paper presented at the ICA Workshop on Generalization and Multiple Representation, Leicester, 20–21 August 2004. Available via: http://ica.ign.fr/Leicester/paper/hampe-v2-ICAWorkshop.pdf. Accessed 12 Feb 2007
8. Viaña, R., Magillo, P., Puppo, E., P.A. Ramos: Multi-VMap: a multi-scale model for vector maps. GeoInformatica, **10**, 359–394 (2006)
9. Anders, K.-H., Bobrich, J.: MRDB Approach for Automatic Incremental Update. Paper presented at the ICA Workshop on Generalization and Multiple Representation, Leicester, 20–21 August 2004. Available via: http://ica.ign.fr/Leicester/paper/Anders-v1-ICAWorkshop.pdf. Accessed 12 Feb 2007
10. Rusak Masur, E., Castner, H.W.: Horton's ordering scheme and the generalisation of river networks. Cartogr. J. **27**, 104–112 (1992)
11. Lehto, L., Sarjakoski, L.T.: Real-time generalization of XML-encoded spatial data for the web and mobile devices. Int. J. Geogr. Inf. Sci. **19**, 957–973 (2005)
12. Sarjakoski, L.T., Sarjakoski, T.: A use case based mobile GI service with embedded map generalisation. Paper presented at the ICA Workshop on Generalization and Multiple Representation, Leicester, 20–21 August 2004. Available via: http://ica.ign.fr/Leicester/paper/Sarjakoski-v2-ICAWorkshop.pdf. Accessed 12 Feb 2007
13. Glover, E., Mackaness, W.A.: Dynamic generalisation from a single detailed database to support web based interaction. In: Proceedings 19th International Cartographic Conference, Ottawa, 14–21 August (1999)
14. Mustafa, N., Krishnan, S., Varadhan, G., Venkatasubramanian, S.: Dynamic simplification and visualization of large maps. Int. J. Geogr. Inf. Sci. **20**, 273–302 (2006)
15. Mackaness, W.A., Ruas, A., Sarjakoski, L.T. (eds.): Generalisation of Geographic Information: Cartographic Modelling and Applications. Elsevier, Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo (2007)
16. Sederberg, T.W., Greenwood, E.: A physically based approach to 2-D shape blending. Comput. Graphics **26**, 25–34 (1992)
17. Guttmann, A.: R-Trees: A dynamic index structure for spatial searching. ACM SIGMOD 13, 47–57 (1984)
18. van Oosterom, P.: Variable-scale topological data structures suitable for progressive data transfer: The GAP-face tree and GAP-edge forest. Cartogr. Geogr. Inf. Sci. **32**, 331–346 (2005)
19. Burghardt, D., Purves, R. S., Edwardes, A.J.: Techniques for on-the-fly generalisation of thematic point data using hierarchical data structures. In: Proceedings of the GIS Research UK 12th Annual Conference, Norwich, 28–30 April (2004)
20. Edwardes, A., Burghardt, D., Dias, E., Purves, R.S., Weibel, R.: Geo-enabling spatially relevant data for mobile information use and visualisation. In: Proceedings of the 5th International Workshop on Web and Wireless GIS, W2GIS 2005. Lecture Notes in Computer Science, vol. 3833, pp. 78–92. Springer-Verlag, Berlin, Heidelberg (2005)

21. Bertolotto, M., Egenhofer, M.J.: Progressive transmission of vector map data over the world wide web. GeoInformatica, **5**, 345–373 (2001)
22. Sester, M., Brenner, C.: Continuous generalization for visualization on small mobile devices. In: Peter Fisher (ed.) Developments in Spatial Data Handling–11th International Symposium on Spatial Data Handling, pp. 469–480. Springer Verlag, Berlin, Heidelberg, New York (2004)
23. Follin, J.-M., Bouju, A., Bertrand, F., Boursier, P.: Multiresolution extension for transmission of geodata in a mobile context. Comput. Geosci. **31**, 179–188 (2005)
24. Yang, B., Purves, R., Weibel, R.: Efficient transmission of vector data over the internet. Int. J. Geogr. Inf. Sci. **21**, 215–237 (2007)
25. Harrie, L, Stigmar, H., Koivula, T., Lehto, L.: An algorithm for icon labelling on a real-time map. In: Fisher, P. (ed.) Developments in Spatial Data Handling, pp. 493–507. Springer-Verlag, Berlin, Heidelberg, New York (2004)
26. Burghardt, D., Neun, M., Weibel, R.: Generalization services on the web—classification and an initial prototype implementation. Cartogr. Geogr. Inf. Sci. **32**, 257–268 (2005)
27. Foerster, T., Stoter, J.: Establishing an OGC web processing service for generalization processes. Paper presented at the ICA Workshop on Generalization and Multiple Representation, Portland, 25 June 2006. Available via: http://ica.ign.fr/Portland/paper/ICA2006-foerster_stoter.pdf Accessed 12 Feb 2007
28. Cecconi, A., Galanda, M.: Adaptive zooming in web cartography. Comput. Graphics Forum **21**, 787–799 (2002)
29. Jiang, B., Yao, X.: Location-based services and GIS in perspective. Comput., Environ. Urban Syst. **30**, 712–725 (2006)
30. Zerger, A., Smith, D.I.: Impediments to using GIS for real-time disaster decision support. Comput., Environ. Urban Syst. **27**, 123–141. (2003)

Generalized Minimum Spanning Tree

► Trip Planning Queries in Road Network Databases

Generalizing

► Hierarchies and Level of Detail

Genome Mapping

► Bioinformatics, Spatial Aspects

Geocollaboration

Shivanand Balram, Suzana Dragićević
Department of Geography, Simon Fraser University,
Burnaby, BC, Canada

Synonyms

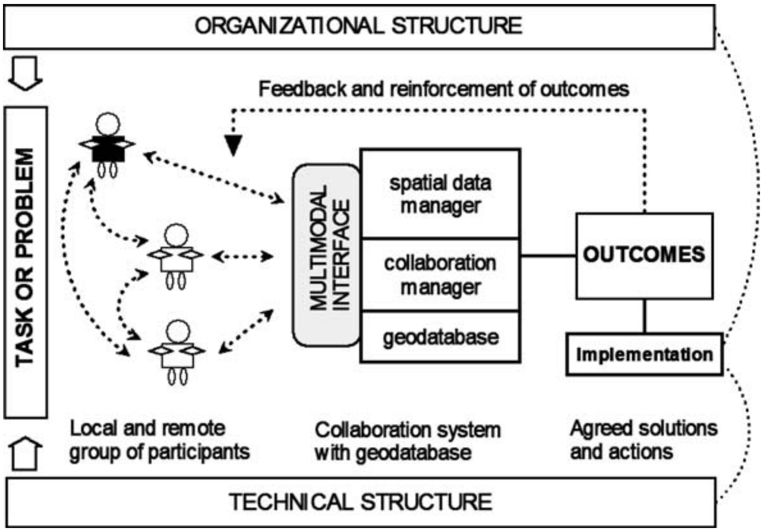
Collaborative geographic information systems; CGIS; Computer supported cooperative work; CSCW; Group spatial decision support systems; GSDSS; PSS; Planning support systems

Definition

Geocollaboration is an emerging area of study examining how spatial information and communication technologies can be designed and adapted to support group interactions that use geographically-referenced data and information [1]. These group interactions normally focus on tasks such as spatial data access and exploration, problem-solving, planning, and decision-making. In a recent classification of knowledge areas within geographic information science, geocollaboration has been interpreted as a specific implementation of group spatial decision support systems (GSDSS) which in turn forms a component of GIS and Society research [2]. In order to support collaborative interactions, group participants need to be able to browse, explore and query spatial data and information. Further, participants must be able to represent knowledge and communicate with each other towards achieving well defined objectives or goals. Hence, key issues of geocollaboration include group support methods such as real time conferencing and sketch mapping, distributed computing using the Web and local area networks, and information communication with maps and scientific visualizations tools [1,3]. Figure 1 shows a general system architecture for geocollaborative interactions consisting of: a group of participants arranged in various configurations of place and time; a computer system for handling geospatial data and group interactions; technical expertise to integrate the system; and organizational expertise to focus the goals of the collaborative process for appropriate implementation.

Historical Background

Geocollaboration as a concept was formally proposed around the year 2000 by researchers from Penn State University (USA) as a focused response to the need for designing and adapting geospatial technologies for supporting group interactions [4]. Geocollaboration represents an important confluence of methods and tools. The need for group based spatial technologies was formally recognized and systematically discussed during the September 2005 Specialist Meeting of the US National Center of Geographic Information Analysis (NCGIA) research initiative on collaborative spatial decision making (CSDM). The CSDM initiative (I-17) investigated the design of interactive group environments for spatial decision making. The research that followed in this area focused on specific issues including structuring the collaborative process [5], embedding spatial data directly into group discussions [6], analyzing map-based group data [7], ensuring



Geocollaboration, Figure 1 General system architecture for geocollaborative interactions

democratic levels of group participation [8], and multiple visual data representation for improved understanding [9]. Efforts on specific aspects of spatial collaboration resulted in many flavors of collaborative system designs such as Spatial Understanding Support Systems, Planning Support Systems, and Collaborative Geographic Information Systems [2]. These systems have their foundations in mainly decision analysis theory, group structuration theory, and geographic information systems and science. Geocollaboration, however, has its foundations in geographic information science, computer supported cooperative work, and distributed computing. With this foundation, geocollaboration seeks to address the impact of technological, social and cognitive factors on group based interactions with geospatial data [1].

Scientific Fundamentals

Figure 1 shows a typical architecture for geocollaborative interactions. Before the system is constructed, well defined tasks or problems to be solved are usually specified by participatory groups and decision makers. A choice amongst the four combinations of same place or time and different place or time designs determines if the collaboration infrastructure will incorporate the Web as a distributed communication medium. Structuring the collaborative interactions follows, and is influenced by the collaborative process focus on either knowledge construction, problem solving, task implementation, data exploration, spatial analysis, decision making, or training. Many structuring approaches are available including shared workspace, argumentation mapping, spatial Delphi, real time conferencing and sketch maps. User interfaces and their designs can vary from a range that includes simple map sketching with digital pens and sophisticated three-dimensional head mounted displays to immersive virtual environments. The capabilities of user interfaces to aid visualization and natural data interactions are constantly being enhanced. These user interfaces connect directly to a distributed or localized computer system that manages the collaborative interactions and geospatial data storage. There are usually three components: a spatial data manager to facilitate data requests and transactions, a collaboration manager to track and monitor the interactions among the group of participants, and a geospatial database to provide spatial data and information as per the request of the spatial data manager component. The collaborating participants eventually generate spatial outputs in the form of maps, scientific visualization products, or geo-referenced data attributes in response to the specified system tasks or problem definition. Iterative analysis of the outputs can generate more robust final outcomes. The geocollaboration system must be embedded in an organizational and technical structure for continued development and support of the local and remote collaborations [6]. Interface design (user-friendly interactions), social dynamics (level of social representation), distributed GIS (equal participation opportunities), and cartography (representation and understanding of reality) principles all have an influence on geocollaboration.

Key Applications

Principles of geocollaboration are applied in many diverse knowledge domains where there is a need to integrate people, technology and data in a spatial context. These applications cover a diverse range from the environmental sciences to engineering, but are more dominant in geography where geospatial data are widely collected and analyzed.

Geography

In geography, geocollaboration principles focus mainly on the people component of the people-technology-data integration. The development of group support methods are of specific interest. Applications for public participation, transportation, and health are outlined.

Public Participation Geocollaboration principles can be applied to designing systems that improve the public's access to information and their contributions to planning and decision making forums. These public interactions usually occur at the same-time/same-place or same-time/different-place [10].

Transportation Transportation planners can use geocollaboration principles to develop systems that reconcile multiple interests and organizational goals. Candidate sites for transportation development can be identified and assessed more effectively for more appropriate spending of public funds [11].

Health In health care and disease management, geocollaboration designs can be used to develop highly interactive systems that allow the public to identify and locate disease incidences thereby allowing more targeted responses from health care professionals.

Engineering

In engineering, geocollaboration principles focus mainly on the technology component of the people-technology-data integration. The development of more user friendly spatial technologies is of particular interest. Applications in disaster management, user interfaces, and distributed databases are outlined.

Disaster Management During times of disaster, the close coordination of resources from multiple organizations is necessary for mitigating and managing the crisis situation. Geocollaboration principles allow for the design and development of systems that can integrate managers, task groups, data resources and collaborative interactions for real-time planning and coordination among teams [12].

User Interfaces Geocollaboration principles are used in the design of natural user interfaces for more embedded interactions and manipulation of geospatial data [13].

Distributed Databases Geocollaboration designs can be used to develop highly accessible and interactive knowledge systems that allow the non-technical individuals to input data into existing spatial databases to improve understanding and awareness of existing conditions [14].

Environmental Sciences

In the environmental sciences, geocollaboration principles focus mainly on the spatial data component of the people-technology-data integration. Developing more accessible spatial databases and providing for greater collaborative data analysis and modeling are of specific interest. Applications in land use, water resources, and forestry are outlined.

Land Use Geocollaboration principles can be used in exploring scenarios and alternative futures for land use planning and analysis [15]. Multiple participants interact with each other and geospatial data towards defining models of reality that best capture common interests, concerns, and goals.

Water Resources Water resource managers can use geocollaboration principles to develop systems that reconcile multiple interests and organizational goals. The influence of technology designs and participation perceptions on the desirable outcomes can be examined experimentally [16].

Forestry The development of participatory model-based forestry tools can benefit from geocollaboration principles for improving model input interfaces and in the visualization of output scenarios.

Future Directions

Any problem or task that requires the integration of people, technology, and geospatial data can benefit from geocollaboration principles in the design of solutions. Applications in national security and human health are new and emerging areas of interest for geocollaboration principles. Collaborative interaction systems with virtual environments can allow military strategists to better simulate command and control operations for training and emergency purposes. Medical personnel can be briefed and deployed in emergencies using distributed collaborative systems that link multiple authorities into a centralized decision-making structure.

On the research frontier, geocollaboration will need to further address key issues including:

- Integrating real-time spatial data from multiple sources and with multiple errors
- Using the Internet as a robust medium to support collaborative communication
- Developing open source toolkits to encourage geocollaboration system designs
- Designing more effective multi-modal interfaces to improve end-user experiences

The application potential and research needs of geocollaboration make it an exciting and rewarding area for study and future enhancements.

Cross References

- ▶ Decision-Making Effectiveness with GIS
- ▶ Environmental Planning and Simulation Tools
- ▶ Multicriteria Decision Making, Spatial
- ▶ Web Mapping and Web Cartography

Recommended Reading

1. MacEachren, A.M., Brewer, I.: Developing a conceptual framework for visually-enabled geocollaboration. Int. J. Geogr. Inf. Sci. **18**, 1–34 (2004)
2. Balram, S., Dragicevic, S.: Collaborative Geographic Information Systems: Origins, Boundaries, and Structures. In: Balram, S., Dragicevic, S. (eds.) Collaborative Geographic Information Systems, pp. 1–22. Idea Group Publishers, Hershey, PA (2006)
3. Schafer, W.A., Ganoe, C.H., Xiao, L., Coch, G., Carroll, J.M.: Designing the next generation of distributed, geocollaborative tools. Cartography Geogr. Inf. Sci. **32**, 81–100 (2005)
4. MacEachren, A., Brewer, I., Cai, G., Chen, C.: Visually enabled Geocollaboration to support data exploration and decision-making. In: Proceedings of the 21st International Cartographic Conference, 10–16 August: Durban, South Africa (2003)
5. Jankowski, P., Nyerges, T.: Geographic information systems for group decision making: Towards a participatory geographic information science. Taylor and Francis, New York (2001)
6. Dragicevic, S., Balram, S.: A Web GIS collaborative framework to structure and manage distributed planning processes. J. Geogr. Syst. **6**, 133–153 (2004)
7. Jankowski, P., Andrienko, N., Andrienko, G.: Map-centred exploratory approach to multiple criteria spatial decision making. Int. J. Geogr. Inf. Sci. **15**, 101–127 (2001)
8. Ghose, R.: Use of information technology for community empowerment: Transforming geographic information systems into community information systems. Trans. GIS **5**, 141–163 (2001)
9. Shiffer, M.J.: Multimedia GIS for planning support and public disclosure. Cartography Geogr. Inf. Syst. **25**, 89–94 (1998)
10. Kingston, R., Turton, I., Carver, S., Evans, A.: Web-based public participation geographical information systems: An aid to local environmental decision-making. Comput. Environ. Urban Syst. **24**, 109–125 (2000)
11. Nyerges, T.L., Montejano, R., Oshiro, C., Dadswell, M.: Group-based geographic information systems for transportation improvement site selection. Transp. Res. Part C-Emerging Technol. **5**, 349–369 (1997)
12. Cai, G.R., MacEachren, A.M., Brewer, I., McNeese, M., Sharma, R., Fuhrmann, S.: Map-mediated GeoCollaborative Crisis Management. In: Intelligence and Security Informatics, Proceedings, vol. 3495. pp. 429–435 (2005)
13. MacEachren, A.M., Cai, G., Brewer, I., Chen, J.: Supporting map-based geocollaboration through natural interfaces to large-screen displays. Cartographic Persp. **54**, 16–34 (2006)
14. Mason, B., Dragicevic, S.: Web GIS and knowledge management systems: An integrated design for collaborative community planning. In: Balram, S., Dragicevic, S. (eds.) Collaborative Geographic Information Systems, pp. 263–284. Idea Group Inc. (2006)
15. Klosterman, R.E.: Computer applications in planning. Environ. Plann. B: Plann. Des. **25**, 32–36 (1998)
16. Nyerges, T., Jankowski, P., Tuthill, D., Ramsey, K.: Collaborative water resource decision support: Results of a field experiment. Ann. Association Am. Geogr. **96**, 699–725 (2006)

GeoDa

▶ Data Analysis, Spatial

Geodemographic Segmentation

AUSTIN TROY
Rubenstein School of Environment and Natural Resources, University of Vermont, Burlington, VT, USA

Synonyms

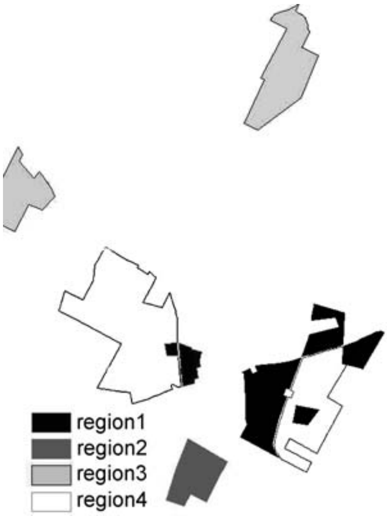
Neighborhood segmentation; Spatially agglomerative clustering; Geographic market segmentation; Regionalization, spatial ontologies; Cluster analysis; Partitioning; Regression, geographically weighted

Definition

Geodemographic segmentation refers to a range of methods used for classifying and characterizing neighborhoods or localities based on the principal that residents living near each other are likely to have similar demographic, socio-economic and lifestyle characteristics. It is used for a wide range of purposes including direct marketing, retail location, service area analysis, housing market analysis, and public service targeting.

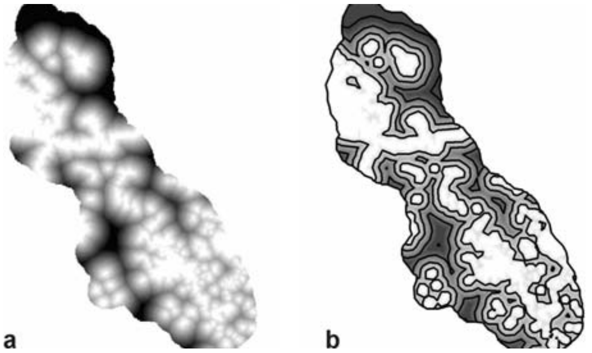
Many methodological approaches exist for geodemographic segmentation. Some form of cluster analysis is generally used to assign entities (e. g. zip codes, property parcels, neighborhoods) to classes which can eventually form larger objects, or segments. Entities are classified so as to minimize heterogeneity within groups relative to between groups. The classification taxonomy can incorporate any number of classes and can also make use of nested or non-nested hierarchies (see ▶ Hierarchical Spatial Models, this volume). The output is the class membership for each constituent object. While cluster analysis is generally aspatial (although geographic coordinates can be included as additional dimensions), adjacency or connectivity rules can be applied afterwards specifying when entities of the same class can be merged. The geographic nature of segments depends on the entities they are built from. Where segments are built from spatially exhaustive polygons, the

Geodemographic Segmentation, Figure 1a and b Aggregation of spatially exhaustive subunit polygons into spatially exhaustive segments



Geodemographic Segmentation, Figure 2 Segmentation of "island polygons"

output may be a landscape of larger spatially exhaustive polygons (Fig. 1a and b). Where they are built from spatially isolated (i. e. "island") polygons, the output may be new multi-part regions whose constituents do not necessarily touch (Fig. 2). Point objects or raster surfaces (Fig. 3a and b) may also be segmented.

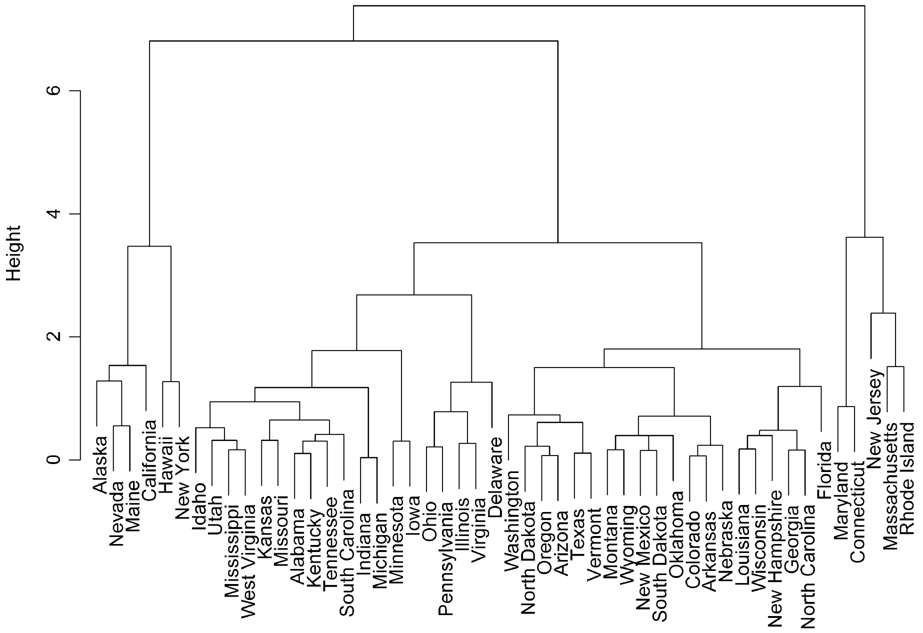


Geodemographic Segmentation, Figure 3 Segementation of continuous fields

Historical Background

The history of cluster analysis dates back to the 1930s, with early works such as R.C. Tryon's *Cluster Analysis* [1], but the greatest advances in the technique's application and development, and much of the literature, comes from the 1960s and after. One of the early landmark texts on cluster methods was *Principles of Numerical Taxonomy*, written by Sokal and Sneath in 1963 [2]. Neighborhood and geodemographic segmentation, dates back even earlier, to the sociologist Charles Booth who, in the early 20th century, developed a multivariate system of classifying neighborhoods to study patterns of poverty in London. In the 1920s sociologists from the "Chicago School," such as Robert E. Park and Ernest Burgess built on this foundation shortly thereafter with a theory of urban "natural areas," which posited that competition for land in cities among social groups led to its division into geographic units similar to ecological niches [3]. Their application of this theory involved considerable pre-digital overlay analysis of maps, including the *Social Science Research Base Map of Chicago* which combined all available spatially referenced data about the city and was used to detect boundaries of the "zones" that their theory predicted. Geodemographic analysis became increasingly sophisticated with the availability of electronic Census data, the enumeration of smaller Census units, such as block groups, and advancement of cluster analysis.

The commercial application of geodemographics has its roots in 1961, when Jonathan Robbin formed the General Analytics Company, a marketing research firm that began researching geodemographic methods. This firm was pur-



Geodemographic Segmentation, Figure 4 An agglomerative hierarchical cluster tree grouping states based on population density and owner occupancy rates

chased and later re-acquired by Robbin in 1971, at which point it became Claritas Inc., the company that created the PRIZM®, the first major geodemographic segmentation, in 1978. It uses Census and other data to classify US Census Block Groups into hierarchically nested segments. This and later products, such as Mosaic by Experian and ACORN™ (A Classification Of Residential Neighbourhoods) by CACI, are used to help target marketing efforts, site retail locations, and generally understand how purchasing behavior varies by neighborhood. Many of the key concepts of geodemographic segmentation in the post-digital age are summarized by Micheal Weiss in a trilogy of books from 1988, 1994 and 2000 [4,5,6] as well as by Debenham et al. [7].

Scientific Fundamentals

Segmentation is fundamentally a classification exercise. In the simplest case, a typology is developed a priori with set criteria for membership in each class. Spatial entities (e. g. zip code, city, house, etc.) are then classified based on meeting these criteria. More commonly, segmentation uses a posterior approach in which classifications are based on cluster analysis or other multivariate methods. Cluster analysis is a data structuring tool that is generally used as an exploratory rather than confirmatory tool. It organizes data into meaningful taxonomies in which groups are relatively homogeneous with respect to a specified set of attributes. That is, it maximizes the association between objects in the same group while minimizing the association between groups. It does so based on the concepts of dissimilarity or distance in n dimensional space, where the axis of each dimension represents some attribute. While the nomenclature of cluster distance sounds spatial, in fact the distances being measured do not represent real space, but rather attribute space.

There are a number of clustering techniques that can be used in geographic segmentation. These can be broadly broken up into partitioning and hierarchical methods. Hierarchical methods work by either dividing (divisive methods) or aggregating (agglomerative methods) groupings of data using a hierarchical structure. Agglomerative clustering starts by treating every data observation as a separate group in itself and then groups those observations into larger groups until there is a single group (Fig. 4). That is, it sequentially lowers the threshold for defining uniqueness, causing progressively fewer observations to be seen as "dissimilar." Dissimilarity or similarity is based on the "distance" between observations, which is represented in the "height" axis in Fig. 4. In multi-dimensional attribute space, distance refers to how dissimilar the attributes of an observation are from another. When attribute variables are in different numeric scales, it is often required to standardize the data so that no one variable is overly weighted. Distance can be measured as Euclidean (straight line) distance (Eq. 1), squared Euclidean distance (Eq. 2), Manhattan (city block) distance (Eq. 3), and Chebychev distance (Eq. 4), among many other approaches.

$$\begin{aligned}
\mathrm{distance}(x, y) &= \Sigma_i (x_i - y_i)^2 && (1)\\
&= \{\Sigma_i (x_i - y_i)^2\}^{1/2} && (2)\\
&= \Sigma_i |x_i - y_i| && (3)\\
&= \mathrm{Max}|x_i - y_i| \,. && (4)
\end{aligned}$$

This approach is useful for *a posteriori* data explorations and allows users to interpret how clusters relate to each other based on the patterns in which they branch. Divisive clustering is the opposite of agglomerative clustering; it starts with all data in a single set, which is successively divided into smaller groups until each group is a single observation.

Partitioning methods are often used when the analyst has some a priori notion about how many or what type of clusters to expect. The clusters can then be analyzed for systematic differences in the distributions of the variables. Partitioning iteratively creates clusters by assigning observations to the cluster centroid that is nearest. The most common partitioning method is k means, popularized by Hartigan [8], which works by randomly generating *k* clusters, determining the location of the cluster center, assigning each point to the nearest cluster center, and then iteratively recomputing the new cluster centers until convergence occurs, which is generally signaled by when point-cluster assignments are no longer changing. Other partitioning approaches use different measures of dissimilarity. For instance, Expectation Maximization (EM) Clustering maximizes the probability of the data given the final clusters by analyzing the data's probability distributions to compute probabilities of cluster memberships. This method is frequently used for market and geodemographic segmentation because it can be applied to both continuous and categorical variables, unlike k means. Yet another partitioning method is partitioning around medoids (PAM) [9]. Rather than minimizing distances, as k means does, PAM minimizes dissimilarities, a more robust measure of difference measured in a dissimilarity matrix. The dissimilarity matrix allows PAM to perform clustering with respect to any distance metric and allows for flexible definition of distance. The use of medoids rather than centroids makes PAM less sensitive to outliers. Finally, PAM has the advantage of producing several outputs that help in assessing the strength of classifications. In particular, silhouette plots and scores (which range from -1 to 1) give an idea of how well an observation is grouped. A score of 1 means it is well classified, -1 poorly classified, and zero means it lies between two clusters. A clusterplot provides a further diagnostic, showing how much dimensional overlap or isolation there is between clusters. These outputs allow one to run PAM several different times with different combinations of variables or different numbers of clusters, to assess how the strength of groupings changes.

A new set of partitioning methods was developed in the nineteen nineties, based on Neural Network (NN) and Artificial Intelligence paradigms [10]. NN essentially uses a nonlinear and flexible regression technique which does not require prior assumptions of the distribution of the data to classify data. NN methods have the advantage of evaluating similarities based on a set of multi-dimensional criteria, as opposed to traditional clustering algorithms which generally use a single measure of dissimilarity. However, they suffer from the problem of some non-hierarchical clustering algorithms of requiring a user defined number of clusters and random starting points in N dimensional space when seeding the clustering algorithm. Hence, there is no clear superior method. One study found that while NNs did not perform superior to traditional k-means classifiers in geodemographic segmentation, a combination of the two, with NNs generating the input seed for k-means resulted in superior classification [11].

Yet another approach to clustering is Multivariate Divisive Partitioning (MDP). Claritas, Inc. used a form of MDP called Classification and Regression Tree (CART) analysis to more functionally classify their market segments. In this more supervised form of clustering, an analyst chooses a dependent variable or behavior they wish to model and then conducts a stepwise process to determine which variables, and which breaks in the values of those variables, best divides a single segment into two segments with the greatest difference in that behavior. Splitting then continues iteratively until a threshold of similarity in the dependent variable is reached.

Where there are a large number of variables being considered, segmentation often makes use of Principal Components Analysis (PCA), a data reduction tool which condenses a large set of variables into a smaller number of discriminating factors, while eliminating ineffectual noise. These factors are represented using standardized linear combinations of the original variables. Generally, most of the original variation in the variables is explained in the first principal component. Because each component is orthogonal, each subsequent component should be uncorrelated with the previous one, and hence explain less variance. Thus, while the number of principal components is equal to the number of variables, only a few of the principal components need be used because they explain most of the variance. Segmentation uses PCA to condense the vast number of socio-economic and demographic variables that differentiate neighborhoods into a much smaller set of largely uncorrelated factors, or principal components. For example, Applied Geographic Solutions (AGS) uses PCA in building its MOSAIC™ geodemographic segmentation. The user's guide for this product states that PCA is an excellent approach because it makes it easier to identify and isolate "site signatures," or combinations of traits that make neighborhood types unique, since each factor is a condensed and distilled representation of many related variables, and each factor is theoretically independent.

Geodemographic Segmentation, Figure 5 Adjacent polygons with the same cluster value are dissolved into single polygons

Once principal components have been generated, observational units can be classified by cluster analysis of these values or through simple numeric breaks.

The methods described up to this point are aspatial. While the observations being analyzed may represent spatial entities (e. g. zip codes or watersheds), cluster analysis does not explicitly consider spatial relationships unless specified. However, it follows logically that when spatial autocorrelation is present in the variables being analyzed, nearby geographic entities will be more likely to have the same cluster value than distant entities. That is because spatial autocorrelation also describes similarity in attribute values, just as cluster analysis does. Regardless of autocorrelation, however, actual spatial variables, such as latitude and longitude coordinates, can be used as variables in cluster analysis.

Even if no spatial variables are used in cluster analysis, spatial rules can be applied afterwards. For instance, a cluster analysis might give a result where very distant entities have the same cluster designation. In this case the analyst may wish to enforce a rule that spatial entities with the same cluster designation that do not touch or are not within a certain distance have different cluster designations. An analyst may also wish to combine entities with the same cluster designation if they are adjacent or within a certain distance. In the case of polygons this can easily be done by dissolving common boundaries to form a single polygon from multiple polygons with the same value, as shown in Fig. 5, or aggregating non-adjacent polygons into topological regions.

The output of segmentation will by definition depend on the spatial nature of entities or observations. Where spatially exhaustive polygons (e. g. zip codes) are used as the initial observational units, the cluster analysis yields a spatially exhaustive landscape classification. While adjacent polygons of the same class can be merged, the end result gives a class for all locations where land is present. Hence there is no need for any interpolation. In some cases, polygons are not actually exhaustive, but are functionally so, such as in the case of property parcels, which include all land except for public rights of way, which likely are irrelevant to most analyses using segmentation. Points may also be classed into segments although by definition the resulting map would not be geographically exhaustive. If one desires to segment a landscape from point data, though, some form of interpolation is necessary. That is, where no areal boundaries exist, they must somehow be detected. In the simplest case, Thiessen polygons/Voronoi diagrams (see ► Voronoi Diagram, this volume) are created around the points to form an exhaustive landscape, with those polygons inheriting all of the overlaying points' attributes. Polygons are then segmented through cluster analysis. In the continuous surface approach, interpolated surfaces are created for each variable relevant to the segmentation. Cluster analysis is then performed on the multiple values associated with each grid cell to yield a cluster designation. Due to the spatial autocorrelation associated with interpolation methods, nearby cells will be likely to be in the same class and in fairly large contiguous areas of the same class, which can then be converted to polygons to form an exhaustive landscape. As described under "housing market segmentation" below, this problem is very relevant to the study of housing markets, where data often come in household-level point format, but where exhaustive regions are desired as output.

Key Applications

Marketing

The most significant application of geodemographic segmentation is for marketers. Marketing is most effective when targeted at the audience most likely to purchase a product or service. Market segmentation products help companies target where and how they spend their marketing and advertising money by comparing the profile of their customer base with the profile of a particular geographic region. This can also help marketers successfully replicate their efforts and expand their campaigns.

A number of proprietary segmentation systems have been developed for this purpose, including Experian's MOSAIC™, Claritas's PRIZM®, and Consolidated Analysis Centers Incorporated's ACORN™. These traditionally use Census data, as well as other ancillary data, such as point-of-sale receipts data, purchasing patterns, mortgage and financial information, and household surveys. Claritas now uses household-level data, including responses from lifestyle surveys. In the United States, most of these systems use Census block groups, which were designed to approximate "neighborhoods" of about 350 households, as the minimum geographic unit of analysis, although the recently released PRIZM® NE used an algorithm to attempt to interpolate PRIZM® segments to individual households. In the United Kingdom, post codes are often used as the minimum mapping unit. Each segment is expected to be relatively homogeneous in its purchasing habits. Segment designations are generally updated periodically to reflect changes in the neighborhood.

Many segmentation systems are hierarchical as well. For instance, PRIZM® NE has three hierarchically nested levels of segment classes. All observational units are first classed into one of 5 "urbanicity" classes (PRIZM® 5), representing degree of urbanization. Those 5 classes are further broken down by socio-economic status to yield 14 "social groups" (PRIZM® 14). Those are then further broken down by lifestyle factors to yield 66 final segments (PRIZM® 66). Each segment and segment aggregation is given a unique name and description. So, for instance, the "Urban" PRIZM® 5 class contains a number of PRIZM® 14 sub-classes, such as "Urban Midtown." Within that are three PRIZM® 66 sub-sub-classes, including "Urban Achievers," "Close in Couples," and "Multi-Culti Mosaic." Detailed socio-economic and demographic generalities are given about each segment, as well as information about their consumption preferences. For instance, "Urban Achievers" tend to have no children in the household, be younger, and pursue active social lives, including going to clubs, bars, and concerts. More specifically, they are more likely to eat pita bread, shop at Nordstroms and Banana Republic, watch tennis, listen to jazz, and go downhill skiing.

Retail Site and Performance Modeling

When retail facilities choose a location, they consider the size and characteristics of the consumer population within a service or trade area. Geodemographic segmentation systems are commonly used for this purpose, especially by companies with large numbers of outlets. Specifically, segmentation systems can be used to assess the penetration rate of different market segments targeted by the retailer. This approach is often combined with service area analysis, in which a network algorithm is used to delineate the geographic region that is within a certain drive time of the facility (the maximum drive time for a facility will depend on the nature of the facility, e. g. home improvement store, versus convenience store). Once this region is established, the number of households belonging to each segment type can be estimated. Sites can be found that maximize the number of households belonging to the segments that are likely to demand the particular good or service. Gravity models can be used to account for competing facilities in the service area. Further, this approach can be used to help stores fine-tune what product lines they should feature based on the characteristics of the surrounding neighborhoods. For instance, a grocery chain may stock a certain store with more organic produce and high-end gourmet lines based on surrounding market segment. This approach can has been used for large facilities like big box store, smaller facilities like coffee shops, and even automated teller machines (ATMs).

Retail stores have also been segmented themselves based on their characteristics and the characteristics of their trade area. Kumar and Karand [12] created such a segmentation for grocery stores, intended to account for the interaction between internal store attributes, such as "scrambling of service," and trade area characteristics. They found that such a segmentation did predict differences in retail performance.

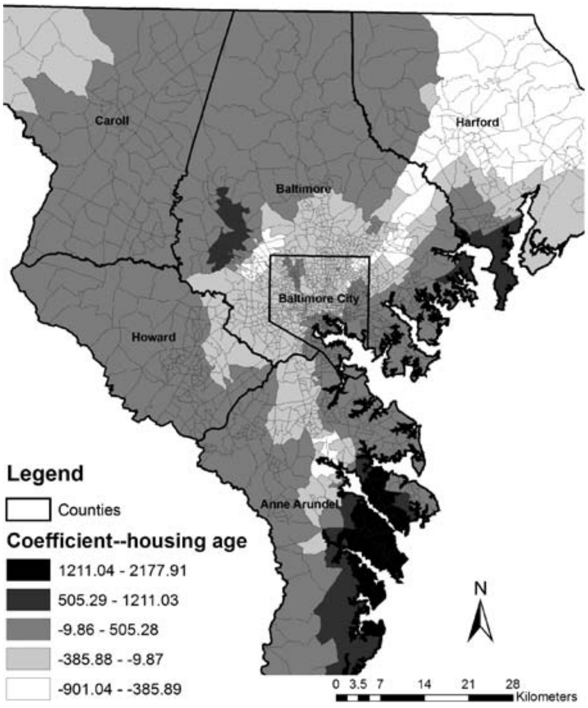
Public Service Planning

Public agencies can use geodemographic approaches to help target the location and scale of public facilities, based on the same principals of retail site modeling, but substituting public need for profit maximization. In some cases this may involve searching out demographic segments that are not well served in the private sector. In others it may involve matching facilities to areas with a certain profile expected to use those facilities. The potential applications range from siting of public transportation facilities, to libraries, to youth centers.

Housing Market Analysis

Economists contend that there are distinct housing "submarkets" within each urban area. These represent areas with relative similarity in housing characteristics, like price, size, type, ownership status, and quality of housing, among other variables. Some also contend that submarkets are defined by the relationship between these variables and price [13]. Accounting for housing submarkets through segmentation can result in great improvements for several applications, including derivation of price indices, automated valuation models (AVM's) or mass appraisal methods for mortgage lenders, and studies of non-market goods using hedonic analysis (in which property price is regressed against explanatory variables in order to disaggregate the contributors to housing price and place a value on characteristics that are not directly priced in the market). That is because these methods all rely on the assumption of a homogeneous housing market. When housing submarkets exist and are accounted for separate models can then be estimated for each market segment, resulting in increased precision. Segmentation has also been found to be necessary for generating welfare measures of benefits from hedonic indices [14,15].

A variety of methods have been proposed for segmenting housing markets, including both geographic and non-geographic methods. There has been a wide range of opinions as to whether housing markets are defined by spatial criteria or not. In some studies, individual non-contiguous properties within the same neighborhood have been classed into different segments based only on the internal characteristics of those properties and disregarding location [16], while in many others geographic partitions have been used. One recent study Bourassa et al. compared submarkets defined by small spatially contiguous geographical units (based on real estate appraisers' data) versus submarkets defined based on the properties' internal characteristics only, disregarding spatial contiguity, and found that stratifying data by the former resulted in more accurate hedonic housing price equations [17]. As the authors point out, housing submarkets matter and it is geography that makes them matter. For those using geographic units to segment housing markets, some use exogenously derived geographies, like Metropolitan Statistical Areas [14] or real estate board jurisdictions [13], while others derive segment boundaries statistically.



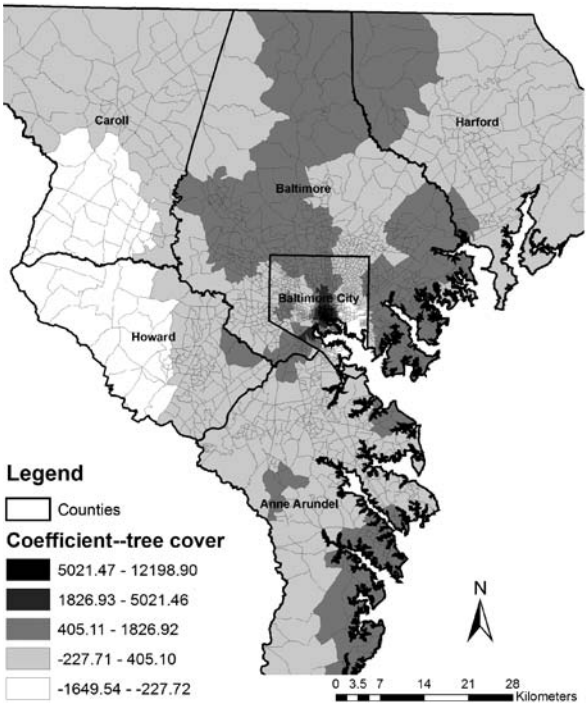
Geodemographic Segmentation, Figure 6 Block groups with grayscale values graduated by GWR coefficients for median house age as a predictor of housing price

Urban and Environmental Planning

Just as geodemographic segmentations help marketers target their clientele, they can also help planners target investments or policies. A recent study found that PRIZM® lifestyle clusters could actually be used to predict differences in urban public right of way vegetation, suggesting that PRIZM® class could be used to target future street tree investments. Other potential applications include targeting outreach efforts towards residential areas that are expected to produce higher levels of non-point source pollution, targeting pedestrian improvements to districts where residents are predicted to commute by foot or bike, helping decide the type of recreational facilities to invest in for neighborhoods based on expected preferences, and locally adjusting parking requirements based on expected car usage by neighborhood.

Future Directions

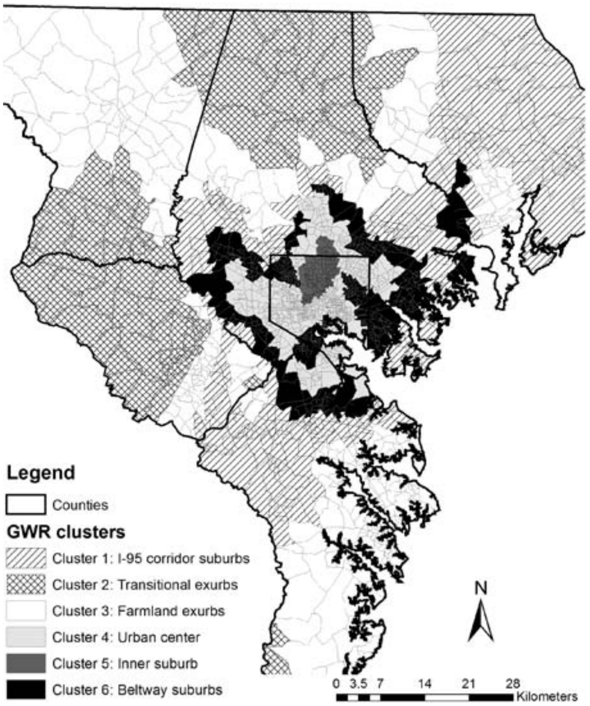
A potentially more sophisticated but rarely used approach towards segmentation is based on the use of Geographically Weighted Regression (GWR; see ► Spatial and Geo-



Geodemographic Segmentation, Figure 7 Block groups with grayscale value graduated by GWR coefficients for mean tree cover as a predictor of housing price

graphically Weighted Regression, this volume) [18]. This approach could be useful in marketing when there is some factor that cannot be easily quantified (e. g. local taste, values, customs) but which mediates the relationship between geodemographic variables (e. g. income, education) and sales. It may also be helpful in defining spatial housing market segments by showing where the changes occur in the relationship between price and attributes.

GWR is a spatial regression method which tests whether model coefficients are non-constant over space, unlike traditional regression which assumes spatial stationarity (that is, the relationship between dependent and independent variables is constant in space and time). In this method, separate regressions are run centered on each observation, with a spatial kernel determining which observations are included in the population of each localized regression and how they are weighted, based on a distance decay function. An adaptive or fixed size kernel can be used to determine the number of local points that will be included, depending on the spacing of the data. The GWR model is tested against a stationary regression model to see if its increased complexity is warranted relative to the increased predictive power. Individual parameters can then be tested for nonstationarity. Those that are can be analyzed by looking at spatial patterns in the distribution of coefficient or test statistic values, which vary continuously across the landscape. In other words, the visual pattern shows how the relationship between the independent and dependent variables is spatially variable. For instance, in certain parts of a city, the presence of parks may be positively associated with property value (where parks are well-maintained) and in other areas it may be negatively associated (where parks are characterized by crime). In other words, the spatial variation in the coefficients may signal important omitted variables. Figures 6 and 7 illustrate this approach for a GWR predicting mean housing price by Census block group; the color is graduated by the GWR coefficient for median house age in Fig. 6 and tree canopy cover in Fig. 7, holding all else equal. As these figures show, the relationships between housing age and price and tree canopy and price are in some places negative and some places positive. Cluster analysis can then be performed on the coefficient values to classify each observation. Because nearby observations will tend to have highly correlated coefficient values (since they are likely to have be part of the same moving-window estimation regression subset), this means also that nearby observations are likely to be classified similarly.



Geodemographic Segmentation, Figure 8 Block groups segmented into clusters representing housing market type using the GWR method

This approach can also be used on any geographical unit. Figure 8 shows the results of an analysis in which median housing value at the block group level was regressed against a number of socio-economic and spatial predictor variables using GWR for the Baltimore, MD metropolitan region. The resulting coefficients were analyzed with cluster analysis and, based on this, each block group was classified into one of six segments. As this map shows, individual members of each segment are generally spatially contiguous because of the spatial autocorrelation of GWR coefficients.

Cross References

- ► Hierarchical Spatial Models
- ► Spatial and Geographically Weighted Regression
- ► Voronoi Diagram

Recommended Reading

1. Tryon, R.C.: Cluster analysis; correlation profile and orthometric (factor) analysis for the isolation of unities in mind and personality., Ann Arbor, Mich.: Edwards brother, inc., lithoprinters and publishers. viii, 122 p (1939)
2. Sokal, R.R., Sneath, P.H.A.: Principles of numerical taxonomy. Freeman W.H., San Francisco, xvi, 359 p (1963)
3. Park, R.E., et al.: The City. The University of Chicago Press, Chicago, Ill 239 p (1925)
4. Weiss, M.J.: The clustering of America, 1st edn. Harper & Row, New York, xvi, 416 p (1988)
5. Weiss, M.J.: Latitudes & attitudes: an atlas of American tastes, trends, politics, and passions: from Abilene, Texas to Zanesville, Ohio. 1st edn. Little, Brown, Boston, 224 p (1994)
6. Weiss, M.J.: The clustered world: how we live, what we buy, and what it all means about who we are. 1st edn. Little, Brown, Boston, viii, 323 p (2000)
7. Debenham, J., Clarke, G., Stillwell, J.: Extending geodemographic classification: a new regional prototype. Environ. Plann. A **35**(6), 1025–1050 (2003)
8. Hartigan, J.A.: Clustering algorithms, Wiley, New York, xiii, 351 p (1975)
9. Kaufman, L., Rousseeuw, P.J.: Finding groups in data: an introduction to cluster analysis. Wiley series in probability and mathematical statistics. Applied probability and statistics.: Wiley, New York, xiv, 342 p (1990)
10. Schèurmann, J.: Pattern classification: a unified view of statistical and neural approaches. Wiley, New York, xvii, 373 p (1996)
11. Balakrishnan, P., et al.: Comparative performance of the FSCL neural net and K-means algorithm for market segmentation. Eur. J. Oper. Res. **93**(2), 346–357 (1996)
12. Kumar, V., Karande, K.: The effect of retail store environment on retailer performance. J. Bus. Res. **49**(2), 167–181 (2000)
13. Palm, R.: Spatial Segmentation of the Urban Housing Market. Econom. Geogr. **54**(3), 210–21 (1978)
14. Brasington, D., Hite, D.: Demand for environmental quality: a spatial hedonic analysis. Reg. Sci. Urban Econ. **35**(1), 57–82 (2005)
15. Zabel, J.E.a.K.A.K.: Demand for air quality. Sage Urban Stud. Abstr. **29**(1), 3–135 (2001)
16. Lipscomb, C., Farmer, M.: Household diversity and market segmentation within a single neighborhood. Ann. Reg. Sci. **39**(4), 791–810 (2005)
17. Bourassa, S., Hoesli, M., Peng, V.: Do housing submarkets really matter? J. Housing Econ. **12**(1), 12–28 (2003)
18. Fotheringham, A.S., Brunsdon, C., Charlton, M.: Geographically weighted regression: the analysis of spatially varying relationships. Wiley, Chichester, England; Hoboken, NJ, USA xii, 269 p (2002)

Geographic Coverage Standards and Services

WENLI YANG, LIPING DI
Center for Spatial Information Science and Systems,
College of Science, George Mason University,
Fairfax, VA, USA

Synonyms

Geographic Phenomena

Definition

A geographic coverage is a representation of a phenomenon or phenomena within a bounded spatiotemporal region by assigning a value or a set of values to each position within the spatiotemporal domain. This definition of geographic coverage implies that there is usually more than one value for the each phenomenon type within its spatiotemporal domain. This is in contrast to a discrete geographic feature to which a single value is assigned for each feature type. For example, a coverage representing the flow rate of a river has different flow rate values at different positions of the river, while a discrete feature representing river type has only one type value for the same river. Geographic coverage standards specify schema and frameworks for geographic coverage or coverage components. They provide common languages in describing, publishing, accessing, and processing geographic coverage data. A service is a distinct part of the functionality that is provided by an entity through interfaces [1]. Geographic coverage services are those parts of functionality that are used to access and process coverage data. While in its generic definition a service can be provided through any interface, geographic coverage services are commonly referred to as those that have standard interfaces defined by widely recognized standardization bodies.

Historical Background

The term coverage has been used in Geographic Information System (GIS) software since the early 1980s to refer to a set of data representing the geographic phenomena in certain spatiotemporal regions. Geographic coverage can be conceptually modeled in many different ways. Thus, there are many technical languages that can be used to describe the complex coverage information and the operations that can be performed on a coverage. The lack of common, consensus based modeling on geographic information results in difficulties in exchanging geographic data among different GISs and in collaborating among different GIS software components. In 1994, the OpenGIS project presented a vision of making different GIS processing software systems able to communicate with each other over the Internet by using a set of open, standard compliant interface protocols. An attempt of formally defining geographic coverage was made in one of its draft Open Geodata Interoperability Specifications. In the same year, the Open GIS Consortium Inc. (OGC), which was renamed to the Open Geospatial Consortium in 2004, was launched. Through multiple revisions, a more complete concept of coverage was drawn in 1999 in the OGC's Abstract Specification Topic 6 [2]. Since the late 1990's, the Technical Committee 211 (TC211) of the International Organization for Standardization (ISO) has been working on a number of standards and technical specifications on geographic coverage and coverage components [3], most notably, the ISO19123 standard, "Schema for coverage geometry and functions" [4]. This standard defines a conceptual schema for the spatial characteristics of coverages, including the concept of geographic coverage, its spatial domain characteristic, and the relationship between the spatiotemporal domain of a coverage and its attribute range. In the meantime, service standards either for or related to coverage or coverage components have been developed, such as the OGC's "Web Coverage Service" [5], and ISO/TC211's "Geographic Information-Service" [6].

Scientific Fundamentals

Geographic phenomena can be observed in two forms, one is discrete and the other is continuous. Discrete phenomena are usually objects that can be directly recognized due to the existence of their geometrical boundaries with other objects, such as a road, a lake, and a building. Continuous phenomena usually do not have observable boundaries and vary continuously over space, such as temperature, air quality, and reflected solar radiation from land surfaces. Due to such differences between discrete and continuous phenomena, the methods of describing these two categories of phenomena are also different. A discrete phenomenon can be described by a curve representing its boundary and some values representing its attributes. The same attribute has the same value for the within the boundary. For example, a field of crop can be described by its boundary and attributes such as its area, estimated yield, name of planted crop, and owner of the field. An attribute value, e. g., estimated yield, is the same at any position in the field because the field is described as one single discrete object. The attribute value for a continuous phenomenon, on the other hand, varies from one location to another and, thus, must be recorded at each direct position. For example, the reflected solar energy from a land surface as shown from a satellite image changes continuously across the land surface and, therefore, each pixel value of the satellite image represents the amount of the reflected solar energy from the location where the pixel locates. When information about a geographic phenomenon is stored in a computer system, different data structures are usually used for discrete and continuous phenomena. Information of a discrete phenomenon is usually stored by recording the attribute values for the phenomenon and a series of spatial coordinate values depicting the spatial geometry of the phenomenon. Information of a continuous phenomenon is usually stored by recording a series of attribute values and a scheme describing how the locations of each attribute value can be determined. The primary difference between the two is that the former stores a usually limited number of values for attributes together with typically much more spatial coordinate values for its geometry, while the latter usually stores a large number of attribute values together with usually relatively small numbers of values describing the scheme of locating the often spatially regularly distributed positions. The information on discrete phenomena and continuous phenomenon are often used differently. The operations performed on the two are also typically different. For example, geometrical overlaying and buffering analyses are frequently performed on the former while numerical computing is often conducted for the latter. In many cases, continuous data are used to derive discrete objects such as classifying a satellite image into land cover types. Thus, there are often differences in data structure designs, data encoding approaches, data accessing and processing methods for these two types of geographic phenomena. Traditionally, data structure for discrete phenomena is referred to as a vector data type, while that for continuous phenomena is referred to as a raster data type. However, the raster data structure is not sufficient to characterize the spatiotemporal domain of continuous geographic phenomena, classify different subtypes of the spatiotemporal domain, depict the relationship between the spatiotemporal domain and the attribute value range, and define appropriate operations that can be performed on coverage data. The term coverage, as herein defined, is intended to provide a concept that can comprehensively describe the continuous geographic phenomena and facilitate the design of standard interfaces for services on data recording of this kind of phenomena.

The standards on geographic coverage define conceptual schema for coverage and analyze coverage types and components. For example, the ISO19123 standard defines fundamental characteristics of coverage, including spatiotemporal domain, attribute range, major coverage types, and operations on coverages. These standards provide a common technology language and guide the development of interoperable services on coverage data. A number of services have been available involving geographic coverage standards, most notably the OGC Web Coverage Service (WCS). The WCS is designed to provide users of geospatial data an interoperable way to obtain coverage data across the Internet. Users can utilize any interoperable client to request from WCS servers of coverage data in interested spatial and temporal extent, with user-specified spatial coordinate reference system and spatial and temporal resolutions, and in a user-defined data encoding format. In contrast to the WCS' ability of providing coverage data, the OGC Web Map Service (WMS) allows clients to obtain coverage data as a map to be viewed or overlaid with other geospatial maps, including both coverage and feature maps. Special purpose service standards and services such as image classification service, image exploitation services, and coverage processing services have either been developed or are currently under development. These can provide more specialized coverage data processing capabilities for different scientific and application communities. Users of such services frequently need more specialized knowledge in their respective areas. For example, the OGC Web Image Classification (WICS) provides interoperable ways for users to perform classification of images, but it requires that a user understand which classifier and/or classification scheme is more suitable for his or her specific purpose.

Key Applications

The concept of geographic coverage and the coverage standards and services are mainly used in geospatial information technology, especially in observing, collecting, processing, archiving, cataloging, and publishing of geographical data for providers, and in searching, discovering, accessing, and acquiring of geospatial data and information for consumers. Primarily because most data and information available are either directly or indirectly related with spatiotemporal dimensions, geographic coverage, coverage standards and services find their applications in many science and application domains, such as atmospheric sciences, environmental sciences, forestry, geography, oceanography, decision making, emergency response, and regional panning. Examples of some specific applications are listed bellow.

Software Industry

With the internet has become an indispensable element of scientific research, in applications as well as in daily life, people increasingly demand that more and more information, including geographic coverage information, be freely exchanged in the Web. Software interoperability is the key for such information exchange. Coverage standards and services play important roles in the development of interoperable software associated with the use or processing of geographic information, such as GISs. Vendors of such software can benefit from following well recognized concepts and standard interface protocols in their designs.

Knowledge Discovery and Data Mining

Knowledge discovery and data mining (KDD) involves identifying novel patterns or models in often a large amount of data. KDD is usually conducted in individual data holdings such as a principal investigator's database. Web technology makes it possible to perform KDD across the entire web as long as data and service providers share common interfaces for accessing the data and services. Coverage standards and services provide essential enabling foundations for such perspectives.

Regional Management

Many geospatial data involved in regional management are historically viewed as relatively static, i. e., not varying constantly with time, and are described by discrete features, for example, transportation routes used in traffic management and stream network lines used in public safety management. Many attributes associated to such seemingly static phenomena are actually quite dynamic if examined from an individual position within the phenomena, such as the traffic flow along a transportation route and the water depth along a river. With the advances of observing capabilities such as real-time imaging, high resolution data at individual positions can be collected and processed. Coverage standards and interoperable coverage services can promote the share and use of data with more dynamic attributes and help regional management.

Government Policy

Geographic coverage standards and services assist in the process of making government policies regarding intra-agency collaboration and information distribution and sharing, for example, assessing the benefits of restructuring a local information system into conformance with a federal information system, and evaluating privacy protection for citizens in designing geospatial databases.

Citizen's Everyday Life

Citizens can benefit from geographic coverage standards and services in their daily life, such as virtual reality landscapes from earth observing data, digital museums and galleries, and education and training. In the aforementioned traffic line versus traffic flow example, a citizen is able to find the shortest route from one location to another when the road is viewed as a discrete feature. She is also able to find the shortest time needed to drive between the two locations when traffic flow rate is represented as a coverage having data at each and every point along the road and when the coverage is constantly accessible to her through standard compliant services.

Future Directions

Standardization of technology language and service interface protocols related to geographic coverage is still in its early stage and the effort is an ongoing one. Most current coverage services are not in operational use. Many factors are related to the degree of maturity of standardization and interoperable services, such as status of adopting and implementing specifications, server and client development speed, cross-community communication, easiness in adapting to existing service interfaces of data distribution systems, and government policies. Future directions shall include: stabilization of current implementation specifications, additional coverage processing specifications, promotion of coverage standards and interoperable services among major geospatial data providers, communications to more science and application communities who are employing the use of geographic data, and government enforcing policies.

Cross References

► OGC's Open Standards for Geospatial Interoperability

Recommended Reading

1. ISO/IEC: ISO/IEC, ISO/IEC TR 14252 Information technology – Guide to the POSIX Open System Environment (1996)
2. OGC: OGC Abstract Specification, Topic 6 – The coverage type and its subtypes, Version 4. Wayland, Massachusetts (1999)
3. Di, L.: The development of remote sensing related standards at FGDC, OGC, and ISO TC211. In: Proceedings of IEEE International Geoscience and Remote Sensing Symposium, Toulouse, France, 21–25 July 2003
4. ISO: ISO/TC211, ISO 19123 Geographic information – Schema for coverage geometry and functions (2005)
5. Evans, J. (ed.): OGC, Web coverage service (WCS), Version 1.0.0. Wayland, Massachusetts (2005)
6. ISO: ISO/TC211, ISO 19119 Geographic information-service (2005)

Geographic Data Management

► Privacy and Security Challenges in GIS

Geographic Data Mining

► Geographic Knowledge Discovery

Geographic Data Reduction

► Abstraction of GeoDatabases

Geographic Database Conceptual Modeling

► Modeling with a UML Profile

Geographic Databases

► Spatio-temporal Database Modeling with an Extended Entity-Relationship Model

Geographic Dynamics, Visualization and Modeling

MAY YUAN
Department of Geography and Center for Spatial Analysis, University of Oklahoma, Norman, OK, USA

Synonyms

Dynamics; Process; Evolution, landscape; Maps, animated; Map, centrographic timeseries; Map, bi-plot; Trend-surface analysis; Time-series correlation graph; Geovista; Agent-based models; Evolutionary algorithms; Mutation

Definition

Generally speaking, *dynamics* is the work of forces (or energy) that produces movement and change in a system and manifests interrelationships among the elements of space, time, and forces. Geographic dynamics centers upon such interrelationships near the surface of the Earth and often across physical, biological, and human domains. While the term, *geographic dynamics* is less popular than terms like *atmospheric dynamics* and *geological dynamics*, the concept of geographic dynamics is deeply rooted in a broad range of geographic studies. Time, like space, is fundamental to time geography, historical geography, cultural geography, environmental geography, biogeography, geomorphology, climatology, hydrology, and many other fields of geography. Spatiotemporal interrelationships among forms, patterns, and processes are central to geographic understanding of physical, biological, and human environments.

Geographic dynamics is multiscalar, multidimensional, and complex. From local and regional to global scales, geographic dynamics is displayed in human activities, urban sprawl, land use and land cover change, transformation of cultural, biological, and physical landscapes, and environmental changes in human dimensions. Geographic processes responsible for these changes are afforded by physical forcing, ecological forcing, anthropogenic forcing, or a combination thereof. Complexity of geographic dynamics emerges when geographic processes interact and drive composite changes that must be explained by all involved processes as a whole. These geographic processes may operate at multiple spatiotemporal scales and across multiple thematic domains. Visualization techniques are critical to inspect the multiscalar, multidimensional, and complex nature of geographic dynamics and in seeking forms and patterns embedded in observations and measurements. The forms and patterns serve as the basis on which to build hypotheses and computational models of geographic dynamics.

Historical Background

Field observations and mapping have been the primary means of studying geographic dynamics, and advances in computing technologies have greatly enhanced visualization and modeling capabilities in the study of geographic dynamics.

As early as 1884, geomorphologist, William Morris Davis developed the Davisian model of fluvial landscape evolution with concepts of dynamic systems used to posit the cyclical nature of erosion in landform development. Visualization was limited to sketches, diagraphs, and graphics to illustrate concepts and stages of landform development. Research interest in the temporal dimension was later incorporated into human and cultural geography. In the early 1950s, cultural geographer, Carl O. Sauer studied agricultural origins and dispersals with emphases on the interactions between physical and cultural environments and historical discourses of the interactions. Sauer's school of cultural–historical geography stressed the importance of maps to record, illustrate, and synthesize geographic patterns and explore geographic processes. Analysis of geographic dynamics in these early geographic studies is mostly qualitative and descriptive.

Besides geography, many Earth science disciplines also embrace the concepts of dynamics in their studies. Geologists, meteorologists, hydrologists, and biologists, for example, all have long histories of investigating how forces interact and result in motion and change in their respective domains of interest. *Earth System Science*, particularly, emphasizes dynamics across scales and subsystems of the Earth (e. g., the geographic system). The past and future possible behavior of the Earth system depends upon positive and negative feedbacks among these processes and human activities. Computer models have been widely developed to address the dynamics in and across the Earth's subsystem and in the Earth system as a whole. For example, there are atmospheric general climate models and ocean general climate models to address climate behavior with emphases on the effects of the atmospheric and oceanic dynamics, respectively. As early as the 1960s, attempts began to couple atmosphere–ocean general climate models to improve predictions of climate change.

In the late 1980s, global impacts of human activities were confirmed through the study of ozone depletion and chlorofluorocarbon (CFC) compounds used in industrial and commercial applications. Increased ultraviolet (UV) exposure results in increases in skin cancer and cataracts in human populations. Research into the interrelationships of human–environmental interactions relies heavily upon visualization and modeling. NASA satellite data from the Total Ozone Mapping Spectrometer (TOMS) show temporal sequences of the ozone hole as a dynamic regional phenomenon in the Antarctic, yet the regional phenomenon has shown long-term global impacts on ecosystems, such as a reduction in yield and an alteration in species competition. Empirical and statistical models have been developed to address various ozone impacts.

Visualization and modeling of geographic dynamics is further leveraged by significant advances in remote sensing and survey technologies along with the astonishing growth of internet and computing technologies in the mid 1990s. The study of geographic dynamics has been transformed from a data-poor to a data-rich environment. New visualization methods are being developed to filter massive amounts of data for pattern detection, and new computational models are being proposed to simulate and predict geographic dynamics. The synergy built upon visualization and computation expands into new modes of scientific practices centering upon data mining (DM) and knowledge discovery in databases (KDD).

Scientific Fundamentals

Visualization and modeling of geographic dynamics is a powerful combination for acheiving insights into geographic understanding. By generating an animated sequence of maps of urban growth in the Detroit region, for example, Tobler [1] invoked "the first law of geography: everything is related to everything else, but near things are more related than distant things."

Animated cartography has been the dominant technique for visualizing geographic dynamics since late 1950s. By animating maps (or images) in a time sequence, how a phenomenon changes over space and time becomes visually evident. There is no shortage of eye-catching animated maps for visualizing climate change, hurricane tracking, population dispersion, land use and land cover change, and many other geographic dynamics. Over the last 30 years, cartographers have been investigating methodological and cognitive issues to promote effective graphical communications about change and facilitate an understanding of processes. [2] summarizes six major research areas in animated mapping: (1) methods for animating time-series data, (2) methods for animating across attributes of data, (3) methods for representing uncertainty and data quality in animated maps, (4) designing effective temporal legends and controls, (5) identifying the visual variables of animation, and (6) methods for temporal interpolation and smoothing of sparse datasets.

Compared to static maps, animated maps provide additional opportunities for and challenges to communicating multidimensional spatiotemporal information. Besides the eight conventional visual variables in static mapping, three dynamic variables (*scene duration*, *rate of change between scenes*, and *scene order*) can be used in the design of map animation to emphasize location, change, and attributes of the geographic phenomenon of interest [3]. The effectiveness of methods and designs for map animation is evaluated by cognitive studies through experiments, interviews, and surveys. Slocum [4] argues strongly that cartographic methods must be iteratively tested within a theoretical cognitive framework and engineering principles.

The design of animated maps should include particular attention to the notion of cognitive friendliness because information load and complexity is much higher in animated maps than in static maps. Ignorance of the effects of the added temporal dimension can induce biases in communication. Based on the theory of human information processing, the human eye–brain system does not instantaneously process patterns from short-term to long-term memories. Therefore, information perceived later in a cartographic sequence in our short-term memory may mask previously viewed patterns that have already entered our long-term memory. Monmonier [5] suggested four animation supplements to bridge the perceptual biases in animated cartography: (1) the centrographic time-series map; (2) the biplot, a joint, two-dimensional representation of time units and places based upon two principal components; (3) canonical trend–surface analysis; and (4) the time-series correlation graph.

Beyond animated maps, geovisualization has emerged as a discipline that connects cartographic principles, multivariate statistics, database systems, data mining, and knowledge discovery. Immersive, interactive, and multimodal software packages have been developed to explore and present dynamic spatiotemporal data. GeoVista is one successful example [6] that provides a suite of toolboxes for spatial data visualization and analysis. The study of geographic dynamics is thriving with the recent surge of integrated visualization–computation approaches. Tobler's urban growth simulation research demonstrated the power of combining visualization and modeling techniques. Hägerstrand's *Time Geography* in the 1970s has been rejuvenated with a growing interest in visualization and geocomputation of human activities [7]. In addition to statistical and mathematical models that describe the behavior or transitions of a system in arithmetic, differential, logistic, or other mathematical terms (such as stochastic models, Markov chains, and Monte Carlo simulation), new approaches to spatiotemporal modeling address a wide range of geographic dynamics that account for local effects, interactions, and complexity in space and time. Examples of popular computational models include distributed modeling, *cellular automata*(CA), *agent-based modeling* (ABM), and evolutionary (or geneitc) algorithms.

These geocomputational models present a new school of empirical modeling in geography. Geocomputational models provide opportunities for thought experiments to shape ideas and build hypotheses. Distinguished from statistical or mathematical models, geocomputational models consider behavior at the individual level and examine emergent properties at aggregation. Common to all geocomputational models is spatial discretization. *Distributed hydrological modeling*, for example, starts with the determination of water flow properties and surface conditions (e. g., amount, flow direction, infiltration, roughness, etc.) at discrete locations. Hydrological patterns (e. g., stream networks) and dynamics (e. g., stream discharge) are aggregated properties of hydrology at these discrete locations. Nevertheless, spatial discretization can go beyond the cell-based approach to allow flexibility in spatial positioning and movement [8].

Geocomputational approaches handle individuality from different perspectives. In addition to distributed modeling, CA and spatial ABM have been broadly applied to urban, ecological, and epidemiological applications. CA emphasize transitional rules for cells to change from one state (e. g., rural area) to another (e. g., urban area) to drive changes in spatial patterns over time (e. g., urban growth). Neighbor effects are often incorporated into the development of transitional rules, such that a cell is likely to become an urban cell when all its neighboring cells have changed to urban cells, and collectively the transition of cells from rural to urban states simulates urban growth. *Spatial* ABM, in contrast, is built upon cognitive mobile decision-making agents that interact with each other and the environment. The collective behavior of these individual agents results in large-scale patterns in space and time [9].

Nevertheless, CA and spatial ABM are complementary bottom-up approaches to revealing geographic dynamics at multiple scales. CA simulate environmental conditions and how the conditions may change over space and time. Spatial ABM mimics the spatial behavior of individuals. CA and spatial ABM together can express geographic dynamics as an interactive system of agents and environments and allows for spatial exploration of dynamics in the context of human–environment interactions [10].

Evolutionary (or genetic) algorithms can add “intelligence” to ABM to allow for adaptability and learning. Spatial evolutionary modeling builds upon Darwinian principles of evolution. Each individual agent possesses a genetic make-up that determines its interactions and responses to the environment and environmental change. In addition, every individual is assigned an objective function and a fitness value as a measurement of the individual’s reach to objectives. Evolutionary principles, such as selection, mutation, cross-over, learning, and adaptation are incorporated into the constituents and make-ups of individual agents. The added intelligence enables agents to cope with the environment by changing their behavior as well as to account for multicriteria evaluation of the problems they face. Furthermore, evolutionary algorithms allow for changes in agent composition and population: those who cannot adapt themselves to the environment may decease, while there will be births of new generation of agents from cross-over and mutations. These evolutionary functions bring a higher level of complexity into geographic dynamic modeling by extending dynamics from the individuals in ABM to the communities by considering evolutionary processes. With incorporation of dynamic properties into both agents and the environment, the combination of CA, ABM, and evolutionary algorithms offers an enormous potential to integrate object–field approaches into the investigation of human–environment interactions.

Key Applications

The world is dynamic, and geographic dynamics is central to understanding the biophysical environment, the human environment, and their interactions. No geographic knowledge is complete without a full comprehension of the spatial processes responsible for forms, patterns, relationships and the pathways by which these processes shape the landscape over time. Hence, visualization and modeling of geographic dynamics have a wide range of applications across biophysical and human domains.

Weather and Climate

Visualization and modeling of atmospheric conditions over space and time is central to weather forecasting and climate prediction. Unidata (http://www.unidata.ucar.edu) has been developing and maintaining impressive suites of tools for display and analysis of weather and climate data. Space–time algorithms using visualization means have been developed to evaluate the spatiotemporal behavior of mesoscale rainstorms and assessment similarity in rainstorm development in space and time [11].

Land Use and Land Cover Change and Urban Development

CA and ABM techniques have been broadly used to simulate land use and land cover change (LULC) and urban development [12]. Scenarios generated by simulations are instrumental in exploring urban growth and LULC possibilities and drawing insights into potential drivers for planning considerations.

Human Activities, Movement, and Interactions

Underpinned by theoretical frameworks from time geography, modeling human activities, movement, and interaction has been carried out by eliciting and analyzing travel patterns, employment choices, and many other human activities with space and time constraints. Visualization is used to track individual movements through space–time paths and to evaluate the individual’s accessibility to locations through space–time prisms [13]. Most existing studies are limited to urban environments. Progress in this research area holds great promises for applications in public safety, emergency preparedness, social program development, and a wider range of urban planning and policy making.

Spread of Species, Wildfire, Diseases, or Pollutants

Visualization and modeling of biological and physical processes expand upon the underlying principles that guide the process development. Different species, for example, have distinctive patterns of movement and migration. Some species are solitary, but others operate in herds. Likewise, wildfires behave differently in grasslands or in forests. The spread of disease and dispersion of pollutants are both constrained by environmental conditions and the nature of the dispersion agents. CA and ABM are common methods to simulate biological or physical spreads with proper rules to capture their distinctive spatial process patterns [14]. Furthermore, modeling of wildfire spread and pollutant dispersion must incorporate the underlying physical mechanisms (e. g., thermal dynamics and atmospheric dynamics) according to scientific understanding.

Future Directions

Geographic dynamics is ubiquitous and essential to advance our understanding of the world. Visualization and modeling are technologies that enable such advancement. As spatiotemporal data acquisition grows exponentially at a rapid pace, powerful and robust visualization and modeling tools are urgently needed to decipher forms, patterns, and relationships embedded in spatiotemporal data to understand geographic dynamics. CA, ABM, and spatial evolutionary algorithms (SEAs), numerical modeling techniques developed for specific domains (such as groundwater flows) have successfully demonstrated capabilities to model geographic dynamics. However, the massive amount of ever-growing spatiotemporal data challenges both the computation and visualization powers of these tools. High-performance computing (HPC) has been a major driver to enable numerical modeling of atmospheric and fluvial dynamics but is still at the experimental stage for CA, ABM, and SEA. HPC is one important future direction to the research on visualization and modeling of geographic dynamics [15]. Other important research topics include the development of visual analytical and methodological frameworks for geographic dynamics across multiple scales and multiple domains.

Cross References

▸ Exploratory Visualization

Recommended Reading

1. Tobler, W. R.: A computer movie simulating urban growth in the Detroit region. Econ. Geogr. **46**, 234–240 (1970)
2. Harrower, M.: A look at the history and future of animated maps. Cartographica **39**, 33–42 (2004)
3. Dibiase, D., Maceachren, A.M., Krygier, J.B., Reeves, C.: Animation and the role of map design in scientific visualization. Cartogr. Geogr. Inf. Syst. **19**, 201–214 (1992)
4. Slocum, T. A., Montello, D.R., Fuhrmann, S., Hedley, N.R., Blok, C., Jiang, B., Koussoulakou, A.: Cognitive and usability issues in geovisualization. Cartogr. Geogr. Inf. Sci. **28**, 61–75 (2001)
5. Monmonier, M.: Summary graphics for integrated visualization in dynamic cartography. Cartogr. Geogr. Inf. Syst.19, 23–36 (1992)
6. Gahegan, M., Hardisty, F., Takatsuka, M., Wheeler, M.: Introducing Geo VISTA Studio: an integrated suite of visualization and computational methods for exploration and knowledge construction in geography. Comput. Environ. Urban Syst. **26**, 267–292 (2002)
7. Kwan, M.-P.: GIS Methods in time-geographic research: geocomputation and geovisualization of human activity patterns. Geogr. Ann., Ser. B Hum. Geogr. **86**, 267–280 (2004)
8. Bithell, M., Macmillan, W.D.: Escape from the cell: Spatially explicit modeling with and without grids. Ecol. Modelling **200**, 59–78 (2007)
9. Batty, M.: Agents, cells, and cities: new representational models for simulating multiscale urban dynamics. Environ. Planning A **37**, 1373–1394 (2005)
10. Torrens, P. M.: Simulating sprawl. Ann. Assoc. Am. Geogr. **96**, 248–275 (2006)
11. McIntosh, J., Yuan, M.: Assessing similarity of geographic processes and events. Trans. GIS **9**, 223–245 (2005)
12. Dietzel, C., Clarke, K.C.: Decreasing computational time of urban cellular automata through model portability. GeoInformatica **10**, 197–211 (2006)
12. Miller, H. J.: Measuring space-time accessibility benefits within transportation networks: basic theory and computational procedures. Geogr. Anal. **31**, 187–212 (1999)
14. Komarov, A. S., Palenova, M.M., Smirnova, O.V.: The concept of discrete description of plant ontogenesis and cellular automata models of plant populations. Ecol. Modelling **170**, 427–439 (2003)
15. Armstrong, M.P., Cowles, M.K., Wang, S.: Using a computational grid for geographic information analysis: a reconnaissance. Prof. Geogr. **57**, 365–375 (2005)

Geographic Information

▸ Market and Infrastructure for Spatial Data

Geographic Information Retrieval

▸ Retrieval Algorithms, Spatial

Geographic Information Sciences

▸ Information Services, Geography

Geographic Information Systems

- ► Information Services, Geography
- ► Spatial Decision Support System

Geographic Knowledge Discovery

HARVEY J. MILLER
Department of Geography, University of Utah,
Salt Lake City, UT, USA

Synonyms

Spatial data mining; Geographic data mining; Exploratory spatial analysis; Pattern recognition in spatial data; Hypothesis validation in spatial data

Definition

Geographic knowledge discovery (GKD) is the human-centered process of using computational methods and visualization to explore massive geo-referenced digital databases for novel and useful geographic knowledge. GKD is an extension of the more commonly known process of knowledge discovery from databases (KDD), with "data mining" being a central activity of this process.

Main Text

Geographic information science exists in an increasingly data-rich environment facilitated by the development of high spatial and temporal resolution environmental monitoring devices, location-aware technologies, information infrastructure for data sharing and interoperability, as well as reductions in data storage costs. Traditional spatial analytical methods were developed when data collection was expensive and computational power was weak. The increasing volume and diverse nature of digital geographic data easily overwhelm techniques that are designed to tease information from small, scientifically sampled and homogeneous datasets. They are also confirmatory and require the researcher to have *a priori* hypotheses, meaning that they cannot discover unexpected or surprising information.

Geographic knowledge discovery (GKD) attempts to leverage the continuing exponential growth in computing power to find novel and useful geographic knowledge hidden in the unprecedented amount and scope of digital geo-referenced data being collected, archived and shared by researchers, public agencies and the private sector. This is a human-centered process that involves activities such as developing background knowledge as a guide, selecting, cleaning and reducing the data to make it usable, data mining through the application of efficient, scalable techniques for extracting patterns, interpreting these patterns and constructing knowledge. These stages are not necessarily sequential and are often applied in an iterative manner.

Cross References

- ► Time Geography

Recommended Reading

1. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery: An overview. In: Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Ulthurusamy, R. (eds.) Advances in Knowledge Discovery and Data Mining, pp. 1–34. MIT Press, Cambridge, MA (1996)
2. Gahegan, M.: On the application of inductive machine learning tools to geographical analysis. Geogr. Anal. **32**, 113–139 (2000)
3. Miller, H.J., Han, J.: Geographic data mining and knowledge discovery: An overview. In: Miller, H.J., Han, J. (eds.) Geographic Data Mining and Knowledge Discovery, pp. 3–32. Taylor and Francis, London (2001)

Geographic Market Segmentation

- ► Geodemographic Segmentation

Geographic Markup Language

- ► Web Feature Service (WFS)

Geographic Metadata

- ► Metadata

Geographic Phenomena

- ► Geographic Coverage Standards and Services

Geographic Profiling

- ► CrimeStat: A Spatial Statistical Program for the Analysis of Crime Incidents

Geographic Resources Analysis Support Software

- ► GRASS

Geographic Spatial Regression

▶ Spatial and Geographically Weighted Regression

Geographic Weighted Regression (GWR)

▶ Spatial Regression Models

Geographical Analysis

▶ Crime Mapping and Analysis
▶ Data Analysis, Spatial

Geographical Information Retrieval

▶ Internet-Based Spatial Information Retrieval

Geographically Weighted Regression

▶ Data Analysis, Spatial

Geography Markup Language (GML)

Raimundo F. Dos Santos Jr., Chang-Tien Lu
Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

Synonyms

GML; Format; Open source; Interoperability; Extensibility; Exchange format; OGC; ISO; Resource description framework (RDF); Properties, simple; Properties, geometric; Reference system, temporal; Reference system, spatial; Features; Schema

Definition

Geography Markup Language (GML) is an open-source encoding based on the eXtensible Markup Language (XML), and suitable for the representation of geographical objects. Organized as a hierarchy of features, collections, and geometries, among other structures, GML objects are modeled after real-world entities characterized by properties and state. In addition, GML has been defined as an information exchange and storage format with which disparate systems can share common geographic data. GML schemas establish a standard blueprint of how geographic objects can be defined by one system and understood by others in a vendor-independent manner.

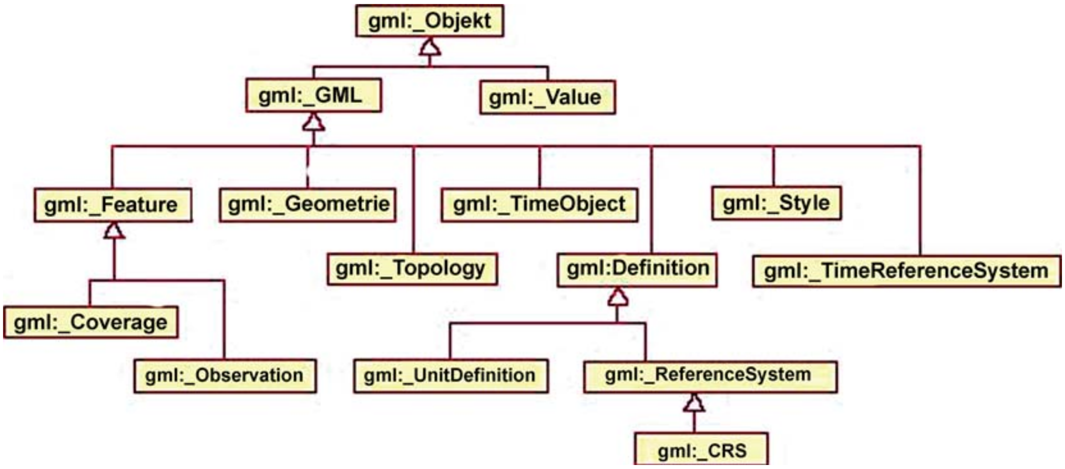
Historical Background

Advances in geospatial applications have promoted the creation of technologies more suitable to handle geographic data. As one of those technologies, GML was developed and released by the Open Geospatial Consortium (OGC), an international entity formed by various public and private organizations. The OGC adopted XML as defined by the World Wide Web Consortium (W3C) as its foundation for the development of GML. By obtaining acceptance from industry, government, and academia, the GML specification has been incorporated as a document of the International Organization for Standardization (ISO).

The development of **GML 1.0** began in 1999. In its first version, it used the *Resource Description Framework* (RDF) metadata model to make statements about resources. Known as triples, RDF statements combine the concepts of subject, predicate, and object to describe a particular object and its characteristics. GML 1.0 defines *Simple Properties* such as strings, numbers, and Boolean values, and *Geometric Properties* such as points and lines. GML 1.0 was a major initiative in the geospatial domain, but more importantly, it served as a stepping stone towards the development of a more robust language.

Under **GML 2.0**, released in 2001, the OGC introduced GML Schemas as a more flexible approach for describing and validating geographic objects. While the language itself was enhanced with more sophisticated structures, the OGC maintained its goal of easy interoperability and extensibility. Three base schemas, namely Geometry, Feature, and XLink, were released with the goal of supporting geospatial applications for specific domains. In addition, it provided a framework for better utilization of linked features in distributed data sets. GML 2.0 addresses mostly two-dimensional simple features, and introduced better support for namespaces to distinguish among features and properties defined in different application domains.

With the advent of **GML 3.0** in 2003, the OGC has made significant additions to the number and capabilities of its schemas. As real-world phenomena is not just limited to the one and two-dimensional spaces, the support for features with complex, non-linear, 3D geometry has been added to the language. Other enhancements include the support of 2D topology and the definition of coverages as subtypes of features. For applications in which historical data and localized timestamps are important, GML 3.0 provides temporal references systems. Units of measurement and default styles for visualization have also been



Geography Markup Language (GML), Figure 1 A Subset of the GML Class Structure [3]

```
<gml:location>
  <gml:Point gml:id="xy1"><element ref="gml:boundedBy"/>
    <gml:pos>-21.4 122.993</gml:pos>
    <gml:LocationString>City of Reston</gml:LocationString>
  </gml:Point>
</gml:location>
```

Geography Markup Language (GML), Example 1 Simple GML Features

defined in this specification. As of this writing, version 3 is the current release.

Scientific Fundamentals

System designers can extend the GML base schemas to create an Application Schema suitable to their specific application domains. While many schemas comprise the GML Specification, a single application normally would only implement a subset of them. For example, Feature is one of the main components of GML, and therefore most applications import the *feature.xsd* schema. The system designer, however, need only import the *topology.xsd* schema if the features in question store topological characteristics. Under GML 3, components are defined according to the Class Hierarchy of Fig. 1 [3]:

The hierarchical structure of Fig. 1 illustrates the parent-child relationship that binds GML components. Each class represents one or more aspects of a real-world entity, helping describe its features, geometric characteristics, location, time of existence, spatial reference system, and many other qualities. Within the various application schemas, Designers may use components "as is" or they may choose to restrict or extend the GML base types to tailor the language to their own needs. "*gml:_Object*" is the abstract class that represents a GML data set or document. Two important components of the GML Specification are the *Feature* and *Feature Collection* elements, represented in the diagram by the *GML:_Feature* class. A feature (or a collection of features) can be a road, a river, a building or any geographic entity that needs representation. Features may have several properties defined, such as *gml:location, gml:boundedBy*, and *gml:LocationString*, as in the code fragment (Example 1):

The *gml:Geometry* abstract class defines objects that can be represented as *points, curves*, and *polygons*. The middle of a region may be described as a point having a *centerOf* property set to a given value. An element may be identified by its *id*, a name, and a description. It may be associated with a spatial reference system (attribute "*gml:SRSReferenceGroup*"). Some of the geometry types currently supported are *gml:Point, gml:Curve, gml:LineString, gml:_Surface, gml:Polygon, gml:_Ring, gml:Circle*, and *gml:LinearRing*. Geometries are often qualified with the Coordinate Reference System to which they belong. In Example 2, the *gml:coordinates* element describes a point in the WGN84 coordinate system having $x = 22.21$ and $y = 31.99$. The point itself may represent an

```
<gml:Point gml:id="gr065" srsName="urn:ogc:def:crs:WGN84">
  <gml:coordinates>22.21, 31.99</gml:coordinates>
</gml:Point>
```

Geography Markup Language (GML), Example 2 Coordinate Reference Systems

```
<gml:TimePeriod gml:id="period01">
  <gml:begin>
    <gml:TimeInstant gml:id="instant01">
      <gml:timePosition>2006-12-04</gml:timePosition>
    </gml:TimeInstant>
</gml:TimePeriod>
```

Geography Markup Language (GML), Example 3 GML Time Elements

```
<gml:FeatureStyle featureType="exp:Town">
  <gml:GeometryStyle>
    <gml:style>fill:yellow</gml:style>
  </gml:GeometryStyle>
</gml:FeatureStyle>
```

Geography Markup Language (GML), Example 4 Styling Components

island in the ocean, a building in the city or just a point on the ground.

Temporal properties are some of the most recent additions to the GML 3 Specification. Features may be associated to time elements with the use of such properties as *gml:timePosition, gml:timePeriod, and timeInstant* (Example 3).

To describe an event, an action that occurs over a period of time, GML provides the abstract *TimeSlice* element. According to the GML Specification, a time slice describes a process-oriented event, such as object changes over time, whereas a timestamp relates to a specific snapshot of the object. The *gml:MovingObjectStatus* element is able to capture the conditions that a moving object experiences over an elapsed period. The *history* and *track* elements may also be used to observe the lifecycle of an object. The components described above do not represent an exhaustive list of time-specific components.

GML makes a strict separation between data and presentation. None of the GML constructs have a native capability of presenting data according to a styling method. For this reason, a default styling mechanism can be used in conjunction with a GML data set to provide proper styling. The term "default" specifies that the styling mechanism that comes with a data set may be used or ignored altogether. The *Feature* class defines certain components useful in the visualization process, such as *GraphStyle, LabelStyle, GeometryStyle*, and *TopologyStyle* objects as exemplified in Example 4:

Querying is one of the most important functions in information retrieval. A GML query language must be flexible enough to support querying and retrieving both spatial and non-spatial data. There have been some attempts at the creation of query languages for GML data. However, most are native to XML, and not necessarily designed for GML. Some examples are XML Query Language (XQL), XML-QL, Quilt, XQuery, and Lorel [2]. XQuery [1], a W3C standard, has been one of the most popular approaches for XML querying, and for GML querying by association [4]. Querying on GML data can be made more efficient when the querying language is able to handle spatial constructs. For example, regions defined by a *Minimum Bounding Rectangles* (MBR) can be manipulated more efficiently with *overlaps, contains, intersects*, and other spatial operators. Similarly, spatial joins are a powerful way of aggregating and linking spatial features that are common in GML data sets.

Key Applications

GML has found its way in many fields of application including, but not limited to, Planning and Resource Management, Mapping and Cartography on Mobile Devices, Real-Time Emergency Response Services, and Image Processing in Data Streams.

Planning and Resource Management

GML is becoming the *de facto* standard for the representation, exchange, and storage of geographic data. Local, state, and federal government agencies have increasingly adopted the use of GML for real estate zoning, area mapping and planning, insurance determination, vegetation monitoring, and coastal erosion. For example, GML applications can be constructed to map and visualize different views of the city, specific sections prone to flooding, sources of pollution, and historical time slices of water bodies. Even though spatial data can be voluminous, GML allows users to work with a small subset of the total data that more closely relates to their interest.

Mapping and Cartography on Mobile Devices

Advances in wireless networks have contributed to the widespread use of mobile devices for both business and personal purposes. GML has the potential to fill the gap for mapping applications in PDAs and cell phones where location-based services use the constant exchange of data from GPS systems to synchronize moving objects with remote servers and perform nearest-neighbor searches.

Real-Time Emergency Response Services

GML is modularized in such a way that the entire data set or just a small subset can be analyzed at a time. By managing Minimum Bounding Rectangles (MBR), users can search and view only the information that pertains to them. For example, highways in the North part of the city can be monitored for unusual activity where fog is prominent during certain hours of the morning. By gathering relevant information and filtering out unimportant features, emergency personnel can focus their attention on a smaller area and respond as needed on a near real-time basis.

Image Processing in Data Streams

GML is designed to support streaming data flows. As such, it provides robust facilities that can handle incoming information from various remote sources. This is especially helpful in dynamic environments where changes occur frequently. Applications for satellite imagery and remote sensing are typical examples. Weather data, for instance, are commonly generated from satellite feeds with large volumes of information in short time intervals. Meteorological agencies rely on the frequent and correct processing of this information for fast and accurate forecasts.

RSS News Feeds

In the recent past, GML has been proposed as a data delivery medium to be used in conjunction with RSS news feeds. The idea has seen considerable attention as the RSS format is boosted by the popularity of blogs and internet news sites. GML geometries along with other components such as temporal constructs may be leveraged to make application data more relevant in specific contexts. Political sites, for example, may want to describe the president's trip according to visited locations, spatial characteristics of the areas in question, and other annotations that support the evolving story. While an official schema has not been published as of this writing, this area of application promises to make geospatial data more readily available not only to members of the geospatial community, but also to casual readers of internet news.

Future Directions

Even though GML has gained increased acceptance, there is still much work that can be done to improve its usefulness and efficiency. GML may benefit from a query language that supports spatial queries such as overlaps, intersections, ranges, and nearest neighbors. The objects represented in GML are often more complex than those typically encoded in XML, since geographic objects have both spatial and non-spatial attributes. The data model for a GML query language has to reflect this complexity. Therefore, current XML query languages must be extended to support the rich array of elements that GML makes available. GML data needs indexing just as much as other spatial data sources. Because it is encoded in semi-structured XML, GML data must be parsed out with tools such as SAX (Simple API for XML) or DOM (Document Object Model). The extracted data in turn can be indexed using one of several approaches such as R Trees or Quad Trees. The exact approach most suitable for GML needs further evaluation. Integrating GML with mobile devices is a promising direction. Mobile environments are inherently challenging due to their constrained nature. Adapting GML for low-bandwidth networks and limited-storage devices will help bring those devices inline with other distributed data sources that hold important geospatial information.

Cross References

- ► Data Models in Commercial GIS Systems
- ► Geospatial Semantic Web
- ► Indexing, Hilbert R-tree, Spatial Indexing, Multimedia Indexing

Recommended Reading

1. XQuery 1.0: An XML Query Language, W3C Working Draft. http://www.w3.org/TR/xquery/. Accessed 3 Dec 2006
2. Bonifati, A., Ceri, S.: Comparative Analysis of Five XML Query Languages, ACM SIGMOD Record, vol. 29, pp. 68–79, March, (2000)
3. Cox, S., Daisey, P., Lake, R., Portele, C., Whiteside, A.: OpenGIS Geography Markup Language (GML 3.0) Implementation Specification, OpenGIS Specifications. http://www.opengis.org/specs/?page=specs. Accessed 3 Dec 2006
4. Li, Q., Moon, B.: Indexing and Querying XML Data for Regular Path Expressions. In: Proceedings of the 27th International Conference on Very Large Data Bases, Rome, Italy, pp. 361–370 (2001)
5. Peng, Z.R., Tsou, M.H.: Internet GIS – Distributed Geographic Information Services for the Internet and Wireless Networks. Wiley Publishing, Hoboken, New Jersey (2003)
6. Shekhar, S., Chawla, S.: Spatial Databases: A Tour. Pearson Education, Inc., Upper Saddle River, New Jersey (2003)

Geoinformatic Surveillance

► Hotspot Detection, Prioritization, and Security

GEOINT

► Intelligence, Geospatial

Geo-Intelligence

► Intelligence, Geospatial

Geolocation

► Indoor Positioning, Bayesian Methods

Geo-Mashups

► Computer Environments for GIS and CAD

Geomedia

► Intergraph: Real Time Operational Geospatial Applications

Geometric Engine Open Source (GEOS)

► Open-Source GIS Libraries

Geometric Fidelity

► Positional Accuracy Improvement (PAI)

Geometric Modeling

► Vector Data

GeoOntologies

► Knowledge Representation, Spatial

Geo-Portal

► deegree Free Software

Geopriv Group, IETF

► Privacy Threats in Location-Based Services

Georectified

► Photogrammetric Products

Georegistration

► Registration

Geo-Role-Based Access Control

► Security Models, Geospatial

GEOS Library

► PostGIS

Geosensor Networks

Silvia Nittel
Department of Spatial Information Science
and Engineering, University of Maine, Orono, ME, USA

Synonyms

Sensor networks; Environmental sensor networks; GSN

Definition

Advances in small, low-cost microelectronic and mechanical systems (MEMS) with limited on-board processing and wireless communication capabilities, and the development of novel sensor materials enables us to build a new generation of technology that consists of large collections of untethered, small-form, battery-powered computing nodes with various sensing functions. These sensor nodes can be densely distributed over a geographic region, and are able to measure environmental processes, such as weather development, seismic activity, or track the movements of toxic fumes at a level of detail that was not possible before. The continued trend towards miniaturization and inexpensiveness of sensor nodes makes it possible that such sensor nodes are less than a cubic millimeter in size, and sensor networks can be made up of thousands or even millions of sensors.

G*eosensor networks* (GSN) are a specialized application of sensor network technology to monitor, observe and track environmental phenomena and processes [11]. Geosensor networks are deployed in geographic space to monitor environmental phenomena at a high spatio-temporal scale and resolution, and the data is available in near real-time. For example, a sensor network was deployed on Great Duck Island in Maine to observe the nesting behavior of an endangered species of storm petrels [1]. Another sensor network consisting of 60 motes was deployed to monitor the microclimate around a single Redwood tree over a period of several months [2].

Historical Background

From telescopes to microscopes, humans have developed instruments to monitor and observe the world in ways that are not obvious to the human eye and senses. Many technological innovations have been made over the past century to sense information within the environment. Some of the most impressive and powerful instruments are, e. g., remote sensing satellites deployed in space. Such instruments enable scientists to monitor processes on the earth at a global scale. With local weather stations, rainfall samples and/or wind direction and speed can be measured. Recording such information allows the analysis of long-term trends in climate development on an entire continent. Today, it is characteristic that sensors are stationary, expensive and thus, sparsely distributed over a large geographic area. Data is logged and stored locally, and often retrieved in batch mode manually or via satellite link. Thus, dynamic complex processes at a small and local scale are more difficult to observe due to the scale of the existing instruments. During the mid-1990s, the miniaturization of computing devices and the development of novel sensor materials and microsensors lead to the technology of wireless sensor networks. Sensor networks consist of a large number of small computing devices with attached sensor boards, and equipped with batteries and wireless communication. They can be deployed in a local environment in an untethered way, and sense the environment at temporally and spatially high-resolution in real-time.

Sensor networks, however, have the following constraints that pose novel challenges from a system and application developmental standpoint:

Power consumption Sensor nodes are limited with regard to their battery supply, and energy conservation is a major system design principle.

Low-range communication The bandwidth and range of wireless communication is limited. Since communication is a much higher drain on the energy consumption than on-board processing, optimizing communication within the sensor network is a major system design consideration.

Limited computing and storage capabilities Sensor nodes have, at least for the foreseeable future, limited on-board computational, volatile, and persistent storage capabilities. Thus, on-board data processing using available memory and CPU is also limited.

Self-organization Due to the large number of sensor nodes, the failure rates of nodes, and the often unattended deployment, task management and handling in sensor networks is decentralized and self-organizing. Thus, some level of local autonomy must be provided for the devices.

Scientific Fundamentals

The deployment of sensor networks provides outstanding ability to monitor discrete and continuous phenomena in physical space.

Over the last years, much research has focused on the design of platforms, programming languages for such constrained environments (e. g., nesC), robust, energy-efficient communication protocols [3], and the development of sensor materials and sensor devices. A major task of sensor networks, however, is their data collection and processing instrumentation. Albeit their powerful and novel capabilities to observe the physical world, today programming sensor networks is cumbersome due to the failure-prone nature of nodes and communication links, and the vast parallel computational nature of such systems.

A user, often a domain scientist, needs to define the necessary tasks in a user-friendly way, and delegate the optimization and ultimately self-adaptive execution to the run-time environment without having to worry about the details. From a database perspective, a sensor network can be viewed as a distributed database system (DBS) with sensor nodes that run lightweight versions of the DBMS themselves. The DBMS supports a single sensor table and the attributes match the attached sensors such as temperature, humidity, location, etc. The query syntax is similar to SQL-style queries and is extended with sampling epochs for continuous queries. Thus, a user can interact with the sensor network as a virtual single sensor, and send declarative queries to it. Each sensor node runs a tiny footprint DBS locally, and participates in global distributed query execution. Tuples for the sensor table are created in an append-only style created by all sensor nodes in the sensor network. The first sensor database system prototypes are TinyDB [4] and Cougar [5].

Typical sensor network queries are spatio-temporal. For example, a user might be interested in hourly measurements of temperature values along the Redwood tree to observe the local microclimate around the tree:

```
SELECT sensor.id, temperature FROM sensors
SAMPLE PERIOD 60min.
```

Take note that queries define the requested data in a declarative way, and the user does not need to know the availability or node identifier of sensor nodes, nor does he/she deal with the details of the execution of the query. The DBMS distributes the query to the sensor nodes of interest using a geo-routing or a tree-based routing protocol. Similarly, query results are accumulated and routed back to the base station using a tree-based protocol. Using an energy-efficient data aggregation tree, nodes combine local partial data aggregation results with routing partial results back to the query originating sensor node.

Currently, spatial queries over a sensor network often retrieve discrete information measured at the location of sensor nodes. Many environmental phenomena, such as a temperature field or a gas concentration field in an open space are spatio-temporally *continuous*. Individual sensor readings, however, are point samples taken at the physical location of the sensor nodes about the underlying phenomenon [5,6]. For geosensor networks, the challenge exists to provide an accurate and precise estimation of all points of dynamic spatial field based on limited discrete point samples collected within a geosensor network. Since sensor nodes are energy- and processing-limited due to their small-form factor and battery power, estimation algorithms have to be lightweight and processed in a distributed manner 'in the network'.

Another setting for geosensor networks are sensor nodes that are *mobile*, such as sensor nodes installed within automobiles or used by pedestrians, attached to animals, floating on the ocean [10] or embedded in the road pavement or traffic lights. Here, up-to-date information is collected and exchanged in an ad-hoc, peer-oriented fashion between nodes in spatial proximity [9]. Neighboring sensor nodes in intelligent transportation systems can inform other existing nodes in the same region about environmental events such as icy patches or other unexpected road hazards that can be sensed on the road.

Key Applications

The application areas for geosensor networks are plentiful and can be found in the following areas: biodiversity, biogeochemical cycles, climate change, infectious diseases, invasive species, carbon cycle, earthquakes and tsunami warning systems, coral reefs observation, animal and habitat monitoring, and coastal and ocean observation. Some applications have been mentioned before [1,2].

It is typical for geosensor networks that the technology is integrated with existing, larger-scale sensing platforms such as remote sensing instruments, buoys, autonomous underwater vehicles, wind mills, or weather stations to integrate sensed data for the same spatial region at various spatial and temporal scale. When combining sensor networks with existing sensing platforms, so-called *sensor webs* are created. Similar to the idea of the World Wide Web, the vision exists that sensors and sensor networks should be accessible and usable in a uniform way so that scientists can find, combine, and query real-time sensors in a geographic region for a specific application. This leads to the fact that (often expensive) sensor platforms are more reusable for different purposes. The OpenGIS Consortium provides a standardization approach for sensor platform interfaces and data exchange (Sensor Web Specifications) for sensor webs. Today, several research networks have been established in the area NEON (National Ecological Observatory Network), GLEON (Global Lake Ecological Observatory Network), CREON (Coral Reef Environmental Observatory Network), and the NEPTUNE Cyberinfrastructure.

Future Directions

Currently, the actual application and deployment of geosensor networks is in its infancy. With the increasing robustness of sensors, wireless communication, improved battery life and technology, and software developments, geosensor networks will be a powerful, added technology to the existing environmental sensing platforms [12].

Cross References

- ► Geosensor Networks, Estimating Continuous Phenonena
- ► Geosensor Networks, Qualitative Monitoring of Dynamic Fields

Recommended Reading

1. Mainwaring, A., Polastre, J., Szewczyk, R., Culler, D.: Wireless sensor networks for habitat monitoring In: 1st Int'l Workshop on Wireless Sensor Networks and Applications, Atlanta, Georgia 2002
2. Tolle, G.,Polastre, J., Szewczyk, R., Culler, D., Turner, N., Tu, K., Burgess, S., Dawson, T., Buonadonna, P., Gay, D., Hong, W.: 2005. A Macroscope in the Redwoods. International Conference on Embedded Networked Sensor Systems (ACMSensys), San Diego, November (2005)
3. Woo, A., Tong, T., Culler, D.: Taming the underlying challenges of reliable multihop routing in sensor networks. In: Proceedings of the 1st International Conference on Embedded Networked Sensor Systems (ACMSensys), pp. 14–27. Los Angeles, November 2003
4. Madden, S., Franklin, M.J., Hellerstein, J., Hong, W.: TAG: A Tiny Aggregation Service for Ad-Hoc Sensor Networks. In: Proceedings of the 5th USENIX Symposium on Operating Systems Design and Implementation (OSDI), pp. 131–146. December 2002
5. Fung, W.F., Sun, D., and J. Gehrke, J.: COUGAR: The Network is the Database. In: Proceedings of International ACM SIG Conference Modeling of Data (SIGMOD), Madison, WI 2002
6. Sharifzadeh, M., Shahabi, C.: Supporting spatial aggregation in sensor network databases. In: Proceedings of ACM Symposium Advanced in Geographic Information Systems (ACM-GIS), pp. 166–175. Washington, DC (2004)
7. Hellerstein, J.M., Hong, W., Madden, S., Stanek, K.: Beyond Average: Towards Sophisticated Sensing with Queries. In: Proceedings of Workshop on Information Processing In Sensor Networks (IPSN), 2003
8. Jin, G., Nittel, S.: NED: Efficient Event Detection in Sensor Network. In: Proceedings of MDM-Workshop "Mobile Location-Aware Sensor Networks", Nara, Japan 2006
9. Nittel, S., Duckham, M., Kulik, L.: Information Dissemination in Mobile Ad-hoc Geosensor Networks. In: Proceedings of the 3rd International Conference on Geographic Information Science (GIScience 2004), Springer LNCS, College Park, Maryland 2004
10. Nittel, S., Trigoni, N., Ferentinos, D., Neville, F., Nural, A., Pettigrew, N.: A drift-tolerant model for data management in ocean sensor networks In: Proceedings of MobiDE'07, in conjunction with SIGMOD, Beijing, China 2007
11. Stefanidis, A., Nittel, S.: Geosensor Networks, CRC Press, Boca Raton (2004)
12. Nittel, S., Labrinidis, A., Stefanidis, A.: Advanced in Geosensor Networks. Springer LNCS (in press 2007)

Geosensor Networks, Estimating Continuous Phenonena

GUANG JIN
Department of Spatial Information Science and Engineering, University of Maine, Orono, ME, USA

Synonyms

Interpolation of continous geofields; Estimation predication; Phenomenon spatial field; Sample trial; Sensor networks; Energy constraint; Estimation, parametric; Estimation, non-parametric; Tiny aggreagation service; Distributed algorithm

Definition

Geosensor networks (GSN) are deployed to monitor different environmental phenomena over a spatiotemporal space. Many environmental phenomena (also called *spatial fields*), such as a temperature field or a gas concentration field in an open space, are spatiotemporally continuous within the spatiotemporal space. Individual sensor readings, however, are point samples taken at the physical location of the sensor nodes about the underlying phenomenon. Thus, neighboring sensor nodes' readings are likely similar.

For GSN, the challenge is to provide an accurate and precise estimation of all points of dynamic spatial field based on limited discrete point samples collected by the network. Since sensor nodes are energy and processing limited due to their small-form factor and battery power, estimation algorithms have to be lightweight and processed in a distributed manner "in the network".

Main Text

Sensor database management systems (SDMS) provide users an intuitive and easy-to-learn programming interface to define observation tasks for phenomena, and can also serve to estimate and track continuous phenomena. One of the most challenging objectives of SDMS is to provide accurate and precise estimation results for continuous phenomena based on limited point samples processed in real-time within the network in a distributed and collaborative way. Since GSN are energy- and computing-constrained, studies on estimating continuous phenomena in GSN have to maintain a graceful balance between the quality of the estimation results and the resource consumption within GSN.

Traditionally, spatiotemporal point samples are collected in advance and processed on a centralized computer. Generally speaking, there are two types of methods on how to estimate continuous phenomena from point samples.

Parametric estimation methods estimate an underlying phenomenon as a function. The function is usually a linear combination of a set of basic functions. For example, in a polynomial regression, $f(x, y) = \beta_0 + \beta_1(x) + \beta_2(y) + \beta_3(xy) + \varepsilon$, different ordered polynomials are the basic functions. The parameters, βs, are chosen to minimize the error, ε.

Nonparametric estimation methods require no prior specified model structure, and estimate an underlying phenomenon directly from the samples. For example, moving average is a simple nonparametric estimation, which estimates the phenomenon reading as the average sample value from a moving region.

Wireless communication between sensor nodes is very expensive, in terms of energy consumption from the battery-driven sensor nodes. A traditional centralized approach to estimating the dynamic field of a continuous phenomenon, however, requires large amounts of wireless communication to transmit raw readings from the distributed nodes to a central base computer. An efficient estimation solution in GSN needs to use distributed algorithms that are executed within the network to minimize the communication cost. Two types of communication cost, the processing and the result representation cost, are necessary for any estimation method in GSN. These two communication costs are also important to evaluating and comparing the efficiency of different algorithms.

In many cases, a global average is useful to estimate a phenomenon. Tiny AGgregation (TAG) service [1] presents an in-network solution to computing an average of sensor readings using a routing tree structure [2]. In TAG, a node computes the average result based on its own reading and its children's readings. The node sends its local partial result to its parent. Finally, the root is able to compute the global average based on all its children nodes. TAG requires only a concise communication package to represent a partial average, and processes the partial results locally. However, if the sensor nodes are not evenly distributed over a monitoring region, the simple average usually returns faulty results. A weighted spatial average [3] applies a weight to each sensor reading, and returns more accurate results. For example, a monitoring region can be partitioned into a Voronoi diagram based on the sensor nodes' location. The reading for a Voronoi cell can be weighted by the size of cell, and aggregated along the routing tree. Generating the Voronoi diagram, however, requires additional processing costs.

Often, users prefer a more fine-grained estimation result than a simple average value over an entire region or a subregion. TAG has been extended [4] to return a fine-grained estimation. For example, each sensor node can represent a grid cell in a monitoring region. A node can fill its reading to the corresponding grid. Nearby similar readings can be represented by a larger polygon region. In this way, the root node can generate a contour map to estimate the underlying phenomenon. However, representing the contour map requires additional costs and has to be simplified to save expensive wireless communication [5]. Parametric estimation methods, such as kernel regression [6], have also been applied to estimate underlying phenomena in GSN. A set of kernel functions are used as the basic functions. In this way, only the estimated values of kernel functions are necessary to represent the underlying phenomena. However, to compute the optimal parameter values, complex matrix operations are required, which consume large amounts of communication in GSN.

The constraints of GSN, especially the expensive wireless communication, challenge efficient phenomenon estimation methods in GSN. An efficient estimation solution has to return high quality results while still preserving the processing and representing requirement.

Cross References

- ► Geosensor Networks
- ► Geosensor Networks, Qualitative Monitoring of Dynamic Fields

Recommended Reading

1. Madden, S., Franklin, M.J., Hellerstein, J.M., Hong, W.: TAG: a Tiny AGgregation service for ad-hoc sensor networks. SIGOPS Operating Systems Review, 36, pp. 131–146. ACM Press, New York (2002)
2. Woo, A., Tong, T., Culler, D.: Taming the underlying challenges of reliable multihop routing in sensor networks. In: Proceedings of the First International Conference on Embedded networked sensor systems, California, USA, 4–7 Nov 2003, pp. 14–27. ACM Press, New York (2003)
3. Sharifzadeh, M., Shahabi, C.: Supporting spatial aggregation in sensor network databases. In: Proceedings of the 12th annual ACM international workshop on Geographic information systems, Washington DC, USA, 12–13 Nov 2004, pp. 166–17. ACM Press, New York (2004)
4. Hellerstein, J.M., Hong, W., Madden, S., Stanek, K.: Beyond average: Towards Sophisticated Sensing with Queries. In: Proceedings of the Second International Workshop on Information Processing in Sensor Networks, pp. 63–79, Palo Alto, CA, USA, 22–23 Apr 2003
5. Harrington, B., Huang, Y.: In-network surface simplification for sensor fields. In: Proceedings of the 13th annual ACM international workshop on geographic information systems, pp. 41–50, Bremen, Germany, 4–5 Nov 2005
6. Guestrin, C., Bodik, P., Thibaux, R., Paskin, M.A., Madden, S.: Distributed regression: an efficient framework for modeling sensor network data. In: Proceedings of the third international symposium on information processing in sensor networks, Berkeley, California, USA, 26–27 Apr 2004, pp. 1–10. ACM Press (2004)

Geosensor Networks, Formal Foundations

LARS KULIK
Department of Computer Science and Software
Engineering, The University of Melbourne,
Melbourne, VIC, Australia

Synonyms

Provable properties; Theoretical analysis; Mathematical theory of geosensor networks

Definition

The geometric nature of geosensor networks (GSNs) has sparked an interest in researching their formal foundations. As the research on formal foundations has just begun, there is no single view of what encompasses a theoretical approach to sensor networks. Typically, a theoretical approach applies techniques from computational geometry, theoretical computer science, or mathematics, in particular (algebraic) topology and graph theory. The goal of a formal analysis is to identify fundamental properties of a sensor network, for example, the total number of sensors required if every point in the sensor field has to be covered by a fixed minimum number of sensors. More generally, a formal analysis studies provable properties of GSNs and algorithms used for GSNs. A formal analysis can compute the computational complexity of an algorithm and decide, if a problem has an efficient solution or if it is intractable.

Historical Background

Current approaches to evaluating algorithms in GSNs are mainly based on extensive simulations or the deployment of test-beds. Extensive simulations can analyze an algorithm under a variety of conditions but can only assess algorithms in terms of their average runtime. Test-beds are ideal to determine the performance of an algorithm under realistic conditions but the results are only applicable to the specific settings of the test-bed. A rigorous algorithmic complexity analysis of a geometric algorithm can not only evaluate the average complexity of an algorithm, but more importantly its worst case complexity. Such a complexity analysis can be vital for the life expectancy of a GSN. For example, a geometric routing algorithm that has a low average energy cost, can have a high worst-case cost in certain situations, which could deplete the batteries of a significant number of nodes and thus lead to the failure of the entire GSN.

Scientific Fundamentals

The monitoring, collection, storage, and processing of geospatial data lies at the core of GSNs [1]. Spatial information is vital for GSNs on two counts: first, in sensor networks it is generally not important which sensor has monitored a geospatial phenomenon, but what type of data and *where* this data is sensed within the network; second, the nodes of a GSN often have the ability to determine their own positions and those of other nodes. Geometric and topological approaches exploit the availability of spatial information to research more efficient algorithms and to discover fundamental characteristics of a GSN.

The features of a GSN are primarily determined by the spatial layout of its deployment and its underlying geography. Its spatial configuration controls the topology of the network, in particular, which nodes are in communication range and, therefore, which routing and data collection algorithms are most effective. The proximity of the nodes influences how well areas are covered by the sensor field and affects the achievable resolution monitoring a geospatial phenomenon. The proximity also determines the spatial correlation of the sensor readings, and in turn decides which and how often sensors need to transmit and provide sensor readings. A formal analysis can identify key parameters that determine the global behavior and properties of a sensor network. Theoretical and algorithmic results can assist in designing, configuring, and administering a GSN, and aid in the selection of appropriate algorithms for tasks such as monitoring, tracking, or routing.

The deployment of a sensor network can be formally represented as a graph: vertices represent the sensor nodes and edges the links between the sensor nodes, i. e., specify whether or not two nodes are in communication range. The edge weights can represent the received signal strength. In a location-aware GSN vertices are not only known by their IDs but also have positions, and their edge weights can reflect the Euclidean distance between two nodes in communication range.

Key Applications

There is a large variety of research areas in GSNs in which formal geometric approaches can be applied. The areas include coverage problems, topology control, data collection and aggregation, routing and dissemination, as well as the discovery of the spatial layout in terms of holes and boundaries. This article surveys these five topics and shows how formal approaches can advance them.

Coverage Problems

Research on coverage problems [2] investigates efficient techniques, where to deploy sensors if a certain data quality is required and how to adapt a sensor network such that each point in the sensor field is covered by a certain number of sensors. Computational geometry provides a large range of techniques that can be used for coverage problems. It is known that in order to guard a room with n corners, at least $n/3$ sensors (guards) are necessary [3]. Similarly, if sensors are randomly placed, then the Voronoi cell of a sensor is given by all points in the sensor field that are closer to this sensor than to any other sensor in the network. Assuming that the sensor closest to a location provides the highest data quality and thus should be responsible monitoring that location, the Voronoi diagram determines the responsibility region of each sensor. Computational geometry provides efficient algorithms to compute the Voronoi cells [4].

Topology Control

The topology of a sensor network is determined by the communication range of each sensor node. As communication is expensive, it is imperative to adjust the communication range to save energy. The key question in topology control [5] is: given a set of sensor nodes, what is the critical transmission range to ensure a connected network? As sensor nodes are cheap devices, most approaches assume that all nodes have the same transmission range. If the locations are precisely known, then controlling the topology of a sensor network can be formally modeled as a minimum spanning tree problem. A minimum spanning tree of a graph is a subgraph, more precisely a tree, with the same vertices that connects all vertices and minimizes the sum over all edge weights. The critical transmission range is the longest edge in the minimum spanning tree. There are a number of algorithms in graph theory to compute the minimum spanning tree efficiently in a distributed manner [6]. The theory of geometric random graphs [7] allows the computation of the critical transmission range even if the node locations are not precisely known, as long as the nodes are randomly and uniformly distributed. Random geometric graphs are generated by randomly uniformly placing points into an area (often the unit square) and connecting any pair of points by an edge if their distance is smaller than a predetermined threshold.

Data Collection and Aggregation

Efficient collection and aggregation algorithms are vital for larger GSNs. A theoretical analysis [8] shows that for any wireless network the throughput per node converges to zero if the number of participating nodes increases in an architecture where each node separately reports to a single processing node. Aggregation algorithms trade communication for computation, as a sensor node consumes significantly less energy for processing than for communication. Instead of transmitting the raw data from each node individually, nodes compute aggregates such as the sum or the average of sensor readings and only transmit those aggregates. As a result, aggregation not only prolongs the life expectancy of a network, but also increases the scalability of the network. The problem of collecting data from a set of sensor nodes in the most efficient way back to a central sink can be regarded as an instance of the minimum Steiner tree problem [9]: given a set of sensor nodes, the Steiner tree is the minimum spanning tree that includes those nodes. This problem is known to be NP-complete (Non-deterministic Polynomial time), which means there is no efficient way known to compute the Steiner tree. This has initiated a variety of local algorithms that can approximate the optimal collection tree [10].

Routing and Dissemination

In routing or dissemination, the task is to route a packet between two or more nodes along an optimal path, which might be the shortest, most reliable, or most energy-efficient path. Geometric routing [11] assumes that a node knows the locations of itself, its neighbors in communication range, and the destination node. Geometric routing protocols are lightweight because they do not need to maintain or discover routes, as is the case for topological routing protocols, which are based on link information only. A criticism often put forward for location awareness is that the energy cost for acquiring coordinates can be high and unreliable. The underlying idea of geometric routing is the use of a local algorithm that forwards a message to a neighbor in communication range that is closer to the destination than itself. A greedy strategy selects among those neighbors in each step the closest one to the destination [12]. Although this greedy strategy is loop-free, it can get stuck if none of its neighbors is closer to the destination than itself. If the set of points that are in communication range of a sensor but are closer to the destination than the sensor itself does not contain any other sensor, then that area is called a void. Since voids can happen frequently, a backup strategy is required. There is a large body of theoretical work on face-routing algorithms [13] that can be used as a fallback strategy. If a graph is planar, it has been shown that face routing along the perimeter of voids can guarantee the delivery of all packets. This is immediately applicable to GSNs as it is possible to compute locally a planar communication subgraph. As location is a powerful concept in routing, current research extends this idea in the form of virtual coordinates to sensor networks that are not location-aware. Virtual coordinates are based on the idea that nearby nodes have similar coordinates, whereas nodes that are far apart have different coordinates. In all sensor networks, nodes have local information about the topology of the network, and the virtual coordinates are computed based on the observed connectivity. Virtual coordinates can be more beneficial as positional information for routing, because nearby nodes might not be connected, for example if they are separated by a wall.

Discovery of the Spatial Layout

Discovering the shape features and spatial layout of a GSN deployment is vital to routing and data collection algorithms. Many algorithms have been evaluated under the assumption that the deployment area of a sensor network is rectangular and the nodes themselves are uniformly distributed. This assumption is often too optimistic, and algorithms that perform well in simulations are less suitable in real GSN deployments. The performance of an algorithm depends on the spatial features of a deployment, in particular its boundary, its medial axis, its communication and deployment holes, its corridors that could bottleneck the transmission of sensor readings, as well as the clustering of the sensor nodes. Algorithms oblivious to those features can prematurely collapse a sensor network. One example: an algorithm that routes information irrespective of the residual energy level always along the boundary of a hole will quickly deplete those nodes and leads to a rapid growth of the hole, which might disconnect the network or exponentially decrease its life expectancy. Furthermore, a dramatic change of the network's boundary could signal a major event such as a fire or a thunderstorm that destroyed significant parts of the network. Algebraic topology [14] and computational geometry both provide techniques to identify shape features of a GSN; however, many techniques assume global knowledge. Research in both areas is currently underway to adapt these techniques for sensor networks.

Future Directions

Although many of the techniques in computational geometry and mathematical topology are often directly applicable to GSNs, they typically assume global knowledge about the sensor network, for example in the computation of a Voronoi diagram or the Steiner tree. However, providing global knowledge at the node level is an exceptionally expensive operation and usually not a viable option in sensor networks. Many algorithms need to be revisited and tailored for local computations that only assume geometric knowledge in the neighborhood of a sensor node. More generally, techniques are sought that can infer from local knowledge of individual nodes, global characteristics of the entire spatial layout of the network, such as its boundary or the number of holes. One approach to inferring global features from local properties is based on combinatorial maps, which can be applied to study spatiotemporal phenomena such as dynamic spatial fields in a qualitative manner [15]. Furthermore, algorithms generally assume that the position of nodes is precisely known. In GSNs, however, node positions are often inexact or not available, and links between nodes can be asymmetric or might be intermittent, posing new challenges for formal approaches to GSNs.

Cross References

- ► Geosensor Networks, Qualitative Monitoring of Dynamic Fields
- ► Mereotopology
- ► Voronoi Diagram

Recommended Reading

1. Stefanidis, A., Nittel, S.: GeoSensor Networks. CRC, Boca Raton, FL (2004)
2. Cardei, M., Wu, J.: Coverage in wireless sensor networks. In: Ilyas, M., Magboub, I. (eds.) Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems, pp. 19:1–19:10. CRC, Boca Raton, FL (2004)
3. O'Rourke, J.: Computational Geometry in C. Cambridge University Press, New York (1988)
4. Berg, M., Kreveld, M., Overmars, M., Schwarzkopf, O.: Computational Geometry: Algorithms and Applications. Springer, New York (2000)
5. Santi, P.: Topology control in wireless ad hoc and sensor networks. ACM Comput. Surveys **37**, 164–194 (2005)
6. Gallager, R., Humblet, P., Spira, P.: A distributed algorithm for minimum-weight spanning trees. ACM Trans. Programm. Languages Syst. **5**, 66–77 (1983)
7. Diaz, J., Penrose, M., Petit, J., Serna, M.: Convergence theorems for some layout measures on random lattice and random geometric graphs. Combinatorics Probability Comput. **9**, 489–511 (2000)
8. Gupta, P., Kumar, P.: The capacity of wireless networks. IEEE Trans. Inf. Theory **46**, 388–404 (2000)
9. Krishnamachari, B., Estrin, D., Wicker, S.: The impact of data aggregation in wireless sensor networks. In: Proceedings of the 22nd International Conference on Distributed Computing Systems, Workshops, Vienna, Austria, 2–5 July 2002
10. Kulik, L., Tanin, E., Umer, M.: Efficient data collection and selective queries in sensor networks. In: Nittel, S., Labrinidis, A., Stefanidis, A. (eds.) Advances in Geosensor Networks. Springer, New York (2007, in press)
11. Kuhn, F., Wattenhofer, R., Zhang, Y., Zollinger, A.: Geometric ad-hoc routing: of theory and practice. PODC '03: Proceedings of the 22nd Annual Symposium on Principles of Distributed Computing, Boston, MA, 13–16 July 2003
12. Karp, B., Kung, H.: GPSR: greedy perimeter stateless routing for wireless networks. In: Proceedings of the 6th International Conference on Mobile Computing and Networking, Boston, MA, 6–11 Aug 2000
13. Bose, P., Morin, P., Stojmenovic, I., Urrutia, J.: Routing with guaranteed delivery in ad hoc wireless networks. Wireless Networks **7**, 609–616 (2001)
14. Ghrist, R., Muhammad, A.: Coverage and hole-detection in sensor networks via homology. In: Proceedings of the 4th International Symposium on Information Processing in Sensor Networks, Los Angeles, CA, 25–27 Apr 2005
15. Duckham, M., Nittel, S., Worboys, M.: Monitoring dynamic spatial fields using responsive geosensor networks. In: GIS '05: Proceedings of the 13th Annual ACM International Workshop on Geographic Information Systems, Bremen, Germany, 4–5 Nov 2005

Geosensor Networks, Qualitative Monitoring of Dynamic Fields

MATT DUCKHAM
Department of Geomatics, University of Melbourne, Melbourne, VIC, Australia

Synonyms

Ambient spatial intelligence; Qualitative spatial reasoning; Discretization of quatitative attributes; Combinatorial map; Qualitative spatial representations

Definition

Environmental phenomena that vary continuously across regions of space and periods of time, such as changing sea temperature, concentrations of gas pollutant in the air, or levels of soil moisture, are called dynamic spatial fields. Information about dynamic spatial fields is important to a wide range of environmental applications. One of the goals of using a geosensor network (GSN) is to enable improved, more detailed monitoring of dynamic spatial fields. Individual sensor nodes in a GSN usually generate *quantitative* information. For example, a node might record a temperature of 23°C located at coordinates 18.04S, 146.49E at time 20:51:06 on January 7th 2007. However, the information needed by decision makers is typically *qualitative*, and concerns the relationships between groups of spatially and temporally nearby records. For example, an environmental manager may be interested in the whether a high temperature "hotspot" has grown or moved.

Generating qualitative information about dynamic spatial fields within a GSN presents a number of challenges. The most important challenge is to achieve qualitative monitoring using only on local communication between nearby nodes. Resource limitations in GSN mean that global communication, where any node can communicate with any other, is not scalable. Thus, studies of qualitative monitoring of dynamic spatial fields usually assume that at any time an individual node in the network does not have access to global knowledge about the state of the entire network, only to local knowledge about the state of its immediate neighbors.

Historical Background

Qualitative spatial reasoning is concerned with discrete, non-numerical properties of space. There are three main reasons for being interested in the qualitative (as opposed to the quantitative) aspects of geographic space [5]:

- Qualitative properties form a small, discrete domain; quantitative properties form a large, continuous domain, often modeled by real numbers. For example, temperatures in degrees Kelvin are modeled using the set of non-negative real numbers. Yet for some applications, temperature may be adequately modeled as an element from the set {hot, cold, warm}.
- Qualitative properties are *supervenient* on, and derivable from, quantitative properties. For example, in a particular application the numerical temperature 35°C may be described qualitatively as "hot."
- The boundaries between qualities normally correspond to salient discontinuities in human conceptualization of quantitative properties. For example, in coral reef monitoring applications, the qualitative boundary between "warm" and "hot" may be set to correspond to the quantitative temperature at which coral reefs are in danger of coral bleaching.

The literature contains many studies of different qualitative aspects of space, including relative distance [6,9] and direction [3,4], and in particular topological relationships between spatial entities [2,7].

Scientific Fundamentals

With respect to GSNs, the three general reasons for being interested in qualitative aspects of geographic space lead directly to three potential advantages of using qualitative monitoring of dynamic spatial fields in GSNs.

- Because qualitative properties form a smaller discrete domain than quantitative properties of space, processing and communication of qualitative information in GSNs can potentially be achieved more efficiently, using less resources, than for quantitative information.
- Any quantitative information generated by sensors nodes can always be converted into a less detailed qualitative representation, although the converse is not true. Further, the inherent imprecision of qualitative information can help make sensor networks more robust to imprecision and other forms of uncertainty in sensor readings.
- Using qualitative representations enables salient entities to be derived from complex dynamic fields, reducing system complexity and resulting in GSNs that are easier to design, construct, and query.

Looking at the problem from the application perspective, it is possible to identify at least five distinct issues facing any GSN for monitoring dynamic spatial fields in the environment.

1. Spatial: The phenomena of interest are spatial, for example involving points and regions; metric and topological relations; and spatial autocorrelation.
2. Temporal: The phenomena of interest change over time. Much of this change is spatiotemporal, including move-

ment and events, such as splitting and merging of regions within the field.

3. Scale dependency: The phenomena of interest are typically scale-dependent, and phenomena observable at one spatial scale may be different from those observable at a different scale.
4. Imperfect knowledge: Our knowledge of geographic information is always in some way imperfect. In particular, the spatial distribution of sensors and temporal frequency of sensing leads to *granularity*, the existence of grains or clumps in data.
5. Local computing: GSN are a type of highly distributed spatial computing, where individual nodes in the network typically only have access to information about their immediate vicinity. The challenge is to construct systems with desirable global properties using only local knowledge and local behaviors.

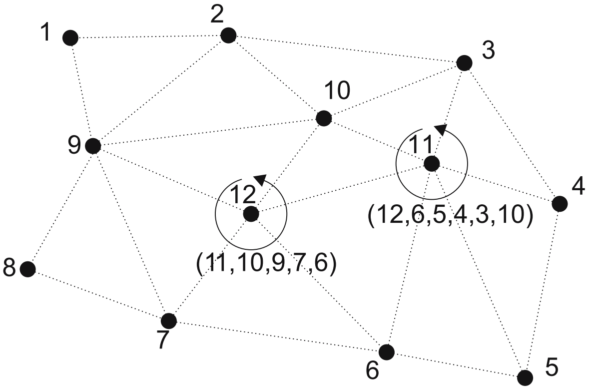
Although each of these five issues has a history of research in isolation, and in some combinations, the combination of all five of these issues makes monitoring of dynamic spatial fields especially complex. This complexity is fundamental to the need for qualitative representation and reasoning techniques. Representing continuous dynamic information using discrete sets of salient symbols is a first step to reducing complexity. The second step is to develop formal techniques for local reasoning about these discrete, salient symbols.

Example: Event-Oriented Georesponsive Sensor Networks

As an example, imagine designing a sensor network tasked with monitoring a dynamic spatial field, such as a network for monitoring sea temperature in marine coral reef environments. Qualitative representation and reasoning can be used to managed the high levels of complexity in such application domains.

As explained above, the first step is to create discrete, qualitative representations of continuous information. Discretizing the values of the field itself starts to reduce complexity. Often, natural and salient categories may already exist for an application domain, such as a classification into "cold," "cool," "warm," and "hot" water. Using less detailed, qualitative representations can immediately increase robustness to imperfection into the system. For example, miscalibrated or inaccurate temperature sensors are much less likely to affect qualitative information about the temperature (e. g., "warm" versus "hot") than quantitative information (e. g., 35.2°C versus 34.9°C).

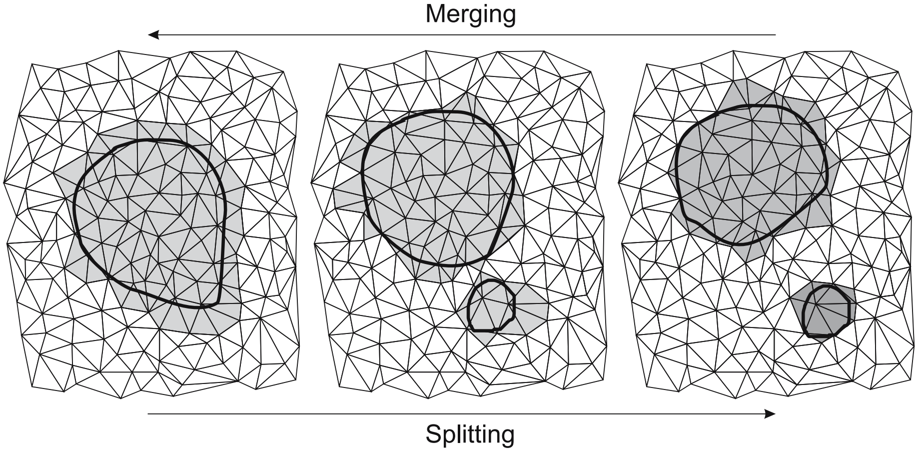
Discretization of the space itself further reduces complexity. The sensor nodes themselves already provide a discrete framework for the space under investigation. One approach to structuring this framework adopted by previous work [1,8] is to use a combinatorial map. Figure 1 illustrates the idea behind a combinatorial map, where each node has a number of defined neighbors and stores the cyclic ordering of those neighbors. Using a combinatorial map allows a variety of qualitative planar structures, such as triangulations, to be built without requiring any quantitative location information, such as geodetic or local coordinates. The structure is purely local: each node only knows information about the relative direction of it's immediate neighbors. Further, the inherent spatial imprecision in combinatorial maps, and related qualitative spatial structures, means that the resulting system can be more tolerant to imperfect information (e. g., the cyclic ordering of neighbors around a node is is much less likely to be subject to inaccuracy that, for example, location systems that rely on exact coordinate locations or bearings from one sensor to another).



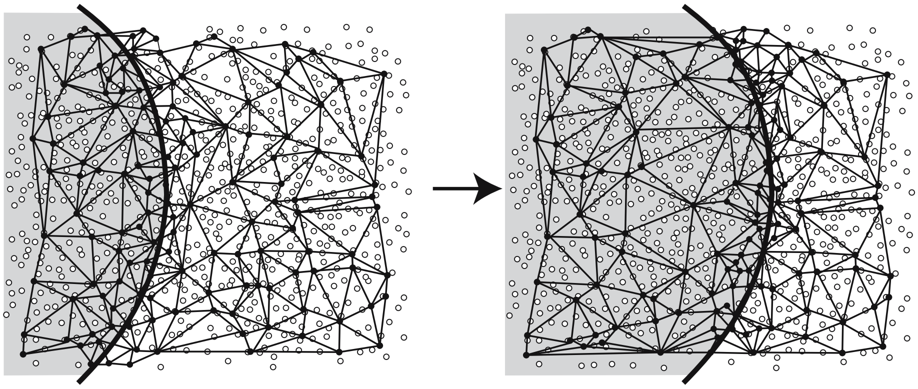
Geosensor Networks, Qualitative Monitoring of Dynamic Fields, Figure 1 Combinatorial map of sensor nodes and neighborhoods, including (counterclockwise) cyclic ordering for nodes 11 and 12

Having created appropriate qualitative representations, the second step is to develop local techniques for reasoning about these qualitative representations. In Fig. 2a region of dynamic spatial field (such as a high temperature "hotspot") is being tracked through a GSN structured as a triangulation (using a combinatorial map). Assuming the region moves continuously, a variety of qualitative spatial events can be locally detected. In order for regions to split or merge, for instance, they must first go through an intermediate stage where a single node connects two distinct parts (Fig. 2, center). As a consequence of the combinatorial map structure, this node can locally detect that a split/merge event is taking place (see [8] for more information).

Because the combinatorial maps is locally constructed, it can be efficiently and dynamically reconfigured. The concept of a *georesponsive* sensor network aims to acti-



Geosensor Networks, Qualitative Monitoring of Dynamic Fields, Figure 2 Local tracking of salient spatial events, such as splitting and merging



Geosensor Networks, Qualitative Monitoring of Dynamic Fields, Figure 3 Georesponsive sensor network, with increased node activation in vicinity of the boundary of large region

vate and deactivate sensors in response to changes in the dynamic field. Figure 3 illustrates the idea, where sensors in the vicinity of the boundary of the region of high temperature are activated to increase spatial resolution in those areas. Elsewhere, sensors are deactivated to increase sensor node lifetimes. Qualitative rules for achieving such behavior can be constructed based purely on qualitative spatial representations, like the combinatorial map (see [1] for further information).

Key Applications

Dynamic spatial fields are of interest across an enormous variety of environmental applications, including meteorology, land cover change, marine science, water resources management, defense, and emergency management and response. In general, applications of qualitative monitoring of dynamic spatial fields can fall into two broad categories. One category of use can be characterized as natural resource management, where decision makers use information gathered by GSN to manage scarce or fragile natural resources. Qualitative monitoring can help provide salient information to decision makers in a form that is more understandable and compatible with human conceptualization of dynamic spatial processes. Ultimately, such information can contribute to improved decision making. A second category of use can be characterized as scientific investigation of natural resources, where GSN are used by scientists to gather more detailed information about the environment than possible with conventional data logging techniques. In such cases, qualitative monitoring can assist in filtering data, screening out irrelevant data and highlighting high-level events of interest that can subsequently be investigated more closely.

Future Directions

As a relatively young area of study, qualitative monitoring of dynamic spatial fields has many important directions for future study, including:

- Sensor mobility: Although regions of a dynamic spatial field may be regarded as mobile, currently sensor nodes within the GSN are typically assumed to be static. Sensor mobility adds another layer of complexity to designing geosensor networks, which qualitative approaches are ideally suited to dealing with.
- Heterogeneity and multi-tasking: Current GSN usually comprise one type of sensor node engaged in a sin-

gle task. Future GSN will need to enable different types of node to interoperate on a variety of tasks, requiring the capability to integrate multiple qualitative queries across multiple different node types.

- Vagueness: Vagueness concerns the existence of boundary cases in information. Many qualitative concepts can be treated as vague and lacking precise boundaries (for example a "hot" region might be regarded as a vague spatial concept, if the boundary between "hot" and "not hot" is not precisely defined). Qualitative reasoning in the presence of vagueness remains an important challenge.
- Ubiquity: The goal of GSN is ultimately to embed spatial intelligence within the environment. Making GSN ubiquitous, unseen helpers that blend seamlessly requires the ability to manage complexity at every system level. Qualitative approaches provide one component of that complexity management, but further tools are required.

Cross References

- ► Distributed Geospatial Computing (DGC)
- ► Geosensor Networks, Estimating Continuous Phenonena
- ► Geosensor Networks
- ► Geosensor Networks, Formal Foundations
- ► Localization, Cooperative

Recommended Reading

1. Duckham, M., Nittel, S., Worboys, M.: Monitoring dynamic spatial fields using responsive geosensor networks. Shahabi, C., Boucelma, O. (eds.) ACM GIS 2005, pp. 51–60. ACM Press, New York (2005)
2. Egenhofer, M.J., Franzosa, R.D.: Point set topological relations. Int. J. Geogr. Inf. Syst. **5**, 161–174 (1991)
3. Frank, A.: Qualitative spatial reasoning about distances and directions in geographic space. J. Visual Lang. Comput. **3**, 343–371 (1992)
4. Freksa, C.: Using orientation information for qualitative spatial reasoning. In: Andrew U.F., Campari, I., Formentini, U. (eds.) Theories and Methods of Spatio-temporal Reasoning. Lecture Notes Computing Science vol. 639, pp. 162–178. Springer, Berlin (1992)
5. Galton, A.: Qualitative Spatial Change. Oxford University Press, Oxford (2000)
6. Hernández, D., Clementini, E., Felice, P.D.: Qualitative distances. In: Frank, A.U., Kuhn, W. (eds.) Proc. COSIT 95, Lecture Notes in Computer Science, vol. 988, pp. 45–57. Springer, Berlin (1995)
7. Randell, D.A., Cui, Z., Cohn, A.G.: A spatial logic based on regions and connection. In: Knowledge Rerpresentation, pp. 165–176. Morgan Kaufmann, San Francisco (1992)
8. Worboys, M.F., Duckham, M.: Monitoring qualitative spatiotemporal change for geosensor networks. Int. J. Geogr. Inf. Sci. **20**(10), 1087–1108 (2006)
9. Worboys, M.F.: Nearness relations in environmental space. Int. J. Geogr. Inf. Sci. **15**(7), 633–651 (2001)

Geospatial Analysis

► Data Analysis, Spatial

Geospatial Authorization

► Security Models, Geospatial

Geospatial Authorizations, Efficient Enforcement

VIJAY ATLURI, HEECHANG SHIN
Department of Management Science & Information Systems, Rutgers University, Newark, NJ, USA

Synonyms

Unified index scheme; Resolution-based matrix quadtree; Spatio-temporal-authorization-driven R-tree; Secure time-parameterized tree; Secure past, present and future tree

Definition

Enforcing security often incurs overheads, and as a result may degrade performance. The problem is exasperated more in geospatial data, which includes, among other things, geospatial image data and moving-object data. Uncontrolled dissemination of geospatial image data may have grave implications for national security and personal privacy. This is because high resolution satellite imagery may be used to identify vital national resources. As a result, this could encourage industrial espionage, terrorism, or cross-border military attacks, and combination of publicly available personal data pools with high resolution image data. This, coupled with the integration and analysis capabilities of modern geographic information systems, can result in a technological invasion of personal privacy. Similarly, the location and tracking of mobile users, required in effective delivery of location-based services, also raises a number of privacy and security issues, because disclosing the location information of mobile users has the potential to allow an adversary to physically locate a person for malicious purposes, and the location history can be used to identify the user's private information such as health status, political affiliations and religious beliefs. Thus, it is essential to have appropriate access control mechanisms in place. Unlike conventional authorizations that can be managed and searched using access control lists, management and searching of geospatial authorizations require a suitable indexing structure since they possess spatiotemporal attributes. As such, serving an access request requires the searching of two indexes: the index for data objects and the index for authorizations. However, the response time can be improved by using a unified index to support the user access requests. Here, the recently proposed unified index schemes are introduced. Specifically, two types of unified index are presented: (1) unified index structures for geospatial images, the resolution-based matrix quadtree (RMX-quadtree) and the spatiotemporal-authorization-driven R-tree (STAR-tree), and (2) unified index structures for moving objects, the secure time-parameterized R-tree (^{S}TPR-tree) and the secure past, present and future tree (S^{PPF}-tree).

Historical Background

Until recently, the access control model has not taken the spatial dimension into consideration. However, the need to control geographical information has increased because high-resolution satellite images have become readily available for public with reduced costs, and new context-aware applications such as location-based services have been developed due to the proliferation of mobile devices and wireless technologies. The *Geospatial Data Authorization Model* (GSAM) [1] is the first access-control model for geospatial data in the literature, and the model controls access to satellite images based on spatial extent, temporal duration, resolution, and other attributes of images. GSAM has been extended in two directions: one direction is support of vector-based spatial data [2,3], and the other is support of mobile data [4,5].

However, most of the proposed access-control models do not support the index structure of authorizations for efficient processing of geospatial access requests: for each access request, all the security policies that have been issued to the user of the access request are linearly searched and evaluated. Thus, in an environments where a high volume of user requests needs to be processed very fast, such as traffic management systems, the security enforcement process becomes the main bottleneck. Only a few authorization models that support the index scheme can be found in the context of a mobile database. An index scheme for moving-object data and user profiles has been proposed in [4]. However, this does not consider authorization. An index structure has been proposed to index authorizations ensuring that the customer profile information be disclosed to the merchants based on the choice of the customers [5]. However, this provides separate index structures for data and authorizations.

The RMX-quadtree [6] is the first proposal in the literature that allows geospatial images to be indexed based on their resolutions as well as their spatial attributes. Although the performance of the RMX-quadtree is efficient, it has two limitations: (1) the shape of the images to be indexed is restricted to a square, and (2) it does not allow overlapping images for a given resolution. The STAR-tree [7] eliminates these limitations. It is a three dimensional variant of R-tree, and it allows any overlapped rectangular shapes of images to be indexed.

In terms of moving-object data, the ^{S}TPR-tree [8] is an extension of the TPR-tree [9], and it is the first proposed unified index structure in literature for both moving objects and authorizations. One main limitation of the ^{S}TPR-tree is that it can only support those security policies based on the current and near-future locations of moving objects. Because the security policies in mobile environment are based on the *past*, *present* and *future* statuses of moving objects, ^{S}TPR-tree cannot fully handle security policies such as *track* because the past location of moving objects is not being stored. This limitation is resolved in the S^{PPF}-tree [10].

Scientific Fundamentals

Authorizations are typically implemented either as access-control lists or capabilities lists for traditional types of data. However, authorization specifications on geospatial data include a spatial dimension (such as authorized geospatial extents) and a temporal dimension (such as valid duration of time for access). Thus, in order to efficiently identify the authorizations that are relevant to an access request, an index structure can be used. Because geospatial data is also organized using index structures, a unified index structure that holds both geospatial data and authorizations can be created for efficient processing of user requests. The basic idea encompassing these unified index structures is to devise an appropriate index structure for the geospatial data, and then overlay authorizations on the relevant nodes of the tree. The construction of the tree and the overlaying process are performed in such a way that the cost of an access-control request is minimal. In the following, two types of unified indexing structures are presented: one for geospatial images, and another for moving objects.

Unified Index Structures for Geospatial Images

RMX-quadtree RMX-quadtree is a variant of the matrix (MX) quadtree: the structure of the RMX-quadtree permits overlaying of geospatial authorizations over nodes of the MX-quadtree. Thus, access requests for geospatial images can be processed more efficiently because only one index is used for evaluation. In order to build RMX-quadtrees, the following assumptions are made: (1) the spatial region represented by each image is a square, (2) images with the same resolution level are non-overlapping, and (3) higher resolution images cover smaller spatial extents.

The search space in the RMX-quadtree is assumed to be square and it is recursively decomposed into quadrants; the quadrants are named northwest (NW), northeast (NE), southwest (SW), and southeast (SE). An authorization is stored into a node *N* if the authorization applies to all the possible images in *N*'s subtrees. The structure satisfies the following properties:

- Images of the same resolution are stored at the same level.
- Each index node includes fixed number of children nodes (NW, NE, SW, and SE).
- Images can be stored in an index node as a result of merging different levels of images. Only the images with the highest resolution are stored in the leaf nodes.
- Each geospatial image corresponds to a 1×1 square, and it can be the spatial extent of index nodes.
- The depth of the quadtree is predefined and the structure is independent of the order of insertion operations.

Authorization specifications on geospatial data include spatial attributes and privilege modes containing geospatial operations. Therefore, to efficiently identify the authorizations relevant to an access request, an index structure is necessary. However, the index structure built based on authorizations would be unbalanced, because given two regions, if one region allows access only at lower resolution and the other region allows access at higher resolution, then the resultant index would be skewed due to the larger number of authorizations in the higher resolutions. Instead of trying to construct a tree for authorizations, RMX-quadtree chooses a different strategy, which is to overlay authorizations on top of the geospatial index tree, MX-quadtree. The overlaying means that the relevant authorization is stored on a node; thus, each node in the tree includes one additional field that points to the group of overlaid authorizations. This overlaying strategy guarantees the RMX-quadtree as a balanced tree because the base tree, the MX-quadtree, is balanced, and authorizations are simply stored on the relevant nodes of the tree; some of nodes may include more than one authorization, but this does not change the balance nature of the tree.

RMX-quadtrees are constructed by (1) creating MX-quadtrees for each resolution level of geospatial images, (2) merging them into one tree, and (3) overlaying authorizations on the tree. The first step is to build MX-quadtrees for each resolution level. Because MX-quadtrees are designed for storing only point data, an image is represented with its center location and edge length. Then, the MX-quadtree for the highest resolution level becomes the base tree, and other MX-quadtrees are merged with the base tree by adding the image data into the index node of the base tree. This is possible because the structure of the MX-quadtree is independent of the order of insertion operations, which becomes the property of the RMX-quadtree. Thus, the different levels of index nodes in the base tree refer to the same spatial region as the leaf nodes of other RMX-quadtrees. The last step is overlaying authorizations on the merged tree. An authorization is overlaid on a node when it applies to all the possible images in its subtrees. The reasoning for this overlaying process is that if a subject is allowed to access an area, it is also allowed to access all the regions at the same level of resolution within it. Representation of images using points creates a challenge for the overlaying process: an image stored in a node may not correspond to the area covered by the node. This is because the tree is constructed based on point information (center location of images) rather than geospatial regions that the image covers. In order to handle this situation, the area covered by a node *N* is expanded by ($p/2$) where p is the edge length of *N*. The authorization is overlaid on *N* if, and only if, the spatial extent of the authorization fully encloses the expanded area of the node *N*. This overlaying process recursively traverses the tree and for each node, the above rule applies. The overlaying process halts when the traversal reaches the level of resolution of the authorization object or both of the spatial regions are disjoint. During the overlaying process, there are two possible cases. The first case is that an authorization is applicable to only parts of an image, and there exist no nodes in the tree that satisfy the overlaying rule; the second is that there is no spatial relationship between the authorization and the node, and thus, there is no reason to traverse down the tree further.

A user request for geospatial images includes the geospatial region of interest with the specific resolution level. The user request evaluation starts from the root node of RMX-quadtree. This process will fall into one of the following cases:

- Stopping evaluation of authorizations: this occurs when the node contains an authorization that includes the resolution level that the user requests, which means that the user is allowed access to all the images in the node's subtree, and therefore, no more authorizations need to be evaluated. However, the search process continues until it reaches the resolution level that the user requests, and returns the images that are enclosed in the spatial extent of the user request.
- Continuing to traverse the tree and check for more authorizations until reaching the node that stores same resolution level as the user requests.
- Halting the process; this happens when the geospatial region of the user request is disjoint with that of a node because all the images stored at the node are disjoint with the spatial region of the user request.

STAR-Tree A STAR-tree is a 3D variant of R-tree [11], and it is a unified index structure for geospatial images and authorizations similar to the RMX-quadtree. R-tree data structure is similar to B-tree, but different in the sense that it is extended to multidimensional space. The spatial extent of each node of R-tree is approximated by the minimum bounding rectangle (MBR) that encloses all its entries' spatial extents tightly, which implies that the spatial extent of the root node will enclose all the space that its children nodes embed. The node structure of R-tree can be summarized as follows: (1) a leaf node contains (a predefined number of) images and MBR, and (2) a non-leaf node contains (a predefined number of) child nodes (which can be another nonleaf nodes or leaf nodes) and MBR that tightly bounds all the MBRs of its child nodes.

By employing the capabilities of the R-tree, STAR-tree can index any rectangular shape of overlapping images, and therefore the native satellite images can be indexed without any preprocessing. In addition, it is capable of handling temporal attributes of the geospatial images: the capture time or the download time of images. Because STAR-tree can index spatiotemporal attributes of geospatial objects, security policies with a specific valid time interval for accessing the data can be supported.

To construct the STAR-tree, first the images of the same resolution are grouped together and a three-dimensional (3D) (x-, y-, and t-dimensions) R-tree is built for each group. Then, the tree with the highest resolution is taken as the base tree, and other trees are merged carefully in the order of their resolution, so that they do not violate the properties of the STAR-tree and until all the single-resolution trees are merged. Then authorizations are appropriately overlaid based on principles similar to those of the RMX-tree, essentially by comparing the bounding region of the node and the spatiotemporal region of authorizations. However, the process is more straightforward since STAR-tree is able to index 3D objects natively, and thus does not need to deal with the transformation between region and point as in the RMX-quadtree.

The following are the main properties of a STAR-tree:

- Unlike the RMX-quadtree, only leaf nodes store geospatial images.
- Images with the same resolution level are stored at the same level in STAR-tree.
- STAR-tree is an unbalanced tree, but this does not degrade performance because the tree operations (insert, delete, and search) will be processed only in the part of the tree from the root node to the leaf nodes with the specified resolution levels for tree operations, instead of the whole tree. This part of the tree is balanced with longer height compared to the R-tree for geospatial images for the specified resolution level.
- The best case for tree operations is $O(h)$, and $O(hm^{h-1}+N)$ is the worst case where h is the height of the STAR-tree, m the maximum number of entries that each node holds, M the number of authorizations overlaid on the tree, and N the number of images. Note that, since the RMX-tree is a more rigid tree structure that does not allow overlapping images, the cost is $O(k+M)$ where k is the height of the highest resolution and M the number of authorizations, which is not surprising.
- Similar to the RMX-tree, the images stored at a higher level in the tree will have lower resolutions. This is because images of lower resolution would cover more spatial region than those with higher resolution when their graphical size is the same.
- Compared to the R-tree, each node includes one additional field that points to the group of overlaid authorizations on the node.

To further improve the response time, authorizations can be preevaluated to compute the set of subjects involved in each authorization. Then the subjects associated with the authorizations overlaid on a node can be indexed using a B^+-tree. As a result, each node would include the B^+-tree whose key is a subject, and whose data is authorizations associated with the subject. This structure makes the evaluation process efficient because only the authorizations that are relevant to the user request would be evaluated instead of going through all the authorizations overlaid on the node.

Unified Index Structures for Moving Objects

^{S}TPR-Tree The ^{S}TPR-tree is an extension of the TPR-tree [9], and it is the first proposed unified index structure for both moving objects and authorizations in the literature. TPR-tree is a variant of R-tree and is designed to answer the queries for supporting present and future locations of moving objects. Because the locations of moving objects are constantly updated, the main challenge for a moving-object database is to minimize the updating cost. For this purpose, in the TPR-tree the moving object is represented as its initial location and its constant velocity vector; thus, a moving object is updated only if it deviates more than the specified tolerance level. This will reduce the necessity for frequent updating.

In order to support the moving-objects queries, a time-parameterized rectangle (TPR) is used for the same purpose of MBR in R-tree: time is parameterized in MBR so that for a given time, all the trajectories of moving objects stay in TPR. Also, if it is required to insert a new moving object, TPR-tree finds the minimum volume enlargement of TPR between the insertion time and the predefined duration of time, called the time horizon, because as time elapses, the future locations of moving objects become less accurate. Thus, the time horizon guarantees the validity of query results. The node structure of a TPR-tree is similar to R-tree: a leaf node contains (a predefined number of) locations of moving objects (represented as the combination of reference position vector and velocity vector), and a nonleaf node contains (a predefined number of) child nodes, which can be other nonleaf nodes or leaf nodes.

The security policies specify the access control rules to profiles and location as well as movement trajectory information of mobile users, or to stationary resources based on the mobile user's location. Thus, either a subject or an object in an authorization specification can be a moving object, which is being indexed.

The ^{S}TPR-tree is constructed via a consecutive insertion operation into an initially empty tree, then overlaying of authorizations over the nodes of the tree. It employs the same overlaying and user request evaluation processes as that of STAR-tree except the facts that (1) it does not need to handle resolutions (it can be considered a STAR-tree with only one resolution), and (2) TPR is used instead of MBR for the overlaying procedure. The spatiotemporal extent of TPR is bounded because the valid time duration of the node is finite, i. e., from the current time to the future time by the time horizon. Therefore, the TPR of a node is considered similar to the MBR of STAR-tree for overlaying and user request evaluation processes, and thus, same procedures as those of STAR-tree can be used.

One main limitation of the ^{S}TPR-tree is that it can only support security policies based on the current and future locations of moving objects. Because the security policies in mobile environments are based on the *past*, *present* and *future* statuses of moving objects, ^{S}TPR-tree cannot fully handle security policies such as *tracking* because the past status of moving objects is not being stored. This limitation is resolved in the S^{PPF}-tree using the concept of partial persistence. Another limitation is that the ^{S}TPR-tree is capable of overlaying authorizations where either subjects or objects in an authorization are moving objects, but not at the same time. This is because mobile subjects and objects would be stored in different nodes of the index, and thus, supporting such authorizations' overlaying may require splitting the subject and object components of the authorization.

S^{PPF}-Tree The previously introduced ^{S}TPR-tree cannot support security policies based on *tracking* of mobile users. It is important to note that tracking information could also be sensitive and therefore security policies are often specified to reflect this. To efficiently enforce these policies, the tree must support this functionality in the sense that all the location history of moving objects is preserved. S^{PPF}-tree, an extension of ^{S}TPR-tree, can maintain past, present and future positions of moving objects along with authorizations, by employing partial persistent storage. Thus, it can support security policies based on tracking of mobile objects.

S^{PPF}-tree is a variant of R^{PPF}-tree [12], which applies the concept of the partial persistence to the TPR-tree in order to preserve the past location of moving objects as well. Partial persistence is the key concept of R^{PPF}-tree, in order to keep past positions as well as present and future positions of moving objects. Observe that there are two kinds of moving objects: one is currently moving objects so that their final location is predicted but not decided (called *alive* moving objects), and another type is objects that have already stopped moving, or have changed their velocity or anticipated future location above the predefined deviation level (called *dead* moving objects). During update (insertion or deletion) of moving objects in the tree, the leaf node where the update occurs are evaluated to see if there still exists a prespecified range of alive moving objects. If the number is out of this specified range, alive objects in the node are copied into a new node (called *time split*). The original node is used for evaluating the past positions of moving objects; the newly created node is for the present and future positions of moving objects, as in ^{S}TPR-tree. A similar process is applied to index nodes: in this case, the number of alive children nodes is checked if it is within the predefined range.

Because S^{PPF}-tree maintains past positions of moving objects as well, the overlaying process is more complicated than that of the ^{S}TPR-tree because authorizations are required to be maintained properly not only for present and future positions but also past positions: in the case of ^{S}TPR-tree, the tree is reconstructed after some reasonable duration of time, and authorizations are batch-overlaid on the tree. Thus, there is no need to deal with maintenance of authorizations during the tree's lifetime. Since the S^{PPF}-tree handles all the history information as well, it is necessary to maintain the overlaid authorizations more carefully in order not to violate the overlaying strategy. An authorization log is introduced to handle this situation; whenever an authorization is applicable to the tree, the authorization log overlays the newly applicable authorization on the alive nodes, and relocates the authorizations from the alive nodes to the dead nodes if they are only applicable to the dead nodes. An authorization log is a data structure constructed by spreading all the authorizations on the time line. As time elapses, a new authorization becomes applicable to the tree when the valid time duration of the authorizations is overlapped with the tree's valid time duration, i. e., between the current time and the time horizon. Then, the authorization log triggers an *auth_begin* event, which will overlay the authorization on the tree. On the other hand, certain overlaid authorizations become invalid when the valid time duration of the authorization is not applicable to the overlaid nodes. In this case, the authorization log triggers an *auth_end* event, which will remove the invalid authorizations from the overlaid nodes and reoverlay on the tree, because the removed ones may satisfy the overlaying conditions of other nodes in the tree. Also, an update must take care of the cases when the time-split occurs. Time-split creates a new node where some authorizations may be eligible to be overlaid on it. The authorization log supports a method, called *find-auth*, which computes all the authorizations overlapping with the valid interval of the newly created node. Then, the authorizations as a result of find-auth, will be overlaid on the new node if it meets the overlaying condition.

A user request evaluation is similar to that of ^{S}TPR-tree except that it can now evaluate a user request that also includes the tracking of moving objects, due to the functionality of holding all the updates history. In this case, only the nodes for which initial creation time and the time when time-split occurs, if time-splitted (otherwise, this time can be considered as current time) are overlapped with the time interval of the user request are evaluated.

Key Applications

Mobile Commerce Marketing

Owing to technological advances in mobile devices with wireless networks and positioning devices such as global positioning systems (GPS), customers' locations are used to provide customized services based on their current positions and movement histories. Mobile-targeted marketing or location-based advertising is one such example. Although there has been a general consensus that mobile-targeted marketing can provide more profits, most customers consider the advertised information spam unless they allow receiving advertisements. Therefore, the mobile service providers should employ a security scheme wherein users' preferences can be specified properly. Because customers are mobile, and their security policies are based on their current locations, tracking the locations of customers and enforcing security policies must be handled efficiently. Unified index schemes such as ^{S}TPR-tree and S^{PPF}-tree can handle the access request more efficiently than traditional approaches such as using two index structures (one for mobile customers and another for authorizations) because access-control requests are processed using just one index structure.

Selective Dissemination of High-Resolution Satellite Images

There are now more than 15 privately owned commercial satellites with resolutions from 1 to 30 m. For example, satellites such as IKONOS (launched in September 1999), ORBVIEW, EROS and QUICKBIRD are privately owned and provide images with resolution of 1 m or smaller. Uncontrolled dissemination of this information may have grave implications for national security and personal privacy, as some groups may exploit this information for aggressive purposes. Thus, formal policies for controlling the release of imagery based on geographical boundaries and resolutions of images have been proposed. However, the number of images being disseminated is tremendous. For example, in case of TerraServer-USA, the average daily image tiles transferred was 3,684,093 during 1998 and 2000. There is a need for effective and efficient schemes for facilitating controlled dissemination of satellite imagery. The proposed unified index schemes, such as RMX-quadtree and STAR-tree, can improve the response time because access requests can be processed more efficiently as only one index structure is used.

Securing Sensitive Resources

Physical location of individuals can be used to secure sensitive resources more effectively. For example, in order to gain the access to the secure repository room, an individual possessing relevant authorizations should be physically present in front of the room while she submits an access request. In order to do so, the location of mobile individuals and relevant security policies must be efficiently processed in large organizations. Unified index schemes such as ^{S}TPR-tree and S^{PPF}-tree can efficiently enforce security policies on mobile objects.

Traffic Management

Information on some traffic may be very sensitive knowledge to disclose because they are carrying dangerous materials. In this case, only an authorized group of people can locate and track the traffic. As the number of traffic and security policies increases, user requests must be handled efficiently. Unified index schemes such as ^{S}TPR-tree and S^{PPF}-tree can efficiently enforce security policies on the traffic management.

Future Directions

None of the above proposed unified index trees support negative authorizations. Providing such support is not trivial since they give rise to conflicts among the authorizations. Moreover, it may require changes to the fundamental assumptions used in the construction and access request evaluation. The overlaying strategy assumes only positive authorizations. Thus, an authorization is overlaid at as high a level as possible in the tree because as long as there exists an authorization that allows the user to access the given region, there will not exist any conflicting negative authorization that will not allow the user to access some parts of the allowed region. Based on this assumption, authorization evaluation halts whenever a relevant authorization is located during the traversal from the root node towards the leaf level. However, if negative authorizations are supported, all the authorizations overlaid on the traversal path need to be evaluated due to the possibility of conflicts among the authorizations: although an authorization that allows a user to access a region is overlaid in an index node, it is possible that another negative authorization that prohibits the user to access a part of the region may exist in the leaf node. Therefore, in order to support negative authorizations, it is necessary to create another copy of the data index and overlay positive authorizations on one index tree and negative authorizations on the other. Then, during the user request process, the result set from the positive authorization index is filtered out by the result set of the second negative authorization index.

Moreover, the ^{S}TPR-tree and S^{PPF}-tree cannot handle overlaying authorizations whose subjects and objects are both moving. As a result, supporting such authorizations' overlaying may require splitting the subjects and objects components.

Cross References

- ► Indexing Schemes for Multi-dimensional Moving Objects
- ► Indexing the Positions of Continuously Moving Objects
- ► Information Services, Geography
- ► Location-Based Services: Practices and Products
- ► Mobile Object Indexing
- ► Privacy Threats in Location-Based Services
- ► Privacy and Security Challenges in GIS
- ► R*-tree
- ► Raster Data
- ► R-Trees – A Dynamic Index Structure for Spatial Searching
- ► Security Models, Geospatial

Recommended Reading

1. Atluri, V., Chun, S.: An authorization model for geospatial data. IEEE Trans. Dependable Sec. Comput. **1**, 238–254 (2004)
2. Belussi, A., Bertino, E., Catania, B., Damiani, M., Nucita, A.: An authorization model for geographical maps. In: GIS Washington, DC, USA, 12–13 Nov 2004
3. Bertino, E., Damiani, M., Momini, D.: An Access Control System for a Web Map Management Service. In: RIDE, Boston, MA, USA, 28–29 March 2004
4. Atluri, V., Adam, N., Youssef, M.: Towards a unified index scheme for mobile data and customer profiles in a location-based service environment. In: Proceedings of theWorkshop on Next Generation Geospatial Information (NG2I'03), Cambridge, MA, USA 19–21 Oct 2003
5. Youssef, M., Adam, N., Atluri, V.: Preserving mobile customer privacy: an access control system for moving objects and customer information. In: International Conference on Mobile Data Management, Ayia Napa, Cyprus, 9–13 May 2005. Lecture Notes in Computer Science (2005)
6. Atluri, V., Mazzoleni, P.: Uniform indexing for geospatial data and authorizations. In: DBSec, Kings College, Cambridge, UK, 28–31 July 2002
7. Atluri, V., Guo, Q.: STAR-Tree: An index structure for efficient evaluation of spatiotemporal authorizations. In: IFIP TC11/WG 11.3 Eighteenth Annual Conference on Data and Applications Security, Sitges, Catalonia, Spain, 25–28 2004
8. Atluri, V., Guo, Q.: Unified index for mobile object data and authorizations. In: ESORICS, Milan, Italy, 12–14 2005
9. Saltenis, S., Jensen, C., Leutenegger, S., Lopez, M.: Indexing the positions of continuously moving objects. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, Dallas, Texas, USA, 14–19 2000
10. Atluri, V., Shin, H.: Efficient enforcement of security policies based on tracking of mobile users. In: DBSec, Sophia, Antipolis, France, 31 Aug 2006
11. Guttman, A.: R-trees a dynamic index structure for spatial searching. In: ACM SIGMOD Conference, Boston, Massachusetts, 18–21 1984
12. Pelanis, M., Saltenis, S., Jensen, C.S.: Indexing the past, present and anticipated future positions of moving objects. ACM Trans. Database Syst. **31**(1), 255–298 (2006). doi: 10.1145/1132863.1132870

Geospatial Computational Grid

► Grid, Geospatial

Geospatial Computing Grid

► Grid, Geospatial

Geospatial Data Alignment

► Conflation of Geospatial Data

Geospatial Data Grid

► Grid, Geospatial

Geospatial Data Reconciliation

▶ Conflation of Geospatial Data

Geospatial Metadata

▶ Metadata

Geospatial Ontology

▶ Geospatial Semantic Web: Applications
▶ Geospatial Semantic Web, Interoperability

Geospatial Semantic Integration

NAIJUN ZHOU
Department of Geography, University of Maryland
College Park, College Park, MD, USA

Synonyms

Interoperability; Semantic Web; Ontology; Similarity, semantic; Features, linguistic; Features, Physical; Similarity discovery; Similarity Representation; Category, geographic; RDF; OWL; Geospatial semantics

Definition

Semantics refers to the meaning of symbols such as words, graphs, and other representations that are employed to describe a real-world object. A geospatial object is defined with its spatial distribution and its attributes. Geospatial semantics indicate the meaning of geospatial terms as attributes of geospatial objects. Geospatial semantic integration provides a global view of diverse terms in different data sources. For example, a geospatial object is described as a *river* in one data source and a *water body* in another source; geospatial semantic integration reveals that the two terms fundamentally mean the same thing in the real world and merge them into a single term such as water body.

Historical Background

Semantic integration, as one task of data integration, has been investigated in the database community since database technology was introduced in the 1960s. Geospatial semantics is concerned with the geospatial representation of reality that is a fundamental question of geospatial information theory. Recently, geospatial semantic integration has received increasing attention because more and more geospatial applications involve diverse data sources employing different terms to describe the same or similar geospatial objects and because new technologies (e. g., Semantic Web, emergency management) demand efficient machine-based communication among data sources. A special meeting, *Interoperating GISs* [1], was held in 1997 with a section focused on geospatial semantics. Early research on semantic integration is found in [2,3,4].

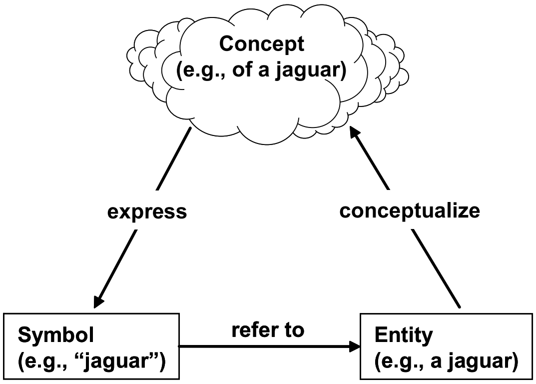
Scientific Fundamentals

Geospatial Semantics

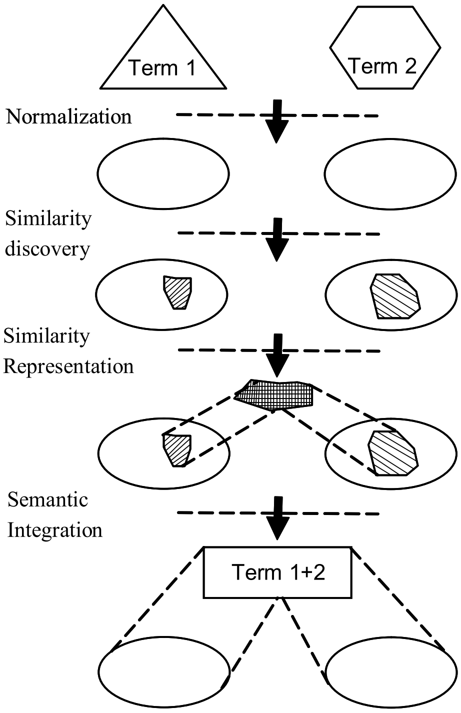
When humans describe and present an object, the object is first conceptualized and then is expressed in terms (Fig. 1). Therefore, the study of geospatial semantics involves the fields of spatial ontology and linguistics. Spatial ontology studies how human beings conceptualize an object, and linguistics investigates how the concepts are represented (e. g., as words).

Methods of Geospatial Semantic Integration

The primary tasks of geospatial semantic integration include understanding the semantics of geospatial terms and measuring their semantic relationships. Geospatial semantics can be obtained from the terms directly, or from additional information as the semantic definition of terms. It is well accepted that semantic integration should be based on semantic definition. Geospatial semantic integration can take a four-step approach: normalizing semantic definitions, measuring semantic similarity, representing and maintaining semantic similarity, and merging terms (Fig. 2) [6]. Semantic normalization may regulate the syntax of semantic definitions and may require that the definitions use linguistic words from the same thesaurus. Semantic similarity is a measure of semantic relationships, and the similarity can be quantitative (e. g., the probability that two terms indicate the same objects), or be qualitative values (e. g., superset, subset, equal, unequal, etc.).



Geospatial Semantic Integration, Figure 1 Semantic triangle [5]



Geospatial Semantic Integration, Figure 2 The four phases of semantic integration (adopted from [6]). Ellipses indicate semantics of terms

One geospatial semantic integration method deploys a feature-based approach originally proposed in psychology and cognitive science [7]. In addition to a geospatial term, a geospatial object also has a set of distinguishing features. Because the features define the meaning of the object, the comparison of the features provides semantic similarity of terms. The features can be physical or linguistic properties of geospatial objects. An example of physical features can be the *part*, *function* and *attributes* [8] of buildings; and two specific types of buildings (e. g., *stadium*, *theater*) can be compared and integrated based on their features. Linguistic features can be the words *branch*, *brook*, *canal*, *creek*, *ditch*, *gutter*, *river*, *rivulet*, *spillway* as the semantic definition of a *watercourse* (Spatial Data Transfer Standard, SDTS). The semantic similarity of the feature-based approach primarily measures how many common features two objects share [8,9,10].

Geospatial semantic integration can also be based on a common data model. Major common data models include Open Geospatial Consortium data standards, FGDC (Federal Geographic Data Committee) Content Standard for Digital Geospatial Metadata, SDTS, WordNet, etc. The common data model approach defines a common thesaurus containing the semantic definition of geospatial terms. Geospatial attributes can only take values from the predefined thesaurus from which semantic similarity is derived.

Ontology has been used in geospatial semantic integration [11]. Taxonomy is a light version ontology and is a major mechanism of defining geospatial terms. The use of taxonomy in geospatial semantic integration is supported by the theory of geographic category. Based on the theory of categorization, the perceived world is made up of structured information (i. e., taxonomy), and humans' categorization aims to reduce the differences among concepts to behaviorally and cognitively usable proportions. Geospatial semantic relationships are derived from taxonomy or from ontology mapping.

Geospatial semantics and ontology can be formally represented in XML, RDF, OWL, etc. Such representations enable the semantic communication among data sources to achieve automated semantic integration.

Key Applications

Geospatial Ontology Mapping

Geospatial ontology usually consists of a set of terms. Semantic integration provides a solution to the semantic relationships of the terms in different ontologies. In fact, semantic integration sometimes is also called ontology mapping (integration).

Geospatial Semantic Web

The communication of different ontologies and semantics are the primary tasks of the geospatial Semantic Web. Automated geospatial semantic integration enables machines to understand and exchange meanings and, thus, can help achieve the goals of the Semantic Web.

Geospatial Data Portals

Geospatial data portals allow users to search geospatial data sources that may use different terms for the same objects. Semantic integration provides a global view of diverse terms by comparing and merging terms. Such global view improves the efficiency and accuracy of data portal service.

Geospatial Database Integration

Geospatial database integration needs to merge schema and domain values [12,13]. The schema integration actually is also a task of semantic integration of database attributes.

Future Directions

Geospatial semantic integration is a much debated research area. There has not been a consensus on the definition of semantic integration. Current research and application activities mainly rely on the results from fields such as artificial intelligence, databases, cognitive science, linguistics, etc. Research is needed to address the uniqueness of geospatial information in defining and comparing the semantics of geospatial objects. There are two complimentary efforts to achieve geospatial semantic integration: defining common semantic models that regulate the use and exchange of semantics (top down approach), and developing automated methods to compare diverse semantics such that semantics can be defined independently by data providers. Finally, in addition to geospatial attributes, geospatial semantics should study geospatial expressions such as spatial topology and spatial operations; for example, the semantics of *close* should be interpreted as a certain spatial distance.

Cross References

► Geospatial Semantic Web

Recommended Reading

1. NCGIA (1997) Interoperating GISs. Report of a Specialist Meeting Held under the Auspices of the Varenius Project. December 5–6, 1997, Santa Barbara, California. http://ncgia.ucsb.edu/conf/interop97/interop_toc.html. Accessed 11 Sept 2007
2. Bishr, Y.: Overcoming the semantic and other barriers to GIS interoperability. Int. J. Geogr. Inf. Sci. **12**(4), 299–314 (1998)
3. Sheth, A.: Changing focus on interoperability in information systems: from system, syntax, structure to semantics. In: Goodchild, M., Egenhofer,M., Fegeas, R., Kottman, C. (eds.) Interoperating Geographic Information Systems, pp. 5–69 Kluwer Academic Publishers, Norwell, MA (1999)
4. Harvey, F., Kuhn, W., Pundt, H., Bishr, Y.: Semantic interoperability: A central issue for sharing geographic information. Ann. Reg. Sci. **33**, 213–232 (1999)
5. Warner, K.: Geospatial Semantics: Why, of What, and How?. J. Data Semant., Lect. Notes Comp. Sci. **3534**, 1–24 (2005)
6. Kalfoglou, Y., Hu, B., Reynolds, D., Shadbolt, N.: Semantic Integration Technologies Survey. CROSI Project, University of Southampton, Technical Report, E-Print No. 10842 May (2005)
7. Tversky, A.: Features of similarity. Psychol. Rev. **84**(4), 327–352 (1977)
8. Rodriguez, M., Egenhofer, M., Rugg, R.: Comparing geospatial entity classes: An asymmetric and context-dependent similarity measure. Int. J. Geogr. Inf. Sci. **18**(3), 229–256 (2004)
9. Kokla, M., Kavouras, M.: Fusion of top-level and geographical domain ontologies based on context formation and complementarity. Int. J. Geogr. Inf. Sci. **15**(7), 679–687 (2001)
10. Zhou, N.: A study on automatic ontology mapping of categorical information. In: Proceedings of National Conference on Digital Government Research, pp. 401–404. 18–21 May, Boston (2003)
11. Fonseca, F., Egenhofer, M., Davis, C., Camara, G.: Semantic granularity in ontology-driven geographic information systems. Ann. Math. Artif. Intell. **36**(1–2), 121–151 (2002)
12. Ziegler, P., Klaus, R., Dittrich, K.:Three Decades of Data Integration – All Problems Solved?". Paper presented at the 18th World Computer Congress, Toulouse, France, 22–27 Aug 2004. http://arvo.ifi.unizh.ch/dbtg/Staff/Ziegler/papers/ZieglerWCC2004.pdf. Accessed 11 Sept 2007
13. Doan, A.,, Halevy, A.Y.: Semantic Integration Research in the Database Community: A Brief Survey. AI Magazine **26**(1), 83–94 (2005)

Geospatial Semantic Interoperability

► Geospatial Semantic Web, Interoperability

Geospatial Semantic Web

FREDERICO FONSECA
College of Information Sciences and Technology,
The Pennsylvania State University,
University Park, PA, USA

Synonyms

Semantic web; Spatial semantic web; Ontologies

Definition

The Geospatial Semantic web addresses the geographic and spatial components of the Semantic Web, which is a project to create the infrastructure necessary for information sharing on the World Wide Web. The sharing should be based on the meaning of the documents. Since the interpretation of meaning is being done by computers another crucial issue is that whatever way this meaning is expressed it has to be machine-readable. Egenhofer highlights the need for a simple canonical form to pose geospatial data queries and methods to evaluate the semantics of the geospatial data sources to fit them to geospatial queries.

Historical Background

In his vision paper about the Semantic Web, Tim Berners-Lee (2001) gives the following example: “at the doctor’s office, Lucy instructed her Semantic Web agent through her handheld Web browser. The agent promptly retrieved information about Mom’s prescribed treatment from the doctor’s agent, looked up several lists of providers, and checked for the ones in-plan for Mom’s insurance within a 20-mile radius of her home and with a rating of excellent or very good on trusted rating services.”

Both space and time can be found in this initial example. These are questions that are familiar to many geographic information science researchers. The future of the Semantic Web includes not only the geographic component but also the spatial component. For instance, Egenhofer addresses the combination of geospatial data with other spatial operators such as *in* for instance. Another perspective for the web user is to look exclusively for geospatial data. This second type of user is looking for images, maps, spatial databases, and tables in general. In this case, the geospatial data was previously classified as such. A third and widely available geospatial information resource on the web is the geographic descriptions present in personal web pages. The geospatial semantic web should be able to combine all these aspects of geographic information on the Web in a meaningful way.

Scientific Fundamentals

The main challenge for researchers in the geospatial semantic web is how to represent in computers what is special about spatial. Concepts of space and time that are intuitive for humans need to be carefully formalized in order to be used by computers.

In order to achieve the objective of the semantic web of exchanging information across disciplines with different world-views, it is necessary to have an agreement on the meaning of words—the categories, relations and instances that form parts of a mental model that in turn represents an understanding of the world. Such an explicit formalization of mental models is called an ontology. The basic description of the real things in the world—the description of what would be so-called objective truth—is called Ontology (with an upper-case O). The result of making explicit the agreement within communities is what the Artificial Intelligence community calls ontology (with a lower-case o). Therefore, there is only one Ontology, but many ontologies.

A geographic ontology has to provide a description of geographical entities, which can be conceptualized in two different views of the world. The field view considers spatial data to be a set of continuous distributions. The object view conceives the world as occupied by discrete, identifiable entities.

Web Services

The W3C definition of Web services is computer application that enables interoperable machine-to-machine interaction over a network. Such an application needs an interface described in a machine-processable format so that other application can interact with the service through the exchange of messages. The OpenGIS consortium has a standard for web services called OGC Web Services (OWS). OWS will define standards and procedures to create interoperable services that range from the acquisition of geospatial data from sensors to the creation of common infrastructure for decision support. Currently it has six major initiatives: (1) Sensor Web Enablement (SWE), designed to integrate a variety of sensors, platforms and management infrastructure into a single enterprise. The main focus of this initiative is on integration of physical sensors and simulators within a realistic operating environment; (2) Geo Processing Workflow (GPW), will interconnect geo-processes following workflow requirements; (3) Information Communities and Semantics (ICS) deals with complex geographic information types. The types will be modeled in UML and GML. After being catalogued, they will be available to be used in geo-processing workflows; (4) Agile Geography is targeted at the integration of processes and the enabling of services that show the possibilities of interoperability and service-oriented architectures using OGC Web Services. This initiative will apply links, bookmarks and Web pages to digital cartography and geospatial information management. It will also enable the integration of user initiatives interested in collaborating on the maintenance of a shared geospatial data sets; (5) Compliance Testing (CITE) will check for proof of compliance of Web Coverage Service and Catalog Service for the Web with OGC specifications. CITE will also require reference implementations of the specifications. Reference implementations are developed using open source standards; (6) CAD / GIS / BIM (CGB) has as an objective bridging the information models and workflows of the various communities involved with the 3-D representation cities. Future uses include location-based services and services of emergency preparedness and response. For this to be feasible it will be necessary to create a services-based architecture integrate information and make it accessible in a secure way.

This shows again the importance of the integration between OpenGIS and the W3C. The geospatial semantic web will need to build a semantic layer over the standards that the OpenGIS is developing or have already deployed.

Geography Markup Language

The OpenGIS Consortium has also created a language based on XML to store and exchange geographic data. The Geography Markup Language, GML (see http://www.opengis.org/docs/02-023r4.pdf for additional details), is able to encode OpenGIS geographic features. According to OGC, GML is “an XML grammar written in XML Schema for the modelling, transport, and storage of geographic information.”

GML will be able to formally describe geographic features, coordinate reference systems, geometry, topology, time, and units of measure. OGC’s view is based on features, which are abstractions of geographic phenomena associated with positions on the surface of the Earth. A computer representation of the Earth is a set of features.

Each feature can be described by a set of properties. Features may represent both vector data and coverages thus enabling a broader spectrum of information integration.

Queries

There are two important aspects to the queries created by users of the geospatial semantic web: the form and the presentation of the results. Egenhofer suggests a canonical form to pose geospatial queries:

<geospatial request> ::= <geospatial constraint>
[<logical connective> <geospatial request>]
<geospatial constraint> ::= <geospatial term>
<geospatial comparator><geospatial term>
<geospatial comparator>
::= ! based on the geospatial-relation ontology used
<geospatial term>
::= <geospatial class> | <geospatial label>
<geospatial class>
::= ! based on a geospatial feature ontology
<geospatial label> ::= ! based on a geospatial gazetteer

The second aspect in a geospatial query is how to represent the results. Depending on the nature of the query the result may be only an address such as an URI or real spatial data. What kind of spatial data and how to represent it is a question that an ontology of spatial data can help to answer. Again ontologies come into the picture. The solution for most semantic web problems involves the user committing to an ontology. This means that a user relies on some previous compilation of some explanation of facts that he/she thinks is truthful. In the query problem, the user will accept as a representation of the answer the representation of geographic data as stated in a specific ontology.

Geospatial Semantic Web Interoperability Experiment (GSWIE)

As a way of gathering all its initiatives, the Open Geospatial Consortium is developing the Geospatial Semantic Web Interoperability Experiment (GSWIE). Accoding to OGC, the GSWIE "will address several important steps towards the development of a Geospatial Semantic Web (GSW), where discovery, query, and consumption of geospatial content are based on formal semantic specification. Many pieces of the GSW puzzle have been worked extensively in the last several years; this experiment aims to augment WFS/FE with a semantic query capability, through the definition of an ontology for the geospatial intelligence community." The experiment will address the creation of formal geospatial ontologies. It will also create and test service interfaces that can reference those ontologies and accept query requests in a semantic query language. It will be necessary to test also tools that will support the generation of semantically expressed geospatial information. The main goal of the experiment is to show that users may be able in the future to pose semantic queries to spatial databases of diverse sorts and receive meaningful answers in response.

Key Applications

The Geospatial Semantic Web is intended to handle information that is meaningful both for human and computers. Therefore, the way potential users understand and use information is very important. In particular, there are three basic dimensions for users of geographic information in the Geospatial Semantic Web context:

1. Professional: highly structured geographic information stored in geographic databases which are indexed, stored, or described in the Web;
2. Naïve: the retrieval of unstructured, subjacent, informal geoinformation in the Web;
3. Scientific: geographic information science papers, models, and theories available on the Web.

Professional: In order to improve the results of queries looking for information stored in geographic databases it is necessary to support better definition for spatial concepts and terms used across different disciplines and the development multiple spatial and terminological ontologies. Here there are also web services: the most common examples are the called location services: through the use of spatial databases, those services are able to locate restaurants, hotels, and other facilities depending on the user's location.

Naïve: This concept is based on Naïve Geography. This case looks for geographic information in the text of web pages. It can be pages with description of geographic features such as cities, geographic phenomena, or facilities. The user may be looking for complementary information or it can even be the main source of data.

A case in which it is possible to mix the content of textual web pages with more structured data such as the geographic distribution of machines on the Web is the ability to answer queries such as "I found this interesting web page, where is its geographic location?" or "Find other web sites that contain information about places close to places mentioned in this web site" or "List (or even graphically display) all the location information on the IBM web site (offices, research centers, etc.)". The user may also want to find pages that describe a special region of San Francisco and uses as a constraint that the web server in which the page is stored is located in San Francisco because he or she thinks that locals will give better naïve information.

Scientific: Here the situation is similar to what Google Scholar does today: a specialized search engine for scientific papers in a specific domain. Besides being able to index GIScience documents it is also necessary to create ontologies that express the different theories expressed in the papers.

Future Directions

The increasing generation and availability of geographic information brings with itself the possibility of using the Web as an easy and immediate channel for its distribution. For the semantic web to work it is necessary to efficiently index, retrieve, and integrate information on the web is the Semantic Web and its geographic counterpart, the Geospatial Semantic Web.

But it is important to recognize that the semantic web requires a level of organization and formalization that is not here yet. Furthermore, there is currently a wealth of geographic information on the Web that is available in different forms ranging from images, maps, spatial databases, and tables to locations of restaurants, description of landscapes, and informal reports of bird watching activities.

Research on the sharing and integration of geographic information through the Web vary from data distribution policy and legal issues, interfaces, performance, information retrieval, mapping, to information brokering and ontologies. Research on the Semantic Web is focused on building highly formalized representations of the available information resources called ontologies. A variety of tools and languages are currently available. But to capture the richness of the available geographic information on the Web it is necessary more than that. It is necessary to create representation of information resources that are more loosely organized. Another important step is how to link and map these more informal resources to the highly formalized representation structures being built by the Semantic Web. The result will be a broader spectrum of data available for users of geographic information. The specific context for this research relates to the understanding of the different kinds of geographic information on the Web and the corresponding representations (geo-ontologies) of these resources.

Recommended Reading

1. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web: A New Form of Web Content That Is Meaningful To Computers Will Unleash A Revolution Of New Possibilities. The Scientific American. **284**(5), 34–43 (2001)
2. Egenhofer, M.J.: Toward the Semantic Geospatial Web. Tthe Tenth ACM International Symposium on Advances in Geographic Information Systems, McLean, Virginia, pp. 1–4 (2002)
3. Fonseca, F.: From Geo-Pragmatics to Derivation Ontologies: New Directions for the GeoSpatial Semantic Web. Transactions in GIS. **11**(3), 313–316 (2007)
4. Fonseca, F., Egenhofer, M., Agouris, P., Câmara, G.: Using Ontologies for Integrated Geographic Information Systems. Transactions in GIS. **6**(3), 231–257 (2002)
5. Goodchild, M.: Geographical Data Modeling. Computers and Geosciences. **18**(4), 401–408 (1992)
6. Smith, B., Mark, D.: Ontology and Geographic Kinds. International Symposium on Spatial Data Handling, Vancouver, BC, Canada, pp. 308–320 (1998)
7. The Open Geospatial Consortium White Papers at www.opengeospatial.org/pressroom/papers

Geospatial Semantic Web: Applications

YASER BISHR
Image Matters LLC, Leesburg, VA, USA

Synonyms

Geospatial Web 3.0; Geospatial semantics web; Geospatial ontology

Definition

The semantic web is an extension of the current web. It is a vision for the future of the web, in which information is given explicit meaning, making it easier for machines to automatically process and integrate information available on the web [1]. The geospatial semantic web adds space and time dimensions to the semantic web. "Web content" is used here to refer to network-accessible distributed digital contents. A geospatial semantic web application is any software that takes advantage of the geospatial semantic web. Ontology is a fundamental building block of the semantic web in general and the geospatial semantic web in particular. Ontology is a partial specification of conceptualization [2,3]. Ontology can be thought of as a data model about a domain of interest. The World Wide Web Consortium (W3C) developed and adopted the Web Ontology Language (OWL) as the standard language to encode and share ontologies.

Enriching geospatial contents with semantics is a two-step process. The first step is to create a geospatial ontology that captures the semantics of the domain. Geospatial ontology allows for definition and reasoning about real world spatial objects by combining information from multiple sources. A geospatial semantic approach to building applications allows for machines to understand and exchange qualitative spatial concepts like "eastbound vehicle" "near the airport" or "above ground" and hence complements the quantitative analytical power of geographic information systems (GIS). It also allows the definition and inference of higher level concepts like "highway" from lower level concepts like "paved road with speed limit 65 miles per hour". These concepts can then be instantiated to refer to actual objects in the database. The second step is to tag database elements with the ontology concepts so that they are accessible through the ontology and consumed by runtime geospatial semantic agents. These agents provide the runtime reasoning, querying, transformation and integration services upon which one can build geospatial semantic applications.

Historical Background

Historically, the use of GIS required specialized skills and therefore was limited to a small group of experts. Since then, GIS vendors and users have focused on adding analytical features to their GIS applications and have paid less attention to issues like interoperability, information sharing and integration. This situation has been further compounded over the last 30 years by an exponential growth of geospatial content, which has led public and private organizations to demand more interoperability, integration, and the ability to serve geospatial knowledge, rather than mere data. Perhaps the first published work on geospatial semantics is [4],which developed the ability to query heterogeneous geospatial database objects that are semantically tagged through a consistent and formal ontology. Some of the potential benefits geospatial semantic applications can provide are greater consistency and accuracy from improved system-wide management, more efficient use and sharing of knowledge, reduced redundancy of data across the system, better use of departmental GIS resources and reduced maintenance and support costs.

In recent years, the proliferation of the web and the latest advancements in what are broadly known as web mashups, which combine contents from different sources, have paved the way for general web users who are not GIS experts to capture, store, process, and display information about space and time. Egenhofer [5] naïve geography approach was proposed to mark the beginning of a research community push to facilitate the use of spatial concepts in nontraditional GIS applications and by general users.

Scientific Fundamentals

Geospatial semantic web applications allow users to use ontologies as a means to map (or tag) underlying data elements and their hidden relationships as explicit ontology concepts and predicates. In this way, users are able to unify data format and semantics regardless of their source. Through a high-level knowledge management interface that hides the complexity of geospatial content and operations, users can access and manipulate heterogeneous data in a unified semantic representation. It is important to note that geospatial semantic web applications are relatively new, and only a limited number of commercial off the shelf products exist. Academic research on the other hand is highly active in this field. Generally speaking, geospatial semantic web applications can be discussed from two distinct viewpoints. In the first, the focus is on the general architectural patterns, while in the second, the focus is on the geospatial ontologies that are consumed by those applications.

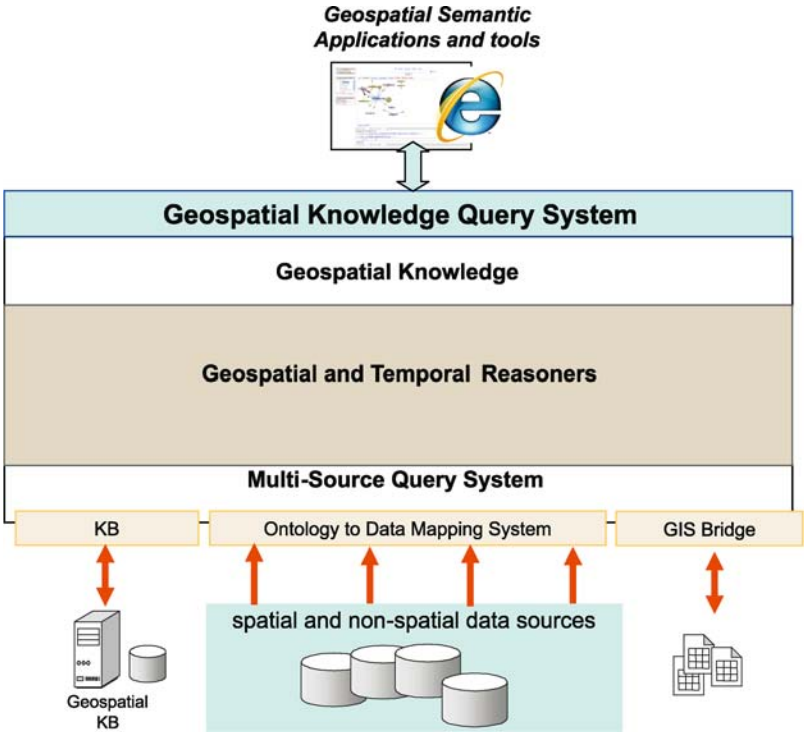
Geospatial Semantic Web Architecture View

A key benefit from geospatial semantic web technology is its ability to unify and fuse data semantics without resorting to changing the underlying data source schemas. This is contrasted to the prevailing data warehousing approach where relevant geospatial contents are mapped, transformed and migrated to a single large and data store. As depicted in Fig. 1, geospatial knowledge applications contain several components:

- Multisource inputs: a collection of semantically and syntactically heterogeneous and distributed spatial and non-spatial data sources.
- Ontology to data mapping system: a library of custom adapters providing consistent access to the underlying multi-source inputs. These adapters enable to explicitly map database features and their implicit relationships in and between the databases as ontological concepts and predicates respectively.
- GIS operators bridge: this is a component that facilitates the execution of traditional geospatial analytics. It returns the analysis results back the geospatial semantic system for further reasoning and querying.
- Multisource query engine: facilitates query processing and optimization; it translates queries based on the geospatial semantics to the underlying data sources.
- Spatial/temporal reasoners: a collection of spatial and temporal qualitative logical reasoners.
- Geospatial knowledge: the integrated virtual knowledge layer created from querying and reasoning about underlying data sources. This knowledge layer is specific to applications domains.
- Geospatial semantic query: provides semantic query interface so that applications can query to retrieve knowledge. This can trigger the reasoners and/or the multi-level query system. The W3C semantic WG has developed SPARQL for this purpose.
- Knowledge base (KB): a persistent knowledge base store that usually stores ontology information.

Geospatial Ontology View

A common underpinning for the semantic web as it is envisioned is that it contain several languages as shown in



Geospatial Semantic Web: Applications, Figure 1

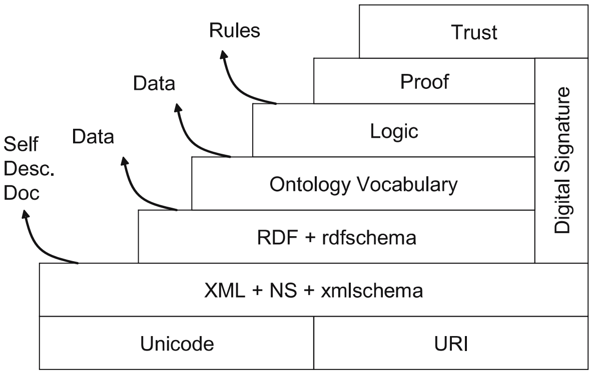
Fig. 2. The diagram depicts a semantic web architecture in which languages of increasing power are layered one on top of the other. The basis of a particular way of providing meaning for data is embodied in the model theory for the Resource Description Framework (RDF), OWL, and Logic Rules.

There are several proposals for describing space, spatial relationships, and relations between entities and their locations that have been developed within broad “ontological” frameworks. Several researchers have investigated the expressiveness of a number of publicly available ontologies. Similar to database schema design, those ontologies do not go beyond simple classifications and property associations. It is important to note here that developing ontologies must be executed with full knowledge of the kinds of inferences that geospatial semantic applications may require. To illustrate the above premise, an example of developing a model of nuclear facilities using XML schema (XML-S)and OWL is shown. Also shown are two different ways of representation in OWL which give completely different inference results.

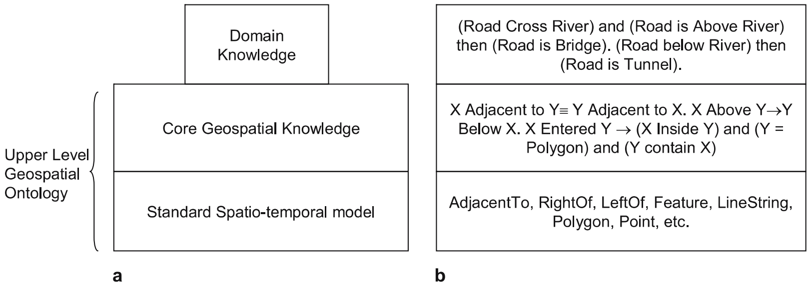
Example 1

```
<element name=“ElectricSubstation”
type=“ElectricSubstationType”/>
<complexType name=“ ElectricSubstation Type”>
<complexContent>
<extension base=“Facility”>
<attribute name=“Produces”
type=“anyURI”
default=“#Electricity ”/>
</extension>
</complexContent>
</complexType>
```

XML-S provides a standardized syntactical way to expose structural information. XML-S allows the definition of a schema for XML documents and provides some level of machine-understandable semantics of data. However, XML-S does not attach meaning to structural information.



Geospatial Semantic Web: Applications, Figure 2 Layers of the semantic web [Tim Berners Lee XML Conference, 2000]



Geospatial Semantic Web: Applications, Figure 3 Layered ontology for geospatial semantic web

As shown in Example 1, ElectricSubstation is a subclass of Facility and has a property Uses which has a default value: Coal.

Example 2

```
<owl:Class
rdf:ID=“ElectricSubstation”>
<rdfs:subClassOf
rdf:resource=“#Facility”/>
<owl:Restriction>
<owl:onProperty
rdf:resource=“#Produces”/>
<owl:hasValue
rdf:resource=“# Electricity”/>
</owl:Restriction>
</owl:Class>
```

Example 2 conveys the same semantics as defined by the XML-S and does not necessarily add any value to the already defined XML-S.

Example 3

```
<owl:Class
rdf:ID=“ElectricSubstation”>
<owl:intersectionOf
rdf:parseType=“Collection”>
<owl:Class
rdf:about=“#Facility”/>
<owl:Restriction>
<owl:onProperty
rdf:resource=“#Produces”/>
<owl:hasValue
rdf:resource=“#Electricity”/>
</owl:Restriction>
</owl:intersectionOf>
</owl:Class>
```

Example 3 shows an OWL representation, which essentially has added semantics. The members of the class are completely specified by the set operation. The example states that ElectricSubstation is the intersection of the class Facility and the set of things that produce Electricity. This means that if something produces Electricity and is a Facility, then it is an instance of ElectricSubstation. Without such a definition it would not be possible to know that Electricity Substations are facilities and they produce Electricity, but not vice versa.

The geospatial semantic web has at its core the basic semantic web layers, and it requires well-defined extensions to the ontology vocabulary and the logic layers. Constructing the core geospatial ontology is the most critical aspect for the success of geospatial semantic web. Application domains are expected to build their specific ontologies-based on the upper level ontology.

Figure 3 depicts a three-layered ontology for the geospatial semantic web:

- In the first layer a model of basic topologic, geometric types are defined, and the notion of features and feature properties are defined. A simple example of this layer is shown in Fig. 3b. The example above defines the concepts of *Adjacent, Above* and *Feature*. These concepts can be encoded in OWL.
- The default spatial inference and rules between features are defined in the core geospatial knowledge. A simple example of this layer is shown in Fig. 3b. Here *X Adjacent to Y ≡ Y Adjacent to X* can be encoded in OWL. *X Above Y → Y Below X and X Entered Y → (X Inside Y) and (Y = Polygon) and (Y contain X)* are conjunctive rules that can be encoded in SWRL.
- The third layer is the application layer. These are specialized ontologies for specific applications, e. g., roads, rail, environmental engineering, etc. A simple example of this layer is shown in shown in Fig. 3b. In this example, applications define the concepts of Road, River and Bridge. These concepts can be defined in OWL. Also, applications define *(Road Cross River) and (Road is Above River) then (Road is Bridge), (Road below River) then (Road is Tunnel)* as conjunctive rules that can be written in SWRL.

Qualitative Spatial and Temporal Reasoning

The most common approach in most geospatial systems (e. g., GIS) relies on computational models of space, which involves a simple rendition of Euclidean geometry to compute spatial relations. This approach ties the spatial relations of features to their geometric rendition. For example if a land parcel is represented as a point on a map, it is not possible to indicate that it shares an edge with a gas station. For this to occur, land parcels must be encoded in the GIS as aerial features. Mathematical theories of space, upon which most GIS are built, has points as the primary primitive of spatial entities. Lines consist of collection of points, and areas, or regions, are defined as sets of lines and points. Simmons [6] rightly mentioned that "No one has ever perceived a point, or ever will do so, whereas people have perceived individuals of finite extent."

For the geospatial ontology to be useful, it is necessary to have some capabilities that can reason with it. As mentioned earlier, spatial and temporal reasoners derive answers from a knowledge base. They provide a methodology for reasoning about the information in the knowledge base, and for formulating conclusions. The knowledge base will naturally be the ontologies as defined before and the actual data instances. The geospatial semantic web is the marriage between GIS and the semantic web. It therefore complements GIS quantitative computational power by providing spatial and temporal qualitative reasoning. The principal goal of qualitative spatial reasoning is to make common-sense spatial knowledge explicit so that, given appropriate reasoning techniques, a computer could predict, diagnose and explain the spatial behavior of real world concepts in a qualitative manner without recourse to quantitative methods. Geometry focuses on quantitative aspects of space. Topology, on the other hand, answers qualitative questions about spaces, for example adjacency and directionality.

Elements of Geospatial Ontologies

Context Is the statement: "All Birds Fly" always true? Penguins are birds but can't fly. This statement, however, might be entirely true if its context was "in Brazil". Here, context was added to the statement, the context being "in Brazil". Several domains have already elaborated their own working definition of context. In a human–machine interaction, a context is a set of information that could be used to define and interpret a situation in which agents interact. In the context-aware applications community, the context is composed of a set of information for characterizing the situation in which humans interact with applications and the immediate environment. Context is considered here as the set of assertions and conditions under which a set of axioms are true. The geospatial ontology framework should indicate that a given set of ontology axioms can only be determined within context

Ontologies are shared models of a domain that encode a view which is common to different communities. Context is a model that cast a local view of shared models, i. e., shared ontologies. Context can be considered as a filter that helps scope this subset of an ontology that is relevant to a given situation. Developing a theory of context for the geospatial domains must satisfy the following requirements:

- Context should allow a simpler formalization of axioms by defining the set of known conditions that are common to the stated axioms.
- Context should allow us to restrict the vocabulary and the facts that are used to solve a problem on a given occasion. This requirement will enable us to scope large ontologies to those subsets that are relevant to the problem at hand.
- The truth values of facts should be dealt with as dependent on a collection of assumptions which implicitly define context.
- There are no absolute, context-independent facts, namely each fact must be stated in the appropriate context.
- Reasoning across different contexts should be modeled. This will enable mapping between ontologies (in context), and hence semantic interoperability.
- A theory of geospatial context must consider the role of time, location, and other spatial and temporal aspects in determining the truth value of a given set of axioms.

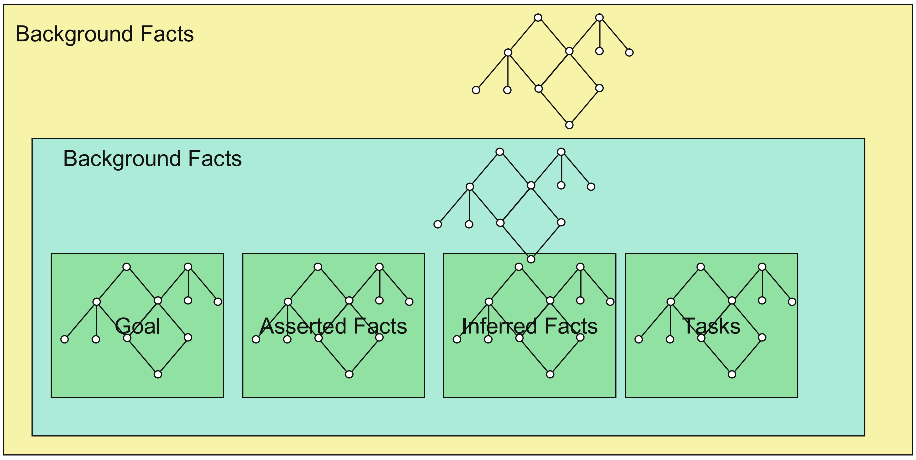
Identity Criteria (IC) IC determine *sufficient* conditions in determining the identity of concepts defined in ontology. IC may affect the taxonomic organization of ontology.

These are believed to be the most important criteria to create classification hierarchies in ontology. From an ontological perspective, IC is to:

- *Classify* an entity as an instance of a class C
- *Individuate* an entity as *distinct* instance of C
- *Identify* two entities at a given time *(synchronic identity)*
- *Re-identify* an instance of C across time (persistence, or *diachronic identity*)

One important issue to note here is the relationship between IC and location. Current GIS and even modern spatial ontologies adopt the premise that an object has to have some location, but that location is, in general, arbitrary. Just where an object can be is constrained by its physical constitution.

The above thesis flies against the realities of geospatial intelligence analysts. In many cases, analysts' determine the identity of objects by analyzing their relative location. For example, a tilted building in a high-resolution satellite image that covers Italy can only be the Leaning Tower of Pisa. On the other hand, a tilted building in Afghanistan can not be the Leaning Tower of Pisa.



Geospatial Semantic Web: Applications, Figure 4 Nested contexts

Following the above example, it then becomes important to always ask which spatial relationships can be used to define distinct ontological entities. Physical entities, for example, are typically distinguished from abstract entities precisely by virtue of their necessary location in time and space. In general, what is required is to be able to characterize precisely in what ways physical objects (or events) can be said to be located at particular locations in space. This raises questions about how the objects concerned are to be identified and how the locations are to be identified.

Spatial Reference System There are two distinct schools in representing location; absolutist and relativist position. Some researchers argue that for the purposes of linguistic semantics it is the relativist view that is compelling; another group favors the absolutist position, ignoring the fact that there is no such a thing as absolute reference system. This is because there is no single point from which the entire environment can be considered. Therefore, geospatial ontology GOF should be able to accommodate other views.

Mereotopology Mereology is the formal theory of part-relations: of the relations of part to whole and the relations of part to part within a whole. Mereotopology is the combined logical account of mereology and topology. Mereotopology is the theory of boundaries, contact, and separation built upon a mereological foundation. In the last few years, several researchers have shown that mereotopology is more than a trivial formal variant of point-set topology; Example 4 shows an example of mereotopological relationship.

Example 4

```
<owl:Class rdf:ID=“RCCSpatialRegion”>
<rdfs:label>RCCSpatialRegion
</rdfs:label>
</owl:Class>
<owl:ObjectProperty
rdf:ID=“connectsWith”>
<rdfs:label>connects
</rdfs:label>
<rdfs:domain
rdf:resource=“#RCCSpatialRegion”/>
<rdfs:range
rdf:resource=“#RCCSpatialRegion”/>
<rdf:type
rdf:resource=“&owl;SymmetricProperty”/>
</owl:ObjectProperty>

<owl:ObjectProperty
rdf:ID=“isDisconnectedFrom”>
<rdfs:label>isDisconnectedFrom
</rdfs:label>
<rdfs:domain
rdf:resource=“#RCCSpatialRegion”/>
<rdfs:range
rdf:resource=“#RCCSpatialRegion”/>
</owl:ObjectProperty>.
```

The central aims of point-set and set-theoretic topology are to investigate and prove results concerning the properties of entire classes of spaces. On the other hand, mereotopology aims are to find perspicuous ways to represent and reason about the topological properties and relations of entities that exist in space.

Boundaries Geospatial semantic ontology should distinguishes between bona-fide and fiat boundaries. Bona fide boundaries are boundaries *in the things themselves*. Bona-fide boundaries exist independently of all human cognitive acts. They are a matter of qualitative differentiations or discontinuities of the underlying reality. Examples are surfaces of extended objects like cars, walls, the floor of a parking lot. Bona-fide boundaries are marked by bold solid lines in Fig. 2. Fiat boundaries exist only in virtue of different sorts of demarcation effected cognitively by human beings. Such boundaries may lie skew to boundaries of the bona fide sort as in the case of the boundaries of a parking spot in the center of a parking lot. They may also, however as in the case of a parking spot at the outer wall of the parking lot, involve a combination of fiat and bona-fide portions such as a wall at its back side. The classification of boundaries generalizes to a classification of objects. Bona-fide objects have a single topologically closed bona-fide boundary. Fiat objects have fiat boundary parts.

A prime example of the distinction between boundary types is colocation. Colocation of spatial objects means that they are located at exactly the same three- or two-dimensional region of space at the same moment in time. Colocation of boundary parts means that two- or one-dimensional boundary parts of spatial objects are located at exactly the same two- or one-dimensional region of space. Distinct bona-fide objects cannot be colocated since they cannot overlap. Owing to their atomic structure the surfaces of bona-fide objects can be brought in contact, i. e., the atoms forming their surface can come close but the atoms forming the surface of one object remain distinct from the atoms of the other. They do not mix and do not form a shared boundary. Distinct fiat objects of the same ontological kind cannot overlap. Bona-fide boundaries are dependent particulars. This approach stands opposed to the set-theoretic conception of boundaries as sets of independent points, each one of which could exist though all around it would be annihilated, and has a number of possible interpretations. The general statement of the fiat and boundaries is that one would be able to assert that the existence of any boundary is such as to imply the existence of some entity of higher dimension which it bounds.

Shape and Size In a purely qualitative ontology it is difficult to describe shape. In such an environment, very limited statements can be made about the shape of a region. Possible statements include whether the feature has holes, hollow, and whether it is one piece or not. One approach that has not been explored by the research community to qualitatively describe shapes is to use analogical reasoning and similarities. For example, one can say that the shape of this feature is analogous to a crown and then one can create assertions about crown-shaped objects.

Suppose that an intelligence analyst runs object extraction software on a satellite image. The software had recognized and identified an object as a sport utility vehicle (SUV). During the course of collecting other evidence about this object from other sources, the analysts get a more accurate dimension of the SUV size. Once the size of the SUV is entered, the system recognizes that these are not the dimension of a SUV, but of a truck used to carry troops and hence it reclassifies the object as a truck.

It is evident from the above example that size plays an important role in the identification of some features. Current GIS is supported by the belief that identity of features does not depend on some physical quantities. There are no GIS that can check the consistency of feature identification with respect to their size. However, the size of an entity can determine its identity: consider a lake versus a pond, an island versus a continent, or a car versus a SUV.

Key Applications

The geospatial semantic web is a horizontal capability that is still in its infancy. It has the potential to penetrate virtually all applications where reasoning about space and time are relevant. This includes but is not limited to:

- Location-based services and advertising
- Spatially enabled tags and folksonomy
- Enterprise information, application integration
- Semantic discovery and mash-up of heterogeneous geospatial content and geospatial web services
- Addition of location- and time-aware services to web-based social networks and other web-based consumer services
- Management of mobile and moving assets
- Workforce management

Future Directions

Geospatial semantic web research has been identified as a short-term priority in 2002. In the longer term the W3C had identified spatial ontologies as a long term research challenge. Key research areas include spatial and temporal reasoning, information retrieval, integration, and semantic ontology development. Geospatial information systems are becoming part of the mainstream IT. The geospatial semantic web is expected to be the main catalyst for such proliferation. Semantic web and GIS capabilities will be buried under the mainstream web technology stack. Semantic web inference engines will be able to natively support spatial and temporal reasoning support. Yahoo maps, Google Earth and many other similar web services have made geospatial web mash-up possible. The geospatial semantic web will surely facilitate automated service discovery and mash-up.

GIS applications are becoming ubiquitous. Geospatial semantic web technology will be the main facilitator that will enable developers to bury and hide GIS complexities under sophisticated and yet simple user interfaces, embedded in new consumer gadgets, or integrated with mainstream IT. Several research efforts are currently underway to standardize on core geospatial ontology to facilitate semantic interoperability.

Research directions from the scientific perspective are geospatial ontology core,the need for uncertainty and fuzzy reasoning, enterprise integration, semantic spatial services, and enterprise integration of spatial information.

Recommended Reading

1. OWL Overview http://www.w3.org/TR/owl-features/
2. Gruber, T.: A translation approach to portable ontology specifications. Knowl. Acquis. **5**, 199–220 (1993)
3. Gruber, T.: Toward principles for the design of ontologies used for knowledge sharing. Int. J. Hum.-Comput. Stud. **43**, 907–928 (1995)
4. Bishr, Y.: Semantic Aspects of GIS. ITC, (1997)
5. Egenhofer, M.: Naïve geography. In: Frank, A., Kuhn, W. (eds.) COSIT '95, Semmering, Austria, September 1995. Lecture Notes in Computer Science, vol. 988, pp. 1–15, Springer-Verlag, (1995)
6. Simmons, P.: Parts: A Study in Ontology. Clarendon, Oxford (1987)
7. Egenhofer, M.: Toward the semantic geospatial web. In: Proceedings of the 10th ACM International Symposium on Advances in Geographic Information Systems,
8. Resource Description Framework http://www.w3.org/RDF/
9. Rule Markup Language (2002) http://www.ruleml.org/
10. Semantic Web Rule Markup Language http://www.daml.org/2003/11/swrl/
11. Web Ontology Language http://www.w3.org/2004/OWL/

Geospatial Semantic Web, Definition

Bhavani Thuraisingham, Latifur Khan,
Ganesh Subbiah, Ashraful Alam,
Murat Kantarcioglu
Department of Computer Science, The University
of Texas at Dallas, Dallas, TX, USA

Definition

Geospatial semantic web refers to an intelligent, machine-understandable web where geospatial data are encoded in a semantic-rich data model to facilitate automated decision making and efficient data integration.

Main Text

Geospatial semantic web is a set of enabling technology that promotes more efficient use of geospatial data. The loose structure of the provided data model ensures a flexible environment for applications outside of a pure geospatial domain to freely mix geospatial data with native data without having to know all the details ahead of time. The geospatial data management gets a boost from the inherent data model and logic-based reasoning framework, since data can be filtered more efficiently upon acquisition. Moreover, implicit knowledge can be discovered through inference. The hundreds of geospatial clearinghouses that have stored volumes of data over the years but cannot disseminate them very effectively due to security vulnerability and data assurance issues can now utilize the semantic web concepts to solve their problems.

Geospatial Semantic Web, Interoperability

Liping Di, Peisheng Zhao
Center for Spatial Information Science
and Systems (CSISS), George Mason University,
Fairfax, VA, USA

Synonyms

Semantic geospatial web; Geospatial semantic interoperability; Geospatial ontology

Definition

According to ISO/IEC 2382-01, *Information Technology Vocabulary, Fundamental Terms*, interoperability is defined as *"the capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units"*. Based on various forms of perspective on information heterogeneity, interoperability can be classified in four different levels [1].

1. System interoperability involves bridging the differences in hardware, system software and communication systems.
2. Syntax interoperability aims to solve the differences in machine-readable aspects of data representation.
3. Structure interoperability constructs a uniform view of data modeling.
4. Semantic interoperability is concerned with having a common understanding of the meaning of the exchanged information in different systems.

The interoperability within the geospatial semantic web is the capability of sharing domain-specific resources and knowledge among the domains of geospatial science in the semantic web environment. There is no general agreement on what exactly constitutes semantic interoperability in the geospatial semantic web. Some elements of semantic interoperability include: 1) top of the interoperability stack for distributed computing; 2) shared concepts; 3) formal, common knowledge representation, usually in graph form, generically termed "ontologies"; 4) actionable (i. e., machine-process-able) relationships between similar/connected concepts using inference, reasoning, pattern matching, rule processing, etc.; and 5) shared models for distributed processes, such as interface contracts and virtual machines [20]. The interoperability in the geospatial semantic web facilitates the cross-domain resource discovery, heterogeneous resource query, and automatic resource translation from one domain to another.

Historical Background

With the rapid development of Web (e. g., Web Service) and distributed computing infrastructures (e. g., GRID), it is easy to access a large amount of independently developed and managed information resources of a broad variety. The success of syntactically interoperable Web service connections has created significant semantic gaps in what can be utilized. Both users and their client software may lack the tools to work with the diverse information which Web service standards have made nominally obtainable. *"The Semantic Web seeks to make the meaning as accessible as the material by enabling connections which are both logical and (machine) actionable between concepts which a user presently understands and those which may be new and foreign"* [21]. Interoperability in the semantic web is the single most important ingredient for satisfying the increasing need to reuse and analyze data in order to create information and knowledge and to subsequently reuse and share it across multiple domains because it makes the meaning understandable by broad user communities.

Tim Berners-Lee, the inventor of the Web, is the driving force of the Semantic Web initiative [2]. Recently a growing number of individuals and groups from academia and industry have been evolving the Web into another level – the Semantic Web, in which the meaning of information plays a far more important role. By representing not just words, but their definitions and contexts, the Semantic Web uses descriptive technologies (e. g., RDF and OWL) to provide a common semantic interoperable framework in which information is given a well-defined meaning such that distributed heterogenous the data and applications can be interoperable regardless of their differences on system, syntax, structure and semantics.

Because of the multi-source, multi-format, multi-scale, and multi-disciplinary nature of geospatial data and processing, the importance of semantics on accessing and integrating distributed heterogeneous geospatial information has long been recognized. The advances in the Semantic Web promise a generic framework to use ontologies to capture information meanings and relationships. Therefore, without doubt, adopting Semantic Web technologies to solve the issues of geospatial semantic interoperability is a good choice. Parallel to the development of the Semantic web, momentum has gained in recent years for the development of the Geospatial Semantic Web (GSW) – a geospatial domain-specific version of the Semantic Web which gives both people and machines true access to a wider range of geospatial knowledge. The GSW provides a common interoperable framework in which geospatial information is given a well-defined meaning such that distributed heterogeneous geospatial data and applications can be interoperable regardless of their differences on system, syntax, structure and semantics.

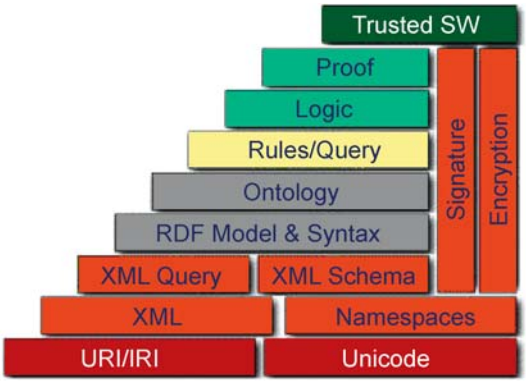
The Semantic Web provides a generic framework for semantic interoperability. It does not explicitly address some of the most basic geospatial entities, properties and relationships that are critical to geospatial information processing tasks, such as direction, orientation, spatial aggregation and topology. To better support the discovery, retrieval and consumption of geospatial content, the GSW creates and manages geospatial ontologies to exploit the logical structure of the geospatial world and allows applications to take advantage of "intelligent" geospatial reasoning capabilities. It does so by incorporating geospatial data semantics and exploiting the semantics of both the processing of geospatial relationships and the description of tightly coupled service content [3,4].

The Geospatial Semantic Web was listed as an immediately-considered research priority in 2002 by the University Consortium for Geospatial Information Science (UCGIS) [5]. Three eminent research questions were identified—creation and management of geo-ontologies, matching geographic concepts in Web pages to geo-ontologies, and ontology integration. Subsequently, a number of geospatial ontologies have been developed, such as the ISO19100 series, RDF geo vocabulary, OWL-Space, and NASA GCMD. In collaboration with industry, government and universities, the Open Geospatial Consortium (OGC) conducted the Geospatial Semantic Web Interoperability Experiment (GSW-IE) in 2005. GSW-IE aims to develop a method for discovering, querying and collecting geospatial content on the basis of formal semantic specifications. Significant achievements of this experiment include [21]: 1) development / encoding of formal geospatial ontologies, including feature type descriptions, 2) geospatial service interfaces which can provide service information formulated in the OWL-S semantic expression language referencing those ontologies, 3) OGC Web Feature Service interfaces which can operate on requests for ontologically expressed service and content descriptions, 4) semantic query language processing interfaces, and 5) tools for generating semantically expressed geospatial information.

Scientific Fundamentals

Compared to the conventional Web, the Semantic Web in advanced in two aspects: 1) common formats for data interchange (the original Web only has interchange of documents), and 2) a language for recording how the data maps to real world objects [6]. With such advances, intelligent systems with the help of reasoning engines and Web-crawling agents can go one step further to answer questions such as "*which airfields within 500 miles of Kandahar support C5A aircraft?*", rather than simply returning Web pages that contain the text "airfield" and "Kandahar", which most search engines can do today.

The hierarchical architecture of the Semantic Web is shown in Fig. 1. At the bottom level, Extensible Markup Language (XML) is a data-centric, customizable language providing syntax to represent structured documents with a user-defined vocabulary, however, it imposes no semantic constraints on the meaning of these documents. XML schema provides a means for restricting the structure of XML documents. Resource Description Framework (RDF) is a basic data model with XML syntax for referring to objects ("resources") and how they are related in order to allow information exchange between applications without loss of meaning. RDF Schema is a semantic extension of RDF for describing the properties of generalization-hierarchies and classes of RDF resources. RDF Schema can be used as a primitive ontology language. Web Ontology Language (OWL) adds more vocabularies for explicitly representing the meaning of terms and the relationships between those terms, such as relations between classes (e. g., disjointness), cardinality (e. g., "exactly one"), equality, richer typing of properties, the characteristics of properties (e. g., symmetry), and enumerated classes. The logic layer represents and derives knowledge, and the proof layer involves deductive process and proof validation. A digital signature can be used to sign and export the derived knowledge. A trust layer provides the trust level or its quality rating. Such a trust mechanism helps in building users' confidence in the process and the quality of the information provided and prevailing the Semantic Web [8].

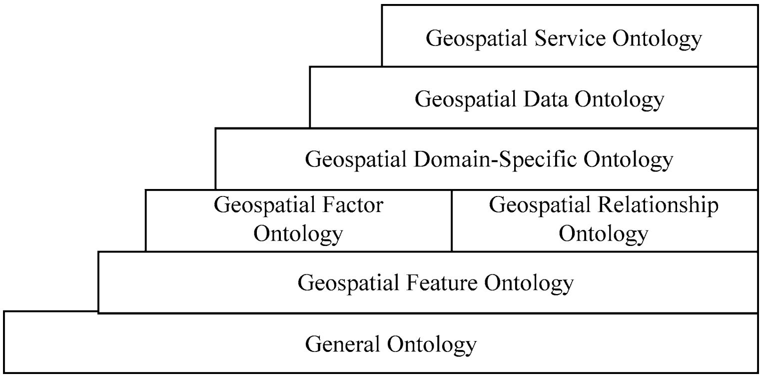


Geospatial Semantic Web, Interoperability, Figure 1 W3C Semantic Stack [7]

The architecture of the GSW is similar to the SW portrayed in Fig. 1, but with extensive domain-specific technology and knowledge. The GSW shares the basic components with the Semantic Web, such as top level (general) ontology, ontological languages, and general reasoning mechanisms. However, it extends the Semantic Web with domain-specific components. Among all the components of the GSW, two are especially unique – geospatial ontology and geospatial reasoning. The former aims at expressing geospatial concepts and relationships, specialized processing of geospatial rules and relationships, and self-described geospatial Web service in a formal way. The latter embraces sets of geospatial inference rules on the basis of geospatial ontologies and techniques to conduct automated geospatial reasoning for deriving geospatial knowledge by machine with less human interaction.

Ontology is the centerpiece of the most prevalent semantic technologies. It enables the semantic interoperability by providing a formal way of representing, acquiring and utilizing knowledge. Ontology originated from philosophy as a reference to the nature and the organization of reality. In general, an ontology is a "*specification of a conceptualization*" [9]. In the domain of computer science, ontology provides a commonly agreed understanding of domain knowledge in a generic way for sharing across applications and groups [10]. In more practical terms, ontology consists of a list of vocabularies and the relationships between them. The vocabularies denote the domain concepts (*classes* of objects), such as aircraft or runway. The *relationships* indicate hierarchies of concepts. They also include property information (an airfield has runways), value restrictions (aircraft can only take off at an airfield), disjointness statements (aircraft and train are disjoint), and specification of logical relationships between objects (a runway must be at least 400 meters long for supporting C5A aircraft).

In comparison to general ontologies, geospatial ontology specifically describes 1) spatial factors, e. g., location and units; 2) spatial relationships, e. g., inside, near, and far; 3) physical facts, e. g., physical phenomena, physical properties, and physical substances; 4) disciplines, e. g., scientific domains and projects; 5) data collection, e. g., collection properties, such as instruments, platforms and sensors; and 6) geospatial computing models, e. g., input, output and preconditions. A well-formatted geospatial ontology is the core for enabling geospatial semantic interoperability.



Geospatial Semantic Web, Interoperability, Figure 2 The Hierarchy of Geospatial Ontology

It enables the following key functions in geospatial semantic Webs:

- **Communication** among humans and information systems. Since the geospatial sciences is about the research across a variety of scales and disciplines, the semantics of geospatial information is essential for the development of interoperable geospatial software and data formats. Geospatial ontology provides a common understanding of not only general geospatial concepts but also in context of geospatial scientific computing
- **Spatial reasoning** about geospatial associations and patterns, e. g., topological relations (connectivity, adjacency and intersection of geospatial objects), cardinal direction (relative directions among geospatial objects, such as east, west and southwest), and proximity relations, such as A is close to B and X is very far from Y), and contextual relations, such as an obstacle separates two objects that would be considered nearby space, but are considered far because of the obstacle [11].

Based on the interactions and the role within the context of the GSW, geospatial ontology can be classified into several large groups. The hierarchical relationships of those groups are shown in Fig. 2.

General ontology is the core upper level vocabulary representing the common human consensus reality that all other ontologies must reference. There are several well-developed general ontologies. The Dublin Core Metadata [12] is a standard of metadata vocabularies describing resources that enables the development of more intelligent information discovery systems. OpenCyc [13] defines more than 47,000 upper level concepts and 306,000 assertions about these concepts.

Geospatial feature ontology, defining geospatial entities and physical phenomena, forms the ontological foundation for geospatial information. It should be coordinated with the development of geospatial standards. The ISO 19100 series and the OGC specifications are very helpful in defining its scope and content.

Geospatial factor ontology describes geospatial location, unit conversions factors and numerical extensions. Geospatial relationship ontology represents geospatial and logical relationships between geospatial features to enable geospatial topological, proximity and contextual reasoning. The RDF geo vocabulary [14] provides a basic RDF vocabulary with a namespace for representing latitude, longitude and other information about spatially-located things using WGS84 as a reference datum. The OWL-Space initiative (formerly DAML-Space) provides the ontologies of comprehensive spatial properties and relations. By incorporating Region Connection Calculus (RCC), the CoBrA [15] and the SWETO-GS [11] define basic relations between two-dimensional areas and relevant rules to enable the reasoning of spatiotemporal thematic proximity.

Geospatial domain-specific ontology represents the domain concepts by using proprietary vocabularies. Sometimes there are some different representations for a geospatial feature in different domains. To achieve interoperability, there must be a link between domain ontology and feature ontology, either by subclassing feature ontology concepts or by mapping from feature concepts to domain concepts. To provide formal semantic descriptions of NASA data sets and scientific concepts, several projects are underway to develop a semantic framework. Defined in OWL, the ontologies within the Semantic Web for Earth and Environmental Terminology (SWEET) [16] contain several thousand terms spanning a broad extent of Earth system science and related concepts (such as NASA GCMD, ESML, ESMF, grid computing, and OGC). The SWEET provides a high-level semantic description of Earth system science.

Geospatial data ontology uses the ontologies below it, as seen in Fig. 2, to provide a dataset description including representation, storage, modeling, format, resources, services and distributions. It ensures that the geospatial data are discoverable, usable, and interoperable semantically.

The ontologies of geographic metadata, such as ISO 19115 ontology [17], add semantic meanings and relationships to the data description by which data sets are explicitly associated with providers, instruments, sensors and disciplines. Geospatial service ontology describes who provides the service, what the service does and what other properties the service has that make it discoverable. It also states the service inputs, outputs, preconditions and effects to ensure the usability of the service. The inputs and outputs of a service are data described by data ontology. To allow dynamic invocation, this ontology includes the concrete service ports, protocols and encodings.

Key Applications

With geospatial ontology-based semantic interoperability, the GSW promises an "intelligent" method to discover, retrieve, integrate, and utilize large and diverse geospatial information and services over the Web. Numerous efforts to approach this "intelligence" are currently active.

Intelligent Search Engine

SWEET is a prototype for showing how the Semantic Web can be implemented in Earth Science. By incorporating the Global Change Master Directory (GCMD), SWEET provides 11 ontologies to describe Earth and environmental concepts and their relations, including Earth realm, nonliving element, living element, physical property, units, numerical entity, temporal entity, spatial entity, phenomena, human activities and data. Aided by these ontologies, an intelligent search engine (http://sweet.jpl.nasa.gov/perl/agents/interface_agent.pl) is implemented as a Web service using the RQDL (RDF Query Language). This engine can locate NASA data by finding related terms instead of having an exact keyword match. These terms may be synonymous (same as), more specific (subclass), or less specific (superclass) than those requested. Once the synonyms and subclass-superclass relationships have been discovered, the engine then submits the union of these terms to the GCMD search tool. The search results are presented as GCMD DIF summaries.

Geospatial Web Service Composition

A geospatial Web service is a modular application designed to deal with geospatial information over the network. In order to solve real-world geospatial problems through Web services, an "intelligent" mechanism is required to facilitate service discovery and integration and to automate the assembly of service chains. In [18], an approach to intelligent geospatial Web service is presented. The approach uses an ontologized "Geo-Object", a component of "Geo-Tree", to integrate the views of geospatial services and make them understandable and inferable. By using a semantic-enabled OGC catalog service, a "Geo-Object" can be easily found based on the flexible semantic match on its inputs and outputs. In fact, a "Geo-Tree" describing a geospatial modeling process can be represented by a service chain. The automatic construction of such a tree can be conducted through backward reasoning (from goal to source) in which all relevant "Geo-Objects" are discovered automatically based on a user's goals. Once all components have been found, the tree is initiated as a service chain in BPEL (Business Process Execution Language) and then sent to the workflow engine for execution.

Future Directions

From representation logic to computational logic, the GSW enhances our ability to express and deduce geospatial concepts and relationships for achieving interoperability among distributed heterogeneous geospatial data and applications. To make the GSW more efficient, practical, and operational, researchers and developers are investigating the following area [5,19]:

- Mappings between semantics and structured data files and databases, e. g., how can an ontology be generated automatically from maps, images and sketch available on the Web
- Ontology integration, e. g., how to assess computational models to specify, represent, access, and share multiple ontologies of geographic information
- Inference engines, e. g., how do users query the GSW using natural language
- Issues of trust, e. g., how to determine ontology reliability.
- Issues of proofs, e. g., what inferences lead to this answer and are the supporting ontologies logically sound.

Cross References

► Geospatial Semantic Web
► OGC's Open Standards for Geospatial Interoperability
► Ontology-Based Geospatial Data Integration

Recommended Reading

1. Sheth, A.: Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics. In: Goodchild, M., Egenhofer, M., Fegeas, R., Kottman, C. (eds.) Interoperating Geographic Information Systems, pp. 5–30. Kluver, New York (1999)

2. Berners-Lee, T. (1998) Semantic Web Road Map. http://www.w3.org/DesignIssues/Semantic.html. Accessed 20 Mar 2007
3. Egenhofer, M. (2002) Toward the Semantic Geospatial Web. In: The Tenth ACM International Symposium on Advances in Geographic Information Systems. Available via. http://www.spatial.maine.edu/~max/SGW02.pdf. Accessed 8-9 Nov 2002
4. Lieberman, J., Pehle, T., Dean, M. (2005) Semantic Evolution of Geospatial Web Services. In: W3C Workshop on Frameworks for Semantic in Web Services. Available via. http://www.w3.org/2005/04/FSWS/Submissions/48/GSWS_Position_Paper.html. Accessed 9-10 June 2005
5. Fonseca, F., Sheth, A. (2002) The Geospatial Semantic Web. University Consortium for Geographic Information Science. http://www.ucgis.org/priorities/research/2002researchPDF/shortterm/e_geosemantic_web.pdf. Accessed 21 March 2007
6. W3C Semantic Web. http://www.w3.org/2001/sw/. Accessed 21 March 2007
7. Berners-Lee, T. (2000) Semantic Web on XML. In: XML 2000. Available via. http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide1-0.html. Accessed 21 March 2007
8. Antoniou, G., Harmelen, F.: A Semantic Web Primer. MIT Press, Cambridge, MA (2004)
9. Gruber, T.R.: A Translation Approach to Portable Ontologies. Knowledge Acquisition **5**, 199–220 (1993)
10. Chandrasekaran, B., Johnson, T., Benjamins, V.: Ontologies: what are they? Why do we need them?. IEEE Intelligent Syst. Their Appl. **14**, 20–26 (1999)
11. Arpinar, A., Sheth, A., Ramakrishnan, C.: Geospatial Ontology Development and Semantic Analytics. In: Wilson, J.P., Fotheringham, A.S. (eds.) Handbook of Geographic Information Science, Blackwell Publishing, Boston (2004)
12. DublinCore Dublin Core Metadata Initiative. http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide1-0.html. Accessed 21 March 2007
13. Cycorp OpenCyc. http://www.cyc.com/opencyc. Accessed 21 March 2007
14. Brickley, D. (2003) Basic Geo (WGS84 lat/long) Vocabulary. http://www.w3.org/2003/01/geo/. Accessed 21 March 2007
15. Chen, H., Finin, T., Joshi, A. (2003) An Intelligent Broker for Context-Aware Systems. In: Ubicomp (2003), Seattle, Washington, USA. Available via. http://ebiquity.umbc.edu/paper/html/id/78/An-Intelligent-Broker-for-Context-Aware-Systems. Accessed 12–15 Oct 2003
16. Raskin, R.: Guide to SWEET Ontologies. http://sweet.jpl.nasa.gov/guide.doc. Accessed 21 March 2007
17. Islam, A.S., Bermudez, L., Beran, B., Fellah, S., Piasecki, M.: Ontology for Geographic Information – Metadata (ISO 19115:2003). http://loki.cae.drexel.edu/~wbs/ontology/iso-19115.htm. Accessed 21 March 2007
18. Di, L., Zhao, P., Yang, W., Yu, G., Yue, P.: Intelligent Geospatial Web Services. In: Proceedings of IEEE 25th Anniversary International Geoscience and Remote Sensing Symposium. Seoul, Korea (2005)
19. Lowe, J.A.: Geospatial Semantic Web. http://www.geospatial-online.com/geospatialsolutions/article/articleDetail.jsp?id=163883. Accessed 21 March 2007
20. Lieberman, J., Lutz, M., Pehle, T., Kolas, D., Morris C., Dean M., Probst, F., Klien, E.: Geospatial Semantic Web: An Interoperability Experience. http://www.semantic-conference.com/Presentations_PDF/Lieberman-Joshua.pdf (2007). Accessed 31 Aug 2007
21. Lieberman, J.: Geospatial Semantic Web Interoperability Experiment Report.http://portal.opengeospatial.org/files/?artifact_id=15198 Accessed 22 March 2007

Geospatial Semantic Web: Personalization

Pragya Agarwal
Department of Geomatic Engineering,
University College London, London, UK

Synonyms

Adaptation; Embodiment, Individualization; Cognitive engineering; Externalization; Reasoning; User-centered; Contextualization; Personalization

Definition

What is the Semantic Web?

The basic principle underlying the semantic web is the idea of having data defined and linked in such a way that it can be used for more effective discovery, automation, integration, and reuse across various applications [7,16]. With the idea of a semantic web in which machines can understand, process and reason about resources to provide better and more comfortable support for humans in interacting with the World Wide Web, the question of personalizing the interaction with web content to user needs and profile is of prime significance.

What is the Geospatial Semantic Web?

Geographic data has meanings associated with it and semantic annotations of geographic data will allow for better retrieval and more effective use of geospatial services. Recent developments in the semantic web have great potential for the geospatial community, specifically because the focus on the incorporation of data semantics will lead to a better retrieval and more reliable integration methods by tapping into the semantics during the search process on the web. There is a distinct move away from structure and syntax in the geospatial community, accompanied by an increased awareness that semantics is the backbone for a successful ontology to enable translation of data from different resources and users. Ontologies are being increasingly used in the geospatial community to enable interoperability between multiple resources, systems, services and semantic conceptualizations. The most commonly used definition of an ontology is that it is a “specification of a conceptualization” [18]. However, the basic semantic web and the technological developments are not targeted to the needs of the geospatial community as spatial data has its own specific needs associated with it. A geospatial domain is characterized by vagueness, especially in the semantic disambiguation of the concepts in the domain, which makes defining universally accepted geoontology an onerous task [2]. This is compounded by the lack of appropriate methods and techniques where the individual semantic conceptualizations can be captured and compared to each other. The idea of a more focused "Geospatial Semantic Web" has been recognized as a research priority within UCGIS (University Consortium for Geographic Information Science) initiatives [17]. Egenhofer [15] identified the need to support queries based on meanings and better definition of spatial terms across a number of disciplines, and the need to integrate multiple terminological ontologies as a backbone for an effective geospatial semantic web (GSW).

What is Personalization?

Nowadays, most large-scale applications are planned and designed for a large variety of end users. Nevertheless, the traditional "one-size-fits-all" approach turns out to be outdated and inappropriate to meet with heterogeneous user needs. "Personalization" is the key word and the specific challenge is to find out what the users are looking for in order to understand how to offer them tailored information. Therefore, personalization of information means to deal with information more efficiently and effectively, performing complex tasks within a reasonable time, in a "user-friendly" environment. The development of any personalization techniques and adaptive information retrieval systems should deal with the users' understanding of domain examined by knowledge capture approaches.

A "closed world" paradigm in which individual preferences and goals are not considered, in which the system requirements are deterministic, and not adapted to changing needs of the community, has meant that the system design and operation work on a fixed set of resources which are normally known to the system designers at design time. On the other hand, personalization facilitates a move away from this deterministic approach that traditional information system design frameworks operate under. The individual requirements and context that the user operates under has to be taken into account to move from a closed-world setting. knowLedge-enhanced web services are normally driven by some description of the world which is encoded in the system in the form of an ontology defined by knowledge engineers. The users' conceptualization of the world may also differ, sometimes significantly, from the conceptualization encoded in the system. If not taken into account, the discrepancies between a user's and a system's conceptualization may lead to the user's confusion and frustration when utilizing semantic web services, which, in turn, can make these services less popular. Semantic personalisation of web-based services is required to exploit user intentions and perspectives and tailor the information accordingly.

Personalization can be defined as a process of gathering user-information during interaction with the user, which is then used to deliver appropriate content and services, tailored to the user's needs. Personalization is a process by which it is possible to give the user optimal support in accessing, retrieving, and storing information, where solutions are built so as to fit the preferences, the characteristics and the taste of the individual. Personalization of information systems aims at developing systems that can autonomously interoperate—either with humans or with other systems—adapting their processing and its outcome to specific requests. Personalized information systems aim to make user needs the center of the interaction process and to optimize user access to the relevant and required information according to individual needs, profiles and particular contexts. This result can be achieved only by exploiting machine-interpretable semantic information, e. g., about the possible resources, about the user him/herself, about the context, about the goal of the interaction. Personalization is realized by a reasoning process applied to the semantic information, which can be carried out in many different ways depending on the specific task.

Although, conventionally Personalization is taken to mean directing the system design to user needs and profiles, it can also mean to adapt according to local browser or regional contexts. The individual requirements of the user are to be taken into account in such different dimensions. Some of these individual user requirements will include the current task, the goal of the user, the context in which the user is requesting the information, the previous information requests or interactions of the user, the working process s/he is involved in, knowledge of the user (an expert will be satisfied by information which is not suitable for a layman), the device s/he is using to display the information, the bandwidth and availability of the communication channel, the abilities/disabilities/handicaps of the user, and the time constraint of the user (whether s/he is under time pressure, or is just browsing some information).

Personalization of Semantic Web

Personalized semantic web is the concept of the semantic web in which machines are enabled to understand the meaning of information and thereby provide better support to individuals in carrying out their tasks, and is aimed at improving the user's experience of a web-based service. In particular, applications that can retrieve, process and present information in enhanced user-adapted ways are interesting.

Many attempts have been made to apply personalization techniques to the World Wide Web as a natural extension of work on hypertext and hypermedia; however, the web is an information space thought for human to human communication, while personalization requires software systems to take part to the interaction and help [13]. Such systems require knowledge to be expressed in a machine-interpretable format, which in the web is not available. The development of languages for expressing information in a machine-processable form is characteristic of the semantic web initiative [6]. Over this knowledge layer, the use of inferencing mechanisms is envisioned as a fundamental means for performing a content-aware navigation, producing an overall behavior that is closer to the user's intuition and desire. This is the reason that the semantic web is the most appropriate environment for realizing personalization. In other words, the semantic web is deeply connected to the idea of personalization in its very nature.

What is Personalization of GSW?

Users' preferences, expectations, goals and tasks differ while using the web for geographic information. Moreover, people form different conceptual models of the world and these models dynamically change over time. In addition, meanings are crucial in distinction of geographic information and people constantly assign meanings to real world objects, while categorizing them as they interact with the world around them. Agarwal [2] discusses in detail the problems associated with ontology development in the geospatial domain primarily due to semantic ambiguities. The knowledge-enhanced web services are normally driven by some description of the world which is encoded in the system in the form of an ontology defined by knowledge engineers. The user's conceptualization of the world may differ, sometimes significantly, from the conceptualization encoded in the system. If not taken into account, the discrepancies between a user's and a system's conceptualization may lead to the user's confusion and frustration when utilizing the web-based geospatial services, which, in turn, can make these services less popular. Indeterminacy and ambiguity in meanings are key issues in the development of ontologies in the geographic domain [2,5]. Empirical results show that individual conceptualizations are characterized by semantic heterogeneity (Agarwal [1]).

With multiple user conceptualizations, efforts towards a reliable geospatial semantic web, therefore, require personalization where user diversity can be incorporated and targeted to multiple needs, semantics and conceptualizations of the real world. Egenhofer [15] identified the need to support queries based on meanings and better definition of spatial terms across a number of disciplines, and the need to integrate multiple terminological ontologies as a backbone for an effective GSW. The success of a standardized geoontology for the semantic web will be determined by the level of acceptance by the users of the services, both expert and naïve, and the level to which the basic geoontology is semantically compatible with the users' conceptualizations.

In this way, personalization is crucial for acknowledging and addressing individual user differences and needs and providing user-friendly user access to geospatial information on the web. To achieve a GSW will require both syntactic and semantic interoperability of resources, and therefore the personalization efforts are essential in making the GSW a reality.

Historical Background

GSW is a new initiative and many concrete applications of it are yet to be seen. Personalization of GSW is a novel concept and has no historical precedents.

Scientific Fundamentals

Many research disciplines have contributed to web personalization research, for example, hypertext research has studied personalization in the area of so-called adaptive hypertext systems, collaborative filtering research has investigated recommender systems, artificial intelligence techniques have been widely used to cluster web data, usage data, and user data, reasoning and uncertainty management has been adopted to draw conclusions on appropriate system behavior, etc. [13]. However, in most systems, there are no mechanisms to capture the interaction and context of the user. There is an urgent need to include the people as an axis in the design, development, and deployment of semantically enriched services, especially for personalization of the GSW where individual user semantics vary. In addition, computational models are needed that can process the different terminological and semantic ontologies and process the semantic incompatibilities between users and the expert's geoontology.

Resolving the discrepancy between psychological user variables and physical system variables in the area of geospatial services goes beyond the user-interface level. Rather than a closed view of the world, the personalization efforts for geospatial services design will ensure that the different perspectives and semantic conceptualizations of the real world are maintained as "open". The underlying principle for the methodology adopted by Agarwal et al. [3] and Huang et al. [20] is that such geospatial services and applications could benefit from personalization efforts where semantic variations and conceptualizations are captured and aligned, and differences with the core expert ontology identified and formalized to provide the basis for user-supported access to such location-based information. This will allow, first, for the development of systems that allow personalization by incorporating user models and diversity and second, as a means to test any core ontologies that are developed as the basis for a geospatial services against user conceptualizations for discrepancies and thereby evaluate its reliability as a standard, reusable ontology. Moreover, the personalization approach allows flexibility and the possibility of using the user models to enrich the available information resources with shared semantics instead of relying on fixed ontologies available to the developers at the design stage. Using this approach, the research towards specification of well-defined standards and ontologies for interoperability in the geospatial domain can be enhanced and personalized to provide extendibility, flexibility, interoperability, and reusability. Automating the mapping of multiple conceptualizations and personalization of web-based services will also facilitate pervasive computing in mobile services and enable more effective use of mobile GIS services.

Key Applications

The availability of geospatial knowledge resources on the web enables members of the public to take advantage of trusted knowledge built by domain experts, e. g., for planning travel routes and for accessing weather information. Geospatial services and systems are also unique in the way that they use data, which are related to locations in space and time, and that the processing of the data with respect to these spatial locations is possible. People's input to geospatial tools have a spatiotemporal context. One can ask "where is a certain object" or "where are all objects with certain properties" at a given time when trying to find the nearest health services for the elderly; or one can ask "what are the properties of a certain area in space (as well as time)" when trying to ascertain the suitability of an environment (for example, crime rates) while renting or buying a property. Users access the web services with different goals and often; these services require integration of the various different resources to provide a comprehensive result for the user search for their specific requirements. For example, in a "what is in my backyard" service provided by the Environment Agency (EA) in the UK, members of the public can see what pollutants may be scattered across their neighborhood. End-users will have their own contexts of use: property evaluation, ecology, etc. and for a member of the public, a general interest (based on a topographic view of different areas in the city). Each could potentially view the data provided by the others but form their own conceptual understanding of the location-based information. Personalization efforts will make it possible to capture and formalize context and thereby provide context-relevant information.

Future Directions

There is no common language for expressing adaptive functionality, hence these systems are difficult to compare and analyze, also suffering to a great extent from lack of reusability—which in fact is the key capability for successful personalization functionality for the GSW. To deal with a diverse user population having different preferences, goals, understanding of tasks and conceptual models, existing design paradigms in geospatial services will have to be redefined. Furthermore, new diagnostic techniques and models are needed to capture the long-term development of users' capabilities, the dynamics of users' goals and conceptual understanding, the uncertainty and inconsistency of naive users' conceptualizations, and so on. The ambitious target is to offer manageable, extendible and standardized infrastructure for complementing and collaborating applications tailored to the needs of individual users. Traditional personalization and adaptation architectures were suited to deal with closed-world assumption, where user modeling methods, such as overlay, bug library, constraint-based modeling and other marked discrepancies in a user and expert's semantics as erroneous, and often called them misconceptions. New approaches for open-world user modeling that facilitate elicitation of extended models of users are needed to deal with the dynamics of a user's conceptualization. Similarly, methods that acknowledge semantic discrepancies and heterogeneity are required for effectively personalizing the web-based services for the geospatial community. Adaptation and personalization methods are not developed to address the temporality inherent in geospatial information and therefore, any specific personalization efforts for the GSW will have to be modified to suit this need. Without the benefit of deeper semantic or ontological knowledge about the underlying domain, personalization systems cannot handle heterogeneous and complex objects based on their properties and relationships. Nor can these systems possess the ability to automatically explain or reason about the user models or user recommendations. This realization points to an important research focus that can combine the strengths of web mining with semantic or ontological knowledge. Development of automated reasoning tools to detect mismatches and discrepancies between the user and the expert ontology forming the backbone for the web-based resources will be a step forward.

Cross References

► Geospatial Semantic Web

Recommended Reading

1. Agarwal, P.: Contested Nature of place. In: Egenhofer, M., Freska, C., Harvey, M. (eds.) 3rd International Conference GIScience, 2004, LNCS 3234, pp. 1–21. Springer, Berlin (2004)
2. Agarwal, P.: Ontological considerations in GIScience. Int. J. Geogr. Inf. Sci. **19**, 501–535 (2005)
3. Agarwal, P., Huang, Y., Dimitrova, V.:, Formal approach to reconciliation of individual ontologies for personalization of geospatial semantic web. In: Proceedings of the First International Conference on Geospatial Semantics, Mexico City, November 2005. Lecture Notes in Computer Science, vol. 3799, pp.202–219, Springer-Verlag, Berlin (2005)
4. Antoniou, G., van Harmelen, F.: A Semantic Web Primer. MIT Press, Cambridge, Massachusetts (2004)
5. Bennett, B.: What is a Forest? On the vagueness of certain geographic concepts. Topoi **20**, 189–201 (2001)
6. Berners-Lee, T.: (1998) The Semantic Web Road Map. World Wide Web Consortium W3C Draft. http://www.w3.org/DesignIssues/Semantic
7. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Sci. Am. May (2001)
8. Brusilovsky, P.: Adaptive Hypermedia. User Modeling User-Adapted Interaction **11**, 87–110 (2001)
9. Cardoso, J.: Semantic Web Services: Theory, Tools and Applications. Idea Group, (2007)
10. Cardoso, J., Sheth, A.: Semantic Web Services, Processes and Applications. Springer-Verlag, USA (2006)
11. Davies, J., Fensel, D., van Harmelen, F.: Towards the Semantic Web: Ontology-Driven Knowledge Management. Wiley, UK (2003)
12. Doan, A., Madhavan, J., Domingos, P., Halvey, A.: Learning to map between ontologies on the semantic web. In: Proceedings of WWW'02, Hanau, 7–11 May 2002
13. Dolog, P., Henze, N., Nejdl, W., Sintek, M.: (2003) Towards the adaptive semantic web. In: Proceedings of the 1st Workshop on Principles and Practice of Semantic Web Reasoning (PPSWR'03), co-located with International Conference on Logic Programming (ICLP'03), Mumbai, India, December 8, 2003
14. Dolog, P., Henze, N., Nejdl, W., Sintek, M.: Personalization in Distributed e-Learning Environments. In: Proceedings of 13th World Wide Web Conference, New York, USA 17–22 May 2004
15. Egenhofer, M.J.: Toward the semantic geospatial web. In: Proceedings of the Tenth ACM International Symposium on Advances in Geographic Information Systems, McLean, Virginia, 8–9 Nov 2002
16. Fensel, D., Wahlster, W., Lieberman, L., Hendler, J.: Spinning the Semantic Web: Bringing the World Wide Web to its Full Potential. MIT Press, Cambridge, MA (2002)
17. Fonseca, F., Sheth, A.: (2002) The geospatial semantic web. UCGIS White Paper. http://www.ucgis4.org/priorities/research/2002researchagenda.htm
18. Gruber, T.R.: A translation approach to portable ontologies. Knowledge Acquisition **5**, 199–220 (1993)
19. Hendler, J., Berners-Lee, T., Miller, E.: Integrating applications on the semantic web (English version). J. Inst. Electrical Eng. Jpn. **122**, 676–680 (2002)
20. Huang, Y., Dimitrova, V., Agarwal, P.: Detecting mismatches between a user's and a system's conceptualizations to empower user modelling for the Semantic Web. In: Proceedings of the Workshop on Personalization in the Semantic Web, User Modelling Conference 2005, Edinburgh, 24–30 July 2005
21. Shadbolt, N., Berners-Lee, T., Hall, W.: The semantic web revisited. IEEE Intell. Syst. **21**, 96–101 (2006)
22. Staab, S., Studer, R.: Handbook on Ontologies. Springer-Verlag, Heidelberg (2004)

Geospatial Semantics

► Geospatial Semantic Integration

Geospatial Semantics Web

► Geospatial Semantic Web: Applications

Geospatial Technology

► Evolution of Earth Observation

Geospatial Web 3.0

► Geospatial Semantic Web: Applications

Geostatistical Models

► Hierarchical Spatial Models

Geostatistics

► Data Analysis, Spatial

Geotemporal Role-Based Access Control

► Security Models, Geospatial

Geovista

► Geographic Dynamics, Visualization And Modeling

Geovisualization

► Exploratory Visualization

Getis-Ord Index G*

► Local and Global Spatial Statistics

GIbb’s Sampling

► Hurricane Wind Fields, Multivariate Modeling

GIS

► Information Services, Geography

GIS Mashups

ILYA ZASLAVSKY
San Diego Supercomputer Center,
University of California San Diego, San Diego, CA, USA

Synonyms

Composite geographic information systems web application; Web application hybrids

Definition

A mashup is an online application or web site that seamlessly combines content from several sources. Geographic information systems (GIS) mashups typically combine spatial data and maps from several web sources to produce composite thematic maps.

Main Text

Mashup is an approach to creating composite dynamic web sites that has quickly gained popularity due to the proliferation of reliable online sources of maps, web feeds, and other content that can be accessed via relatively simple published interfaces (APIs). Online maps and GIS services available through Google, Windows Live, and Yahoo APIs, among others, can be combined with real time traffic information, weather information, various third party databases, blogs, information feeds, search engines, forums, etc. Mashups are also thought of as “Web 2.0 applications”, a somewhat hyped notion of a next generation internet application platform that emphasizes openness of data and services, lightweight application interfaces and the ability to mix and match different services to create a rich application experience for web users.

Cross References

► Cyberinfrastructure for Spatial Data Integration
► Grid
► Web Services

GIS-Based Hydrology

► Distributed Hydrologic Modeling

GIS-Based Multicriteria Decision Analysis

► Multicriteria Decision Making, Spatial

GIS Software

► ArcGIS: General Purpose GIS Software System

GIService

► Information Services, Geography

GiST Index

► PostGIS

Global and Local Spatial Modeling

► Spatial and Geographically Weighted Regression

Global Positioning System

► Location-Aware Technologies

Global Sensitivity Analysis

XIAOBO ZHOU, HENRY LIN, HENRY LIN
Department of Crop and Soil Sciences, The Pennsylvania State University, University Park, PA, USA

Definition

Global sensitivity analysis is the process of apportioning the uncertainty in outputs to the uncertainty in each input factor over their entire range of interest. A sensitivity analysis is considered to be global when all the input factors are varied simultaneously and the sensitivity is evaluated over the entire range of each input factor.

Main Text

In many complex and nonlinear spatial phenomenon or processes, input factors usually interact with each other and therefore it is inappropriate to evaluate the impact of one input factor on the model output with other factors being constant. Global sensitivity analysis quantifies the importance of model inputs and their interactions with respect to model output. It provides an overall view on the influence of inputs on outputs as opposed to a local view of partial derivatives as in local sensitivity analysis. One of the most challenging issues for global sensitivity analysis is the intensive computational demand for assessing the impact of probabilistic variations. There are many global sensitivity analysis methods including the Sobol's sensitivity estimates, the Fourier amplitude sensitivity test (FAST), and the Monte-Carlo-based regression–correlation indices. The global sensitivity analysis is suitable for a nonlinear input–output relationship, and is more realistic to the real world since it allows all input factors to be varied simultaneously.

Cross References

- ► Local Sensitivity Analysis
- ► Screening Method
- ► Sensitivity Analysis

GML

- ► deegree Free Software
- ► Geography Markup Language (GML)
- ► Web Feature Service (WFS)

GNU

- ► deegree Free Software

Gnu Public License (GPL)

- ► Open-Source GIS Libraries

Google Earth

- ► Multimedia Atlas Information Systems

G-Polygon

- ► Spatial Data Transfer Standard (SDTS)

GPS

- ► Evolution of Earth Observation
- ► Indoor Localization
- ► Location-Aware Technologies
- ► Moving Object Uncertainty
- ► Privacy Preservation of GPS Traces

GPSBabel

- ► Open-Source GIS Libraries

Graph

- ► Road Network Data Model

Graph Theory, Konigsberg Problem

Betsy George
Department of Computer Science,
University of Minnesota, Minneapolis, MN, USA

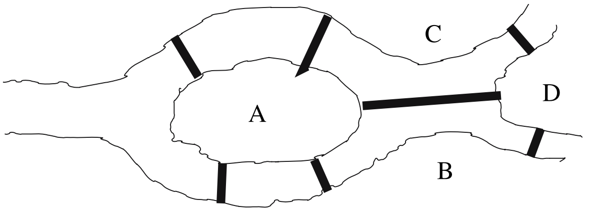
Synonyms

Problem of seven bridges of Konigsberg; Euler's Konigsberg's bridges problem; Edge routing problems

Definition

In Geographic Information Systems, concepts from graph theory are extremely useful in expressing the spatial structure of entities seen as points, lines, areas and solids, after the geometrical details of these entities are removed. For example, in transportation and river networks, the topological properties of their structures can be represented using graphs. This article describes the origins of graph theory and the impact it has on various fields ranging from geography to economics.

The Konigsberg Bridge Problem is a classic problem, based on the topography of the city of Konigsberg, formerly in Germany but now known as Kalingrad and part of Russia. The river Pregel divides the city into two islands and two banks as shown in Fig 1. The city had seven



Graph Theory, Konigsberg Problem, Figure 1 Layout of the city of Konigsberg showing the river, bridges, land areas

bridges connecting the mainland and the islands (represented by thick lines in the figure). [1,2,3,4]. The problem asks whether there is a walk that starts at any island, traverses every bridge exactly once, and returns to the start point.

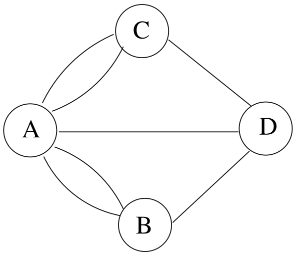
The solution proposed by a Swiss Mathematician, Leonhard Euler, led to the birth of a branch of mathematics called graph theory which finds applications in areas ranging from engineering to the social sciences. Euler proved that there is no solution to the problem based on the number of bridges connecting each land area.

The results from the solution of the Konigsberg problem have been extended to various concepts in graph theory. In graph theory a path that starts and ends at the same node and traverses every edge exactly once is called an Eulerian circuit. The result obtained in the Konigsberg bridge problem has been generalized as Euler's theorem, which states that a graph has an Eulerian circuit if and only if there are no nodes of odd degree. Since the graph corresponding to Konigsberg has four nodes of odd degree, it cannot have an Eulerian circuit. Subsequently the concept of Eulerian paths was introduced, which deals with paths that traverse every edge exactly once. It was proved that such a path exists in a graph if and only if the number of nodes of odd degree is 2 [3,4,5,8,9].

While studying the Konigsberg bridge problem, Euler also observed that the number of bridges at every land area would add up to twice the number of bridges. This result came to be known as the hand-shaking lemma in graph theory, which states that the sum of node-degrees in a graph is equal to twice the number of edges. This result is the first formulation of a frequently used result in graph theory that states that the sum of node-degrees in a graph is always even [8,9].

Historical Background

The Konigsberg bridge problem was formulated based on the layout of the city of Konigsberg around the river Pregel. The problem was to find a tour that starts at any point in the city, crosses each bridge exactly once, and returns to the starting point. No one succeeded in doing this. A Swiss mathematician, Leonhard Euler formulated this problem by abstracting the scenario. He formulated the problem as finding a sequence of letters A, B, C, D (that represent the land areas) such that the pairs (A,B) and (A,C) appear twice (thus representing the two bridges between A and B, and A and C) and the pairs (A,D), (B,D), (C,D) appear only once (these pairs would represent the bridges between A and D, B and D, C and D). Euler used a counting argument to prove that no such sequence exists thus proving that there the Konigsberg Bridge Problem has no solution. Euler presented this result in the paper, "The Solution of Problem Relating to the Geometry of Position" at the Academy of Sciences of St. Petersburg in 1735. This paper, in addition to proving the non-existence of solution to the Konigsberg Bridge Problem, gave some general insights into arrangements of bridges and land areas [5,6,8].



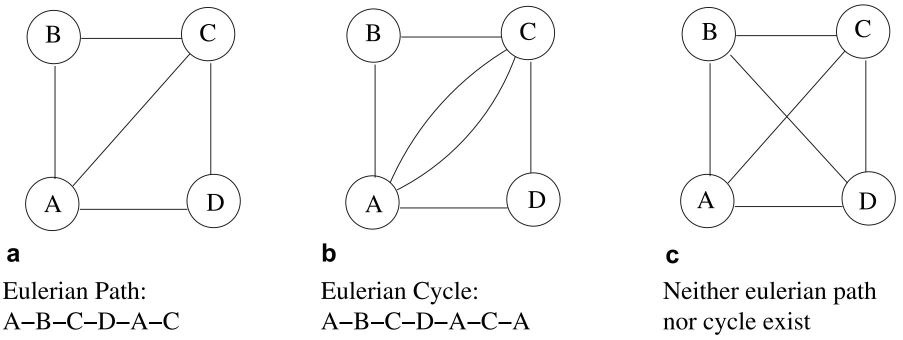
Graph Theory, Konigsberg Problem, Figure 2 Graph Representation ot the city of Konigsberg

Euler summarized his main conclusions as follows:

1) If there is any land area that is connected by an odd number of bridges, then a cyclic journey that crosses each bridge exactly once is impossible.
2) If the number of bridges is odd for exactly two land areas, then there is a journey that crosses each bridge exactly once is possible, if it starts at one of these areas and ends in the other.
3) If there are no land areas that are connected by odd number of bridges, the journey can start and end at any land area [8].

Euler gave heuristic reasons for the correctness of the first conclusion. To complete a cyclic journey around the land areas, crossing each bridge exactly once, there must be a bridge to leave the area for every bridge to enter it. This argument was generalized to the conclusion that a cyclic journey is possible if every island is connected by an even number of bridges. Formal proofs for the conclusions were not proposed until the year 1871, in a posthumous paper by Carl Hierholzer [2,5].

The paper presented by Euler on the Konigsberg bridge problem can be considered to mark the birth of graph theory, in general. Later, a diagrammatic representation evolved which involved nodes or vertices and the connecting lines that are called edges. Using this representation, the Konigsberg problem is modeled as shown in Fig. 2.



a Eulerian Path: A–B–C–D–A–C

b Eulerian Cycle: A–B–C–D–A–C–A

c Neither eulerian path nor cycle exist

Graph Theory, Konigsberg Problem, Figure 3 Illustration of an eulerian path and eulerian cycle

Circles, called nodes, represent the islands and the banks and connecting lines called edges represent the bridges. The number of edges that are incident on a node is called the degree of the node [8].

In the Konigsberg bridge problem, the number of bridges connecting a land area would be the degree of the node representing the land area.

Scientific Fundamentals

In an undirected graph, a cycle that traverses every edge exactly once is called an Euler tour or Euler cycle. Any graph that possesses an Euler cycle is called an eulerian graph. A path that traverses each edge exactly once with different starting point and end point is called an eulerian path. An undirected multi-graph has an Eulerian circuit (path) if and only if it is connected and the number of vertices with odd degree is zero (two).

Figure 3 illustrates eulerian path and eulerian cycle in a graph. In Fig. 3a, an eulerian path exists and it can be observed that the graph has exactly two, odd degree vertices, which would be the start and end vertices of the eulerian path, A-B-C-D-A-C. Figure 3b does not have vertices with odd degree and has an eulerian cycle whereas Fig. 3c has neither an eulerian path nor an eulerian cycle.

Finding an Eulerian Circuit in a Graph

The method successively finds cycles in the graph. At each step the edges that are in the already discovered cycles are removed and the cycle is spliced with the one discovered in the previous step. This process is continued until all edges are exhausted. These basic ideas were formalized into an algorithm in [10]. The algorithm maintains a list L with each vertex x such that the kth entry in the list indicates the vertex to visit when vertex x is reached the kth time.

Algorithm

Step 1: Select any vertex $v1$. $v = v1$; $k[v] = 0$ for all vertices v. Label all edges as unvisited.

Step 2: Select an unvisited edge e incident to v. Mark this edge "visited". Let w be the other end vertex of e. Increment k_v by 1 and $L_v[k_v] = w$. If w has an unvisited incident edge, go to step 3. If not, y will be $v1$. Then, go to step 4.

Step 3: Set $v = w$ and go to step 2.

Step 4: Find a vertex $v1$ such that there is at least one visited edge and one unvisited edge incident at $v1$. Set $v = v1$ and go to step 2. If no such vertex exists, go to step 5.

Step 5: To construct the Eulerian circuit, start at $v1$. First time a vertex u is reached, proceed to the vertex $L_u[k_u]$. Decrement k_u and continue.

Trace of the algorithm for Figure 4

Step 1: $v1 = 1 = v$; $k_x = 0$ for $x = 1, 2, 3, 4$.

Step 2: Select edge a. $w = 2$; $k_2 = 1$; visited$(a) = 1$.

Step 3: $v = 2$; Select edge b. $w = 3$; $k_3 = 1$; visited$(b) = 1$.

Step 4: $v = 3$; Select edge c. $w = 4$; $k_4 = 1$; visited$(c) = 1$.

Step 5: $v = 4$; Select edge d. $w = 1$; $k_1 = 1$; visited$(d) = 1$.

Step 6: $v = 2$;

Step 7: Select edge e; $w = 4$; $k_4 = 2$; visited$(e) = 1$

Step 8: $v = 4$;

Step 9: Select edge f; $w = 2$; $k_2 = 2$; visited$(f) = 1$

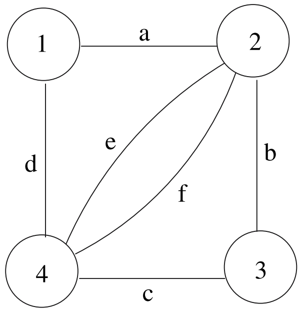
Step 10: Construct the cycle as 1–2–4–2–3–4–1.

Key Applications

Eulerian cycles find applications in problems where paths or cycles need to be found that traverse a set of edges in a graph. Such problems are generally called edge routing problems.

Snow Plow Problem

This problem requires finding the least distance route in the road network that starts and ends at the station so that snow



Graph Theory, Konigsberg Problem, Figure 4 Illustration of eulerian algorithm

can be cleared from the streets at the minimum cost. The minimum distance route is obviously the eulerian cycle since this cycle traverses each edge exactly once. However, it is unlikely that any real road network would happen to satisfy the necessary conditions that make it Eulerian. In that case, the problem moves to the realm of "the Chinese postman problem" [10,11,12].

Chinese Postman Problem

A postman delivers mail everyday in a network of streets. It is useful to know whether or not, the postman can traverse the network and return to the mail station without driving the length of any street more than once. If the network is not eulerian, the problem gets modified to the one where it is required to find the shortest path, which visits each edge at least once. This problem statement requires a parameter to be associated with each edge that represents the cost of traversing that edge. For example, cost can be the represented in terms of the length of the street, which the edge represents.

In a non-eulerian graph, the postman's circuit, shortest or otherwise, will repeat one or more edges. Every vertex is entered the same number of times that it is left so that any vertex of odd degree has at least one incident edge that is traversed at least twice. Chinese postman problem is formulated as an optimization problem where the total cost of repeated edges in minimized.

Algorithm

Step 1: Find the shortest path between each pair of odd-degree.

Step 2: Find the sub-graph G' with the odd degree vertices.

Step 3: Find the minimum weight matching of all the edges in G'. The edges in the shortest path connecting a matched pair of odd degree vertices should be repeated.

Figure 5 shows a sample graph with edge weights and the Chinese postman algorithm finds the least cost (minimum edge weight) path in the graph such that every edge is traversed at least once. Table 1 shows the shortest path costs between every pair of vertices, which is used by the algorithm to find the minimum weight matchings on edges (the matchings and their costs are listed in Table 2). The algorithm finds that the paths from vertex 1 to vertex 3, and the path from 2 to 4 must be repeated since this is the minimum cost matching (cost is 5). The algorithm finds the optimal route to be 1–2–3–4–2–4–1–3–1 in the graph shown in Fig. 5.

Capacitated Chinese Postman Problem

This problem arises where each edge has a demand and vehicles to be routed have finite capacities. For example, in applications involving road salting in winter season, there is a limit on the maximum amount of salt that a truck can carry. The amount of salt required is fixed for a road segment. The Capacitated Chinese postman problem finds the sets of routes from a single station that service all the road segments in the network at a minimal cost and subject to the constraint that the total demand of each route does not exceed the capacity of each truck. Christofides' proposed an algorithm to solve this problem.

Capacitated Arc Routing Problem

This problem is different from the Capacitated Chinese postman problem in that demands of some of the road segments can be zero. This situation can arise in road salting scenarios where state highways can be used for traveling, but need not be salted. These edges can be used to traverse between the edges that require the service.

Both the Capacitated Chinese Postman Problem and Capacitated Arc Routing Problem are NP-hard [12] and heuristic methods are normally used to obtain solutions.

Graph Theory

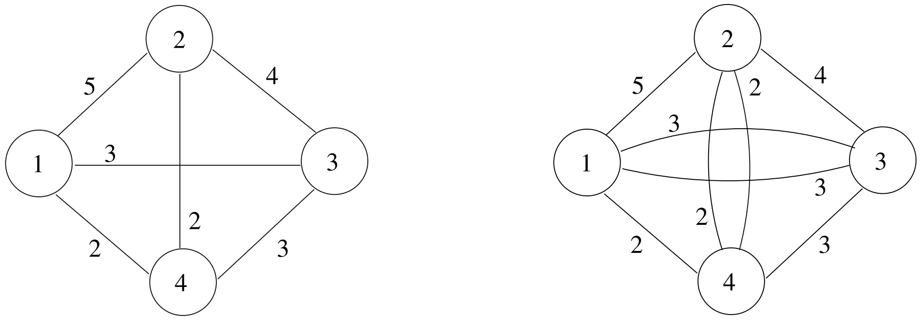
The Konigsberg problem had a powerful impact on mathematics, paving the way for the creation of a new modeling

	1	2	3	4
1	0	4	3	2
2	4	0	4	2
3	3	4	0	3
4	2	2	3	0

Graph Theory, Konigsberg Problem, Table 1 Shortest Path Cost between the pairs

Matching	Cost
(1,2),(3,4)	4+3=7
(1,4),(2,3)	2+4=6
(1,3),(2,4)	3+2=5

Graph Theory, Konigsberg Problem, Table 2 Matchings and Costs



Road Graph with edge weights

Optimal Route:
1–2–3–4–2–4–1–3–1

Graph Theory, Konigsberg Problem, Figure 5 Illustration of Chinese postman problem algorithm

theory called graph theory. The applications of graph theory are numerous in science and engineering. A few are listed below

Graph Theory in Spatial Networks The very fact that graph theory was born when Euler solved a problem based on the bridge network of the city of Konigsberg points to the apparent connection between spatial networks (e. g. transportation networks) and graphs. In modeling spatial networks, in addition to nodes and edges, the edges are usually qualified by adding weights that encode information like the length of the road segment that the edge represents. Connectivity and shortest paths in spatial networks have been extensively studied using graphs [13].

Graph Theory in Geography Graphs are also widely applied in geography in the modeling of stream systems. Streams have been modeled as hierarchical graphs and random graphs in the literature [14].
In addition to the applications described above, graphs find other wide applications, including modeling of social networks, molecular structures in Chemistry, computer networks, electrical networks and syntax structures in linguistics.

Future Directions

Relationships between eulerian graphs and other graph properties such as Hamiltonian property are being studied [7]. Graph, the mathematical model which owes its origin to Konigsberg bridge problem, is being increasingly applied to several evolving domains such as spatio-temporal networks, which has necessitated the incorporation of temporal dimension in graphs.

Recommended Reading

1. Chartrand, G.: The Königsberg Bridge Problem: An Introduction to Eulerian Graphs. In: Introductory Graph Theory. Dover, New York (1985)
2. Harary, F.: Graph Theory, pp. 1–2. Addison-Wesley, Reading (1994)
3. Biggs, N.L., Lloyd, E.K., Wilson, R.J.: Graph Theory 1736–1936. Oxford University Press, Oxford (1976)
4. Bollobás, B.: Graph Theory: An Introductory Course, p.~12. Springer, Berlin (1979)
5. Steinhaus, H.: Mathematical Snapshots, 3rd edn. Dover, New York (1999), pp. 256–259
6. Hartsfield, R.: Pearls in Graph Theory. Academic Press, San Diego (1990)
7. Fleischner, H.: (Some of) the many uses of Eulerian Graphs in graph theory, Discrete Math. **230**, 23–43 (2001)
8. Biggs, N.L., Lloyd, E.K., Wilson, R.J.: Graph Theory. Oxford University Press, Oxford (1998)
9. Wilson, R.J.: Introduction to Graph Theory. Longman Inc., New York (1985)
10. Edmonds, J., Johnson, E.L.: Matching, Euler tours and Chinese Postman, Math. Program. **5**, 88–124 (1973)
11. Christofides, N.: The optimum traversal of a graph, Int. J. Manage. Sci. **1**(6) (1973)
12. Evans, J.R., Minieka, E.: Optimization Algorithms for Networks and Graphs. Marcel Dekker Inc., New York (1992)
13. Shekhar, S., Chawla, S.: Spatial Databases: A Tour. Prentice Hall, New Jersey (2003)
14. Wilson, R.J., Beineke, L.W. (eds.): Applications of graph theory. Academic Press, London, New York (1979)

GRASS

MALTE HALBEY-MARTIN
Institute for Geographic Sciences, Freie Universität
Berlin, Berlin, Germany

Synonyms

Geographic Resources Analysis Support Software; GRASS-GIS; Visualization tool

Definition

GRASS-GIS (Geographic Resources Analysis Support Software) is a powerful open source software for geospatial analysis and modeling that can manage both raster and vector data. In addition it supports three dimensional modeling with 3D raster voxel or 3D vector data and contains several image processing modules to manipulate remote sensing data. It includes visualization tools and interacts with other related software such as the statistical software package R, gstat and Quantum GIS. GRASS supports a wide variety of GIS formats through the use of the GDAL/OGR library. It also supports Open Geospatial Consortium (OGC)—conformal Simple Features and can connect to spatial databases such as PostGIS via ODBC (Open DataBase Connectivity). GRASS datasets can be published on the internet with the UMN Mapserver software.

The software is published under the terms of the GNU General Public Licence (GPL). Anyone can access the source code, the internal structure of the software and the algorithms used. Therefore any user can improve, modify or extend the program according to his own needs. No licence fees have to be paid under the terms of the GPL. Programmers all over the world contribute to GRASS, one of the largest Open Source projects in the world with more than a million lines of source code.

GRASS runs on a variety of platforms including GNU/Linux, MS-Windows, MacOS X and POSIX compliant systems. It is completely written in C although a Java version also exists (JGRASS).

Historical Background

The history of GRASS dates back to the early 1980s when it was developed by the U.S. Army Construction Engineering Research Laboratory (CERL), Champaign, Illinois, USA to meet their needs for land management and environmental planning tools at military installations. Emphasis was placed on raster analysis and image processing because a principal goal was estimation of the impact of actions on continuous surfaces like elevation or soils [27] and there was no adequate raster GIS software on the market at that time. Modules for vector processing were added later.

The first version of GRASS was released in 1984 and because its development was financed by federal funds, US law required the release into the public domain. The complete source code was published on the Internet during the late eighties, a period during which there was significant improvement in its capabilities. CERL withdrew from GRASS development in 1995 and an international team took over this task. In 1997, GRASS 4.2 was published by Baylor University, Waco, Texas, USA. In 1999, GRASS 4.2.1 was released by the Institute of Physical Geography and Landscape Ecology, University of Hannover, Germany. Since GRASS version 4.2.1, GRASS has been published under the terms of the GPL of the Free Software Foundation. In 1999 the work on version 5.0 was started and the headquarters of the "GRASS Developer Team" moved to the Instituto Trentino di Cultura (ITC-irst), Trento, Italy. GRASS 5.0 was released in 2002, followed by version 6.0 in March 2005 which included a complete rewrite of the GRASS vector engine. The current stable version is 6.2 which was released at the end of October 2006.

GRASS is a founding project of the Open Source Geospatial Foundation (OSGeo) which was established in February 2006 to support and build high-quality open source geospatial software.

Scientific Fundamentals

Philosophy of GRASS

The most distinguishing feature of GRASS in comparison to other GIS software is that the source code can be explored without restriction so that anyone can study the algorithms used. This open structure encourages contributions by the user community to the source code in order to improve existing features or to extend it in new directions. For this purpose GRASS provides a GIS library and a free programming manual, which can be downloaded from the GRASS project site [16]. Under the terms of the GPL these contributions can not be distributed in proprietary software unless free access to the source code is granted to the end user. Any published code which is based on GPL licensed code must be licensed under the GPL as well.

GRASS offers the user a wide range of GIS functions. Together with other (free) software tools it provides a complete and powerful GIS software infrastructure at very low cost. GRASS is available on the project's homepage. The design is modular, consisting of more than 350 stand alone modules which are loaded when they are called into a GRASS session.

Interoperability: GIS and Analysis Toolchain

GRASS is designed in a way that offers a high level of robust interoperability with external applications, offering the user tremendous flexibility and efficiency in accomplishing analytic tasks.

Programming and Extending GRASS

GRASS is written in C and comes along with a sophisticated and well documented C/C++ application programming interface (API). As a side benefit of the open source philosophy, the user has the opportunity to learn how to develop new applications by using existing modules as examples and exploring their source code.

In addition, GRASS functions can be called with high level programming languages like Python using, for example, the GRASS-SWIG (Simplified Wrapper and Interface Generator) which translates ANSI C/C++ declarations into multiple languages (Python, Perl). An integrated parser is provided for scripting languages.

Extensions can be created easily using the extension manager, so that no programming of source code is needed to build additional GRASS modules. Moreover, this modular design helps the user to add new contributions to GRASS without affecting the software suite as a whole.

To automate repeating tasks in GRASS shell scripts can be written.

Relational Database Management Systems

GRASS can connect directly to relational database management systems (RDBMS) such as SQlite, MySQL and PostgreSQL. It supports PostGIS, the spatial extension of PostgreSQL and can connect to other external RDBMS via ODBC interface.

Statistical Analysis

R, a software environment for statistical computing and graphics, can be called within a GRASS session for statistical analysis of geodatasets. Similarly, there is a GRASS interface to gstat, a multivariable geostatistical modeling, prediction and simulation software package. Therefore gstat and R can access both GRASS raster and vector datasets for computations within the spatial region defined in GRASS. This capability creates the potential for both simple and complex geostatistical analysis as shown by [4] and [5]. GRASS can import and export Matlab® binary (.mat) files (version 4) for processing numerical calculations outside GRASS.

Interoperability with Other GIS Software

GRASS supports nearly all common GIS file formats enabling the use of other GIS applications or external datasources. Its binding to the GDAL/OGR library and the support of OGC Simple Features ensure that data exchange between various applications and between multiple users is straightforward. The internal file structure implemented in GRASS, coupled with UNIX-style permissions and file locks, allows concurrent access to any project. Several individuals can share the resources of a single machine and dataset.

GRASS works closely with Quantum GIS. GRASS modules are accessible through a GRASS plugin in Quantum GIS.

2D and 3D Visualization

GRASS comes with fully functional 2D cartography and 3D visualization software (NVIZ). It also interacts with other software tools to enable production of maps or visualization of geographic data sets. GRASS contains an export filter for Generic Mapping Tool (GMT) files and various image formats so that high quality maps for publication can be generated with external image manipulation software.

3D vector and raster datasets can be exported from GRASS as VTK (Visualization ToolKit) files which can be viewed in Paraview, a large data set visualization software package. Script files for Povray, a raytracer to design 3D graphics can be produced, as can VRML (Virtual Reality Modeling Language) files. Animations can be built with NVIZ or the external programs mentioned above.

Web Mapping

The UMN MapServer is able to connect to GRASS and can read GRASS geodatasets directly. With the help of PyWPS (Python Web Processing Service, an implementation of the Web Processing Service standard from the Open Geospatial Consortium) GRASS modules are accessible via web interface and can serve as a backbone in WebGIS applications.

Key Applications

GRASS is currently used around the world in academic and commercial settings as well as by many governmental agencies and environmental consulting companies. Due to the wide variety of applications for spatial data analysis, the following selection gives only a brief overview of situations where GRASS has been adopted. A collection of papers highlighting GRASS implementation can be found here [7].

Archaeology

GIS is of growing importance in this field. In fact GRASS has been widely used in archaeology to support the survey of excavation areas or to simulate ancient processes. GRASS has been used to model the most suitable place to conduct a survey in the Netherlands by [6]. Based on the assumption that settlement actions of ancient people show regional patterns, locations most suitable for archaeologic sites can be deduced. [11] used artificial neural networks as a tool to predict archaeological sites in East Germany and [22] extended GRASS to automate cumulative viewshed analysis. These examples also shows how the potential power of GIS increases when the software is modified by its users for specific needs. Pedestrian hunters and gatherers can be modelled in GRASS using MAGICAL, which consists of three separate GRASS modules [21] to simulate multiagent spatial behavior. [24] proposed how a GRASS based multidimensional GIS framework for archaeological excavations can be developed.

Biology

[31] used GRASS to model the distribution of three bird species in north-east England using a Bayesian rule-based approach. They linked data about habitat preferences and life-history of the birds against physiogeographic and satellite data using GRASS.

On the Iberian Peninsula [15] used GRASS to model the potential area of *Pinus sylvestris*. They predicted the habitat suitability with a machine learning software suite in GRASS incorporating three learning techniques (Tree-based Classification, Neural Networks and Random Forest) in their GIS-based analysis. All three techniques show a larger potential area of *P. sylvestris* than the present model. In the Rocky Mountains National Park tree population parameters have been modeled by [2] for the forest-tundra ecotone.

Environmental Modeling

GRASS is widely used in environmental modeling because of its strong raster and voxel processing capabilities. It offers a variety of techniques to conduct environmental modeling as described in [25]. Besides the use of custom-written models, GRASS includes a large number of models already implemented that can be used for hydrological analysis (Topmodel, SWAT, Storm Water Runoff, CASC2D), watershed calculations and floodplain analysis as well as erosion modeling (ANSWERS, AGNPS 5.0, KINEROS). Models for landscape ecological analysis and wildfire spread simulation also exist within GRASS.

Geography (Human/Physical)

GIS is used in a wide range of analyzes in human and physical geography because both subjects make extensive use of geodata or spatial geodatabases. GRASS is the GIS software of choice in many geographic surveys worldwide.

Geology/Planetary Geology

[20] used a digital elevation model (DEM) and a logical model of the geological structure to derive the surface boundaries of each geologic unit in their study area located in the Izumi mountain range. From these data they built a 3D model of the local geology.

GRASS has also been used in planetary geology. [13] identified Wrinkle Ridges on Mars which can be an evidence of existing subsurface ice on the planet. They used Mars MGS and Viking Mission data to perform their study. The mapping of geologic features from Mars data was done by [10]. The authors detected tectonic surface faults and assigned them to a geologic Mars region. The ability to import the raw data from various Mars datasets and to reproject them quickly and accurately is seen as a great benefit by the authors.

Geomorphology / Geomorphometry

Modules for surface analyzes in GRASS offer the possibility to derive terrain parameters like slope, aspect, pcurve and tcurve in one step. [3] has shown how the geomorphology of a exemplary study area in Kosovo can be statistically analysed with GRASS and R. From a subset of GTOPO30 elevation data he performed various statistic computations on selected relief parameters leading to a classification of geomorphologic units. [17] has used the combination of GRASS and R to perform morphometric analysis of a mountainous terrain in Brazil. With this package he derived morphometric parameters (hypsometry, slope, aspect, swat profiles, lineament and drainage density, surface roughness, isobase and hydraulic gradient) from DEMs and analysed these parameters statistically.

GRASS has also been used to define landslide susceptibility areas by [8]. They used a combination of GRASS with the gawk programming language to create landslide susceptibility maps of Parma River basin in Italy. They showed that very large datasets can be processed in GRASS quickly without problems.

The characterization of landscape units which are not only used in geomorpholgy but also in other scientific fields such as soil science and environmental modeling has benefited tremendously from GRASS in the past.

Geostatistics

[5] used a combination of GRASS, R and PostgreSQL to analyze various geodatasets. They showed that these techniques provide a powerful toolbox to analyze natural phenomena as well as socio-economic data.

Hydrologic Modeling

Hydrologic models like the USDA-Water Erosion Prediction Project (WEPP) model can be easily parameterized with GRASS as shown by [30]. [9] calculated a more appropriate flow time as an input for the flow analysis of a river in East Germany based on WaSiM-ETH. Besides the existing models incorporated in GRASS, custom-written models can be created as shown by [12]. They incorporated a Soil Moisture Routing model which combines elevation, soil and landuse data and predicts soil moisture, evapotranspiration, saturation-excess overland flow and interflow for a watershed.

Oceanography

For nautical hydrographic surveys GRASS offers some helpful modules to generate bathymetric surfaces by the interpolation of sounding data. [19] built up an environmental GIS database for the White Sea based on GRASS incorporating several hydrological and chemical parameters to validate numerical ecosystem modeling with the purpose to evaluate effects of climate change and human impact on this ecosystem.

Landscape Epidemiology and Public Health

With the help of GIS the spread of epidemics can be analysed or predicted. The outbreak of avian influenza in northern Italy in winter 1999–2000 was examined by [23]. GRASS and R were used to map the distribution of the outbreaks of highly pathogenic avian influenza which was caused by a H7N1 subtype virus.

To predict the risk of Lyme Disease for the Italian province of Trento GRASS has been used in several studies. The distribution of ticks infected with *Borrelia burgdorferi* s.l. was analysed by [28] with a bootstrap aggregation model of tree based classifiers in GRASS. The occurrence of ticks were cross-correlated with environmental data in GIS. [14] developed a spatial model of the probability of tick presence using machine learning techniques incorporated in GRASS and R.

A combination of GRASS, Mapserver and R is used by the Public health Applications in Remote Sensing (PHAiRS) NASA REASoN project. The objective of this project is to offer federal state and local government agencies dynamic information that might impact the spread of illnesses. Environmental and atmospheric conditions which affect public health are derived from NASA data sets and presented in a way that local public health officials can use them for decision making.

Remote Sensing

GRASS with its sophisticated raster processing capability and the already implemented image processing modules offer the user sophisticated tools for processing remote sensing data at low cost. The existing modules include functions for image preparation, image classification and image ratios. The software has also some functions for creating orthophotos and image enhancement. [26] give an overview of the tools for image processing in GRASS.

The potential to detect objects from airborne Laser Scanning data for urban mapping is described in [18].

Soil Science

GRASS is used in this field for several tasks because several useful tools have been developed by soil scientists.

Terrain parameters are important input to soil modeling and are widely used to assist in the mapping of soil properties. The aspect angle is commonly used by soil scientists as a proxy for the variation in surface moisture. Together with climatic data it is possible to derive a quantitative model of the surface soil moisture status of a landscape. The components of the solar radiation budget for each cell can be computed by GRASS using modules where site specific solar radiation is considered. [29] improved the predictive potential of pedotransfer functions. These are the basis of some hydrologic models with which soil hydraulic behavior can be characterized on a large scale. These terrain parameters included topographic information that was processed with the help of GRASS.

[1] derived a three dimensional continuous soil model with the help of GRASS. He used fuzzy sets to represent soil-landscape relations as fuzzy rules. With these rules he examined landscape information data which led to a three dimensional soil model.

Education

The GRASS community actively promotes the teaching of GRASS and other FOSSGIS (Free and Open Source Software GIS) programs to train the next generation in these forward looking techniques. Educational materials are available on the GRASS wiki.

Future Directions

The ongoing development of GRASS as a native Windows application and the implementation of a new unified Graphical User Interface for Linux, Mac, Windows and Unix using WxWidgets and Python will certainly increase distribution and use of the program. The prototype code is already written and is being tested. Its advantages in modeling, price and adaptability makes GRASS a strong alternative to other GIS software. Increasing popularity will logically lead to increased development efforts for the software. More people will contribute to the source code, bugtracking and documentation. GRASS has already some innovative functions implemented (e. g. functions for network analysis like shortest path, route planing), waiting for new applications to be developed on top. For 3D modeling the infrastructure and modules are in place for raster, vector and site data leading to an increasing usage in spatial modeling.

Cross References

- ▶ Open Geospatial Consortium
- ▶ PostGIS
- ▶ Quantum GIS
- ▶ University of Minnesota (UMN) Map Server

Recommended Reading

1. Ameskamp, M.: Three-dimensional Rule-based Continuous Soil Modelling. PhD thesis, Christian-Albrechts-Universitaet Kiel (1997)
2. Baker, W.L., Weisberg, P.J.: Using GIS to model tree population parameters in the rocky mountain national park forest-tundra ecotone. J. Biogeogr. **24**, 513–526 (1997)
3. Bivand, R.S.: Integrating grass 5.0 and r: GIS and modern statistics for data analysis. 7th Scandinavian Research Conference on Geographical Information Science. Aalborg, Denmark (1999)
4. Bivand, R.S.: Using the r statistical data analysis language on grass 5.0 GIS database files. Comp. & Geosci. **26**, 1043–1052 (2000)
5. Bivand, R.S., Neteler, M.: Open source geocomputation: using the r data analysis language integrated with grass GIS and postgresql data base systems (2000)
6. Brandt, R., Groenewoudt, B.J., Kvamme, K.L.: An experiment in archaeological site location: Modeling in the netherlands using GIS techniques. World Arch. **24**(2), 268–282 (1992)
7. Ciolli,M. Zatelli, P.: http://www.ing.unitn.it/~grass/conferences/grass2002/proceedings/proceedings/abstracts.html. online, 12 (2006)
8. Clerici, A., Perego, S., Tellini, C., Vescovi, P.: A procedure for landslide susceptibility zonation by the conditional analysis method. Geomorphol. **48**(4), 349–364 (2002)
9. Cullmann, J., Mishra, V., Peters, R.: Flow analysis with wasimeth model parameter sensitivity at different scales. Adv. Geosci. **9**, 73–77 (2006)
10. Deuchler, C., Wählisch, M., Gehrke, S., Hauber, E., Oberst, J., Jaumann, R.: Combining mars data in grass GIS for geological mapping. ISPRS Congress, 12–23 July 2004 Istanbul, Turkey, vol.#160;35, ISPRS, Istanbul (2004)
11. Ducke, B.: Archaeological predictive modelling in intelligent network structures. Doerr, M., Sarris, A. (eds.): CAA 2002 The Digital Heritage of Archaeology. Proceedings of the 30th CAA conference held at Heraklion, Crete, Greece, 2–6 April 2002, pages 267–273, Athens (2003)
12. Frankenberger, J.R., Brooks, E.S., Walter, M.T., Walter, M.F., Steenhuis, T.S.: A GIS-based variable source area hydrology model. Hydrol. Process **13**, 805–822 (1999)
13. Frigeri, A., Federico, C., Pauselli, C., Minelli, G.: Identifying wrinkle ridges from mars mgs and viking mission data: Using grass GIS in planetary geology. Trans. GIS **8**(2), 255–262 (2004)
14. Furlanello, C., Neteler, M., Merler, S., Menegon, S., Fontanari, S., Donini, A., Rizzoli, A., Chemini, C.: GIS and the Random Forest Predictor: Integration in R for Tick-borne Disease Risk Assessment. In: Hornik, K., Leisch, F. (eds.) Proceedings of the 3rd International Workshop on Distributed Statistical Computing, March 20–22, Vienna, Austria (2003)
15. Garzon, M.B., Blazek, R., Neteler, M., Dios, R.S., Ollero, H.S., Furlanello, C.: Predicting habitat suitability with machine learning models: The potential area of pinus sylvestris l. in the iberian peninsula. Ecol. Modell. **197**(3–4), 383–393 (2006)
16. GRASS Development Team. http://grass.itc.it/. online, 12 (2006)
17. Grohmann, C.H.: Morphometric analysis in geographic infromation systems: applications of free software grass and r. Comput. & Geosci. **30**, 1055–1067 (2004)
18. Höfle, B., Rutzinger, M., Geist, T., Stötter, J.: Using airborne laser scanning data in urban data management – set up of a flexible information system with open source components. In: Proceedings UDMS 2006: Urban Data Management Symposium, Aalborg, Denmark (2006)
19. Kaitala, S., Shavykin, A., Volkov, V.A.: Environmental GIS database for the white sea. In: Proceedings of the Open source GIS – GRASS users conference, 11–13 September 2002, Trento, Italy (2002)
20. Kajiyama, A., Ikawa, N., Masumoto, S., Shiono, K., Raghavan, V.: Three-dimensional geological modeling by foss grass GIS – using some field survey data –. In: Proceedings of the FOSS/GRASS Users Conference – Bangkok, 12–14 September 2004, Bangkog, Thailand (2004)
21. Lake, M.W.: The use of pedestrian modelling in archaeology, with an example from the study of cultural learning. Environ. Plann. B: Plann. Des. **28**, 385–403 (2001)
22. Lake, M.W., Woodman, P.E., Mithen, S.J.: Tailoring GIS software for archaeological applications: An example concerning viewshed analysis. J. Arch. Sci. **25**(1), 27–38 (1998)
23. Mannelli, A., Ferre, N., Marangon, S.: Analysis of the 1999 2000 highly pathogenic avian influenza (h7n1) epidemic in the main poultry-production area in northern Italy. Prev. Vet. Med., **73**, 273–285 (2006)
24. Merlo, S., Colin, A.: Developing a multidimensional GIS framework for archaeological excavations. In: CIPA 2005, XX International Symposium Torino, Torino (2005)
25. Mitasova, H., Mitas, L., Brown, W.M., Gerdes, D.P., Kosinovsky, I., Baker, T.: Modelling spatially and temporally distributed phenomena: new methods and tools for grass GIS. Int. J. Geogr. Inf. Syst. **9**(4), 433–446 (1995)
26. Neteler, M., Grasso, D., Michelazzi, I., Miori, L., Merler, S., Furlanello, C.: An integrated toolbox for image registration, fusion and classification. Int. J. Geoinf. Special Issue on FOSS/GRASS 2004 & GIS-IDEAS 2004, **1**(1), 51–60 (2005)
27. Neteler, M., Mitasova, H.: Open Source GIS: A GRASS GIS Approach. Number 773 in SECS. Kluwer Academic Publishers/Springer, Boston, first edition, June 2001. ISBN: 1–4020-8064-6; also published as eBook: ISBN 1-4020-8065-4
28. Rizzoli, A., Merler, S., Furlanello, C., Genchi, C.: Geographical information system and bootstrap aggregation (bagging) of tree-based classifiers for lyme disease risk prediction in trentino, italian alps. J. Med. Entomol. **39**(3), 485–492 (2002)
29. Romano, N., Palladino, M.: Prediction of soil water retention using soil physical data and terrain attributes. J. Hydro. **265**, 56–75 (2002)
30. Savabi, M.R., Flanagan, D.C., Hebel, B., Engel, B.A.: Application of WEPP and GIS-grass to a small watershed in indiana. J. Soil Water Conserv. **50**(5), 477–483 (1995)
31. Tucker, K., Rushton, S.P., Sanderson, R.A., Martin, E.B., Blaiklock, J.: Modelling bird distributions – a combined GIS and bayesian rule-based approach. Landsc. Ecol. **12**(2), 77–93 (1997)

GRASS-GIS

► GRASS

Grid

ILYA ZASLAVSKY
San Diego Supercomputer Center, University
of California San Diego, San Diego, CA, USA

Synonyms

Computational grid; Data grid; Computational infrastructure

Definition

Grid computing is a model for organizing large-scale computation and data management such that process execution tasks are distributed across many network computers, and data transmission occurs over high-speed parallel network connections, in a system that resembles parallel computer architecture. Grids rely on "resource virtualization", that is, organize different locally-managed processing and data nodes as relatively homogeneous "grid resources" that follow open standards and enable quality of service through resource sharing and service-level agreements.

Main Text

Grid computing is defined via a series of specifications being developed within the Open Grid Forum (OGF), a collaboration between industry and academia. Globus Toolkit is a middleware component implementing OGF standards. Initial ideas of grids, and grid service specifications, have been explored in seminal articles by Ian Foster, Carl Kesselman and Steve Tuecke. In recent years, grid services and open grid services architecture (OGSA) have evolved towards better compatibility with web services and service-oriented architecture principles. The Web Service Resource Framework (WSRF) is a set of web service specifications describing OGSA implementation using web services.

Several research projects have applied grid computing principles to spatial information processing. In particular, they include development of stateful grid services for assembling thematic maps within the Geosciences Network (GEON) project (support for stateful services environment is a distinct feature of grid services compared with web services). Another direction is relegating spatial information processing to P2P networks, in particular, development of decentralized indexing and routing techniques for spatial data. The goal is to enable dynamic insertion or deletion of spatial information so that processing is adjusted to the set of currently available nodes in a P2P network. Grids can primarily focus on aggregating processing power of distributed computers (computational grids), or on secure management and sharing of large data volumes (data grids), or on remote control of scientific instruments (equipment grids). However, many grid implementations share features of all the three types.

Cross References

- ► Cyberinfrastructure for Spatial Data Integration
- ► Grid
- ► Spatial Information Mediation
- ► Web Services

Grid Computing

► Grid, Geospatial

Grid, Geospatial

AIJUN CHEN, LIPING DI, YAXING WEI, YUQI BAI
Center for Spatial Information Science and Systems,
George Mason University, Fairfax, VA, USA

Synonyms

Geospatial data grid; Geospatial computing grid; Geospatial computational grid; Grid computing; Distributed computing

Definition

Geospatial grid refers to the *grid*, an emerging cyberinfrastructure, dealing with geospatial information (aka spatial information grid). The geospatial grid includes geospatial data grids (GDGs) and geospatial computational grids (GCGs) [1,2]. A geospatial grid leverages geospatial standards and interfaces to handle geospatial data, information, and processes. A GDG is used for registering, managing, retrieving, and securely copying and moving geospatial data and its replicas among data centers distributed around the world. A GCG focuses on processing and analyzing geospatial data and their replicas by utilizing distributed geospatial resources in the grid environment. Both GDG and GCG are based on the fundamental grid standards, the Open Grid Services Architecture (OGSA) and the Web Service Resource Framework (WSRF) [3]. A geospatial grid is an advanced, open, service-oriented, and distributed geospatial data and computing system. Figure 1 in the later of the paper illustrates an integrated system of a geospatial data grid and a geospatial computing grid. Geospatial grid services can be combined to form a service chain, which acts as a new grid service, for providing new geospatial functions. *Geospatial semantic grid* and *geospatial knowledge grid* are used to refer to the application of ontology and artificial intelligence to geospatial grids.

Historical Background

The grid is used for the next generation data- and computing-intensive information infrastructure providing a number of new capabilities that emerged in the late 1990s [4]. The Globus Toolkit, including (a) security, (b) data management, (c) information services, (d) execution management, is a de facto standard middleware for implementing grids. The grid was applied to many disciplines and research areas, such as the the Earth System Grid (ESG) project for Earth science [5], the Grid Physics Network and the Particle Physics Data Grid projects for physics [6,7]; Austria's BioGrid, the United States' Bioportal, Europe's Bioinformatics Grid, and South Africa's National Bioinformatics Grid projects for bioinformatics; and the National Virtual Observatory (NVO) for astronomy [8]. The application of the grid in geospatial field originates the term geospatial grid.

In 2001, the Chinese Ministry Of Science and Technology (MOST) sponsored a spatial information grid. The San Diego Super Computing Center developed the Spatial Resource Broker for constructing a geospatial data grid. In the early 2000s, the National Science Foundation (NSF) supported several geospatial-related information technology researches, such as the Geosience Network (GEON) (www.geongrid.org/) and Network for Earthquake Engineering Simulation (NEES) (www.nees.org/). Di [1,2] introduced the concept of a geospatial grid in 2004 through the project on "The Integration of Grid and OGC (Open Geospatial Consortium) (IGO) Technologies for Earth Science Modeling and Applications" sponsored by the National Aeronautics and Space Administration (NASA) Advanced Information System Technology (AIST) Program. Following these activities, the geospatial grid concept was used in many journal and conference papers [2,10].

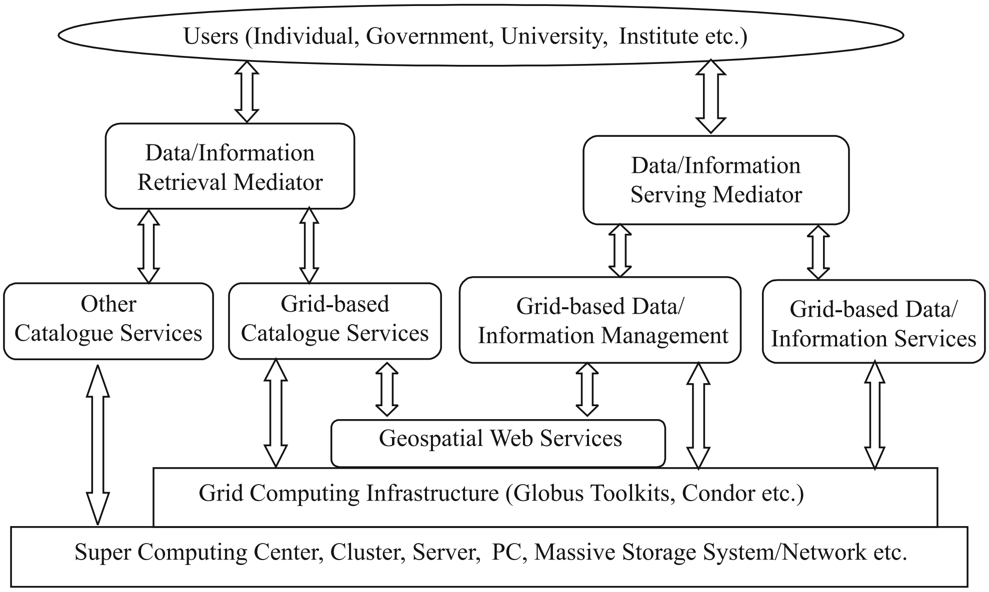
The IGO project addressed how to effectively and efficiently share, process, visualize, and serve geospatial resources using grid technology and OGC web services standards. In the project, a number of OGC web services were implemented as grid services and deployed in the Globus-based grid environment. Those OGC services included the web mapping service (WMS), the web coverage service (WCS), the catalogue service/web profile (CS/W), the web image classification service (WICS), and the web coordinate transform service (WCTS). They formed the foundation of the geospatial grid for providing basic geospatial functions such as registration, updating and retrieval of geospatial data and services, and subsetting, reprojection, reformatting, resampling, and visualization of distributed geospatial data. In addition to fully integrating grid services from the Globus Toolkit into the geospatial grid, a number of new generic grid services have also been devised and implemented to strengthen the geospatial grid.

Scientific Fundamentals

The geospatial grid provides a promising mechanism to share worldwide huge volumes of heterogeneous geospatial computing and data resources to facilitate important research programs, such as climate change, weather forecasting, and natural hazard prediction. To take full advantages of grid technology for effectively sharing distributed geospatial resources and facilitating their interoperability, a geospatial grid must be established following geospatial and Grid technology standards.

A geospatial data grid is for managing and sharing large volumes of geospatial data. The top level of a GDG architecture consists mainly of such grid services as a secure data transfer service (GridFTP), a data replica service (DRS), a replica location service (RLS), and a reliable file transfer service (RFT). A geospatial computational grid (GCG) is focused on the sharing and utilization of geospatial computing capabilities. It provides distributed supercomputing, high-throughput, on-demand, data-intensive and collaborative computing. Most of the geospatial-related applications demonstrate a tremendous appetite for computational resources (CPU, memory, storage, etc.) that cannot be met in a timely fashion by expected growth in single-system performance. OGC web service standards play a very important role in a GCG. (Fig. 1).

The architecture includes a geospatial data grid and a geospatial computational grid. Grid computing infrastructure consists of the popular grid middleware: Globus Toolkit and Condor. Geospatial management and serving grid services can be developed using geospatial web services and other geospatial services, e. g., Geographic Resources Analysis Support System (GRASS), and following both grid standards and geospatial standards. A geospatial grid catalogue service can be developed following the geospatial catalogue standard for registering and querying geospatial grid services in the grid environment. A grid retrieval mediator can be designed to coordinate the user's query for multiple catalogue services and data replica index services. All geospatial



Grid, Geospatial, Figure 1 The architecture of a geospatial grid

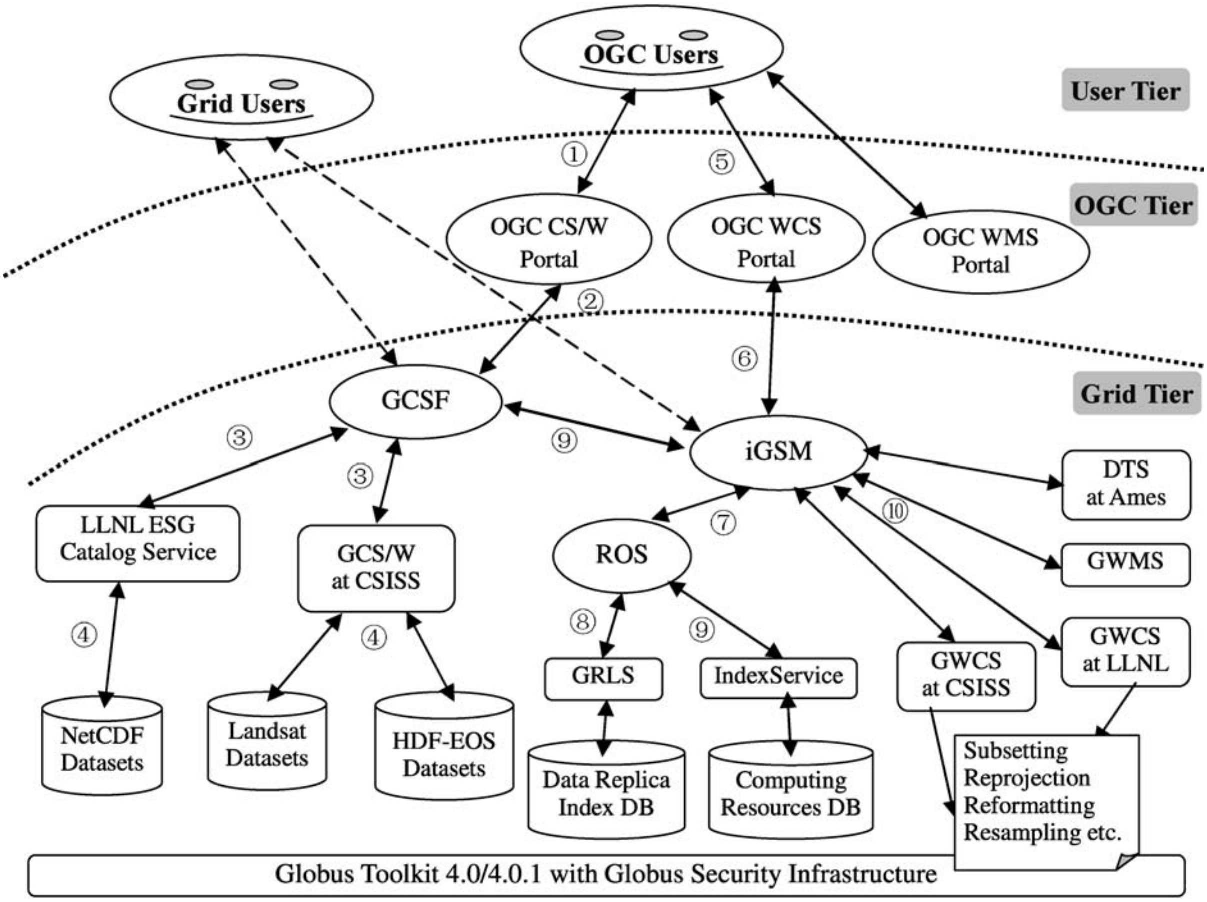
data/information and grid services are registered to catalog services. Geospatial data/information serving mediators are used to coordinate users' requests among all grid services and send each request to the grid services with best performance. Optimized grid service can be a pure grid service and implemented based on grid software. All components in the geospatial grid are grid service-oriented. A service to manage job submission [e. g., Web Service Globus Resource Allocation Manager (WS GRAM)] and schedule grid services (e. g., Condor) can be used to manage all submitted jobs.

Prototype

OGC has developed a set of widely accepted spatial web services for geospatial interoperability, such as the WMS, WCS, and CS/W specifications. Currently, numerous services complying with those specifications have been developed and deployed worldwide [10]. By adopting OGC specifications and providing OGC services in the grid environment, the geospatial grid not only extends the grid capabilities to geospatial disciplines and enrich the grid computing technology, but it also makes it possible to easily use the newest computer technology available and to resolve some large geospatial questions that are difficult to resolve without the geospatial grid.

Figure 2 is the architecture and dataflow diagram of a standards-compliant geospatial grid system prototype. The prototype consists of three tiers: the user tier, the OGC tier, and the grid tier [11]. The OGC tier provides OGC-compliant clients with standard OGC interfaces that are supported by grid technology. To an OGC client, this tier acts as a set of OGC compliant servers, but to the grid tier, the OGC tier operates as an authorized user of the grid tier. Any user request compliant with OGC standards can be submitted through these interfaces and processed through grid computing resources. The grid tier provides grid users with standard grid service interfaces. Any authorized grid user in a virtual organization (VO) can access any grid resources in the same VO or any other authorized VO. Therefore, the geospatial grid hides the complexity of the grid and the managed resources. It provides standards-compliant interfaces for geospatial users and gives them transparent access to grid-enabled distributed resources.

The OGC Standard-compliant web portals, including a CS/W portal, a WCS portal and a WMS portal, are based on grid technology. Through those portals, any OGC-compliant clients can use all the resources managed by the grid. Grid enabled GSW (GCS/W) [12] is based on the OGC catalog service specification. The information model of OGC CS/W has been extended to satisfy the requirement for describing Earth observation data while complying with the International Organization for Standardization (ISO), the Federal Geographic Data Committee (FGDC), and NASA EOS (Earth Observation System) Core System (ECS) metadata standards [13]. The Grid-Enabled Catalog Service Federation (GCSF) acts as the catalogue mediator. It aims to federate different catalog services into one logical service that provides a standard uniform interface to users. The GCSF is designed to integrate all grid-enabled



Grid, Geospatial, Figure 2 Architecture and dataflow of an open standard compliant geospatial grid. *CS/W* Catalog service for web, *WCS* web coverage service, *GWCS* grind-enabled WCS *WMS* web map service, *GWMS* grid-enabled WMS, *GCSF*: grid-enabled Catalog Service Federation, *iGSM* intelligent grid service mediator, *ROS* replica and optimization service, *DTS* data transfer service, *GRLS* grid replica location service, *ESG* Earth System Grid, *LLNL* Lawrence Livermore National Laboratory, *Ames* NASA Ames Research Center, *CSISS* Center for Spatial Information Science and Systems

catalog services within a Virtual Organization (VO). In the prototype system, the GCSF has federated the ESG catalog services from LLNL (Lawrence Livermore National Laboratory), GCS/W from CSISS at GMU, and NASA ECHO (EOS ClearingHOuse). The GCSF mainly maps metadata, transfers protocol-based XML documents, and merges results.

The grid-enabled web coverage service (GWCS) can be accessed indirectly by any OGC user through the WCS portal and directly by authorized grid users. The main customizable data access interface of GWCS is getCoverage. GWCS contains some built-in geoprocessing functions, such as subsetting, reprojection, reformatting, and georectification. These functions provide flexible operations with which the users can customize their desired data. The grid-enabled web mapping service (GWMS), on users' requests, dynamically produces static maps from spatially referenced geospatial data managed by the grid. The getMap operation is grid-enabled to produce maps in the formats of PNG (Portable Network Graphics), GIF (Graphics Interchange Format), or JPEG (Join Photographic Experts Group).

With the grid computing infrastructure, all OGC standard interfaces are implemented in the prototype system. By collaborating with the intelligent grid service mediator (iGSM), the replica and optimized service (ROS) and the data transfer service (DTS), accelerated processing speed, better serving performance, and data quality are obtained.

The intelligent grid service mediator (iGSM) is a grid service. It receives and responds to user requests from the OGC WCS/WMS portal, and mediates the received user requests among ROS, DTS, GCS/W, GWCS, and GWMS. The iGSM service sends the user request, including the logical data address, to ROS to get back the best performing service that serves real data. If no service is returned by ROS, iGSM will query GCS/W to get all the GWCS/GWMS services in the VO and all replicas of the required data and then select the best service to serve the data with the best performance. When data transfer between two machines in the VO is required, DTS is invoked with the authentication operation.

ROS is used to optimize user requests for data and services based on the grid technology. It provides a mechanism for managing grid-enabled geospatial data replicas and the best service and data selection mechanism. It integrates the Globus Toolkit's IndexService and Replica Location Service (RLS). The IndexService is used to obtain dynamic information for every machine that hosts GWCS/GWMS. Meanwhile, RLS manages all distributed replicas of every registered geospatial data item. ROS will identify all machines that can serve the data the user requires and then select the one most appropriate for their current resources available, e. g., CPU, memory, storage.

DTS works as a grid service to be invoked by other grid services when secure data transfer between different machines in the same VO or a different VO is needed. This data transfer is secured strictly through the grid security infrastructure (GSI). Both a user certificate and a host certificate are involved in identifying the authorization and authentication. It is based on the GridFTP, which is a command line component of the Globus Toolkit. Grid-enabled RLS (GridRLS), is based on the Globus RLS and is a grid service that enables RLS to orchestrate with any other grid services. RLS is a component of the Globus Toolkit. It can be run from the command line, but it cannot be run as a grid service in a Globus Toolkit 3.0 or above environment.

Key Applications

The geospatial grid can be applied in many geospatial science domains for large-scale data- and computing-intensive research and applications to acquire data, deploy services, construct models, execute workflows, and disseminate results. Some applications are global climate change, weather forecasting, crop yield, natural hazard, and air quality, etc.

Global Climate Change

Global climate change is a change in the long-term weather patterns that characterize the regions of the world. Research on global climate change requires analysis of large volumes of historical and current data from sensors aboard various platforms in addition to field observations. Computing-intensive models are routinely used in such research. The required massive quantities of geospatial data and large computing capabilities can be made available through the geospatial grid.

Geospatial Product Virtualization

In the geospatial grid, users can register their own geospatial grid services in the system catalog. They can use existing grid services to construct service chain to express geoprocessing workflows and register the workflows into the catalog as virtual geospatial products [14]. The virtual geospatial products can be materialized by executing the workflows in the grid-enabled workflow engine when users request them. The workflow associated with a virtual geospatial product can be considered to be a new grid service that can be chained into another workflow. For example, in calculating landslide susceptibility, a user can create Slope and Aspect grid services and register them into the catalogue. Then, the user can compose a landslide susceptibility workflow based on the Slope service, Aspect service, GCS/W, GWCS, and GWCTS. The composition of the workflow is based on users' knowledge: GCS/W can help users to find the digital elevation model (DEM) data they need, and GWCS can serve this DEM data to the Slope and Aspect services. The intermediate results produced can be further processed through the landslide susceptibility service, and finally, the landslide susceptibility parameters can be obtained for landslide forecasting [14].

Future Directions

The advancement of information technology has meant the accumulation of increasing amount of digitized data/information/knowledge (DIK), of which 80% are geospatially related. Currently both DIK and computing resources are not fully utilized. It is expected that research facilitating the application of grid technology to derive DIK, especially geospatial, will become increasingly important.

The grid-enabled collaborative and intelligent environments for geospatial computing and modeling that can handle huge amount of shared geospatial resources will be a focus of future research. Ontology-driven semantic workflow is a key technology for modeling geospatial applications in the Geospatial Grid.

The Geospatial Grid not only offers conventional functions of geographical information systems, it also provides an open, ubiquitous and intelligent platform for authorized users to share their knowledge and innovative ideas with the community and quickly convert the knowledge and ideas into operational capabilities within the Geospatial Grid.

Recommended Reading

1. Di, L.: Geospatial grids, geospatial modeling, and virtual geospatial products. In: Earth Observation and Geospatial Technology

for Civil Society, University College London, London, 23–25 June 2004
2. Di, L.: The geospatial grid. In: S. Rana, J. Sharma (eds.) Frontiers of Geographic Information Technology, pp 121–138. Springer-Verlag, Berlin, Heidelberg (2006)
3. Globus (2006) Globus Toolkit 4.0 Release Manuals. http://www.globus.org/toolkit/docs/4.0/. Accessed 22 Sept 2006
4. Chen, A., Di, L., Bai, Y., Wei, Y.: Grid computing and its application to geoinformatics. In: Karimi, H. (ed.) The Encyclopedia of Geoinformatics. Idea Group Inc., Hershey, PA, USA (2006, in press)
5. ESG (2006) Earth System Grid. https://www.earthsystemgrid.org/home/publicHomePage.do. Accessed 10 Sept 2006
6. GriPhyN (2006) Grid Physics Network. http://www.griphyn.org/projinfo/intro/appl_domains.php. Accessed 25 Sept 2006
7. PPDG (2006) Particle Physics Data Grid. http://www.ppdg.net/. Accessed 26 Sept 2006
8. NVO (2006) National Virtual Observatory. http://www.us-vo.org/grid.cfm. Accessed 28 Sept 2006
9. Di, L., Yang, W., Chen, A., Liu, Y., Wei, Y., Mehrotra, P., Hu, C., Williams, D.: The development of a geospatial grid by integrating OGC Web Services with Globus-based grid technology. In: GridWorld/GGF15, Boston, MA, October 3–6, 2005
10. Chen, A., Di, L., Bai, Y., Wei, Y.: An optimized grid-based, OGC standards-compliant collaborative software system for serving NASA geospatial data. In: The 2nd IEEE Systems and Software Week (SASW 2006), 30th Annual Software Engineering Workshop. Loyola College Graduate Center, Columbia, MD, USA. April 25–27, 2006
11. Di, L., Chen, A., Yang, W., Liu, Y., Wei, Y., Mehrotra, P., Hu, C., Williams, D.: The development of a geospatial data grid by integrating OGC Web Services with Globus-based grid technology. Concurrency and Computation: Practice and Experience (2006)
12. Chen, A., Di, L., Wei, Y., Liu, Y., Bai, Y., Hu, C., Mehrotra, P.: Grid enabled geospatial catalogue web services. In: American Society for Photogrammetry and Remote Sensing 2005, Baltimore, MD, March 7–11, 2005
13. Wei, Y., Di, L., Zhao, B., Liao, G., Chen, A.: Transformation of HDF-EOS metadata from the ECS model to ISO 19115-based XML. Comput. Geosci. **33**(2):238–247 (2007)
14. Di, L., Chen, A., Bai, Y., Wei, Y.: Implementation of geospatial product virtualization in grid environment. In: The Sixth Annual Earth Science Technology Conference – ESTC2006, College Park, MD, June 27–29, 2006

Gridded Data

► Oracle Spatial, Raster Data

G-Ring

► Spatial Data Transfer Standard (SDTS)

Group Decisions

► Participatory Planning and GIS

Group Spatial Decision Support Systems

► Geocollaboration

GSDSS

► Geocollaboration

GSN

► Geosensor Networks