A Framework for Name Placement Problem Solving in GIS

Vladimir Miroshnikov
Department of Computer Systems and Technologies,
Moscow State Engineering Physics Institute
(Technical University)
Moscow, Russia
vladmir@mail.vivos.glasnet.ru

Evgueni Tchepine
Department Computer Systems and Technologies,
Moscow State Engineering Physics Institute
(Technical University)
Moscow, Russia
jeka@minas.rosmail.com
chepin@vist.dozen.mephi.ru

Abstract

A map is a medium of communication that uses graphical symbols to represent geographical features and their relationships. One important part and a key factor in a map's ultimate effectiveness as a communication medium is the labeling of the feature, referred as name placement. This paper is concerned with the matter of name placement problem. Authors present general formulation and constraints and briefly overview the history of the problem. Complexity of the name placement is noted, which lead to a use the system approach to achieve map's effectiveness. This work is an attempt to summarize recent research and construct a framework for solve problem at different levels of system using classification, geodata modelling, generalization, computational geometry algorithms, range indexing and query optimization, rules methodology, evolutionary algorithms.

1. Introduction

An efficient interactive work with electronic map depends on different parameters of GIS functionality. Correctly generated electronic map is a result of thorough geodata modeling and applying a quantity of algorithms to process heterogeneous cartographic information. Such a map presents information both in explicit form by determining location of the features, and in implicit form by indicating relationships between different features. Feature labeling plays a vital role in the increasing of the map readability. Label (or name) defines a proper feature, feature semantics and shows up feature shape and relationships with other features. Working with a map user firstly relates to the textual information, so correctly placement of such an information is a key factor of the interaction effectiveness.

The first attempt to automate the name placement problem was addressed more than 25 years ago (Yoeli, 1972). This research served as a starting point to the further explorations, which was resumed in early 1980s, when at first Hirsh, Kelly, and then Basolgu, Ahn, Freeman and Balodis, defined principles and developed first name placement programs. Thus they laid the good foundation for further investigation.

All the early approaches were limited regarding the degree of the feature density (and thus names) with which they could cope. Clearly, as the density of features increases, the degree of freedom for satisfactorily placing the names decreases, that lead to the dramatically degradation of placement quality.

Most of the first name placement systems determine location of the name analytically by examining the position of other features and names, which may influence the placement [3]. But the degree of feature density of the real map may be much more than the one in a such systems, so this approach isn't practical because of large number of names and features that can directly or indirectly influence the placement. In early 1990s there was some research that account for map density problem [3]. Undoubtedly, taking into account the map density constraint was lead to more significant complexity in name placement system design.

This article outlines the problem complexity and proposes to use the system approach to solve it. According to it we will find solution at a different levels of the system. We will describe different solving techniques for each level.

Proceedings of the Workshop on Computer Science and Information Technologies CSIT'99
Moscow, Russia, 1999

2. Problem formulation

A map is a set of features that represents a geographic region. Each map feature is represented and modeled by an object, in this article we will refer it also as geoobject. Each geoobject has name and is labeled by its name on the map. Common label shape must denote the shape of its geoobject and show up explicit and implicit relationships with other geoobjects. Each labeled geoobject is represented on the map by a point sign, line, region, or its combination. So, there are three basic object's geometry types. Correspondingly, there are three basic name placement rules. Name placement is a function of proper geoobject (i.e. of its type, location, and semantics), of other nearby geoobjects, and of other names already on the map.

There are several basic requirements with no relation to geometry type of labeled object introduced by Imhof [4]:

- Readability. The names should be easy to read and to find on map.
- Definite attachment. The name and object it labeled should be fine to recognize.
- Avoidance of overlaps. One should avoid an overlapping between the name being placed and other features and names.
- Spatial integration. The name should clearly promote to show up spatial location, extent, relationships, significance and differentiation of the objects.
- Site identification. The name type should be concerned with classification and hierarchy of the objects.
- Overall aesthetics. The names on the map should be nor too rarefied, nor too thickened.

There are some additional constraints imposed by geometry types of objects (according to [3]).

Rules for place names of the region objects:

- The name should occupy the entire region and correspond to the shape of region, leaving free space on the both ends. It's preferably horizontal location.
- Non horizontal name location should not be rectilinear, but curvilinear with the curvature at the most 60 degrees.
- The name that is read away from the horizontal is preferably than one that is read towards the horizontal.

Rules for place names of the line objects:

- The shape of the label should be the same as the line contour.
- One should avoid complex contours.
- Characters of the name shouldn't be rarefied. Instead, the label may be repeated along the line with some interval.
- The name should be above the line for horizontal objects. This rule is more complex for vertical objects. If line is in the left part of the map than the name should be sited at the left of the line from bottom to top. For the opposite case, the name should be sited at the right of the line from bottom to top.
- One should avoid siting the name near the end point of the line.

Rules for place names of the point objects:

- The label should be horizontal.
- Names shouldn't be rarefied.
- The name should be located near point object it labeled, but with some minimal distance between them.
- It's desirable to position name a little above and at the right of the object.

Depending on the area of application GIS may impose some specific constraints. Thus, the constraint domain should be opened to assimilate future insertion and modification.

3. Problem solving methodology

As a rule the first approaches were based on a specific database that was excessive with respect to the basic GIS database. The placement program was an isolated part of the GIS, or mostly belonged to the automated cartography systems. In general, such programs are used at the special workstations for cartographers and other specialists in field of preparing cartographic information for end users. So, the first solutions were intended for production of the static cartographic compositions (i.e. for printing maps or for making raster map files for the documentation use). In many respects such specialization is a result of a significant resource use to solve the problem, and insufficient perfection of proper GIS.

During the last decade GIS technologies became one of the most dynamically developing. Software integration led to the merging and close interaction between different software components processed spatial information, which caused a lot of additional difficulties. In particular, the problem of the high-quality representation of the textual information on the map comparatively solved in automated cartography systems arose in face of developers of the GIS. But the constraint domain of the problem is increased: besides the static factor was added dynamic one. The problem was complicated in many times.

The dynamic component of the problem includes the fact of frequently changed map scale during the real work and temporal dynamics of the geographical objects. The influence of the dynamics is tried to reduce bringing the problem to the static variant where it may be solved more easy and quickly. On the other hand, it's obviously, that a full projection of the problem to the static component without risk of considerable loss of effectiveness is impossible. Thus, one should balance the solution optimally without deviation to the one of the two extremes.

It's impossible to solve the problem considering it as a separate monolith block which only take into input a one set of data and return an another one. GIS should be able to support the problem solving at different levels, i.e. the problem should be deeply introduce into the system. So, the use of the system approach for effective problem solving it's necessary. The aim of this article isn't an introduction of the use of system approach relative to GIS (for this aim see [5]); we will only take advantage of the results of such a use.

According to system approach there are three basic levels in GIS: accumulation, storage and representation. We will discuss the second and the third levels. Storage level requires taking into account the increasing significance of names with respect to the traditional approach when model data. Complicating of the data model is consisted in further formalization of name attribute information with the aim to get such description of label position and shape, that allows quickly computing real screen representation parameters at the subsequent levels. In comparison with modeling features the peculiarity consists in combining the plane coordinates and the geodesic coordinates of labels. There is also some mathematics description of label shape, e.g. description of label curve (see above). So, it's observed a shift from the simple objects (or even attributes) represented the name toward the full-fledged complex objects, that allows producing label with complex shape specified in combined reference system. In this sense labels and geoobjects become the objects with the same weight.

The modeling at the representation level implies a requirement of close interaction of two threads processed geoobjects and names. The first thread fetched geoobject labeled generates the second one. As in other respects the labels and geoobjects are equal in rights, as it's possible to apply the methodology of process of geoobjects with respect to names considering its specific

Thus, there are two opposite methods of problem solving as a whole. The first case is in the developing isolated software interacting with GIS through some program interface, and the second is in the close integration of name placement software with GIS. But the practice suggests to take into account both cases, i.e. as far as the problem developed introduce support of solving name placement into basic GIS at different levels modeling name storage at data storehouse and presenting program interfaces for handling fetched names and modeling annotation map layer. Such approach requires an exact object scheme of developing GIS as a whole. Having presented a common direction of problem solving, it's necessary to describe in details the solution architecture and discuss some methods of solving particular subtasks.

4. The common framework for problem solving

Let us start from the modeling names in cartographic data storehouse. In general name's object consist of the reference to the geoobject, position relative to the geoobject and font attributes. But in fact only name position is concerned with name placement problem directly. It's very difficult to formalize this attribute. Perhaps, it's impossible to exactly describe a spatial location of the label at the storage level, which is the reason of introducing dynamics. The rest of attributes are under jurisdiction of the GIS as a whole. All the attributes are modeled based on both classification and group processing principles (see bellow). *Thus, the first task is the formalization of description of the name location within limits of storage model.*

The modeling of the map representation (or rather map of the geographic region) is preceded by fetching geoobjects from the cartographic database. Both processes handling geoobjects and names must close interact during the fetch stage. From the point of view of the group handling both processes should be based on the principles of classification and generalization of data, using correspondingly model of the classification data (such as the system of classifiers or metadata) and generalization algorithms. From the point of view of the personal processing it may take into account informational attributes of the personal objects. The intermediate representation model, describing the entire label attributes except for a precise position (i.e. a text of label and font attributes) is a result of name processing at this stage. This model must include some range index, which will be used at subsequent stage to perform geometric operations. *Thus, an effective process organization at fetch stage is the second key task.*

Finally, *the third task is an organization of the name placement process* based on the model obtained at the previous stage. The label representation model for the selected map region (or map annotation layer) is a result of problem having been successfully solved. A precise position of each fetched label is computed at this stage.

Thus, we solve problem at this three levels. Today the most investigated part is the last part. But having neglected any component from this scheme we will shift the centroid of processing toward another component, which causes the decreasing of total efficiency. Let us discuss some techniques to solve formulated tasks.

5. Problem solving techniques

Modeling of label storage should be considered within common context of a whole system. The requirement of handling of a large body of data involves the necessity of using group processing everywhere it possible and justified. Group processing in GIS is based on preliminary geoobject classification, which in turn results in classifier system or metadata (with regard to geoobjects). From the other hand, generalization process also used within group handling utilizes classification. So, classification support is one of a key factor in an effective work of system as a whole and name placement processor in particular.

As names is derived form geoobjects, as they could be linked to geoobject classes. Geoobject class description is within classifier's jurisdiction. It's convenient to correlate an entire name attributes (except for position attribute) and references to the handling algorithms of second and third stages with the geoobject class entity. So, having used classifiers there is a good chance to begin develops rule-based system (or subsystem). An example of such system is discussed in [3]. As for modeling of label position one ought to note the fuzzy nature of such kind of knowledge. Perhaps the methodology of fuzzy systems may help to solve problem at this level. Thus, at the stage of storage modeling we lay the foundation of successful solving of subsequent tasks and provide a full support for modeling entire label attributes except for position, which modeled by fuzzy techniques.

An implementation of name handling process at the fetch stage in many respects is determined by an organization of geoobject fetching, which in turn depends on the system purpose and its “mightiness”. But the link between two process of handling geoobjects and names is become invariant. If geoobject fetching is made in parallel than label handling may also be organized as a parallel process. For large amount of data this results in performance benefit. Names of fetched geoobjects are filtered by the name generalization algorithm, which finally forms a names set and computes name attributes. A generalization algorithm utilizes the rules of generalization and font attribute computing, linked to the geoobject classes. It’s computed a part of name placement rules to determine an initial location of the each label from a set. Finally it builds a range index under that label set. As an index will be used to determine geometrical collision during placement process at the next stage it must support dynamics.

Let us discuss the third stage by introducing a methodology of rule-based systems (see above). As mentioned group processing is based on classification of geoobjects. There are many different principles of classification. We will concern only with geometric (geometry types) and semantic (information coding) principles. As classification groups geoobjects (and thus labels), it needs to order the handling of such groups. On the one hand, labeling order should be determined by three types of geoobject geometries (see above), on the other hand, it’s convenient to link it with classifier system. The first link is concerned with differences in degree of freedom of each geometry type (relative to the name placement). Region geoobjects has the least degree of freedom, but line geoobjects has the largest one. So, regions are labeled firstly, points are labeled at the second hand, and finally lines are labeled. The labeling order within each of these three parts is given by the classification system. As the first order technique sets a global labeling sequence while the second one details it within each stage. The first technique may be hard-coded by GIS software, but the second one supposes flexible tuning.

After each name is placed, a quality evaluation algorithm(s) is used to provide feedback to the name placement processor. If the quality is unacceptable an additional placement are tried (perhaps under another algorithm). During the process the quality of an earlier placement may become unacceptable which causes reposition. Quality overvaluation may affect only labels of certain classes, group of classes or all classes within the current geometry type bounds.

Thus, there are two kinds of rule at the third level, one for name placement and another for quality evaluation of name placement. The second always followed the first for each label. There are many rule sets in a system, for example, there is a pair rule sets for each geoobject class, one to place name and another to evaluate quality. In general, rule may be organized as a control structure, determining algorithms and parameters. A parameters are used to label geoobject (such as default font attributes) or to execute algorithms computed real label visualization parameters (such as label text, character capitalization, font attributes) and positions of each label character (this is depend on implementation).

One ought to note a combinatorial nature of the name placement problem. So, it could apply evolutionary computations to solve the problem. Let us briefly examine the problem in that context. To determine correct label position an algorithm must look over a quantity of different cases, taking into account positions of nearby labels and geoobjects. Each time a label position is computed (i.e. after yielding a sequence of evolutionary operators) an algorithm evaluates placement quality (or yields fitness function), which serves as a criterion for further surviving of a position. As a rule algorithm simultaneously handles a group of labels (rather a group of positions), which is corresponded to one or more geoobject classes. During the yielding of evolutionary operators a population size (amount of positions) may be invariant or variable depending on algorithm. Also there are one or more placement strategies (or rules). It’s typical, a label displacement is bounded by some area (e.g. by geoobject neighbourhood), so coordinates of each label computed by mutation operator must be from the bounded domain of coordinates. Obviously, name placement for point objects is more preferable to solve it by applying evolutionary computations. For an example of such an application see [2].

As for applying the parallel computations to represent high-density maps there are at least two directions. Firstly, this is a parallel execution of name placement process and quality evaluation process, which results in stepwise improvement of the quality of textual information representation within the map annotation layer. Secondly, this is parallel processing based on spatial localization nature of labels, which result in dividing map into several responsibility zones, each of which is handled by separate computational thread. Both techniques suppose close coordination between computational threads.

It’s important to discuss the usefulness of name placement forecasting. Such approach bases on parallel processing and must not influence the total performance. There are two parts of forecasting - constant map scale forecasting (when user walks through the map at constant scale) and variable map scale forecasting. But there is no point to use the placement forecasting if an entire system does not support this process for geoobjects.

6. Conclusion

Increased requirements to the performance and representation quality of textual information at electronic map caused to use the system approach to solve the problem. According to it we defined three levels of problem solving and discussed some solving techniques. Future work will be based on the presented scheme.

Acknowledgements

The comments from our colleagues of National Institute for Fiber Optic and Data Processing Systems have greatly improved the readability. We also thank they for their help with other aspects.

References

1. Ahn J., Freeman H. "A program for automatic name placement". *Cartographica*, 21(2-3), 1984.
2. Djouadi Y. "CARTAGE: A Cartographic Layout System Based on Genetic Algorithms". In: *Proceedings of the Fifth European Conference and Exhibition on Geographical Information Systems (EGIS/MARI’94)*, (http://wwwsgi.ursus.maine.edu/gisweb/spatdb/egis/eg94006.htm), 1994.
3. Doerschler J.S., Freeman H. "A rule-based system for dense-map name placement". *Communications of the ACM*, V 35, N 1, 1992.
4. Imhof E. "Positioning names on maps". *The American Cartographer*, 2(2), 1975.
5. Öâåòêîâ Â.ß. "Ãåîèíôîðìàöèîííûå ñèñòåìû è òåõíîëîãèè". Ì.: Ôèíàíñû è ñòàòèñòèêà, - 1998.