Dodging the Analysis Paralysis Pitfalls in CyberGIS

Daniel W. Goldberg
Department of Geography, Texas A&M University
College Station, Texas
daniel.goldberg@tamu.edu

ABSTRACT
Geographic data and processing capabilities are now available to researchers, practitioners, and consumers in an on-demand basis like never before. The web-enabled Geographic Information Systems (GIS), data, processes, and services which compose modern Cyber-Enabled GIS (CyberGIS) infrastructures now permit distributed computing and commodity geoprocessing at the push of a button. Designers and providers of these services struggle with two opposing goals - exposing the complex capabilities and parameters of geographic processes that are the powerful engine of GIS, while at the same time providing the casual, non-expert user with the ability to intuitively make use of the same system. This tradeoff presents an interesting and important design challenge that can only be overcome by the convergence of HCI and GIS research, practice, and education. This paper details one such CyberGIS example system that typifies what can go wrong when this tradeoff tips too far in one direction. This paper's goal is to spark discussion that leads to a research agenda for blending these two worlds.

Author Keywords
CyberGIS; Analysis paralysis; Over-specification.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Design.

INTRODUCTION – POWER VERSUS SIMPLICITY

Like any well-designed system, the draw of modern geoprocessing systems lies partially in the fact that they abstract away many of the low-level details that users do not need to care about in order to accomplish a specific task at hand. True, many would argue that it would be useful for a user to understand the details of the algorithm underlying a particular geographic operation in order to know how, when, and under what circumstances one particular operation should be applied instead of another.

However, the ubiquity of web-based GIS data sources and processing capabilities and services – known collectively in many circles as the Cyber-Enable Geographic Information Systems (CyberGIS) infrastructure – has made GIS processing, data, and services a commodity for the majority of the internet-connected world. CyberGIS is a part of most people's daily lives, whether they know it or not.

CyberGIS is called upon every time a web search is performed to locate a retail outlet, identify locations around oneself in an area, or detect a fraudulent credit card transaction by analyzing purchases in both space and time. Mobile technologies have thrust CyberGIS into the forefront of the modern software ecosystem; just consider how many times you have been asked "this website wants to use your computer's location – allow/deny?" when you launch an HTML5 website that includes a map. Or, inspect the security settings of the applications installed on your phone for which have permission to access your location. These settings/permissions often help to improve the user experience in these instances by starting a search from a user's location, rather than requiring its manual entry.

In a typical usage scenario, the web-surfing public does not even know that CyberGIS is involved in their applications. But, when using a geo-enabled or mapping application directly, CyberGIS tools are expected to be one-click, dead-simple, easy-to-use operations despite the fact that CyberGIS and the data, processes, and services which power them are complex animals. GIS data are massive, GIS operations are rife with computational complexity, and GIS services are more often than not composed of numerous sub-processes, each with fine grained parameters to tune delicate models susceptible to minute assumptions.

Purists – developers of these CyberGIS tools especially – are often loath to hide these complexities due to the fact that they have worked with these processes for so long in development, analyzed the minor (or major) impacts of varying one parameter ever so slightly, and agonized over getting that last little bit of performance enhancement in specific scenarios by exposing a knob to allow the expert user to account for and cleverly handle the full spectrum of scenarios that one could encounter.

The tradeoff that CyberGIS system designers face is clear: balance system flexibility – allowing the user to change parameters in order to tune the processing to their specific scenario; with usability – not overly burdening the user with decisions and information and/or choices they do not care about or hope to understand. Although not all unique to CyberGIS, this issue is particularly relevant and timely in this context due to the recent explosion of CyberGIS data and services made available to support the meteoric rise of mobile location based services (LBS) as the computing platform of choice in the majority of peoples' pockets.

GeoHCI Workshop at CHI 2013, April 27–28, 2013, Paris, France.

GUILTY AS CHARGED

Case in point: The geoprocessing services offered by the Texas A&M University (TAMU) GeoServices website [1]. This website (created and maintained by the author of the present work) offers a series of web-based geoprocessing services such as geocoding, address parsing, and point-in-polygon intersection which can be run in batch mode over a user-uploaded database or through a set of application programmer interfaces (APIs). Each of the services offered on the TAMU GeoServices website is the result of in-depth research into one or more particular areas of geographic data processing, querying, representation, accuracy, precision, etc. representing a considerable investment of time and energy.

This author spent immeasurable hours tuning the underlying algorithms to the utmost extent possible, parameterizing models and optimizing queries to account for the smallest, most intricate details in order to squeeze out the last, tiniest percentage increase in performance. As a result, the geocoding service for example – a service whose only job is to turn text into geographic coordinates; an operation that has been in use since the 1970's – has no less than 20 input parameters and 127 output parameters. The other services available on the site are similarly ridiculous in terms of fine grained processing control available.

While Google Analytics tracking of the website reveals that the various API documentation pages and technical detail pages describing the TAMU GeoServices do continuously receive page views, this author is continuously inundated with technical support queries from current and potential users who are interested, yet completely baffled by the level of detailed technical tuning that can be performed.

"What does the 'souatts' parameter do and why would I ever want to or not want to use it"? - (it controls the fields which are allowed to match on SOUNDEX equivalents rather than just full-text matches);

"Why are there four different output fields that describe output quality"? - (well if you look close there are actually nine, AND one cannot really tell the quality unless you combine them all).

The developer of the TAMU GeoServices website created the majority of the services offered therein during the completion of his PhD. Ironically, it has been mentioned on several occasions (by this author as well as by current, former, and potential users) that while TAMU GeoServices provides exceptionally useful services, one typically needs a PhD in order to be able to use them.

This is the challenge faced by many GIScience researchers who seek to advance the state-of-the-art in a particular research domain: We are a highly skilled group of researchers making important contributions towards the next generation of tools and techniques that will help change the world; but few of us are trained in what a good interface is, does, or looks like, many times resulting in the comical case of 127 output parameters for a service that simply produces a latitude/longitude pair from an input postal address.

TOWARDS A RE-SOLULTION

As mentioned previously, this situation is not new and is not unique to GIS, CyberGIS, or any flavor of geographic, geospatial, spatio-temporal, or spatial sciences (yes, for those in HCI who are unaware, our community has an identity crisis). The field of Human Computer Interaction (HCI) has been developing methods, techniques, and tools to help address these problems since the earliest inceptions of user interfaces, and even before that if we consider the development and application intelligent design principles in general.

The broad and rich set of theories and methods for evaluating and designing interfaces provided by the HCI community can and must be applied to the development and presentation of GIS and CyberGIS services if we are to avoid the analysis paralysis that users far too often end up suffering from in situations like the TAMU GeoServices example. Many of us GIScientists who develop and attempt to offer our technology to the broader scientific community know we are guilty as charged and have committed horrific HCI offenses, but we are actively seeking to build bridges and right our wrongs.

HCI needs to become an integral part of the standard GIS curriculum to a greater extent than currently observed in most programs. The GIS Body of Knowledge (BoK) [2], one of the primary documents that describes the knowledge a GIS professional should have, does include user interface design in several key areas (Topic CV4-3 Dynamic and interactive displays; Topic CV4-5 Web mapping and visualizations; Topic CV6-5 Evaluation and testing; Topic DA6-2 User interfaces).

However, the links between HCI and GIS training could and should be stronger by the inclusion of in-depth case studies and other resources highlighting good and bad designs for those GIScientists who seek to be GIS developers. Conferences and workshops that draw members from each of these communities are desperately needed in order to identify major challenges and start plotting a path towards addressing them in the short and long terms. GeoHCI offers such a starting place and will undoubtedly result in the production of new knowledge, research agendas, and the application of time-tested techniques to solving problems in new domains.

REFERENCES

1. Goldberg, D.W. Texas A&M GeoServices. https://geoservices.tamu.edu.

2. DiBiase, D., DeMers, M., Johnson, A. B., Kemp, K. K., Plewe, B. P and Wentz, E. A., eds. 2006. *The Geographic Information Science and Technology Body of Knowledge*, Washington, DC: Association of American Geographers.