DESIGN ISSUES FOR AN INTELLIGENT NAMES PROCESSING SYSTEM

Stephen A. Hirsch and Barry J. Glick
PAR Technology Corporation
P.O. Box 2950
Landover Hills, Maryland 20784

ABSTRACT

An overall systems design approach for a names processing system includes design issues concerning a system configuration, names and spatial databases, and names processing operations (selection, positioning, and plotting). A names processing system requires interactive graphics display, data storage, data manipulation, computer processing, and plotting capabilities. Names and spatial databases require an integrated design with respect to the requirements and goals of names processing. Selection consists of the interactive/automatic retrieval of names and spatial data. Positioning consists of a control structure that integrates automatic multi-method positioning algorithms and a human operator (i.e. map editor) in an interactive processing environment. In the names plotting operation softcopy and hardcopy products may be generated on standard display and plotting devices.

INTRODUCTION

Historically, experimental systems have addressed only parts of automatic names processing, because only parts of the process have been approachable given available technologies. These early systems[1] focused on the names processing operations of selection and plotting, and omitted the difficult positioning operation.

More recently, government agencies (e.g. DMA and USGS) have begun to support the large scale development of spatial databases, digital gazetteers, and the hardware and software necessary for automatic names processing (Schmidt et al., 1982; Caldwell, 1982). In addition, technological strides in computer capabilities and software models have supported attempts that have focused on automating the positioning operation (Yoeli, 1972; Hirsch, 1980 and 1982). Currently, all the components of a names processing system are in hand, or within reach; however, what is still lacking is an overall systems design approach. This paper presents such an approach including design issues concerning: 1) a system configuration; 2) names and spatial databases; and 3) names processing operations (selection, positioning, and plotting).

HYPOTHETICAL SYSTEM CONFIGURATION

A typical names processing system consists of a minicomputer processor, disk drives, a tape drive, an interactive graphics work station, plotting capabilities, and a printer terminal (Figure 1). The processor must be able to manage characteristically large volumes of names/ spatial data and relatively complex names positioning software. The disk drives should have adequate capacity for storing the large names/spatial databases, font libraries, and operating system, utility, and names processing software. The tape drive is utilized for general communication, archiving, generating plot tapes for a plotting device, and so forth. The graphics work station consists of an interactive graphics display, keyboard, menu, digitizing table, and cursor. Finally, the system should be able to generate hardcopy products on several types of plotting devices and be able to produce lettering in several fonts.

In a typical session, the operator (map editor) selects names and spatial data from databanks. The system then automatically positions names relative to their respective spatial features, displays results, and prompts the operator for review and/or modification. The operator may then direct the system to generate a hardcopy plot of the spatial data and positioned names.

SPATIAL/NAMES DATABASE DESIGN

Two databases fundamental to a names processing system are a names database (NDB) and a spatial database(SDB). A NDB contains mostly descriptive or attribute data in reference to spatial features. A SDB, for the most part, contains spatial feature coordinate descriptions with some attribute data.

A names database or gazetteer (Figure 2) typically contains the feature name, spatial location, feature type (e.g. populated place), and feature attributes (e.g. population). Spatial location is represented in the database for a point feature simply by a spatial coordinate, for an areal feature perhaps by some arbitrary point inside the area, and for a linear feature by, for example, the mouth or source coordinate of a river.

Existing large scale names databases (e.g. the Geographic Names Information System at the U.S. Geological Survey and the Foreign Names Place File at the Defense Mapping Agency) are limited in scope in terms of automated names processing. Much of the information in these databases is duplicated in the associated spatial database, and it is often not possible to directly link feature names in the names database with their spatial features in the spatial database.

Spatial databases and structures have received considerable attention in the literature, (e.g. Peucker and Chrisman, 1975; Shapiro and Haralick, 1980). However, in the context of names processing, the integrated design of spatial and names databases has received relatively little attention.

An integrated design would allow for the identification of required information unique to both names and spatial databases and information that could be shared. By constructing direct linkages between the NDB and SDB, an integrated design would allow for the sharing of information, the construction of relations, and the avoidance of excessively redundant information. Linkages between the NDB and SDB may be implemented by several direct or indirect addressing schemes, for example, pointers, feature identification codes, and feature names.

An integrated design would provide an opportunity to consider tradeoffs between storing or processing information, and to consider strategies for capturing and building information. For example, one would want to store rather than repeatedly compute the area of a feature, and one would prefer to capture primitive geometric entities and from these build more complex relationships.

On the whole, an integrated design of names and spatial databases would optimize data capture, building, storage, maintenance, and processing operations. With an integrated design, the combination of information contained within a NDB and a SDB would represent an efficient store of information that as yet has not been available to the cartographic community.

NAMES PROCESSING OPERATIONS

Building on the description of a names processing system and the relevant databases, we now address names processing operations: 1) selection - the selective retrieval of names and spatial data; 2) positioning - the positioning of names amidst spatial data; and 3) plotting - the generation of softcopy and hardcopy names and/or spatial data products.

Selection

The selection operation consists of two sub-operations: spatial data selection and names selection. The map editor interactively specifies selection parameters (e.g. area, scale, resolution and theme) and then names and spatial data are automatically retrieved from the database. In spatial data selection the area will indicate the coverage, and the theme, scale, and resolution will determine the kind and detail of spatial information that will be retrieved (Figure 3). For example, the theme might effect the retrieval of a particular overlay, or certain spatial data within an overlay, and the scale and resolution will dictate the detail of spatial information and generalization. The retrieved spatial data will be stored in an application file where it can be efficiently accessed by the positioning and plotting operations.

The same set of spatial data selection parameters is utilized for names selection (Figure 4). Similarly, based on these parameters, a subset of information will be retrieved from the NDB and stored in an application file for the subsequent positioning and plotting operations. As the NDB application file is built, names will be classified according to theme, scale, feature type, and other criteria. For example, depending on the scale, a feature will be classified as a point or area feature type. In another example, if the theme is population of U.S. cities, each city will be classified according to some classification scheme (e.g. percentiles).

Other name parameters requiring specification in the application NDB are lettering characteristics such as font (style), size, and spacing. These parameters will be set automatically or interactively for each class of names as selected from the NDB. For example, all cities within a population range will be in one class and can be assigned a unique characteristic. Lettering characteristics, especially size and spacing, are required information for the positioning operation. Size, spacing, and font characteristics are required for the plotting operation.

During the selection operation the map editor may perform editing operations. He may examine and modify the names and spatial data, the classification of names, lettering characteristics, and selection parameters. The map editor will also be prompted for selected names and spatial data discrepancies. Discrepancy checking will be supported by the linkage structure which integrates the NDB and SDB. The linkage structure will assist in verifying the selection of names and associated features, and assist in monitoring feature type change, for example area feature change to point feature as a consequence of scale.

Positioning

Positioning consists of a control structure that integrates automatic multi-method positioning algorithms, spatial processing techniques, and map analysis and interactive capabilities (Figure 5). Methods for point, linear, and areal features will automatically position selected names amidst selected spatial data. The control structure, upon analyzing initial and iterative map conditions, utilizes decision criteria to determine the appropriate method or sequence of methods, or to transfer to interactive control where the final results are displayed on the graphics terminal. In this interactive environment the map editor is prompted for review and/or modification of the selected positions.

Point Feature Positioning. Figure 6 shows a scheme developed by Hirsch (1980, 1982) to position names around points. Overlaps between point feature names and other names or spatial data generate vectors indicating direction of movement toward open space. This information is translated into a particular movement strategy depending on the method of movement.

Area Feature Positioning. Figure 7 illustrates a method for positioning a name within an areal feature. First a minimum area encasing rectangle (Freeman and Shapiro, 1975) is derived. Then n line segments (n = to the number of name characters) are constructed equally spaced and orthogonal to the rectangle orientation. Mid-points between line segment and polygon boundary intersections mark the positions for each character.

The orientation of each character may be orthogonal to the horizontal x axis, the rectangle orientation, or to a trend line fitted to the mid-point positions. This method produces a name placement solution that is representative of the area's shape and extent.

Linear Features Positioning. Figure 8 displays a scheme for positioning a linear feature name along a river. The name may slide from left to right, above or below the river depending on the information content of an overlap vector (i.e. an indication of symbol density in the local area).

Interactive Adjustment. When decision criteria indicate that position processing is completed, the control structure directs control to the operator. At this time the spatial and names data may be displayed on the graphics terminal for review and/or modification. The results may be saved on disk or directed to a plotting device.

Plotting

The plotting operation consists of the capability to generate softcopy and hardcopy products on a host of standard display and plotting devices. Pens, scribing tools, photoheads, CRT photoheads, electron beam recorders, laser, and electrostatic matrix plotters are examples of standard plotting devices. Speed and accuracy tradeoffs between these different devices must be considered. Softcopy products on display devices are desired for quick review and analysis or to be copied by screen scanning copiers.

Input required by the plotting operation is accessed from the application names and spatial data files. This input includes a full specification of names to be plotted (text, starting location, orientation, and lettering characteristics). A comprehensive selection of lettering styles or fonts should be available from a font library. Examples of products are final names overlays, spatial data overlays and combinations.

CONCLUSION

All of the major components of a names processing system (the proposed system configuration, names and spatial databases, and names processing operations) are now approachable given available technologies. An overall systems design approach is necessary for the design and development of an intelligent system. Names and spatial databases require an integrated design in order to optimize data capture, building, storage, maintenance, and processing operations. The automation of names processing operations requires the support of a map editor in an interactive processing environment. Finally, further research is required for the positioning operation, in particular for the development of a decision-based control structure.

FOOTNOTE

[1] The Oxford System (Bickmore and Boyle, 1964); The Automatic Type Placement System (ATPS) developed by the U.S. Army Engineer Topographic Laboratories for the Defense Mapping Agency; and cartographic name processing system efforts at the University of Saskatchewan, Canada (Boyle, 1973 and 1974; Wilkie, 1973).

REFERENCES

Bickmore, D., and Boyle, A. R. (1964), "The Oxford System of Automatic Cartography, "Proceedings of Technical Symposium, International Cartographic Association, Edinburgh, England.

Boyle, A. R. (1973), "Computer Aided Map Compilation," Electrical Engineering Department, University of Saskatchewan, Canada.

Boyle, A. R. (1974), "Report on Symbol and Name Manipulation and Placement," Electrical Engineering Department, University of Saskatchewan, Canada.

Caldwell, D. R. (1982), "A Names Input Station for Automated Foreign Names Production at the United States Defense Mapping Agency," Proceedings, AUTO CARTO V.

Freeman, H. and Shapiro, R. (1975) "Determining the Minimum-Area Encasing Rectangle for an Arbitrary Closed Curve," Communications of the ACM, Vol. 18, pp. 409-413.

Hirsch, S. A. (1980), "An Algorithm for the Automatic Placement of Point Data," Master's thesis project, Department of Geography, State University of New York at Buffalo, New York.

Hirsch, S. A. (1982), "An Algorithm for Automatic Name Placement Around Point Data," The American Cartographer, Vol. 9, pp. 5-19.

Peucker, T. K. and Chrisman, N. (1975), "Cartographic Data Structures," The American Cartographer, Vol. 2, pp. 55-69.

Shapiro, L. G. and Haralick, R. M. (1980), "A Spatial Data Structure," Geoprocessing, Vol. 1, pp. 313-337.

Schmidt, W. E. et al. (1982), "Automated Names Processing Research," Proceedings, AUTO-CARTO V.

Wilkie, W. T. (1973), "Computerized Cartographic Name Processing," Master's thesis, Department of Electrical Engineering, University of Saskatchewan, Canada.

Yoeli, P. (1972), "The Logic of Automated Map Lettering," The Cartographic Journal, Vol. 9, pp. 99-108.

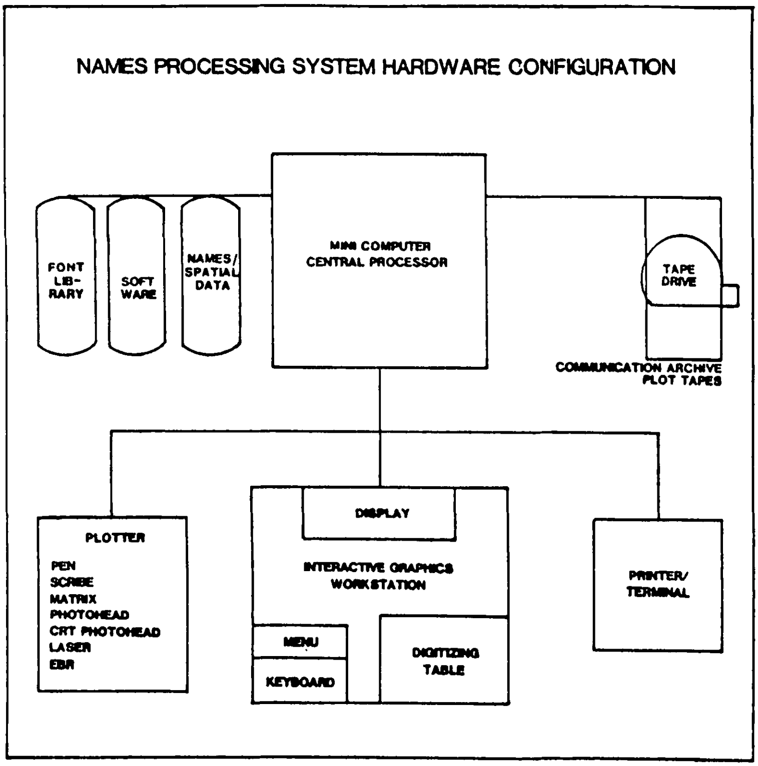

Figure 1. Components of a typical names processing system.

NAMES DATABASE

NAME	LOC.	FEATURE TYPE	SUBTYPE	ATTRIBUTE
HUDSON RIVER	X,Y	LINEAR	RIVER/BROOK	LENGTH/LOAD
BOSTON	X,Y	AREA	STATE CITY/CAPITAL	AREA/POPULATION

Figure 2. Standard information contained in a names database.

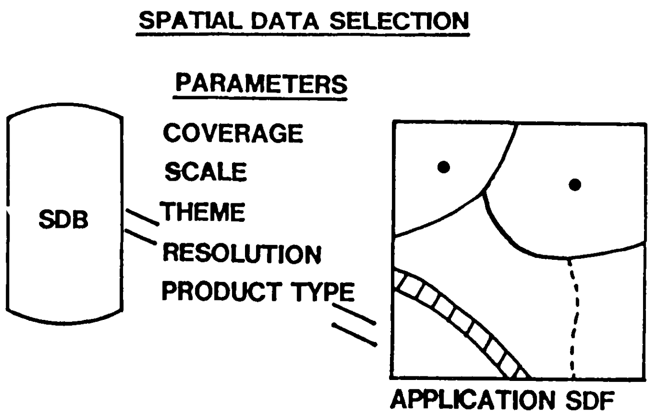

Figure 3. Based on selection parameters, spatial data are retrieved from a spatial database and stored in an application spatial data file (SDF).

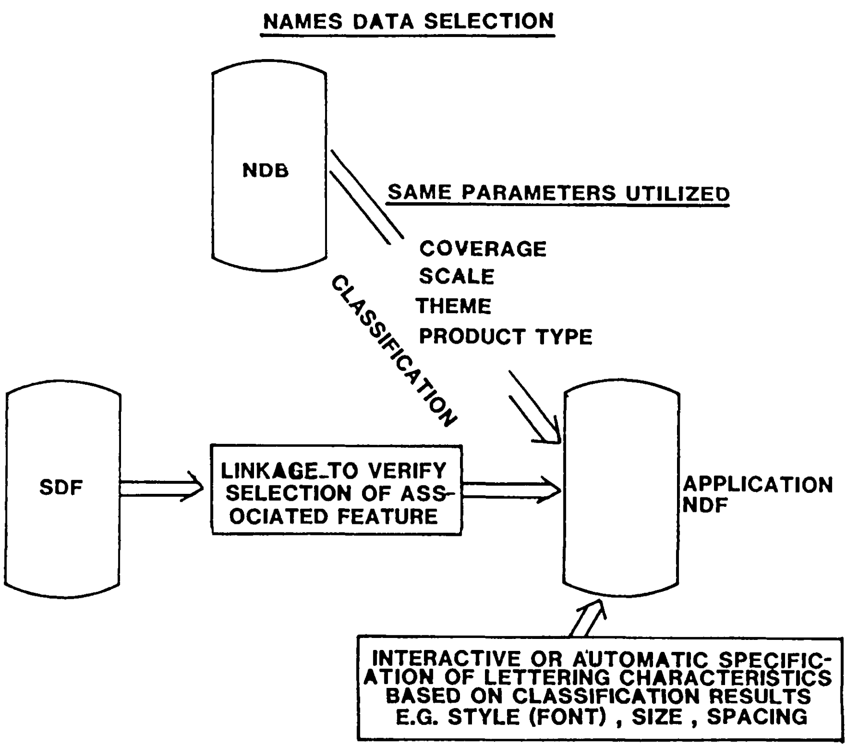

Figure 4. Based on the same selection parameters, names data are retrieved from a names database (NDB).

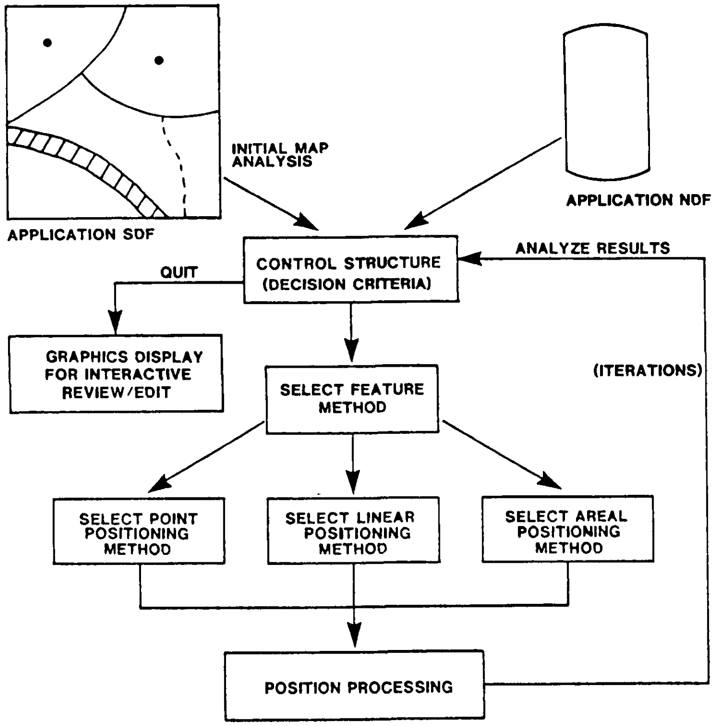

Figure 5. In the names positioning operation, a control structure integrates automatic multi-method positioning algorithms, spatial processing techniques, and map analysis and interactive capabilities.

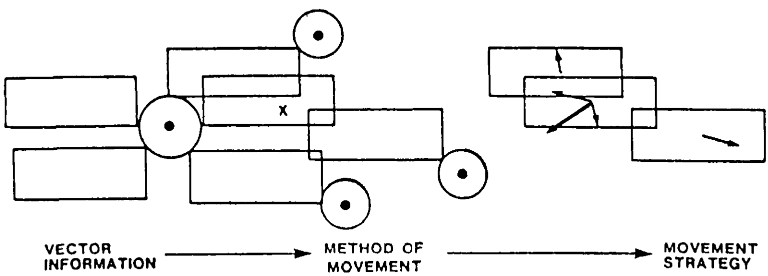

Figure 6. An example of point feature positioning utilizing overlap vectors.

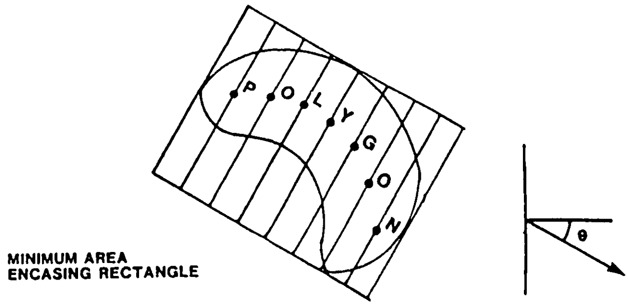

Figure 7. An example of area feature positioning.

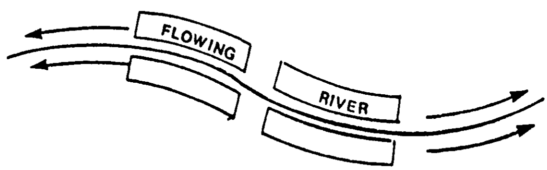

Figure 8. An example of linear feature positioning.