DEVELOPMENT, AUTOMATION, AND DOCUMENTATION OF GEOINFORMATICS PROCESSES USING A KNOWLEDGE-BASED ASSISTANT

Kurt T. Rudahl
and
Sally E. Goldin
Department of Computer Engineering
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
email: kurt (at) cpe.kmutt.ac.th

KEY WORDS:

Abstract: Effective use of geoinformatics software requires substantial training and expertise. This has proved to be an obstacle to realizing the full benefits from geospatial data and geoinformatics software in government and industry. Professional users of geoinformatics packages, both novice and experienced, find it difficult to understand how the capabilities of the software tools can be used to obtain the geospatial products which they need. This derives from the difficulty of matching their geographical or other domain-specific knowledge to the characteristics of the software packages they are using. We present a prototype expert system which uses knowledge of both geoinformatics analysis objectives and specific geoinformatics software packages to help the user to define a model; that is, a sequence of operations which will yield the desired goals. Having defined the model, the user can then refine it or execute it repeatedly over different geographic regions. Finally, the user can generate a document describing the sequence of operations and the rationale behind it.

1. BACKGROUND AND OBJECTIVES

Geospatial information processing has become a pivotal tool for decision making and planning in municipalities, government and non-profit agencies, and private business. In addition, it is generally agreed that geospatial processing systems have significant potential for improving business performance through increases in efficiency and more accurate market analysis (Walker, 1990). Geoinformatics is also a critical tool in current programs to comprehend and control global and local environmental change (Moore, 1990; Pradhan, 1991; Stetina, 1991). Remote sensing and GIS are currently used for analyzing biodiversity (Glasgow, 1992; Kempka, 1992), monitoring water quality (Joao, 1992; Thewessen, 1992), epidemiological tracking (Chan, 2011), managing forest resources (Jaakkola, 1990; Singh, 1990), predicting natural disasters (Bitters, 1991), and studying climate and ecological change (Corbett, 1992; Erdenetuya, 2011), to mention only a few applications.

Although geospatial data processing has been productively applied in a broad range of domains, obstacles to the widespread use of geoinformatics technology remain. One of the most serious is the inherent complexity of the analysis techniques and systems (Goldin, 1997). Almost anyone can understand and appreciate the results of a geospatial analysis presented as a map or other graphic; this is one of the advantages of the technology. However, highly-trained technicians or analysts are usually required to produce those results. Analysts typically need some knowledge of geography, cartography, mathematics, computer science, and graphic design, as well as specific expertise in using their particular GIS software package.

One major difficulty for professional users of geoinformatics packages, both novice and experienced, lies in understanding how the capabilities of the software tools can be used to obtain the geospatial products which they need. This derives from the difficulty of matching their geographical or other domain-specific knowledge to the characteristics of the software packages they are using.

Our objective is to provide a software system which uses knowledge of geospatial outcomes, such as (but not limited to) map products, combined with knowledge about the capabilities of some *target* geoinformatics software package (or packages), to assist these users.

Broadly speaking, we would like to create a system in which a non-expert in geoinformatics can describe her objectives in geographic rather than geoinformatics software terms. For example, she might say that she needs a location for a new hazardous waste center with certain specified geographic constraints. Then the system should produce that result, or at least explain in detail how to achieve that result, in the context of the target software.

We consider that this objective is not realistic at present. Therefore we have adopted more conservative objectives to produce a system with the following behavior:

- The system allows the user to specify, from a hierarchical list of choices, the geographic products (which we call *results*) needed.
- For each needed result, the system offers a list of candidate operations, or *actions*, available in the target software, which can produce that product. The user chooses an action.
- For the chosen action, the system knows what type of *intermediate* or *origin* data sets or other inputs are required. If there is a choice, the user makes that choice. The system also knows about input parameters for that action.
- For each intermediate data set, the above two steps can be iterated.
- For each action or data product specified, the user can add an arbitrarily detailed commentary, in free form text, of why it is needed or chosen.
- The user can save, visualize, and/or revise the entire set of processes (called a *model*) derived in the above manner,
- The user can execute the model, specifying origin data sets from different places or times at each execution.
- The system can generate a document derived from the above commentaries plus built-in commentaries based on the target software. This document describes the entire process including the user's rationale for each step.

2. EXPERT SYSTEMS

An *expert system* is a software system which can solve problems that normally would be considered to require human expertise. For instance, expert systems have been developed to do medical diagnosis, computer configuration management, discovery of molecular structures, and interpretation of oil exploration data (Winston, 1984).

Typically, expert systems consist of a knowledge base of *rules* relating to the problem domain, plus some sort of *deduction engine* for applying rules and drawing conclusions (Buchanan,1984). Researchers in GIS and the related fields of remote-sensing image interpretation and environmental modeling have been exploring the utility of expert systems concepts for quite some time (Walker, 1990).

3. APPROACH AND RESULTS

Our software includes detailed knowledge about the operations available on the target geoinformatics software. In principle, this could be any geoinformatics package (or even a collection of packages from different manufacturers). The ability to execute the model, however, requires that the target geoinformatics software package has a scripting or command-line interface.

Creating this detailed knowledge base is a non-trivial exercise which depends on the quality of the documentation provided by the manufacturer. In the development of our prototype, we have worked with the free OpenDragon remote sensing package (Rudahl, 2010).

Although nothing in our list of objectives mandates a graphical user interface, we consider that ease of use and visualization make an interactive design tool desirable. Figure 1 shows an example model. As described below, the model is created in reverse order, beginning with the desired outcomes at the bottom and moving upwards. In this case, the user specified two desired outcomes: a *suitability map* of locations where a new park might be located, and a *region report* describing, for example, the size and exact location of each parcel in the suitability map.

At this point, the system suggested an operation called **ITEMIZE** as being useful for creating both of the desired products. ITEMIZE separates a classified image or other raster GIS coverage into separately-identified polygons. Its only input is a layer file*

* **ITEMIZE** is currently only available as an experimental operation in OpenDragon.

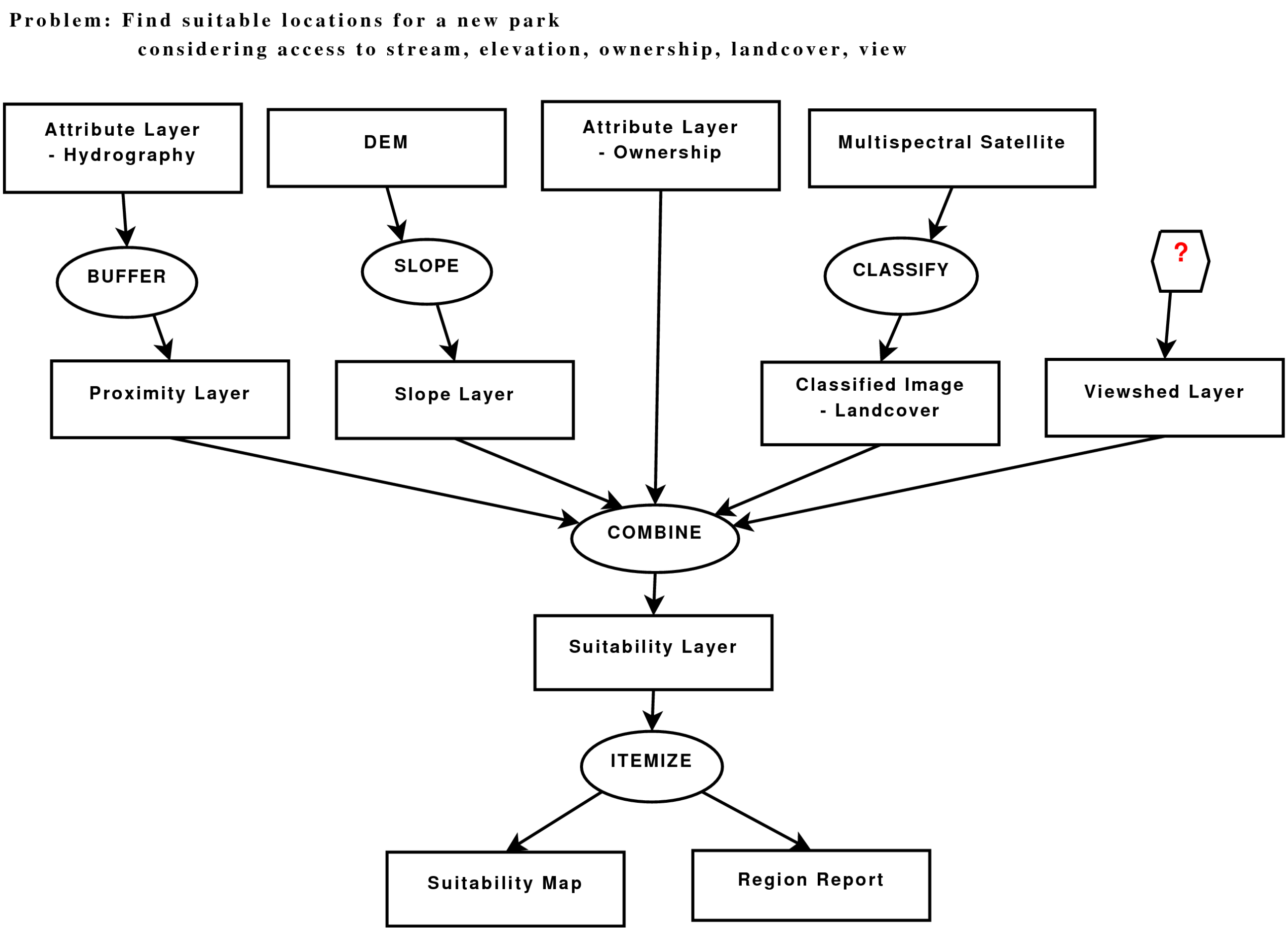

Figure 1: Example model to locate new park

The system recognized that the intermediate suitability layer could be provided by a number of different operations, and offered the choice to the user. In this case, the user chose the **COMBINE** operation. COMBINE provides the ability to combine up to twelve data products based on logical rules such as shown in Figure 2. The inputs are all of the different layers shown in Figure 1, plus the combination rules themselves. The form in Figure 3 is used to define the combination rules.

The user continued to iterate in this way, alternately adding intermediate data sets and actions to create those sets, until she reached data sets which she already had available, or knew how to obtain. However, an interesting example appears on the far right of Figure 1. She found a need for a viewshed layer, but OpenDragon does not currently provide a viewshed operation. Although this is a setback in her current analysis, it actually provides her with a good opportunity. It will be easy for her to document, using our prototype, the need for funds to acquire this data layer or the software to produce it.

```
for each pixel
   result = 2
   if band0 between 32 and 255
      and band 5 between 64 and 255
      and ...
```

Figure 2: Sample combination rule for COMBINE operation.

Up to this point, no specific geographic data sets have been provided. Once the user is satisfied with the flow of work represented by the model, and has saved it to disk, she can choose to *execute* it. At this time, the system will ask for specific *bindings* (file names, etc) for each of the origin data sets and for the result sets.

The user can also choose to create a document, describing the data, operations, and justifications. The document can either include or exclude the model bindings.

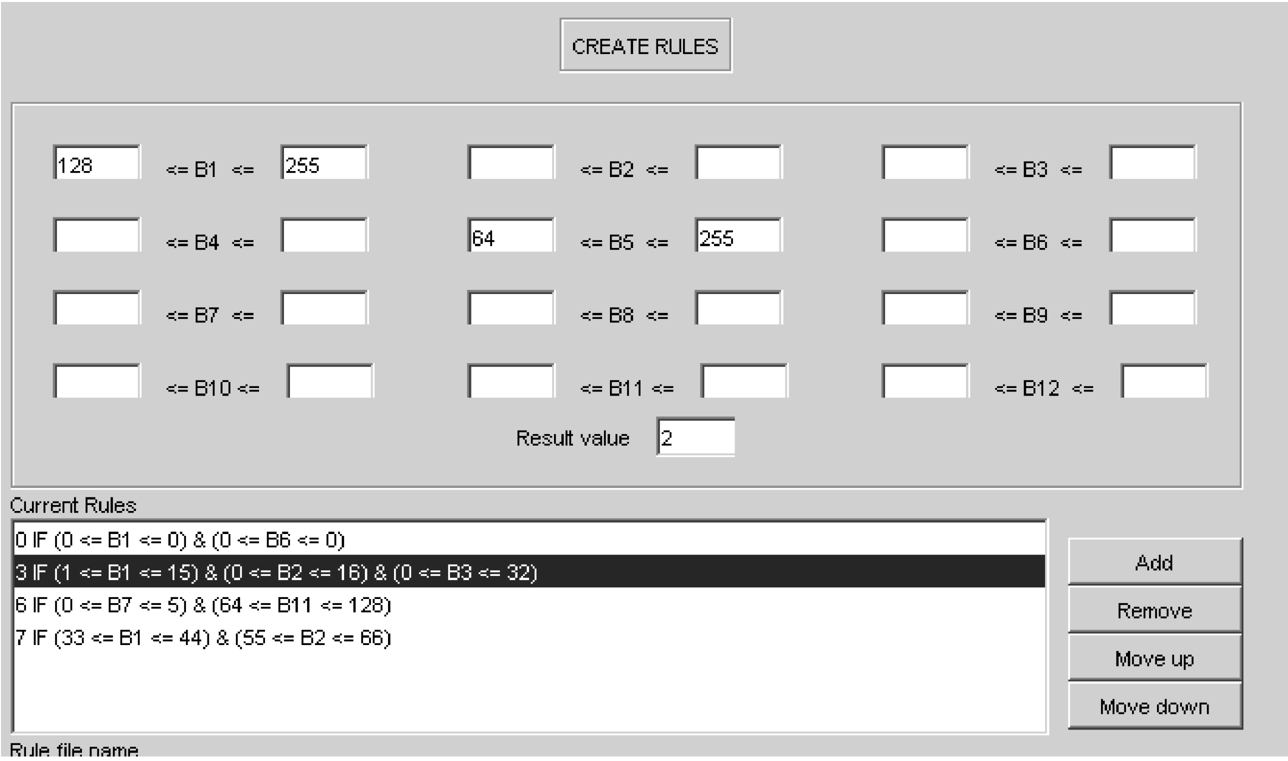

Figure 3: Panel for adding combination rules

4. BENEFITS

Since we have argued that users of geospatial data are already confused by the software they have, it is reasonable to ask whether one more software package will not simply further confuse them.

The first answer is that our software is intended to make the user's life easier by providing easily accessible information and advice. Users of geospatial data report (Goldin, 1997) that their interest and knowledge is focused on the *data*, while the software manufacturers, perhaps quite naturally, tend to focus more on the *processes*. Our software is an attempt to partly overcome this disparity in focus.

We further believe that the capabilities of our software to provide a unified view, even of operations which might span several geoinformatics packages, will help to clarify the processes for the user.

A further result from (Goldin,1997) is that users feel a great need to create notes or commentary about what they are doing, both as future reminder for themselves and as preparation for formal documentation. Thus, we view the documentation capabilities of our software as having central importance.

Somewhat less obvious benefits which may come from this software include:

- Justify acquisitions. Geoinformatics software and geoinformatics data can represent substantial financial expenditures. Our software provides an easy, and politically neutral, way to marshal and present evidence about the need for new purchases

- Educate management. In any industry, management is rarely as knowledgeable about the underlying technology as are the professional experts. That is not a criticism, and it is probably inevitable. Being able to see a succinct summary of each stage of the analysis process could be more informative for a manager than simply being told that you put ***X*** information in and get ***Y*** answers out.

5. FUTURE PLANS

5.1 Extending the Rules Sets

The prototype expert system incorporates three types of knowledge:

- Knowledge about the types of data sets, or data products, relevant to geoinformatics.

- Knowledge about the operations, or actions, which are possible methods to produce such data sets.

- Knowledge about the specific syntax and parameters needed for each action.

The knowledge about data products should be generic to all types of geoinformatics processing. Because our prototype is focused on a remote sensing package, the current set of knowledge in this area is weighted toward that specialty. It should be expanded. We believe it would also be desirable for the user to be able to add new product types.

The other two types of knowledge are specific to a particular geoinformatics software package. We need to define an external format for specifying this knowledge. This format must be general enough to adapt to any geoinformatics package, so that someone with detailed knowledge of a particular package could easily create the rules needed for that package. Data in this format would then be loaded to initialize the expert system.

5.2 Extending the Software

Expert system software works by making inferences based on rules. To the extent that more rules become available, the software becomes more useful.

Other ways to extend the software involve improving the inferencing engine or adding new rule types. We are particularly interested in exploring the possibility of having the system improve the sequence of choices it offers, and the default values of parameters, by learning the work habits of its users. This would make the system more convenient for the individual user. Other system enhancements could focus on

- improving consistency among multiple users at a particular organization;

- facilitating technology transfer to new employees;

- reducing time and costs spent on employee education.

6. CONCLUSION

We believe that our software points the way toward making geoinformatics software easier to use and, even more importantly, easier to use correctly. The gains from this process are not merely greater efficiency and productivity for individual employees, but also for organizations, allowing them to make better use of the wealth of geoinformatics data that is becoming available. At the same time, our software should also facilitate use of geoinformatics processing in organizations which do not currently use it, or which use it to only a slight degree.

Our purpose in creating the software was to support the more efficient utilization of (possibly expensive) geoinformatics resources: the data, the geoinformatics systems, and the trained analysts. Along the way, we have come to realize that it has a possible additional benefit. By displaying the logical sequence of operations in a geospatial analysis, the system may help non-users of geoinformatics, whether an organization's management or the general public, to understand the nature of geoinformatics processing without needing to become immersed in the technical details.

REFERENCES

Bitters, J.D., Restrepo, P.J., and Jourdan, M.P., 1991. Using GIS to predict effects of flooding on the Han River, South Korea. Proceedings of the 1991 ASPRS-ACSM Convention. Bethesda, MD: American Society for Photogrammetry and Remote Sensing, Vol. 4, pp. 11-20

Buchanan, B.G. and Shortliffe, E.H., 1984. Rule-based Expert Systems. Reading, MA: Addison-Wesley Publishing

Chan, T.C., King, C.C., Rudahl, K.T. and Goldin, S.E., 2011. Crowd-Sourced Acquisition of Geographic Information for Influenza Epidemic Surveillance. Proceedings of the 32th Asian Conference on Remote Sensing. Asian Association for Remote Sensing, 2011.

Corbett, J.D., 1992. An evaluation of global climate GIS data bases for identification of crop environments: a case study for northern Latin America. Proceedings of the 1992 ASPRS-ACSM Convention. Bethesda, MD: American Society for Photogrammetry and Remote Sensing, Vol. 1, pp. 51-61.

Erdenetuya, M., Bulgan, D. Erdenetsetseg, B., 2011. Drought Monitoring and Assessment Using Multi Satellite Data in Mongolia. Proceedings of the 32th Asian Conference on Remote Sensing. Asian Association for Remote Sensing

Glasgow, K.C., Franklin, J. and Wright, R., 1992. Use of GIS data analysis in mapping biological diversity for preservation by local open space policies. Proceedings of the 1992 ASPRS-ACSM Convention. Bethesda, MD: American Society for Photogrammetry and Remote Sensing, Vol. 1, pp. 109-110.

Goldin, S.E. and Rudahl,K.T.,1997. Why is GIS Difficult? Proceedings of the 18th Asian Conference on Remote Sensing, Asian Association for Remote Sensing.

Jaakkola, S., 1990. Managing data for the monitoring of tropical forest cover: the Global Resource Information Database approach. Photogrammetric Engineering and Remote Sensing, Vol. 56, No. 10, October 1990, pp. 1355-1357.

Joao, E.M. and Walsh, S.J., 1992. GIS implications for hydrologic modeling: simulation of non-point pollution generated as a consequence of watershed development scenarios. Computers, Environment and Urban Systems, Vol. 16, No.l, 1992, pp. 43-63.

Kempka, R., Kollasch, P., and Koeln, G.T., 1992. Ducks Unlimited - Using GIS to preserve the Pacific Flyway's wetland reserve. GIS World, Vol. 5, No. 1, February 1992, pp. 46-52.

Moore, D.G., Howard, S.M. and Tappan, G.G., 1990. Geographic Information System as country level development and monitoring tool, Senegal example. Proceedings of the 23rd International Symposium on Remote Sensing of Environment, Vol. 2, pp. 683-694. Environmental Research Institute of Michigan.

Pradhan, V.C., Sheinkman, M., Otoma, S., and Johnson, G.E., 1991. South East Asian one kilometer mosaic from NOAA AVHRR data. Proceedings of the 12th Asian Conference on Remote Sensing. Asian Association for Remote Sensing, Vol. 1, pp. E-2-1 to E-2-6.

Rudahl, K.T. and Goldin, S.E., 2010. Learning Remote Sensing Using Dragon/ips and OpenDragon, Global Software Institute.

Singh, K.D., 1990. Design of a global tropical forest resource assessment. Photogrammetric Engineering and Remote Sensing, Vol. 56, No. 10, October 1990, pp. 1353-1354.

Stetina, F. and Vermillion, C., 1991. Development of a uniform global environmental data gathering and distribution system to support global change studies. Proceedings of the 12th Asian Conference on Remote Sensing. Asian Association for Remote Sensing, Vol. 1, pp. S-3-1 to S-3-5.

Thewessen, T., Van de Velde, R., and Verlouw, H., 1992. European groundwater threats analyzed with GIS. GIS Europe, Vol. 1, No.3, April 1992, pp. 28-33.

Walker, T.C. and Miller, R.K., 1990. Geographic Information Systems: An Assessment of Technology, Applications and Products. Madison, GA: SEAI Technical Publications.

Winston, P.H., 1984. Artificial Intelligence. Reading, MA: Addison- Wesley Publishing.